Università L. Bocconi

Dipartimento di Scienze delle Decisioni
Ph.D. in Statistics, XX Ciclo

Ph.D. Thesis

Contributions to the Dirichlet process and related classes of random probability measures

Stefano Favaro
Matricola: 20057

Supervisor

Prof. Stephen G. Walker

January, 2009

Author's e-mail: stefano.favaro@unito.it

Author's address:

Dipartimento di Statistica e Matematica Applicata
Università di Torino
Corso Unione Sovietica 218/bis
10134 Torino, Italia

UNIVERSITÀ L. BOCCONI

DIPARTIMENTO DI SCIENZE DELLE DECISIONI
PH.D. IN STATISTICS, XX Ciclo

The thesis *"Contributions to the Dirichlet process and related classes of random probability measures"* by Stefano Favaro is recommended for acceptance by the members of the delegated committee, as stated by the enclosed reports, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

DATE: January 2009

Interlan examiner: Prof. Donato Michele Cifarelli

Internal examiner: Prof. Eugenio Melilli

External examiner: Prof. Stephen G. Walker

UNIVERSITÀ L. BOCCONI

DIPARTIMENTO DI SCIENZE DELLE DECISIONI
PH.D. IN STATISTICS, XX Ciclo

DATE: January 2009

Signature

To Tranquillo, Ada, Lino, Giuseppina, Vittorio, Sandra ed Elena

Acknowledgments

It is a great pleasure for me to thank the many people who have made this Ph.D. thesis possible.

It is difficult to overstate my gratitude to my Ph.D. supervisor, Professor Stephen G. Walker. With his enthusiasm, his inspiration, and his attitude to explain things clearly and simply, he helped to make mathematics fun for me. Throughout the last two years I spent at the Institute of Mathematics, Statistics and Actuarial Sciences at University of Kent, he provided encouragement, sound advice, good teaching, good company, and lots of good ideas. I would have been lost without him.

I am indebted to Professor Pietro Muliere for making possible these four important years of Ph.D. in Statistics at the Istituto di Metodi Quantitativi at Università "L. Bocconi" in Milano. In particular, I would like to thank him for disclosing me, with his rare and admirable enthusiasm, the foundation of Bayesian theory.

My gratitude goes also to Professor Alessandra Guglielmi, Professor Antonio Lijoi and Professor Igor Prünster, co-supervisors of this Ph.D. thesis, for the unique and valuable opportunity to collaborate with them and for their uncountable suggestions and helpful advice in the last two years. Moreover, I want to thank other people I had the pleasure to work with: Ramsés H. Mena, Matteo Ruggiero and Dario Spanò.

Lastly, but not less importantly, I wish to conclude these academic acknowledgments by expressing my deep gratitude to him who is known to many of those I have worked as 'My Professor', Professor Stefano F. Tonnellato. Thank you for your genuine friendship, which has always guided me towards the wisest choice.

Contents

Introduction

In Bayesian nonparametrics the role of the parameter appearing in a statistical model is taken by a probability distribution; therefore, the parameter space becomes a class of probability distributions defined on a given sample space. A Bayesian usually considers the class of all probability measures on the sample space and defines a probability on this class, the so-called prior distribution.

A precise description of a Bayesian nonparametric problem, the first to our knowledge, appears in de Finetti [68] in a paper about the problem of fitting a smooth curve onto an empirical distribution. At that time workable prior distributions on spaces whose elements are probabilities were not known. For these we have to wait until the 60' when a few schemes were put forward. Freedman [72] and Fabius [48] introduced a class of random probability measures termed tailfree. Unfortunately, these papers were not directly aimed at Bayesian nonparametric analysis so that their importance with respect to this topic was not immediately recognized. Rolph [168] suggested a scheme for constructing random probability distributions based on moments evaluation. A definite impetus to nonparametric inference within the Bayesian approach to statistics came eventually with the papers by Ferguson [61] (see also Ferguson [62]) and Doksum [25] based on the cited papers of Fabius and Freedman as well as on the paper by Dubins and Freedman [35]. In these papers a particular tailfree random probability measures, termed Dirichlet process, is presented. In particular, the Dirichlet process is characterized by the double advantage of having a large support, with respect to a suitable topology on the space of probability measures on the sample space, and of being analytically manageable for Bayesian posterior computations.

Since the introduction of the Dirichlet process, the literature on Bayesian nonparametrics has grown enormously and need for solutions of new problems has caused the introduction of new prior distribution.

I.1 Exchangeability and de Finetti's representation theorem

The notion of exchangeability, due to Bruno de Finetti, represents a cornerstone of Bayesian statistics. Indeed, the power of exchangeability derives from the celebrated representation theorem for sequences of exchangeable observations, provided by de Finetti [67] (see also de Finetti [69]). At this point it seems appropriate to give a brief description of these concepts and results, summarizing parts of Regazzini [164], key references for foundation issues concerning the Bayesian paradigm.

Suppose the observations take values in a complete and separable metric space, i.e. a Polish space, to be denoted by $\mathbb{X}$ with corresponding Borel σ-field $\mathscr{X}$. Let us consider the space $(\mathbb{X}^\infty, \mathscr{X}^\infty)$ where $\mathbb{X}^\infty$ indicates the set of all sequence $\{x_i, i \geq 1\}$ of elements of $\mathbb{X}$ whereas $\mathscr{X}^\infty$ is the σ-field generated by the subset of $\mathbb{X}^\infty$ of the type $A_1 \otimes \cdots \otimes A_n \otimes \mathbb{X}^\infty = \{x \in \mathbb{X}^\infty : x_1 \in A_1, \ldots, x_n \in A_n\}$ with $n \geq 1$ and $A_1, \ldots, A_n \in \mathscr{X}$. By assuming that the observation process is extendible to infinity, each observation X_i can be viewed as a measurable function from $\mathbb{X}^\infty$ into $\mathbb{X}$ according to the definition $X_i(x) = x_i$ for $i \geq 1$ and for every $x = \{x_j, j \geq 1\} \in \mathbb{X}^\infty$. According to the Ionesco-Tulcea theorem it is known that for all $p \in \mathcal{P}_{\mathbb{X}}$, there exists a unique probability measure p^∞ on $(\mathbb{X}^\infty, \mathscr{X}^\infty)$ such that

$$p^\infty(A_1 \otimes \cdots \otimes A_n \otimes \mathbb{X}^\infty) = \prod_{i=1}^n p(A_i)$$

for any $n \in \mathbb{N}$ and $A_1, \ldots, A_n \in \mathscr{X}$. In particular, a probability measure τ on $(\mathbb{X}^\infty, \mathscr{X}^\infty)$ makes the X_i exchangeable if the distribution of $\{X_{\pi(i)}, i \geq 1\}$ is the same as the distribution of $\{X_i, i \geq 1\}$ for any finite permutation π. In the present framework, it is well-known that the set $\mathcal{P}_{\mathbb{X}}$ of all probability measures on $(\mathbb{X}, \mathscr{X})$ is also a Polish space if endowed with a metric ρ, termed Prokhorov metric, which metrizes the topology of weak convergence of probability measures (see Prokhorov [158]). In particular, let $\mathscr{P}_{\mathbb{X}}$ be the σ-field on $\mathcal{P}_{\mathbb{X}}$ generated by the weak convergence topology. Any random element from a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ into $(\mathcal{P}_{\mathbb{X}}, \mathscr{P}_{\mathbb{X}})$ is said to be a random probability measure on $\mathbb{X}$.

The starting point for de Finetti's representation of an exchangeable law is the statement of the existence of a random probability measure P on $\mathbb{X}$ which is the limit (in distribution) of the empirical distribution of $X_1, \ldots, X_n$ an n tends to infinity. Denote by δ_x the probability measure with unit mass at $x \in \mathbb{X}$. Then the empirical distribution

$$e_n(X_1, \ldots, X_n) := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

is an example of random probability measure, and the following proposition makes the previous remarks precise:

"If $\{X_n, n \geq 1\}$ is exchangeable with respect to τ, then there is a random probability measure P on $\mathbb{X}$ such that

$$\lim_{n \to +\infty} \rho(e_n(X_1, \ldots, X_n), P) = 0 \qquad a.s.\text{-}\mathbb{P} \tag{I.1.1}$$

where ρ is the Prokhorov metric."

Furthermore, under the hypothesis that p is a realization of P, i.e. p is the "real" distribution of the characteristic under observation, one expects, since the observations are collected under analogous environmental conditions, that the X_i's are independent and identically distributed with common distribution p. This fact is the essence of the following version of de Finetti's representation theorem:

"If $\{X_n, n \geq 1\}$ in exchangeable with respect to τ, then for any $B \in \mathscr{X}^\infty$

$$\tau(X_1, X_n, \ldots \in B|P) = P^\infty(B) \qquad a.s.\text{-}\mathbb{P} \tag{I.1.2}$$

where given any probability measure μ on $(\mathbb{X}, \mathscr{X})$, μ^∞ denotes the distribution of $\{X_n, n \geq 1\}$ when the X_n's are supposed to be independent and identically distributed with common distribution μ. Conversely, if there is a random probability measure P such that(5.1.4) *holds, then the X_n's are exchangeable".*

The following statement represents another equivalent version of de Finetti's representation theorem:

"$\{X_n, n \geq 1\}$ is exchangeable if and only if there is a probability measure Q on $(\mathcal{P}_\mathbb{X}, \mathscr{P}_\mathbb{X})$ such that for any $B \in \mathscr{X}^\infty$

$$\tau(X_1, X_n, \ldots \in B) = \int_{\mathcal{P}_\mathbb{X}} P^\infty(B) Q(dP) \tag{I.1.3}$$

where Q, the so-called de Finetti's measure of $\{X_n, n \geq 1\}$, is uniquely determined and coincides with the distribution of P."

The Bayesian paradigm can be introduced by considering a process of observations $\{X_n, n \geq 1\}$ taking values on some space $\mathbb{X}$ and by assuming that the observations are taken under homogeneous physcal conditions, i.e. the order in which the observations are detected is not relevant for the inferential purpose. Therefore, given a sample $X_1, \ldots, X_n$ of size n taking values in $\mathbb{X}$, these are assumed to be realizations of n random variables belonging to an infinite exchangeable sequence $\{X_n, n \geq 1\}$ of random variables with values in $\mathbb{X}$. Exchangeability and de Finetti's representation theorem imply the existence of a random element $P \in \mathcal{P}_\mathbb{X}$ such that, conditionally on $P = p$, the random variables $\{X_n, n \geq 1\}$ are independent and identically distributed with probability distribution p. Initial opinion about the sequence $\{X_n, n \geq 1\}$ are expressed through the probability distribution Q of P and updated by means of Bayes theorem; all inferential questions related to the sequence $\{X_n, n \geq 1\}$ are the answered on the basis on the conditional probability distribution of P, given the observed sample from $\{X_n, n \geq 1\}$. Since we are concerned with an inferential problem whose basic ingredients are an infinite exchangeable sequence $\{X_n, n \geq 1\}$ of random variables with values in $\mathbb{X}$ and a random probabilily measure P, the natural space where to embed it is the product space $(\mathbb{X}^\infty \times \mathcal{P}_\mathbb{X}, \mathscr{X}^\infty \otimes \mathscr{P}_\mathbb{X})$ where $\mathbb{X}^\infty \times \mathcal{P}_\mathbb{X} = \{(x,p) : x \in \mathbb{X}^\infty, p \in \mathcal{P}_\mathbb{X}\}$ while $\mathscr{X}^\infty \otimes \mathscr{P}_\mathbb{X}$ is the smallest σ-field of subsets of $\mathbb{X}^\infty \times \mathcal{P}_\mathbb{X}$ containing the family of measurable rectangle $\{A \otimes B : A \in \mathscr{X}^\infty, B \in \mathscr{P}_\mathbb{X}\}$. The probability distribution Q of P, often called the prior probability distribution, plays a pivotal role for the definition of a probability measure on $(\mathbb{X}^\infty \times \mathcal{P}_\mathbb{X}, \mathscr{X}^\infty \otimes \mathscr{P}_\mathbb{X})$. In fact, given a probability measure Q defined on the space $(\mathcal{P}_\mathbb{X}, \mathscr{P}_\mathbb{X})$, we set

$$\pi(C) = \int_{\mathcal{P}_\mathbb{X}} p^\infty(\{x \in \mathbb{X}^\infty : (x,p) \in C\})Q(dp) \tag{I.1.4}$$

for all $C \in \mathbb{X}^\infty \times \mathcal{P}_\mathbb{X}$. Note that, when $C = (A_1 \otimes \cdots \otimes A_n \otimes \mathbb{X}^\infty) \times B$ for any $n \in \mathbb{N}$, $A_i \in \mathscr{X}$ for $i = 1, \ldots, n$ and $B \in \mathcal{P}_\mathbb{X}$, then

$$\pi(C) = \int_B \prod_{i=1}^n p(A_i)Q(dp)$$

It can be checked that π is probability measure on $(\mathbb{X}^\infty \times \mathcal{P}_\mathbb{X}, \mathscr{X}^\infty \otimes \mathscr{P}_\mathbb{X})$. Therefore, given a probability measure Q on $(\mathcal{P}_\mathbb{X}, \mathscr{P}_\mathbb{X})$, we call statistical space (or statistical model) the triple $(\mathbb{X}^\infty \times \mathcal{P}_\mathbb{X}, \mathscr{X}^\infty \otimes \mathscr{P}_\mathbb{X}, \pi)$ where π is the probability measure defined as in (5.1.2).

Once given the statistical space it is possible to tackling any inferential problem: the hypotetical inference if the aim is to infer on the random element $P \in \mathcal{P}_\mathbb{X}$ and the predictive inference if the aim is to infer on the random element $X \in \mathbb{X}$ given an observed sample $X_1, \ldots, X_n$ from $\{X_n, n \geq 1\}$. Since $(\mathbb{X}^\infty, \mathcal{P}_\mathbb{X})$ is a Polish space, for any $n \in \mathbb{N}$ and $(x_1, \ldots, x_n) \in \mathbb{X}^n$ there exists a regular conditional probability q on $(\mathscr{X}^\infty \otimes \mathscr{P}_\mathbb{X})$ given $X_1 = x_1, \ldots, X_n = x_n$. This means that for any $n \in \mathbb{N}$ there exists a function $q : \mathbb{X}^n \otimes (\mathscr{X}^\infty \otimes \mathscr{P}_\mathbb{X}) \to [0,1]$ such that

i) for all $(x_1, \ldots, x_n) \in \mathbb{X}^n$, $q((x_1, \ldots, x_n), \cdot)$ is a probability on $(\mathbb{X}^\infty \times \mathcal{P}_\mathbb{X}, \mathscr{X}^\infty \otimes \mathscr{P}_\mathbb{X})$;

ii) for all $C \in \mathscr{X}^\infty \otimes \mathscr{P}_\mathbb{X}$, $q(\cdot, C)$ is a measurable map from $(\mathbb{X}^n, \mathscr{X}^n)$ to $([0,1], \mathscr{R} \cap [0,1])$

iii) for all $A \in \mathscr{X}^n$, $C \in \mathscr{X}^\infty \otimes \mathscr{P}_\mathbb{X}$

$$\pi(C \cap (X_1, \ldots, X_n)^{-1}(A)) = \int_A q((x_1, \ldots, x_n), C)\tau_n(dx_1, \ldots, dx_n)$$

where τ_n is the probability distribution induced on $(\mathbb{X}^n, \mathscr{X}^n)$ by the random vector $(X_1, \ldots, X_n)$

Given $n \in \mathbb{N}$ and $(x_1, \ldots, x_n) \in \mathbb{X}^n$, the probability $Q(\cdot|X_1 = x_1, \ldots, X_n = x_n)$ on $(\mathcal{P}_\mathbb{X}, \mathscr{P}_\mathbb{X})$, defined by setting

$$Q(B|X_1 = x_1, \ldots, X_n = x_n) = q((x_1, \ldots, x_n), \mathbb{X}^\infty \times B)\pi(\mathbb{X}^\infty \times B|X_1 = x_1, \ldots, X_n = x_n)$$

for any $B \in \mathcal{P}_{\mathbb{X}}$, is called posterior probability distribution of P: this is the conditional distribution of P given $X_1 = x_1, \dots, X_n = x_n$. Analogously, given $m, n \in \mathbb{N}$ and $(x_1, \dots, x_n) \in \mathbb{X}^n$, the probability $\tau_m(\cdot | X_1 = x_1, \dots, X_n = x_n)$ on $(\mathbb{X}^m, \mathscr{X}^m)$ defined by setting for any $A \in \mathscr{X}^m$

$$\tau_m(A | X_1 = x_1, \dots, X_n = x_n) = q((x_1, \dots, x_n), (A \otimes \mathbb{X}^\infty) \otimes \mathcal{P}_{\mathbb{X}})$$

is called predictive probability distribution of $X_{n+1}, \dots, X_{n+m}$: this is the conditional probability distribution of $X_{n+1}, \dots, X_{n+m}$ given the observed sample $X_1 = x_1, \dots, X_n = x_n$.

So far we introduced the Bayesian paradigm and we highlighted how in a Bayesian nonparametric inferential problem the role of the parameter appearing in the statistical model is taken by a probability distribution. Therefore, the parameter space of the statistical model is represented by the class of all probability measures on a given sample space and the problem of placing a probability distribution on this class, the so-called prior distribution, becames crucial. In light of these considerations, it becomes evident that the notion of exchangeability combined with de Finetti's representation theorem cannot be reduced to a mere synonym of conditional independence and identity in distribution; indeed, it provides a coherent subjective solution to the inferential problem, which includes in a unifying way both the "parametric" and the "nonparametric" case.

I.2 Prior distribution on spaces of probability measures

As far as the problem of assigning prior distribution on spaces of probability measures is concerned, it is convenient to start with the case of a finite $\mathbb{X}$, i.e. $\mathbb{X} := \{a_1, \dots, a_k\}$, k being a fixed element of $\mathbb{N}$. Hence, the law of any random sequence $\{X_n, n \geq 1\}$ with values in $(\mathbb{X}^\infty, \mathscr{X}^\infty)$ is completely characterized by its restriction to all events $\{X_1 = x_1, \dots, X_n = x_n\}$ defined for all $(x_1, \dots, x_n)$ in $\mathbb{X}^n$ and for all $n \in \mathbb{N}$. In particular, if $\{X_n, n \geq 1\}$ is exchangeable with respect to τ, then by (I.1.3), there exists one and only one probability distribution function Q such that

$$\tau(X_1 = x_1, \dots, X_n = x_n) = \int_{\Delta^{(k-1)}} \prod_{i=1}^{k-1} v_i^{n(a_i)} \left(1 - \sum_{i=1}^{k-1} v_i\right)^{n(a_k)} Q(dv_1, \dots, dv_{k-1})$$

with support included in $\Delta^{(k-1)} := \{(v_1, \dots, v_{k-1}) : v_i \geq 0, i = 1, \dots, k-1, \sum_{1 \leq i \leq k-1} v_i \leq 1\}$ holds for any $(x_1, \dots, x_n) \in \mathbb{X}^n$ and for any $n \in \mathbb{N}$ with $n(a_i) = \sum_{1 \leq j \leq n} \delta_{a_i}(x_j)$ for $i = 1, \dots, k$, $\sum_{1 \leq i \leq k} n(a_i)$ and $v^0 := 1$ if $v = 0$. Obviously there is a one-to-one correspondence between the set of all probability measures on $(\mathcal{P}_{\mathbb{X}}, \mathscr{P}_{\mathbb{X}})$ and the set of all distribution functions Q with support included in $\Delta^{(k-1)}$. Thus, one immediately sees that the representation theorem provides a coherent framework for the usual implementation of the Bayesian paradigm for the multinomial model. In this case the de Finetti measure Q is the so called prior distibution.

A statistician, due to convinction or lack of information, is often not able to commit himself to assuming finite $\mathbb{X}$. If this is the case, assessing Q can become cumbersome. Indeed, if a quite deep knowledge of the random phenomenon under consideration is available, the statistician may be able to isolate the randomness (e.g. through a functional relation) and, consequently, to restrict the class of admissible laws. The concrete implementation of statistical methods greatly benefits from such restrictions which usually, but not necessarily, are reflected upon the assumption that de Finetti's measures are supported by parametric families of probabilities and parameters belong a subset of some finite dimensional Eucledian space. If one set $\mathcal{P}'_{\mathbb{X}} = \{P_\theta : \theta \in \Theta\}$ and assigns a de Finetti measure such that $Q(\mathcal{P}'_{\mathbb{X}}) = 1$, there exists a bijection $f : \Theta \to \mathcal{P}'_{\mathbb{X}}$ such that Q induces a probability measure q on Θ by means of $q(B) = Q(f(B))$ for every $B \in \Theta$. The general question to be answered by the statistician is wheather, on the basis of the available information, it is coherent to reduce the class of all formally admissible distributions on a given sample space to some distiguished parametric family. From a concrete point of view, parametric structures would stem from distinguished assumptions formulated about the probabilistic modelling of the random phenomenon under study. The reader is referred to, e.g, Bernardo and Smith [6] and Shervish [175] and reference therein.

In cases, in which the statistician is in a position to assume neither a finite $\mathbb{X}$ or a restriction of the class of admissible distributions, a so called nonparametric approach has to be undertaken. Such an approach essentially corresponds to the impossibility of a commitment about a probabilistic model underlying the phenomenon at issue. Thus, it should be clear that the term nonparametric, commonly used to designate it, is vacuous since it does not reflect the non-informativeness which leads to adopt this very same approach. Nonetheless, it will be sometimes employed in the sequel, however bearing in mind its correct interpretation. In this case, as far as the problem of assessing a de Finetti measure is concerned, it seems natural to regard it as a problem of extending the multinomial model. The prominent role, played by the Dirichlet distribution in the multinomial case, is held by an appropriately defined functional Dirichlet distribution or Dirichlet process prior. This is due to the fact that the relevan advantages regarding the mathematical tractability of the Dirichlet distribution, in particular, the solution it provides to the problem of prevision and its reproducibility property in passing to the posterior distribution, carry over to the functional Dirichlet distribution. Other de Finetti measure are attempts to generalize the Dirichlet process prior in various directions.

It has to be said immediately that the major drawback inherent to a Bayesian nonparametric approach is represented by the analytica formulation of priors. From a conceptual point of view, the assumption of exchangeability makes the assessment of a particular Q simpler, since it can be relied upon (I.1.1)-(I.1.3). Indeed, one can always go back to the case of exchangeable observations, according to the following line of reasoning. Any Q on $(\mathcal{P}_{\mathbb{X}}, \mathscr{P}_{\mathbb{X}})$ can be thought of as the de Finetti measure of an exchangeable τ on $(\mathbb{X}^{\infty}, \mathscr{X}^{\infty})$, according to (I.1.3). Hence, in order to specify the analytica form of Q, it is convenient to think of Q as the probability law of the unknown distribution of a characteristic, with values in $\mathbb{X}$, of the elements of a population that is sampled with replacement. So, Q will be assessed as if it were the limiting distribution of the empirical distribution connected with the above fictitious sequence of trials.

I.3 Outline of the thesis

On the basis of the previous paragraphs, one can state that random probability measures represent the essence of Bayesian nonparametrics both from a conceptual and practical perspective. In particular, they are the starting point of the present treatment which aims at providing some contributions to the Dirichlet process and to some related classes of random probability measures obtained by transformations of increasing additive processes.

Parts of this thesis are based upon Favaro and Walker [50], Favaro et al. [55], Favaro et al. [52], Favaro et al. [53], Favaro and Walker [51], Favaro et al. [54], Favaro et al. [56] and Favaro and Walker [57]. The first paper represents the first seminal work: there, a new distributional equation having as unique solution the Dirichlet process is proved and it is used to provide an alternative series representation of the Dirichlet process. The second paper, moving from the new distributional equation introduced in Favaro and Walker [50], defines and investigates a new measure-valued Markov chain which generalizes the well known Feigin-Tweedie Markov chain widely used to provide properties of linear functionals of the Dirichlet process and approximation procedures for estimating the law of the mean of the Dirichlet process. The third and the fourth paper consider the new distributional equation introduced in Favaro and Walker [50] in order to provide two characterization of a Fleming-Viot processes which is a wide class of measure-valued diffusion processes arising as large population limits of so-called particle processes. The fifth paper represents the second seminal work: there, some developments for a class of random probability measures, termed generalized Dirichlet processes, which induces exchangeable sequences which are characterized by a more elaborated predictive structure that those one arising from Gibbs-type random probability measures. The sixth paper investigates in details the posterior behavior of the generalized Dirichlet process and it provides conditional distributions and the corresponding Bayesian nonparametric estimators derived from the random partition structure characterized by the generalized Dirichlet process. The last one considers some developments on the Bayesian nonparametric inference for species sampling problems based on the two parameter Poisson-Dirichlet process. The dissertation is structured in two parts.

The first part of the dissertation deals with the Dirichlet process and with some classes of measure-valued random processes having as unique invariant measure the law of the Dirichlet process. Given a general Polish space $\mathbb{X}$, we start by considering a new distributional equation defined on a the set of all probability measures on $(\mathbb{X}, \mathscr{X})$ and having as unique solution the Dirichlet process on $\mathbb{X}$; some applications of this new distributional equation are then considered in order to provide new characterizations of the Dirichlet process. In particular, a generalization of the constructive definition of the Dirichlet process proposed by Sethuraman is derived and investigated with emphasis on its application to the Dirichlet process approximations. Moreover, the new distributional equation for the Dirichlet process is considered in order to provide some contributions on measure-valued random processes having as unique invariant measure the law of the Dirichlet process: the first contribution regards the definition and the study of a measure-valued Markov chain which generalizes the well known Feigin-Tweedie Markov chain and which still has the law of the Dirichlet process as unique invariant measure; the second contribution regards a characterization of the so-called particle process driving a wide class of measure-valued diffusion processes termed Fleming-Viot processes and having as unique invariant measure the law of the Dirichlet process.

The second part of the dissertation deals with random probability measures derived by normalization of increasing additive process and in particular with the class of normalized random measures with independent increments (NRMIs). The subclass on NRMIs with logarithmic singularity is defined and an appealing example, termed generalized Dirichlet process; by considering its characterization in terms of normalized superposition of independent Gamma processes, the generalized Dirichlet process is deeply investigated and a comprehensive treatment in terms of finite dimensional distributions, moments, predictive distributions and posterior distributions. In particular, such process induces exchangeable random partitions which are characterized by a more elaborate clustering structure than those arising from Gibbs-type random probability measures. A natural area of application of these random probability measures is represented by species sampling problems and, in particular, prediction problems in genomics. To this end we study both the distribution of the number of distinct species present in a sample and the distribution of the number of new species conditionally on an observed sample. We also provide the nonparametric Bayesian estimator, under quadratic loss, for the number of new species in an additional sample of given size and for the discovery probability as function of the size of the additional sample. Some developments on the conditional distributions and the corresponding

Bayesian nonparametric estimators recently obtainied for the class of Gibbs-type random probability measures are also provided.

A brief review of the main results to be presented seems appropriate at this point.

In Chapter 1, after introducing the definition and the main properties of the Dirichlet process, we define a generalization of the constructive definition of the Dirichlet process proposed by Sethuraman. In particular, the new constructive definition for the Dirichlet process is obtained from the original Sethuraman's construction nesting an appropriate random convex linear combination of random variables from a Blackwell-MacQueen Pólya sequence. Properties of the new definition of the Dirichlet process are studied.

In Chapter 2 we provide some remarks on the approximation of the Dirichlet process. Moving from a series representation for the Dirichlet process which includes the Sethuraman series representation as particular case, we consider the application of two stopping procedures: a random stopping procedure which corresponds to the truncation of the series at a random number of terms and the almost sure stopping procedure which corresponds to the truncation of the series at a fixed number ot terms. In particular, the accuracy in these approximations is given with respect to the corresponding approximations obtained using the Sethuraman series representation of the Dirichlet process. A straightforward extension of the random stopping procedure to the more general class of infinite dimensional stick-breaking random measures is considered. As a by-product, we also obtain some interesting results related to the convolution of distributions belonging to the class of generalized convolutions of mixtures of Exponential distributions.

In Chapter 3, moving from a distributional equation having as unique solution the Dirichlet process, we define and we investigate a new measure-valued Markov chain having as unique invariant measure the law of a Dirichlet process. This Markov chain generalizes the well known Feigin-Tweedie Markov chain which has been widely used to provide properties of linear functionals of the Dirichlet process and approximation procedures for estimating the law of the mean of the Dirichlet process. Our main aim in this chapter is to show that the Feigin-Tweedie chain sits in a large class of chains indexed by an integer $n \in \mathbb{N}$ and they worked solely on the case $n = 1$, where n can be viewed as a sample size. We provide properties of this new class of Markov chain.

In Chapter 4 we consider a different class of random processes having as unique invariant measure the law of a Dirichlet process. This class of random process is termed Fleming-Viot processes and it is a wide class of measure-valued diffusion processes arising as large population limits of so-called particle processes. Here we invert the procedure and show that a countable population process can be derived directly from the neutral diffusion model, with no arbitrary assumptions. We study the atomic structure of the neutral diffusion model, and elicit a finite dimensional particle process from the time-dependent random measure, for any chosen population size. The static properties are consequences of the fact that its stationary distribution is the Dirichlet process, and rely on a new representation for it. The dynamics are derived directly from the transition function of the neutral diffusion model. As by-product we also obtain a new constructive definition of the Dirichlet process.

In Chapter 5 we consider priors obtained by normalizing random measures with independent increments (NRMIs) we define a new class on NRMIs, the so-called NRMI with logarithmic singularity. The class of NRMIs with logarithmic singularity includes as particular case the celebrated Dirichlet process and on the other hand it does not include the normalized generalized Gamma process recently introduced in the context of mixture models and species sampling problems. In particular, we are interested in a generalization of the Dirichlet process which has been recently introduced in the literature and which is in the class of NRMIs with logarithmic singularity, the so-called generalized Dirichlet process. Some developments of the generalized Dirichlet process are presented in terms of its finite dimensional distributions, moments, predictive distributions and posterior distributions.

In Chapter 6 we investigate a class of random probability measures, termed generalized Dirichlet processes, which has been recently introduced in the literature and further investigated in Chapter 5. Such processes induce exchangeable random partitions which are characterized by a more elaborate clustering structure than those arising from Gibbs-type random probability measures. A natural area of application of these random probability measures is represented by species sampling problems and, in particular, prediction problems in genomics. To this end we study both the distribution of the number of distinct species present in a sample and the distribution of the number of new species conditionally on an observed sample. Some developments on the conditional distributions and the corresponding Bayesian nonparametric estimators recently obtainied for the class of Gibbs-type random probability measures are also provided.

I

Contributions to the Dirichlet process and some measure-valued random processes

1

A generalized constructive definition for Dirichlet processes

In this chapter we introduce a generalization of the constructive definition of the Dirichlet process proposed by Sethuraman. The new constructive definition for the Dirichlet process is obtained from the original Sethuraman's construction nesting an appropriate random convex linear combination of random variables from a Blackwell-MacQueen Pólya sequence. Properties of the new definition of the Dirichlet process are studied.

1.1 Introduction

Let us consider a topology $\mathcal{T}$ such that $(\mathbb{X}, \mathcal{T})$ is a Polish space endowed with the Borel σ-field $\mathscr{X}$ and let $\mathcal{P}_{\mathbb{X}}$ be the space of probability measures on $(\mathbb{X}, \mathscr{X})$ endowed with its σ-field $\mathscr{P}_{\mathbb{X}}$ generated by the weak convergence topology[1] $\mathcal{W}$ which makes $(\mathcal{P}_{\mathbb{X}}, \mathcal{W})$ a Polish space. The first example of probability measure on $(\mathcal{P}_{\mathbb{X}}, \mathscr{P}_{\mathbb{X}})$ useful as prior distribution for Bayesian nonparametrics is the probability measure induced by the celebrated Dirichlet process whose characterization and essential properties were extensively presented by Ferguson [61] and Ferguson [62] and further investigated by Blackwell [9] and Blackwell and MacQueen [10]. Various authors have considered other characterizations and properties of the Dirichlet process. In particular, in this chapter we are interested in the constructive definition of the Dirichlet process proposed by Sethuraman [174] and here recalled.

Let α be a finite measure with total mass a and define $\alpha_0 := \alpha/a$. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space supporting two independent sequence of r.v.s $\theta := \{\theta_i, i \geq 1\}$ and $Y := \{Y_i, i \geq 1\}$ such that the application $(\theta, Y) : \Omega \to ([0,1] \times \mathbb{X})^\infty$ is $(\mathscr{F}, (\mathscr{R} \cap [0,1] \otimes \mathscr{X})^\infty)$-measurable. In particular, the sequence θ is a sequence of independent r.v.s identically distributed a ccording to a Beta distribution function with parameter $(1, a)$ for $i \geq 1$

[1] The weak convergence topology is the topology induced by the measurable projection mappings $\wp_B : \mathcal{P}_{\mathbb{X}} \to [0,1]$, with $B \in \mathscr{X}$, defined for all $p \in \mathcal{P}_{\mathbb{X}}$ by $\wp_B(P) := P(B)$.

and the sequence Y is a sequence of independent r.v.s identically distributed according to α_0. The condition of independence between the sequence of r.v.s θ and Y and the usual construction of a product measure implies the existence of the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ supporting the r.v. (θ, Y) and does not require any restrictions on $\mathbb{X}$, such as it being a Polish space. We now consider the sequence of r.v.s $\{p_i, i \geq 1\}$ obtained from the sequence θ, by the so-called stick-breaking construction

$$p_1 = \theta_1$$

$$p_i = \theta_i \prod_{j=1}^{i-1}(1-\theta_j) \qquad i \geq 2$$

$p_1 = \theta_1$ and The stick-breaking construction implies that as $n \to +\infty$, $\sum_{1\leq i\leq n} p_i = 1 - \prod_{1\leq i\leq n}(1-\theta_i) \to 1$ a.s. For any $B \in \mathbb{X}$ let $P : \Omega \to \mathcal{P}_{\mathbb{X}}$ be the map defined by

$$\omega \mapsto \sum_{i\geq 1} p_i(\omega)\delta_{Y_i(\omega)}(B) \tag{1.1.1}$$

which is clearly a measurable map and takes values in the subset of discrete probability measures on $(\mathbb{X}, \mathscr{X})$. In particular, Sethuraman [174] proved that for any finite measurable partition $B_1, \ldots, B_k$ of $\mathbb{X}$, $(P(\cdot, B_1), \ldots, P(\cdot, B_k))$ is distributed according to a Dirichlet distribution function with parameter $(\alpha(B_1), \ldots, \alpha(B_k))$. This establishes that P defined in (1.1.1) is a Dirichlet process with parameter α. A more direct way to describe the constructive definition (1.1.1) is as follows. Let $\{Y_i, i \geq 1\}$ be a sequence of r.v.s i.i.d. with common distribution α_0. Let $\{p_i, i \geq 1\}$ be a sequence of probabilities from a discrete distribution on the integers with discrete failure rate $\{\theta_i, i \geq 1\}$ which are i.i.d. distributed according to a Beta distribution function with parameter $(1, a)$. Let P be the r.p.m. that puts weights p_n at the r.p.m. δ_{Y_n} for $n \geq 1$.

In this chapter our main aim is to introduce a new constructive definition for the Dirichlet process on $\mathbb{X}$ with parameter α that includes as particular case the constructive definition proposed by Sethuraman. For any fixed integer-valued sequence $n_\bullet := \{n_i, i \geq 1\}$ we denote the new constructive definition of the Dirichlet process by $P^{(n_\bullet)}$ which is obtained from the original Sethuraman's construction nesting at each an appropriate random convex linear combination of r.v.s instead of the r.p.m. δ_{Y_i} for $i \geq 1$. We prove that $P^{(n_\bullet)}$ is still a r.p.m. on $\mathbb{X}$ which gives probability one to the subset of the discrete probabily measures on $(\mathbb{X}, \mathscr{X})$. Moreover, we will prove that for any finite measurable partition $B_1, \ldots, B_k$ of $\mathbb{X}$, the r.v. $(P^{(n_\bullet)}(\cdot, B_1), \ldots, P^{(n_\bullet)}(\cdot, B_k))$ is distributed according to a Dirichlet distribution function with parameter $(\alpha(B_1), \ldots, \alpha(B_k))$ for any finite measure α on $(\mathbb{X}, \mathscr{X})$. This establishes that $P^{(\tilde{n})}$ is a Dirichlet process with parameter α. Finally, following similar arguments to those used in Sethurman [174], we provide a posterior analysis based on the generalized series representation of the Dirichlet process.

The chapter is structured as follows. In Section 1.2, we recall the definition of the Dirichlet process in terms of the family of its finite dimensional distributions as originally defined by Ferguson [61] and we consider some of its properties. Section 1.3 contains the main results of the chapter; in particular we provide a new constructive definition of the Dirichlet process and we discuss some of its properties. In Section 1.5 we provide the posterior characterization of the Dirichlet process based on the new constructive definition of the Dirichlet process. Section 1.4 is devoted to a discussion of the results.

1.2 The Dirichlet process

In the introduction we recalled the constructive definition of the Dirichlet process introduced by Sethuraman [174]. In this section we recall the definition of the Dirichlet process via its family of finite dimensional distribution as originally defined by Ferguson [61]. Some of the main properties of the Dirichlet process are also recalled.

In order to define the Dirichlet process, let $\mathcal{A}_{\mathbb{X}}$ be the space of locally finite non-negative measures on $(\mathbb{X}, \mathscr{X})$ endowed with the σ-field $\mathscr{A}_{\mathbb{X}}$ generated by the vague topology[2] $\mathcal{V}$ which makes $(\mathcal{A}_{\mathbb{X}}, \mathcal{V})$ a Polish space. Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure with total mass $a := \alpha(\mathbb{X})$ and for any finite measurable partition $B_1, \ldots, B_k$ of $\mathbb{X}$ such that $\alpha(B_i) > 0$ for $i = 1, \ldots, k$ and $k \geq 2$ we introduced the Dirichlet distribution with parameter $(\alpha(B_1), \ldots, \alpha(B_k))$

$$\Pi_k(A; \alpha(B_1), \ldots, \alpha(B_k)) \tag{1.2.1}$$
$$:= \frac{\Gamma(a)}{\prod_{i=1}^k \Gamma(\alpha(B_i))} \int_{\tilde{A}\cap\Delta^{(k-1)}} \prod_{i=1}^{k-1} x_i^{\alpha(B_i)-1} \left(1 - \sum_{i=1}^{k-1} x_i\right)^{\alpha(B_k)-1} dx_1 \cdots dx_{k-1}$$

for any $A \in (\mathscr{R}\cap[0,1])^k$ where $\tilde{A} := \{(x_1, \ldots, x_{k-1}) : (x_1, \ldots, x_{k-1}, 1-\sum_{1\leq i\leq k-1} x_i) \in A\}$. Moreover, we set

$$\Pi_1(A; \mathbb{X}) = \delta_1(A) \qquad A \in \mathscr{R} \cap [0,1]$$

Let $A_1, \ldots, A_n$ be n distinct ordered elements in $\mathscr{X}$ and let $B_1, \ldots, B_k$ be the finite partition generated by $A_1, \ldots, A_n$. If $\alpha(B_i) > 0$ for $i = 1, \ldots, k$ we denote by $q_{A_1,\ldots,A_n}$ the probability measure on $((\mathbb{R} \cap [0,1])^n, \mathscr{R} \cap [0,1])^n$ defined by

$$\tilde{q}_{A_1,\ldots,A_n}(B) := \Pi_k\left(\left\{(x_1, \ldots, x_k) : \tilde{X} \in B\right\}; B_1, \ldots, B_k\right) \tag{1.2.2}$$

[2] The vague topology is the topology induced by the mappings $\pi_g : \mu \mapsto \int_{\mathbb{X}} g d\mu$ with f a continuous function $g : \mathbb{X} \to \mathbb{R}^+$.

where

$$\tilde{X} := \left(\sum_{i:B_i \subset A_1} x_i, \ldots, \sum_{i:B_i \subset A_n} x_i \right)$$

for any $B \in (\mathscr{R} \cap [0,1])^n$. If $\alpha(B_i) = 0$ for some i, then we consider $B_1^{'}, \ldots, B_{k'}^{'}$ obtained from the finite measurable partition $B_1, \ldots, B_k$ of $\mathbb{X}$ by removing the set B_i such that $\alpha(B_i) = 0$. Then we denote by $q_{A_1,\ldots,A_n}$ the probability measure on $((\mathbb{R} \cap [0,1])^n, \mathscr{R} \cap [0,1])^n$ defined by

$$\tilde{q}_{A_1,\ldots,A_n}(B) := \Pi_{k'} \left(\left\{ (x_1, \ldots, x_{k'}) : \tilde{X}^{'} \in B \right\}; B_1^{'}, \ldots, B_{k'}^{'} \right) \tag{1.2.3}$$

where

$$\tilde{X}^{'} := \left(\sum_{i:B_i^{'} \subset A_1} x_i, \ldots, \sum_{i:B_i^{'} \subset A_n} x_i \right)$$

for any $B \in (\mathscr{R} \cap [0,1])^n$, where $\sum_{i:C_i^{'} \subset A_j} x_i := 0$ if $\{i : B_i^{'} \subset A_j\} = \emptyset$ and $\Pi_1(A; B) = \delta_1(A)$ if $\alpha(B) = a$. We now define the following finite dimensional Kolmogorov spaces $([0,1]^n, (\mathscr{R} \cap [0,1])^n, \tilde{q}_{A_1,\ldots,A_n})$ and the family of finite dimensional probability distribution $\mathbb{Q} := \{\tilde{q}_{A_1,\ldots,A_n} : A_1, \ldots, A_n \in \mathscr{X}, n \in \mathbb{N}\}$. We can show that $\mathbb{Q}$ satisfies the following conditions:

C1) for any $n \in \mathbb{N}$ and any finite permutation σ of $\{1, \ldots, n\}$

$$\tilde{q}_{A_1,\ldots,A_n}(B) = \tilde{q}_{A_{\sigma(1)},\ldots,A_{\sigma(n)}}(\sigma B) \qquad \forall B \in (\mathscr{R} \cap [0,1])^n$$

where $\sigma B = \{(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) : (x_1, \ldots, x_n) \in B\}$;

C2) $\tilde{q}_{\mathbb{X}} = \delta_1$, where δ_x is the point mass at x;

C3) for any family of sets $A_1, \ldots, A_n$ in $\mathscr{X}$, let $D_1, \ldots, D_k$ be a finite measurable partition of $\mathbb{X}$ such that it is finer than the finite partition generated by $A_1, \ldots, A_n$. Then for any $B \in (\mathscr{R} \cap [0,1])^n$

$$\tilde{q}_{A_1,\ldots,A_n}(B) = \tilde{q}_{D_1,\ldots,D_k}(B^{'})$$

where

$$B^{'} = \left\{ (x_1, \ldots, x_k) \in [0,1]^k : \left(\sum_{i:D_i \subset A_1} x_i \ldots, \sum_{i:D_i \subset A_n} x_i \right) \in B \right\}$$

C4) for any sequence $\{A_n, n \geq 1\}$ of measurable subset of $\mathbb{X}$ such that $A_n \downarrow \emptyset$

$$\tilde{q}_{A_n} \Rightarrow \delta_0.$$

Hence, according to Proposition 3.9.2 of Regazzini [164] (see also Regazzini and Petris [162]), there exists a unique r.p.m. P on $\mathbb{X}$ admitting $\mathbb{Q}$ as its family of finite dimensional distribution. We term such a r.p.m. P a Ferguson-Dirichlet r.p.m. with parameter α (or simply a Dirichlet process with parameter α) since it was originally introduced by Ferguson [61]. Finally, we denote by Π the distribution of the Dirichlet process with parameter α.

We now consider some basic properties of the Dirichlet process with parameter α which make it an appealing prior distribution for Bayesian nonparametric statistics. In particular we consider the predictive distributions, the posterior process and the support. Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure with total mass a and for any sequence $\{x_n, n \geq 1\} \in \mathbb{X}^\infty$ we set

$$\mu_{n+1}^{(1,\ldots,n)}(x_1, \ldots, x_n; B) := \frac{\alpha(B) + \sum_{i=1}^n \delta_{x_i}(B)}{a+n} \qquad B \in \mathscr{X}. \tag{1.2.4}$$

For any $B \in \mathscr{X}$ we can rewrite the right-hand side of(1.2.4) as a convex linear combination of a probability measure and of an empirical distribution

$$\frac{a}{a+n}\mu_1(B) + \frac{1}{a+n}\sum_{i=1}^n \delta_{x_i}(B) \qquad B \in \mathscr{X}$$

where $\mu_1 := \alpha/a$. It can be easily checked that $\{\mu_1, \mu_2^{(1)}, \mu_3^{(1,2)}, \ldots\}$ is a sequence of probability measures on $(\mathbb{X}, \mathscr{X})$ and for any $B \in \mathscr{X}$ and for any $n \in \mathbb{N}$

$$(x_1, \ldots, x_n) \mapsto \mu_{n+1}^{(1,\ldots,n)}(x_1, \ldots, x_n; B)$$

is a measurable function with respect to $(\mathbb{X}^n, \mathscr{X}^n)$. Then, according to the Ionescu-Tulcea theorem, there exists a unique probability measure μ on the space $(\mathbb{X}^\infty, \mathscr{X}^\infty)$ such that

$$\begin{aligned} &\mu(A_1 \otimes \cdots \otimes A_n \otimes \mathbb{X}^\infty) \qquad\qquad (1.2.5)\\ &\quad = \int_{A_1} \mu_1(dx_1) \int_{A_2} \mu_2^{(1)}(x_1; dx_2) \cdots \int_{A_n} \mu_n^{(1,\ldots,n-1)}(x_1, \ldots, x_{n-1}; dx_n) \end{aligned}$$

for any $A_1, \ldots, A_n$ in $\mathscr{X}$ and for any $n \in \mathbb{N}$.
We observe that with respect to μ, μ_1 represents the distribution of a certain r.v. X_1 on $\mathbb{X}$ and $\mu_n^{(1,\ldots,n-1)}$ represent the regular conditional probability of a certain r.v. X_n given $X_1, \ldots, X_{n-1}$ for any $n \geq 2$. The sequence $\{X_n, n \geq 1\}$ is usually known as the Blackwell-MacQueen Pólya sequence with parameter α since it was first investigated by Blackwell and MacQueen [10].

Theorem 1.2.1. *(cfr. Blackwell and MacQueen [10]) Let $\{X_n, n \geq 1\}$ be the Blackwell-MacQueen Pólya sequence with parameter α.*

i) as $n \to +\infty$, $\mu_n^{(1,\ldots,n-1)}$ converges with probability 1 to a limiting discrete measure P^*;

ii) P^* *is a Dirichlet process with parameter* α*;*

iii) *given* P^**, the sequence* $\{X_n, n \geq 1\}$ *is a sequence of independent r.v. with distribution* P^**.*

Theorem 1.2.2. *(cfr. Regazzini [161]) Let* $\{X_n, n \geq 1\}$ *be the Blackwell-MacQueen Pólya sequence with parameter* α*. Then, the unique probability measure* μ *of* $\{X_n, n \geq 1\}$ *such that* (1.2.4) *is a sistem of predictive probability measures is exchangeable and its de Finetti measure is* Π*.*

Theorem 1.2.1 and Theorem 1.2.2 represent two important characterizations of the Blackwell-MacQueen Pólya sequence. In particular, we observe that if $\{X_n, n \geq 1\}$ is the Blackwell-MacQueen Pólya with parameter α, then it can be easily checked that X_i is distributed according to α/a. The Dirichlet process is useful only if we can do the necessary calculations for making inference. The most crucial is updating in the light of data. In the next theorem we consider the conditional distribution of a Dirichlet process P on $\mathbb{X}$ with parameter α given a sample $X_1, \ldots, X_n$ from P. It turns out that this conditional distribution is also Dirichlet process.

Theorem 1.2.3. *(cfr. Ferguson [61]) Let* P *be a Dirichlet process on* $\mathbb{X}$ *with parameter* α*, and let* $X_1, \ldots, X_n$ *be a sample of size* n *from* P*. Then the conditional distribution of* P *given* $X_1, \ldots, X_n$ *is as a Dirichlet process with parameter* $\alpha + \sum_{1 \leq i \leq n} \delta_{X_i}$*.*

In particular, the distribution of the posterior r.p.m. found in Theorem 1.2.3 is a regular conditional distribution.

Ferguson [61] and Blackwell [9] proved that there is a set of discrete distributions $\tilde{\mathcal{P}}_{\mathbb{X}} \subseteq \mathcal{P}_{\mathbb{X}}$ such that the Dirichlet process on $\mathbb{X}$ with parameter α assigns probability 1 to $\tilde{\mathcal{P}}_{\mathbb{X}}$. Note that for any given $p \in \mathcal{P}_{\mathbb{X}}$, the set $\{x \in \mathbb{X} : p(\{x\}) > 0\} \in \mathscr{X}$ since $\mathbb{X}$ is separable; moreover p is discrete if and only if $p(\{x \in \mathbb{X} : p(\{x\}) > 0\}) = 1$. One can in fact prove that the set $\{p \in \mathcal{P}_{\mathbb{X}} : p(\{x \in \mathbb{X} : p(\{x\}) > 0\}) = 1\}$ of all discrete probability distributions on $(\mathbb{X}, \mathscr{X})$ is an element of $\mathscr{P}_{\mathbb{X}}$. Then, it can be easily checked that

$$\Pi(\{p \in \mathcal{P}_{\mathbb{X}} : p(\{x \in \mathbb{X} : p(\{x\}) > 0\}) = 1\}) = 1. \tag{1.2.6}$$

Sethuraman [174] proved an alternative theorem, which not only shows that the Dirichlet process on $\mathbb{X}$ with parameter α is a r.p.m. on discrete distributions, but also gives the useful series representation (1.1.1) which provide an "algorithm" for approximately simulating a cumulative distribution function with distribution Π.

1.3 A generalized Sethuraman's construction

In this section, moving from the constructive definition of the Dirichlet process on $\mathbb{X}$ introduced by Sethuraman [174] we propose a more general construction and we investigate its properties.

Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure with total mass a and let $n_\bullet$ be a fixed integer-valued sequence. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space supporting three independent sequences of r.v.s $\theta := \{\theta_i, i \geq 1\}$, $q := \{(q_{i,1}, \ldots, q_{i,n_i}), i \geq 1\}$ and $Y := \{(Y_{i,1}, \ldots, Y_{i,n_i}), i \geq 1\}$, such that the application $(\theta, q, Y) : \Omega \to \times_{i\geq 1}([0,1] \times \Delta^{(n_i-1)} \times \mathbb{X}^{n_i})$ is $(\mathscr{F}, \otimes_{i\geq 1}(\mathscr{R} \cap [0,1] \otimes \mathscr{R}^{n_i} \cap \Delta^{(n_i-1)} \otimes \mathscr{X}^{n_i}))$-measurable. In particular, we assume the sequence θ to be a sequence of independent r.v.s distributed according to a Beta distribution function with parameter (n_i, a) for $i \geq 1$, the sequence q to be a sequence of independent r.v.s identically distributed according to a Dirichlet distribution function with parameter $(1, \ldots, 1)$ and the sequence Y to be a sequence of independent r.v.s (samples of size n_i for $i \geq 1$) from a Blackwell-MacQueen Pólya sequence with parameter α, i.e. if P_i are independent Dirichlet processes with parameter α for $i \geq 1$, then for $i \geq 1$, $Y_{i,1}, \ldots, Y_{i,n_i}|P_i$ are i.i.d. from P_i. The condition of independence between the sequence of r.v.s θ, q and Y and the usual construction of a product measure implies the existence of the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ supporting the r.v (θ, q, Y) and does not require any restrictions on $\mathbb{X}$, such as it being a Polish space. As for the Sethuraman constructive definition, we now consider the sequence of r.v.s $\{p_i, i \geq 1\}$ obtained from the sequence θ, by the usual stick-breaking construction $p_1 = \theta_1$ and $p_i = \theta_i \prod_{1\leq j\leq i-1}(1 - \theta_j)$ for $i \geq 2$. In particular, the stick-breaking construction implies that as $n \to +\infty$, $\sum_{1\leq i\leq n} p_i = 1 - \prod_{1\leq i\leq n}(1-\theta_i) \to 1$ a.s. For any $B \in \mathbb{X}$ we consider the map $P^{(n_\bullet)} : \Omega \to \mathcal{P}_{\mathbb{X}}$ defined by

$$\omega \mapsto \sum_{i\geq 1} p_i(\omega) \sum_{j=1}^{n_i} q_{i,j}(\omega)\delta_{Y_{i,j}(\omega)}(B) \tag{1.3.1}$$

which is clearly a measurable map and takes values in the subset of discrete probability measures on $(\mathbb{X}, \mathscr{X})$. In Theorem 1.3.1 we prove that for any finite measure $\alpha \in \mathcal{A}_{\mathbb{X}}$ and for any finite measurable partition $B_1, \ldots, B_k$ of $\mathbb{X}$, the r.v. $(P^{(n_\bullet)}(\cdot, B_1), \ldots, P^{(n_\bullet)}(\cdot, B_k))$ is distributed according to a Dirichlet distribution function with parameter $(\alpha(B_1), \ldots, \alpha(B_k))$. This establishes that $P^{(n_\bullet)}$ defined in (1.3.1) is a Dirichlet process with parameter α. Before proving Theorem 1.3.1 let us cosider the following lemma which introduce a distributional equation having as unique solution the Dirichlet process with parameter α. For similar distributional equation for the Dirichlet process see James [98] and for more general r.p.m., using the duality with the posterior distribution, see James [96].

Lemma 1.3.1. *Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure with total mass a and let $n_\bullet$ be a fixed* *integer-valued sequence. If (θ, q, Y) is the r.v. above described, then the distributional equation*

$$Q_i \stackrel{d}{=} \theta_i \sum_{j=1}^{n_i} q_{i,j}\delta_{Y_{i,j}} + (1-\theta_i)Q_i \qquad i \geq 1 \tag{1.3.2}$$

has as its unique solution the Dirichlet process with parameter α.

Proof. For any $i \geq 1$, let $\xi_{i,1}, \dots, \xi_{i,n}$ be n independent r.v.s such that $\xi_{i,j}$ is distributed according to a Beta distribution function with parameter $(1, n-j)$. From Theorem 1 in Jambunathan [95] it follows that the marginal distributions of the r.v. $(q_{i,1}, \dots, q_{i,n})$ for $i \geq 1$, can be represented as $q_{i,1} = \xi_{i,1}$ and $q_{i,j} = \xi_{i,j} \prod_{1 \leq l \leq j-1}(1-\xi_{i,l})$ for $j \geq 2$. By using this stick-breaking construction it follows by induction that $1 - \sum_{1\leq j \leq n-1} q_{i,j} = \prod_{1 \leq j \leq n-1}(1-\xi_{i,j})$. Now, if $B_1, \dots, B_k$ is any finite measurable partition of $\mathbb{X}$, then it follows that

$$\xi_{i,n}(\delta_{Y_{i,n}}(B_1), \dots, \delta_{Y_{i,n}}(B_k)) \qquad i \geq 1$$

is distributed according to a Dirichlet distribution function with the following parameter $(\delta_{Y_{i,n}}(B_1), \dots, \delta_{Y_{i,n}}(B_k))$. Using the stick-breaking construction for the marginal distributions of the r.v. $(q_{i,1}, \dots, q_{i,n})$, it follows by induction the following identity

$$\begin{aligned} &\sum_{j=1}^{n} q_{i,j}(\delta_{Y_{i,j}}(B_1), \dots, \delta_{Y_{i,j}}(B_k)) \\ &\quad = \sum_{j=1}^{n-1} q_{i,j}(\delta_{Y_{i,j}}(B_1), \dots, \delta_{Y_{i,j}}(B_k)) + \left(1 - \sum_{j=1}^{n-1} q_{i,j}\right)(\xi_{i,n}(\delta_{Y_{i,n}}(B_1), \dots, \delta_{Y_{i,n}}(B_k))). \end{aligned} \tag{1.3.3}$$

Since by construction $\sum_{1 \leq j \leq n} q_{i,j} = 1$, then it follows that given $Y_{i,1}, \dots, Y_{i,n}$

$$\sum_{j=1}^{n} q_{i,j}(\delta_{Y_{i,j}}(B_1), \dots, \delta_{Y_{i,j}}(B_k))$$

is distributed according to a Dirichlet distribution function with the following parameter $(\alpha(B_1) + \sum_{1\leq j\leq n} \delta_{Y_{i,j}}(B_1), \dots, \alpha(B_k) + \sum_{1 \leq j \leq n} \delta_{Y_{i,j}}(B_k))$. This argument shows that the Dirichlet process with parameter α satisfies the distributional equation (1.3.2). The uniqueness of the solution follows by Lemma 3.3 in Sethuraman [174] (see also Vervaat [185], Section 1). □

Lemma 1.3.1 proves that conditional given $Y_1, \dots, Y_n$ the r.p.m. $\sum_{1 \leq j \leq n} q_j \delta_{Y_j}$ is a Dirichlet process with parameter $\sum_{1 \leq i \leq n} \delta_{Y_i}$ which has been considered as a sort of Bayesian bootstrap empirical measure by Rubin [170]. Thus, if we denote by $P|Y_1, \dots, Y_n$

the posterior Dirichlet process with parameter $\alpha + \sum_{1\leq i\leq n} \delta_{Y_i}$, then as the total mass $a \to 0$

$$P|Y_1,\dots,Y_n \Rightarrow \sum_{j=1}^{n} q_j \delta_{Y_j}|Y_1,\dots,Y_n \qquad \text{a.s.-}\mathbb{P} \tag{1.3.4}$$

i.e., the r.p.m. $\sum_{1\leq j\leq n} q_j\delta_{Y_j}$ is the weak limit a.s.-$\mathbb{P}$ of the posterior Dirichlet process as the total mass tends to zero. In particular, the limit in (1.3.4) has been taken as a justification of the use of the Dirichlet process with parameter $\sum_{1\leq i\leq n} \delta_{Y_i}$ as a non-informative posterior. We are now in the position to prove the following theorem which represents the main result of this section.

Theorem 1.3.1. *Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure with total mass a, let $n_\bullet$ be a fixed integer-valued sequence. Let $P^{(n_\bullet)}$ the measurable map defined by* (1.3.1)*. Then, for any finite measurable partition $B_1,\dots,B_k$ of $\mathbb{X}$, $(P^{(n_\bullet)}(\cdot,B_1),\dots,P^{(n_\bullet)}(\cdot,B_k))$ is a r.v. distributed according to a Dirichlet d.f with parameter $(\alpha(B_1),\dots,\alpha(B_k))$.*

Proof. From Lemma 1.3.1, the distributional equation

$$P \stackrel{d}{=} \theta_1 \sum_{j=1}^{n_1} q_{1,j}(\delta_{Y_{1,j}}(B_1),\dots,\delta_{Y_{1,j}}(B_k)) + (1-\theta_1)P$$

has as unique solution the Dirichlet process with parameter α. Then, to prove the theorem, we can use arguments similar to those used in Lemma 1.3.1. In particular for any finite measurable partition $B_1,\dots,B_k$ of $\mathbb{X}$, if we define $\tilde{P} := (P(B_1),\dots,P(B_k))$, then

$$\begin{aligned}
&\sum_{i=1}^{m} p_i \sum_{j=1}^{n_i} q_{i,j}(\delta_{Y_{i,j}}(B_1),\dots,\delta_{Y_{i,j}}(B_k)) + \left(1-\sum_{i=1}^{m} p_i\right)\tilde{P} \\
&\quad = \sum_{i=1}^{m-1} p_i \sum_{j=1}^{n_i} q_{i,j}(\delta_{Y_{i,j}}(B_1),\dots,\delta_{Y_{i,j}}(B_k)) \\
&\qquad + \left(1-\sum_{i=1}^{m-1} p_i\right)\left(\theta_m \sum_{j=1}^{n_m} q_{m,j}(\delta_{Y_{m,j}}(B_1),\dots,\delta_{Y_{m,j}}(B_k)) + (1-\theta_m)\tilde{P}\right)
\end{aligned}$$

where

$$\theta_m \sum_{j=1}^{n} q_{m,j}(\delta_{Y_{m,j}}(B_1),\dots,\delta_{Y_{m,j}}(B_k)) + (1-\theta_m)\tilde{P}$$

is a r.v. distributed according to a Dirichlet distribution function and its parameter is $(\alpha(B_1),\dots,\alpha(B_k))$. Then, it follows that

$$\sum_{i=1}^{m} p_i \sum_{j=1}^{n_i} q_{i,j}(\delta_{Y_{i,j}}(B_1),\dots,\delta_{Y_{i,j}}(B_k)) + \left(1-\sum_{j=1}^{m} p_i\right)\tilde{P}$$

is a r.v. distributed according to a Dirichlet distribution function and its parameter is $(\alpha(B_1), \ldots, \alpha(B_k))$. Then, the result follow by taking the limit for $m \to +\infty$. □

Theorem 1.3.1 proved that for any fixed integer-valued sequence $n_\bullet$

$$P^{(n_\bullet)} = \sum_{i \geq 1} p_i \sum_{j=1}^{n_i} q_{i,j} \delta_{Y_{i,j}}$$

is a new constructive definition of the Dirichlet process on $\mathbb{X}$ with parameter α. The original constructive definition of the Dirichlet process obtained by Sethuraman can be recoverd as particular case by setting $n_\bullet = 1_\bullet$, i.e.

$$P^{(1_\bullet)} = \sum_{i \geq 1} p_i \sum_{j=1}^{1} q_{i,j} \delta_{Y_{i,j}} = \sum_{i \geq 1} p_i \delta_{Y_{i,1}}.$$

Moreover, if we fix $n_\bullet$ to be a constant sequence and equal to $n \in \mathbb{N}$ then, as a consequence of Theorem 1.3.1 we have the following result.

Corollary 1.3.1. *Let $n_\bullet$ be a fixed integer-valued constant sequence equal to $n \in \mathbb{N}$. Then, $P^{(n_\bullet)} \Rightarrow P^*$ a.s.-$\mathbb{P}$ as $n \to +\infty$ where P^* is a Dirichlet process with parameter α.*

Proof. If we define $H_n := \sum_{1 \leq j \leq n} q_{1,j} \delta_{Y_{1,j}}$, we need to prove that as $n \to +\infty$, the sequence $\{H_n, n \geq 1\}$ converges weakly a.s.-$\mathbb{P}$ to a r.p.m. P^* and P^* is a Dirichlet process with parameter α. For a bounded and continuous function $g : \mathbb{X} \to \mathbb{R}$ define $G_n := \int_{\mathbb{X}} H_n(dx)$. First of all we prove that for all bounded and continuous $g : \mathbb{X} \to \mathbb{R}$, the sequence of r.v.s $\{G_n, n \geq 1\}$ converges a.s. Indeed, let us consider for any $k = 1, 2, \ldots$ and $m \in \mathbb{N}_0$,

$$G_{m+k} = \left(\prod_{i=m}^{m+k} (1 - \xi_i) \right) G_{m-1} + \sum_{j=0}^{k} \left(\xi_{m+j} \prod_{i=m+j+1}^{m+k} (1 - \xi_i) \right) g(Y_{m+j}).$$

where $\{\xi_n, n \geq 1\}$ is a sequence of independent r.v. such that ξ_n is distributed according to a Beta distribution function with parameter $(1, n-1)$. Let us denote by $W_{m,m+k} := \prod_{m \leq i \leq m+k} (1 - \xi_i)$ and by K a positive constant such that $|g| \leq K$, so that

$$\begin{aligned} |G_{m+k} - G_{m-1}| &\leq (1 - W_{m,m+k}) G_{m-1} + \sum_{j=0}^{k} \left(\xi_{m+j} \prod_{i=m+j+1}^{m+k} (1 - \xi_i) \right) |g(Y_{m+j})| \\ &\leq (1 - W_{m,m+k}) K + (1 - W_{m,m+k}) K = 2K(1 - W_{m,m+k}). \end{aligned}$$

Since $1 - W_{m,m+k} = \sum_{m \leq i \leq m+k} p_i$, where $\{p_i, i \geq 1\}$ is a sequence of r.v.s defined by the usual stick-breaking construction $p_1 = 1 - \xi_1$ and $p_i = \xi_i \prod_{2 \leq j \leq i-1} (1 - \xi_j)$ for $i \geq 2$. In particular, $\sum_{i\geq 1} p_i = 1$ for almost all $\omega \in \Omega$ since $\sum_{i\geq 1} \mathbb{E}(\log(\xi_i)) = -\sum_{i\geq 1}(i-1)/i^2 = -\infty$ (see Ishwaran and James [90], Lemma 1), then $1 - W_{m,m+k} \to 0$ for almost all $\omega \in \Omega$. We conclude that the sequence $\{G_n, n \geq 1\}$ is Cauchy a.s., therefore it converges a.s. to some r.v. By Theorem 2.2 in Berti et al. [7] it follows that there exists a r.p.m. P^* on $\mathbb{X}$, defined on $(\Omega, \mathscr{F}, \mathbb{P})$, such that, as $n \to +\infty$,

$$H_n \Rightarrow \mathbb{P}^* \qquad \text{a.s.-}\mathbb{P}$$

Using Theorem 4.2 in Kallenberg [104], the second part of the proof is devoted to prove that for any finite measurable partition $B_1, \ldots, B_k$ of $\mathbb{X}$ we have $(H_n(B_1), \ldots, H_n(B_k)) \Rightarrow (P^*(B_1), \ldots, P^*(B_k))$. In particular, we observe that conditioning to $Y_1, \ldots, Y_n$, the r.v. $(H_n(B_1, \ldots, H_n(B_k))$ is distributed according to a Dirichlet distribution funtion with parameter $(\sum_{1\leq i\leq n} \delta_{Y_i}(B_1), \ldots, \sum_{1\leq i\leq n} \delta_{Y_i}(B_k), n - \sum_{1\leq j\leq k}\sum_{1\leq i\leq n} \delta_{Y_i}(B_j))$. Then, the $(r_1, \ldots, r_k)$-th moment of $(H_n(B_1, \ldots, H_n(B_k))$ can be calculated as

$$\mathbb{E}[H_n^{r_1}(B_1)\cdots H_n^{r_k}(B_k)]$$
$$= \sum_{(j_1,\ldots,j_k)\in\mathcal{D}_{k,n}^{(0)}} \binom{n}{j_1 \cdots j_k} \frac{(\alpha(B_1))_{j_1\uparrow 1}\cdots(\alpha(B_k))_{j_k\uparrow 1}}{(a)_{n\uparrow 1}} \frac{(j_1)_{r_1\uparrow 1}\cdots(j_k)_{r_k\uparrow 1}}{(n)_{(r_1+\cdots+r_k)\uparrow 1}}$$

where we defined $\mathcal{D}_{k,n}^{(0)} := \{(j_1, \ldots, j_k) \in \{0, \ldots, n\}^k : \sum_{1\leq i\leq k} j_i = n\}$. In particular, we observe that

$$\sum_{(j_1,\ldots,j_k)\in\mathcal{D}_{k,n}^{(0)}} \binom{n}{j_1 \cdots j_k} \frac{(\alpha(B_1))_{j_1\uparrow 1}\cdots(\alpha(B_k))_{j_k\uparrow 1}}{(a)_{n\uparrow 1}} (j_1)_{r_1\uparrow 1}\cdots(j_k)_{r_k\uparrow 1}$$

is the $(r_1, \ldots, r_k)$-th ascending factorial moment for a r.v. distributed according to a multivariate Pólya distribution function with parameter $(n, \alpha(B_1), \ldots, \alpha(B_k))$. Finally, observe that

$$\mathbb{E}[\mathrm{e}^{\sum_{i=1}^k it_i P(B_i)}] = \Phi_2^{(k)}\left(\alpha(B_1), \ldots, \alpha(B_k); \sum_{i=1}^k \alpha(B_i); it_1, \ldots, it_k\right)$$
$$= \sum_{(r_1,\ldots,r_k)\in(\mathbb{N}_0)^k} \frac{(it_1)^{r_1}\cdots(it_k)^{r_k}}{r_1!\cdots r_k!} \frac{(\alpha(B_1))_{r_1\uparrow 1}\cdots(\alpha(B_k))_{r_p\uparrow 1}}{(a)_{(r_1+\cdots+r_k)\uparrow 1}}$$

Let us consider $k = 2$ and observe that

$$\frac{1}{(n)_{(r_1+r_2)\uparrow 1}} \sum_{(j_1,j_2)\in\mathcal{D}_{2,n}^{(0)}} \binom{n}{j_1, j_2} \frac{(\alpha(B_1))_{j_1\uparrow 1}(\alpha(B_2))_{j_2\uparrow 1}}{(a)_{n\uparrow 1}} (j_1)_{r_1\uparrow 1}(j_2)_{r_2\uparrow 1}$$
$$= \frac{1}{(n)_{(r_1+r_2)\uparrow 1}} \sum_{t_1=0}^{r_1} |s(r_1, t_1)| \sum_{s_1=0}^{t_1} S(t_1, s_1) \sum_{t_2=0}^{r_2} |s(r_2, t_2)| \sum_{s_2=0}^{t_2} S(t_2, s_2)$$
$$\times \sum_{(j_1,j_2)\in\mathcal{D}_{2,n}^{(0)}} \binom{n}{j_1, j_2} \frac{(\alpha(B_1))_{j_1\uparrow 1}(\alpha(B_2))_{j_2\uparrow 1}}{(a)_{n\uparrow 1}} (j_1)_{s_1\downarrow 1}(j_2)_{s_2\downarrow 1}$$

where $s(\cdot,\cdot)$ is the Stirling number of the first kind and $S(\cdot,\cdot)$ is the Stirling number of the second kind. In particular, we have

$$\begin{aligned} C_{n+1}^{(s_1,s_2)} &:= \sum_{(j_1,j_2)\in\mathcal{D}_{2,n+1}^{(0)}} \binom{n+1}{j_1,j_2} \frac{(\alpha(B_1))_{j_1\uparrow 1}(\alpha(B_2))_{j_2\uparrow 1}}{(a)_{(n+1)\uparrow 1}} (j_1)_{s_1\downarrow 1}(j_2)_{s_2\downarrow 1} \\ &= \sum_{j_1=0}^{n} \binom{n+1}{j_1+1} \frac{(\alpha(B_1))_{(j_1+1)\uparrow 1}(\alpha(B_2))_{(n-j_1)\uparrow 1}}{(a)_{(n+1)\uparrow 1}} (j_1+1)_{s_1\downarrow 1}(n-j_1)_{s_2\downarrow 1} \\ &= \frac{n+1}{(a+n)} \sum_{j_1=0}^{n} \binom{n}{j_1} (\alpha(B_1)+s_1-1+j_1-s_1+1) \\ &\qquad\qquad \times \frac{(\alpha(B_1))_{j_1\uparrow 1}\alpha(B_2)_{(n-j_1)\uparrow 1}}{(a)_{(n)\uparrow 1}} (j_1)_{(s_1-1)\downarrow 1}(n-j_1)_{s_2\downarrow 1} \\ &= \frac{(n+1)(\alpha(A_1)+s_1-1)}{(a+n)} C_n^{(s_1-1,s_2)} + \frac{n}{(a+n)} C_n^{(s_1,s_2)} \end{aligned}$$

The last recursive equation we obtained provides by induction on n, an expression for the descending factorial moment of order (s_1,s_2) of a r.v. distributed according to a bivariate Pólya distribution function with parameter $(n,\alpha(B_1),\alpha(B_2))$, i.e. $C_n^{(s_1,s_2)} = (\alpha(B_1))_{s_1\uparrow 1}(\alpha(B_2))_{s_2\uparrow 1}(n)_{(s_1+s_2)\downarrow 1}/(a)_{(s_1+s_2)\uparrow 1}$ where the starting point is

$$\begin{aligned} C_n^{(1,1)} &= \sum_{i=0}^{n-1} \frac{(i+1)\alpha(A_1)}{(a+i)} C_i^{(0,1)} \prod_{j=i+1}^{n-1} \frac{j+1}{a+j} \\ &= \frac{\Gamma(n+1)\alpha(A_1)}{\Gamma(a+n)} \sum_{i=0}^{n-1} \frac{\Gamma(a+i)}{\Gamma(i+1)} C_i^{(0,1)} \\ &= \frac{\Gamma(n+1)\alpha(A_1)}{\Gamma(a+n)} \sum_{i=0}^{n-1} \frac{\Gamma(a+i)}{\Gamma(i+1)} \frac{i\alpha(A_2)}{a} = \frac{\alpha(A_1)\alpha(A_2)(n)_{2\downarrow 1}}{(a)_{2\uparrow 1}} \end{aligned}$$

The general case can be easily proved by induction hypothesis. As for the case $k=2$, in general we can write

$$\begin{aligned} &\frac{1}{(n)_{(r_1+\cdots+r_k)\uparrow 1}} \sum_{(j_1,\ldots,j_k)\in\mathcal{D}_{k,n}^{(0)}} \binom{n}{j_1\cdots j_k} \frac{(\alpha(B_1))_{j_1\uparrow 1}\cdots(\alpha(B_k))_{j_k\uparrow 1}}{(a)_{n\uparrow 1}} (j_1)_{r_1\uparrow 1}\cdots(j_k)_{r_k\uparrow 1} \\ &= \frac{1}{(n)_{(r_1+\cdots+r_k)\uparrow 1}} \sum_{t_1=0}^{r_1} |s(r_1,t_1)| \sum_{s_1=0}^{t_1} S(t_1,s_1)\cdots \sum_{t_k=0}^{r_k} |s(r_k,t_k)| \sum_{s_k=0}^{t_k} S(t_k,s_k) \\ &\qquad \times \sum_{(j_1,\ldots,j_k)\in\mathcal{D}_{k,n}^{(0)}} \binom{n}{j_1\cdots j_k} \frac{(\alpha(B_1))_{j_1\uparrow 1}\cdots(\alpha(B_k))_{j_k\uparrow 1}}{(a)_{n\uparrow 1}} (j_1)_{s_1\downarrow 1}\cdots(j_k)_{s_k\downarrow 1} \end{aligned}$$

In particular, we have

$$\begin{aligned}
C_{n+1}^{(s_1,\ldots,s_k)} &= \sum_{(j_1,\ldots,j_k)\in\mathcal{D}_{k,n}^{(0)}} \binom{n}{j_1\cdots j_k} \frac{(\alpha(B_1))_{j_1\uparrow 1}\cdots(\alpha(B_k))_{j_p\uparrow 1}}{(a)_{n\uparrow 1}} (j_1)_{s_1\downarrow 1}\cdots(j_k)_{s_k\downarrow 1} \\
&= \sum_{j_k=0}^{n} \binom{n}{j_p} \frac{(\alpha(B_k))_{j_p\uparrow 1}(a-\alpha(B_k))_{(n-j_k)\uparrow 1}(j_k)_{s_k\downarrow 1}}{(a)_{n\uparrow 1}} \\
&\qquad \times \sum_{(j_1,\ldots,j_{k-1})\in\mathcal{D}_{k-1,n-j_k}^{(0)}} \binom{n-j_k}{j_1\cdots j_{k-1}} \frac{(\alpha(B_1))_{j_1\uparrow 1}\cdots(\alpha(B_{k-1}))_{j_{k-1}\uparrow 1}}{(a-\alpha(B_k))_{(n-j_k)\uparrow 1}} \\
&\qquad \times (j_1)_{s_1\downarrow 1}\cdots(j_{k-1})_{s_{k-1}\downarrow 1} \\
&= \sum_{j_k=0}^{n} \binom{n}{j_k} \frac{(\alpha(B_k))_{j_k\uparrow 1}(a-\alpha(B_k))_{(n-j_k)\uparrow 1}(j_k)_{s_k\downarrow 1}}{(a)_{n\uparrow 1}} \\
&\qquad \times \frac{(\alpha(B_1))_{s_1\uparrow 1}\cdots(\alpha(B_{k-1}))_{s_{k-1}\uparrow 1}(n-j_k)_{(s_1+\cdots+s_{k-1})\downarrow 1}}{(a-\alpha(B_k))_{(s_1+\cdots+s_{k-1})\uparrow 1}} \\
&= \frac{(n+1)(\alpha(B_k)+s_k-1)}{a+n} C_n^{(s_1,\ldots,s_k-1)} + \frac{n+1}{a+n} C_n^{(s_1,\ldots,s_k)}
\end{aligned}$$

The last recursive equation we obtained provides by induction on n, an expression for the descending factorial moment of order $(s_1,\ldots,s_k)$ of a r.v. distributed according to a multivariate Pólya distribution function with parameter $(n,\alpha(B_1),\ldots,\alpha(B_k))$, i.e. $C_n^{(s_1,\ldots,s_k)} = (\alpha(B_1))_{s_1\uparrow 1}\cdots(\alpha(B_k))_{s_k\uparrow 1}(n)_{(s_1+\cdots+s_k)\downarrow 1}(a)_{(s_1+\cdots+s_k)\uparrow 1}$ where the starting point is

$$\begin{aligned}
C_n^{(1,\ldots,1)} &= \sum_{i=0}^{n-1} \frac{(i+1)\alpha(B_k)}{(a+i)} C_i^{(1,\ldots,1,0)} \prod_{j=i+1}^{n-1} \frac{j+1}{a+j} \\
&= \frac{\Gamma(n+1)}{\Gamma(a+n)} \frac{(\alpha(B_1))_{s_1\uparrow 1}\cdots(\alpha(B_{k-1}))_{s_{k-1}\uparrow 1}\alpha(B_k)}{(a-\alpha(B_k))_{(s_1+\cdots+s_{k-1})\uparrow 1}} \sum_{i=0}^{n-1} \frac{\Gamma(a+i)}{\Gamma(1+i)} C_i^{(1,\ldots,1,0)} \\
&= \frac{\Gamma(n+1)}{\Gamma(a+n)} \frac{(\alpha(B_1))_{s_1\uparrow 1}\cdots(\alpha(B_{k-1}))_{s_{k-1}\uparrow 1}\alpha(B_k)}{(a-\alpha(B_k))_{(s_1+\cdots+s_{k-1})\uparrow 1}} (-1)^{p+1} \frac{\Gamma(-k-a+2)}{\Gamma(-a+\alpha(B_k)+2-k)} \\
&\qquad \times \sum_{i=0}^{n-1} \frac{\Gamma(a+i)}{\Gamma(1+i)} \frac{(a-\alpha(B_k))_{i\uparrow 1}(-i)_{(k-1)\uparrow 1}}{(a)_{i\uparrow 1}} \frac{\Gamma(-a+\alpha(B_k)-i+1)}{\Gamma(-a-i+1)} \\
&= \frac{\alpha(B_1)\cdots\alpha(B_k)(n)_{k\downarrow 1}}{(a)_{k\uparrow 1}}
\end{aligned}$$

Then, we have

$$\begin{aligned}
&\lim_{n\to+\infty}\frac{1}{(n)_{(r_1+\cdots+r_k)\uparrow 1}}\sum_{t_1=0}^{r_1}|s(r_1,t_1)|\sum_{s_1=0}^{t_1}S(t_1,s_1)\cdots\sum_{t_k=0}^{r_k}|s(r_k,t_k)|\sum_{s_k=0}^{t_k}S(t_k,s_k)\\
&\qquad\times\sum_{(j_1,\dots,j_k)\in\mathcal{D}^{(0)}_{k,n}}\binom{n}{j_1,\dots,j_k}\frac{(\alpha(B_1))_{j_1\uparrow 1}\cdots(\alpha(B_k))_{j_k\uparrow 1}}{(a)_{n\uparrow 1}}(j_1)_{s_1\downarrow 1}\cdots(j_k)_{s_k\downarrow 1}\\
&=\lim_{n\to+\infty}\frac{1}{(n)_{(r_1+\cdots+r_k)\uparrow 1}}\sum_{t_1=0}^{r_1}|s(r_1,t_1)|\sum_{s_1=0}^{t_1}S(t_1,s_1)\cdots\sum_{t_k=0}^{r_k}|s(r_k,t_k)|\sum_{s_k=0}^{t_k}S(t_k,s_k)\\
&\qquad\times\frac{(\alpha(B_1))_{s_1\uparrow 1}\cdots(\alpha(B_k))_{s_k\uparrow 1}(n)_{(s_1+\cdots+s_k)\downarrow 1}}{(a)_{(s_1+\cdots+s_k)\uparrow 1}}\\
&=\sum_{t_1=0}^{r_1}|s(r_1,t_1)|\sum_{s_1=0}^{t_1}S(t_1,s_1)\cdots\sum_{t_k=0}^{r_k}|s(r_k,t_k)|\sum_{s_k=0}^{t_k}S(t_k,s_k)\\
&\qquad\times\lim_{n\to+\infty}\frac{(\alpha(B_1))_{s_1\uparrow 1}\cdots(\alpha(B_k))_{s_k\uparrow 1}(n)_{(s_1+\cdots+s_k)\downarrow 1}}{(a)_{(s_1+\cdots+s_k)\uparrow 1}(n)_{(r_1+\cdots+r_k)\uparrow 1}}\\
&=|s(r_1,r_1)|S(r_1,r_1)\cdots|s(r_k,r_k)|S(r_k,r_k)\frac{(\alpha(B_1))_{r_1\uparrow 1}\cdots(\alpha(B_k))_{r_k\uparrow 1}}{(a)_{(r_1+\cdots+r_k)\uparrow 1}}\\
&=\frac{(\alpha(B_1))_{r_1\uparrow 1}\cdots(\alpha(B_k))_{r_k\uparrow 1}}{(a)_{(r_1+\cdots+r_k)\uparrow 1}}
\end{aligned}$$

completing the proof. □

1.4 Posterior characterization

Let $X_1,\dots,X_n$ be a sample of size n from a Dirichlet process with parameter α. For any fixed integer-valued sequence $n_\bullet$ we are now interested in providing the distribution of $P^{(n_\bullet)}|X_1,\dots,X_n$. Since we proved that $P^{(n_\bullet)}$ is a Dirichlet process with parameter α, then from Theorem 1.2.3 it is known that $P^{(n_\bullet)}|X_1,\dots,X_n$ is a Dirichlet process with parameter $\alpha+\sum_{1\le i\le n}\delta_{X_i}$. However, we show how to prove the posterior distribution starting from the new constructive definition of the Dirichlet process. Let I be a r.v. having support $\mathbb{N}$ and such that $p_i=\mathbb{P}(I=i|\theta,q,Y)$ for $i\ge 1$. Moreover, let X be a r.v. having support $\mathbb{X}$ and such that, for any fixed integer-valued sequence $n_\bullet$ and for any $B\in\mathscr{X}$

$$\mathbb{P}(X\in B|\theta,q,Y,I)=\sum_{j=1}^{n_I}q_{I,j}\mathbb{P}(Y_{I,j}\in B).$$

Thus, the r.v. X is a function of (θ,q,Y,I). The next lemma represent the extension of Lemma 4.1. in Sethuraman [174].

Lemma 1.4.1. *Let $n_\bullet$ be a fixed integer-valued sequence $n_\bullet$. Then, for any $B \in \mathscr{X}$, the distribution of X given $P^{(n_\bullet)}$ is equaivalent to the distribution of $P^{(n_\bullet)}$.*

Proof. We have

$$\begin{aligned}\mathbb{P}(X \in B|\theta, q, Y) &= \sum_{i\geq 1} \mathbb{P}(X \in B|I = i, \theta, q, Y)\mathbb{P}(I = i|\theta, q, Y) \\ &= \sum_{i\geq 1} p_i \sum_{j=1}^{n_i} q_{i,j}\delta_{Y_{i,j}}(B) = P^{(n_\bullet)}(B).\end{aligned}$$

□

We note that in the constructive definition of the Dirichlet process proposed by Sethuraman [174], the r.v. X was defined such that we have $\mathbb{P}(X \in B|\theta, q, Y, I) = \mathbb{P}(Y_I \in B)$, that is $X \stackrel{d}{=} Y_I$. We now consider the conditional distribution of $P^{(n_\bullet)}$ given X. Let $n^*_\bullet = \{n_i, i \geq 2\}$ and let $\theta^*_i = \theta_{i+1}$, $(q^*_{i,1}, \dots, q^*_{i,n_i}) = (q_{i+1,1}, \dots, q_{i+1,n_i})$ and $(Y^*_{i,1}, \dots, Y^*_{i,n_i}) = (Y_{i+1,1}, \dots, Y_{i+1,n_i})$ for $i \geq 1$. For any fixed integer-valued sequence $n_\bullet$, the r.p.m. $P^{(n_\bullet)}$ defined via the measurable map (1.3.1) satisfies

$$P^{(n_\bullet)} = \theta_1 \sum_{j=1}^{n_1} q_{1,j}\delta_{Y_{1,j}} + (1 - \theta_1)P^{(n^*_\bullet)}_{(\theta^*,q^*,Y^*)} \tag{1.4.1}$$

where

$$P^{(n^*_\bullet)}_{(\theta^*,q^*,Y^*)} = \sum_{i\geq 1} p^*_i \sum_{j=1}^{n_i} q_{i,j}\delta_{Y_{i,j}}$$

with $p^*_1 = \theta^*_1$ and $p^*_i = \theta^*_i \prod_{1\leq j\leq i-1}(1-\theta^*_j)$ for $i \geq 2$. As originally proposed by Sethuraman [174], it is possible to obtain the posterior distribution separately using the conditional distribution of the r.v. (θ, q, Y) given $I = 1$ and then given $I > 1$. Before doing this, we consider the following lemma which extends Lemma 4.2. in Sethuraman [174]. Given a r.v. X and a r.v. Y we will use the notation $\mathscr{L}(X)$ and $\mathscr{L}(X|Y)$ to denote the distribution of X and the conditional distribution of X given Y, under $\mathbb{P}$, respectively.

Lemma 1.4.2. *For any fixed integer-valued sequence $n_\bullet$, let $\tilde{Y}_i := (Y_{i,1}, \dots, Y_{i,n_i})$ and $\tilde{q}_i := (q_{i,1}, \dots, q_{i,n_i})$. Then the conditional distribution of the r.v. (θ, q, Y, I) given $I = 1$ is*

$$\mathscr{L}((\theta_1, \tilde{q}_1, \tilde{Y}_1), \{(\theta_i, \tilde{q}_i, \tilde{Y}_i), i \geq 2\}|I = 1) = \mathscr{L}(W)\mathscr{L}((\theta, q, Y)) \tag{1.4.2}$$

where W is a r.v. distributed according to a Beta distribution function with parameter $(n_1 + 1, a)$. The conditional distribution of the r.v. (θ, q, Y, I) given $I > 1$ is

$$\mathscr{L}((\theta_1, \tilde{q}_1, \tilde{Y}_1), \{(\theta_i, \tilde{q}_i, \tilde{Y}_i), i \geq 2\}, I - 1|I > 1) = \mathscr{L}(V)\mathscr{L}((\theta, q, Y, I)) \tag{1.4.3}$$

where V is a r.v. distributed according to a Beta distribution function with parameter $(n_1, a + 1)$.

Proof. The proof is along lines similar to the proof of Lemma 4.2. in Sethuraman [174]. If $A_i \in \mathscr{R} \cap [0,1]$, $B_i \in \mathscr{X}^{n_i}$, $C_i \in \mathscr{R}^{n_i} \cap \Delta^{(n_i-1)}$ for $i = 1, \dots, m$, then we consider the joint distribution

$$\begin{aligned}
&\mathbb{P}(\theta_i \in A_i, (q_{i,1}, \dots, q_{i,n_i}) \in C_i, (Y_{i,1}, \dots, Y_{i,n_i}) \in B_i, i = 1, \dots, m, I = 1) \\
&\quad \propto \int_{\times_{i=1}^m ([0,1] \times \Delta^{(n_i-1)} \times \mathbb{X}^{n_i})} \prod_{i=1}^m x_i^{n_i - 1 + \mathbb{1}_{\{i=1\}}} (1 - x_i)^{a-1} \\
&\qquad \times \mathbb{P}((q_{i,1} Y_{i,1}) \in (dz_{i,1}, dy_{i,1}), \dots, (q_{i,n_i}, Y_{i,n_i}) \in (dz_{i,n_i}, dy_{i,n_i})) \\
&\qquad\qquad \times \mathbb{1}_{\{x_i \in A_i, (q_{i,1}, \dots, q_{i,n_i}) \in C_i, (y_{i,1}, \dots, y_{i,n_i}) \in B_i, i=1,\dots,m\}} dx_i.
\end{aligned}$$

This implies that, conditional on $I = 1$, the r.v. θ_1 is distributed according to a Beta distribution function with parameter $(n_1 + 1, a)$ and the distributions of θ_i, $(q_{i,1}, \dots, q_{i,n_i})$ and $(Y_{i,1}, \dots, Y_{i,n_i})$ for $i = 1, \dots, m$ do not change. In the same way we consider the joint distribution,

$$\begin{aligned}
&\mathbb{P}(\theta_i \in A_i, (q_{i,1}, \dots, q_{i,n_i}) \in C_i, (Y_{i,1}, \dots, Y_{i,n_i}) \in B_i, i = 1, \dots, m, I > 1) \\
&\quad \propto \int_{\times_{i=1}^m ([0,1] \times \Delta^{(n_i-1)} \times \mathbb{X}^{n_i})} \prod_{i=1}^m x_i^{n_i - 1} (1 - x_i)^{a-1+\mathbb{1}_{\{i=1\}}} \\
&\qquad \times \mathbb{P}((q_{i,1} Y_{i,1}) \in (dz_{i,1}, dy_{i,1}), \dots, (q_{i,n_i}, Y_{i,n_i}) \in (dz_{i,n_i}, dy_{i,n_i})) \\
&\qquad\qquad \times \mathbb{1}_{\{x_i \in A_i,, (q_{i,1}, \dots, q_{i,n_i}) \in C_i, (y_{i,1}, \dots, y_{i,n_i}) \in B_i, i=1,\dots,m\}} dx_i
\end{aligned}$$

so that (1.4.3) follows by the same arguments used for (1.4.2), since we have that $\mathbb{P}(I > 1|\theta, q, Y) = (1 - \theta_1)$. □

Proposition 1.4.1. *For any fixed integer-valued sequence $n_\bullet$, the r.p.m. $P^{(n_\bullet)}$ given X is a Dirichlet process with parameter $\alpha + \delta_X$.*

Proof. The proof is along lines similar to the proof of Theorem 4.3. in Sethuraman [174]. We separate the case when $I = 1$ and $I > I$. When $I = 1$ we are interested in the first part of equation (1.3.1) and we have

$$\begin{aligned}
\mathscr{L}(P^{(n_\bullet)}|X, I = 1) &= \mathscr{L}\left(\theta_1 \sum_{j=1}^{n_1} q_{1,j} \delta_{Y_{1,j}} + (1 - \theta_1) P^{(n_\bullet)}_{(\theta^*, q^*, Y^*)} | X, I = 1\right) \qquad (1.4.4) \\
&= \mathscr{L}\left(\theta_1' \delta_X + (1 - \theta_1') P^{(n_\bullet^{**})}_{(\theta^{**}, q^{***}, Y^{**})}\right)
\end{aligned}$$

where for the previous lemma θ_1' is a r.v. distributed according to a Beta distribution function with parameter $(n_1 + 1, a)$ and for any finite measurable partition $B_1, \dots, B_k$ of $\mathbb{X}$ we have that δ_X given X is distribuited as a Dirichlet distribution with parameter $(\delta_X(B_1), \dots, \delta_X(B_k))$, which is equivalent to a Dirichlet distribution with parameter $((n_1+$

$1)\delta_X(B_1),\dots,(n_1+1)\delta_X(B_k))$. Moreover $P^{(n_\bullet^{***})}_{(\theta^{**},q^{**},Y^{**})}$ is a Dirichlet process independent of $\theta_1^{'}$ with parameter α. Then $P^{(n_\bullet)}|X, I=1$ is a Dirichlet process with parameter $\alpha+(n_1+1)\delta_X$. When $I>1$ we are interested in the second part of equation (1.4.1), then we have

$$\begin{aligned}\mathscr{L}(P^{(n_\bullet)}|X, I>1) &= \mathscr{L}\left(\theta_1\sum_{j=1}^{n_1} q_{1,j}\delta_{Y_{1,j}} + (1-\theta_1)P^{(n_\bullet^{*})}_{(\theta^{*},q^{*},Y^{*})}\right) \\ &= \mathscr{L}\left(\theta_1^{''}\sum_{j=1}^{n_1} q_{1,j}\delta_{Y_{1,j}} + (1-\theta_1^{''})P^{(n_\bullet^{***})}_{(\theta^{***},q^{***}Y^{***})}\right)\end{aligned} \tag{1.4.5}$$

where for the previous lemma $\theta_1^{''}$ is a r.v. distributed according to a Beta distribution function with parameter $(n_1, a+1)$ and $P^{(n_\bullet^{***})}_{(\theta^{***},q^{***}Y^{***})}$ is a r.p.m., independent of $\theta_1^{''}$ and $Y_1,\dots,Y_{n_1}$ and having distribution the same distribution of $P^{(n_\bullet)}|X$. We combine Equations (1.4.4) and (1.4.5) to obtain a distributional equation for $P^{(n_\bullet)}|X$. In particular, we have

$$\begin{aligned}P^{(n_\bullet)}|X \overset{d}{=} A\left(\theta_1^{'}\delta_X + (1-\theta_1^{'})P^{(n_\bullet^{**})}_{(\theta^{**},q^{**},Y^{**})}\right) \\ + (1-A)\left(\theta_1^{''}\sum_{j=1}^{n_1} q_{1,j}\delta_{Y_{1,j}} + (1-\theta_1^{''})P^{(n_\bullet^{***})}_{(\theta^{***},q^{***},Y^{***})}\right)\end{aligned} \tag{1.4.6}$$

where the r.v.s on the right are independent and the r.v. A taking value 1 and 0 with probabilities $1/a{+}1$ and $a/a{+}1$, respectively. Notice that the distribution of $P^{(n_\bullet^{***})}_{(\theta^{***},q^{***},Y^{***})}$ is the same of the $P^{(n_\bullet)}|X$ which makes (1.4.6) a distributional equation. Finally, we want to verify that the Dirichlet process with parameter $\alpha+\delta_X$ is a solution of the distributional equation (1.4.6); then the uniqueness of the solution follows by Lemma 3.3 in Sethuraman [174] (see also Vervaat [185], Section 1). Let $Q_{\alpha+\delta_X+\sum_{1\le j\le n_1}\delta_{Y_{1,j}}}$ be a Dirichlet process on $\mathbb{X}$ with parameter $\alpha+\delta_X+\sum_{1\le j\le n_1}\delta_{Y_{1,j}}$, then conditional to $Y_{1,1},\dots,Y_{1,n_1}$ and taking expectation with respect to $Y_{1,1},\dots,Y_{1,n_1}$ we have

$$\begin{aligned}\mathbb{E}[Q_{\alpha+\delta_X+\sum_{j=1}^{n_1}\delta_{Y_{1,j}}}] &= \theta_1^{''}\sum_{j=1}^{n_1} q_{1,j}\delta_{Y_{1,j}} + (1-\theta_1^{''})P^{(n_\bullet^{***})}_{(\theta^{***},q^{***},Y^{***})} \\ &= \sum_{(n_{1,1},\dots,n_{1,k})\in\mathcal{D}_{k,n_1}} \frac{\alpha(B_1)_{n_{1,1}\uparrow 1}\cdots\alpha(B_k)_{n_{1,k}\uparrow 1}}{(a)_{n_1\uparrow 1}} Q_{\alpha+\delta_X+\sum_{j=1}^{n_1}\delta_{Y_{1,j}}}\end{aligned}$$

where $(x)_{y\uparrow 1}$ stands for the Pochhammer symbol for the ascending factorial of x of order y (see Appendix A) and where for any finite measurable partiton $B_1,\dots,B_k$, $n_{1,i}=\sharp\{j: Y_{i,j}\in B_i\}$ for $i=1,\dots,k$ and $\sum_{1\le j\le k} n_{1,j}=n_1$ and $\mathcal{D}_{k,n_1} := \{(n_{1,1},\dots,n_{1,k})\in$

$\{1, \ldots, n_1\}^k : \sum_{i=1}^k n_{1,i} = n_1\}$. We consider now the r.v.'s $Z_{1,1}, \ldots, Z_{1,n_1}$ having the following distribution

$$\frac{1}{a+1}\delta_X + \frac{a}{a+1}\frac{\alpha(B_1)_{n_{1,1}\uparrow 1}\cdots\alpha(B_k)_{n_{1,k}\uparrow 1}}{(a)_{n_1\uparrow 1}}.$$

and let $\tilde{Q}_{\alpha+\delta_X+\sum_{1\leq j\leq n_1}\delta_{Z_{1,j}}}$ be a Dirichlet process on $\mathbb{X}$ with parameter given by $\alpha+\delta_X+\sum_{1\leq j\leq n_1}\delta_{Y_{1,j}}$. Then

$$\begin{aligned} P^{(n_\bullet)}|X &\stackrel{d}{=} \mathbb{E}[\tilde{Q}_{\alpha+\delta_X+\sum_{j=1}^{n_1}\delta_{Z_{1,j}}}] \\ &= 1 + \frac{a}{a+1}\sum_{(n_{1,1},\ldots,n_{1,k})\in\mathcal{D}_{k,n_1}}\frac{\alpha(B_1)_{n_{1,1}\uparrow 1}\cdots\alpha(B_k)_{n_{1,k}\uparrow 1}}{(a)_{n_1\uparrow 1}}Q_{\alpha+\delta_X+\sum_{j=1}^{n_1}\delta_{Y_{i,j}}} \end{aligned}$$

where $P^{(n_\bullet)}|X$ is a Dirichlet process with parameter $\alpha+\delta_X$. □

1.5 Discussion

The constructive definition of the Dirichlet process proposed by Sethuraman [174] was presented at an invited paper of an IMS meeting in 1980 and also announced in Sethuraman and Tiwari [173] which dealt with the convergence of Dirichlet processes. This definition has since been used by several authors to simplify previous calculations and to obtain new results involving Dirichlet processes. For instance, see Doss [32], Ferguson [63], Ferguson et al. [66] and Kummar and Tiwari [115].

More recently, Sethuraman's series representation of the Dirichlet process has been widely used in several areas of Bayesian nonparametric methods. In particular, it has been used to simulate the Dirichlet process (see Ishwaran and James [90], Ishwaran and Zarepour [92] and the reference therein), to find almost sure approximations and random approximations of the Dirichlet processes (see Ishwaran and Zarepour [94], Muliere and Tardella [140] and the reference therein) and to define new approaches for computing some functionals of the Dirichlet processes (see Diaconis and Kemperman [24], Guglielmi [80] and Guglielmi and Tweedie [81]).

Furthermore, the intuitive idea of the stick-breaking construction for the random weights in Sethuraman's series representation has made possible interesting generalizations of the Dirichlet process. We remind here the class of stick-breaking priors (see Ishwaran and James [90]), the classes of dependent and order-based dependent Dirichlet processes (see MacEachern [130], MacEachern [131] and Griffin and Steel [78]) and the class of spatial Dirichlet processes (see Gelfand et al. [73] and Duan et al. [34]).

In this chapter, moving from the constructive definition of the Dirichlet process proposed by Sethuraman [174]

$$P = \sum_{i \geq 1} p_i \delta_{Y_i}$$

we introduced the new constructive definition

$$P^{(n_\bullet)} = \sum_{i \geq 1} p_i \sum_{j=1}^{n_i} q_{i,j} \delta_{Y_{i,j}}$$

for any fixed integer-valued sequence $n_\bullet$, by nesting in the Sethuraman's series representation the r.p.m. $\sum_{1 \leq j \leq n_i} q_{i,j} \delta_{Y_{i,j}}$ (a random convex linear combination of r.v.s from a Blackwell-MacQueen Pólya sequence) instead of the r.p.m. δ_{Y_i} for $i \geq 1$. By its definition, the new series representation is a r.p.m. which gives probability one to the subset of discrete probability measures. Moreover, we proved that the finite dimensional marginal distribution for the new series representation $P^{(n_\bullet)}$ are Dirichlet distribution. Therefore, we proved that the new series representation is a new constructive definition of the Dirichlet process which includes Sethuraman's as a particular case. It also include the Blackwell and MacQueen result as a special case. Finally, following similar arguments to those one used in Sethuraman [174], we proved that the posterior process is also a Dirichlet process.

On the basis of the large number of applications of the Sethuraman's series representation, it seems natural to investigate the consequences of the new series representation in Bayesian nonparametric methods. In particular, in Chapter 2 we will focus on the application of the new series representation of the Dirichlet process to find more flexible approximations of the Dirichlet process and in Chapter 3 we will use the new series representation of the Dirichlet process in order to generalizes some approaches introduced in the literature for computing functionals of the Dirichlet processes.

2

Some remarks on Dirichlet process approximation via series truncations

In this chapter we provide some remarks on the approximation of the Dirichlet process. Moving from the generalized Sethuraman's series representation introduced in Chapter 1, we consider the application of two stopping procedures: a random stopping procedure which corresponds to the truncation of the series at a random number of terms and the almost sure stopping procedure which corresponds to the truncation of the series at a fixed number ot terms. In particular, the accuracy in these approximations is given with respect to the corresponding approximations obtained using the original Sethuraman series representation of the Dirichlet process. A straightforward extension of the random stopping procedure to the more general class of infinite dimensional stick-breaking random measures is considered. As a by-product, we also obtain some interesting results related to the convolution of distributions belonging to the class of generalized convolutions of mixtures of Exponential distributions.

2.1 Introduction

In this chapter we are interested in the series representation of the Dirichlet process proposed by Sethuraman [174] and in the generalization proposed in Section 1.3 of Chapter 1 and here recalled. Let $(\mathbb{X}, \mathcal{T})$ be the usual Polish space endowed with the Borel σ-field $\mathscr{X}$ and consider the following associated spaces of measures $\mathcal{A}_{\mathbb{X}}$ and $\mathcal{P}_{\mathbb{X}}$. In particular, $\mathcal{A}_{\mathbb{X}}$ is the space of locally finite non-negative measures on $(\mathbb{X}, \mathscr{X})$ endowed with the σ-field $\mathscr{A}_{\mathbb{X}}$ generated by the vague topology $\mathcal{V}$ which makes $(\mathcal{A}_{\mathbb{X}}, \mathcal{V})$ a Polish space, and $\mathcal{P}_{\mathbb{X}}$ is the space of probability measures on $(\mathbb{X}, \mathscr{X})$ endowed with its σ-field $\mathscr{P}_{\mathbb{X}}$ generated by the weak convergence topology $\mathcal{W}$ which makes $(\mathcal{P}_{\mathbb{X}}, \mathcal{W})$ a Polish space. Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure with total mass a, let $n_{\bullet} := \{n_i, i \geq 1\}$ be a fixed integer-valued sequence and let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space supporting three independent sequences of random variables (r.v.s) $\theta := \{\theta_i, i \geq 1\}$, $q := \{(q_{i,1}, \ldots, q_{i,n_i}), i \geq 1\}$ and $Y := \{(Y_{i,1}, \ldots, Y_{i,n_i}), i \geq 1\}$. The sequence θ is a sequence of independent r.v.s distributed according to a Beta distribution function with parameter (n_i, a), for $i \geq 1$, the sequence q is a sequence of independent r.v.s identically distributed according to a Dirichlet distribution function with parameter $(1, \ldots, 1)$ and the sequence Y is a sequence of independent r.v.s (samples of size n_i for $i \geq 1$) from a Blackwell-MacQueen Pólya sequence with parameter α, i.e. if P_i for $i \geq 1$ are independent Dirichlet processes with parameter α, then for any $i \geq 1$, $Y_{i,1}, \ldots, Y_{i,n_i} | P_i$ are independent and identically distributed (i.i.d.) from P_i. The condition of independence between the sequence of r.v.s θ, q and Y and the usual construction of a product measure implies the existence of the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ supporting the r.v.(θ, q, Y) and does not require any restrictions on $\mathbb{X}$, such as it being a Polish space. For any $B \in \mathscr{X}$ consider the measurable map $P^{(n_\bullet)} : \Omega \to \mathcal{P}_{\mathbb{X}}$ defined by

$$\omega \mapsto \sum_{i \geq 1} p_i(\omega) \sum_{j=1}^{n_i} q_{i,j}(\omega) \delta_{Y_{i,j}(\omega)}(B) \tag{2.1.1}$$

where $\{p_i, i \geq 1\}$ is a sequence of r.v.s obtained from the sequence of r.v.s θ by the usual stick-breaking construction, i.e. $p_1 = \theta_1$ and $p_i = \theta_i \prod_{1 \leq j \leq i-1} (1 - \theta_j)$ for $i \geq 2$. Then, in Section 1.3 of Chapter 1 it is proved that $P^{(n_\bullet)}$ in (2.1.1) is a random probability measure (r.p.m.) on $\mathbb{X}$ and in particular it is a Dirichlet process with parameter α. The measurable map (2.1.1) generalizes the measurable map used for the Sethuraman series representation which can be recoverd setting $n_\bullet = 1_\bullet$, where $1_\bullet$ is defined as a sequence of one. In particular, under the condition $n_\bullet = 1_\bullet$, for any $B \in \mathscr{X}$ (2.1.1) reduces to

$$\omega \mapsto \sum_{i \geq 1} p_i(\omega) \delta_{Y_i(\omega)}(B) \tag{2.1.2}$$

and it is proved by Sethuraman [174] that P in (2.1.2) is a r.p.m. on $\mathbb{X}$ and in particular P is a Dirichlet process with parameter α.

The Sethuraman series representation has been widely used in the literature in order to simplify previous calculations and to obtain new results involving the Dirichlet process. For instance, see Doss [32], Ferguson [63], Ferguson et al. [66] and Kummar and Tiwari [115]. More recently, the Sethuraman series representation has been used to simulate the Dirichlet process (see Ishwaran and James [90], Ishwaran and Zarepour [92] and the reference therein) and to find approximations of the Dirichlet process (see Muliere and Tardella [140], Ishwaran and Zarepour [94] and the reference therein).

From the Sethuraman series representation it is immediate to note that simulating the Dirichlet process requires the simulation of an infinite number of r.v.s. To avoid that, it is natural to construct new r.p.m.s which approximate, under an appropriate metric on $(\mathcal{P}_{\mathbb{X}}, \mathcal{W})$, the Dirichlet process. In Muliere and Tardella [140] a random stopping procedure for the Sethuraman series representation has been proposed in order to approximate the Dirichlet process. This procedure is based on stopping the Seturaman series representation at a random number of terms, assigning the remaining probability mass to a random point Y chosen independently according to the distribution $\alpha_0 := \alpha/a$. In particular, for any $\varepsilon \in (0,1]$ the ε-Dirichlet process with parameter (α, ε) is defined as the r.p.m. $P_\varepsilon : \Omega \to \mathcal{P}_{\mathbb{X}}$ such that for any $B \in \mathscr{X}$

$$\omega \mapsto \sum_{i=1}^{M_\varepsilon(\omega)} p_i(\omega)\delta_{Y_i(\omega)}(B) + R_\varepsilon(\omega)\delta_{Y_0(\omega)}(B) \tag{2.1.3}$$

where

$$M_\varepsilon := \inf\left\{m \in \mathbb{N}\colon \sum_{i=1}^{m} p_i > 1 - \varepsilon\right\}$$

$$R_\varepsilon := 1 - \sum_{i=1}^{M_\varepsilon} p_i$$

and Y_0 is a r.v. distributed according to α_0. As shown by Mulere and Tardella [140], an interesting feature of the ε-Dirichlet process is that under an appropriate metric on $(\mathcal{P}_{\mathbb{X}}, \mathcal{W})$, the random stopping procedure can fix in advance the closeness ε between the Dirichlet process and the ε-Dirichlet process. In particular, this appealing feature can also be extended to the distributions of several fanctionals of the Dirichlet process.

A different type of approximation for the Dirichlet process has been proposed by Ishwaran and Zarepour [94] (see also Ishwaran and Zarepour [93]) and it is still based on the truncation of the Sethuraman series representation of the Dirichlet proccess. Differently from the approach proposed by Muliere and Tardella [140], an almost sure stopping procedure for the Sethuraman series representation is proposed. In particular, for any $M \in \mathbb{N}$ the M-Dirichlet process with parameter (α, M) is defined as the r.p.m. $P_M : \Omega \to \mathcal{P}_{\mathbb{X}}$ such that for any $B \in \mathscr{X}$

$$\omega \mapsto \sum_{i=1}^{M} p_i(\omega)\delta_{Y_i(\omega)}(B) \tag{2.1.4}$$

where $p_M := 1 - \sum_{1\leq i\leq M-1} p_i$ to ensure that P_M is a well-defined r.p.m. As shown by Ishwaran and Zarepour [94] the closeness between the M-Dirichlet process and the Dirichlet process increase exponentially fast in M, and thus a moderate M should be able to achieve an accurate approximation.

In this chapter our aim is to apply the random stopping procedure and the almost sure stopping procedure to the generalized Sethuram series representation (2.1.1). As regard the application of the random stopping procedure to (2.1.1), we consider the definition of a new truncated r.p.m. which is the truncation of the series (2.1.1) at a random number of terms. The main problem in the application of the random stopping procedure to the more elaborated series representation of the Dirichlet process (2.1.1), is due to the characterization of the distribution of the random number of terms of the truncated r.p.m. As shown in Muliere and Tardella [140], for the random truncation of the Sethuraman series representation (2.1.1), the distribution of the random number of terms can be easily computed and it coincides with the Poisson distribution with parameter $-a\log(\varepsilon)$. Differently, for the truncation of the generalized Sethuraman series representation (2.1.1) the distribution of the random number of terms of the truncated r.p.m. has not always a simple expression and in general it involves the convolution of distributions belonging to the class of generalized convolutions of mixtures of Exponential distributions (see Bondesson [11] and Bondesson [12]). As regard the application of the almost sure stopping procedure to (2.1.1), we consider the definition of a new truncated r.p.m. which is the almost sure truncation of the series (2.1.1) at a fixed number of terms. Moving from these two new truncated series representation for the Dirichlet process, we consider their accuracy in the approximation of the Dirichlet process and we provide a comparison with the corresponding accuracy in the approximation of the Dirichlet process using the truncated series representations obtained by the Sethuramn series representation. Our goal is to verify if the more flexible series representation can be useful to obtain better approximations of the Dirichlet process.

Recently, the almost sure stopping procedure introduced by Ishwaran and Zarepour [94] (see also Ishwaran and Zarepour [93]) has been extetended by Ishwaran and James [90] to a more general class of r.p.m. which includes the Dirichlet process as particular case, the so-called class of infinite dimensional stick-breaking random measures. In particular, for the class of infinite dimensional stick-breaking random measures, in Ishwaran and James [90] is underlined the problem of finding an appropriate method for selecting the truncation level M which may be very large in order to obtain a reasonable approximation. After a brief introduction of the class of infinite dimensional stick-breaking random measures, we propose the application of the the random stopping procedure to the class of infinite dimensional stick-breaking random measures in in order to provide a solution to the problem of the selection of an appropriate truncation level M.

The chapter is structured as follow. In Section 2.2 we define a new random truncated Dirichlet process and we provide a comparison in terms of accuracy in the approximation with the ε-Dirichlet process. In particular, as a by-product, we also obtain some interesting results related to the convolution of distributions belonging to the class of generalized convolutions of mixtures of Exponential distributions. In Section 2.3 we define a new almost sure truncated Dirichlet process and we provide a comparison in terms of accuracy in the approximation with the M-Dirichlet process. In Section 2.4 we define the class of random truncated stick-breaking random measures. Section 2.5 is devoted to a discussion of the results.

2.2 The random stopping procedure

In this section, using the generalized series representation for the Dirichlet process (2.1.1), we define by its random truncation a new r.p.m. which approximates, under an appropriate metric on $(\mathcal{P}_{\mathbb{X}}, \mathcal{W})$, the Dirichlet process. This new r.p.m. generalizes the ε-Dirichlet process proposed by Muliere and Tardella [140]; in particular, a comparison between the new random approximation of the Dirichlet process and the ε-Dirichlet process is given.

Definition 2.2.1. *Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure and let $n_\bullet$ be a fixed integer-valued sequence. For any $\varepsilon \in (0,1]$, the generalized ε-Dirichlet processes with parameter $(\alpha, \varepsilon, n_\bullet)$ is the r.p.m. $P_\varepsilon^{(n_\bullet)} : \Omega \to \mathcal{P}_{\mathbb{X}}$ such that for any $B \in \mathscr{X}$*

$$\omega \mapsto \sum_{i=1}^{M_\varepsilon^{(n_\bullet)}(\omega)} p_i(\omega) \sum_{j=1}^{n_i} q_{i,j}(\omega) \delta_{Y_{i,j}(\omega)}(B) + R_\varepsilon^{(n_\bullet)}(\omega) \sum_{j=1}^{n_0} q_{0,j} \delta_{Y_{0,j}(\omega)}(B) \tag{2.2.1}$$

where

$$M_\varepsilon^{(n_\bullet)} := \inf\left\{ m \in \mathbb{N} : \sum_{i=1}^{m} p_i > 1 - \varepsilon \right\}$$

$$R_\varepsilon^{(n_\bullet)} := 1 - \sum_{i=1}^{M_\varepsilon^{(n_\bullet)}} p_i$$

and for $n_0 \in \mathbb{N}$, $(q_{0,1}, \ldots, q_{0,n_0})$ is a r.v. distributed according to a Dirichlet distribution function with parameter $(1, \ldots, 1)$ and $(Y_{0,1}, \ldots, Y_{0,n_0})$ is a r.v. from a Blackwell-MacQueen Pólya sequence with parameter α.

As we can see from Definition 2.2.1 the random stopping procedure follows the original idea proposed by Muliere and Tardella [140]. In particular, it is based on stopping the generalized series representation of the Dirichlet process at a random number of terms, assigning the remaining probability mass to n_0 random points $Y_{0,1}, \ldots, Y_{0,n_0}$ chosen idependently from a from a Blackwell-MacQueen Pólya sequence. As a consequence of the generalized series representation of the Dirichlet process, the definition of ε-Dirichlet process with parameter (α, ε) can be recoverd from Definiton 2.2.1 setting $n_\bullet = 1_\bullet$ in (2.2.1), i.e. the ε-Dirichlet process with parameter (α, ε) corresponds to the generalized ε-Dirichlet process with parameter $(\alpha, \varepsilon, 1_\bullet)$.

As for the ε-Dirichlet process, the Definition 2.2.1 implies the closeness, in the total variation metric ρ_V on $(\mathcal{P}_{\mathbb{X}}, \mathcal{W})$, between the generalized ε-Dirichlet processes and the

Dirichlet process, i.e. for any $\omega \in \Omega$ $\rho_V(P_\varepsilon^{(n\bullet)}(\omega,\cdot), P(\omega,\cdot)) \leq \varepsilon$ on a set of $\mathbb{P}$-probability 1. As regard the the closeness, in the Prokhorov metric[1] ρ_P on $(\mathcal{P}_{\mathbb{X}}, \mathcal{W})$, between the generalized ε-Dirichlet processes and the Dirichlet process we consider the following lemma.

Lemma 2.2.1. *For any $\omega \in \Omega$ and for any $\varepsilon \in (0,1]$*

$$\rho_P(P_\varepsilon^{(n\bullet)}(\omega,\cdot), P(\omega,\cdot)) \leq \varepsilon.$$

Proof. The proof follows by the same arguments used in Lemma 2 in Muliere and Tardella [140]. □

As for the ε-Dirichlet process, Lemma 2.2.1 implies that the random stopping procedure for the generalized series representation of the Dirichlet process (2.1.1) can fix in advance the closeness ε between the generalized ε-Dirichlet process and the Dirichlet process.

Now, moving from Definition 2.2.1 our aim is to investigate about the distribution of the stopping time r.v. $M_\varepsilon^{(n\bullet)}$ and provide a comparison with the Poisson distribution with parameter $-a\log(\varepsilon)$ obtained by Muliere and Tardella [140] for the stopping time of the ε-Dirichlet process. We write

$$R_\varepsilon^{(n\bullet)} = 1 - \sum_{i=1}^{M_\varepsilon^{(n\bullet)}} p_i = \prod_{i=1}^{M_\varepsilon^{(n\bullet)}} (1-\theta_i)$$

then the r.v. $M_\varepsilon^{(n\bullet)}$ can be written as

$$M_\varepsilon^{(n\bullet)} = \inf\left\{ m \in \mathbb{N} : \sum_{i=1}^{m} \log(1-\theta_i) < \log(\varepsilon) \right\}. \tag{2.2.2}$$

As we can see from (2.2.2) the r.v. $M_\varepsilon^{(n\bullet)}$ involves the distribution of the r.v. $\sum_{1\leq i\leq m}\log(1-\theta_i)$. In particular, if we find the distribution of $\sum_{1\leq i\leq m}\log(1-\theta_i)$, then using some arguments related to the point processes we can determine the disribution of $M_\varepsilon^{(n\bullet)}$. The distribution of the r.v. $-\log(1-\theta_i)$ is a known distribution belonging to the class of generalized convolutions of mixtures of Exponential distributions (see Bondesson [11] and Bondesson [12]) which is a class of distributions closed with respect to positive translation, change of scale, convolution and convolution roots. In particular, if θ_i is distributed according to a Beta distribution function with parameter $(1,a)$, then $-\log(1-\theta_i)$ is distributed according to an Exponential distribution function with parameter a. Here, we are interested in investigating the distribution of the r.v. $\sum_{1\leq i\leq m}\log(1-\theta_i)$ under the more general assumption that θ_i is distributed according to a Beta distribution function with

[1] The Prokhorov metric ρ_P on $(\mathcal{P}_{\mathbb{X}}, \mathcal{W})$ is a metric such that $\rho_P(P_n, P) \to 0$ as $n \to +\infty$ if and only if $P_n \Rightarrow P$ as $n \to +\infty$ (see Section 6 in Billingsley [8]).

parameter (n_i, a) for $i = 1, \dots, m$.

We first consider some general results about the convolution of m negative logarithm of r.v.s distributed according to a Beta distribution function with parameter (a_i, b_i) for $i = 1, \dots, m$. Some of the results of the next lemma are expressed in terms of the Gauss hypergeometric function ${}_2F_1$ (see Appendix C).

Lemma 2.2.2. *For any $m \in \mathbb{N}$, let $Y_1, \dots, Y_m$ be m independent r.v.s distributed according to a Beta distribution function with parameter (a_i, b_i) for $i = 1, \dots, m$. If $X^{*m} := \sum_{1 \le i \le m} -\log(1 - Y_i)$, then*

i) for any $z \ge 0$

$$\mathbb{P}(X^{*m} \le z) = \left(\Gamma\left(\sum_{k=1}^{m} a_k\right)\right)^{-1} \prod_{k=1}^{m} \frac{\Gamma(a_k + b_k)}{\Gamma(b_k)} \sum_{k \ge 0} \rho_k \int_{e^{-z}}^{1} y^{b_m - 1}(1-y)^{k + \sum_{k=1}^{m} a_k - 1} dy \tag{2.2.3}$$

where $\rho_0 := 1$ and

$$\rho_k := \sum_{(j_1,\dots,j_{m-1}) \in [k]_0^{m-1}} \prod_{i=1}^{m-1} \frac{(a_{i+1} - d_i)_{j_i \uparrow 1}(a_i + \sum_{l=1}^{i-1} a_l + j_l)_{j_i \uparrow 1}}{j_i!(\sum_{k=1}^{m} a_k + \sum_{l=1}^{i-1} j_l)_{j_i \uparrow 1}} \qquad k \ge 1$$

where $(x)_{y\uparrow 1}$ stands for the Pochhammer symbol for the ascending factorial of x of order y (see Appendix A) and where $d_i := b_i - b_{i+1}$ for $i = 1, \dots, m-1$;

ii) if $a_i, b_i \in \mathbb{N}$ for $i = 1, \dots, m$ for any $z \ge 0$

$$\mathbb{P}(Y^{*m} \le z) = \sum_{k=1}^{n} \sum_{j=0}^{e_k - 1} \int_{e^{-z}}^{1} \frac{K_{k,j} x^{d_k - 1}(-\log(x))^{e_k - j - 1}}{(e_k - j - 1)! j!} dx \tag{2.2.4}$$

where

$$K_{k,0} := \sum_{q \ne k} (d_q - d_k)^{-e_q}$$

and

$$K_{k,j} := \sum_{r=0}^{j-1} \sum_{q \ne k} (-1)^{r+1} \binom{j-1}{r} \frac{r! K_{k,j-r-1}}{(d_q - d_k)^{r+1}} \qquad j = 1, \dots, e_k - 1$$

where d_k denotes the n different integers that occour with multiplicity e_k among $b_i - 1$, b_i, $b_i + 1, \dots, a_i + b_i - 2$ for $i = 1, \dots, m$;

iii) if $a_i \in \mathbb{N}$ for $i = 1, \ldots, m$ for any $z \geq 0$

$$\begin{aligned}
&\mathbb{P}(X^{*m} \leq z) \qquad (2.2.5)\\
&= \prod_{k=1}^{m} \frac{\Gamma(a_k + b_k)}{\Gamma(a_k)\Gamma(b_k)} \sum_{(i_1,\ldots,i_{m-1}) \in \times_{k=1}^{m-1}[a_k-1]_0} \left(\prod_{k=1}^{m-1} \binom{a_{k+1}-1}{i_k} (-1)^{i_k} (a_k - i_{k-1} + i_k)^{-1} \right)\\
&\times \int_{e^{-z}}^{1} x^{b_m - 1}(1-x)^{\sum_{k=1}^{m} a_k - 1}\\
&\qquad \times \prod_{k=1}^{m-1} {}_2F_1(a_k - i_{k-1} + i_k, -d_k + a_{k+1}; a_k - i_{k-1} + i_k + 1; 1-x)dx
\end{aligned}$$

where $d_k := b_k - b_{k+1}$ for $k = 1, \ldots, m-1$;

iv) if $b_{i+1} = b_i + a_i$ for $i = 1, \ldots, m-1$ for any $z \geq 0$

$$\mathbb{P}(X^{*m} \leq z) = \frac{\Gamma(b_1 + \sum_{k=1}^{m} a_k)}{\Gamma(b_1)\Gamma(\sum_{k=1}^{m} a_k)} \int_{e^{-z}}^{1} x^{b_1 - 1}(1-x)^{\sum_{k=1}^{m} a_k - 1} dx. \qquad (2.2.6)$$

Proof. As regard the point i), for any $z \geq 0$ we have

$$\mathbb{P}(X^{*m} \leq z) = \mathbb{P}\left(-\sum_{i=1}^{m}(\log(1-Y_i)) \leq z \right) = \mathbb{P}\left(\prod_{i=1}^{m}(1-Y_i) \geq \mathrm{e}^{-z} \right)$$

which is the probability that the product of m independent r.v.s $(1-Y_1), \ldots, (1-Y_m)$ distributed according to a Beta distribution function with parameter (b_i, a_i), for $i = 1, \ldots, m$ is greater than e^{-z}. An explicit expression for the density function of the product of m independent r.v.s distributed according to a Beta distribution function with parameter (b_i, a_i) was provided by Nandi [143] and later by Tang and Gupta [178]. Then, the result follows.

As regard the point ii) it follows by the same arguments of point i). In particular we need to evaluate the probability that the product of m independent r.v.s $(1-Y_1), \ldots, (1-Y_m)$ distributed according to a Beta distribution function with parameter (b_i, a_i) with $a_i, b_i \in \mathbb{N}$ for $i = 1, \ldots, m$ is greater than e^{-z}. An explicit expression for the density function of the product of m independent r.v.s distributed according to a Beta distribution function with parameter (b_i, a_i) with $a_i, b_i \in \mathbb{N}$ for $i = 1, \ldots, m$ was provided by Springer and Thompson [176]. Then, the result follows.

As regard the point iii), for any $z \geq 0$ we have

$$\mathbb{P}(X^{*m} \leq z) = \mathbb{P}\left(-\sum_{i=1}^{m}(\log(1-Y_i)) \leq z \right) = \mathbb{P}\left(\prod_{i=1}^{m}(1-Y_i) \geq \mathrm{e}^{-z} \right)$$

which is the probability that the product of m independent r.v.s $(1-Y_1), \ldots, (1-Y_m)$ distributed according to a Beta distribution function with parameter (b_i, a_i) with $a_i \in \mathbb{N}$ is greater than e^{-z}. We consider a general expression for the density function of the product of m independent r.v.s distributed according to a Beta distribution function with parameter (b_i, a_i) for $i = 1, \ldots, m$ provided by Wilks [197] in terms of a multiple integral

$$\begin{aligned}
\mathbb{P}(X^{*m} \leq z) &= \prod_{k=1}^{m} \frac{\Gamma(a_k + b_k)}{\Gamma(a_k)\Gamma(b_k)} \int_{e^{-z}}^{1} x^{b_m - 1}(1-x)^{\sum_{k=1}^{m} a_k - 1} \\
&= \int_{(0,1)^{m-1}} \prod_{k=1}^{m-1} w_k^{a_k - 1}(1-\omega_k)^{\sum_{i=k+1}^{m} a_k - 1} \\
&\qquad \times \left(1 - (1-x)\left(1 - \prod_{i=1}^{k}(1-\omega_i)\right)\right)^{b_k - b_{k+1} - a_{k+1}} d\omega_k dx.
\end{aligned}$$

First, we consider the change of variable $v_k = (1 - \prod_{1 \leq i \leq k}(1-\omega_i))$ for $k = 1, \ldots, m-1$. It can be easily checked that the absolute value of the determinant of the Jacobian matrix is $\prod_{1 \leq k \leq m-2}(1-v_i)^{-1}$. Second, since $|\arg(1 + (-1+x))| < \pi$, we can then use Equation 3.194.1 in Gradshteyn and Ryzhik [77] to solve the multiple integral. In particular, using the Binomial theorem we have

$$\begin{aligned}
\mathbb{P}(X^{*m} \leq z) &= \prod_{k=1}^{m} \frac{\Gamma(a_k + b_k)}{\Gamma(a_k)\Gamma(b_k)} \int_{e^{-z}}^{1} x^{b_m - 1}(1-x)^{\sum_{k=1}^{m} a_k - 1} \\
&\qquad \times \int_{(0,1)^{m-1}} \prod_{k=1}^{m-1} v_k^{a_k - i_{k-1} + i_k - 1}(1 - (1-x)v_k)^{b_k - b_{k+1} - a_{k+1}} dv_k dx \\
&\qquad \times \sum_{(i_1, \ldots, i_{m-1}) \in \times_{k=1}^{m-1}[a_k - 1]_0} \prod_{k=1}^{m-1} \binom{a_{k+1} - 1}{i_k}(-1)^{i_k} \\
&= \prod_{k=1}^{m} \frac{\Gamma(a_k + b_k)}{\Gamma(a_k)\Gamma(b_k)} \int_{e^{-z}}^{1} x^{b_m - 1}(1-x)^{\sum_{k=1}^{m} a_k - 1} \\
&\qquad \times \sum_{(i_1, \ldots, i_{m-1}) \in \times_{k=1}^{m-1}[a_k - 1]_0} \prod_{k=1}^{m-1} \binom{a_{k+1} - 1}{i_k}(-1)^{i_k} \\
&\qquad \times \int_{(0,1)^{m-1}} \prod_{k=1}^{m-1} v_k^{a_k - i_{k-1} + i_k - 1}(1 + (-1+x)v_k)^{b_k - b_{k+1} - a_{k+1}} dv_k dx \\
&= \prod_{k=1}^{m} \frac{\Gamma(a_k + b_k)}{\Gamma(a_k)\Gamma(b_k)} \int_{e^{-z}}^{1} x^{b_m - 1}(1-x)^{\sum_{k=1}^{m} a_k - 1} \\
&\qquad \times \sum_{(i_1, \ldots, i_{m-1}) \in \times_{k=1}^{m-1}[a_k - 1]_0} \prod_{k=1}^{m-1} \binom{a_{k+1} - 1}{i_k}(-1)^{i_k}(a_k - i_{k-1} + i_k)^{-1} \\
&\qquad \times {}_2F_1(a_k - i_{k-1} + i_k, -b_k + b_{k+1} + a_{k+1}; a_k - i_{k-1} + i_k + 1; 1-x) dx.
\end{aligned}$$

As regard the point iv) it follows by the same arguments of point i). In particular we need to evaluate the probability that the product of m independent r.v.s $(1-Y_1),\dots,(1-Y_m)$ distributed according to a Beta distribution function with parameter (b_i,a_i) with $b_{i+1}=b_i+a_i$ for $i=1,\dots,m-1$ is greater than e^{-z}. It is known from Jambunathan [95] that product of m independent r.v.s $(1-Y_1),\dots,(1-Y_m)$ with parameter (b_i,a_i) with $b_{i+1}=b_i+a_i$ for $i=1,\dots,m-1$ is still distributed according to a Beta distribution function with parameter $(b_1,\sum_{1\le k\le m}a_i)$. □

As corollary of Lemma 2.2.2 we provide a result related to the convolution of Exponential distributions. In particular, given $m\in\mathbb{N}$ independent r.v.s $X_1,\dots,X_m$ distributed according to an Exponential distribution function with parameter b_i for $i=1,\dots,m$, we consider the distribution of the r.v. $X^{*m}=\sum_{1\le i\le m}X_i$ which correspond to the distribution computed in point iii) of Lemma 2.2.2 under the condition $a_i=1$ for $i=1,\dots,m$. The convolution of Exponential distributions and its importance in applications is known in the literature and it has been recently reviewed by Nadarajah [142]. Here, we provide a different approach to compute the convolution of Exponential distributions and in particular we provide a more simple expression for it.

Corollary 2.2.1. *For any $m\in\mathbb{N}$, let $X_1,\dots,X_m$ be m independent r.v.s distributed according to an Exponential distribution function with parameter b_i for $i=1,\dots,m$. If $X^{*m}:=\sum_{1\le i\le m}X_i$, then*

i) for any $z\ge 0$

$$\mathbb{P}(X^{*m}\le z)=b_m\left(\prod_{k=1}^{m-1}b_k(-b_k(-1)^{\mathbb{1}_{\{d_k\neq 0\}}}-b_{k+1})^{-\mathbb{1}_{\{d_k\neq 0\}}}\right)$$
$$\times\sum_{(i_1,\dots,i_{m-1})\in\times_{k=1}^{m-1}[\mathbb{1}_{\{d_k\neq 0\}}]_0}\left(\prod_{k=1}^{m-1}\binom{\mathbb{1}_{\{d_k\neq 0\}}}{i_k}(-1)^{-i_k}\right)\rho_{(i_1,\dots,i_{m-1})}(z)$$

where, if $b_m+\sum_{1\le k\le m-1}i_kd_k\neq 0$

$$\rho_{(i_1,\dots,i_{m-1})}(z):=\left(b_m+\sum_{k=1}^{m-1}i_kd_k\right)^{-\sum_{k=1}^{m-1}\mathbb{1}_{\{d_k=0\}}-1}$$
$$\times\gamma\left(1+\sum_{k=1}^{m-1}\mathbb{1}_{\{d_k=0\}},z\left(b_m+\sum_{k=1}^{m-1}i_kd_k\right)\right)$$

and if $b_m+\sum_{1\le k\le m-1}i_kd_k=0$

$$\rho_{(i_1,\dots,i_{m-1})}(z):=\frac{z^{1+\sum_{k=1}^{m-1}\mathbb{1}_{\{d_k=0\}}}}{1+\sum_{k=1}^{m-1}\mathbb{1}_{\{d_k=0\}}}$$

where $d_k:=b_k-b_{k+1}$ for $k=1,\dots,m-1$ and where in general $\gamma(x,y):=\int_0^y t^{x-1}e^{-t}dt$ is the lower incomplete Gamma function;

ii) if $b = b_i$ for $i = 1, \ldots, m$ for any $z \geq 0$

$$\mathbb{P}(X^{*m} \leq z) = \frac{1}{(m-1)!}\gamma(m, bz)$$

where in general $\gamma(x,y) := \int_0^y t^{x-1}e^{-t}dt$ is the lower incomplete Gamma function;

iii) if $b_i \neq b_j$ for any $i \neq j$ for any $z \geq 0$

$$\mathbb{P}(X^{*m} \leq z) = \prod_{k=1}^{m-1} b_k \sum_{k=1}^{m} \frac{1-e^{-b_k z}}{b_k \prod_{j\neq k}(b_j - b_k)}.$$

Proof. As regard the point i), for any $z \geq 0$

$$\begin{aligned}\mathbb{P}(X^{*m} \leq z) &= \mathbb{P}\left(\sum_{k=1}^{m} X_i \leq z\right) \\ &= \mathbb{P}\left(-\sum_{i=1}^{m}(\log(1-Y_i)) \leq z\right) = \mathbb{P}\left(\prod_{i=1}^{m}(1-Y_i) \geq \mathrm{e}^{-z}\right)\end{aligned}$$

where $Y_1, \ldots, Y_m$ are m independent r.v.s distributed according to a Beta distribution function with parameter $(1, b_i)$. Then, following the approach used in Lemma 2.2.2 we have

$$\begin{aligned}P(X^{*m} \leq z) &= \prod_{k=1}^{m} b_k \int_{\mathrm{e}^{-z}}^{1} x^{b_m-1}(1-x)^{m-1}\int_{(0,1)^{m-1}}\prod_{k=1}^{m-1}(1-(1-x)v_k)^{b_k-b_{k+1}-1}dv_k dx \\ &= \prod_{k=1}^{m} b_k \int_{\mathrm{e}^{-z}}^{1} x^{b_m-1}(1-x)^{m-1}\prod_{k=1}^{m-1}\int_0^1(1+(-1+x)v_k)^{b_k-b_{k+1}-1}dv_k dx \\ &= \prod_{k=1}^{m} b_k \int_{\mathrm{e}^{-z}}^{1} x^{b_m-1}(1-x)^{m-1}\prod_{k=1}^{m-1}{}_2F_1(1, -b_k+b_{k+1}+1; 2; 1-x)dx\end{aligned}$$

We can write the last equation as

$$\begin{aligned}&\prod_{k=1}^{m} b_k \int_{\mathrm{e}^{-z}}^{1} x^{b_m-1}(1-x)^{m-1}\prod_{k=1}^{m-1}\left(-\frac{\log(x)}{1-x}\right)^{\mathbb{1}_{\{b_k=b_{k+1}\}}} \\ &\qquad\times\prod_{k=1}^{m-1}\left(\frac{x^{b_k-b_{k+1}}-1}{(1-x)(b_k(-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}+b_{k+1})}\right)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}} dx \\ &= \prod_{k=1}^{m} b_k \int_{\mathrm{e}^{-z}}^{1} x^{b_m-1}(1-x)^{m-1} \\ &\qquad\times\prod_{k=1}^{m-1}\frac{(-1)^{\mathbb{1}_{\{b_k=b_{k+1}\}}}(x^{b_k-b_{k+1}}-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}}{(1-x)(b_k(-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}+b_{k+1})^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}(\log(x))^{-\mathbb{1}_{\{b_k=b_{k+1}\}}}}dx\end{aligned}$$

$$
\begin{aligned}
&= \prod_{k=1}^{m} b_k \int_{\mathrm{e}^{-z}}^{1} x^{b_m-1}(-1)^{m-1} \\
&\quad \times \prod_{k=1}^{m-1} \frac{(x^{b_k-b_{k+1}}-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}}{(-b_k(-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}-b_{k+1})^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}(\log(x))^{-\mathbb{1}_{\{b_k=b_{k+1}\}}}}dx \\
&= (-1)^{m-1}\int_{\mathrm{e}^{-z}}^{1} b_m x^{b_m-1}\prod_{k=1}^{m-1} b_k(-b_k(-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}-b_{k+1})^{-\mathbb{1}_{\{b_k\neq b_{k+1}\}}} \\
&\quad \times \prod_{k=1}^{m-1}(x^{b_k-b_{k+1}}-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}(\log(x))^{\mathbb{1}_{\{b_k=b_{k+1}\}}}dx \\
&= (-1)^{m-1}b_m\int_{\mathrm{e}^{-z}}^{1} x^{b_m-1}\prod_{k=1}^{m-1} b_k(-b_k(-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}-b_{k+1})^{-\mathbb{1}_{\{b_k\neq b_{k+1}\}}} \\
&\quad \times \sum_{(i_1,\ldots,i_{m-1})\in\times_{k=1}^{m-1}[\mathbb{1}_{\{d_k\neq 0\}}]_0}\prod_{k=1}^{m-1}\binom{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}{i_k}(-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}-i_k} \\
&\quad \times x^{i_k(b_k-b_{k+1})}(\log(x))^{\mathbb{1}_{\{b_k=b_{k+1}\}}}dx \\
&= (-1)^{m-1}b_m\prod_{k=1}^{m-1} b_k(-b_k(-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}-b_{k+1})^{-\mathbb{1}_{\{b_k\neq b_{k+1}\}}} \\
&\quad \times \sum_{(i_1,\ldots,i_{m-1})\in\times_{k=1}^{m-1}[\mathbb{1}_{\{d_k\neq 0\}}]_0}\left(\prod_{k=1}^{m-1}\binom{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}{i_k}(-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}-i_k}\right) \\
&\quad \times \int_{\mathrm{e}^{-z}}^{1} x^{b_m-1+\sum_{k=1}^{m-1}i_k(b_k-b_{k+1})}(\log(x))^{\sum_{k=1}^{m-1}\mathbb{1}_{\{b_k=b_{k+1}\}}}dx
\end{aligned}
$$

Then $\mathbb{P}(X>\mathrm{e}^{-z})$ can be computed via a change of variable and using Equation 3.381.1 in [77]. In particular, we have

$$
\begin{aligned}
\mathbb{P}(X>\mathrm{e}^{-z}) &= (-1)^{m-1}b_m\prod_{k=1}^{m-1} b_k(-b_k(-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}-b_{k+1})^{-\mathbb{1}_{\{b_k\neq b_{k+1}\}}} \\
&\quad \times \sum_{(i_1,\ldots,i_{m-1})\in\times_{k=1}^{m-1}[\mathbb{1}_{\{d_k\neq 0\}}]_0}\left(\prod_{k=1}^{m-1}\binom{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}{i_k}(-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}-i_k}\right) \\
&\quad \times \int_{\mathrm{e}^{-z}}^{1} x^{b_m-1+\sum_{k=1}^{m-1}i_k(b_k-b_{k+1})}(\log(x))^{\sum_{k=1}^{m-1}\mathbb{1}_{\{b_k=b_{k+1}\}}}dx \\
&= (-1)^{m-1}b_m\prod_{k=1}^{m-1} b_k(-b_k(-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}-b_{k+1})^{-\mathbb{1}_{\{b_k\neq b_{k+1}\}}} \\
&\quad \times \sum_{(i_1,\ldots,i_{m-1})\in\times_{k=1}^{m-1}[\mathbb{1}_{\{d_k\neq 0\}}]_0}\left(\prod_{k=1}^{m-1}\binom{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}{i_k}(-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}-i_k}\right)
\end{aligned}
$$

$$\times (-1)^{\sum_{k=1}^{m-1} \mathbb{1}_{\{b_k=b_{k+1}\}}} \int_0^z \mathrm{e}^{-t(b_m+\sum_{k=1}^{m-1} i_k(b_k-b_{k+1}))} t^{\sum_{k=1}^{m-1} \mathbb{1}_{\{b_k=b_{k+1}\}}} dt$$

$$= (-1)^{m-1} b_m \prod_{k=1}^{m-1} b_k(-b_k(-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}} - b_{k+1})^{-\mathbb{1}_{\{b_k\neq b_{k+1}\}}}$$

$$\times \sum_{(i_1,\dots,i_{m-1})\in \times_{k=1}^{m-1}[\mathbb{1}_{\{d_k\neq 0\}}]_0} \left(\prod_{k=1}^{m-1} \binom{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}{i_k} (-1)^{1-i_k} \right)$$

$$\times \int_0^z \mathrm{e}^{-t(b_m+\sum_{k=1}^{m-1} i_k(b_k-b_{k+1}))} t^{\sum_{k=1}^{m-1} \mathbb{1}_{\{b_k=b_{k+1}\}}} dt$$

$$= b_m \prod_{k=1}^{m-1} b_k(-b_k(-1)^{\mathbb{1}_{\{b_k\neq b_{k+1}\}}} - b_{k+1})^{-\mathbb{1}_{\{b_k\neq b_{k+1}\}}}$$

$$\times \sum_{(i_1,\dots,i_{m-1})\in \times_{k=1}^{m-1}[\mathbb{1}_{\{d_k\neq 0\}}]_0} \left(\prod_{k=1}^{m-1} \binom{\mathbb{1}_{\{b_k\neq b_{k+1}\}}}{i_k} (-1)^{-i_k} \right)$$

$$\times \int_0^z \mathrm{e}^{-t(b_m+\sum_{k=1}^{m-1} i_k(b_k-b_{k+1}))} t^{\sum_{k=1}^{m-1} \mathbb{1}_{\{b_k=b_{k+1}\}}} dt$$

where, in particular

$$\int_0^z \mathrm{e}^{-t(b_m+\sum_{k=1}^{m-1} i_k(b_k-b_{k+1}))} t^{\sum_{k=1}^{m-1} \mathbb{1}_{\{b_k=b_{k+1}\}}} dt$$

$$= \left(b_m + \sum_{k=1}^{m-1} i_k(b_k - b_{k+1}) \right)^{-\sum_{k=1}^{m-1} \mathbb{1}_{\{b_k=b_{k+1}\}}-1}$$

$$\times \gamma \left(1 + \sum_{k=1}^{m-1} \mathbb{1}_{\{b_k=b_{k+1}\}}, z \left(b_m + \sum_{k=1}^{m-1} i_k(b_k - b_{k+1}) \right) \right)$$

if $b_m + \sum_{1\leq k\leq m-1} i_k(b_k - b_{k+1}) \neq 0$ and

$$\int_0^z \mathrm{e}^{-t(b_m+\sum_{k=1}^{m-1} i_k(b_k-b_{k+1}))} t^{\sum_{k=1}^{m-1} \mathbb{1}_{\{b_k=b_{k+1}\}}} dt = \frac{z^{1+\sum_{k=1}^{m-1} \mathbb{1}_{\{b_k=b_{k+1}\}}}}{1+\sum_{k=1}^{m-1} \mathbb{1}_{\{b_k=b_{k+1}\}}}$$

if $b_m + \sum_{1\leq k\leq m-1} i_k(b_k - b_{k+1}) = 0$, where in general $\gamma(x, y) := \int_0^y t^{x-1}\mathrm{e}^{-t} dt$ is the lower incomplete Gamma function. Then the result follows defining $d_k := b_k - b_{k+1}$ for $k = 1, \dots, m-1$ and by substitution.

The point ii) can be easily obtained as a special case of i). In particular it is known that the r.v. X^{*m} is distributed according to the Erlang distribution function with parameter (m, b).

The point iii) can be easily obtained as a special case of iii). However, we provide an alternative proof by induction following the hint in Problem 12 of Chapter 1 in Feller [60]. In particular, Let $f_{Y_i}(y_i)$ be the density function induced by the Exponential distribution function with parameter b_i. We prove the statement by induction on m. In particular, it can be easily checked that for $m = 2$

$$f_{Y^{*2}}(y) = b_1 b_2 \left(\frac{\mathrm{e}^{-b_1 y}}{b_2 - b_1} + \frac{\mathrm{e}^{-b_2 y}}{b_1 - b_2} \right).$$

Now, fix $m \geq 3$, and assume the statement of the lemma is true for $m-1$. Then we have

$$\begin{aligned} f_{Y^{*m}}(y) &= \prod_{i=1}^{m-1} b_i \sum_{j=1}^{m-1} \frac{\mathrm{e}^{-b_j y}}{\prod_{k \neq j}(b_k - b_j)} \\ &= \prod_{i=1}^{m} \sum_{j=1}^{m-1} \frac{\mathrm{e}^{-b_m} y - \mathrm{e}^{-b_j} y}{(b_j - b_m) \prod_{k \neq j}(b_k - b_j)} = \prod_{i=1}^{m} \left(\sum_{j=1}^{m-1} \frac{\mathrm{e}^{-b_j} y}{\prod_{k \neq j}(b_k - b_j)} - \frac{\mathrm{e}^{-b_m} y}{\prod_{k \neq j}(b_k - b_j)} \right). \end{aligned}$$

Then the proof is completed if we show that the coefficient of $\mathrm{e}^{-b_m y}$ in the last equation fits the coefficients $1/\prod_{1 \leq k \leq n-1}(b_k - b_m)$ i.e. or, equivalently, $-\sum_{1 \leq j \leq m} 1/\prod_{k \neq j}(b_k - b_j) = 0$. Then, we write

$$\sum_{j=1}^{m} \frac{1}{\prod_{k \neq j}(b_k - b_j)} = \sum_{j=1}^{m} \frac{\prod_{k \neq l \neq j}(b_k - b_j)}{\prod_{k \neq l}(b_k - b_l)}$$

which is zero if and only if $\sum_{1 \leq j \leq m} \prod_{k \neq l \neq j}(b_k - b_j)(b_k - b_l) = 0$. In particular, from the last equation we obtain

$$\begin{aligned} \sum_{j=1}^{m} \prod_{k \neq l \neq j} (b_k - b_j)(b_k - b_l) &= \sum_{j=1}^{m} \prod_{j \neq k \neq l \neq j} (b_k - b_l) \prod_{k = j \neq l} (b_k - b_l) \\ &= \pm \sum_{j=1}^{m} \prod_{j \neq k > l \neq j} (b_k - b_l)^2 \prod_{k = j > l} (b_k - b_l) \prod_{k = j < l} (b_k - b_l) \\ &= \pm \sum_{j=1}^{m} \prod_{j \neq k > l \neq j} (b_k - b_l)^2 \prod_{j = k > l} (b_k - b_l) \prod_{k > l = j} (b_k - b_l)(-1)^{n-j} \\ &= \pm \prod_{k > l}(b_k - b_l) \sum_{j=1}^{m} \prod_{j \neq k > l \neq j} (b_k - b_l)(-1)^{n-j} \end{aligned}$$

which is zero if and only if $\sum_{1 \leq j \leq m} \prod_{j \neq k > l \neq j}(b_k - b_l)(-1)^j = 0$. In particular, we observe that the product $\prod_{j \neq k > l \neq j}(b_k - b_l)(-1)^j$ is a Vandermonde determinant of

$$\begin{pmatrix} 1 & b_1 & b_1^2 & \cdots & b_1^{n-2} \\ 1 & b_2 & b_2^2 & \cdots & b_2^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & b_{j-1} & b_{j-1}^2 & \cdots & b_{j-1}^{n-2} \\ 1 & b_{j+1} & b_{j+1}^2 & \cdots & b_{j+1}^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & b_n & b_n^2 & \cdots & b_n^{n-2} \end{pmatrix}$$

and hence the last equation is nothing else but the expansion of the determinant of

$$\begin{pmatrix} 1 & 1 & b_1 & b_1^2 & \cdots & b_1^{n-2} \\ 1 & 1 & b_2 & b_2^2 & \cdots & b_2^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & b_n & b_n^2 & \cdots & b_n^{n-2} \end{pmatrix}$$

with respect to its second column. Since this determinant is zero, then the statement is proven. □

Lemma 2.2.2 provide a general results on the convolution of m negative logarithm of r.v.s distributed according to a Beta distribution function with parameter (a_i, b_i) for $i = 1, \dots, m$ under some different assumptions on the parameter (a_i, b_i) for $i = 1, \dots, m$. For our purpose we are interested in the distribution of $\theta^{*m} := -\sum_{1 \le k \le m} \log(1-\theta_i)$ when θ_i is distributed according to a Beta distribution function with parameter (n_i, a) for $i = 1, \dots, m$, which corresponds to (2.2.5) under the condition $a_i = a$ for $i = 1, \dots, m$. In particular, we observe equation (2.2.5) can be further simplified by the application of the distant neighbors property of the Gauss hypergeometric function. For any $m \in \mathbb{N}$ and for any vector $(i_1, \dots, i_{m-1})$ if we define $n_k := a_k - i_{k-1} + i_k$ for $k = 1, \dots, m$, then

$$\begin{aligned}
&\prod_{k=1}^{m-1} {}_2F_1(n_k, -d_k + a_{k+1}; n_k + 1; 1 - x) \\
&\quad = \prod_{k=1}^{m-1} \frac{(-n_k)_{(n_k-1)\uparrow 1}(1-x)^{-(n_k-1)}}{(1-n_k)_{(n_k-1)\uparrow 1}} \\
&\qquad \times \sum_{(l_1,\dots,l_{m-1}) \in \times_{k=1}^{m-1}[n_k-1]_0} \prod_{k=1}^{m-1} (-1)^{l_k} \binom{n_k - 1}{l_k} {}_2F_1(1, -d_k + a_{k+1} - l_k; 2; 1 - x).
\end{aligned}$$

Then, following the same arguments used in the proof of point i) of Corollary 2.2.1, it can be easily checked that for any $z \ge 0$

$$\begin{aligned}
&\mathbb{P}(\theta^{*m} \le z) \qquad (2.2.7) \\
&\quad = \prod_{k=1}^{m} \frac{\Gamma(a+n_k)}{\Gamma(a)\Gamma(n_k)} \sum_{(i_1,\dots,i_m) \in \times_{k=1}^{m-1}[n_k-1]_0} \left(\prod_{k=1}^{m} \binom{n_k - 1}{i_k} (-1)^{-i_k} \right) \\
&\qquad \times (a + i_m) \left(\prod_{k=1}^{m-1} (a+i_k)(-(a+i_k)(-1)^{\mathbb{1}\{i_k - i_{k+1} \neq 0\}} - (a + i_{k+1}))^{-\mathbb{1}\{i_k - i_{k+1} \neq 0\}} \right) \\
&\qquad \times \sum_{(j_1,\dots,j_{m-1}) \in \times_{k=1}^{m-1}[i_k - i_{k+1} \neq 0]_0} \left(\prod_{k=1}^{m-1} \binom{\mathbb{1}\{i_k - i_{k+1} \neq 0\}}{j_k} (-1)^{-j_k} \right) \rho_{(j_1,\dots,j_{m-1})}(z)
\end{aligned}$$

where, if $a+i_m+\sum_{1\le k\le m-1} j_k(i_k-i_{k+1})\neq 0$

$$\rho_{(j_1,\dots,j_{m-1})}(z)=\left(a+i_m+\sum_{k=1}^{m-1} i_k(i_k-i_{k+1}))\right)^{-\sum_{k=1}^{m-1}\mathbb{1}_{\{i_k-i_{k+1}=0\}}-1}$$
$$\times\gamma\left(1+\sum_{k=1}^{m-1}\mathbb{1}_{\{i_k-i_{k+1}=0\}},z\left(a+i_m+\sum_{k=1}^{m-1} i_k i_k-i_{k+1})\right)\right)$$

and if $a+i_m+\sum_{1\le k\le m-1} j_k(i_k-i_{k+1})=0$

$$\rho_{(j_1,\dots,j_{m-1})}(z)=\frac{z^{1+\sum_{k=1}^{m-1}\mathbb{1}_{\{i_k-i_{k+1}=0\}}}}{1+\sum_{k=1}^{m-1}\mathbb{1}_{\{i_k-i_{k+1}=0\}}}.$$

Using the distribution of the convolution of m negative logarithm of r.v.s distributed according to a Beta distribution function with parameter (n_i,a) for $i=1,\dots,m$ and some arguments related to the point processes we obtain the following result which generalizes Lemma 3 in [140].

Lemma 2.2.3. *For any $\varepsilon\in(0,1]$, $M_\varepsilon^{(n_\bullet)}$ is a r.v. on $\mathbb{N}$ such that*

$$\mathbb{P}(M_\varepsilon^{(n_\bullet)}=m)=\begin{cases}\mathbb{P}(\theta^{*m}\le-\log(\varepsilon)) & m=1\\ \mathbb{P}(\theta^{*m}\le-\log(\varepsilon))-\mathbb{P}(\theta^{*(m+1)}\le-\log(\varepsilon) & m>1\end{cases}\tag{2.2.8}$$

where $\mathbb{P}(\theta^{*m}\le z)$ is given by (2.2.7).

Proof. Using the definition of the random stopping time given by (2.2.2) we define the following point process

$$M_\varepsilon^{(n_\bullet)}(-\log(\varepsilon)):=\inf\left\{m\in\mathbb{N}_0:\sum_{i=1}^{m}\log(1-\theta_i)<\log(\varepsilon)\right\}.$$

In particular, we have $M_\varepsilon^{(n_\bullet)}(-\log(\varepsilon))=M_\varepsilon^{(n_\bullet)}-1$, i.e. $M_\varepsilon^{(n_\bullet)}-1$ is the number of occurrences at time $-\log(\varepsilon)$ in a point process. Because in a point process the number of occurrences by time $-\log(\varepsilon)$ is greater than or equal to m if and only if the m-th occurrence occurs before of at time $-\log(\varepsilon)$, then

$$\mathbb{P}(M_\varepsilon^{(n_\bullet)}(-\log(\varepsilon))=m)=\mathbb{P}(M_\varepsilon^{(n_\bullet)}(-\log(\varepsilon))<m+1)-\mathbb{P}(M_\varepsilon^{(n_\bullet)}(-\log(\varepsilon))<m).$$

Then, the probability distribution (2.2.8) follows by the fact

$$\mathbb{P}(M_\varepsilon^{(n_\bullet)}(-\log(\varepsilon))<m)=1-\mathbb{P}(\theta^{*m}<\log(\varepsilon))$$

where $\theta^{*m}=\sum_{1\le k\le m}-\log(1-\theta_k)$ with θ_k is a r.v. distributed according to a Beta distribution function with parameter (n_i,a) for $i=1,\dots,m$. □

By construction, the distribution of the r.v. $M_\varepsilon^{(n_\bullet)}$ generalized the Poisson distribution with parameter $-a\log(\varepsilon)$ obtained in Lemma 3 in Muliere and Tardella [140] which can be recovered setting $n_\bullet = 1_\bullet$ in (2.2.8). We now compare the random truncation of the Sethuraman series represetaion (2.1.3) with the random truncation of the generalized Sethuraman series representation (2.2.1). Since we know the distribution of the of $M_\varepsilon^{(n_\bullet)}$, a comparison based on computer simulation it is possible and requires to implement in a source code the formula for (2.2.8). In the next propositions we provide an analytic approach to compare the random truncation of the Sethuraman series represetaion (2.1.3) with the random truncation of the generalized Sethuraman series representation (2.2.1). In particular, we show that the best choice of $n_\bullet$ consists in $n_\bullet = 1_\bullet$, i.e. the best approximation of the Dirichlet process is obtained using the random truncation of the generalized Sethuraman series representation (2.2.1).

Theorem 2.2.1. *For any $\varepsilon \in (0,1]$, let $M_\varepsilon^{(n_\bullet)}$ be a r.v. on $\mathbb{N}$ distributed according to (2.2.8) and let M_ε be a r.v. on $\mathbb{N}$ distributed according to a Poisson distribution function with parameter $-a\log(\varepsilon)$. Then*

$$\sum_{i\geq 1}\left(\sum_{j=1}^{i} n_i\right)\mathbb{P}(M_\varepsilon^{(n_\bullet)} = i) > \sum_{i\geq 1}\left(\sum_{j=1}^{i} n_i\right)\mathbb{P}(M_\varepsilon = i). \tag{2.2.9}$$

Proof. First of all we can write the left side of (2.2.9) in terms of $\mathbb{P}(\tilde{\theta}^{*i} \leq -\log(\varepsilon))$ for $i \geq 1$, where $\tilde{\theta}^{*i} := -\sum_{1\leq j\leq i}\log(1-\tilde{\theta}_j)$ where $\tilde{\theta}_1,\dots,\tilde{\theta}_i$ are i r.v.s distributed according to a Beta distribution function with parameter (n_j, a) for $j = 1,\dots,i$. In particular, we have

$$\begin{aligned}\sum_{i\geq 1}\left(\sum_{j=1}^{i} n_i\right)\mathbb{P}(M_\varepsilon^{(n_\bullet)} = i) &= \sum_{i\geq 1} n_{i+1}\left(1 - \sum_{j=1}^{i}\mathbb{P}(M_\varepsilon^{(n_\bullet)} = i)\right)\\ &= n_1 + \sum_{i\geq 1} n_{i+1}\mathbb{P}(\tilde{\theta}^{*i} \leq -\log(\varepsilon)).\end{aligned}$$

In the same fashion we can write the right side of (2.2.9) in terms of $\mathbb{P}(\theta^{*i} \leq -\log(\varepsilon))$ for $i \geq 1$, where $\theta^{*i} := -\sum_{1\leq j\leq i}\log(1-\theta_j)$ where $\theta_1,\dots,\theta_i$ are i r.v.s distributed according to a Beta distribution function with parameter $(1, a)$ for $j = 1,\dots,i$.

$$\begin{aligned}\sum_{i\geq 1}\left(\sum_{j=1}^{i} n_i\right)\mathbb{P}(M_\varepsilon = i) &= \sum_{i\geq 1} n_{i+1}\left(1 - \sum_{j=1}^{i}\mathbb{P}(M_\varepsilon = i)\right)\\ &= n_1 + \sum_{i\geq 1} n_{i+1}\mathbb{P}(\theta^{*i} \leq -\log(\varepsilon)).\end{aligned}$$

We know that if $n_1 = n_2 = \ldots = n_{M_\varepsilon^{(n_\bullet)}}$, then the right side of (2.2.9) coincides with $1-\log(\varepsilon)$. Therefore, we need to prove the that

$$n_1 + \sum_{i\geq 1} n_{i+1}\mathbb{P}(\tilde{\theta}^{*i} \leq -\log(\varepsilon)) > 1 - \log(\varepsilon).$$

Without lost of generality, for any fixed $j \in \mathbb{N}$ we can prove the above condition considering that at least one n_i increase. Therefore, we can see that we can concentrate in the index $i^* := \inf\{i \in \mathbb{N} : n_i > 1\}$, i.e. in the index i^* such that $n_i > 1$ for $i = i^*$ and $n_i \in \mathbb{N}$ for $i > i^*$and $n_i = 1$ for $i < i^*$. Therefore for any $\varepsilon \in (0,1]$ and for a fixed $i^* \in \mathbb{N}$ we need to prove that for $i^* = 1$

$$\sum_{i\geq 1} n_{i+1}\mathbb{P}(\tilde{\theta}^{*i} \leq -\log(\varepsilon)) > 1 - log(\varepsilon) - n_1$$

and, for $i^* > 1$

$$\sum_{i\geq i^*} n_{i+1}\mathbb{P}(\tilde{\theta}^{*i} \leq -\log(\varepsilon)) > -\log(\varepsilon) - n_{i^*}\mathbb{P}(\theta^{*(i^*-1)} \leq -\log(\varepsilon)).$$

If we consider the two extreme points $\varepsilon = 1$ and the limit $\varepsilon \to 0^+$, then for any i^* the above conditions are verified. This is sufficient to assert that for $i^* = 1$ and for any $\varepsilon \in (0,1]$

$$\sum_{i\geq 1} n_{i+1}\mathbb{P}(\tilde{\theta}^{*i} \leq -\log(\varepsilon)) > 1 - log(\varepsilon) - n_1$$

and for $i^* > 1$ and for any $\varepsilon \in (0,1]$

$$\sum_{i\geq i^*} n_{i+1}\mathbb{P}(\tilde{\theta}^{*i} \leq -\log(\varepsilon)) > -\log(\varepsilon) - n_{i^*}\mathbb{P}(\theta^{*(i^*-1)} \leq -\log(\varepsilon)).$$

□

Theorem 2.2.2. *For any $\varepsilon \in (0,1]$, let $M_\varepsilon^{(n_\bullet)}$ be a r.v. on $\mathbb{N}$ distributed according to* (2.2.8) *and let M_ε be a r.v. on $\mathbb{N}$ distributed according to a Poisson distribution function with parameter $-a\log(\varepsilon)$. Then*

$$\sum_{i\geq 1}\left(\sum_{j=1}^{i} n_i\right)\mathbb{P}(M_\varepsilon^{(n_\bullet)} = i) > \sum_{i\geq 1} i\mathbb{P}(M_\varepsilon = i)$$

for some $n_j \to +\infty$ for $j = 1,\ldots,i$.

Proof. To prove this proposition we use a contraddiction that is, we prove that if $n_{j^*} \to +\infty$ for some $j = 1,\ldots,i$ then negative logic of

$$\sum_{i\geq 1}\left(\sum_{j=1}^{i} n_i\right)\mathbb{P}(M_\varepsilon^{(n_\bullet)} = i) > \sum_{i\geq 1} i\mathbb{P}(M_\varepsilon = i)$$

is false. Using the same idea of Theorem 2.2.1 we focus on the case $i^* = 1$. For $i^* > 1$ follows by the same arguments. The negative logic of the last equation is that for $n_1 \to +\infty$ there exists an $\varepsilon \in (0,1]$ such that $\forall \varepsilon \in (0,1]$

$$\sum_{i\geq 1} \mathbb{P}(\tilde{\theta}^{\star i} \leq -\log(\varepsilon)) \leq 1 - \log(\varepsilon) - n_1$$

where again $\theta^{*i} := -\sum_{1\leq j\leq i} \log(1-\theta_j)$ where $\theta_1, \dots, \theta_i$ are i r.v.s distributed according to a Beta distribution function with parameter $(1, a)$ for $j = 1, \dots, i$. The last equation implies that if $n_1 \to +\infty$ there exists an $\varepsilon \in (0,1]$ such that $\forall \varepsilon \in (0,1]$

$$\mathbb{P}(\tilde{\theta}_1 \leq -\log(\varepsilon)) = (1-\varepsilon)^{n_1} \leq 1 - \log(\varepsilon) - n_1$$

and it is false. □

2.3 The almost sure stopping procedure

In this section, we still use the generalized series representation of the Dirichlet process (2.1.1) and we consider its almost sure truncation in order to define a new r.p.m. which approximate, under an appropriate metric on $(\mathcal{P}_{\mathbb{X}}, \mathcal{W})$, the Dirichlet process. This new r.p.m. generalizes the M-Dirichlet process proposed by proposed by Ishwaran and Zarepour [94] (see also Ishwaran and Zarepour [93]); in particular, a comparison between the new almost sure approximation of the Dirichlet process and the M-Dirichlet process is given.

Definition 2.3.1. *Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure and let $n_\bullet$ be a fixed integer-valued sequence. For any $M \in \mathbb{N}$, a generalized M-Dirichlet processes with parameter $(\alpha, M, n_\bullet)$ is the r.p.m. $P_M^{(n_\bullet)} : \Omega \to \mathcal{P}_{\mathbb{X}}$ such that for any $B \in \mathscr{X}$*

$$\omega \mapsto \sum_{i=1}^{M} p_i(\omega) \sum_{j=1}^{n_i} q_{i,j}(\omega) \delta_{Y_{i,j}(\omega)}(B) \tag{2.3.1}$$

where we necessarily set $p_M := 1 - \sum_{1\leq i\leq M-1} p_i$ to ensure that $P_M^{(n_\bullet)}$ is a well defined r.p.m.

As a consequence of the generalized series representation of the Dirichlet process, the definition of M-Dirichlet process can be recoverd from Definiton 2.3.1 setting $n_\bullet = 1_\bullet$ in (2.3.1), i.e. the M-Dirichlet process with parameter (α, M) corresponds to the generalized M-Dirichlet process with parameter $(\alpha, M, 1_\bullet)$.

In the same way as in Section 2.2, we now provide a comparison between the generalized M-Dirichlet process and the M-Dirichlet process. In other word, we provide a comparison between the r.p.m. $P_M^{(n_\bullet)}$ defined by (2.3.1) and the r.p.m. P_M defined by (2.1.4). Following the approach proposed in [90] we consider the following proposition.

Proposition 2.3.1. *For any $M \in \mathbb{N}$ let $p_1, \ldots, p_M$ be the random weights of the generalized M-Dirichlet process. Then, for any $k_0 \in \mathbb{N}$*

$$\mathbb{E}\left[\sum_{i\geq M} n_i p_i\right]^{k_0} = \prod_{i=1}^{M-1} \frac{(a)_{k_0\uparrow 1}}{(n_i+a)_{k_0\uparrow 1}} \prod_{i\geq 0} \sum_{k_{i+1}}^{k_i} \binom{k_i}{k_{i+1}} \frac{n_{M+i}^{k_i-k_{i+1}}(a)_{k_{i+1}\uparrow 1}(n_{M+i})_{(k_i-k_{i+1})\uparrow 1}}{(n_{M+i}+a)_{k_i\uparrow 1}} \tag{2.3.2}$$

and

$$\mathbb{E}\left[\sum_{i\geq M} n_i^{k_0} p_i^{k_0}\right] = \prod_{i=1}^{M-1} \frac{(a)_{k_0\uparrow 1}}{(n_i+a)_{k_0\uparrow 1}} \sum_{i\geq M} \frac{n_i^{k_0}(n_i)_{k_0\uparrow 1}((a)_{k_0\uparrow 1})^{i-M}}{\prod_{j=M}^{i}(n_j-a)_{k_0\uparrow 1}}. \tag{2.3.3}$$

Proof. If we prove (2.3.2) then the proof of (2.3.3) follows similar arguments. In particular, for any $k_0 \in \mathbb{N}$ we need to compute

$$\mathbb{E}\left[\sum_{i\geq M} n_i p_i\right]^{k_0} = \prod_{i=1}^{M-1} \frac{(a)_{k_0\uparrow 1}}{(n_i+a)_{k_0\uparrow 1}} \mathbb{E}\left[n_M\theta_M + n_{M+1}\theta_{M+1}(1-\theta_M)\cdots\right]^{k_0}$$

where, to compute the expected value $\mathbb{E}\left[n_M\theta_M + n_{M+1}\theta_{M+1}(1-\theta_M)\cdots\right]^{k_0}$ we can use the Binomial theorem. Then, we define $\mathbb{E}_M[\tilde{\theta}]^{k_0} := E\left[n_M\theta_M + n_{M+1}\theta_{M+1}(1-\theta_M)\cdots\right]^{k_0}$ and we obtain

$$\begin{aligned}
\mathbb{E}_M[\tilde{\theta}]^{k_0} &= \sum_{k_1=0}^{k_0} \binom{k_0}{k_1} \frac{n_M^{k_0-k_1}(a)_{k_1\uparrow 1}(n_M)_{(k_0-k_1)\uparrow 1}}{(n_M+a)_{k_0\uparrow 1}} \mathbb{E}_{M+1}[\tilde{\theta}]^{k_1} \\
&= \sum_{k_1=0}^{k_0} \binom{k_0}{k_1} \frac{n_M^{k_0-k_1}(a)_{k_1\uparrow 1}(n_M)_{(k_0-k_1)\uparrow 1}}{(n_M+a)_{k_0\uparrow 1}} \\
&\qquad \times \sum_{k_2=0}^{k_1} \binom{k_1}{k_2} \frac{n_{M+1}^{k_1-k_2}(a)_{k_2\uparrow 1}(n_{M+1})_{(k_1-k_2)\uparrow 1}}{(n_{M+1}+a)_{k_1\uparrow 1}} \mathbb{E}_{M+2}[\tilde{\theta}]^{k_2} \\
&= \prod_{i\geq 0} \sum_{k_{i+1}}^{k_i} \binom{k_i}{k_{i+1}} \frac{n_{M+i}^{k_i-k_{i+1}}(a)_{k_{i+1}\uparrow 1}(n_{M+i})_{k_i-k_{i+1}\uparrow 1}}{(n_{M+i}+a)_{k_i\uparrow 1}}.
\end{aligned}$$

Then, equation (2.3.3) can be easily proved following the same argument used for equation (2.3.2). Then, we need to compute

$$\begin{aligned}
\mathbb{E}[n_M^{k_0}\theta_M^{k_0} + n_{M+1}^{k_0}\theta_{M+1}^{k_0}(1-\theta_M)^{k_0} + \cdots] &= \mathbb{E}[n_M^{k_0}\theta_M^{k_0} + (1-\theta_M)^{k_0}(n_{M+1}^{k_0}\theta_{M+1}^{k_0} + \cdots)] \\
&= \sum_{i\geq M} \frac{n_i^{k_0}(n_i)_{k_0\uparrow 1}(a)_{k_0\uparrow 1})^{i-M}}{\prod_{j=M}^{i}(n_j-a)_{k_0\uparrow 1}}
\end{aligned}$$

and the proof is completed. □

Equation (2.3.2) and (2.3.3) decrese in M. In particular, for a moderate M we can obtain an accurate approximation of the Dirichlet process. From Proposition 2.3.1 it can be easily checked that under the condition $n_\bullet = 1_\bullet$, then

$$\mathbb{E}\left[\sum_{i\geq M} p_i\right]^{k_0} = \left(\frac{a}{a+k_0}\right)^{M-1} \tag{2.3.4}$$

and

$$\mathbb{E}\left[\sum_{i\geq M} p_i^{k_0}\right] = \left(\frac{a}{a+k_0}\right)^{M-1} \frac{\Gamma(k_0)\Gamma(a)}{\Gamma(a+k_0)}. \tag{2.3.5}$$

We can compare formula (2.3.2) with (2.3.4) and formula (2.3.3) with (2.3.5) with respect to the accuracy of approximation. It is obvious from (2.3.2) and (2.3.3) that for a fixed M we can improve the accuracy of approximation increasing an n_i for $i = 1, \ldots, M$.

Nevertheless, fixing M we do not have a well-balanced comparison because for a fixed M we have to consider as benchmark the total number of r.v.'s involved in the series representation that is $2M$ for the M-Dirichlet process of Ishwaran and Zarepour [94]. Using this comparison we can easly verify that the best strategy is that proposed by Ishwaran and Zarepour [94]. In fact for any $M \in \mathbb{N}$

$$\left(\frac{a}{a+k_0}\right)^{M-1} \leq \frac{(a)_{k_0\uparrow 1}}{(M-2+a)_{k_0\uparrow 1}}\mathbb{E}[(M-1)\theta_M]^{k_0}$$

and

$$\left(\frac{a}{a+k_0}\right)^{M-1} \frac{\Gamma(k_0)\Gamma(a)}{\Gamma(a+k_0)} \leq \frac{(a)_{k_0\uparrow 1}}{(M-2+a)_{k_0\uparrow 1}}\mathbb{E}[(M-1)^{k_0}\theta_M^{k_0}]$$

under the constraint that the total number of r.v.s involved in the series represetation of the truncated r.p.m. is $2M$.

2.4 The random stopping procedure for stick-breaking random measures

The class of infinite dimensional stick-breaking random measures has been introduced by Ishwaran and James [90] as a class of all a.s. discrete r.p.m.s $Q : \Omega \to \mathcal{P}_{\mathbb{X}}$ such that for any $B \in \mathscr{X}$

$$\omega \mapsto \sum_{i\geq 1} p_i \delta_{Y_i(\omega)}(B) \tag{2.4.1}$$

where $\{p_i, i \geq 1\}$ a sequence of r.v.s chosen independent of the sequence $\{Y_i, i \geq 1\}$ and so that $0 \leq p_i \leq 1$ and $\sum_{i\geq 1} p_i = 1$. In particular, given $\alpha \in \mathcal{A}_{\mathbb{X}}$ a finite measure with total mass a, it is assumed that $\{Y_i, i \geq 1\}$ is a sequence of i.i.d. r.v.s from α_0. The method of construction for the random weights $\{p_i, i \geq 1\}$ is what sets stick-breaking random measure apart from general r.p.m. Q expressible as (2.4.1). Here, a formal definition of an infinite dimensional stick-breaking random measures.

Definition 2.4.1. *Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure and let $a_\bullet$ and $b_\bullet$ be two sequences of positive real number. An infinite dimensional stick-breaking random measure on $\mathbb{X}$ with parameter $(\alpha, a_\bullet, b_\bullet)$ is an a.s. discrete r.p.m. Q on $\mathbb{X}$ of the form (2.4.1) such that the sequence $\{p_i, i \geq 1\}$ is obtained by stick-breaking construction from a sequence $\{\theta_i, i \geq 1\}$ of independent r.v.s distributed according to a Beta distribution function with parameter (a_i, b_i) for $i \in \mathbb{N}$.*

The stick-breaking notion for constructing random weights has a very long story. For examples, see Halmos [82], Freedman [72], Fabius [48], Connor and Mosimann [21] and Kingman [109]. In particular, the definition of stick-breaking random measure gives one a unified way of connecting together a collection of seemingly unrelated measures scattered throughout the literature. These include the Dirichlet process, the two parameter Poisson-Dirichlet process introduced by Pitman and Yor [154], the Dirichlet multinomial process introduced by Muliere and Secchi [139], the m-spike models introduced by Liu [127], the finite dimensional Dirichlet process introduced by Ishwaran and Zarepour [94] (see also Ishwaran and Zarepour [93]) and the Beta two parameter process introduced by Ishwaran and Zarepour [92].

From Definition 2.4.1 it follows that an infinite dimensional stick-breaking random measure Q is only well defined if its random weights sum to one with probability one. In particular it can be easilty cheked

$$\sum_{i \geq 1} p_i = 1 \qquad \text{a.s.} \qquad \Leftrightarrow \qquad \sum_{i \geq 1} \mathbb{E}(\log(1 - \theta_i)) = -\infty$$

or, alternativley $\sum_{i \geq 1} \log(1 + a_i/b_i) = +\infty$. In particular, the Dirichlet process with parameter α can be recoverd as particular case of the infinite dimensional stick-breaking measure Q having random weights $\{p_i, i \geq 1\}$ obtained by the stick-breaking construction from a sequence $\{\theta_i, i \geq 1\}$ of i.i.d. r.v.s distruted according to a Beta distribution function with parameter $(1, a)$ where a is the total mass of the measure α. In particular, this corresponds to the Sethuraman series representation of the Dirichlet process (2.1.2).

The two parameter Poisson-Dirichlet process and the Beta two parameter process are further examples of infinite dimensional stick-breaking measures. In particular, the two parameter Poisson-Dirichlet process introduced by Pitman and Yor [154] has been the subject of a considerable amount of reseach interest. A key property of the two parameter Poisson-Dirichlet process is its characterization as an infinite dimensional stick-breaking random measure due to Pitman [149] and Pitman [151]. As shown there, the size-biased random permutation for the ranked random weights from the measure produces a sequence of random weights derived using a residual allocation scheme (see Pitman [152]), or what we are calling a stick-breaking scheme. This then identify the two parameter Poisson-Dirichlet process as an infinite dimensional stick-breaking random measure where the random weights $\{p_i, i \geq 1\}$ are obtained by the stick-breaking construction with $\{\theta_i, i \geq 1\}$ a sequence of independent r.v.s distributed according to a Beta distribution function with parameter $(1-\sigma, b+i\sigma)$ where $0 \leq \sigma \leq 1$ and $b > -\sigma$. The Dirichlet process with parameter α can be recovered from the two parameter Poisson-Dirichlet process by setting $\sigma = 0$ and $b = a$ where a is the total mass of α. Another important example is the process obtained from the two parameter Poisson-Dirichlet process by setting $b = 0$ which correspond to a measure whose random weights are based on a stable law with index σ.

As shown by Ishwaran and James [90] an interesting subclass of the class of infinite dimensional stick-breaking random measures is the class of r.p.m.s obtained by the almost sure tructation of infinite dimensional stick-breaking random measures. As we introduced in Section 2.1, the almost sure truncation was originally proposed by Ishwaran and Zarepour [94] (see also Ishwaran and Zarepour [93]) for the Sethuraman series representation of the Dirichlet process. For any $M \in \mathbb{N}$, the trunctaion is applied by discarding the $M+1, M+2, \ldots$ terms in Q and replacing p_M with $1 - p_1 - \ldots - p_{M-1}$. Notice that this also corresponds to setting $\theta_M = 1$ in the stick-breaking construction. In particular, given an infinite dimensional stick-breaking random measure Q with parameter $(\alpha, a_\bullet, b_\bullet)$, for any $M \in \mathbb{N}$ we denote by Q_M its almost truncation and we call it the M-stick-breaking random measure with parameter $(\alpha, a_\bullet, b_\bullet, M)$. The M-Dirichlet process with parameter (α, M) proposed by Ishwaran and Zarepour [94] (see also Ishwaran and Zarepour [93]) can be recoverd as an almost sure truncation of the Sethuraman series representation of the Dirichlet process with parameter α. Determining an appropriate truncation level can be based on the moments of the random weights. For instance, the following theorem can be used to determining an appropriate truncation lavel for the two parameter Poisson-Dirichlet process.

Theorem 2.4.1. *(cfr. Ishwaran and James [90]) Let $\{p_i, i \geq 1\}$ be the sequence of random weights characterizing the stick-breaking representation of the two parameter Poisson-Dirichlet process. For any $M \in \mathbb{N}$ and any $r \in \mathbb{N}$*

$$\mathbb{E}\left[\sum_{k \geq M} (p_k)^r\right] = \prod_{k=1}^{M-1} \frac{(b+k\sigma)_{r\uparrow 1}}{(\sigma + (k-1)\sigma + 1)_{r\uparrow 1}} \tag{2.4.2}$$

and

$$\mathbb{E}\left[\sum_{k \geq M} p_k^r\right] = \mathbb{E}\left[\sum_{k \geq M} (p_k)^r\right] \frac{(1-\sigma)_{(r-1)\uparrow 1}}{(b + (M-1)\sigma + 1)_{(r-1)\uparrow 1}}. \tag{2.4.3}$$

Given a measure $\alpha \in \mathcal{A}_{\mathbb{X}}$ a finite measure with total mass a, as particular case of equation (2.4.2) and equation (2.4.3) we can obtain the equation (2.3.4) and the equation (2.3.5) for the Dirichet process setting $\sigma = 0$ and $b = a$. In particular equation (2.3.4) and the equation (2.3.5) underline that a moderate M should be able to achive an accurate approximation. In general, for an infinite dimensional stick-breaking random measure Q a very large value for M may be needed for reasonable accuracy. For instance, it the particular case of the two parameter Poisson-Dirichlet process obtained by setting $b = 0$, we have

$$\mathbb{E}\left[\sum_{k \geq M} p_k\right] = \frac{\sigma^{M-1}(M-1)!}{(\sigma+1)\cdots((M-2)\sigma+1)}.$$

Note that the value of M needed to keep this value small rapidly increases as σ approaches one. Thus it may not be feasible to approximated the corresponding process over all values of σ.

Here, in order to solve the problem of the selection of an appropriate $M \in \mathbb{N}$, we propose the application of the random stopping rule to the class of the infinite dimensional stick-breaking random measures as an alternative to the almost sure truncation. In particular, we extend to the class of the infinite dimensional stick-breaking random measures the random stopping rule proposed by Muliere and Tardella [140] for the Sethuraman series representation of the Dirichlet process.

Definition 2.4.2. *Let Q be an infinite dimensional stick-breaking random measure on $\mathbb{X}$ with parameter $(\alpha, a_\bullet, b_\bullet)$. For any ε, the ε-stick-breaking random measure with parameter $(\alpha, a_\bullet, b_\bullet, \varepsilon)$ is a r.p.m. $Q_\varepsilon : \Omega \to \mathcal{P}_{\mathbb{X}}$ such that, for any $B \in \mathscr{X}$*

$$\omega \mapsto \sum_{i=1}^{M_\varepsilon(\omega)} p_i(\omega)\delta_{Y_i(\omega)}(B) + R_\varepsilon(\omega)\delta_{Y_0(\omega)}(B) \tag{2.4.4}$$

where

$$M_\varepsilon := \inf\left\{m \in \mathbb{N}\colon \sum_{i=1}^{m} p_i > 1 - \varepsilon\right\}$$

$$R_\varepsilon = 1 - \sum_{i=1}^{M_\varepsilon} p_i$$

and Y_0 is a r.v. distributed according to α_0.

As a consequence of the definition of infinite dimensional stick-breking random measure, the definition of ε-stick breaking random measure, the definition of ε-Dirichlet process with parameter (α, ε) can be recoverd from Definiton 2.4.2 when Q is the Dirichlet process with parameter α, i.e. when Q is an infinite infinite dimensional stick-breaking random measure with parameter $(\alpha, 1_\bullet, a_\bullet)$ where $a_\bullet$ is the sequence of the total mass of α.

The Definition 2.4.2 implies the closeness, in the total variation metric ρ_V on $(\mathcal{P}_{\mathbb{X}}, \mathcal{W})$, between the infinite dimensional sitck-breaking random measure Q and its corresponding ε-stick-breaking random measure Q_ε, i.e. for any $\omega \in \Omega$ $\rho_V(Q_\varepsilon(\omega, \cdot), Q(\omega, \cdot)) \leq \varepsilon$ on a set of $\mathbb{P}$-probability 1. As regard the the closeness, in the Prokhorov metric between the infinite dimensional sitck-breaking random measure Q and its corresponding ε-stick-breaking random measure Q_ε we consider the following lemma.

Lemma 2.4.1. *For any $\omega \in \Omega$ and for any $\varepsilon \in (0, 1]$*

$$\rho_P(Q_\varepsilon(\omega, \cdot), Q(\omega, \cdot)) \leq \varepsilon.$$

Proof. The proof follows by the same arguments used in Lemma 2 in Muliere and Tardella [140]. □

In particular, Lemma 2.4.1 implies that the application of the random stopping procedure to an infinite dimensional stick-breaking random measure Q can fix in advance the closeness ε with its corresponding random approximation Q_ε. Thus, the random stopping procedure for the class of infinite dimensional stick-breaking random measures provide a solution to the problem of the selection of an appropriate M in the class of the almost sure truncated stick-breaking random measures. In particular, this appealing feature can also be extended to the distributions of several fanctionals of an infinte dimensional stick-breaking random measure.

In the definition of the ε-stick-breaking random measure the role of the stopping time M_ε it to allow generating a random probability as close as one wants (in the total variation distance) to the corresponding infinite dimensional stick-breaking random measure. We write

$$R_\varepsilon = 1 - \sum_{i=1}^{M_\varepsilon} p_i = \prod_{i=1}^{M_\varepsilon} (1 - \theta_i)$$

then the r.v. M_ε can be written as

$$M_\varepsilon = \inf \left\{ m \in \mathbb{N} : \sum_{i=1}^{m} \log(1 - \theta_i) < \log(\varepsilon) \right\}.$$

As we can see, the r.v. M_ε involves the distribution of the r.v. $\sum_{1 \leq i \leq m} \log(1 - \theta_i)$ which has been computed in Lemma 2.2.2 point i). Then, the following lemma provide the distribution of M_ε.

Lemma 2.4.2. *For any $\varepsilon \in (0, 1]$, M_ε is a r.v. on $\mathbb{N}$ such that*

$$\mathbb{P}(M_\varepsilon = m) = \begin{cases} \mathbb{P}(\theta^{*m} \leq -\log(\varepsilon)) & m = 1 \\ \\ \mathbb{P}(\theta^{*m} \leq -\log(\varepsilon)) - \mathbb{P}(\theta^{*(m+1)} \leq -\log(\varepsilon) & m > 1 \end{cases} \tag{2.4.5}$$

where $\mathbb{P}(\theta^{*m} \leq z)$ *is given by* (2.2.3).

Proof. The proof follows by the same arguments used in Lemma 2.2.3. □

2.5 Discussion

In the first part of this chapter, moving from the alternative constructive definition of the Dirichlet process (2.1.1) which includes the Sethuraman constructive definition as particular case, we considered the application of two different stopping procedure. The first stopping procedure we applied to the series representation (2.1.1) was originally proposed by Muliere and Tardella [140] for the Sethuraman series representation and it corresponds to the truncation of the series at a random number of terms. The second stopping procedure we applied to the series representation (2.1.1) was originally proposed by Ishwaran and Zarepour [94] (see also Ishwaran and Zarepour [93]) for the Sethuraman series representation and it corresponds to the truncation of the series at a fixed number ot terms.

As regard the application of the random stopping procedure to the series representation (2.1.1) we defined the new random truncated r.p.m. (2.2.1), the so-called generalized ε-Dirichlet process and we provided a comparison, in terms of accuracy in the Dirichlet process approximation, with respect to the ε-Dirichlet defined by Muliere and Tardella [140]. In particular, Theorem 2.2.1 and Theorem 2.2.2 have shown that using the random stopping procedure of the series representation (2.1.1) the best approximation of the Dirichlet process corresponds with the choice of $n_{\bullet} = 1_{\bullet}$. In other words, the best approximation of the Dirichlet process is obtained using the random truncation of the Sethuraman series representation as proposed by Muliere and Tardella [140]. As a by-product of this analysis, we also obtain in Lemma 2.2.2 and in Corollary 2.2.1 some interesting results related to the convolution of distributions belonging to the class of generalized convolutions of mixtures of Exponential distributions (see Bondesson [11] and Bondesson [12]). In particular Corollary 2.2.1 provide a simple expression for the convolution of Exponential distributions whose importance in applications is known in the literature and it has been recently reviewed by Nadarajah [142].

As regard the application of the almost sure stopping procedure to the series representation (2.1.1) we defined the new random truncated r.p.m. (2.3.1), the so-called generalized M-Dirichlet process and we provided a comparison, in terms of accuracy in the Dirichlet process approximation, with respect the M-Dirichlet defined by Ishwaran and Zarepour [94] (see also Ishwaran and Zarepour [93]). In particular, Proposition 2.3.1 has shown that using the random stopping procedure of the series representation (2.1.1) the best approximation of the Dirichlet process corresponds with the choice of $n_{\bullet} = 1_{\bullet}$. In other words, the best approximation of the Dirichlet process is obtained using the random truncation of the Sethuraman series representation as proposed by Muliere and Tardella [140].

In the second part of we provided a straightforward extension of the random stopping procedure to the more general class of infinite dimensional stick-breaking random measures introduced by Ishwaran and James [90]. In particular, in Ishwaran and James [90] is considered the subclass of the almost sure trunctation of infinite dimensional stick-breaking random measures and the roblem of finding an appropriate method for selecting the truncation level M for an infinite dimensional stick-breaking random measure is underlined. In order to give a solution to this problem we considered the definition of the class of ε-stick-breaking random measures which is defined as the class of r.p.m. obtained by the random trunction of the infinite dimensional stick-breaking random measures. In particular, differently from the almost sure truncatd stick-breaking random measures, the the ε-stick-breaking random measures admit the possibility to fix in advance the closeness ε with respect to the corresponding infinite dimensional stick-breaking random measure.

3

Some developments on the Feigin-Tweedie Markov chain

In this chapter, moving from the distributional equation for the Dirichlet process introduced in Chapter 1, we define and we investigate a new measure-valued Markov chain having as unique invariant measure the law of a Dirichlet process. This Markov chain generalizes the well known Feigin-Tweedie Markov chain which has been widely used to provide properties of linear functionals of the Dirichlet process and approximation procedures for estimating the law of the mean of the Dirichlet process. Our main aim in this chapter is to show that the Feigin-Tweedie chain sits in a large class of chains indexed by an integer $n \in \mathbb{N}$ and they worked solely on the case $n = 1$, where n can be viewed as a sample size. We provide properties of this new class of Markov chain.

3.1 Introduction

In this chapter we focus on an interesting result obtained by Feigin and Tweedie [58] which characterizes the law of the Dirichlet process as the unique invariant measure of a certain measure-valued Markov chain. Let $(\mathbb{X}, \mathcal{T})$ be the usual Polish space endowed with the Borel σ-field $\mathscr{X}$ and consider the following associated spaces of measures $\mathcal{A}_{\mathbb{X}}$ and $\mathcal{P}_{\mathbb{X}}$. In particular, $\mathcal{A}_{\mathbb{X}}$ is the space of locally finite non-negative measures on $(\mathbb{X}, \mathscr{X})$ endowed with the σ-field $\mathscr{A}_{\mathbb{X}}$ generated by the vague topology $\mathcal{V}$ which makes $(\mathcal{A}_{\mathbb{X}}, \mathcal{V})$ a Polish space, and $\mathcal{P}_{\mathbb{X}}$ is the space of probability measures on $(\mathbb{X}, \mathscr{X})$ endowed with its σ-field $\mathscr{P}_{\mathbb{X}}$ generated by the weak convergence topology $\mathcal{W}$ which makes $(\mathcal{P}_{\mathbb{X}}, \mathcal{W})$ a Polish space. Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure with total mass a and let P be a Dirichlet process on $\mathbb{X}$ with parameter α. Furthermore, let Y be a random variable (r.v.) distributed according to $\alpha_0 := \alpha/a$ and let θ be a r.v. distributed according to a Beta distribution function with parameter $(1, a)$ and independent of Y. Then, according to Theorem 3.4 in Sethuraman [174], the Dirichlet process on $\mathbb{X}$ with parameter α is the unique solution of the distributional equation

$$P \overset{d}{=} \theta\delta_Y + (1-\theta)P, \tag{3.1.1}$$

where all the r.v.s on the right hand-side are independent. Equation (3.1.1) has been widely used in the literature in order to provide characterizations both of the Dirichlet process and of the mean of the Dirichlet process. See Hjort and Ongaro [85] and references therein. In particular, in Feigin and Tweedie [58], equation (3.1.1) is recognised as the distributional equation for the unique invariant measure of a Markov chain $\{P_m, m \geq 0\}$ on $\mathcal{P}_{\mathbb{X}}$ defined via the recursion

$$P_m = \theta_m\delta_{Y_m} + (1-\theta_m)P_{m-1} \qquad m \geq 1 \tag{3.1.2}$$

where $P_0 \in \mathcal{P}_{\mathbb{X}}$ is arbitrary and $\{Y_m, m \geq 1\}$ and $\{\theta_m, m \geq 1\}$ are two independent sequences of independent and identically distributed (i.i.d.) r.v.s distributed as Y and θ, respectively. In Feigin and Tweedie [58] it is shown that the Markov chain $\{P_m, m \geq 0\}$ has as unique invariant measure the law of a Dirichlet process P on $\mathbb{X}$ with parameter α. This approach has been particularly convenient for analyzing the existence and properties of linear functional of the Dirichlet process because they also turn out to be derivable as strong limits of Markov chain on $\mathbb{R}$. In particular, in Feigin and Tweedie [58], by investigating the linear functional $\{G_m, m \geq 0\}$, with $G_m := \int_{\mathbb{X}} g(x)P_m(\cdot, dx)$ for $m \geq 0$ and for any real-valued measurable function $g : \mathbb{X} \to \mathbb{R}$, properties of the corresponding linear functional of a Dirichlet process are given. The existence of the linear functional $G := \int_{\mathbb{X}} g(x)P(\cdot, dx)$ of the Dirichlet process P is characterized according to the condition $\int_{\mathbb{X}} \log(1 + |g(x)|)\alpha(dx) < +\infty$. These functionals were considered by Hannum et al. [83] and their existence was also investigated by Doss and Sellke [33] who referred to them as moments.

Further developments of the linear functional Markov chain $\{G_m, m \geq 0\}$ are provided by Guglielmi and Tweedie [81] and by Jarner and Tweedie [103]. In Guglielmi and Tweedie [81], the distribution $\mathscr{M}$ of the mean functional $M := \int_{\mathbb{X}} xP(\omega, dx)$ of a Dirichlet process on $\mathbb{R}$ with parameter α is characterized as the unique limiting distribution of the real-valued Markov chain $\{M_m, m \geq 0\}$ with $M_m := \int_{\mathbb{X}} xP_m(\omega, dx)$ for $m \geq 0$. Guglielmi and Tweedie [81] proved that the rate of convergence in total variation of the Markov chain $\{M_m, m \geq 0\}$ to M is geometric if α_0 admits finite expectation and they found bounds on the rate of convergence. These results are then used to simulate effectively from the distribution $\mathscr{M}$ by the empirical distribution of a sample from the m-th step distribution of the approximating Markov chain $\{M_m, m \geq 0\}$ starting from any point $x \in \mathbb{R}$. In particular, the characterization of the Dirichlet process given in Feigin and Tweedie [58] is used to develop a Markov chain Monte Carlo (MCMC) algorithm which provides a useful tool to evaluate the approximation error in simulating $\mathscr{M}$. In general, all the results obtained by Guglielmi and Tweedie [81] can be applied to any functional G of a Dirichlet process P on $\mathbb{X}$ since the distribution of G corresponds to the distribution of $G \circ \alpha_g$ where $\alpha_g(B) := \alpha(g^{-1}(B))$ for any $B \in \mathscr{R}$.

From the distributional equation (3.1.1), a constructive definition of the Dirichlet process has been proposed by Sethuraman [174]. In particular, if $\alpha \in \mathcal{A}_{\mathbb{X}}$ is a finite measure with total mass a and P is a Dirichlet process with parameter α, then

$$P = \sum_{i \geq 1} p_i \delta_{Y_i} \tag{3.1.3}$$

where $\{p_i, i \geq 1\}$ is a sequence of r.v.s obtained by the stick-breaking construction $p_1 = w_1$ and $p_i = w_i \prod_{1 \leq j \leq i-1}(1 - w_j)$ for $i > 1$ with $\{w_i, i \geq 1\}$ a sequence of i.i.d. r.v.s distributed according to a Beta distribution function with parameter $(1, a)$, and $\{Y_i, i \geq 1\}$ is a sequence of i.i.d. r.v.s distributed according to α_0. Then, equation (3.1.1) arises by considering

$$P = p_1 \delta_{Y_1} + (1 - w_1) \sum_{i \geq 2} \tilde{p}_j \delta_{Y_i}$$

where now $\tilde{p}_2 = w_2$ and $\tilde{p}_i = w_i \prod_{2 \leq j \leq i-1}(1 - w_j)$ for $i > 2$. Thus it is easy to see that

$$\tilde{P} := \sum_{i \geq 2} \tilde{p}_i \delta_{Y_j} \tag{3.1.4}$$

is also a Dirichlet process with parameter α and it is independent of (p_1, Y_1). If we want to extend this idea to n initial samples, we would consider writing

$$P = \sum_{i=1}^{n} p_i \delta_{Y_i} + \sum_{i \geq n+1} p_i \delta_{Y_j}$$

which can be written as

$$P = \theta \sum_{i=1}^{n} \left(\frac{p_i}{\theta} \right) \delta_{Y_i} + (1 - \theta) \tilde{P}$$

where $\theta = \sum_{1 \leq i \leq n} p_i = 1 - \prod_{1 \leq i \leq n}(1 - w_i)$ and $\tilde{P}$ is a Dirichlet process with parameter α independent of the r.v. $((p_1, Y_1), \ldots, (p_n, Y_n))$. However this is not an easy extension since the distribution of θ is unclear and moreover θ and $\sum_{1 \leq i \leq n} (p_i/\theta) \, \delta_{Y_i}$ are not independent. For this reason we consider an alternative distributional equation for the n which has been introduced in Section 1.3 of Chapter 1 and here recalled.

In particular, let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure with total mass a and let P be a Dirichlet process on $\mathbb{X}$ with parameter α. Furthermore, let $\{Y_j, j \geq 1\}$ be a Blackwell-MacQueen Pólya sequence with parameter α, i.e. if P is a Dirichlet processes with parameter α, then for any $n \in \mathbb{N}$, $Y_1, \ldots, Y_n | P$ are i.i.d. from P. For any $n \in \mathbb{N}$ let $(q_1, \ldots, q_n)$ be a r.v.

distributed according to Dirichlet distribution function with parameter $(1, \ldots, 1)$ and let θ be a r.v. distributed according to a Beta distribution function with parameter (n, a) such that P, $\{Y_j, j \geq 1\}$, $(q_1, \ldots, q_n)$ and θ are mutually independent. Moving from such collection of random elements and assuming independence between them, in Section 1.3. of Chapter 1 it is shown that for any $n \in \mathbb{N}$, the distributional equation

$$P \stackrel{d}{=} \theta \sum_{j=1}^{n} q_j \delta_{Y_j} + (1-\theta) P \tag{3.1.5}$$

has as unique solution the Dirichlet process on $\mathbb{X}$ with parameter α. It can be easily checked that the distributional equation (3.1.5) generalizes the distributional equation (3.1.2) which can be recovered setting $n = 1$. So where precisely does this distributional equation come from with solution the Dirichlet process with parameter α? Since it is clear it does not come from the Sethuraman series representation. In fact it comes from a posterior representation of the Dirichlet process. In particular, it is well known that if P is a Dirichlet process with parameter α then $P|Y_1, \ldots, Y_n$ is a Dirichlet process with parameter $\alpha + \sum_{1 \leq i \leq n} \delta_{Y_i}$ and the law of $\int [P|Y_1, \ldots, Y_n] Q(dY_1, \ldots, dY_n)$ is the law of a Dirichlet process with parameter α when $Y_1, \ldots, Y_n$ has the distribution of the first n samples of a Blackwell-MacQueen Pólya sequence with parameter α (see Antoniak [2]). Thus, the key now is to write

$$P|Y_1, \ldots, Y_n \stackrel{d}{=} \theta \sum_{j=1}^{n} q_j \delta_{Y_j} |Y_1, \ldots, Y_n + (1-\theta) P$$

where θ is distributed according to a Beta distribution function with parameter (n, a). Hence, the distributional equation (3.1.5) follows.

In this chapter, following the original idea of Feigin and Tweedie [58], our aim is to use the more general distributional equation (3.1.5) in order to define a Markov chain $\{P_m^{(n)}, m \geq 0\}$ on $\mathcal{P}_{\mathbb{X}}$ which generalizes the Feigin-Tweedie Markov chain and which still has as unique invariant measure the law of a Dirichlet process with parameter α. In particular, we are interested in providing a detailed analysis of the Markov chain $\{P_m^{(n)}, m \geq 0\}$ in order to verify if it preserves all the properties of the original Markov chain $\{P_m, m \geq 0\}$ proposed by Feigin and Tweedie [58]. Furthermore, a special case of the Markov chain $\{P_m^{(n)}, m \geq 0\}$ is considered by assuming as state space the subset of $\mathcal{P}_{\mathbb{X}}$ of all probability measures with a finite number $k \in \mathbb{N}$ of massees. Under this assumption on the state space, the Markov chain $\{P_m^{(n)}, m \geq 0\}$ is interpreted as a discrete time stochastic model describing the evolution of the proportions of k distinct types in a population of a certain constant size. Comparisons, in terms of the transion probabilities, are given with respect to similar models known in the literature (see Ewens [45]).

Moving from the Markov chain $\{P_m^{(n)}, m \geq 0\}$, we are interested in providing properties of the associated linear functional Markov chain $\{G_m^{(n)}, m \geq 0\}$ with $G_m^{(n)} := \int_{\mathbb{X}} g(x) P_m^{(n)}(\omega, dx)$ for $m \geq 0$ and for any real-valued measurable function $g : \mathbb{X} \to \mathbb{R}$ such that $\int_{\mathbb{R}} \log(1 + |g(x)|)\, \alpha(dx) < +\infty$. These properties represent the starting point to study the rates of covergence in total variation of the new mean functional Markov chain $\{M_m^{(n)}, m \geq 0\}$ with $M_m^{(n)} := \int_{\mathbb{X}} x P_m^{(n)}(\omega, dx)$ for $m \geq 0$ to the mean functional M of a Dirichlet process on $(\mathbb{R}, \mathscr{R})$ with parameter α. In particular, following the same idea proposed by Guglielmi and Tweedie [81] we provide an approximation of the distribution $\mathscr{M}$ using the empirical distribution of a sample from the m-th step distribution of the approximating Markov chain $\{\tilde{M}_m, m \geq 0\}$ starting from any point $x \in \mathbb{R}$.

Before proceeding we make a remark about an interesting connection with the Markov chain $\{P_m, m \geq 0\}$ defined via recursion (3.1.5). The sample $Y_1, \ldots, Y_n$ can be obtained by considering a Dirichlet process P with parameter α and then taking $Y_1, \ldots, Y_n | \tilde{P}$ to be i.i.d. $\tilde{P}$ defined in (3.1.4). Hence we could consider the distributional equation

$$P \stackrel{d}{=} \theta \sum_{j=1}^{n} q_j \delta_{Y_j} + (1 - \theta) \tilde{P} \tag{3.1.6}$$

where the $Y_1, \ldots, Y_n$ are i.i.d. from P and $\tilde{P}$ is a Dirichlet process with parameter α with P, θ and $\tilde{P}$ being mutually independent. We have essentially switched $\tilde{P}$ and P around since they are independent and both Dirichlet processes with parameter α. Then the solution of the of distributional equation (3.1.6) would again be a Dirichlet process with parameter α and the Markov chain which corresponds to (3.1.2) would turn out to be a Wright-Fisher model (see Ewens [45] and references therein).

The chapter is structured as follows. In Section 3.2, moving from the distributional equation (3.1.5), we define the Markov chain $\{P_m^{(n)}, m \geq 0\}$ on $\mathcal{P}_{\mathbb{X}}$ and we prove that it has as unique invariant measure the law of a Dirichlet process with parameter α. In Section 3.3 we consider the transition densities of the Markov chain $\{P_m^{(n)}, m \geq 0\}$ on the subset of $\mathcal{P}_{\mathbb{X}}$ of all probability measures with a finite number $k \in \mathbb{N}$ of massees. In Section 3.4 a detailed analysis of the mean functional Markov chain $\{M_m^{(n)}, m \geq 0\}$ and we discuss some illustrative examples. In Section 3.5 we consider a different application of the distributional equation (3.1.5) related to the mean functional of a Dirichlet process with parameter α. Section 3.6 is devoted to a discussion of the results.

3.2 A generalized Feigin-Tweedie Markov chain

In this section our aim is to use the distributional equation (3.1.5) in order to define a Markov chain on the space $\mathcal{P}_{\mathbb{X}}$ that generalizes the Markov chain proposed by Feigin and

Tweedie [58] and which still has as unique invariant measure the law of a Dirichlet process with parameter α.

Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure with total mass a and let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space supporting a sequence $\theta := \{\theta_m, m \geq 1\}$ and for any $n \in \mathbb{N}$ the sequences $q := \{(q_{m,1}, \ldots, q_{m,n}), m \geq 1\}$ and $Y := \{(Y_{m,1}, \ldots, Y_{m,n}), m \geq 1\}$. The sequence θ is a sequence of independent r.v.s distributed according to a Beta distribution function with parameter (n, a) while q is a sequence of independent r.v.s identically distributed according to a Dirichlet distribution function with parameter $(1, \ldots, 1)$ and Y is a sequence of independent r.v.s (samples of size n) from a Blackwell-MacQueen Pólya sequence with parameter α, i.e. if P_m, for any $m \in \mathbb{N}$ are independent Dirichlet processes with parameter α, then for any $m \in \mathbb{N}$, $Y_{m,1}, \ldots, Y_{m,n} | P_m$ are i.i.d. from P_m. The condition of independence between the sequence of r.v.s θ, q and Y and the usual construction of a product measure implies the existence of the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ supporting the r.v. (θ, q, Y) and does not require any restrictions on $\mathbb{X}$, such as it being a Polish space. Moving from such collection of random elements, for any $n \in \mathbb{N}$ we define the Markov chain $\{P_m^{(n)}, m \geq 0\}$ on $\mathcal{P}_{\mathbb{X}}$ via the recursion

$$P_m^{(n)} = \theta_m \sum_{j=1}^{n} q_{m,j} \delta_{Y_{m,j}} + (1 - \theta_m) P_{m-1}^{(n)} \qquad m \geq 1 \tag{3.2.1}$$

where $P_0^{(n)} \in \mathcal{P}_{\mathbb{X}}$ is arbitrary.

By construction, the Markov chain $\{P_m, m \geq 0\}$ proposed by Feigin and Tweedie [58] and defined via the recursion (3.1.2) can be recoverd from the Markov chain $\{P_m^{(n)}, m \geq 0\}$ by setting $n = 1$. Following the original idea of Feigin and Tweedie [58] by equation (3.2.1) we define the Markov chain $\{P_m^{(n)}, m \geq 0\}$ from a distributional equation having as unique solution the Dirichlet process. In particular, the Markov chain $\{P_m^{(n)}, m \geq 0\}$ is defined from the distributional equation (3.1.5) which generalizes the distributional equation (3.1.1) by the substitution of the random probability measure (r.p.m.) δ_Y with the random convex linear combination $\sum_{1 \leq j \leq n} q_j \delta_{Y_j}$. In the next theorem we provide an alternative proof for the solution of the distributional equation (3.1.5). The proof easily follows from three lemmas here: the first one recalls a property of the Dirichlet distribution function and its proof can be found in Wilks [196], Section 7; the second one can be derived by a simple transformation of r.v.s; the third one is related to the uniqueness of the solution of the distributional equation (3.1.5) and its proof can be found in Sethuraman [174] (see also Vervaat [185], Section 1).

Lemma 3.2.1. *For any $k \in \mathbb{N}$, let $\beta = (\beta_1, \ldots, \beta_k)$ and $\gamma = (\gamma_1, \ldots, \gamma_k) \in \mathbb{R}^k$. Let U, V be independent k-dimensional r.v.s distributed according to a Dirichlet distribution function with parameter β and γ, respectively. Let W be a r.v. independent of (U, V) and* *distributed according to a Beta distribution function with parameter* $\sum_{1\le i\le k}\beta_i, \sum_{1\le i\le k}\gamma_i$. *Then,* $WU+(1-W)V$ *is distributed according to a Dirichlet distribution with parameter* $\beta+\gamma$.

Lemma 3.2.2. *For any* $n\in\mathbb{N}$, *let* $\xi_1,\dots,\xi_n$ *be* n *independent r.v.s such that* ξ_i *is distributed according to a Beta distribution function with parameter* $(1,n-i)$ *for* $i=1,\dots,n$ $(\xi_n=1$ *a.s.). If* $q_1=\xi_1$ *and* $q_j=\xi_j\prod_{1\le i\le j-1}(1-\xi_i)$ *for* $j=2,\dots,n$, *then* $(q_1,\dots,q_n)$ *is a r.v. with* $\sum_{1\le i\le n}q_i=1$ *distributed according to a Dirichlet distribution function with parameter* $(1,\dots,1)$.

Lemma 3.2.3. *(cfr. Sethuraman [174]) Let* W, U *be a pair of r.v.s where* W *takes values in* $[-1,1]$ *and* U *takes values in a linear space. Suppose that* V *is a r.v. taking values in the same linear space as* U *and which is independent of* (W,U) *and satisfies the distributional equation*

$$V\stackrel{d}{=}U+WV. \tag{3.2.2}$$

Suppose that $\mathbb{P}(|W|=1)\neq 1$. *Then there is only one distribution for* V *that satisfies* (3.2.2).

Theorem 3.2.1. *The distributional equation* (3.1.5) *has the Dirichlet process with parameter* α *as its unique solution.*

Proof. From Skorohod's theorem and Lemma 3.2.2 it follows that there exist n independent r.v.s $\xi_1,\dots,\xi_n$ such that ξ_i is distributed according to a Beta distribution function with parameter $(1,n-i)$ for $i=1,\dots,n$ and $q_1=\xi_1$ and $q_j=\xi_j\prod_{1\le i\le j-1}(1-\xi_i)$ for $j=2,\dots,n$. Further, since $\xi_n=1$ a.s., then $\sum_{1\le j\le n}q_j=1$ a.s. and it can be verified by induction that

$$1-\sum_{j=1}^{i}q_j=\prod_{j=1}^{i}(1-\xi_j)\qquad i=1,\dots,n-1.$$

Let $B_1,\dots,B_k$ be a finite measurable partition of $\mathbb{X}$. We first prove that conditionally on $Y_1,\dots,Y_n$, the finite dimensional distribution of the r.p.m. $\sum_{1\le j\le n}q_j\delta_{Y_j}$ is the Dirichlet distribution with parameter given by the following $(\sum_{1\le j\le n}\delta_{Y_j}(B_1),\dots,\sum_{1\le j\le n}\delta_{Y_j}(B_k))$. Actually, since

$$\begin{aligned}\left(\left(\sum_{j=1}^{n}q_j\delta_{Y_j}\right)(\cdot,B_1),\dots,\left(\sum_{j=1}^{n}q_j\delta_{Y_j}\right)(\cdot,B_k)\right)&=\left(\sum_{j=1}^{n}q_j\delta_{Y_j}(B_1),\dots,\sum_{j=1}^{n}q_j\delta_{Y_j}(B_k)\right)\\&=\left(\sum_{j:Y_j\in B_1}q_j,\dots,\sum_{j:Y_j\in B_k}q_j\right),\end{aligned}$$

conditionally on $Y_1, \ldots, Y_n$, the r.v. $(\sum_{j:Y_j \in B_1} q_j, \ldots, \sum_{j:Y_j \in B_k} q_j)$ is distributed according to a Dirichlet distribution function with parameter given by $(n_1, \ldots, n_k)$, where $n_j = \sum_{1\leq i\leq n} \delta_{Y_i}(B_j)$ for $j = 1, \ldots, n$. Then, using Lemma 3.2.1, it follows that, conditionally on $Y_1, \ldots, Y_n$, the finite dimensional distributions of the right hand-side of (3.1.5) are Dirichlet with updated parameter $((\alpha(B_1) + \sum_{1\leq j\leq n} \delta_{Y_j}(B_1), \ldots, \alpha(B_k) + \sum_{1\leq j\leq n} \delta_{Y_j}(B_k))$. This argument verifies that the Dirichlet process with parameter α satisfies the distributional equation (3.1.5). This solution is unique by Lemma 3.2.3. □

Using similar arguments to those used in Feigin and Tweedie [58] we now provide a detailed analysis of the Markov chain $\{P_m^{(n)}, m \geq 0\}$. To this aim, we make use of some results on limit theory for Markov chains on general state space (see Meyn and Tweedie [141]) together with a result in Kallenberg [104] about converge of r.p.m.s. We denote by $C(\mathbb{R})$ the set of bounded and continuous functions $g : \mathbb{X} \to \mathbb{R}$. The following lemma reduces the problem of convergence of the Markov chain $\{P_m^{(n)}, m \geq 0\}$ to the problem of weak convergence of its bounded linear functionals.

Lemma 3.2.4. *Let* $\{P_m^{(n)}, m \geq 0\}$ *be the Markov chain defined by* (3.2.1). *Then, if as* $m \to +\infty$

$$\int_{\mathbb{X}} g(x) P_m^{(n)}(\cdot, dx) \Rightarrow \text{ some } X_g \qquad \forall g \in C(\mathbb{R}), \tag{3.2.3}$$

then there exists a r.p.m. P^*, *such that* $P_m^{(n)} \Rightarrow P^*$ *as* $m \to +\infty$ *and* $X_g \overset{d}{=} \int_{\mathbb{X}} g(x) P^*(\cdot, dx)$ *for all* $g \in C(\mathbb{R})$.

Proof. The proof is an application of Lemma 5.1 in Kallenberg [104] to the Markov chain $\{P_m^{(n)}, m \geq 0\}$. □

Lemma 3.2.5. *Let* $\{P_m^{(n)}, m \geq 0\}$ *be the Markov chain defined by* (3.2.1) *and let* $g \in C(\mathbb{R})$. *Then* $\{G_m^{(n)}, m \geq 0\}$ *is a Markov chain on* $\mathbb{R}$ *whose unique invariant measure* Π_g *is the distribution of the random Dirichlet functional* G. *In particular,* $\{G_m^{(n)}, m \geq 0\}$ *converges weakly for* Π_g*-almost all starting points* $G_0^{(n)}$.

Proof. The proof is along lines similar to the proof of Lemma 2 in Feigin and Tweedie [58]. From the recursive equation (3.2.1),

$$\begin{aligned} G_m^{(n)} &= \int_{\mathbb{X}} g(x) P_m^{(n)}(\cdot, dx) \\ &= \theta_m \int_{\mathbb{X}} g(x) \sum_{j=1}^{n} q_{m,j} \delta_{Y_{m,j}}(dx) + (1-\theta_m) G_{m-1}^{(n)} = \theta_m \sum_{j=1}^{n} q_{m,j} g(Y_{m,j}) + (1-\theta_m) G_{m-1}^{(n)}. \end{aligned}$$

Then $\{G_m^{(n)}, m \geq 1\}$ is a Markov chain on $\mathbb{R}$ and in particular is a Markov chain restricted to the compact set $[-||g||, ||g||]$ where $||g|| = \sup_{\mathbb{X}} |g(x)|$. From Meyn and Tweedie [141] a

Markov chain on a compact space has at least one finite invariant measure if it is a weak Feller Markov chain, that is if the transition probability $\mathbb{P}(G_m^{(n)} \in \cdot | G_{m-1}^{(n)} = x^*)$ is a lower semicontinuous function in x^*. For a fixed $y \in \mathbb{R}$ we have

$$\begin{aligned}
&\liminf_{x \to x^*} \mathbb{P}(G_m^{(n)} \leq y | G_{m-1}^{(n)} = x) \\
&\qquad = \liminf_{x \to x^*} \mathbb{P}\left(\theta_m \sum_{j=1}^n q_{m,j} g(Y_{m,j}) \leq y - x(1-\theta_m)\right) \\
&\qquad = \liminf_{x \to x^*} \int_{(0,1)} \mathbb{P}\left(\sum_{j=1}^n q_{m,j} g(Y_{m,j}) \leq \frac{y - x(1-z)}{z}\right) \mathbb{P}(\theta_m \in dz) \\
&\qquad \geq \int_{(0,1)} \liminf_{x \to x^*} \mathbb{P}\left(\sum_{j=1}^n q_{m,j} g(Y_{m,j}) \leq \frac{y - x(1-z)}{z}\right) \mathbb{P}(\theta_m \in dz) \\
&\qquad = \int_{(0,1)} \mathbb{P}\left(\sum_{j=1}^n q_{m,j} g(Y_{m,j}) \leq \frac{y - x^*(1-z)}{z}\right) \mathbb{P}(\theta_m \in dz) \\
&\qquad = \mathbb{P}(G_m^{(n)} \leq y | G_{m-1}^{(n)} = x^*),
\end{aligned}$$

since the distribution of $\sum_{1 \leq j \leq n} q_{n,j} g(Y_{m,j})$ has at most a countable numbers of atoms and θ_m is absolutely continuous. This prove that $\{G_m^{(n)}, m \geq 0\}$ is a weak Feller Markov chain. Now, if we show that $\{G_m^{(n)}, m \geq 0\}$ is ϕ-irreducible for a finite measure ϕ, then the Markov chain is Harris positive recurrent and the invariant measure is unique (see Proposition 4.3. in Tweedie [183]). Let us consider the following event $E := \{Y_{1,1} = Y_{1,2} = \cdots = Y_{1,n}\}$. Then for a finite measure ϕ we have to prove that if $\phi(A) > 0$, then $\mathbb{P}(G_1^{(n)} \in A | G_0^{(n)}) > 0$ for any $G_0^{(n)}$. We observe that

$$\begin{aligned}
\mathbb{P}(G_1^{(n)} \in A | G_0^{(n)}) &= \mathbb{P}(G_1^{(n)} \in A | G_0^{(n)}, E)\mathbb{P}(E | G_0^{(n)}) + \mathbb{P}(G_1^{(n)} \in A | G_0^{(n)}, E^c)\mathbb{P}(E^c | G_0^{(n)}) \\
&= \mathbb{P}(G_1^{(n)} \in A | G_0^{(n)}, E)\mathbb{P}(E) + \mathbb{P}(G_1^{(n)} \in A | G_0^{(n)}, E^c)\mathbb{P}(E^c) \\
&\geq \mathbb{P}(G_1^{(n)} \in A | G_0^{(n)}, E)\mathbb{P}(E).
\end{aligned}$$

Therefore, since $\mathbb{P}(E) > 0$, we have to prove that

$$\mathbb{P}(G_1^{(n)} \in A | G_0^{(n)}, E) > 0$$

for any $G_0^{(n)}$ and for any A such that $\phi(A) > 0$. Nevertheless, given E and $G_0^{(n)}$,

$$G_1^{(n)} = \theta_1 g(Y_1) + (1-\theta_1) G_0^{(n)}$$

where θ_1 is distributed according to a Beta distribution function with parameter (n, a), so that the distribution of $G_1^{(n)} | G_0^{(n)}, E$ admits a density with respect to the Lebesgue measure on $[G_0^{(n)}, g(X_1)]$ for $G_0^{(n)} < g(X_1)$ (and similarly for $G_0^{(n)} > g(X_1)$). Then, using the same argument in Lemma 2 in Feigin and Tweedie [58], we can conclude that $\mathbb{P}(G_1^{(n)} \in A|G_0^{(n)}, E) > 0$ for a suitable measure ϕ such that $\phi(A) > 0$. We prove the aperiodicity of $\{G_m^{(n)}, m \geq 0\}$ by contradiction. If the chain is periodic with period $d > 1$, the exist d disjoint sets $D_1, \ldots, D_d$ such that for $i = 1, \ldots d-1$

$$\mathbb{P}(G_m^{(n)} \in D_{i+1}|G_{m-1}^{(n)} = x) = 1 \qquad \forall x \in D_i$$

which implies

$$\mathbb{P}\left(z\sum_{j=1}^{n} q_{m,j}g(Y_{m,j}) + (1-z)x \in D_{i+1}\right) = 1 \qquad z \text{ almost anywhere-}\lambda_{|(0,1)},$$

where $\lambda_{|(0,1)}$ is the Lebesgue measure restricted to $(0,1)$. We have

$$\mathbb{P}\left(\sum_{j=1}^{n} q_{m,j}g(Y_{m,j}) \in D_{i+1}\right) = 1 \qquad \forall i = 0, \ldots, d-1.$$

For generic α and g this is in contradiction with the assumption $d > 1$. By Theorem 13.3.4. in Meyn and Tweedie [141], there exists a unique invariant probability measure Π_g for $\{G_m^{(n)}, m \geq 0\}$. $\square$

The convergence result in Lemma 3.2.5 does not depend on $P_0^{(n)}$, so we can state the following proposition which prove that the Markov chain $\{P_m^{(n)}, m \geq 0\}$ has as unique invariant measure the law of a Dirichlet process with parameter α.

Theorem 3.2.2. *The Markov chain $\{P_m^{(n)}, m \geq 0\}$ has a unique invariant measure Π which is the law of a Dirichlet process with parameter α.*

Proof. From Lemma 3.2.5, as $m \to +\infty$, $G_m^{(n)} \Rightarrow G$ for Π_g-almost starting points $G_0^{(n)}$. Moreover, for Lemma 3.2.4 there exists a r.p.m. P^* such that $P_m^{(n)} \Rightarrow P^*$ as $m \to +\infty$. The law of P^* is the invariant measure for the Markov chain $\{P_m^{(n)}, m \geq 0\}$. Then, for Lemma 3.2.4, as $m \to +\infty$

$$\int_{\mathbb{X}} g(x) dP_m^{(n)}(\cdot, dx) \Rightarrow \int_{\mathbb{X}} g(x) P^*(\cdot, dx)$$

and the limit is unique for any $g \in C(\mathbb{R})$. Since for any random measure ζ_1 and ζ_2 we know that

$$\zeta_1 \overset{d}{=} \zeta_2 \Leftrightarrow \int_{\mathbb{X}} g(x)\zeta_1(\cdot, dx) = \int_{\mathbb{X}} g(x)\zeta_2(\cdot, dx)$$

for any $g \in C(\mathbb{R})$ (see Theorem 3.1. in Kallenberg [104]), the invariant measure for the Markov chain $\{P_m^{(n)}, m \geq 0\}$ is unique. By the definition of $\{P_m^{(n)}, m \geq 0\}$ it is straightforward to show that the limit P^* must satisfy (3.1.5) so that P^* is the Dirichlet process with parameter α. $\square$

So far we extended the convergence result proved in Feigin and Tweedie [58] to the Markov chain $\{P_m^{(n)}, m \geq 0\}$. We conclude this section by giving a result on the Harris ergodicity of the Markov chain $\{G_m^{(n)}, m \geq 0\}$. In Feigin and Tweedie [58] it is shown that $\int_{\mathbb{X}} |g(x)| P(\cdot, dx)$ is finite or infinite a.s. according to the condition $\int_{\mathbb{X}} \log(1 + |g(x)|)\alpha(dx)$ (see also Hannum et. al [83], Doss and Sellke [33], Yamato [188] and Cifarelli and Regazzini [18]). In particular in Feigin and Tweedie [58], the sufficiency is proved using arguments involving the Harris ergodicity of the Markov chain $\{G_m, m \geq 0\}$. In the next theorem we have the same sufficient condition for the Harris ergodicity of the Markov chain $\{G_m^{(n)}, m \geq 0\}$.

Theorem 3.2.3. *Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure with total mass a and $g : \mathbb{X} \to \mathbb{R}$ any measurable function. If*

$$\int_{\mathbb{X}} \log(1 + |g(x)|)\alpha(dx) < +\infty \tag{3.2.4}$$

then the Markov chain $\{G_m^{(n)}, m \geq 0\}$ is Harris ergodic with unique invariant measure Π_g.

Proof. The proof is along lines similar to the proof of Theorem 2 in Feigin and Tweedie [58]. The proof of ϕ-irreducibility is virtually identical to that of Lemma 3.2.5. On the other hand, if $f(u) = \log(1 + |u|)$, if there exists a compact set $K \subset \mathbb{R}$ and $\varepsilon > 0$ such that

$$\mathbb{E}[f(G_1^{(n)})|G_0^{(n)}] < +\infty \qquad G_0^{(n)} \in K$$

and

$$\mathbb{E}[f(G_1^{(n)})|G_0^{(n)}] \leq f(G_0^{(n)}) - \varepsilon \qquad G_0^{(n)} \in K^c,$$

Harris ergodicity follows. We have

$$\begin{aligned}
f(G_1^{(n)}) &= \log\left(1 + \left|\theta_1 \sum_{j=1}^{n} q_{1,j} g(Y_{1,j}) + (1-\theta_1) G_0^{(n)}\right|\right) \\
&= \log\left(1 + \left|(1-\theta_1^*) \sum_{j=1}^{n} q_{1,j} g(Y_{1,j}) + \theta_1^* G_0^{(n)}\right|\right) \\
&\leq \log\left(1 + (1-\theta_1^*) \left|\sum_{j=1}^{n} q_{1,j} g(Y_{1,j})\right| + \theta_1^* |G_0^{(n)}|\right) \\
&\leq \log\left(1 + \left|\sum_{j=1}^{n} q_{1,j} g(Y_{1,j})\right|\right) + \log\left(1 + \theta_1^* |G_0^{(n)}|\right) \\
&\leq \sum_{j=1}^{n} \log\left(1 + |g(Y_{1,j})|\right) + \log\left(1 + \theta_1^* |G_0^{(n)}|\right).
\end{aligned}$$

Observe that the marginal distribution of each $Y_{1,j}$ is α_0. Since $-\infty < \mathbb{E}[\log(\theta_1^*)] < +\infty$ and (3.2.4) holds, $\mathbb{E}[f(G_1^{(n)})|G_0^{(n)}] < +\infty$ for $G_0^{(n)} \in K$. Moreover, if $K = [-k, k]$, $k > 0$, since

$$f(G_1^{(n)}) \leq f\left((1-\theta_1^*)\left|\sum_{j=1}^{n} q_{1,j} g(Y_{1,j})\right| + \theta_1^*|G_0^{(n)}|\right)$$
$$= f(G_0^{(n)}) + \log(\theta_1^*) + \log\left(\frac{1 + (1-\theta_1^*)\left|\sum_{j=1}^{n} q_{1,j} g(Y_{1,j})\right| + \theta_1^*|G_0^{(n)}|}{(1+|G_0^{(n)}|)\theta_1^*}\right),$$

then, if we choose $\varepsilon = -1/2\mathbb{E}[\log(\theta_1^*)]$ and k large enough such that the following condition is true,

$$\mathbb{E}\left[\log\left(\frac{1 + (1-\theta_1^*)\left|\sum_{j=1}^{n} q_{1,j} g(Y_{1,j})\right| + \theta_1^*|K|}{(1+|K|)\theta_1^*}\right)\right] < \varepsilon,$$

we have

$$\mathbb{E}[f(G_1^{(n)})|G_0^{(n)}] \leq f(G_0^{(n)}) - \varepsilon \qquad G_0^{(n)} \in K^c.$$

Therefore, by Harris ergodicity $\{G_m^{(n)}, m \geq 0\}$ has a unique invariant distribution. Since the Dirichlet process P is the unique solution of (3.1.5) and by Theorem 3.2.2 its distribution is the unique invariant measure of $\{P_m^{(n)}, m \geq 0\}$, G must satisfy the distributional equation

$$G = \theta \sum_{j=1}^{n} q_j g(Y_j) + (1-\theta)G.$$

We conclude that the law of G is the unique invariant distribution for the chain $\{G_m^{(n)}, m \geq 0\}$. □

Condition (3.2.4) corresponds to the original condition discovered in Feigin and Tweedie [58]. Then, Propostion 3.2.3 show how, for a more general Markov chain having the law of a Dirichlet process with parameter α as invariant measure, the condition required for the existence of a functional $g \in C(\mathbb{R})$ is the same.

Theorem 3.2.4. *Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure with total mass a and let P be a Dirichlet process with parmatere α. If for any $n \in \mathbb{N}$ any $g \in C(\mathbb{R}^+)$*

$$\exists i \leq n : \mathbb{E}\left[\log\left(1 + |g(Y_{1,j})|\right)\right] = +\infty \tag{3.2.5}$$

then $\int_{\mathbb{X}} g(x)P(\cdot, dx) = +\infty$ a.s.

Proof. The r.v. $(Y_{1,1}, \ldots, Y_{1,n})$ is a sample of size n from a Blackwell-MacQueen Pólya sequence with parameter α and conditional to P, the marginal distribution of each $Y_{1,j} \sim$

α_0. Then the proof follows from Theorem 3 in Feigin and Tweedie [58] where it is proved that if

$$\int_{\mathbb{X}} \log(1 + |g(x)|)\alpha(dx) = +\infty$$

then

$$\int_{\mathbb{X}} g(x)P(\cdot, dx) = \infty \qquad \text{a.s.-}\Pi$$

□

Theorem 3.2.3 and Theorem 3.2.4 provide a condition for the existence of linear functionals of a Dirichlet process on $(\mathbb{X}, \mathscr{X})$ with parameter α and they can be summarized by the following proposition.

Theorem 3.2.5. *Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure with total mass a, let P be a Dirichlet process on $\mathbb{X}$ with parameter α and let $g \in C(\mathbb{R})$. Then $\int_{\mathbb{X}} |g(x)|P(\cdot, dx)$ is finite a.s.-Π or infinite a.s.-Π according to the condition $\int_{\mathbb{X}} \log(1 + |g(x)|)\alpha(dx)$ is finite or infinite.*

Proof. The proof follows immediately from Theorem 3.2.3 and Theorem 3.2.4 Note that, in the infinite case, $\int_{\mathbb{X}} g(x)P(\cdot, dx)$ may be undefined, or may be infinite, a.s. □

We conclude this section providing a discussion about the conjecture of using a random parameter n for the Markov chain $\{P_m^{(n)}, m \geq 0\}$ in order to prove the existence of a continuous time version of the Markov chain $\{P_m^{(n)}, m \geq 0\}$. A different approach in order to construct a continuous time version of the Markov chain $\{P_m^{(n)}, m \geq 0\}$ is described in chapter 5 where also an interesting connection with a class of measure-valued diffusion process is considered.

In order to make the Markov chain $\{P_m^{(n)}, m \geq 0\}$ in continuous time we consider the parameter n to be random. In particular, for any $t \geq 0$ we consider a r.v. N_t on $\mathbb{N}$ and we define its distribution by $d_n(t) := \mathbb{P}(N_t = n)$. We now consider a conjecture about the existence of a continuous time version $\{P_t, t \geq 0\}$ of the Markov chain $\{P_m^{(n)}, m \geq 0\}$ by looking at the Chapman-Kolmogorov equation

$$p(t + s, \mu, d\nu) = \int_{\mathcal{P}_{\mathbb{X}}} p(t, \nu, \cdot)p(s, \mu, d\nu)$$

where in general $p(t, \mu, d\nu) := \mathbb{P}(P_t \in d\nu | P_0 = \mu)$. The idea is to consider a Chapman-Kolmogorov equation from the recursive equation (3.2.1). In particular, at time $t + s$

$$P_{t+s} = \theta_t \sum_{j=1}^{N_{t+s}} q_{t,j}\delta_{Y_{t,j}} + (1 - \theta_t)P_s$$

and at time s

$$P_s = \theta_s \sum_{j=1}^{N_s} q_{s,j}\delta_{Y_{s,j}} + (1 - \theta_s)P_0$$

where we remember that now θ_t is a r.v. distributed according to a Beta distribution function with parameter (N_t, a). Then if we substitute P_{t+s} with its formal expression we can derive the followig Chapman-Kolmogorov condition

$$\theta_{t+s} \sum_{j=1}^{N_{t+s}} q_{t+s,j} \delta_{Y_{t+s,j}} + (1-\theta_{t+s}) P_0 = \theta_t \sum_{j=1}^{N_t} q_{t,j} \delta_{Y_{t,j}} + (1-\theta_t) \left(\theta_s \sum_{j=1}^{N_s} q_{s,j} \delta_{Y_{s,j}} + (1-\theta_s) P_0 \right) \tag{3.2.6}$$

What we want to find is the distribution of N_t that satisfies the equation (3.2.6). If we find the distribution of N_t such that is true the following

$$(1-\theta_{t+s}) \stackrel{d}{=} (1-\theta_t)(1-\theta_s) \tag{3.2.7}$$

then we can write

$$\theta_t \sum_{j=1}^{N_t} q_{t,j} \delta_{Y_{t,j}} + \theta_s (1-\theta_t) \sum_{j=1}^{N_s} q_{s,j} \delta_{Y_{s,j}} + (1-\theta_t)(1-\theta_s) P_0 = \theta_t \sum_{j=1}^{N_t} q_{t,j} \delta_{Y_{t,j}} + (1-\theta_t) \left(\theta_s \sum_{j=1}^{N_s} Q_{s,j} \delta_{Y_{s,j}} + (1-\theta_s) P_0 \right)$$

Then, the Chapman Kolmogorov equation (3.2.6) holds true if we prove that

$$\theta_{t+s} \sum_{j=1}^{N_{t+s}} q_{t+s,j} \delta_{Y_{t+s,j}} = \theta_t \sum_{j=1}^{N_t} q_{t,j} \delta_{Y_{t,j}} + \theta_s (1-\theta_t) \sum_{j=1}^{N_s} q_{s,j} \delta_{Y_{s,j}}$$

Equation (3.2.7) implies that $\theta_{t+s} \stackrel{d}{=} \theta_t + \theta_s(1-\theta_t)$ then to satisfy the last equation we need that $N_{t+s} \stackrel{d}{=} N_t + N_s$ that is, the number of atoms at time $t+s$ is equal to the sum of the number of atoms at time t and at time s. Therefore we conjecture that in order to prove the Chapman Kolmogorov equation (3.2.6) we need to find the distribution of N_t such that the masses and the number of locations of the random measures of the right side of (3.2.6) are equal in distribution respectively to the masses and the number locations of the random measures of the left side of (3.2.6). This mean that we have to find the distribution for N_t that satisfies the following distributional equation

$$(1-\theta_{t+s}) \stackrel{d}{=} (1-\theta_t)(1-\theta_s)$$

under the constraint on the number of location $N_{t+s} \stackrel{d}{=} N_t + N_s$.

3.3 The Feigin-Tweedie Markov chain on the finite dimensional symplex

Moving from the distributional equation (3.1.5), in Section 3.2 we provided a detailed analysis of the Markov chain $\{P_m^{(n)}, m \geq 1\}$ having state space $\mathcal{P}_{\mathbb{X}}$ and unique invariant measure the law of a Dirichlet process with parameter $\alpha \in \mathcal{A}_{\mathbb{X}}$ with α a finite measure. In this section we restrict the attention on a Markov chain $\{Q_m^{(n)}, m \geq 1\}$ which can be recoverd from the Markov chain $\{P_m^{(n)}, m \geq 1\}$ assuming the state space to be the finite dimensional space $\Delta^{(k-1)} \subset \mathcal{P}_{\mathbb{X}}$ corresponding to the space of all probability measure on $([k], \mathscr{K})$ where $\mathscr{K}$ is the σ-field of all subsets of $[k]$. In particular, we are interested to obtain the transition density of the Markov chain $\{Q_m^{(n)}, m \geq 1\}$ in order to compare it with the transition density of other Markov chain $\Delta^{(k-1)}$ with invariant measure the law of the Dirichlet distribution.

For any $k \in \mathbb{N}$, let $(a_1, \ldots, a_k)$ be a vector of positive real number such that $a := \sum_{1 \leq i \leq k} a_i$ and for any $n \in \mathbb{N}$, let $\{(Y_{m,1}, \ldots, Y_{m,n}), m \geq 0\}$ be a sequence of independent r.v.s each one a sample of size n from a Pólya sequence with parameter $(a_1, \ldots, a_k)$. Moreover, for any $n \in \mathbb{N}$, let $\{(q_{m,1}, \ldots, q_{m,n}), m \geq 1\}$ be a sequence of independent r.v.s identically distribuited according to a Dirichlet distribution function with parameter $(1, \ldots, 1)$. Finally, let $\{\theta_m, m \geq 1\}$ be sequence of independent r.v.s such that θ_m is distribuited according to a Beta distribution function with parameter (n, a). We further assume that the sequences $\{(Y_{m,1}, \ldots, Y_{m,n}), m \geq 0, j \geq 1\}$, $\{(q_{m,1}, \ldots, q_{m,n}), m \geq 1\}$ and $\{\theta_m, m \geq 1\}$ are mutually independent. Moving from such collection of random sequences we then define the Markov chain $\{Q_m^{(n)}, m \geq 0\}$ on $\Delta^{(k-1)}$ via the recursion

$$Q_m^{(n)} = \theta_m \sum_{j=1}^{n} q_{m,j} \delta_{Y_{m,j}} + (1 - \theta_m) Q_{m-1}^{(n)} \qquad m \geq 1 \tag{3.3.1}$$

where $Q_0^{(n)} \in \Delta^{(k-1)}$ is arbitrary. Since $\Delta^{(k-1)} \subset \mathcal{P}_{\mathbb{X}}$, then it can be checked that the Markov chain $\{Q_m^{(n)}, m \geq 1\}$ on $\Delta^{(k-1)}$ has the same properties of the Markov chain $\{P_m^{(n)}, m \geq 1\}$ on $\mathcal{P}_{\mathbb{X}}$ and in particular, it has as unique invariant measure, the measure the law of a Dirichlet distribution with parameter $(a_1, \ldots, a_k)$.

The Markov chain $\{Q_m^{(n)}, m \geq 1\}$ can be interpreted as a discrete time stochastic model describing the evolution of the proportions of k distinct types $A_1, \ldots, A_k$ in a population of a certain constant size. We can describe this discrete time stochastic model as follows. At each time m we associate a Pólya urn with n balls of k different colours $A_1, \ldots, A_k$ such that for any $a \in \mathbb{R}^+$

$$a_i = a\mathbb{P}(\text{observing a ball of colour } A_i) \qquad i = 1, \ldots, k$$

Therefore, $Q_m^{(n)} \in \Delta^{(k-1)}$ describes the proportions at the generation m as a mixture with random weights θ_m between the proportions $Q_{m-1}^{(n)} \in \Delta^{(k-1)}$ at the generation $m-1$ and the proportions induced by a sample $Y_{m,1}, \dots, Y_{m,n}$ of size n from a Pólya sequences, i.e. for any $m \in \mathbb{N}$

$$\mathbb{P}(Y_{m,n} = A_i | Y_{m,1}, \dots, Y_{m,n-1}) = \frac{a_i + \sum_{j=1}^{n-1} \delta_{Y_{m,j}}(A_i)}{a+n-1}.$$

The weights of the mixutre are given by a r.v. distributed according to a Beta distribution function with parameter (n, a). In the discrete time stochastic model above described the role of the Pólya urn is of primary importance. In particular, the proportions in the population changes according to the Pólya mechanism which drives the discrete time stochastic model. The interesting feature of this discrete time stochastic model is that as $m \to +\infty$ the proportions of individual types are distributed according to a Dirichlet distribution function with parameter $(a_1, \dots, a_k)$.

In the literature, the most studied class of discrete time stochastic model describing the evolution of the proportions in a population of constant size is the Wright-Fisher models (see Ewens [45]). Consider a population of fixed size n in any generation, and a single locus at which k alleles are possible. Each individual in the population is one of the k allelic types, denoted by $A_1, \dots, A_k$. Let $Q_{m,j}^{(n)}$ be the proportion of individuals of allelic type A_j at time m. Clearly

$$Q_{m,j}^{(n)} \in [0,1], \qquad \sum_{j=1}^{k} Q_{m,j}^{(n)} = 1 \qquad m \geq 0.$$

Several Markov chain models are used to describe the evolution of $Q_m^{(n)}$. We will assume selectively neutrality, and allow mutation between types. To this we define a matrix M with the following elements

$$m_{i,j} := \mathbb{P}\{\text{allele of type } A_i \text{ mutates to } A_j\} \qquad i \neq j$$

and set $m_{i,i} = 1 - \sum_{j \neq i} m_{i,j}$. If the current value of $Q_m^{(n)}$ is q, then for $j = 1, 2, \dots, k$ the fraction of allelic type A_j in the gene pool after mutation is the j-th element of the vector qM and we denote this element by π_j. The Wright-Fisher model precribes the transition probabilities

$$\mathbb{P}(Q_{m+1}^{(n)} = p | Q_{n-1}^{(n)} = q) = \binom{n}{nq_1, \dots, nq_k} \pi_1^{nq_1} \cdots \pi_k^{nj_k} \tag{3.3.2}$$

where (3.3.2) corresponds to random mating and multinomial sampling of the gene pool which is divided into fractions π_j of allelic type A_j for $j = 1, 2, \dots, k$. This model has non-overlapping generations. An analogus process in which generations overlap, can is described in Karlin and McGregor [105] and Kelly [106].

In this section our aim is to investigate the transition probabilities of the Markov chain $\{Q_m^{(n)}, m \geq 1\}$ described via the recursion (3.3.1). In particular, we are interested in finding the transition probabilities of the Markov chain $\{Q_m^{(n)}, m \geq 1\}$ and in comparing it with the transition probabilities of the Wright-Fisher model. We start considering a population of fixed size n in any generation, and a single locus at which 2 alleles are possible. Each individual in the population is one of the 2 allelic types denoted by A_1, A_2. Let $x, y \in \Delta^{(1)}$, then the model precribes the transition probabilities

$$\begin{aligned}
&\mathbb{P}(Q_m^{(n)} \leq y | Q_{m-1}^{(n)} = x) \\
&\quad = \mathbb{P}\left(\theta_m \sum_{j=1}^{n} q_{m,j}\delta_{Y_{m,j}} + (1-\theta_m)Q_{m-1}^{(n)} \leq y | Q_{m-1}^{(n)} = x\right) \\
&\quad = \mathbb{P}\left(\sum_{j=1}^{n} q_{m,j}\delta_{Y_{m,j}} \leq \frac{y-(1-\theta_m)Q_{m-1}^{(n)}}{\theta_m} | Q_{m-1}^{(n)} = x\right) \\
&\quad = \int_{(0,1)} \mathbb{P}\left(\sum_{j=1}^{n} q_{m,j}\delta_{Y_{m,j}} \leq \frac{y-(1-z)Q_{m-1}^{(n)}}{z} | Q_{m-1}^{(n)} = x\right) \mathbb{P}(\theta_m \in dz) \\
&\quad = \sum_{(n_1,n_2)\in\mathcal{D}_{2,n}} \mathbb{P}(Y_{m,1},\ldots,Y_{m,n} | Q_{m-1}^{(n)} = x) \\
&\qquad \times \int_{(0,1)} \mathbb{P}\left(\sum_{j=1}^{n} q_{m,j}\delta_{Y_{m,j}} | Y_{m,1},\ldots,Y_{m,n} \leq \frac{y-(1-z)x}{z}\right) \mathbb{P}(\theta_m \in dz)
\end{aligned}$$

where, $\sum_{1\leq j\leq n} q_{m,j}\delta_{Y_{m,j}} | Y_{m,1},\ldots,Y_{m,n}$ is a r.v. distributed according to a Beta distribution function with parameter (n_1, n_2) where $n_1 := \sum_{1\leq j\leq n_m} \delta_{Y_{m,j}}(A_1)$ and $n_2 := n - n_1$. Before considering a more explicit expression for the transition probabilities of the Markov chain $\{Q_m^{(n)}, m \geq 1\}$ we states the following lemma which relates the cumulative distribution function of a Beta distribution function with integer parameter with the cumulative distribution function of a Binomial distribution function. The results of the next lemma are expressed in terms of the Gauss hypergeometric function ${}_2F_1$ (see Appendix C).

Lemma 3.3.1. *Let X be a r.v. distributed according to a Beta distribution function with parameter (a, b) with $a, b \in \mathbb{N}$ and let Y be a r.v. distributed according to a Binomial distribution function with parameter $(a+b-1, 1-x)$ with $a, b \in \mathbb{N}$. Then*

$$\mathbb{P}(X \leq x) = \mathbb{P}(Y \leq b-1). \tag{3.3.3}$$

Proof. The proof is based on the application of a simple property of the Gauss hypergeometric function ${}_2F_1$ (see Appendix C)

$${}_2F_1(a, 1-b; a+1; x) = \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b)} \sum_{j=a}^{a+b-1} \binom{a+b-1}{j} x^{j-a}(1-x)^{a+b-1-j}.$$

Then we have

$$\begin{aligned}\mathbb{P}(X \leq x) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1}(1-t)^{b-1}dt \\ &= \frac{x^a}{a} {}_2F_1(a, 1-b; a+1; x) = (1-x)^b {}_2F_1(1, a+b; a+1; x) \\ &= \sum_{j=a}^{a+b-1} \binom{a+b-1}{j} x^j (1-x)^{a+b-1-j} = \sum_{j=0}^{b-1} \binom{a+b-1}{j} x^{a+b-1-j}(1-x)^j\end{aligned}$$

where the last step follows for a known reduction result for the Gauss hypergeometric function with integer parameter. Then, the last expression is the cumulative distribution function of a Binomial distribution with parameter $(a+b-1, 1-x)$. $\square$

Using Lemma 3.3.1, we can compute an explicit expression for the transition probabilities of the Markov chain $\{Q_m^{(n)}, m \geq 1\}$. In particular, we have

$$\begin{aligned}&\mathbb{P}(Q_m^{(n)} \leq y | Q_{m-1}^{(n)} = x) \\ &\quad = \sum_{(n_1,n_2)\in\mathcal{D}_{2,n}} \mathbb{P}(Y_{m,1}, \ldots, Y_{m,n} | Q_{m-1}^{(n)} = x) \\ &\quad \times \int_{(0,1)} \mathbb{P}\left(\sum_{j=1}^{n} q_{m,j}\delta_{Y_{m,j}} | Y_{m,1}, \ldots, Y_{m,n} \leq \frac{y-(1-z)x}{z} \right) \mathbb{P}(\theta_m \in dz) \\ &\quad = \sum_{(n_1,n_2)\in\mathcal{D}_{2,n}} \mathbb{P}(Y_{m,1}, \ldots, Y_{m,n} | Q_{m-1}^{(n)} = x) \\ &\quad \times \int_{(0,1)} \left(1 - \mathbb{P}\left(\sum_{j=1}^{n} q_{m,j}(1-\delta_{Y_{m,j}}) | Y_{m,1}, \ldots, Y_{m,n} \leq 1 - \frac{y-(1-z)x}{z} \right) \right) \mathbb{P}(\theta_m \in dz) \\ &\quad = \sum_{(n_1,n_2)\in\mathcal{D}_{2,n}} \mathbb{P}(Y_{m,1}, \ldots, Y_{m,n} | Q_{m-1}^{(n)} = x) \\ &\quad \times \sum_{j=n_1}^{n-1} \binom{n-1}{j} \int_{(0,1)} \left(\frac{y-(1-z)x}{z} \right)^j \left(1 - \frac{y-(1-z)x}{z} \right)^{n-j-1} \mathbb{P}(\theta_m \in dz).\end{aligned}$$

we can obtain the transition density for the Markov chain $\{Q_m^{(n)}, m \geq 1\}$. In particular, we have

$$\begin{aligned}
&\frac{d}{dy}\mathbb{P}(Q_m^{(n)} \leq y | Q_{m-1}^{(n)} = x) \\
&\quad = \sum_{(n_1,n_2)\in\mathcal{D}_{2,n}} \mathbb{P}(Y_{m,1},\dots,Y_{m,n}|Q_{m-1}^{(n)} = x) \\
&\qquad \times \sum_{j=n_1}^{n-1} \binom{n-1}{j} \int_{(0,1)} \frac{d}{dy}\left(\frac{y-(1-z)x}{z}\right)^j \left(1-\frac{y-(1-z)x}{z}\right)^{n-j-1} \mathbb{P}(\theta_m \in dz) \\
&\quad = \sum_{(n_1,n_2)\in\mathcal{D}_{2,n}} \mathbb{P}(Y_{m,1},\dots,Y_{m,n}|Q_{m-1}^{(n)} = x) \\
&\times \int_{(0,1)} \frac{1}{z}\left(\sum_{j=n_1}^{n-1}\binom{n-1}{j} j \left(\frac{y-(1-z)x}{z}\right)^{j-1}\left(1-\frac{y-(1-z)x}{z}\right)^{n-1-j}\right. \\
&\qquad \left. - \sum_{j=n_1}^{n-1}\binom{n-1}{j}(n-j-1)\left(\frac{y-(1-z)x}{z}\right)^{j}\left(1-\frac{y-(1-z)x}{z}\right)^{n-2-j}\right) \mathbb{P}(\theta_m \in dz) \\
&\quad = \sum_{(n_1,n_2)\in\mathcal{D}_{2,n}} \mathbb{P}(Y_{m,1},\dots,Y_{m,n}|Q_{m-1}^{(n)} = x) \\
&\qquad \times \int_{(0,1)} \frac{1}{z}\frac{\Gamma(n)}{\Gamma(n_1)\Gamma(n_2)}\left(\frac{y-(1-z)x}{z}\right)^{n_1-1}\left(1-\frac{y-(1-z)x}{z}\right)^{n_2-1} \mathbb{P}(\theta_m \in dz).
\end{aligned}$$

We observe that the transition densities of the Markov chain $\{Q_m^{(n)}, m \geq 1\}$ correspond to the densities of a suitable transormation of mixtures of Beta distribution function with parameter (n_1, n_2). In particular, if $Q_{m-1}^{(n)}$ is distributed according to a Beta distribution function with parameter (a_1, a_2), then $\mathbb{P}(Y_{m,1},\dots,Y_{m,n}|Q_{m-1}^{(n)} = x)$ is distributed according to a Binomial distribution function with parameter (x, n). Therefore, we have

$$\begin{aligned}
&\frac{d}{dy}\mathbb{P}(Q_m^{(n)} \leq y | Q_{m-1}^{(n)} = x) \\
&\quad = \sum_{(n_1,n_2)\in\mathcal{D}_{2,n}} \binom{n}{n_1} x^{n_1}(1-x)^{n_2} \\
&\qquad \times \int_{(0,1)} \frac{1}{z}\frac{\Gamma(n)}{\Gamma(n_1)\Gamma(n_2)}\left(\frac{y-(1-z)x}{z}\right)^{n_1-1}\left(1-\frac{y-(1-z)x}{z}\right)^{n_2-1} \mathbb{P}(\theta_m \in dz).
\end{aligned}$$

We now consider the extension of the above calculation to a population of sized n with k distinct types. The population has a fixed size n in any generation, and a single locus at which k alleles are possible. Each individual in the population is one of the k allelic types denoted by $A_1, \dots, A_k$. Let $x, y \in \Delta^{(k-1)}$, then the model precribes the transition probabilities

$$\begin{aligned}
&\mathbb{P}(Q_m^{(n)} \leq y | Q_{m-1}^{(n)} = x)\\
&\quad = \mathbb{P}\left(\theta_m \sum_{j=1}^{n} q_{m,j}\delta_{Y_{m,j}} + (1-\theta_m)Q_{m-1}^{(n)} \leq y | Q_{m-1}^{(n)} = x\right)\\
&\quad = \mathbb{P}\left(\sum_{j=1}^{n} q_{m,j}\delta_{Y_{m,j}} \leq \frac{y-(1-\theta_m)Q_{m-1}^{(n)}}{\theta_m} | Q_{m-1}^{(n)} = x\right)\\
&\quad = \int_{(0,1)} \mathbb{P}\left(\sum_{j=1}^{n} q_{m,j}\delta_{Y_{m,j}} \leq \frac{y-(1-z)Q_{m-1}^{(n)}}{z} | Q_{m-1}^{(n)} = x\right)\mathbb{P}(\theta_m \in dz)\\
&\quad = \sum_{(n_1,\dots,n_k)\in\mathcal{D}_{n,k}} \mathbb{P}(Y_{m,1},\dots,Y_{m,n}|Q_{m-1}^{(n)} = x)\\
&\qquad\quad \times \int_{(0,1)} \mathbb{P}\left(\sum_{j=1}^{n} q_{m,j}\delta_{Y_{m,j}}(A_1)|Y_{m,1},\dots,Y_{m,n} \leq \frac{y-(1-z)x}{z}\right)\mathbb{P}(\theta_m \in dz)
\end{aligned}$$

where $\sum_{1\leq j\leq n} q_{m,j}\delta_{Y_{m,j}}|Y_{m,1},\dots,Y_{m,n}$ is a r.v. distributed according to a Dirichlet distribution function with parameter $(n_1,\dots,n_k)$ where $n_i := \sum_{1\leq j\leq n}\delta_{Y_{m,j}}(A_i)$ for $i = 1,\dots,k-1$ and $n_k := n - \sum_{1\leq i\leq k-1} n_i$. Before considering a more explicit expression for the transition probabilities of the Markov chain $\{Q_m^{(n)}, m\geq 1\}$ we states the following lemma which relates the cumulative distribution function of a Dirichlet distribution function with integer parameter to the cumulative distribution function of a Multinomial distribution function.

Lemma 3.3.2. *Let $(X_1,\dots,X_{k-1})$ be a r.v. distributed according to a Dirichlet distribution function with parameter $(a_1,\dots,a_k)$ with $a_i \in \mathbb{N}$ for $i = 1,\dots,k$. Then*

$$\begin{aligned}
&\mathbb{P}(X_1 \leq x_1,\dots,X_{k-1} \leq x_{k-1}) &(3.3.4)\\
&\quad = \sum_{j_1=0}^{a_k-1}\sum_{j_2=0}^{j_1}\cdots\sum_{j_k=0}^{j_{k-1}} M \prod_{i=1}^{k-2} x_i^{a_i+j_{k-i}-j_{k-i+1}} x_{k-1}^{a_k+a_{k-1}-j_1-1}\left(1-\sum_{i=1}^{k-1}x_i\right)^{j_k}
\end{aligned}$$

where

$$M := \binom{\sum_{i=1}^{k} a_i - 1}{j_k, a_1+j_{k-1}-j_k,\dots,a_{k-1}+j_1-j_2, a_k+a_{k-1}-j_1-1}. \tag{3.3.5}$$

Proof. The proof applied recursively the equality for Gauss hypergeometric function with integer parameters used in Lemma 3.3.1.

$$
\begin{aligned}
&\mathbb{P}(X_1 \le x_1, \ldots, X_{k-1} \le x_{k-1}) \\
&\quad = \frac{\Gamma(\sum_{i=1}^k a_i)}{\prod_{i=1}^k \Gamma(a_i)} \int_0^{x_1} \cdots \int_0^{x_{k-1}} \prod_{i=1}^{k-1} t_i^{a_i-1} \left(1 - \sum_{i=1}^{k-1} t_i\right)^{a_k-1} dt_i \\
&\quad = \frac{\prod_{i=1}^k \Gamma(a_i)}{\Gamma(\sum_{i=1}^k a_i)} \int_0^{x_1} \cdots \int_0^{x_{k-2}} {}_2F_1\left(a_{k-1}, 1-a_k; a_{k-1}+1; \frac{x_{k-1}}{1-\sum_{i=1}^{k-2} t_i}\right) \frac{x_{k-1}^{a_{k-1}}}{a_{k-1}} \\
&\qquad\quad \times \prod_{i=1}^{k-2} t_i^{a_i-1} \left(1 - \sum_{i=1}^{k-2} t_i\right)^{a_k-1} dt_i \\
&\quad = \frac{\Gamma(\sum_{i=1}^k a_i)}{\prod_{i=1}^{k-2} \Gamma(a_i)\Gamma(a_{k-1}+a_k)} \sum_{j_1=1}^{a_k} \binom{a_{k-1}+a_k-1}{j_1} x_{k-1}^{a_{k-1}+a_k-j-2} \\
&\qquad\quad \times \int_0^{x_1} \cdots \int_0^{x_{k-2}} \prod_{i=1}^{k-2} t_i^{a_i-1} \left(1 - \sum_{i=1}^{k-2} t_i - x_{k-1}\right)^{j_1-1} dt_i \\
&\quad = \frac{\Gamma(\sum_{i=1}^k a_i)}{\Gamma(a_1)\Gamma(a_{k-1}+a_k)} \sum_{j_1=1}^{a_k} \binom{a_{k-1}+a_k-1}{j_1} x_{k-1}^{a_{k-1}+a_k-j_1-2} \\
&\qquad\quad \times \sum_{j_2=1}^{j_1} \binom{a_{k-2}+j_1-1}{j_2} x_{k-2}^{a_{k-2}+j_1-j_2-2} \frac{\Gamma(j_1)}{\Gamma(a_{k-2}+j_1)} \cdots \\
&\qquad\quad \cdots \times \sum_{j_{k-1}=1}^{j_{k-2}} \binom{a_2+j_{k-2}-1}{j_{k-1}} x_2^{a_2+j_{k-2}-j_{k-1}-2} \frac{\Gamma(j_{k-2})}{\Gamma(a_2+j_{k-2})} \\
&\qquad\quad \times \int_0^{x_1} t_1^{a_1-1}(1-t_1-x_2-\cdots-x_{k-1})^{j_{k-1}-1} dt_1 \\
&\quad = \frac{\Gamma(\sum_{i=1}^k a_i)}{\Gamma(a_1)\Gamma(a_{k-1}+a_k)} \sum_{j_1=1}^{a_k} \binom{a_{k-1}+a_k-1}{j_1} x_{k-1}^{a_{k-1}+a_k-j_1-2} \\
&\qquad\quad \times \sum_{j_2=1}^{j_1} \binom{a_{k-2}+j_1-1}{j_2} x_{k-2}^{a_{k-2}+j_1-j_2-2} \frac{\Gamma(j_1)}{\Gamma(a_{k-2}+j_1)} \cdots \\
&\qquad\quad \times \sum_{j_{k-1}=1}^{j_{k-2}} \binom{a_2+j_{k-2}-1}{j_{k-1}} x_2^{a_2+j_{k-2}-j_{k-1}-2} \frac{\Gamma(j_{k-2})}{\Gamma(a_2+j_{k-2})} \\
&\qquad\quad \times (1-\sum_{i=1}^{k-1} x_i)^{j_{k-1}-1} \frac{x_1^{a_1}}{a_1} {}_2F_1\left(a_1, 1-j_{k-1}; a_1+1; \frac{x_1}{1-\sum_{i=2}^{k-1} x_i}\right) \\
&\quad = \frac{\Gamma(\sum_{i=1}^k a_i)}{\Gamma(a_{k-1}+a_k)} \sum_{j_1=1}^{a_k} \binom{a_{k-1}+a_k-1}{j_1} x_{k-1}^{a_{k-1}+a_k-j_1-2} \\
&\qquad\quad \times \sum_{j_2=1}^{j_1} \binom{a_{k-2}+j_1-1}{j_2} x_{k-2}^{a_{k-2}+j_1-j_2-2} \frac{\Gamma(j_1)}{\Gamma(a_{k-2}+j_1)} \cdots
\end{aligned}
$$

$$\cdots \times \sum_{j_{k-1}=1}^{j_{k-2}} \binom{a_2 + j_{k-2} - 1}{j_{k-1}} x_2^{a_2+j_{k-2}-j_{k-1}-2} \frac{\Gamma(j_{k-2})}{\Gamma(a_2 + j_{k-2})}$$
$$\times \sum_{j_k=1}^{j_{k-1}-1} \binom{a_1 + j_{k-1} - 1}{j_k} x_1^{a_1+j_{k-1}-j_k-2} \frac{\Gamma(j_{k-1})}{\Gamma(a_1 + j_{k-1})} \left(1 - \sum_{i=1}^{k-1} t_i\right)^{j_k}$$
$$= \sum_{j_1=0}^{a_k-1} \sum_{j_2=0}^{j_1} \cdots \sum_{j_k=0}^{j_{k-1}} \binom{-1 + \sum_{i=1}^{k} a_i}{j_k, a_1 + j_{k-1} - j_k, \ldots, a_k + a_{k-1} - j_1 - 1}$$
$$\times \prod_{i=1}^{k-2} x_i^{a_i+j_{k-i}-j_{k-i+1}} x_{k-1}^{a_k+a_{k-1}-j_1-1} (1 - |\mathbf{x}|)^{j_k} .$$

□

Lemma 3.3.2 gives an alternative representation for the cumulative distribution function of a Dirichlet distribution function with integer parameters in terms of a cumulative distribution function of a Multinomial distribution function. It extends to the k-dimensional case Lemma 3.3.1.

In the next lemma we give a further representation for the cumulative distribution function of a Dirichlet distribution function in terms of first Lauricella multiple hypergeometric functions (see Appendix C). Then, as corollary we obtain an apparentely new representation for the first Lauricella multiple hypergeometric functions with integer parameters.

Lemma 3.3.3. *Let $(X_1, \ldots, X_{k-1})$ be a r.v. distributed according to a Dirichlet distribution function with parameter $(a_1, \ldots, a_k)$ with $a_i \in \mathbb{N}$ for $i = 1, \ldots, k$. Then*

$$\mathbb{P}(X_1 \le x_1, \ldots, X_{k-1} \le x_{k-1})$$
$$= \frac{\Gamma(\sum_{i=1}^{k} a_i)}{\prod_{i=1}^{k} \Gamma(a_i)} \int_0^1 \cdots \int_0^1 \prod_{i=1}^{k-1} t_i^{a_i-1} \left(1 - \sum_{i=1}^{k-1} t_i x_i\right)^{a_k-1} dt_1 \cdots dt_{k-1}.$$

Proof. Following [47] we have

$$\frac{\Gamma(\sum_{i=1}^{k} a_i)}{\prod_{i=1}^{k} \Gamma(a_i)} \mathbb{P}(X_1 \le x_1, \ldots, X_{k-1} \le x_{k-1}) = \int_0^{x_1} \cdots \int_0^{x_{k-1}} \prod_{i=1}^{k-1} t_i^{a_i-1} \left(1 - \sum_{i=1}^{k-1} t_i\right)^{a_k-1} .$$

If the integrand is expanded by the Multinomial theorem, the multiple series formed converges uniformly over the range of integration, so that term by term integration is justified. Then we have

$$
\begin{aligned}
&\frac{\Gamma(\sum_{i=1}^k a_i)}{\prod_{i=1}^k \Gamma(a_i)} \mathbb{P}(X_1 \le x_1, \dots, X_{k-1} \le x_{k-1}) \\
&\qquad = \sum_{(j_1,\dots,j_{k-1}) \in \mathcal{D}_{a_k,k-1}} \frac{(1-a_k)_{a_{k-1}\uparrow 1}}{j_1! \cdots j_{k-1}!} \int_0^{x_1} t_1^{a_1+j_1-1} dt_1 \cdots \int_0^{x_{k-1}} t_{k-1}^{a_{k-1}+j_{k-1}-1} dt_{k-1} \\
&\qquad = \frac{1}{\prod_{i=1}^{k-1} n_i} F_A^{(k-1)}(1-a_k, a_1, \dots, a_{k-1}; a_1+1, \dots, a_{k-1}+1; x_1, \dots, x_{k-1}).
\end{aligned}
$$

where $(x)_{y\uparrow 1}$ stands for the Pochhammer symbol for the ascending factorial of x of order y (see Appendix A). Then the result follows using the the integral rapresentation of Eulero-type for the first Lauricella multiple hypergeometric function. □

Corollary 3.3.1. *Let* $(a_1, \dots, a_k)$ *be* k *dimesional vector such that* $a_i \in \mathbb{N}$ *for* $i = 1, \dots, k$ *and let* M *be the multinomial coefficent defined by* (3.3.5)*. Then*

$$
\begin{aligned}
&F_A^{(k-1)}(1-a_k, a_1, \dots, a_{k-1}; a_1+1, \dots, a_{k-1}+1; x_1, \dots, x_{k-1}) \\
&\quad = \frac{\prod_{i=1}^k \Gamma(a_i)}{\Gamma(\sum_{i=1}^k a_i)} \prod_{i=1}^{k-1} a_i \sum_{j_1=0}^{a_k-1} \sum_{j_2=0}^{j_1} \cdots \sum_{j_k=0}^{j_{k-1}} M \prod_{i=1}^{k-2} x_i^{a_i+j_{k-i}-j_{k-i+1}} x_{k-1}^{a_k+a_{k-1}-j_1-1} \left(1-\sum_{i=1}^{k-1} x_i\right)^{j_k}.
\end{aligned}
$$

Proof. The proof follows from the two different represions for the cumulative distribution function of a Dirichlet distribution with integer parameter obtained in Lemma 3.3.2 and Lemma 3.3.3. □

On the basis of the result in Lemma 3.3.2 it is possible to obtain an explicit expression for the transition densities of the Markov chain. In the unidimensional setting we proceeded using a series representation for the cumulative distribution function of a Beta distribution function. Here we want to extend that approach using the result on the first Lauricella multiple hypergeometric functions. Using Lemma 3.3.2 we can give an expression for the transition functions for the Markov chain $\{Q_m^{(n)}, m \ge 1\}$. In particular, we have

$$
\begin{aligned}
&\mathbb{P}(Q_m^{(n)} \le y | Q_{m-1}^{(n)} = x) \\
&\quad = \sum_{(n_1,\dots,n_k) \in \mathcal{D}_{n,k}} \mathbb{P}(Y_{m,1}, \dots, Y_{m,n} | Q_{m-1}^{(n)} = x) \\
&\qquad \times \int_{(0,1)} \sum_{j_1=0}^{n_k-1} \sum_{j_2=0}^{j_1} \cdots \sum_{j_k=0}^{j_{k-1}} M \prod_{i=1}^{k-2} \left(\frac{y_i - (1-\theta_m) x_i}{\theta_m}\right)^{n_i+j_{k-i}-j_{k-i+1}} \\
&\qquad \times \left(\frac{y_{k-1} - (1-z) x_{k-1}}{z}\right)^{n_k+n_{k-1}-j_1-1} \left(1 - \sum_{i=1}^{k-1} \frac{y_i - (1-z) x_i}{z}\right)^{j_k} \mathbb{P}(\theta_m \in dz).
\end{aligned}
$$

Using the Leibnitz rule of integration we can obtain the transition density for the Markov chain $\{Q_m^{(n)}, m \geq 1\}$. In particular we have

$$
\begin{aligned}
&\frac{d^{k-1}}{dy_1 \cdots, dy_{k-1}} \mathbb{P}(Q_m^{(n)} \leq y | Q_{m-1}^{(n)} = x) \\
&\quad = \sum_{(n_1,\ldots,n_k)\in\mathcal{D}_{n,k}} \mathbb{P}(Y_{m,1}, \ldots, Y_{m,n} | Q_{m-1}^{(n)} = x) \\
&\qquad \times \frac{(n-1)!}{(n-k)!} \int_{(0,1)} \frac{1}{z^{k-1}} \binom{n-k}{n_1 - 1, \ldots, n_k - 1} \\
&\qquad \times \prod_{i=1}^{k-1} \left(\frac{y_i - (1-z)x_i}{z} \right)^{n_i - 1} \left(1 - \sum_{i=1}^{k-1} \frac{y_i - (1-z)x_i}{z} \right)^{n_k - 1} \mathbb{P}(\theta_m \in dz) \\
&\quad = \sum_{(n_1,\ldots,n_k)\in\mathcal{D}_{n,k}} \mathbb{P}(Y_{m,1}, \ldots, Y_{m,n} | Q_{m-1}^{(n)} = x) \\
&\qquad \times \frac{\Gamma(n)}{\Gamma(n_1) \cdots \Gamma(n - \sum_{i=1}^{k-1} n_i)} \\
&\qquad \times \int_{(0,1)} \frac{1}{z^{k-1}} \prod_{i=1}^{k-1} \left(\frac{y_i - (1-z)x_i}{z} \right)^{n_i - 1} \left(1 - \sum_{i=1}^{k-1} \frac{y_i - (1-z)x_i}{z} \right)^{n_k - 1} \mathbb{P}(\theta_m \in dz).
\end{aligned}
$$

Thus, we can conclude with the following propoposition.

Proposition 3.3.1. *For any $k \in \mathbb{N}$, let $\{Q_m^{(n)}, m \geq 1\}$ be a Markov chain on $\Delta^{(k-1)}$ defined via the recursion* (3.3.1)*. Then, for any $m \in \mathbb{N}$ and for any $x, y \in \Delta^{(k-1)}$, the transition densities of $\{Q_m^{(n)}, m \geq 1\}$*

$$
\begin{aligned}
&f(x, m; y) \qquad\qquad (3.3.6) \\
&\quad = \sum_{(n_1,\ldots,n_k)\in\mathcal{D}_{k,n}} \mathbb{P}(Y_{m,1}, \ldots, Y_{m,n} | Q_{m-1}^{(n)} = x) \frac{\Gamma(n)}{\Gamma(n_1) \cdots \Gamma(n - \sum_{i=1}^{k-1} n_i)} \\
&\qquad \times \int_{(0,1)} \frac{1}{z^{k-1}} \prod_{i=1}^{k-1} \left(\frac{y_i - (1-z)x_i}{z} \right)^{n_i - 1} \left(1 - \sum_{i=1}^{k-1} \frac{y_i - (1-z)x_i}{z} \right)^{n_k - 1} \mathbb{P}(\theta_m \in dz).
\end{aligned}
$$

Therefore, from Proposition 3.3.1 we observe that the transition densities of the Markov chain $\{Q_m^{(n)}, m \geq 1\}$ correspond to the densities of a suitable transormation of mixtures of a Dirichlet distribution function with parameter $(n_1, \ldots, n_k)$. In particular, if $Q_{m-1}^{(n)}$ is distributed according to a Dirichlet distribution function with parameter $(a_1, \ldots, a_k)$, then $\mathbb{P}(Y_{m,1}, \ldots, Y_{m,n} | Q_{m-1}^{(n)} = x)$ is distributed according to a Multinomial distribution function with parameter $(x_1, \ldots, x_{k-1}, n)$. Therefore, we have

$$\begin{aligned}
&\frac{d^{k-1}}{dy_1 \cdots , dy_{k-1}} \mathbb{P}(Q_m^{(n)} \leq y | Q_{m-1}^{(n)} = x) \\
&\quad = \sum_{(n_1,\ldots,n_k)\in \mathcal{D}_{k,n}} \binom{n}{n_1, \ldots, n_k} x_1^{n_1} \cdots x_{k-1}^{n_{k-1}} \left(1 - \sum_{i=1}^{k-1} x_i\right)^{n_k} \\
&\qquad \times \frac{\Gamma(n)}{\Gamma(n_1)\cdots\Gamma(n - \sum_{i=1}^{k-1} n_i)} \\
&\qquad \times \int_{(0,1)} \frac{1}{z^{k-1}} \prod_{i=1}^{k-1} \left(\frac{y_i - (1-z)x_i}{z}\right)^{n_i - 1} \left(1 - \sum_{i=1}^{k-1} \frac{y_i - (1-z)x_i}{z}\right)^{n_k - 1} \mathbb{P}(\theta_m \in dz).
\end{aligned}$$

3.4 Rates of convergence of the mean functional Markov chain

In this section our aim is to investigate the rate of convergence in total variation of the mean functional Markov chain $\{M_m^{(n)}, m \geq 0\}$ to the mean functional of a Dirichlet process on $\mathbb{R}$ with parameter α.

Let $\mathbb{X} = \mathbb{R}$ and let $\{P_m^{(n)}, m \geq 0\}$ be the Markov chain on $\mathcal{P}_{\mathbb{R}}$ defined by (3.2.1). We consider the mean functional Markov chain $\{M_m^{(n)}, m \geq 0\}$ defined recursively by

$$M_m^{(n)} = \theta_m \sum_{j=1}^{n} q_{m,j} Y_{m,j} + (1-\theta_m) M_{m-1}^{(n)} \qquad m \geq 1 \tag{3.4.1}$$

where $M_0^{(n)} \in \mathbb{R}$ is arbitrary. For a detailed analysis of rates of convergence in total variation of the Markov chain $\{M_m, m \geq 0\}$ to the mean functional M of a Dirichlet process see Guglielmi and Tweedie [81] and Jarner and Tweedie [103]. From Theorem 3.2.3 we know that under the condition

$$\int_{\mathbb{R}} \log(1+|x|)\alpha(dx) < +\infty \tag{3.4.2}$$

the Markov chain $\{M_m^{(n)}, m \geq 0\}$ has the distribution $\mathscr{M}$ of M as unique invariant measure. We denote the smallest interval containing the support of α by $[L, U]$, $-\infty \leq L < U \leq +\infty$, that is,

$$\alpha((-\infty, L) \cup (U, +\infty)) = 0 \tag{3.4.3}$$

and no smaller closed interval has this property. We recall that for a real-valued measurable function $g : \mathbb{X} \to \mathbb{R}$, all the results in this section can be applied to more general linear functional G of a Dirichlet process on an arbitrary Polish space $(\mathbb{X}, \mathscr{X})$, since G has the same distribution as $G \circ \alpha_g$ where $\alpha_g(B) := \alpha(g^{-1}(B))$ for any $B \in \mathscr{R}$.

Before introducing the main result of this section, we recall some definition and some results recently developed for Markov chain with general state-space. Fop definition not given below, see Nummelin [146] and Meyn and Tweedie [141]. We consider a Markov chain $\{\Phi_m, m \geq 0\}$ with state space $\mathbb{S}$, $\mathbb{S}$ being a Borel subset of $\mathbb{R}$, not necessarily countable, endowed with σ-field $\mathscr{R} \cap \mathbb{S}$. Such Markov chain $\{\Phi_m, m \geq 0\}$, with transition probabilities $p^m(x, A) = \mathbb{P}(\Phi_m \in A)|\Phi_0 = x)$ for $m \geq 1$, is called Harris ergodic if and only if there exists a probability measure π on $(\mathbb{S}, \mathscr{R} \cap \mathbb{S})$, called an invariant or limiting distribution, such that

$$\lim_{m \to +\infty} ||p^m(x, \cdot) - \pi|| = 0 \qquad \forall x \in \mathbb{S}$$

where $\|P_1(\cdot) - P_2(\cdot)\| := \sup_{A \in \mathscr{R}} |P_1(A) - P_2(A)|$, P_1 and P_2 being probability measures. A Harris ergodic chain $\{\Phi_m, m \geq 0\}$ with invariant distribution π is said to be geometrically ergodic, or to converge geometrically, it there exists $0 < \rho < 1$ and a non-negative function R on $\mathbb{S}$ with $\int_{\mathbb{S}} R(x)\pi(dx) < +\infty$, such that

$$||p^m(x, \cdot) - \pi|| \geq R(x)\rho^m \qquad \forall x \in \mathbb{S}$$

The chain is said to be uniformly ergodic, or to converge uniformly, if as $m \to +\infty$

$$\sum_x ||p^m(x, \cdot) - \pi|| \to 0$$

or, equivalently, if there exist $0 < \rho < 1$ and a positive constant R independent of x, such that

$$||p^m(x, \cdot) - \pi|| \leq R\rho^m \qquad \forall x \in \mathbb{S}.$$

Finally, we consider the definition of small set and drift function. A set C is called small if there exists $n_0 \in \mathbb{N}$, $\varepsilon > 0$ and a probabilty measure φ on $(\mathbb{R}, \mathscr{R})$ such that

$$p^{m_0}(x, A) \geq \varepsilon\varphi(A) \qquad \forall A \in \mathscr{R}, \forall x \in C.$$

The chain $\{\Phi_m, m \geq 0\}$ satisfies the geometric Foster-Lyapunov drift condition if there exist a function $V : \mathbb{R} \to [1, +\infty)$, a small set C and constants $b < +\infty$, $0 < \lambda < 1$, such that, for all $x \in \mathbb{S}$

$$PV(x) := \int_{\mathbb{S}} P(x, dy)V(y) \leq \lambda V(x) + b\mathbb{1}_C(x). \tag{3.4.4}$$

When (3.4.4) holds then the chain is known to be geometrically ergodic. Finally, we say that a real-valued chain $\{\Phi_m, m \geq 0\}$ is stocastically monotone if $x \mapsto \mathbb{P}(\Phi_m < s|\Phi_0 = x)$ is a non-increasing function for $m \geq 1$ and any $s \in \mathbb{R}$.

The next theorem is an extension of Theorem 1 in Guglielemi and Tweedie [81] to the Markov chain $\{M_m^{(n)}, m \geq 0\}$. In particular, using results developed for Markov chains on a general state space (see Meyn and Tweedie [141]) we provide some further properties for the Markov chain $\{M_m^{(n)}, m \geq 0\}$. For any $x \in \mathbb{R}$ and for any $A \in \mathscr{R}$, we define $p^m(x, A) := \mathbb{P}(M_m^{(n)} \in A | M_0^{(n)} = x)$. Then, our aim is to show that, provided $\int_{\mathbb{R}} |x|\alpha(dx) < +\infty$, we have for some $\rho < 1$

$$||p^m(x, \cdot) - \mathscr{M}(\cdot)|| < R(x)\rho^n \qquad \forall x \in \mathbb{R} \tag{3.4.5}$$

Theorem 3.4.1. *The Markov chain $\{M_m^{(n)}, m \geq 0\}$ has the following properties*

i) $\{M_m^{(n)}, m \geq 0\}$ *is a stocastically monotone Markov chain;*

ii) if further

$$\mathbb{E}[|Y_{1,1}|] = \int_{\mathbb{R}} |x|\alpha(dx) < +\infty, \tag{3.4.6}$$

then $\{M_m^{(n)}, m \geq 0\}$ *is a geometrically ergodic Markov chain;*

iii) if the support of α is bounded then $\{M_m^{(n)}, m \geq 0\}$ *is an uniformly ergodic Markov chain.*

Proof. Given the definition of stochastically monotone Markov chain, we have that for $z_1 < z_2$, $s \in \mathbb{R}$,

$$\begin{aligned} p^1(z_1, (-\infty, s)) &= \mathbb{P}\left(\theta_1 \sum_{j=1}^{n} q_{1,j} Y_{1,j} + (1-\theta_1) z_1 < s\right) \\ &\geq \mathbb{P}\left(\theta_1 \sum_{j=1}^{n} q_{1,j} Y_{1,j} + (1-\theta_1) z_2 < a\right) = p^1(z_2, (-\infty, s)). \end{aligned}$$

We now show that under condition (3.4.6) the Markov chain $\{M_m^{(n)}, m \geq 0\}$ satisfies the Foster-Lyapunov condition for the function $V(x) = 1 + |x|$. This fact implies the geometric ergodicity of the $\{M_m^{(n)}, m \geq 0\}$ (see Meyn and Tweedie [141], Chapter 15). We have

$$\begin{aligned} pV(x) &= \int_{\mathbb{X}} (1 + |y|) p(x, dy) \\ &= 1 + \mathbb{E}\left[\left|\theta_1 \sum_{j=1}^{n} q_{1,j} Y_{1,j} + (1-\theta_1)x\right|\right] \\ &\leq 1 + \mathbb{E}[\theta_1] \sum_{j=1}^{n} E\left[|q_{1,j} Y_{1,j}|\right] + |x|\mathbb{E}[1-\theta_1] \\ &\leq 1 + \frac{n}{n+a} \sum_{j=1}^{n} \mathbb{E}\left[|q_{1,j} Y_{1,j}|\right] + \frac{a}{n+a}|x| = 1 + \frac{n}{n+a}\mathbb{E}[|Y_{1,1}|] + \frac{a}{n+a}|x|. \end{aligned}$$

Now we have to find the small set $C^{(n)}$ such that the Foster-Lyapunov condition holds, i.e. we have to find the small set $C^{(n)}$ such that

$$1 + \frac{n}{n+a}\mathbb{E}[|Y_{1,1}|] + \frac{a}{n+a}|x| \leq \lambda(1+|x|) + b\mathbb{1}_{C^{(n)}}(x) \tag{3.4.7}$$

for some constant $b < +\infty$ and $0 < \lambda < 1$. Then if we consider the following set

$$C^{(n)} = [-K^{(n)}(\lambda), K^{(n)}(\lambda)], \tag{3.4.8}$$

where

$$K^{(n)}(\lambda) := \frac{1 - \lambda + n/(n+a)\mathbb{E}[|Y_{1,1}|]}{\lambda - a/(n+a)}, \tag{3.4.9}$$

then, condition (3.4.7) holds for all

$$\lambda \in \left(\frac{a}{n+a}, 1\right), \qquad b \geq 1 - \lambda + \frac{n}{n+a}\mathbb{E}[|Y_{1,1}|].$$

As in Lemma 3.2.5 we can prove that the Markov chain $\{M_m^{(n)}, m \geq 0\}$ is weak Feller; then, since $C^{(n)}$ is a compact set, it is a small set (see Tweedie [182]).

Finally, we prove that the Markov chain $\{M_m^{(n)}, m \geq 0\}$ is uniformly ergodic when the support of α is bounded. Since for $\lambda \to a/(n+a)$ the small set $C^{(n)} = [-\infty, +\infty]$, then we can choose λ such that $C^{(n)} \supseteq [L, U]$, where $[L, U]$ is the support of α. Then the uniform ergodicity holds by the same arguments used in Theorem 1 in Guglielmi and Tweedie [81]. □

We briefly comment on the results obtained in Theorem 3.4.1. Theorem 3.4.1 shows that under the same condition as in Theorem 1 in Guglielmi and Tweedie [81], the Markov chain $\{M_m^{(n)}, m \geq 0\}$ is geometrically or uniformly ergodic. In particular, the small sets $C^{(n)}$ generalize the corresponding small set C obtained in Theorem 1 in Guglielmi and Tweedie [81] which can be recovered by setting $n = 1$

$$C = [-K(\lambda), K(\lambda)] \tag{3.4.10}$$

where

$$K(\lambda) := \frac{1 - \lambda + 1/(1+a)\mathbb{E}[|Y_{1,1}|]}{\lambda - a/(1+a)}. \tag{3.4.11}$$

In that paper, it is clear from the illustrative examples that the convergence of the chain $\{M_m, m \geq 0\}$ is slow when the total mass a is large. For the Markov chain $\{M_m^{(n)}, m \geq 0\}$ the size of the small set $C^{(n)}$ can be controlled by a further parameter n; this fact suggests that the upper bounds of the rate of convergence of the Markov chain $\{M_m^{(n)}, m \geq 0\}$ depends of n. In particular, we observe that the small sets $C^{(n)}$ correspond to the small sets C when the total mass of the measure α is chosen to be a/n. According to the rate of convergence obtained in Guglielmi and Tweedie [81] these observations imply that as n increases the Markov chain $\{M_m^{(n)}, m \geq 0\}$ converges geometrically faster (in total variation) to M than the Markov chain $\{M_m, m \geq 0\}$ (this is not surprising since a greater number of r.v.s have been sampled per iteration).

Remark 3.4.1. *Condition* (3.4.6) *can be relaxed. If the following condition holds*

$$\mathbb{E}[|Y_{1,1}|^s] = \int_{\mathbb{R}} |y|^s \alpha_0(dx) < +\infty \qquad \text{for some } 0 < s < 1, \tag{3.4.12}$$

then the Markov chain $\{M_m^{(n)}, m \geq 0\}$ *is geometrically ergodic. In this case, as in the proof of Theorem 2.3. in Jarner and Tweedie [103], let* $V(x) = 1 + |x|^s$. *Then, if* $\mathbb{E}[(1 + |\sum_{1\leq j\leq n} q_{1,j} Y_{1,j}|)^s] < +\infty$, *it is straightforward to prove that the Foster-Lyapunov condition* $PV(x) \leq \lambda V(x) + b\mathbb{1}_{\tilde{C}^{(n)}}(x)$ *for some constant* $b < +\infty$,

$$\mathbb{E}[(1-\theta_1)^s] = \frac{\Gamma(a+s)\Gamma(a+n)}{\Gamma(a)\Gamma(a+s+n)} < \lambda < 1$$

and for some compact set $\tilde{C}^{(n)}$. *Of course* (3.4.12) *implies*

$$\mathbb{E}[(1 + |\sum_{j=1}^n q_{1,j} Y_{1,j}|)^s] < +\infty$$

if fact conditioning on the random number N *of distinct values* $Y_{1,1}^*, \ldots Y_{1,N}^*$ *in* $Y_{1,1}, \ldots, Y_{1,n}$, $1 \leq N \leq n$, *we have*

$$\left|\sum_{j=1}^n q_{1,j} Y_{1,j}\right| \leq \sum_{j=1}^N \tilde{q}_{1,j} |Y_{1,j}^*| \leq \max\{|Y_{1,1}^*|, \ldots, |Y_{1,N}^*|\}.$$

The $\{Y_{1,1}^*, \ldots, Y_{1,N}^*\}$ *are i.i.d. according to* α_0, *so that*

$$\begin{aligned}\mathbb{E}\left[\left|\sum_{j=1}^n q_{1,j} Y_{1,j}\right|^s\right] &\leq \int_0^{+\infty} y^s N (A_0(y))^{N-1} \alpha_0(dy) \\ &\leq N \int_0^{+\infty} y^s \alpha_0(dy) \leq n\mathbb{E}[|Y_{1,j}|^s] < +\infty,\end{aligned}$$

where A_0 *is the distribution function corresponding to the probability measure* α_0, *and this is equivalent to* $\mathbb{E}[(1 + |\sum_{1\leq j\leq n} q_{1,j} Y_{1,j}|)^s] < +\infty$.

For instance if α_0 *is a Cauchy standard distribution and* $a > 0$, *condition* (3.4.12) *is fullfilled so that the Markov chain* $\{M_m^{(n)}, m \geq 0\}$ *will turn out to be geometrically ergodic for any fixed* n.

Following a similar construction to that one proposed in Guglielmi and Tweedie [81] we now establish analytic upper bounds on the rate of convergence. If we consider the event $E := \{Y_{1,1} = Y_{1,2} = \cdots = Y_{1,n}\}$ then, for any $A \in \mathscr{R}$ and any $x \in \mathbb{R}$,

$$\begin{aligned} p^1(x, A) &\geq \mathbb{P}(M_1^{(n)} \in A | E, M_{0,}^{(n)} = x)\mathbb{P}(E) \\ &= \frac{\Gamma(n-1)\Gamma(a)}{\Gamma(a+n)}(\alpha_0(\{x\})\delta_x(A) + (1 - \alpha_0(\{x\})) \\ &\qquad \times \mathbb{P}\left(\theta_1 Y_{1,1} + (1-\theta_1)x \in A | Y_{1,1} \neq x\right)) \\ &\geq \varrho \frac{\Gamma(n-1)\Gamma(a)}{\Gamma(a+n)} \mathbb{P}\left(\theta_1 Y_{1,1} + (1-\theta_1)x \in A | Y_{1,1} \neq x\right) \end{aligned}$$

where

$$0 < \varrho := 1 - \sup\left(\alpha_0(\{x\})\right) \leq 1$$

and the supremum being taken on the set of discontinuity points of $A_0(x) := \alpha_0((-\infty, x])$ for any $x \in \mathbb{R}$. Thus, we can consider the conditional denstiy $f_{M_1^{(n)}|M_0^{(n)},Y_{1,1},E}(z)$ of the r.v. $M_1^{(n)}$ given $M_0^{(n)} = x$, $Y_{1,1} \neq x$ and E, which is

$$\begin{aligned} &f_{M_1^{(n)}|M_0^{(n)},Y_{1,1},E}(z) \qquad (3.4.13) \\ &= \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \int_{\mathbb{R}\cap\{x\}^c} \frac{(z-x)^{n-1}(y_{1,1}-z)^{a-1}}{(y_{1,1}-x)^{a+n-2}|y_{1,1}-x|} \mathbb{1}_{(0,1)}\left(\frac{z-x}{y_{1,1}-x}\right) A_0(dy_{1,1}) \\ &= \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \left(\int_{(-\infty,x)} \frac{(x-z)^{n-1}(z-y_{1,1})^{a-1}}{(x-y_{1,1})^{a+n-1}} \mathbb{1}_{(0,1)}\left(\frac{x-z}{x-y_{1,1}}\right) A_0(dy_{1,1}) \right. \\ &\qquad \left. + \int_{(x,+\infty)} \frac{(z-x)^{n-1}(y_{1,1}-z)^{a-1}}{(y_{1,1}-x)^{a+n-1}} \mathbb{1}_{(0,1)}\left(\frac{z-x}{y_{1,1}-x}\right) A_0(dy_{1,1})\right). \end{aligned}$$

Therefore, if $z < x$

$$f^*_{M_1^{(n)}|M_0^{(n)},Y_{1,1},E}(z) = \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \int_{(-\infty,z)} \frac{(x-z)^{n-1}(z-y_{1,1})^{a-1}}{(x-y_{1,1})^{a+n-1}} A_0(dy_{1,1})$$

and if $z > x$

$$f^*_{M_1^{(n)}|M_0^{(n)},Y_{1,1},E}(z) = \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \int_{(z,+\infty)} \frac{(z-x)^{n-1}(y_{1,1}-z)^{a-1}}{(y_{1,1}-x)^{a+n-1}} A_0(dy_{1,1}).$$

Now, if we find that $\inf_{x\in C^{(n)}} f_{M_1^{(n)}|M_0^{(n)},Y_{1,1},E}(z) \geq p_0(z)$, where $p_0(z)$ is some density, then

$$\varepsilon_0 := \int_{\mathbb{R}} p_0(z)dz > 0$$

implies

$$p^1(x, A) \geq \varrho\varepsilon_0 \int_A \frac{p_0(z)}{\varepsilon_0} dz = \varepsilon \int_A \frac{p_0(z)}{\varepsilon_0} dz, \qquad A \in \mathbb{R}, \qquad (3.4.14)$$

where $\varepsilon := \varrho\varepsilon_0$. Observe that for any $\eta < 1/2$, if

$$S_1 := \left\{(z,y) \in \mathbb{R}^2 : z < -K(\lambda)^{(n)}, \frac{z}{\eta} + K(\lambda)^{(n)}\left(\frac{1}{\eta} - 1\right) < y < z\right\},$$

and

$$S_2 := \left\{(z,y) \in \mathbb{R}^2 : z > K(\lambda)^{(n)}, z < y < \frac{z}{\eta} - K(\lambda)^{(n)}\left(\frac{1}{\eta} - 1\right)\right\},$$

we have

$$\begin{aligned}\int_{\mathbb{R}} \inf_{x\in C^{(n)}} f_{M_1^{(n)}|M_0^{(n)},Y_{1,1},E}(z)dz &\geq \int_{S_1} \eta^{n-1}\frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)}\frac{(z-y_{1,1})^{a-1}}{(x-y)^a}A_0(dy_{1,1})dz \\ &\quad + \int_{S_2} \eta^{n-1}\frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)}\frac{(y_{1,1}-z)^{a-1}}{(y-x)^a}A_0(dy_{1,1})dz \\ &\geq \int_{S_1} \eta^{n-1}\frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)}\frac{(z-y_{1,1})^{a-1}}{(K^{(n)}(\lambda)-y)^a}A_0(dy_{1,1})dz \\ &\quad + \int_{S_2} \eta^{n-1}\frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)}\frac{(y_{1,1}-z)^{a-1}}{(y-K^{(n)}(\lambda))^a}A_0(dy_{1,1})dz,\end{aligned}$$

which is defined to be ε_0. Therefore, if

$$\begin{aligned}p_0(z) := &\int_{(z/\eta+K^{(n)}(\lambda)(1/\eta-1),z)} \eta^{n-1}\frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)}\frac{(z-y_{1,1})^{a-1}}{(K^{(n)}(\lambda)-y)^a}A_0(dy_{1,1})\mathbb{1}_{(-\infty,-K(\lambda)^{(n)})}(z) \\ &+ \int_{(z,z/\eta-K^{(n)}(\lambda)(1/\eta-1))} \eta^{n-1}\frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)}\frac{(y_{1,1}-z)^{a-1}}{(y-K^{(n)}(\lambda))^a}A_0(dy_{1,1})\mathbb{1}_{(K(\lambda)^{(n)},+\infty)}(z)\end{aligned}$$

then condition (3.4.14) is verified and

$$\begin{aligned}\varepsilon = (1-\sup(\alpha_0(\{x\})))\Bigg(&\int_{S_1} \eta^{n-1}\frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)}\frac{(z-y_{1,1})^{a-1}}{(K^{(n)}(\lambda)-y)^a}A_0(dy_{1,1})dz \\ &+ \int_{S_2} \eta^{n-1}\frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)}\frac{(y_{1,1}-z)^{a-1}}{(y-K^{(n)}(\lambda))^a}A_0(dy_{1,1})dz\Bigg).\end{aligned} \tag{3.4.15}$$

Once ε has been determined, we can use the minorization to establish rate of geometric convergence of the Markov chain $\{M_m^{(n)}, m \geq 0\}$. In paticular we can apply results in Section 4 of Guglielmi and Tweedie [81] to recover the actual bound on on $\|p^m(x,\cdot) - \mathscr{M}(\cdot)\|$.

We first consider the geometric ergodicity under the condition of "one side" parameter α and stochastic monotonicity. Suppose that the interval $[L,U]$ on one side or the other: for convenience, assume that $-\infty < L$. Then the set $[L,+\infty)$ is again absorbing and we can restrict attention to this set. We will use the drift condition in (3.4.7) and also the stochastic monotonicity of Theorem 3.4.1. We take $C = [L, K^{(n)}(\lambda)]$ and apply the results of Roberts and Tweedie [167] to the Markov chain $\{M_m^{(n)}, m \geq 0\}$ in this case. Take $d := 1 + K^{(n)}(\lambda)$, and ε as defined in (3.4.15). We define

$$J := d + \lambda^{-1}(b - \varepsilon)$$

$$\eta := \frac{\log(J/(1-\varepsilon))}{\log(\lambda^{-1})}$$

$$\zeta(x) := V(x)\mathscr{M}((-\infty, x]) + \int_{(z,+\infty)} V(z)\mathscr{M}(dz)$$

$$\xi(x) := \frac{\log(\zeta(x))}{\log(\lambda^{-1})}$$

With these values we have

Theorem 3.4.2. *Suppose* $\mathbb{E}[|Y_{1,1}|] < +\infty$ *and let* $M_0 = x \in [L, +\infty)$. *Then*

i) if $J < 1$, *then for* $m > \xi(x) + \eta(1-\varepsilon)/(\lambda^\eta - (1-\varepsilon))$,

$$||p^m(x,\cdot) - \mathscr{M}(\cdot)|| \leq \zeta(x)\frac{\varepsilon}{1-J}\lambda^m$$

ii) if $J \geq 1$, *and* $\rho = (1-\varepsilon)^{\eta^{-1}}$, *for* $m > \xi(x) + \eta(1-\varepsilon)/\varepsilon$

$$||p^m(x,\cdot) - \mathscr{M}(\cdot)|| \leq \zeta(x)^{\log(\rho)/\log(\lambda)}\frac{e\varepsilon(m - \xi(x) + \eta)}{\eta}\rho^m$$

iii) if $J \geq 1$, *then for any* $1 > w > \rho = (1-\varepsilon)^{\eta^{-1}}$

$$||p^m(x,\cdot) - \mathscr{M}(\cdot)|| \leq \frac{(1-(1-\varepsilon)/w)w^{-\xi(x)}}{1-(1-\varepsilon)w^\eta}w^m$$

Proof. i) and ii) are a direct corollary of Theorem 2.2 in Roberts and Tweedie [167]. For iii), we can use the same method of proof as in that theorem, but use Equation 33 in Roberts and Tweedie [166] rather than Theorem 5.1 of Roberts and Tweedie [166]. □

We observe that $\zeta(x)$ is generally unknown, even if it can bounded by

$$\zeta(x) \leq V(x) + \int_{\mathbb{R}} V(z)\mathscr{M}(dz) \leq V(x) + \frac{b}{1-\lambda} < +\infty \tag{3.4.16}$$

as noted by Roberts and Tweedie [167]. This bound is very rough and in many cases it might be better to estimate $\zeta(x)$ by

$$Z_V^{m^*} = V(x)\frac{1}{m}\sum_{i=1}^{m^*}\mathbb{1}_{(-\infty,x]}(M_i) + \frac{1}{m}\sum_{i=1}^{m}\mathbb{1}_{(x,+\infty)}(M_i)(1+|M_i|)$$

for m big enough, using the Law of Large Numbers for Markov chains.

We now consider the geometric rate of convergence and other rate of convergences when the parameter α has doubly infinite support. When α has support that is unbounded on both side, we cannot use the stochastic monotonicity results and we turn to the results on Roberts and Tweedie [166]. There are two different approaches given in that paper. First, we can use the results of Theorem 5.1 and Theorem 5.2 of Roberts and Tweedie [166]. As they show, we can construct a bivariate test function $h(x, y) := (V(x) - V(y))/2$, which satisfies the drift condition (for the bivariate kernel denote P_2 and the test set $C_2^{(n)} = C^{(n)} \times C^{(n)}$)

$$P_2 h \leq \lambda_2 h + b_2 \mathbb{1}_{C_2^{(n)}}$$

where

$$\lambda_2 := \lambda + \frac{2}{2(1+d)}, \qquad b_2 := \frac{b(1+2d)}{2+2d}. \tag{3.4.17}$$

For $\lambda_2 < 1$ we require $d = 1 + K^{(n)}(\lambda) > b/2(1-\lambda) - 1$. The definition of $K^{(n)}(\lambda)$ shows that if we choose λ close enough to the minimal value $a/a+n$ then this can always be achived although the resulting set $C^{(n)}$ may ger to be somewhat large and this might result in ε becoming unacceptably small. With these values we now construct

$$J_2 := d + \lambda_2^{-1}(b_2 - \varepsilon)$$

$$\eta_2 := \frac{\log(J_2/(1-\varepsilon))}{\log(\lambda_2^{-1})}$$

$$\zeta_2(x) := V(x)\mathscr{M}((-\infty, x]) + \int_{(z,+\infty)} V(z)\mathscr{M}(dz)/2$$

$$\xi_2(x) := \frac{\log(\zeta_2(x))}{\log(\lambda_2^{-1})}$$

Following Roberts and Tweedie [166] let $m' := m - \xi_2(x)$ and define $\hat{\beta}_m$ as

$$\hat{\beta}_m = \begin{cases} \lambda_2^{-1} & J_2 < 1 \\ (1-\varepsilon)^{-1/\eta_2}(1 + \eta_2/m')^{-1/\eta_2} & J_2 \geq 1 \end{cases}$$

As corollary of Theorem 5.1 and of Equation 33 in Roberts and Tweedie [166] we then have the general bounds.

Theorem 3.4.3. *Suppose* $\mathbb{E}[|Y_{1,1}| < +\infty]$ *Then,*

(i) if $J_2 < 1$*, then for* $m > \xi_2(x) + \eta_2(1-\varepsilon)/(\lambda_2^{\eta_2} - (1-\varepsilon))$*,*

$$||p^m(x, \cdot) - \mathscr{M}(\cdot)|| \leq \zeta_2(x)\frac{1 - \lambda_2(1-\varepsilon)}{1 - J_2}\lambda_2^m$$

(ii) if $J_2 \geq 1$*, then for* $m > \xi_2(x) + \eta_2(1-\varepsilon)/\varepsilon$

$$||p^m(x, \cdot) - \mathscr{M}(\cdot)|| \leq (1 - \hat{\beta}_n(1-\varepsilon))\left(1 + \frac{m'}{\eta_2}\right)(\hat{\beta}_m)^{-m'}$$

(iii) if $J_2 \geq 1$, then for any $1 > w > \rho = (1-\varepsilon)^{\eta_2^{-1}}$

$$||p^m(x,\cdot) - \mathscr{M}(\cdot)|| \leq \frac{(1-(1-\varepsilon)/w)w^{-\xi_2(x)}}{1-(1-\varepsilon)w^{\eta_2}} w^m$$

Note again that in these bounds we need the value $\zeta_2(x)$. This can be bounded, again inefficiently in many contexts, by $V(x)+b/(1-\lambda)/2$; alternatively it could be estimated by the Law of Large Numbers using the sum $V(x) + \sum_{i=1}^n V(M_i)/n$. As a second approach, from Theorem 4.2 of Roberts and Tweedie [166], we have

Theorem 3.4.4. *The partial sums of the transient laws converge to to $\mathscr{M}_\alpha$ with a bound given by*

$$\left\|\frac{1}{m}\sum_{k=1}^{m} p^{m^*}(x,\cdot) - \mathscr{M}(\cdot)\right\| \leq \frac{1}{m}\left\{2 + \frac{(\log(V(x)) + \log(b/1-\lambda) + 2/\varepsilon\log(J/1-\varepsilon))}{\log(\lambda^{-1})}\right\} \tag{3.4.18}$$

Although this is not a bound on geometric convergence, it still serves the purpose from the point of simulations. If we set m large enough that the right side of (3.4.18) is small enough to meet our approximation criteria, then we can draw from the distribution $m^{-1}\sum_{1\leq i\leq m} p^{m^*}(x,\cdot)$ by first drawing a value of m^* uniformly in the range $\{1,2,\ldots,m\}$ and then drawing from the resultant distribution $p^{m^*}(x,\cdot)$.

Using the above results, $\mathscr{M}$ can be estimate by generating from the Markov chain $\tilde{M}_m$, with m fixed, and the approximation error between the empirical distribution function of a sample from $M_m^{(n)}$ and the distribution function $\mathscr{M}$ decreases geometrically. In particular we can approximate $\mathscr{M}$ using the empirical distribution of a sample of size k from the m-th step distribution of the approximating Markov chain $\{M_m^{(n)}, m \geq 0\}$, starting from any point x. Using the same notation of Guglielmi and Tweedie [81] we denote by $\mathscr{M}^{m,x}$ the distribution of $M_m^{(n)}$ given $\tilde{M}_0 = x$ and by $\mathscr{M}_k^{m,x}$ the empirical distribution of a sample of size k from the distribution of $M_m^{(n)}$ given $\tilde{M}_0 = x$. Then, for any fixed $n \in \mathbb{N}$

$$\sup_{t\in\mathbb{R}} |\mathscr{M}_k^{m,x}(t) - \mathscr{M}(t)| \leq \sup_{t\in\mathbb{R}} |\mathscr{M}_k^{m,x}(t) - \mathscr{M}^{m,x}(t)| + \sup_{t\in\mathbb{R}} |\mathscr{M}^{m,x}(t) - \mathscr{M}_\alpha(t)|$$

As regard the first term on the right side, there are some inequality to handle it. Since our purpose is to give a comparison with respect to the results obtained in Guglielmi and Tweedie [81] we use the Dvoretzy-Kiefer-Wolfowitz inequality. If F is any distribution function on $\mathbb{R}$, and F_k is the empirical distribution of a sample of size k from F, then this inequality states that

$$\mathbb{P}\left(\sup_x |F_k(x) - F(x)| < \frac{s}{\sqrt{k}}\right) \geq 1 - 58\mathrm{e}^{2s^2} \qquad s > 0$$

(see for instance Serfling [172]). As regard the second term on the right is obviously less than $||p^m(x,\cdot) - \mathscr{M}(\cdot)||$, which can be bounded as in (3.4.5) under appropriate circumstances.

Here we map out some chains for particular choices of α_0. In particular, we highlight the known and obvious result that the convergence of the Markov chain $\{M_m^{(n)}, m \geq 0\}$ improves as n increases.

Example 3.4.1. *Let α_0 be a Uniform distribution on $(0,1)$ and let a be the total mass. In this case $\mathbb{E}[|Y_{1,j}|] = 1/2$ so that for any fixed n, the chain will be geometrically ergodic; moreover it can be proved that $(0,1)$ is small so that the chain is uniformly ergodic. When $a = 1$, Guglielmi and Tweedie [81] showed that the convergence of $\{M_m, m \geq 0\}$ is a very good and there is no need to consider the chain with $n > 1$. We consider the cases $a = 10$, 50 and 100, and for each of them we run the chain for $n = 1$, 2, 10 and 20.*

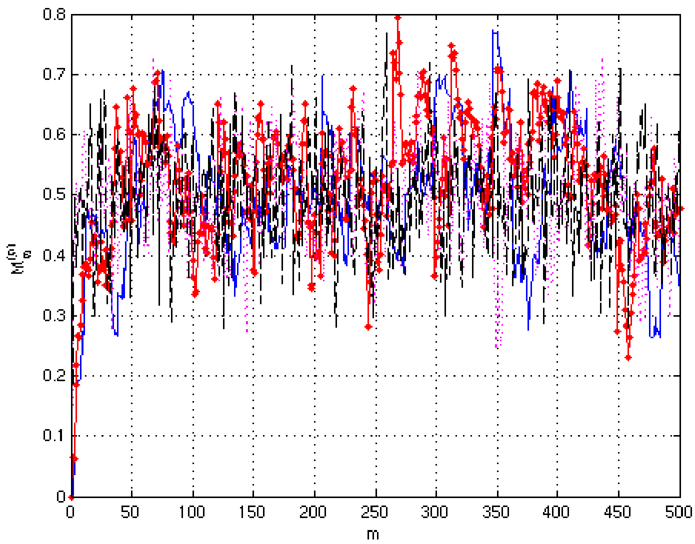

Figure 3.1: Traceplots of the Markov chain $\{M_m^{(n)}, m \geq 0\}$ with α_0 the Uniform distribution on $(0,1)$, $a = 10$ and $n = 1$ (solid blue line), $n = 2$ (dashdot red line), $n = 10$ (dashed magenta line) and $n = 20$ (dotted black line).

We found that the traceplots do not depend on the initial values. In Figures 3.1, 3.2 and 3.3 we give the traceplots of $M_m^{(n)}$ when $M_0^{(n)} = 0$; observe that convergence improves as n increses and it is more evident for larger a. When $a = 50$ the convergence of the chain $\{M_m^{(n)}, m \geq 0\}$ seems to happen before $m = 50$. When $a = 100$ the convergence of the chain for $n = 1$ seems to occur at about $m = 350$, while $m = 20$ the convergence is at about a value between 50 and 75.

Example 3.4.2. *Let α_0 be a Gaussian distribution with parameter $(0,1)$ and let $a = 10$.*

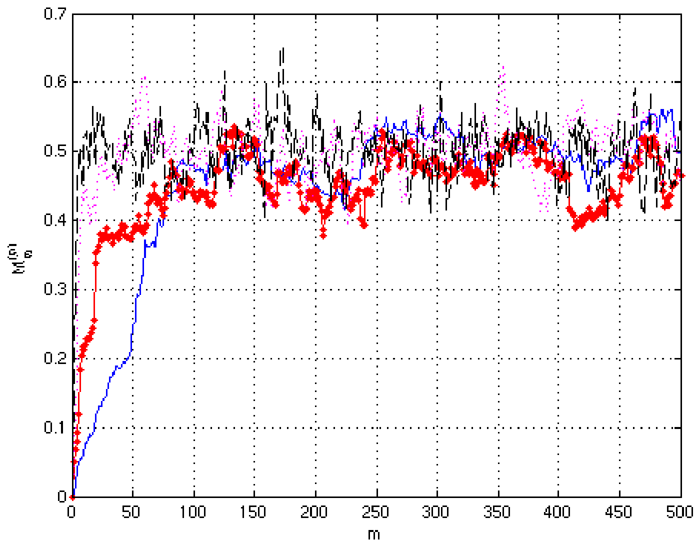

Figure 3.2: Traceplots of the Markov chain $\{M_m^{(n)}, m \geq 0\}$ with α_0 the Uniform distribution on $(0, 1)$, $a = 50$ and $n = 1$ (solid blue line), $n = 2$ (dashdot red line), $n = 10$ (dashed magenta line) and $n = 20$ (dotted black line).

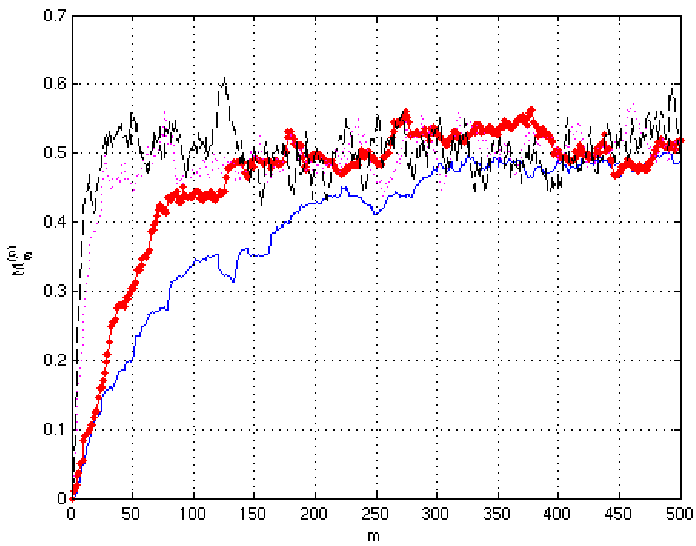

Figure 3.3: Traceplots of the Markov chain $\{M_m^{(n)}, m \geq 0\}$ with α_0 the Uniform distribution on $(0, 1)$, $a = 100$ and $n = 1$ (solid blue line), $n = 2$ (dashdot red line), $n = 10$ (dashed magenta line) and $n = 20$ (dotted black line).

The behaviour of $M_m, m \geq 0$ has been considered in Guglielmi and Tweedie [81].

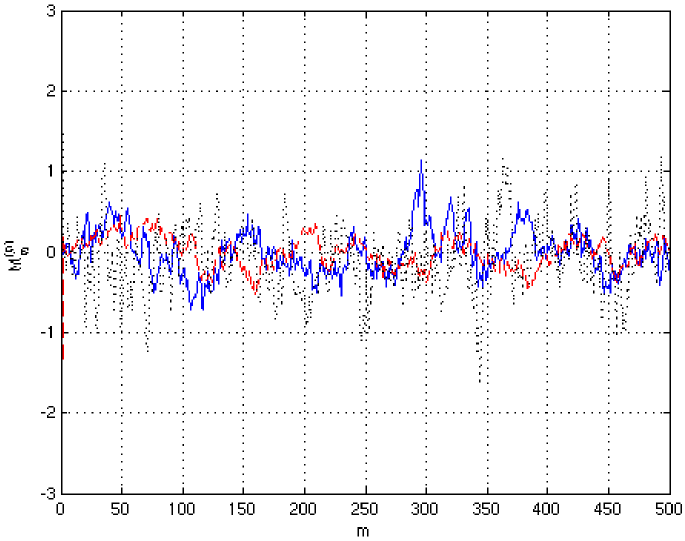

Figure 3.4: Traceplots of the Markov chain $\{M_m^{(n)}, m \geq 0\}$ with α_0 the Gaussian distribution with parameter $(0,1)$, a=10, $n = 1$ and $M_0^{(1)} = -3$ (dashed red line), $M_0^{(1)} = 0$ (solid blue line) and $M_0^{(1)} = 3$ (dotted black line).

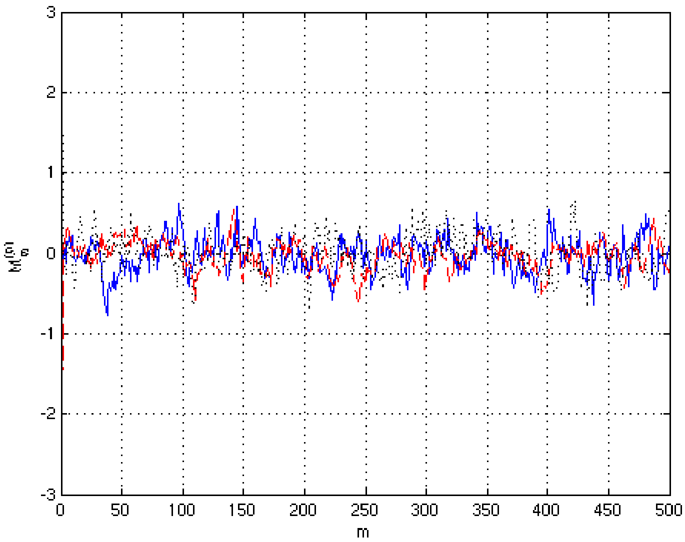

Figure 3.5: Traceplots of the Markov chain $\{M_m^{(n)}, m \geq 0\}$ with α_0 the Gaussian distribution with parameter $(0,1)$, a=10, $n = 10$ and $M_0^{(1)} = -3$ (dashed red line), $M_0^{(1)} = 0$ (solid blue line) and $M_0^{(1)} = 3$ (dotted black line).

Figures 3.4, 3.5 and 3.6 displays the traceplots of $M_m^{(n)}, m \geq 0$ *for three different initial values* ($M_0^{(n)} = -3, 0, 3$)*. Also in this case, it is clear that the convergence improves as* n *increases.*

Example 3.4.3. *Let* α_0 *be distributed according to a Cauchy standard distribution and let* $a > 1$*. In this case it is known that* $\mathscr{M}$ *is the Cauchy standard distribution. Of course,* $\mathbb{E}[|Y_{1,j}|^s] < +\infty$ *for any* $0 < s < 1$*, so that the Markov chain* $\{M_m^{(n)}, m \geq 0\}$ *is geometrically ergodic for any fixed* n*. We consider the cases* $a = 10$*,* 50 *and* 100*, and for each of them we run the chain for* $n = 1$*,* 2*,* 10 *and* 20*.*

We checked that the empirical distribution from $M_m^{(1)}$ *for* $m =$ *well approximates the exact distribution.*

3.5 On a Volterra equation associated to the mean of a Dirichlet process

In Section 3.5 we consider consider the problem of approximating the distribution $\mathscr{M}$ of the mean M of a Dirichlet process with parameter α. In particular we used the the empirical distribution of a sample from the m-th stepdistribution of the approximating Markov chain $\{M_m, m \geq 0\}$ starting from any point $x \in \mathbb{R}$. An exact expression of the distribution function $\mathscr{M}$ was obtained by Cifarelli and Regazzini [18]. Their procedure is

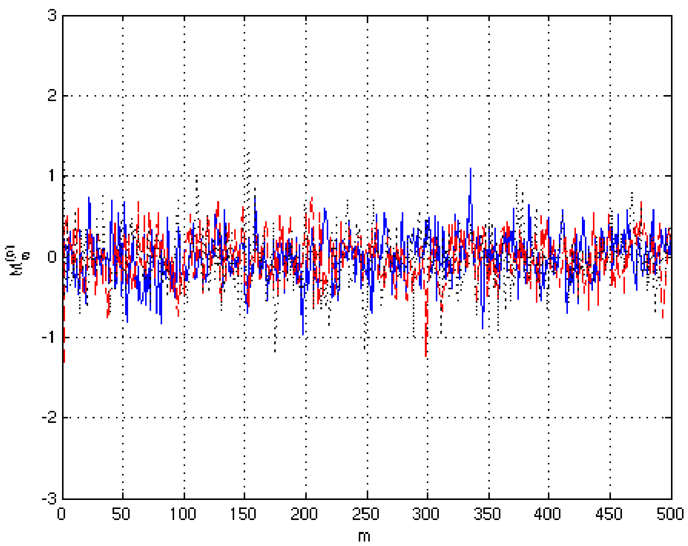

Figure 3.6: Traceplots of the Markov chain $\{M_m^{(n)}, m \geq 0\}$ with α_0 the Gaussian distribution with parameter $(0, 1)$, a=10, $n = 20$ and $M_0^{(1)} = -3$ (dashed red line), $M_0^{(1)} = 0$ (solid blue line) and $M_0^{(1)} = 3$ (dotted black line).

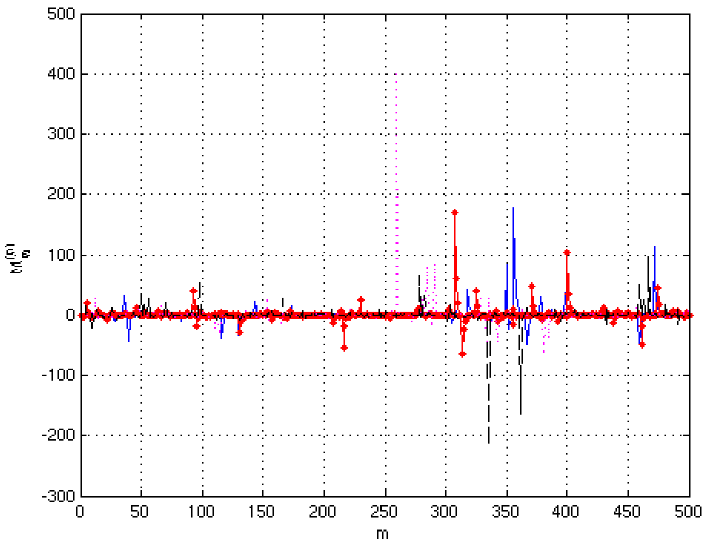

Figure 3.7: Traceplots of the Markov chain $\{M_m^{(n)}, m \geq 0\}$ with α_0 the standard Cauchy distribution, $a = 10$ and $n = 1$ (solid blue line), $n = 2$ (dashdot red line), $n = 10$ (dashed magenta line) and $n = 20$ (dotted black line).

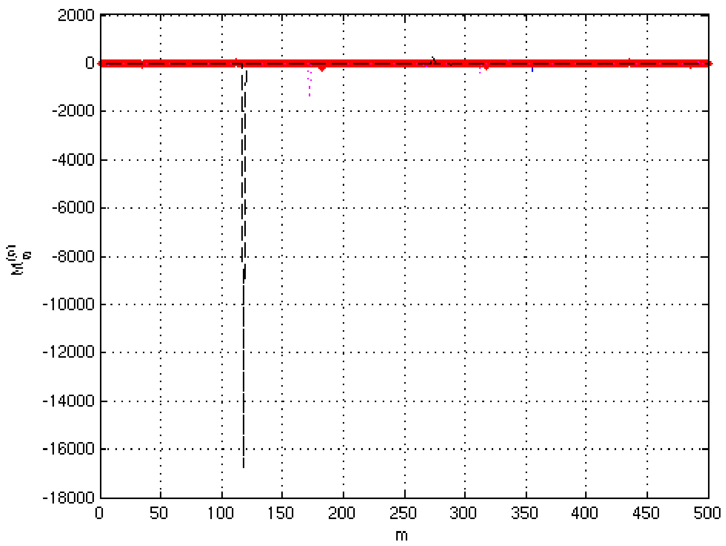

Figure 3.8: Traceplots of the Markov chain $\{M_m^{(n)}, m \geq 0\}$ with α_0 the standard Cauchy distribution, $a = 50$ and $n = 1$ (solid blue line), $n = 2$ (dashdot red line), $n = 10$ (dashed magenta line) and $n = 20$ (dotted black line).

based on the use of the generalized Stieltjes transform, of order a, of M that we denote by $\mathscr{S}$.

Let P be a Dirichlet process on $\mathbb{R}$ with parameter $\alpha \in \mathcal{A}_{\mathbb{X}}$ where α is a finite measure and denote by $A : \alpha((-\infty, x])$. Then $M = \int_{-\infty}^{0} F(\cdot, x)dx + \int_{0}^{+\infty}(1 - F(\cdot, x))dx$ where $F(\cdot, x) := P(\cdot, (-\infty, x])$ is an extended real r.v. defined on $(\Omega, \mathscr{F})$. The starting point in Cifarelli and Regazzini [18] is the explicit expression for the generalized Stieltjes transform of the truncated r.v. $U(\tau, T) := \int_{\tau}^{T}(1-F(\cdot, x))dx$, $\tau < T$, which yields $\mathscr{S}$ by the well known relation between $U(\tau, T)$ and M. First of all, they observe that $\{F(t), -\infty < t < +\infty\}$ is a Markov process and by the Markov property they are able to find a recursive relation for the moments $\mu(\tau, F(\tau); T)$ of the conditional distritution of $U(\tau, T)$ given $F(\tau)$. More precisely, it turns out that

$$\mu_n(\tau, F(\tau); T) = (1 - F(\tau))^n \mu_n^*(\tau, T)$$

where $\mu_n^*(\tau, T)$ are positive constants, recursively defined. If $M_T(\tau, F(\tau); \cdot)$ is the distribution function of $U(\tau, T)$ given $F(\tau)$, and G_T is the integral transform of $M_T(\tau, F(\tau); \cdot)$ defined by

$$G_T(\tau; z) := \int_{[0,+\infty)} (1 + (yz/(1 - x)))^{A(\tau)-a} dM_T(\tau, x; dy)$$

for $z \geq 0$, $x \in [0, 1]$ then for each n, the n-th coefficient of the power series expansion of $G_T(\tau; \cdot)$ contains $\mu_n^*(\tau, T)$; therefore it can be checked that G_T satisfies a first order partial

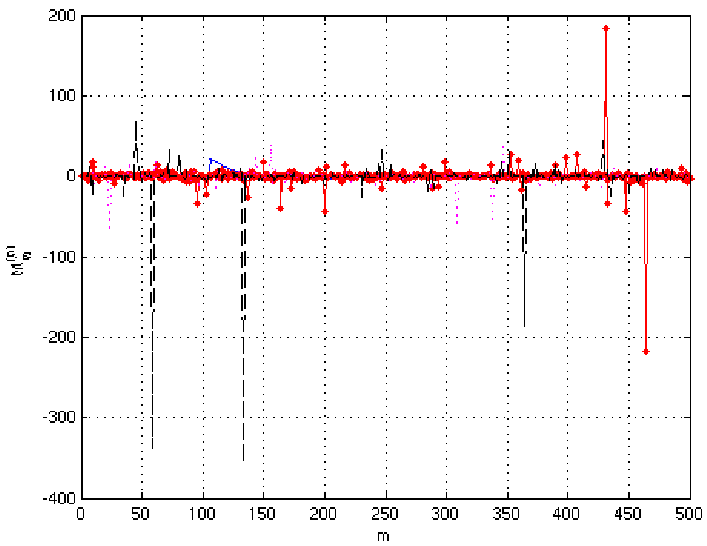

Figure 3.9: Traceplots of the Markov chain $\{M_m^{(n)}, m \geq 0\}$ with α_0 the standard Cauchy distribution, $a = 100$ and $n = 1$ (solid blue line), $n = 2$ (dashdot red line), $n = 10$ (dashed magenta line) and $n = 20$ (dotted black line).

differential equation, whose solution is explicitely given as a function of α. Besides, since the moments of $U(\tau,T)$ are linked to $\mu_n^*(\tau,T)$ by $\mathbb{E}[U(\tau,T)]^n = b_n\mu_n^*(\tau,T)$ for all $n \in \mathbb{N}$, where b_n are suitable coefficients, G_T can be expressed as a power series with coefficients depending on $\mathbb{E}[U(\tau,T)]^n$. Thus,

$$\int_{[0,+\infty)} \frac{1}{(s+y)^a} M_T(\tau; dy) = \frac{1}{s^a} \mathrm{e}^{-\int_\tau^T (a - A(v)/s+v-\tau)dv} \tag{3.5.1}$$

denoting distribution function of $U(\tau,T)$ by $M_T(\tau;\cdot)$, so that the left hand side of (3.5.1) represents the generalized Stieltjes transform of $M_T(\tau;\cdot)$.

In Guglielmi [80] a simple and direct procedure to calculate $\mathscr{S}$, when the support of α is bounded from below, is proposed. In particular, when the support of α is bounded from below, using the distributional equation (3.1.1) for the Dirichlet process on $\mathbb{R}$, a first-type Volterra equation for the Laplace transform m of M, is provided. From this Volterra equation, it follows that the Laplace transform of $x^{a-1}m(x)$, say $f(t;m)$, that is equal to the generalized Stieltjes transform of order a of M, satisfies the first order ordinary differential equation

$$-\frac{d}{dt} f(t;m) = f(t;m) \int_0^{+\infty} \mathrm{e}^{-tx}\hat{\alpha}(x)dx \qquad t > -\tau \tag{3.5.2}$$

whose solution is explicitely given. Then, this expression is extended to a more general case, when α has support bounded from below, if almost all trajectories of P are probability measures with finite expectation.

In this section we consider the same approach proposed by Guglielmi [80] using the more general distributional equation (3.1.5) for the Dirichlet process on $\mathbb{R}$. When the support of α is bounded from below, using the distributional equation (3.1.5) we recover a new second-type homogeneous Volterra equation for the Laplace transform m of M. In particular, from this Volterra equation, it follows that the Laplace transform of $x^{a-1}m(x)$, satisfies a first order ordinary differential equation whose associated homogeneous differential equation is somehow reminiscent of (3.5.1).

For any $n \in \mathbb{N}$ we denote by $\tilde{A}$ the distribution function of the r.v. $\sum_{1\le j\le n} q_j Y_j$ and in particular, we recall that conditionally to $Y_1,\ldots,Y_n$, $\tilde{A}$ is the distribution function of the mean of a Dirichlet process with parameter $\sum_{1\le j\le n}\delta_{Y_j}$. For any $t \in \mathbb{R}$, let $\hat{\alpha} := \int_{\mathbb{X}} \mathrm{e}^{-tx}\tilde{A}(dx)$, $m(t) := \int_{\mathbb{X}} \mathrm{e}^{-tx}\mathscr{M}(dx)$ and let S be the support of α.

Proposition 3.5.1. *If $-\infty < \tau \le \inf S$, $\sup S \le T < \infty$, the Laplace transform m of M satisfies the following second-type Volterra equation*

$$tm(t) = \frac{\Gamma(n+a)}{\Gamma(n)\Gamma(a)} \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k \int_0^t \left(\frac{x}{t}\right)^{a+k-1} m(x)\hat{\alpha}(t-x)dx \tag{3.5.3}$$

for $t \in \mathbb{R}$.

Proof. Moving from the distributional equation (3.1.5) which has as unique solution the Dirichlet process with parameter α, we can obtain the following equation for the mean of the Dirichlet process

$$\begin{aligned} M &= \int_{\mathbb{X}} xP(\cdot, dx) \\ &= \int_{\mathbb{X}} x\theta \sum_{j=1}^{n} q_j \delta_{Y_j}(dx) + \int_{\mathbb{X}} x(1-\theta)P(\cdot, dx) \\ &= \theta \sum_{j=1}^{n} q_j \int_{\mathbb{X}} x\delta_{Y_j}(dx) + (1-\theta)\int_{\mathbb{X}} xP(\cdot, dx) \\ &= \theta \sum_{j=1}^{n} q_j Y_j + (1-\theta)M. \end{aligned}$$

If we define $\tilde{\theta} = 1 - \theta$, then

$$\begin{aligned} m(t) &= \mathbb{E}[\mathrm{e}^{-t((1-\tilde{\theta})\sum_{j=1}^{n} q_j Y_j + \tilde{\theta} M)}] \\ &= \frac{\Gamma(n+a)}{\Gamma(n)\Gamma(a)} \int_0^1 x^{a-1}(1-x)^{n-1} \int_{\mathbb{R}} \mathbb{E}\left[\mathrm{e}^{-t((1-y)x+yM)} | \sum_{j=1}^{n} q_j Y_j = x, \tilde{\theta} = y\right] \tilde{A}(dx)dy \\ &= \frac{\Gamma(n+a)}{\Gamma(n)\Gamma(a)} \int_0^1 y^{a-1}(1-y)^{n-1} m(ty)\hat{\alpha}(t-ty)dy \\ &= \frac{\Gamma(n+a)}{\Gamma(n)\Gamma(a)} \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k \int_0^t \left(\frac{u}{t}\right)^{a+k-1} \frac{1}{t} m(u)\hat{\alpha}(t-u)du. \end{aligned}$$

□

Equation (3.5.3) is a second-type Volterra equation of the form with factor $\Gamma(n+a)/\Gamma(n)\Gamma(a)$ and kernel

$$K(t,x) = \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k \left(\frac{x}{t}\right)^{a+k-1} \hat{\alpha}(t-x).$$

It can be easly checked that if we set $n = 1$ we obtain the first-type Volterra equation for the Laplace transform m of M proposed by Gugliemi [80], i.e. a first-type Volterra equation with kernel

$$K(t,x) = \left(\frac{x}{t}\right)^{a-1} \hat{\alpha}(t-x).$$

Then, we can consider the following result.

Proposition 3.5.2. *If* $-\infty < \tau \le \inf S$, $\sup S \le T < +\infty$, *the generalized Stieltjes transform of order* a *of* M *is*

$$\int_{[0,+\infty)} \frac{1}{(x+s)^a} M(d(\tau+x)) = e^{-\int_{[0,T-\tau)} \log(s+x) A(d(\tau+x))} \qquad s > 0. \tag{3.5.4}$$

Proof. From Equation (3.5.3) we have

$$\begin{aligned}
&\int_0^{+\infty} \mathrm{e}^{-tx} x^a m(x)dx \\
&\quad = \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \sum_{k=0}^{n-1} (-1)^k \int_0^{+\infty} \mathrm{e}^{-tx} x^{a-1} \int_0^x \left(\frac{v}{x}\right)^{a+k-1} m(v)\hat{\alpha}(x-v)dvdx \\
&\quad = \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \sum_{k=0}^{n-1} (-1)^k \int_0^{+\infty} \mathrm{e}^{-tx} x^{a+k-1} m(x)dx \int_0^{+\infty} \mathrm{e}^{-tu}\hat{\alpha}(u)du \\
&\quad = \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \int_0^{+\infty} \mathrm{e}^{-tx} x^{a-1} m(x)dx \int_0^{+\infty} \mathrm{e}^{-tu}\hat{\alpha}(u)du \\
&\qquad + \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \sum_{k=1}^{n-1} \binom{n-1}{k} (-1)^k \int_0^{+\infty} \mathrm{e}^{-tx} x^{a+k-1} m(x)dx \int_0^{+\infty} \mathrm{e}^{-tu}\hat{\alpha}(u)du
\end{aligned}$$

i.e., for $t > -\tau$

$$\begin{aligned}
&-\frac{d}{dt} \int_0^{+\infty} \mathrm{e}^{-tx} x^{a-1} m(x)dx \\
&\quad = \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \int_0^{+\infty} \mathrm{e}^{-tx} x^{a-1} m(x)dx \int_0^{+\infty} \mathrm{e}^{-tx}\hat{\alpha}(x)dx \\
&\qquad + \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \sum_{k=1}^{n-1} \binom{n-1}{k} (-1)^k \int_0^{+\infty} \mathrm{e}^{-tx} x^{a+k-1} m(x)dx \int_0^{+\infty} \mathrm{e}^{-tx}\hat{\alpha}(x)dx
\end{aligned}$$

which is a first order ordinary differential equation whose associated homogeneous differential equation is somehow reminiscent of (3.5.1). In particular, we can write the first order ordinary differential equation as

$$\begin{aligned}
-\frac{d}{dt} f(t,m) = f(t,m) \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \int_0^{+\infty} \mathrm{e}^{-tx}\hat{\alpha}(x)dx + \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \sum_{k=1}^{n-1} \binom{n-1}{k} (-1)^k \\
\times \int_0^{+\infty} \mathrm{e}^{-tx} x^{a+k-1} m(x)dx \int_0^{+\infty} \mathrm{e}^{-tx}\hat{\alpha}(x)dx
\end{aligned}$$

whose solution is

$$\begin{aligned}
f(t,m) &= f(t_0,m)\mathrm{e}^{S(t)} + \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \sum_{k=1}^{n-1} \binom{n-1}{k} (-1)^k \int_{t_0}^t \mathrm{e}^{S(t)-S(w)} \\
&\qquad \times \int_0^{+\infty} \mathrm{e}^{-wx} x^{a+k-1} m(x)dx \int_0^{+\infty} \mathrm{e}^{-wx}\hat{\alpha}(x)dx \\
&= f(t_0,m) t^{-a(\Gamma(a+n)/\Gamma(a)\Gamma(n))} \mathrm{e}^{-(\Gamma(a+n)/\Gamma(a)\Gamma(n)) \int_{[t,\tau]} \log(1+x/t/t_0+x)dA(x)} \\
&\qquad + \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \sum_{k=1}^{n-1} \binom{n-1}{k} (-1)^k \int_{t_0}^t \mathrm{e}^{S(t)-S(w)} \\
&\qquad \times \int_0^{+\infty} \mathrm{e}^{-wx} x^{a+k-1} m(x)dx \int_0^{+\infty} \mathrm{e}^{-wx}\hat{\alpha}(x)dx
\end{aligned}$$

where

$$S(t) := \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \log(t^{-a} \mathrm{e}^{-\int_{[t,\tau]} \log(1+x/t/t_0+x) dA(x)}).$$

Since as $t \to +\infty$

$$f(t,m) = \Gamma(a) \int_{[\tau,T]} \frac{1}{(t+y)^a} M(dy) \to 1$$

then

$$1 = \lim_{t\to+\infty} f(t_0,m)\mathrm{e}^{S(t)} + \lim_{t\to+\infty} f(t_0,m) \frac{\Gamma(a+n)}{\Gamma(a)\Gamma(n)} \sum_{k=1}^{n-1} \binom{n-1}{k} (-1)^k \int_{t_0}^{t} \mathrm{e}^{S(t)-S(w)} \times \int_0^{+\infty} \mathrm{e}^{-wx} x^{a+k-1} m(x) dx \int_0^{+\infty} \mathrm{e}^{-wx} \hat{\alpha}(x) dx$$

$$= \frac{f(t_0,m)}{\Gamma(a)} \mathrm{e}^{\int_{[\tau,T]} \log(t_0+x) A(dx)}.$$

□

In particular, as proved by Guglielmi [80], Equation (3.5.4) is true under more general α, i.e. when the support of α is only bounded from below.

Proposition 3.5.3. *(cfr. Guglielmi [80]) If $-\infty < \tau \leq \inf S$, and*

$$\mathbb{P}\left(\left\{\omega : \int_{\mathbb{X}} |x| P(\omega, dx) < +\infty\right\}\right) = 1$$

then

$$\int_{[0,+\infty)} \frac{1}{(s+x-\tau)^a} M(dx) = e^{-\int_{[\tau,+\infty)} \log(s+x-\tau) A(dx)} \qquad s > 0. \tag{3.5.5}$$

Expression (3.5.5) is equal to the one obtained by Cifarelli and Regazzini [18] from (3.5.1).

3.6 Discussion

In this chapter we defined and we investigated a new measure-valued Markov chain $\{P_m^{(n)}, m \geq 0\}$ having as unique invariant measure the law of a Dirichlet process. For any fixed $n \in \mathbb{N}$, the Markov chain is defined via the recursion

$$P_m^{(n)} = \theta_m \sum_{j=1}^{n} q_{m,j} \delta_{Y_{m,j}} + (1-\theta_m) P_{m-1}^{(n)} \qquad m \geq 1$$

which generalizes the recursion

$$P_m = \theta_m \sum_{j=1}^{n} \delta_{Y_{m,}} + (1-\theta_m) P_{m-1} \qquad m \geq 1$$

originally used by Feigin and Tweedie [58] to define a measure-valued Markov chain $\{P_m, m \geq 0\}$ having as unique invariant measure the law of a Dirichlet process. In particular, the Markov chain $\{P_m, m \geq 0\}$ sits in the large class of Markov chains indexed by an integer $n \in \mathbb{N}$ $\{\{P_m^{(n)}, m \geq 0\} : n \in \mathbb{N}\}$ and they worked solely on the case $n = 1$, where n can be viewed as a sample size.

For any real-valued measurable function $g : \mathbb{X} \to \mathbb{R}$ we proved that the linear functional Markov chain $\{G_m^{(n)}, m \geq 0\}$ associated to $\{P_m^{(n)}, m \geq 0\}$ keeps all the properties of the linear functional Markov chain $\{G_m, m \geq 0\}$ associated to $\{P_m, m \geq 0\}$. We used this properties in order to study the rate of convergence in total variation of the mean functional Markov chain $\{M_m^{(n)} \geq 0\}$ to the mean functional of a Dirichlet process and in particular we observed that the upper bounds of the rate of convergence depends of the parameter n: as n increases the mean functional Markov chain$\{M_m^{(n)}, m \geq 0\}$ converges geometrically faster (in total variation) to the mean functional of a Dirichelt process than the Markov chain $\{M_m, m \geq 0\}$.

Recently a multidimensional version of the Markov chain $\{G_m, m \geq 0\}$ has been proposed by Erhardsson [38] and applied to develop a new method to carry out Bayesian inference for the inverse problem of estimating a finite measure μ from noisy observations of a finite number of integrals. In particular, let the function $g_i : \mathbb{X} \to \mathbb{R}$ for $i = 1, \ldots, n$, be measurable and linearly independent, and define $\bar{g} : \mathbb{X} \to \mathbb{R}^n$ by $\bar{g} = (g_1, \ldots, g_n)$. Let $\{P_m, m \geq\}$ the Markov chain defined via recursion (3.1.2) and let $\{\bar{G}_m, m \geq 0\}$ be the $\mathbb{R}^n$-valued Markov chain defined via the recursion

$$\bar{G}_m = \theta_m Y_m + (1 - \theta_m)\bar{G}_{m-1} \qquad m \geq 1 \tag{3.6.1}$$

where $\bar{G}_m := \int_{\mathbb{X}} \bar{g}(x) P_m(\cdot, dx)$ for $m \geq 0$. Then in Erhardsson [38] the following theorem is proved

Theorem 3.6.1. *(cfr. Erhardsson [38]) Let $\alpha \in \mathcal{A}_{\mathbb{X}}$ be a finite measure and let P be a Dirichlet process with parameter α. Let $g_i : \mathbb{X} \to \mathbb{R}$ be measurable for each $i = 1, \ldots, n$ and define $\bar{g} : \mathbb{X} \to \mathbb{R}^n$ by $g = (g_1, \ldots, n_n)$. Assume $\alpha \circ g^{-1}$ is not supported on a hyperplane of dimension $n - 1$, and that the following condition is satisfied*

$$\int_{\mathbb{X}} \log(1 + ||g(x)||)\alpha(dx) < +\infty$$

where $|| \cdot ||$ denotes the Euclidean distance in $\mathbb{R}^n$. Define the $\mathbb{R}^n$-valued Markov chain $\{\bar{G}_m, m \geq 0\}$ as in (3.6.2). Then $\{\bar{G}_m, m \geq 0\}$ is positive Harris recurrent with stationary distribution the law of $\bar{G} := \int_{\mathbb{X}} \bar{g}(x) P(\cdot, dx)$.

We conjecture that for any $n \in \mathbb{N}$, the results in Theorem 3.6.1 follows also for the

$\mathbb{R}^n$-valued Markov $\{\bar{G}_m^{(n\bullet)}, m \geq 0\}$ defined via the recursion

$$\bar{G}_m^{(n)} = \theta_m \sum_{j=1}^{n} Y_{m,j} + (1-\theta_m)\bar{G}_{m-1}^{(n)} \qquad m \geq 1 \tag{3.6.2}$$

where $\bar{G}_m^{(n)} := \int_{\mathbb{X}} \bar{g}(x) P_m^{(n)}(\cdot, dx)$ for $m \geq 0$ and $\{P_m^{(n)}, m \geq\}$ is the Markov chain defined via recursion (3.2.1). As observed by Erhardsson [38], since the law of $\alpha_0 \circ g^{-1}$ does not in general have a density with respect to the Lebesgue measure on $\mathbb{R}^n$, then it is not possible to use directly Proposition 7.1.5 in Meyn and Tweedie [141] in order to to prove that the Markov chain is ϕ-irriducible. In particular, the crucial point is to prove that the Markov chain $\{\bar{G}_m^{(n\bullet)}, m \geq 0\}$ is a T-chain, then from Proposition 6.2.1 in Meyn and Tweedie [141] it follows that $\{\bar{G}_m^{(n\bullet)}, m \geq 0\}$ is ϕ-irriducible and from Theorem 6.2.5 in Meyn and Tweedie [141] if follows that all compact subsets of $\mathbb{R}^n$ are small. The rest of the proof follows the same argument used in Theorem 3.2.3.

4

On a Fleming-Viot process and Bayesian nonparametrics

Fleming-Viot processes are a wide class of measure-valued diffusion processes which often arise as large population limits of so-called particle processes. Here we invert the procedure and show that a countable population process can be derived directly from the neutral diffusion model, with no arbitrary assumptions. We study the atomic structure of the neutral diffusion model, and elicit a finite dimensional particle process from the time-dependent random measure, for any chosen population size. The static properties are consequences of the fact that its stationary distribution is the Dirichlet process, and rely on a new representation for it. The dynamics are derived directly from the transition function of the neutral diffusion model. As by-product we also obtain a new constructive definition of the Dirichlet process.

4.1 Introduction

Bayesian nonparametric statistics and population genetics have a common interest in providing suitable countable representations for the law of random probability measures (r.p.m.s). The most studied class of r.p.m.s in Bayesian nonparametrics is the Dirichlet process, whose characterization and properties were presented by Ferguson [61] and Ferguson [62] and further investigated by Blackwell [9] and by Blackwell and MacQueen [10]. Let $(\mathbb{X}, \mathcal{T})$ be the usual Polish space endowed with the Borel σ-field $\mathscr{X}$ and consider the following associated spaces of measures $\mathcal{A}_{\mathbb{X}}$ and $\mathcal{P}_{\mathbb{X}}$. In particular, $\mathcal{A}_{\mathbb{X}}$ is the space of locally finite non-negative measures on $(\mathbb{X}, \mathscr{X})$ endowed with the σ-field $\mathscr{A}_{\mathbb{X}}$ generated by the vague topology $\mathcal{V}$ which makes $(\mathcal{A}_{\mathbb{X}}, \mathcal{V})$ a Polish space, and $\mathcal{P}_{\mathbb{X}}$ is the space of probability measures on $(\mathbb{X}, \mathscr{X})$ endowed with its σ-field $\mathscr{P}_{\mathbb{X}}$ generated by the weak convergence topology $\mathcal{W}$ which makes $(\mathcal{P}_{\mathbb{X}}, \mathcal{W})$ a Polish space. As shown by Blackwell and MacQueen

[10], the Dirichlet process with parameter ν[1] can be alternatively defined as the r.p.m. P derived as a.s. weak limit of the empirical distribution of a sequence $\{X_n, n \geq 1\}$ generated by the sampling scheme

$$\mathbb{P}(X_{n+1} \in \cdot \,|X_1, \dots, X_n) = \frac{1}{\theta + n}\nu_0(\cdot) + \frac{1}{\theta + n}\sum_{j=1}^{n} \delta_{X_j}(\cdot) \qquad n \geq 1. \tag{4.1.1}$$

where $\nu_0 := \nu/\theta$. The sampling scheme (4.1.1) generalizes the Pólya-urn schem and it is usually called the Blackwell-MacQueen Pólya-urn scheme. In particular, Blackwell and MacQueen [10] showed that a sequence of observations $\{X_n, n \geq 1\}$ drawn according to the sampling scheme (4.1.1) is equivalent in distribution to a sequence $\{X_n, n \geq 1\}$ of independent and identically distributed (i.i.d.) observations from P, where P is a Dirichlet process with parameter ν. A sequence $\{X_n, n \geq 1\}$ generated by the Blackwell-MacQueen Pólya-urn scheme (4.1.1) is exchangeable then, denoting by Q the law of a Dirichlet process P with parameter ν, by the de Finetti representation theorem for any $n \in \mathbb{N}$ and for any collection of sets $A_1, \dots, A_n \in \mathscr{X}$

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \int_{\mathcal{P}_{\mathbb{X}}} \prod_{i=1}^{n} P(\cdot, A_i)\Pi(dP)$$

where, as $n \to +\infty$

$$e_n(X_1, \dots, X_n) \Rightarrow P \qquad \text{a.s.-}\mathbb{P} \tag{4.1.2}$$

and in particular $\{e_n(X_1, \dots, X_n), n \geq 1\}$ is a random sequence of de Finetti measures with coordinates in $\mathscr{P}_{\mathbb{X}}$.

In Population genetics, the Dirichlet process arises as the stationary distribution of a particular measure-valued diffusion process, the so-called neutral diffusion model, which describes the evolution of the allele frequencies of a population of genes under the hypothesis of neutral, non-recurrent, parent independent mutation (see Ethier and Kurtz [42]). The neutral diffusion model has continuous sample paths which are functions from $[0, +\infty)$ to $\mathscr{P}_{\mathbb{X}}$ and it is characterized in terms of the infinitesimal generator

$$(\mathcal{L}\varphi)(\mu) = \sum_{i=1}^{k} \langle A_i f, \mu^k \rangle + \frac{1}{2} \sum_{1 \leq j \neq i \leq k} (\langle \Phi_{j,i}^{(k)} f, \mu^{k-1} \rangle - \langle f, \mu^k \rangle) \tag{4.1.3}$$

where the domain $\mathscr{D}(\mathcal{L})$ is taken to be the set of all bounded functions on $\mathscr{P}_{\mathbb{X}}$ of the form $\varphi(\mu) = \langle f, \mu^k \rangle$, for f a bounded measurable function on $\mathbb{X}^k$, $\langle f, \mu \rangle$ denoting $\int_{\mathbb{X}} f d\mu$ and μ^k being an k-fold product measure. Here A_i is the mutation operator

$$Af(x) := \frac{1}{2}\theta \int (f(z) - f(x))\nu_0(dz) \tag{4.1.4}$$

[1] In order to have a notation consistent with respect to the notation in population genetics literature, in this chapter we use the symbol ν instead of the usual symbol α to denote the parameter of the Dirichlet process and the symbol θ instead of the usual symbol a to denote to total mass of α.

applied to the i-th argument of the function f, where $\theta \in \mathbb{R}^+$ and $\nu_0 \in \mathscr{P}_{\mathbb{X}}$ is the non-atomic probability measure obtained by the normalization ν/θ. Also we consieder the definition $(\Phi_{j,i}^{(k)} f)(x_1, \dots, x_k) := f(x_1, \dots, x_{i-1}, x_j, x_{i+1}, \dots, x_k)$

The transition density of the neutral diffusion model is provided by Ethier and Griffiths [41] in terms of a mixture of Dirichlet processes, showing an interesting connection with the Bayesian framework. This is given by

$$p(t, \mu, d\nu) = \sum_{n \geq 0} d_n(t) \int_{\mathbb{X}^n} \Pi \left(d\nu | \theta\nu_0 + \sum_{i=1}^{n} \delta_{X_i} \right) \mu^n(dx_1, \dots, dx_n) \tag{4.1.5}$$

where μ^n denotes the n-fold product measure and $\Pi(\cdot | \theta\nu_0 + \sum_{1 \leq i \leq n} \delta_{X_i})$ denotes the law of a Dirichlet process, conditional on the observations $X_1, \dots, X_n$ each one sampled from μ. Moreover, in Equation (4.1.5), $d_n(t) := \mathbb{P}(D_t = n)$, where $\{D_t, t \geq 0\}$ is a death process with rate

$$\lambda_k = \frac{1}{2} k(\theta + k - 1) \tag{4.1.6}$$

and such that $D_0 = +\infty$ almost surely. In particular, in Tavarè [179] an explicit expression for $d_n(t)$ is given for $n \geq 1$

$$d_n(t) = \sum_{k \geq n} (-1)^{k-n} \binom{k}{n} (\theta + n)_{(k-1)\uparrow 1} (k!)^{-1} \gamma_{k,t,\theta} \tag{4.1.7}$$

where

$$\gamma_{k,t,\theta} := (\theta + 2k - 1) \mathrm{e}^{-\lambda_k t}$$

and

$$d_0(t) = 1 - \sum_{k \geq 1} (-1)^{k-1} (\theta)_{(k-1)\uparrow 1} (k!)^{-1} \gamma_{k,t,\theta}.$$

where $(x)_{y\uparrow 1}$ stands for the Pochhammer symbol for the ascending factorial of x of order y (see Appendix A). Further connections between the Bayesian nonparametric framework and the neutral diffusion model are recently established. In particular Walker et al. [193] provided a construction of the neutral diffusion model via its transition function using ideas on Gibbs sampler based Markov processes. Ruggiero and Walker [171] propose a construction of the neutral diffusion model with selection based on a generalised Blackwell-MacQueen Pólya urn scheme, obtained from the Bayesian hierarchical mixture model introduced by Lo [128].

In the first part of this chaper we define a particular continuous time version of the generalized Feigin-Tweedi Markov chain defined in Chapter 3 and we prove that as its parameter $n \to +\infty$ it converge in distribution (in the Skorohod topology) to the neutral diffusion model. Thus, the construction provides a countable representation of the neutral diffusion model such that the n-sized population version of the measure-valued process has the same stationary distribution as the limiting process, which does not hold for the process of empirical measures. More specifically, after defining a pure jump-type $\mathbb{X}^n$-valued Markov process based on the Blackwell-MacQueen Pólya-urn scheme, we define a continuous time version of the n-idexed Feigin-Tweedie Markov chain, such that at any time $t \geq 0$ such continuous time process is distributed according to a posterior Dirichlet process conditional on the pure jump-type $\mathbb{X}^n$-valued Markov process at time t. The proposed continuous continuous time version of the generalized Feigin-Tweedie does converge, as its parameter $n \to +\infty$, to the neutral diffusion model, and in addition its invariant measure is the law of a Dirichlet process for every size of the population. This offers a slightly different perspective on the theoretical side of the neutral diffusion model, by making more explicit the role of the measure-valued random element which drives the exchangeable sequence of individuals with respect to the relative frequencies. Furthermore, this approach can be useful when handling a finite population approximation of the neutral diffusion model for simulation and inference purposes. The classical approach of taking the process of empirical measures of the particle types is recovered in a special case.

The Fleming-Viot processes, introduced by Fleming and Viot [70] represent a general class of measure-valued diffusion processes (or superprocesses or diffusion processes in the Kingman symplex) which includes the neutral diffusion model as particular case. In particular, the Fleming-Viot processes are generally viewed as limit approximations of the behavior of finite populations of say n alleles, as n goes to infinity. The model of reproduction of the n-alleles population is often represented by a n-dimensional particle process $\{(X_{t,1}, \dots, X_{t,n}), t \geq 0\}$ with sample paths on the space $\mathcal{D}_{\mathbb{X}^n}^{[0,+\infty)}$ of càdlàg functions[2] from $[0, +\infty)$ to $\mathbb{X}^n$. In this case $\{(X_{t,1}, \dots, X_{t,n}), t \geq 0\}$ is a countable representation of the Fleming-Viot process $\{\mu_t, t \geq 0\}$ in the sense that the process of allele frequencies $\{e_{t,n}(X_{t,1}, \dots, X_{t,n}), t \geq 0\}$, where at any $t \geq 0$

$$e_{t,n}(X_{t,1}, \dots, X_{t,n}) := \frac{1}{n} \sum_{j=1}^{n} \delta_{X_{t,j}}$$

converges in distribution, in the Skorohod topology to $\{\mu_t, t \geq 0\}$ as n grows to infinity. A general theory for a countable representation of Fleming-Viot processes is provided by Donelly and Kurtz [29] and by Donelly and Kurtz [30].

In the second part of this chapter we consider the opposite problem. Given a measure-valued diffusion process, and in particular given the neutral diffusion model, we investigate how a particle process should be in order to be a suitable representation for a finite-population extract from the limiting diffusion $\{e_{t,\infty}, t \geq 0\}$. The main point is of

[2]A càdlàg function is a function that is right continuous and has a left limit. The acronym càdlàg comes from the French "continue droite, limite gauche".

course what suitable means. A reasonable criterion seems be that the defining properties of the particle process be derived only from the intrinsic features of the neutral diffusion model itself, with no further arbitrary assumptions. In our case, the static properties of the particle process will be consequences of the fact that the stationary distribution of the neutral diffusion model is the Dirichlet process. The dynamic properties will be derived directly from the transition function (4.1.5) and its implications.

Here the focus is on the almost sure discreteness of the Dirichlet process. This suggests that instead of adopting the usual approach by proposing a population process and show that this converges in distribution to the measure-valued diffusion, we can invert the procedure and derive the population process directly from the diffusion. That is, we investigate the properties of the atoms which give the time-dependent random measure, and show that for any chosen population size $n \in \mathbb{N}$ we can elicit n atoms from the random measure; then their properties automatically define a particle process, each atom being a particle, with sample paths in $\mathcal{D}_{\mathbb{X}^n}^{[0,+\infty)}$. When n grows to infinity the infinite population process, summarized by its empirical distribution, is equivalent to the neutral diffusion model. We can thus talk of a population process underlying the neutral diffusion model, in the sense that all properties of the former are derived by the latter. Such constructive approach brings new evidence, once again, of the key role played by Blackwell-MacQueen urn schemes in explaining the fundamental structure of Ferguson-Dirichlet populations.

The chapter is organized as follows. Section 4.2 provides a general introduction on the Fleming-Viot processes in population genetics. In particular we review some detailed steps for the construction of the neutral diffusion models and its countable representation. Section 4.3 describes a continuos time version of the n indexed Feigin-Tweedie Markov chain introduced in Chapter 3. Section 4.4 states the main result in the finite case, which is the most relevant here, that is for an arbitrary population size, which determines the size of the finite dimensional particle process. In particular, the main result is derived by a new constructive definition of the Dirichlet process which is used to elicit the particles from the random measure. Section 4.5 provides some discussion and deals with the infinite population case.

4.2 Fleming-Viot processes in population genetics

This section review the detailed steps for the construction of the neutral diffusion models as special case of the Fleming-Viot process introduced by Fleming and Viot [70]. In particular, we recall the countable representation of the neutral diffusion models originally proposed by Donelly and Kurtz [29] and further developed by Donelly and Kurtz [30].

The neutral diffusion model in population genetics, in which each individual is of some type and the set $\mathbb{X}$ of the types is finite, has state space the finite dimensional symplex $\Delta^{(\mathbb{X})} := \left\{\{p_i, i \in \mathbb{X}\} \in (0,1]^{\mathbb{X}} : \sum_{i\in\mathbb{X}} p_i = 1\right\}$ where p_i denotes the proportion of the population that is of type i. Its infinitesimal generator is

$$L = \frac{1}{2} \sum_{i,j\in\mathbb{X}} p_i(\delta_{i,j} - p_j)\frac{d^2}{dp_i dp_j} + \sum_{j\in\mathbb{X}} \left(\sum_{i\in\mathbb{X}} q_{i,j} p_i\right) \frac{d}{dp_j} \tag{4.2.1}$$

where $\delta_{i,j}$ is the Kronecker delta and $\{q_{i,j}, i, j \in \mathbb{X}\}$ is the infinitesimal matric for a Markov process in $\mathbb{X}$; for $i \neq j$, $q_{i,j}$ is interpreted as the intensity of a mutation from type i to type j. In particular the domain of the infinitesimal generator L is $\mathscr{D}(L) = \{F_{\Delta^{(\mathbb{X})}} : F \in C^2(\mathbb{R}^m)\}$ where $C^2(\mathbb{R}^m)$ denotes the set of all twice-differentiable continuous function with value in $\mathbb{R}^m$.

Except for some technical requirements on $\{q_{i,j}, i, j \in \mathbb{X}\}$, the same description is valid when $\mathbb{X}$ is countably infinite. One such example is the stepwise-mutation model proposed by Otha and Kimura [147], in which $\mathbb{X} = \mathbb{Z}$ and

$$q_{i,j} = \begin{cases} \theta/2 & \text{if } j = i \pm 1 \\ -\theta & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

ofr some $\theta \in \mathbb{R}^+$.

The case in which $\mathbb{X}$ is uncountably infinite, however, requires a different approach. The key idea, due to Fleming and Viot [70], is to topologize $\mathbb{X}$ by a topology $\mathcal{T}$ such that $(\mathbb{X}, \mathcal{T})$ is a Polish space endowed with the Borel σ-field $\mathscr{X}$ and to replace $\Delta^{(\mathbb{X})}$ by $\mathcal{P}_{\mathbb{X}}$, the set of all probability measures on $(\mathbb{X}, \mathscr{X})$ endowed with its σ-field $\mathscr{P}_{\mathbb{X}}$ generated by the weak convergence topology $\mathcal{W}$ which makes $(\mathcal{P}_{\mathbb{X}}, \mathcal{W})$ a Polish space. Then, (4.2.1) becomes

$$(\mathcal{L}\varphi)(\mu) = \frac{1}{2} \int_{\mathbb{X}} \int_{\mathbb{X}} \mu(dx)(\delta_x(dy) - \mu(dy)) \frac{d^2\varphi(\mu)}{d\mu(x)d\mu(y)} + \int_{\mathbb{X}} \mu(dx) A \left(\frac{d\varphi(\mu)}{d\mu(\cdot)}\right)(x) \tag{4.2.2}$$

where

$$\frac{d\varphi(\mu)}{d\mu(x)} = \lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon}(\varphi(\mu + \varepsilon\delta_x) - \varphi(\mu))$$

and A is the generator of a Feller semigroup on $C(\mathbb{X})$ where $C(\mathbb{X})$ denotes the set of all differentiable continuous function with value in $\mathbb{X}$. In particular the domain of the infinitesimal generator $\mathcal{L}$ is

$$\mathscr{D}(\mathcal{L}) = \{\varphi : \varphi(\mu) = F(\langle f_1, \mu\rangle, \ldots, \langle f_m, \mu\rangle), F \in C^2(\mathbb{R}^m), f_1, \ldots, f_m \in \mathscr{D}(A)\}$$

for any $m \in \mathbb{N}$, with $\langle f, \mu\rangle := \int_{\mathbb{X}} f d\mu$ and the domain $\mathscr{D}(A)$ of the Feller semigroup A on $C(\mathbb{X})$ corresponds to a subspace of $C(\mathbb{X})$. The resulting measure-valued diffusion process is reffered to as a Fleming-Viot process. The space $\mathbb{X}$ is called the type space and A is known as the mutation operator. Terms corresponding to recombination and selection can also be included in (4.2.2).

The Fleming-Viot process arises most naturally as the limit in distribution of certain sequences of Markov chains occurring in population genetics known as Wright-Fisher models. The Wright-Fisher model is mathematically simpler than the Fleming-Viot process, but less reasonable biologically. In a diploid population $x_1, \dots, x_n$ of size n there are m gametes and we consider the empirical distribution of the gametic types. It is unnecessary to require that m be even, so let $m \in \mathbb{N}$, and define the application $e_m : \mathbb{X}^m \to \mathcal{P}_{\mathbb{X}}$ by $e_m(x_1, \dots, x_m) := 1/m \sum_{1 \le i \le m} \delta_{x_i}$. The state space for this model is $e_m(\mathbb{X}^m) \subset \mathcal{P}_{\mathbb{X}}$. Given $\mu \in \mathcal{P}_{\mathbb{X}}$, we define $\mu_m^* \in \mathcal{P}_{\mathbb{X}^2}$ and $\mu_m^{**}, \mu_m^{***} \in \mathcal{P}_{\mathbb{X}}$ by

$$\mu_m^*(dx \times dy) := \frac{w_m(x,y)\mu^2(dx \times dy)}{\langle w_m, \mu^2 \rangle} \tag{4.2.3}$$

$$\mu_m^{**}(dx') := \int_{\mathbb{X}^2} R_m((x,y), dx' \times \mathbb{X})\mu_m^*(dx \times dy) \tag{4.2.4}$$

and

$$\mu_m^{***}(dx') := \int_{\mathbb{X}} Q_m(x, dx')\mu_m^{**}(dx) \tag{4.2.5}$$

where for any $m \in \mathbb{N}$ let w_m be a postive, symmetric, bounded, Borel function on $\mathbb{X}^2$, $R_m((x,y), dx' \times dy')$ be a one-step transition function on $\mathbb{X} \times \mathscr{X}^2$ satisfying $R_m((x,y), dx' \times dy') = R_m((y,x), dy' \times dx')$ and let $Q_m(x, dx')$ be a one-step transition function on $\mathbb{X} \times \mathscr{X}$. The functions w_m, R_m and Q_m involve selection, recombination and mutation, respectively. The Wright-Fisher model we are considering is a Markov chain having one-step transition function $p_m(\mu, d\nu)$ on $e_m(\mathbb{X}^m) \times \mathscr{B}(e_m(\mathbb{X}^m))$ given by

$$p_m(\mu, d\nu) = \int_{\mathbb{X}} (\mu^{***})^m (dx_1 \times \cdots \times dx_m)\delta_{e_m}(d\nu) \tag{4.2.6}$$

where $\mathscr{B}(e_m(\mathbb{X}^m))$ is the Borel σ-field generated by $e_m(\mathbb{X}^m)$. The present formulation of the model is from Ethier and Kurtz [42].

For any $m \in \mathbb{N}$, let $\{\nu_\tau^{(m)}, \tau \in \mathbb{Z}^+\}$ be a Wright-Fisher model. Under suitable condition on w_m, R_m and Q_m and assuming weak convergence of initial distributions, it can be shown that as $n \to +\infty$

$$\{\nu_{\lfloor mt \rfloor}^{(m)}, t \ge 0\} \Rightarrow \{\mu_t, t \ge 0\} \tag{4.2.7}$$

where $\lfloor mt \rfloor$ denotes the integer part on mt and where $\{\mu_t, t \ge 0\}$ is diffusion process in $\mathcal{P}_{\mathbb{X}}$. Let $\mathcal{L}_m$ be the discrete generator of the n-th rescaled Markov chain $\{\nu_{\lfloor mt \rfloor}^{(m)}, t \ge 0\}$

$$(\mathcal{L}_m \varphi)(\mu) = m \int_{e_m(\mathbb{X}^m)} (\varphi(\nu) - \varphi(\mu)) p_m(\mu, d\nu) \tag{4.2.8}$$

where $p_m(\mu, d\nu)$ is given by (4.2.6) and we regard (4.2.8) as being defined on all of $\mathcal{P}_{\mathbb{X}}$, not just on $e_m(\mathbb{X}^n)$. We restrict the attention to test functions φ of the form

$$\varphi(\mu) = \langle f_1, \mu\rangle \cdots \langle f_k, \mu\rangle \tag{4.2.9}$$

where $m \in \mathbb{N}$ and $f_1, \dots, f_k \in B(\mathbb{X})$ with $B(\mathbb{X})$ the set of all bounded function with value in $\mathbb{X}$. By (4.2.6)

$$\begin{aligned}(\mathcal{L}_m\varphi)(\mu) &= m\int_{e_m(\mathbb{X}^m)}\left(\prod_{i=1}^{k}\langle f_i, \nu\rangle - \prod_{i=1}^{k}\langle f_i, \mu\rangle\right)p_m(\mu, d\nu) \qquad (4.2.10)\\ &= m\left(\int_{\mathbb{X}^m}\prod_{i=1}^{k}\langle f_i, e_m\rangle(\mu_m^{***})^m(dx_1\times\cdots\times dx_m) - \prod_{i=1}^{k}\langle f_i, \mu\rangle\right)\\ &= m\left(m^{-k}\int_{\mathbb{X}^m}\sum_{j_1=1}^{m}\cdots\sum_{j_k=1}^{m}\prod_{i=1}^{k}f_i(x_{j_i})(\mu_m^{***})^m(dx_1\times\cdots\times dx_m) - \prod_{i=1}^{k}\langle f_i, \mu\rangle\right)\\ &= \sum_{1\le i<j\le k}(\langle f_if_j, \mu_m^{***}\rangle - \langle f_i, \mu_m^{***}\rangle\langle f_j, \mu_m^{***}\rangle)\prod_{l\neq j,i}\langle f_l, \mu_m^{***}\rangle + \sum_{i=1}^{k}m(\langle f_i, \mu_m^{***}\rangle\\ &\qquad - \langle f_i, \mu\rangle)\prod_{l<i}\langle f_l, \mu_m^{***}\rangle + O(m^{-1})\end{aligned}$$

uniformly in $\mu \in \mathcal{P}_{\mathbb{X}}$. To ensure tha this converges, wa assume the existence of $\sigma \in B_{\text{sym}}(\mathbb{X}^2)$ (the selection intensity function) where $B_{\text{sym}}(\mathbb{X}^2)$ is the set of all bounded symmetric function with value in $\mathbb{X}^2$, a bounded linear transformaion B from $B(\mathbb{X})$ to $B(\mathbb{X}^2)$ (the recombination operator) of the form

$$(Bf)(x, y) = \alpha\int_{\mathbb{X}}(f(x') - f(x))R((x, y), dx') \tag{4.2.11}$$

where $\alpha \in \mathbb{R}$ (the recombination intensity) and $R((x, y), dx')$ is a one-step transition function on $\mathbb{X}^2 \times \mathscr{X}$, and a possibly unbounded operator A on $B(\mathbb{X})$ (the mutation operator), defined only on a subspace $\mathscr{D}(A)$ such that

$$w_m(x, y) = 1 + m^{-1}\sigma(x, y) + o(m^{-1})$$

$$\int_{\mathbb{X}} f(x')R_m((x, y), dx' \times \mathbb{X}) = f(x) + m^{-1}(Bf)(x, y) + o(m^{-1})$$

and

$$\int_{\mathbb{X}} f(x')Q_m(x, dx') = f(x) + m^{-1}(Af)(x) + o(m^{-1})$$

for all $f \in \mathscr{D}(A)$, respectively, uniformly in $x, y \in \mathbb{X}$. This implies that

$$\langle f, \mu_m^{***}\rangle = \langle f, \mu\rangle + m^{-1}(\langle Af, \mu\rangle + \langle Bf, \mu^2\rangle + \langle (f\circ\pi)\sigma, \mu^2\rangle - \langle f, \mu\rangle\langle\sigma, \mu^2\rangle) + o(m^{-1}) \tag{4.2.12}$$

for all $f \in \mathscr{D}(A)$ uniformly in $\mu \in \mathcal{P}_{\mathbb{X}}$, where π is the projection operator of $\mathbb{X}^2$ onto its first coordinate. Thus, if for $\varphi \in B(\mathcal{P}_{\mathbb{X}})$ of the form (4.2.9) with $k \in \mathbb{N}$ and $f_1, \ldots, f_k \in \mathscr{D}(A)$ we define

$$(\mathcal{L}\varphi)(\mu) = \sum_{1 \le i < j \le k} (\langle f_i f_j, \mu\rangle - \langle f_i, \mu\rangle\langle f_j, \mu\rangle) \prod_{l \ne i,j} \langle f_l, \mu\rangle \tag{4.2.13}$$
$$+ \sum_{i=1}^{k} (\langle A f_i, \mu\rangle + \langle B f_i, \mu^2\rangle) \prod_{k \ne i} \langle f_l, \mu\rangle + \sum_{i=1}^{k} (\langle (f_i \circ \pi)\sigma, \mu^2\rangle - \langle f_i, \mu\rangle\langle \sigma, \mu^2\rangle) \prod_{l \ne i} \langle f_l, \mu\rangle$$

then, assuming that $\mathscr{D}(A)^c$ is an algebra, (4.2.10) and (4.2.12) imply that $(\mathcal{L}_m\varphi)(\mu) = (\mathcal{L}\varphi)(\mu) + o(1)$ uniformly in $\mu \in \mathcal{P}_{\mathbb{X}}$. More generally, we define $\mathcal{L}$ by

$$\begin{aligned}(\mathcal{L}\varphi)(\mu) = \frac{1}{2} \int_{\mathbb{X}} \int_{\mathbb{X}} \mu(dx)(\delta_x(dy) - \mu(dy)) \frac{d^2\varphi(\mu)}{d\mu(x) d\mu(y)} \\ + \int_{\mathbb{X}} \mu(dx) A \left(\frac{d\varphi(\mu)}{d\mu(\cdot)}\right)(x) + \int_{\mathbb{X}} \int_{\mathbb{X}} \mu(dx)\mu(dy) B \left(\frac{d\varphi(\mu)}{d\mu(\cdot)}\right)(x, y) \\ + \int_{\mathbb{X}} \int_{\mathbb{X}} \mu(dx)\mu(dy)(\sigma(x, y) - \sigma\langle \sigma, \mu^2\rangle) \frac{d\varphi(\mu)}{d\mu(x)}\end{aligned} \tag{4.2.14}$$

where $d\varphi(\mu) d\mu(x) = \lim_{\varepsilon \to 0^+} \varepsilon^{-1}(\varphi(\mu + \varepsilon\delta_x) - \varphi(\mu))$, and we take $\mathscr{D}(\mathcal{L})$ to be the set of all $\varphi \in B(\mathcal{P}_{\mathbb{X}})$ of the form

$$\varphi(\mu) = F(\langle f_1, \mu\rangle, \ldots, \langle f_k, \mu\rangle) = F(\langle f, \mu\rangle) \tag{4.2.15}$$

where $k \in \mathbb{N}$, $f_1, \ldots, f_k \in \mathscr{D}(A)$ and $F \in C^2(\mathbb{R}^k)$. For such φ

$$\begin{aligned}(\mathcal{L}\varphi)(\mu) = \frac{1}{2} \sum_{i,j=1}^{k} (\langle f_i f_j, \mu\rangle - \langle f_i, \mu\rangle\langle f_j, \mu\rangle) F_{z_i, z_j}(\langle f, \mu\rangle) \\ + \sum_{i=1}^{k} (\langle A f_i, \mu\rangle + \langle B f_i, \mu^2\rangle) F_{z_i}(\langle f, \mu\rangle) \\ + \sum_{i=1}^{k} (\langle (f_i \circ \pi)\sigma, \mu^2\rangle - \langle f_i, \mu\rangle\langle \sigma, \mu^2\rangle) F_{z_i}(\langle f, \mu\rangle).\end{aligned} \tag{4.2.16}$$

The formulation (4.2.16) is from Fleming and Viot [70], whereas (4.2.14) is due to Dawson and Hochberg [23].

Another choice for the domain of $\mathcal{L}$ that is often useful is the set of all $\varphi \in B(\mathcal{P}_{\mathbb{X}})$ of the form $\varphi(\mu) = \langle f, \mu^m\rangle$, where $k \in \mathbb{N}$ and $f \in B(\mathbb{X}^k)$ satisfies certain conditions. To describe these conditions precisely, we need to be more specific about the assumptions on A. We assume that $\mathbb{X}$ is locally compact and that the closure of A generates a Feller semigroup $\{T_t, t \ge 0\}$ on $\hat{C}(\mathbb{X})$, the space of real continuous functions on $\mathbb{X}$ vanishing at infinity (if $\mathbb{X}$ is compact, then $\hat{C}(\mathbb{X}) = C(\mathbb{X})$). Note that $\{T_t, t \geq 0\}$ is given by a transition function $p(t, x, d\xi)$, that is,

$$T_t f(x) = \int_{\mathbb{X}} f(\xi) p(t, x, d\xi).$$

For each $k \in \mathbb{N}$, we define the semigroup $\{T_{k,t}, t \geq 0\}$ on $B(\mathbb{X}^k)$ by

$$T_{k,t} f(x_1, \ldots, x_k) = \int_{\mathbb{X}^k} f(\xi_1, \ldots, \xi_k) p(t, x_1, d\xi_1) \cdots p(t, x_k, d\xi_k)$$

and let $A^{(k)}$ denote its generator. In addition, for each $k \geq 2$ and $1 \leq i \leq j \leq k$, we define $\Phi_{i,j}^{(k)} : B(\mathbb{X}^k) \to B(\mathbb{X}^{k-1})$ by letting $\Phi_{i,j}^{(k)} f$ be the function obtained from f by replacing x_j by x_i and renumbering the variables

$$(\Phi_{i,j}^{(k)} f)(x_1, \ldots, x_{k-1}) = f(x_1, \ldots, x_{j-1}, x_i, x_j, \ldots, x_{k-1}).$$

For each $k \in \mathbb{N}$ and $1 \leq i \leq k$, we define $H_i^{(k)} : B(\mathbb{X}^k) \to B(\mathbb{X}^{k+1})$ by

$$(H_i^{(k)} f)(x_1, \ldots, x_{k+1}) = \int_{\mathbb{X}} f(x_1, \ldots, x_{i-1}, \xi, x_{i+1}, \ldots, x_k) R((x_{i, x_{k+1}}), d\xi)$$

and $K_i^{(k)} : B(\mathbb{X}^k) \to B(\mathbb{X}^{k+2})$ by

$$(K_i^{(k)} f)(x_1, \ldots, x_{k+2}) = \frac{\bar{\sigma} + \sigma(x_i, x_{k+1}) - \sigma(x_{k+1}, x_{k+2})}{2\bar{\sigma}} f(x_1, \ldots, x_k)$$

where $\bar{\sigma} = \sup_{x,y,z \in \mathbb{X}} |\sigma(x, y) - \sigma(y, z)|$ and $0/0 = 0$. For each $k \in \mathbb{N}$ and $f \in \mathscr{D}(A^{(k)})$, we define $\varphi_f \in B(\mathcal{P}_{\mathbb{X}})$ by

$$\varphi_f(\mu) = \langle f, \mu^k \rangle \tag{4.2.17}$$

and we note that (4.2.14) reduces to

$$(\mathcal{L}\varphi)(\mu) = \sum_{1 \leq i < j \leq k} (\langle \Phi_{i,j}^{(k)} f, \mu^{k-1} \rangle - \langle f, \mu^k \rangle) + \langle A^{(k)} f, \mu^k \rangle + \alpha \sum_{i=1}^{k} (\langle H_i^{(k)} f, \mu^{k+1} \rangle - \langle f, \mu^k \rangle)$$
$$+ 2\bar{\sigma}(\langle K_i^{(k)} f, \mu^{k+2} \rangle - \langle f, \mu^k \rangle) + \bar{\sigma} k \langle f, \mu^k \rangle. \tag{4.2.18}$$

Note that since $\langle \Phi_{j,i}^{(k)} f, \mu^{k-1} \rangle = \langle \Phi_{i,j}^{(k)} f, \mu^{k-1} \rangle$ implies

$$\frac{1}{2} \sum_{1 \leq i \neq j \leq k} (\langle \Phi_{i,j}^{(k)} f, \mu^{k-1} \rangle - \langle f, \mu^k \rangle) = \sum_{1 \leq i < j \leq k} (\langle \Phi_{i,j}^{(k)} f, \mu^{k-1} \rangle - \langle f, \mu^k \rangle).$$

Then (see Ethier and Kurtz [42]), the generator (4.2.18) can be written as

$$(\mathcal{L}\varphi)(\mu) = \sum_{1\le i\neq j\le k} (\langle \Phi^{(k)}_{i,j} f, \mu^{k-1}\rangle - \langle f, \mu^k\rangle) + \sum_{i=1}^{k}\langle A_i f, \mu^k\rangle + \alpha \sum_{i=1}^{k}(\langle H^{(k)}_i f, \mu^{k+1}\rangle - \langle f, \mu^k\rangle) \tag{4.2.19}$$

$$+ \sum_{i=1}^{k}(\langle \sigma(x_i, x_{k+1}) f, \mu^{k+1}\rangle - \langle \sigma(x_{k+1}, x_{k+2}) f, \mu^{k+2}\rangle).$$

Moreover, if we do not have recombination, i.e. $\alpha = 0$ and the selection function is haploid instead of diploid, i.e.

$$\sigma(x,\mu) = \int_{\mathbb{X}} \sigma(x,y)\mu(dy)$$

then (see Donelly and Kurtz [30])

$$(\mathcal{L}\varphi)(\mu) = \sum_{i=1}^{k}\langle A_i f, \mu^k\rangle + \frac{1}{2}\sum_{1\le i\neq j\le k} (\langle \Phi^{(k)}_{i,j} f, \mu^{k-1}\rangle - \langle f, \mu^k\rangle) \tag{4.2.20}$$

$$+ \sum_{i=1}^{k}(\langle \sigma(x_i, \mu) f, \mu^{k}\rangle - \langle \sigma(x_{k+1}) f, \mu^{k+1}\rangle).$$

From (4.2.20), the neutral diffusion model (see Ethier and Kurtz [41] and Donelly and Kurtz [29]) is recoverd when $\sigma = 0$ and

$$Af(x) = \frac{1}{2}\theta \int_{\mathbb{X}} (f(y) - f(x))p(x,dy).$$

The next result is essentially from Ethier and Kurtz [40]

Theorem 4.2.1. *(cfr. Ethier and Kurtz [40]) Let $\mathbb{X}$ be locally compact and suppos that the closure of A generates a Feller semigroup on $\hat{C}(\mathbb{X})$. Define B in terms of $\alpha \in \mathbb{R}$ and a transition function $R((x,y),dx')$ on $\mathbb{X}^2 \times \mathscr{X}$ by* (4.2.11)*, and let $\sigma \in B_{sym}(\mathbb{X})^2$. Then the martingale problems for $\mathcal{L}$ defined by* (4.2.9) *and* (4.2.13)*, by* (4.2.15) *and* (4.2.16) *and by* (4.2.17) *and* (4.2.18) *are equivalent.*

Moving from the above constructions and following Donelly and Kurtz [29] and Donelly and Kurtz [30] we now recall a further important result which is the countable representation of the Fleming-Viot process. In particular, Donlly and Kurtz [29] introduced an $\mathbb{X}^n$-valued process $\{(X_{t,1},\dots,X_{t,n}), t \ge 0\}$ which represents the evolution in time of a population of size n, such that the empirical measure $e_{t,n}$ in the infinite population limit is a Fleming-Viot process. A process of this type, usually called particle process, appeares implicitly in Dawson and Hochberg [23].

Let $\mathbb{X}$ be a locally compact space. With reference to (4.2.18), observe that in the case of no recombination or selection

$$(\mathcal{L}\varphi)(\mu) = \sum_{1\leq i<j\leq n} (\langle \Phi_{i,j}^{(n)} f, \mu^{n-1}\rangle - \langle f, \mu^n\rangle) + \langle A^{(n)} f, \mu^n\rangle = \langle Cf, \mu^n\rangle \tag{4.2.21}$$

for all $f \in \mathscr{D}(A^{(n)})$, where

$$(Cf)(x_1, \ldots, x_n) = \sum_{1\leq i<j\leq n} (f(\theta_{i,j}(x_1, \ldots, x_n)) - f(x_1, \ldots, x_n)) + (A^{(n)} f)(x_1, \ldots, x_n)$$

$\theta_{i,j}(x_1, \ldots, x_n)$ being the element of $\mathbb{X}^n$ obtained from $(x_1, \ldots, x_n)$ by replacing the j-th component by the i-th. We interpret C as an operator with domain in $B(\mathbb{X}^\infty)$. It is clear that C is the generator for an $\mathbb{X}^\infty$-valued process $\{X_n, n \geq 1\}$ whose j-th component (the particle at level j) evolves according to the mutation process until it "looks down" to level i for some $i < j$ and changes its value to the value on level i.

Let $\{S_t, t \geq 0\}$ denote the Feller semigroup defined on $C(\mathbb{X}^\infty)$ corresponding to C. Then, viewing $C(\mathbb{X}^n)$ as a closed subspace of $C(\mathbb{X}^\infty)$, the following identity holds

$$\mathbb{E}_\mu[\langle f, \mu_t^n\rangle] = \langle S_t f, \mu^n\rangle = \mathbb{E}_{\mu^\infty}^{\{X_n, n\geq 1\}}[f(X_{t,1}, \ldots, X_{t,n})] \tag{4.2.22}$$

for all $f \in C(\mathbb{X}^n)$ and $t \geq 0$. Here $\mathbb{E}_\mu$ denotes the expectation for the Fleming-Viot process under the assumption that the initial state is μ, and $\mathbb{E}_{\mu^\infty}^{\{X_n, n\geq 1\}}$ denotes the expectation for the particle system under the assumption that $\{X_{0,n}, n \geq 1\}$ are i.i.d. with common distribution μ. Let $\nu \in \mathcal{P}_\mathbb{X}$. By (4.2.22)

$$\int_{\mathcal{P}_\mathbb{X}} \mathbb{E}_\mu[\langle f, \mu_t^n\rangle]\nu(d\mu) = \int_{\mathcal{P}_\mathbb{X}} \mathbb{E}_{\mu^\infty}^{\{X_n, n\geq 1\}}[f(X_{t,1}, \ldots, X_{t,n})]\nu(d\mu) \tag{4.2.23}$$

for all $f \in B(\mathbb{X}^n)$ and $t \geq 0$. The left-hand side of (4.2.23) is the expectation for a Fleming-Viot process with initial distribution ν, and the right-hand side of (4.2.23) is the expectation for the particle system under the assumption that $\{X_{0,n}, n \geq 1\}$ is an exchangeable sequence with

$$\mathbb{P}(X_{0,1} \in A_1, \ldots, X_{0,n} \in A_n) = \int_{\mathcal{P}_\mathbb{X}} \prod_{i=1}^n \mu(A_i)\nu(d\mu). \tag{4.2.24}$$

In particular, identity (4.2.23) implies that

$$\mathbb{P}(X_{t,1} \in A_1, \ldots, X_{t,n} \in A_n) = \int_{\mathcal{P}_\mathbb{X}} \prod_{i=1}^n \mu(A_i)\nu_t(d\mu) \tag{4.2.25}$$

for all $t \geq 0$, where ν_t is the distribution at time t of the Fleming-Viot process with initial distribution ν. Therefore, if $\{X_{0,n}, n \geq 1\}$ is an exchangeable sequence, then $\{X_{t,n}, n \geq 1\}$

is an exchangeable sequence, and the corresponding de Finetti measure

$$\hat{\mu}_t := \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{X_{t,i}} \tag{4.2.26}$$

has the same distribution as μ_t. In fact, the next theorem states that the proces $\{\hat{\mu}_t, t \geq 0\}$ is a version of the Fleming-Viot process.

Theorem 4.2.2. *(cfr. Donelly and Kurtz [29]) Let $\mathbb{X}$ be a compact space and suppose that the closure of A generates a Feller semigroup on $C(\mathbb{X})$. Let $\{X_n, n \geq 1\}$ be a Markov process in $\mathbb{X}^\infty$ with generator C and suppose that $\{X_{0,n}, n \geq 1\}$ is exchangeable. Then, for each $t > 0$, $\{X_{t,n}, n \geq 1\}$ is exchangeable and the process given by the de Finetti measures* (4.2.26) *is a Fleming-Viot process with type space $\mathbb{X}$ and mutation operator A.*

4.3 The generalized Feigin-Tweedie Markov chain and its diffusion limit

We consider a pure jump-type $\mathbb{X}^n$-valued Markov process defined by $\{(Y_{1,t}, \dots, Y_{n,t}), t \geq 0\}$ with sample paths in space $\mathcal{D}_{\mathbb{X}^n}^{[0,\infty)}$ as follows. In particular, for any $t \geq 0$, let the r.v. $(Y_{1,t}, \dots, Y_{n,t}) \in \mathbb{X}^n$ be n exchangeable particles, which represents a population of size n. For any $t \geq 0$, at every transition one particle is selected with uniform probability and removed. Conditionally on the other $n-1$ particles, whose value is set equal to their previous one until the next transition, the removed particle is replaced with a sample of size one from the predictive density associated with the Blackwell-MacQueen Pólya urn scheme. In particular, for any $t \geq 0$, given $X_{i,t}$ is removed, say, the incoming particle is sampled from

$$\mathbb{P}(Y_{i,t} \in dy_i|, Y_{1,t}, \dots, Y_{i-1,t}, Y_{i+1,t}, \dots, Y_{n,t}) = \frac{\theta\, \nu_0(dy_i) + \sum_{k\neq i} \delta_{Y_{k,t}}(dy_i)}{\theta + n - 1} \tag{4.3.1}$$

The sojourn times between renewals are driven by a Poisson process with intensity $\lambda_n = n(\theta+n-1)/2$. Let $Y := \{(Y_{1,t}, \dots, Y_{n,t}), t \geq 0\}$, then it can be checked that the infinitesimal generator associated to the proces $\{(Y_{1,t}, \dots, Y_{n,t}), t \geq 0\}$ is given by

$$\begin{aligned}(\mathcal{L}f)(Y) &= \sum_{i=1}^{n} \frac{\lambda_n}{n} \int_{\mathbb{X}} (f(\eta_i(Y|X)) - f(y)) \frac{\theta\nu_0(dx) + \sum_{k\neq i}^{n} \delta_{Y_k}(dx)}{\theta + n - 1} \\ &= \sum_{i=1}^{n} A_i f(Y) + \frac{1}{2} \sum_{1\leq k\neq i\leq n} (f(\eta_i(Y|Y_k)) - f(Y))\end{aligned} \tag{4.3.2}$$

where $\eta_i(Y|X)$ denotes the vector obtained from $(Y_{1,t}, \dots, Y_{n,t})$ by replacing Y_i with X.

At any time $t \geq 0$, the transitions of the process can be seen as the implementation of a Gibbs sampler on the distribution of the r.v. $(Y_{1,t},\dots,Y_{n,t})$, (4.3.1) being the full conditional distribution of $Y_{i,t}$. Hence, the sampler generates a Markov chain which is embedded at jump times (a pure jump-type $\mathbb{X}^n$-valued Markov process) and for any $t \geq 0$, the law of the r.v. $(Y_{1,t},\dots,Y_{n,t})$ is its stationary distribution. Since the Poisson rates depend neither on departure nor on arrival states, the Markov property for the particle process follows immediately. Furthermore, given that between consecutive jumps the value of the vector is set to be constant, the law of $Y_1,\dots,Y_n$ is also the stationary distribution of the pure jump-type $\mathbb{X}^n$-valued Markov process.

From the pure jump-type $\mathbb{X}^n$-valued Markov process above defined $\{(Y_{1,t},\dots,Y_{n,t}), t \geq 0\}$, we consider the follwing theorem which is the main result for the construction for the continuous time version of the generalized Feigin-Tweedie Markov chain described in Chapter 3. In particular, using the definition of the generalized Feigin-Tweedie Markov chain via recursion (3.2.1) we substitute the sample $Y_{m,1},\dots,Y_{m,n}$ from a Blackwell-MacQueen Pólya urn scheme with the pure jump-type $\mathbb{X}^n$-valued Markov process $\{(Y_{1,t},\dots,Y_{n,t}), t \geq 0\}$.

Theorem 4.3.1. *Let P be a Dirichlet process with parameter ν. For any $n \in \mathbb{N}$ let $\{(Y_{1,t},\dots,Y_{n,t}), t \geq 0\}$ be the pure jump-type $\mathbb{X}^n$-valued Markov process above described, $(q_1,\dots,q_n)$ be a r.v. distributed according to a Dirichlet distribution function with parameter $(1,\dots,1)$ independent of $\{(Y_{1,t},\dots,Y_{n,t}), t \geq 0\}$ and let θ_n be a r.v. distributed according to a Beta d.f. with parameter (n,θ) independent of $\{(Y_{1,t},\dots,Y_{n,t}), t \geq 0\}$ and of $(q_1,\dots,q_n)$. Then, for any $t \geq 0$*

$$P_t^{(n)} = \theta_n \sum_{j=1}^{n} q_j \delta_{Y_{j,t}} + (1-\theta_n)P \tag{4.3.3}$$

is a Dirichlet process with parameter ν.

Proof. Since for any $t \geq 0$, $(Y_{1,t},\dots,Y_{n,t})$ is a sample of size n from a Blackwell-MacQueen Pólya urn scheme, then the proof follows similar argument of the proof of Theorem 3.2.1 in Chapter 3. □

From the representation of $P_t^{(n)}$ given by Theorem 4.3.1 we can observe that as $n \to +\infty$

$$\mathbb{E}[\theta_n] = \frac{n}{n+\theta} \to 1$$

and as $n \to +\infty$

$$Var(\theta_n) = \frac{n\,\theta}{(n+\theta)^2(n+1+\theta)} \to 0$$

so that $\theta_n \to 1$ in probability for $n \to \infty$. Then we have the following diffusion limit result.

Theorem 4.3.2. *Let $\{\mu(t), t \geq 0\}$ be the neutral diffusion model and let $\{P_t^{(n)}, t \geq 0\}$ be defined by* (4.3.3)*. Then, as $n \to +\infty$*

$$\{P_t^{(n)}, t \geq 0\} \Rightarrow \{\mu(t), t \geq 0\} \tag{4.3.4}$$

In order to prove Theorem 4.3.2 we need a few intermediate results. The following lemma states that when $\{(Y_{1,t}, \dots, Y_{n,t}), t \geq 0\}$ it the pure jump-type $\mathbb{X}^n$-valued Markov process above described, then for any $t \geq 0$ $P_t^{(n)}|Y_{1,t}, \dots, Y_{n,t}$ is a Dirichlet process with parameter $\nu + \sum_{1 \leq j \leq n} \delta_{Y_{i,t}}$. Loosely speaking, the measure-valued process parallels the particle process in that the changes in the r.v. $(Y_{1,t}, \dots, Y_{n,t})$ affect instantaneously the law of the r.p.m. $P_t^{(n)}$.

Lemma 4.3.1. *Let $\{P_t^{(n)}, t \geq 0\}$ be defined by equation* (4.3.3)*, and let $\{(Y_{1,t}, \dots, Y_{n,t}), t \geq 0\}$ be the pure jump-type $\mathbb{X}^n$-valued Markov process above described. For any $s \geq 0$, $P_{t+s}^{(n)}|Y_{1,t+s}, \dots, Y_{n,t+s}$ is distribuited according to a Dirichlet process with parameter $\nu + \sum_{1 \leq j \leq n} \delta_{Y_{i,t+s}}$ and*

Proof. Suppose the first renewal after time t occurs in $t+\tau$, for some $\tau > 0$. Then the results is trivial for any $0 < s < \tau$. Also, if we show that it holds at $t+\tau$, then it is straightforward to extend it to any $t+u$, for $u > \tau$. Then, without loss of generality, set $s = \tau$. At time $t+\tau$ a particle Y_i, say, is replaced with Z, where Z is either $Z \sim \nu_0$ or $Z \sim \delta_{Y_k}$, for some $k \in \{1, \dots, i-1, i+1, \dots, n\}$. Then we can write $P_{t+\tau}^{(n)}$ as

$$\begin{aligned} P_{t+\tau}^{(n)} &= \theta_n \sum_{j=1}^n q_j \delta_{Y_{j,t+\tau}} + (1-\theta_n) P \\ &= \theta_n \left(\sum_{j \neq i}^n q_j \delta_{Y_{j,t}} + w_{i,n}(t+\tau)\delta_Z \right) + (1-\theta_n) P \end{aligned}$$

given that when Y_i is replaced no other coordinate change. Then it follows that $P_{t+\tau}^{(n)}$ is a Dirichlet process with parameter $\nu + \delta_Z + \sum_{j \neq i} \delta_{Y_{j,t}}$, which is equivalent to a Dirichlet process with parameter $\nu + \sum_{1 \leq j \leq n} \delta_{Y_{j,t+\tau}}$. □

We now derive the generator of the measure-valued process $\{P_t^{(n)}, t \geq 0\}$, and we show that it converges in distribution (in the Skorohod topology) to the neutral diffusion model. Also, we show that the Dirichlet process is the stationary distribution of the finite population version of the process. That is, the n-th version of the process has the same stationary distribution of the limiting diffusion. Consider the vector of random weights $(q_1, \dots, q_n)$ defined in Theorem 4.3.1 and for any $m < n$, let $q_{j_1,\dots,j_m}$ be a weight defined by

$$q_{j_1,\dots,j_m} := q_{j_1} \frac{q_{j_2}}{1-q_{j_1}} \cdots \frac{q_{j_m}}{1-\sum_{l=1}^{m-1} q_{j_l}} \tag{4.3.5}$$

for $1 \leq j_1 \neq \cdots \neq j_m \leq n$. For any $m < n$, the weights defined by (4.3.5) correspond to the probability of picking m elements from an n-dimensional vector without replacement, once each Y_i is assigned a weight q_i. That is, the probability of picking the first is q_{j_1}, then the weights are normalised, so that the probability of picking the second is $q_{j_2}/(1-q_{j_1})$, and so on. If we pick all n elements, the last weight is obviously one. For any $m < n$ we define the probability measure on $\mathbb{X}^m$

$$P_{m,t}^{(n)} = \theta_n \sum_{1\leq j_1\neq\cdots\neq j_m\leq n} q_{j_1,\dots,j_m}\delta_{(Y_{j_1,t},\dots,Y_{j_m,t})} + (1-\theta_n)P^m \tag{4.3.6}$$

and

$$\phi(P_t^{(n)}) := \langle f, P_{m,t}^{(n)}\rangle \tag{4.3.7}$$

Finally we set

$$Z_{m,t}^{(n)} := \sum_{1\leq j_1\neq\cdots\neq j_m\leq n} q_{j_1,\dots,j_m}\delta_{(Y_{j_1,t},\dots,Y_{j_m,t})}$$

Note that if the weights are uniform distributed, then $q_{j_1,\dots,j_m}$ simplifies to $1/(n)_{m\downarrow 1}$. Therfore, equation (4.3.6) can be seen as a generalisation of the probability measure

$$\mu^{(m)} := \frac{1}{(n)_{m\downarrow 1}} \sum_{1\leq i_1\neq\cdots\neq i_m\leq n} \delta_{(Y_{i_1},\dots,Y_{i_m})} \tag{4.3.8}$$

used by Donelly and Kurtz [30], which is recovered with uniform weights and $\theta_n \sim \delta_1$. Under the same conditions, (4.3.7) can be seen as a generalisation of the function

$$(\Gamma^{(m)}f)(Y) := \frac{1}{(n)_{m\downarrow 1}} \sum_{1\leq i_1\neq\ldots\neq i_m\leq n} f(Y_{i_1},\dots,Y_{i_m}) \tag{4.3.9}$$

used in Donelly and Kurtz [29], since $\langle f, P_{m,t}^{(n)}\rangle$ equals to

$$\int_{\mathbb{X}^m} f(Y_1,\dots,Y_m)$$
$$\times \left(\theta_n \sum_{1\leq i_1\neq\cdots\neq i_m\leq n} q_{i_1,\dots,i_m}\delta_{(Y_{i_1,t},\dots,Y_{i_m,t})}(dx_1,\dots,dx_m) + (1-\theta_n)P(dx_1,\dots,dx_m)\right)$$

which, for $\theta_n \sim \delta_1$, reduces to

$$\sum_{1\leq i_1\neq\cdots\neq i_m\leq n} q_{i_1,\dots,i_m} f(Y_{i_1,t},\dots,Y_{i_m,t})$$

and with uniform weights to (4.3.9). Note that including replacement in (4.3.5) yields the function $(\Gamma^{(m)}f)(Y)$, obtained by replacing $(n)_{m\downarrow 1}$ with n^m in (4.3.9) which is asymptotically equivalent to (4.3.9). Let ξ be a r.v. distributed according to ν_0, then the generator of the measure-valued process $\{P_t^{(n)}, t \geq 0\}$ can be written as

$$(\mathcal{L}\phi)(P^{(n)}) = \lim_{\tau \downarrow 0} \frac{1}{\tau} \int_{\mathcal{P}_{\mathbb{X}}} (\phi(P_{t+\tau}^{(n)}) - \phi(P_t^{(n)}))(1 - e^{-\lambda_n \tau}) \tag{4.3.10}$$
$$\times \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\theta}{\theta + n - 1} \Pi(dP_{n,t+\tau}^{(n)} | Y_{1,t}, \ldots, Y_{i-1,t}, Y_{i+1,t}, \ldots, Y_{n,t}, \xi) \right.$$
$$\left. + \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\theta + n - 1} \sum_{k \neq i}^{n} \Pi(dP_{n,t+\tau}^{(n)} | Y_{1,t}, \ldots, Y_{i-1,t}, Y_{i+1,t}, \ldots, Y_{n,t}, Y_{k,t}) \right)$$

Denote with a prime a variable computed in $t + \tau$ when $\tau \downarrow 0$. Since the probability of having two jumps in $[t, t + \tau]$ is $o(\tau)$, we obtain

$$(\mathcal{L}\phi)(P^{(n)}) = \sum_{i=1}^{n} \frac{\theta \, \lambda_n}{n(\theta + n - 1)} \int_{\mathcal{P}_{\mathbb{X}}} [\langle f, P_n^{'(n)} \rangle - \langle f, P_{n,t}^{(n)} \rangle]$$
$$\times \Pi(dP_n^{'(n)} | Y_{1,t}, \ldots, Y_{i-1,t}, Y_{i+1,t}, \ldots, Y_{n,t}, \xi)$$
$$+ \sum_{1 \leq k \neq i \leq n} \frac{\lambda_n}{n(\theta + n - 1)} \int_{\mathcal{P}_{\mathbb{X}}} [\langle f, P_n^{'(n)} \rangle - \langle f, P_{n,t}^{(n)} \rangle]$$
$$\times \Pi(dP_n^{'(n)} | Y_{1,t}, \ldots, Y_{i-1,t}, Y_{i+1,t}, \ldots, Y_{n,t}, Y_{k,t})$$

$$= \sum_{i=1}^{n} \frac{\theta \, \lambda_n}{n(\theta + n - 1)} \int_{\mathbb{X}} \int_{\mathcal{P}_{\mathbb{X}}} [\langle f, P_n^{'(n)} \rangle - \langle f, P_{n,t}^{(n)} \rangle]$$
$$\times \Pi(dP_n^{'(n)} | Y_{1,t}, \ldots, Y_{i-1,t}, Y_{i+1,t}, \ldots, Y_{n,t}, Z) \nu_0(dz)$$
$$+ \sum_{1 \leq k \neq i \leq n} \frac{\lambda_n}{n(\theta + n - 1)} \int_{\mathbb{X}} \int_{\mathcal{P}_{\mathbb{X}}} [f, P_n^{'(n)} \rangle - \langle f, P_{n,t}^{(n)} \rangle]$$
$$\times \Pi(dP_n^{'(n)} | Y_{1,t}, \ldots, Y_{i-1,t}, Y_{i+1,t}, \ldots, Y_{n,t}, Z) \delta_{Y_{k,t}}(dz)$$

Recalling the rate $\lambda_n = 2^{-1} n(\theta + n - 1)$, then we can write the infinitelsimal generator as follows

$$
\begin{aligned}
(\mathcal{L}\phi)(P^{(n)}) &= \sum_{i=1}^{n} \frac{1}{2}\theta \int_{\mathbb{X}} [\langle f, \theta_n Z_n^{'(n)} + (1-\theta_n)P^n\rangle - \langle f, \theta_n Z_n^{(n)} + (1-\theta_n)P_n^n\rangle]\nu_0(dz) \qquad (4.3.11)\\
&+ \frac{1}{2}\sum_{1\le k\ne i\le n} \int_{\mathbb{X}} [\langle f, \theta_n Z_n^{'(n)} + (1-\theta_n)P^n\rangle - \langle f, \theta_n P^{(n)} + (1-\theta_n)P^n\rangle]\delta_{Y_{k,t}}(dz)\\
&= \sum_{i=1}^{n} \left[\frac{1}{2}\theta \int_{\mathbb{X}} \theta_n \langle f, Z_n^{'(n)}\rangle \nu_0(dz) - \theta_n\langle f, Z_n^{(n)}\rangle\right]\\
&+ \frac{1}{2}\sum_{1\le k\ne i\le n} \left[\int_E \alpha_n \langle f, Z_n^{'(n)}\rangle \delta_{Y_{k,t}}(dz) - \theta_n \langle f, Z_n^{(n)}\rangle\right]
\end{aligned}
$$

Now, if P_i is defined as $Pg(y) = 2^{-1}\theta\int_{\mathbb{X}} g(y)\nu_0(dy)$ applied to the i-th coordinate, we have that $2^{-1}\theta\int_{\mathbb{X}}\langle f, Z_n^{'(n)}\rangle\nu_0(dz)$ corresponds to

$$
\begin{aligned}
&\frac{1}{2}\theta \int_{\mathbb{X}^{n+1}} f(y_1,\ldots,y_n)[Z_n^{'(n)}(dy_1,\ldots,dy_n)]\nu_0(dz) \qquad (4.3.12)\\
&= \frac{1}{2}\theta \int_{\mathbb{X}^{n+1}} f(X_{1,t},\ldots,X_{n,t})\\
&\qquad \times \sum_{1\le j_1\ne\cdots\ne j_n\le n} q_{j_1,\ldots,j_n}\delta_{(Y_{j_1,t},\ldots,Y_{j_{i-1},t},Z,Y_{j_{i+1},t},\ldots,Y_{j_n,t})}(dx_1,\ldots,dx_n)\nu_0(dz)\\
&= \frac{1}{2}\theta \int_{\mathbb{X}} \sum_{1\le j_1\ne\cdots\ne j_n\le n} q_{j_1,\ldots,j_n} f(Y_{j_1,t},\ldots,Y_{j_{i-1},t},Z,Y_{j_{i+1},t},\ldots,Y_{j_n,t})\nu_0(dz)\\
&= \sum_{1\le j_1\ne\cdots\ne j_n\le n} q_{j_1,\ldots,j_n} P_{j_i} f(Y_{j_1,t},\ldots,Y_{j_n,t})\\
&= \int_{\mathbb{X}^n} P_i f(X_1,\ldots,X_n) \sum_{1\le j_1\ne\cdots\ne j_n\le n} q_{j_1,\ldots,j_n}\delta_{(Y_{j_1,t},\ldots,Y_{j_n,t})}(dx_1,\ldots,dx_n)\\
&= \langle P_i f, Z_n^{(n)}\rangle
\end{aligned}
$$

In the same way $\int_{\mathbb{X}} \langle f, Z_n^{'(n)} \rangle \delta_{Y_{k,t}}(dZ)$ corresponds to

$$\begin{aligned}
&\int_{\mathbb{X}^{n+1}} f(X_1, \ldots, X_n)[Z_n^{'(n)}(dx_1, \ldots, dx_n)]\delta_{Y_{k,t}}(dx) \qquad (4.3.13)\\
&= \int_{\mathbb{X}^{n+1}} f(X_1, \ldots, X_n) \\
&\qquad \times \sum_{1 \le j_1 \ne \cdots \ne j_n \le n} q_{j_1,\ldots,j_n} \delta_{(Y_{j_1,t},\ldots,Y_{j_{i-1},t},Z,Y_{j_{i+1},t},\ldots,Y_{j_n,t})}(dx_1, \ldots, dx_n)\delta_{Y_{j_k,t}}(dz) \\
&= \int_{\mathbb{X}} \sum_{1 \le j_1 \ne \cdots \ne j_n \le n} q_{j_1,\ldots,j_n} f(Y_{j_1,t}, \ldots, Y_{j_{i-1},t}, Z, Y_{j_{i+1},t}, \ldots, Y_{j_n,t})\delta_{Y_{j_k,t}}(dz) \\
&= \sum_{1 \le j_1 \ne \cdots \ne j_n \le n} q_{j_1,\ldots,j_n} \Phi_{j_k,j_i} f(Y_{j_1,t}, \ldots, Y_{j_n,t}) \\
&= \int_{\mathbb{X}^n} \Phi_{k,i} f(X_1, \ldots, X_n) \sum_{1 \le j_1 \ne \cdots \ne j_n \le n} q_{j_1,\ldots,j_n} \delta_{(Y_{j_1,t},\ldots,Y_{j_n,t})}(dx_1, \ldots, dx_n) \\
&= \langle \Phi_{k,i} f, Z_n^{(n)} \rangle
\end{aligned}$$

Using (4.3.12) and (4.3.13), (4.3.11) becomes

$$\sum_{i=1}^{n} \theta_n[\langle P_i f, Z_n^{(n)} \rangle - \langle f, Z_n^{(n)} \rangle] + \frac{1}{2} \sum_{1 \le k \ne i \le n} \alpha_n[\langle \Phi_{ki} f, Z_n^{(n)} \rangle - \langle f, Z_n^{(n)} \rangle]$$

Note that for any $m < n$, when $Y_{i,t}$ is not an argument of f we have $P_i f = f$ and $\Phi_{k,i} f = f$, so that

$$\sum_{i=m+1}^{n} \theta_n[\langle P_i f, Z_n^{(n)} \rangle - \langle f, Z_n^{(n)} \rangle] = 0$$

and

$$\sum_{i=m+1}^{n} \sum_{k=1,k \ne i}^{n} \theta_n[\langle \Phi_{k,i} f, Z_n^{(n)} \rangle - \langle f, Z_n^{(n)} \rangle] = 0.$$

Further, when $Y_{i,t}$ is an argument of f but $x_{k,t}$ is not, we have $\langle \Phi_{k,i} f, \mu \rangle = \langle f, \mu \rangle$, so that

$$\sum_{i=1}^{m} \sum_{k=m+1}^{n} \theta_n[\langle \Phi_{k,i} f, Z_n^{(n)} \rangle - \langle f, Z_n^{(n)} \rangle] = 0.$$

Hence, we have

$$(\mathcal{L}\phi)(P^{(n)}) = \sum_{i=1}^{m} \theta_n \langle A_i f, Z_n^{(m)} \rangle + \frac{1}{2} \sum_{1 \le k \ne i \le m} \theta_n[\langle \Phi_{k,i} f, Z_n^{(m)} \rangle - \langle f, Z_n^{(m)} \rangle]$$

Moving from the above remarks, se can now proceed to prove the result given in Theorem 4.3.2.

Proof. Call P its weak limit for $n \to \infty$. For $\mu^{(m)}$ as in Donelly and Kurtz [30], it is easy to check that

$$\sup_{y \in \mathbb{X}^n} \left| \langle f, \mu^{(m)} \rangle - \langle f, \mu^m \rangle \right| \tag{4.3.14}$$

$$= \sup_{y \in \mathbb{X}^n} \frac{1}{n^m} \left| \sum_{1 \leq i_1 \neq \cdots \neq i_m \leq n} f(Y_{i_1}, \ldots, Y_{i_m}) - \sum_{1 \leq i_1, \ldots, i_m \leq n} f(Y_{i_1}, \ldots, Y_{i_m}) \right| \to 0$$

Recalling (4.3.5), observe now that for large n, $q_{i_1,\ldots,i_m}$ behaves like $1/(n)_{m\downarrow 1}$, since $\mathbb{E}[q_i] = 1/n$ and $Var(q_i) = (n-1)/n^2(n+1)$. Since furthermore θ_n converges to one in probability it follows that we can replace $\mu^{(m)}$ with $P_n^{(m)}$ in (4.3.14), yielding

$$\sup_{y \in \mathbb{X}^n} \left| \langle f, P_n^{(m)} \rangle - \langle f, \mu^m \rangle \right| \to 0.$$

This easily implies

$$\sup_{y \in E^n} |(\mathcal{L}\phi)(P_n) - (\mathcal{L}\varphi)(P)| \to 0$$

where $(\mathcal{L}\varphi)(P)$ is the generator of the neutral diffusion model. Since $\mu_n(0)$ converges weakly (and in particular to a Dirichlet process) the above computation implies (4.3.4). □

4.4 The particle process

Before stating the main result, we give the following lemma, which beside having a key role in the construction, provides some intuition into the problem. The lemma provides a new constructive definition of the Dirichlet process. As recalled in the introduction, the Dirichlet process has been characterised via the Blackwell-MacQueen Pólya urn scheme, i.e. a sequence of observations $\{X_n, n \geq 1\}$ from the Blackwell-MacQueen Pólya urn scheme (4.1.1) is equivalent to a sequence $\{X_n, n \geq 1\}$ of i.i.d. observations from P, where P is a Dirichlet process with parameter ν.

Lemma 4.4.1. *Let $\{X_n, n \geq 1\}$ be a sequence of random variables (r.v.s) from the Blackwell-MacQueen Pólya urn scheme* (4.1.1) *and for any $n \in \mathbb{N}$, let $(X_1, \ldots, X_n)$ be a sample of size n from $\{X_n, n \geq 1\}$. Let $w_1, \ldots, w_n$ be n independent r.v.s distributed according to a Beta distribution function with parameters $(1, \theta + n - i)$ for $i = 1, \ldots, n$. If P is a Dirichlet process on $\mathbb{X}$ with parameter ν independent of $w_1, \ldots, w_n$ and for any $n \in \mathbb{N}$, $\omega \in \Omega$ and $B \in \mathscr{X}$ we consider the measurable map $P_n : \Omega \to \mathcal{P}_{\mathbb{X}}$ defined by*

$$P_n(\omega, B) := \sum_{i=1}^{n} p_i(\omega) \delta_{X_i(\omega)}(B) + \left(1 - \sum_{i=1}^{n} p_i(\omega)\right) P(\omega, B) \tag{4.4.1}$$

where $p_1 = w_1$ *and* $p_i = w_i \prod_{1\le j\le i-1}(1-w_j)$ *for* $i = 2,\dots,n$, *then* P_n *is a Dirichlet process with parameter* ν.

Proof. From the definition of Dirichlet process on a Polish space it follows that for any $k \in \mathbb{N}$ it suffices to prove that statement for an h-dimensional vector $(P(\cdot,A_1),\dots,P(\cdot,A_k))$, where $A_1,\dots,A_k$ is any finite measurable partition of $\mathbb{X}$. For all $n \ge 1$ we have that $1-\sum_{1\le i\le n} p_i = \prod_{1\le i\le n}(1-w_i)$. Using the stick-breaking of the r.v.s $p_1,\dots,p_n$ we have

$$\begin{aligned}
&\sum_{i=1}^{n} p_i(\delta_{X_i}(A_1),\dots,\delta_{X_i}(A_k)) + \left(1-\sum_{i=1}^{n} p_i\right)(P(\cdot,A_1),\dots,P(\cdot,A_k))\\
&\quad= \sum_{i=1}^{n-1} p_i(\delta_{X_i}(A_1),\dots,\delta_{X_i}(A_k)) + \left(1-\sum_{i=1}^{n-1} p_i\right)\\
&\qquad\qquad \times (w_n(\delta_{X_n}(A_1),\dots,\delta_{X_n}(A_k)) + (1-w_n)(P(\cdot,A_1),\dots,P(\cdot,A_k))).
\end{aligned}$$

Then, it follows by induction that conditionally on $(X_1,\dots,X_n)$,

$$\sum_{i=1}^{n} p_i(\delta_{X_i}(A_1),\dots,\delta_{X_i}(A_k)) + \left(1-\sum_{i=1}^{n} p_i\right)(P(\cdot,A_1),\dots,P(\cdot,A_k))$$

is a r.v. distributed according to a Dirichlet distribution function with parameters $(\nu(A_1)+\sum_{1\le i\le n}\delta_{X_i}(A_1),\dots,\nu(A_h)+\sum_{1\le i\le n}\delta_{X_i}(A_k))$. The result follows integrating out the r.v.s $X_1,\dots,X_n$. □

The sample path of the neutral diffusion model $\{\mu_t, t\ge 0\}$ at stationarity is such that at each time point the state of the process is a Dirichlet process. From Lemma 4.4.1 it follows that a representation alternative to (4.1.1)-(4.1.2) of a Dirichlet process is given by (4.4.1), which can thus be used, once indexed by time, to describe any instant state of the neutral diffusion model. Given the almost sure discreteness of the Dirichlet process, the connection between two states of the process at different time points, say without loss of generality 0 and $t > 0$, can be expressed according to how many atoms $\mu_{0,n}$ and $\mu_{t,n}$ share, for arbitrary $n \ge 1$, where for any $t \ge 0$

$$\mu_{t,n} := \sum_{i=1}^{n} p_{t,i}\delta_{X_{t,i}} + \left(1-\sum_{i=1}^{n} p_{t,i}\right)\mu_t.$$

Thus, the change in time of $X_1,\dots,X_n$ in (4.4.1) provides an approximation of the change undergone by μ_n. Given a sequence of r.v.s from the Blackwell-MacQueen Pólya urn scheme (4.1.1) a sample $X_1,\dots,X_n$ from this sequence is then a natural candidate for a finite-dimensional particle process whose components in any instant are from the population μ_n. Since the dynamics of the particle process reflect to a certain extent those of the measure-valued process, $X_{0,1},\dots,X_{0,n}$ will remain fixed at their state during the interval $[0,t)$ so long as $X_1,\dots,X_n$ remain atoms of $\mu_{s,n}$ for $0\leq s<t$. When one of the atoms drops out, the state of this $\mathbb{X}^n$-valued random process changes, so it is componentwise piecewise constant with jumps. We are then interested in the distribution of interarrival times between jumps, that is the holding times between any atom change. We will show that the atoms change one at a time, and the holding times are exponential with parameter λ_n given in (4.1.6). Once again we remark that these results on the dynamic properties of the particle process will rely only on the transition function (4.1.5) of the neutral diffusion model, with no further assumptions. The next theorem, which is the main result of the chapter, formalizes the above heuristics. It will be proved by means of several lemmas in the remainder of the section.

Theorem 4.4.1. *For any arbitrary $n\in\mathbb{N}$, let $(\mu_{t,n},t\geq 0)$ be the neutral diffusion model with infinitesimal generator* (4.1.3)*. Then, $\{(X_{t,1},\dots,X_{t,n}),t\geq 0\}$ is a n-dimensional particle process with sample paths in $D_{\mathbb{X}^n}^{[0,+\infty)}$ and jumps at exponential times of parameter λ_n, given by* (4.1.6)*, such that at each jump at most one coordinate at a time is updated according to* (4.1.1)*.*

Lemma 4.4.2. *(cfr. Walker et al. [193]) For any $k\in\mathbb{N}_0$, let $d_k(t)$ be* (4.1.7)*. Then*

$$\sum_{k\geq n}\frac{(k)_{n\downarrow 1}}{(\theta+k)_{n\uparrow 1}}\,d_k(t)=e^{-\lambda_n t} \tag{4.4.2}$$

and

$$\sum_{k\geq n-1}\frac{(k)_{(n-1)\downarrow 1}}{(\theta+k)_{n\uparrow 1}}d_k(t)=\frac{e^{-\lambda_{n-1}t}-e^{-\lambda_n t}}{2(\lambda_n-\lambda_{n-1})}. \tag{4.4.3}$$

The following lemma provides a useful result that will be used later.

Lemma 4.4.3. *Let $\theta>0$ and $k,n\in\mathbb{N}$, with $n\leq k$. Then*

$$\sum_{n=1}^{k}\frac{\Gamma(\theta+k-n)}{\Gamma(1+k-n)}=\frac{\Gamma(\theta+k)}{\theta\Gamma(k)}.$$

Proof. We have

$$\begin{aligned}\sum_{n=1}^{k}\frac{\Gamma(\theta+k-n)}{\Gamma(1+k-n)}&=\frac{\Gamma(\theta+k)}{\Gamma(1+k)}\sum_{n=1}^{k}\frac{(k)_{n\downarrow 1}}{(\theta+k-1)_{n\downarrow 1}}\\&=\frac{\Gamma(\theta+k)}{\Gamma(1+k)(\theta+k-1)_{k\downarrow 1}}\left(\sum_{n=1}^{k-1}(\theta+k-1-n)_{(k-n)_{\downarrow 1}}(k)_{n\downarrow 1}+(k)_{k\downarrow 1}\right)\\&=\frac{\Gamma(\theta+k)}{\Gamma(1+k)(\theta+k-1)_{k\downarrow 1}}k(\theta+k-1)_{(k-1)\downarrow 1}=\frac{\Gamma(\theta+k)}{\theta\Gamma(k)}.\end{aligned}$$

□

We have now all the ingredients to show that the interarrival times between successive jumps, that is single atom updates, are exponential with parameter λ_n. This will be proved by means of the following three propositions. Let $\{\mu_t, t \geq 0\}$ be a neutral diffusion model, so that the transitions of μ_t are described by (4.1.5). The form of the transition function yields that conditionally on the starting state μ_0, the arrival state $\mathrm{d}\mu_t$ after a time interval t is obtained as follows. An n-sized sample $(X_1, \dots, X_n)$ is drawn from μ_0, where the sample size n is governed by a death process $\{D_t, t \geq 0\}$ starting from infinity, so that the probability of sampling m variables from μ_0 for an interval of lag t is $d_m(t)$. Then μ_t is sampled from a posterior Dirichlet process, conditionally on the vector $(X_1, \dots, X_n)$. Hence the m-sized vector sampled from the starting state μ_0 carries m atoms of information about μ_0, which are taken into account when sampling μ_t.

We exploit these intrinsic features of the transition function (4.1.5) for computing the probability that respectively none, one or two atoms of $\mu_{0,n}$ among those in $(X_1, \dots, X_n)$ drop in the interval dt. These three cases will be examined separately in Proposition 4.4.1, 4.4.2 and 4.4.3 below.

Proposition 4.4.1. *Let $\{\mu_t, t \geq 0\}$ be a neutral diffusion model with transition function* (4.1.5)*, and suppose the time interval $[0, s]$ is of infinitesimal length. Then the probability of $(X_{0,1}, \dots, X_{0,n})$ being atoms of μ_s is $e^{-\lambda_n s}$, where λ_n is* (4.1.6)*.*

Proof. Call $n_1, \dots, n_n$ the multiplicity of $X_{0,1}, \dots, X_{0,n}$ respectively in an k-sized sample from μ_0, where μ_0 is given by (4.4.1). A necessary condition for $X_{0,1}, \dots, X_{0,n}$ to be in the k-sized sample from μ_0, and hence possibly be atoms of μ_s, is that k be not smaller than n, and that $\sum_{1 \leq i \leq n} n_i \leq k$. Hence we have to integrate: over the random weights $p_1, \dots, p_n$ associated to the atoms $X_1, \dots, X_n$, whose distribution is derived by the stick-breaking procedure, also known as residual allocation model, in Lemma 4.4.1; over all possible combinations of multiplicities of atom draws in a sample of size k, so that $n_1 \in \{1, \dots, k\}$, $n_2 \in \{1, \dots, k - n_1\}$, and so on up to $n_n \in \{1, \dots, k - \sum_{i=1}^{n} n_i\}$, so that $\sum_{1 \leq i \leq n} n_i \leq k$; and over the sample size for $k \geq n$. Hence we have that the probability of $(X_{0,1}, \dots, X_{0,n})$ being atoms in μ_s is

$$\begin{aligned}\mathbb{P}((X_{0,1}, \dots, X_{0,n}) \in X_{s,\infty}) = \sum_{k \geq n} d_k(s) \sum_{n_1=1}^{k} \sum_{n_2=1}^{k-n_1} \cdots \sum_{n_n=1}^{k-n_1-\ldots-n_{n-1}} \binom{k}{n_1, \dots, n_n} \\ \times \int_{(0,1)^n} \prod_{i=1}^{n} \left(w_i^{n_i} \prod_{j=1}^{i-1} (1 - w_j)^{n_i} \right) (1 - w_i)^{k - \sum_{h=1}^{n} n_h} \\ \times \prod_{l=1}^{n} (\theta + n - l)(1 - w_l)^{\theta + n - l - 1} dw_1 \dots dw_n\end{aligned}$$

which simplifies to

$$\sum_{k\geq n}(\theta)_{n\uparrow 1}d_k(s)\sum_{n_1=1}^{k}\sum_{n_2=1}^{k-n_1}\cdots\sum_{n_n=1}^{k-n_1-\ldots-n_{n-1}}\binom{k}{n_1,\ldots,n_n}$$
$$\times\int_{(0,1)^n}\prod_{i=1}^{n}w_i^{n_i}(1-w_i)^{\theta+k-\sum_{h=1}^{i}n_h+n-i-1}dw_1\ldots dw_n.$$

By solving the integrals, the previous equals

$$\sum_{k\geq n}(\theta)_{n\uparrow 1}d_k(s)\sum_{n_1=1}^{k}\sum_{n_2=1}^{k-n_1}\cdots\sum_{n_n=1}^{k-n_1-\ldots-n_{n-1}}\binom{k}{n_1,\ldots,n_n}$$
$$\times\prod_{i=1}^{n}\frac{\Gamma(n_i+1)\Gamma(\theta+k-\sum_{h=1}^{i}n_h+n-i)}{\Gamma(\theta+k-\sum_{h=1}^{i-1}n_h+n-i)}$$

and simplifying the product with the multinomial coefficient gives

$$\sum_{k\geq n}(\theta)_{n\uparrow 1}\frac{\Gamma(k+1)}{\Gamma(\theta+k+n)}d_k(s)\sum_{n_1=1}^{m}\sum_{n_2=1}^{k-n_1}\cdots\sum_{n_n=1}^{k-n_1-\ldots-n_{n-1}}\frac{\Gamma(\theta+k-\sum_{h=1}^{n}n_h)}{\Gamma(k-\sum_{h=1}^{n}n_h+1)}.$$

Applying Lemma 4.4.3 yields

$$\sum_{k\geq n}(\theta)_{n\uparrow 1}\frac{\Gamma(k+1)}{\Gamma(\theta+k+n)}d_k(s)\sum_{n_1=1}^{m}\cdots\sum_{n_{n-1}=1}^{k-n_1-\ldots-n_{n-2}}\frac{\Gamma(\theta+k-\sum_{h=1}^{n-1}n_h)}{\theta\,\Gamma(k-\sum_{h=1}^{n-1}n_h)}.$$

Take now $\theta':=\theta+1$ and $k':=k-1$, so that the last ratio in the previous corresponds to $\Gamma(\theta'+k'-\sum_{1\leq h\leq n-1}n_h)/\theta\Gamma(k'-\sum_{1\leq h\leq n-1}n_h+1)$ and apply again Lemma 4.4.3 to get

$$\sum_{k\geq n}(\theta)_{n\uparrow 1}\frac{\Gamma(k+1)}{\Gamma(\theta+k+n)}d_k(s)\sum_{n_1=1}^{m}\cdots\sum_{n_{n-2}=1}^{k-n_1-\ldots-n_{n-3}}\frac{\Gamma(\theta'+k'-\sum_{h=1}^{n-2}n_h)}{\theta\,\theta'\,\Gamma(k'-\sum_{h=1}^{n-2}n_h)}$$
$$=\sum_{k\geq n}(\theta)_{n\uparrow 1}\frac{\Gamma(k+1)}{\Gamma(\theta+k+n)}d_k(s)\sum_{n_1=1}^{k}\cdots\sum_{n_{n-2}=1}^{k-n_1-\ldots-n_{n-3}}\frac{\Gamma(\theta+k-\sum_{h=1}^{n-2}n_h)}{\theta(1+\theta)\Gamma(k-\sum_{h=1}^{n-2}n_h-1)}.$$

Repeat the procedure other $n-2$ times, taking $\theta'':=\theta'+1$, $k'':=k'+1$ and so on, yielding

$$\sum_{k\geq n}(\theta)_{n\uparrow 1}\frac{\Gamma(k+1)}{\Gamma(\theta+k+n)}d_k(s)\sum_{n_1=1}^{k}\cdots\sum_{n_{n-3}=1}^{k-n_1-\ldots-n_{n-4}}\frac{\Gamma(\theta+k-\sum_{h=1}^{n-3}n_h)}{\theta(1+\theta)(2+\theta)\Gamma(k-\sum_{h=1}^{n-3}n_h-1)}$$
$$\vdots$$
$$=\sum_{k\geq n}(\theta)_{n\uparrow 1}\frac{\Gamma(k+1)}{\Gamma(\theta+k+n)}d_k(s)\frac{\Gamma(\theta+k)}{\theta(\theta+1)\ldots(\theta+n-1)\Gamma(k-n+1)}$$
$$=\sum_{k\geq n}\frac{(k)_{n\downarrow 1}}{(\theta+k)_{n\uparrow 1}}\,d_k(s)$$

which by means of (4.4.2) gives the result. □

The following proposition gives the probability that one atom update occurs in an infinitesimal lag.

Proposition 4.4.2. *Let $\{\mu_t, t \geq 0\}$ be a neutral diffusion model with transition function* (4.1.5)*, and suppose the time interval* $[0, s]$ *is of infinitesimal length. The probability that exactly $n-1$ particles of the vector $(X_{0,1}, \ldots, X_{0,n})$ are atoms in μ_s is $\lambda_n s + o(s)$.*

Proof. Consider the setting of the proof of Proposition 4.4.1. If the atom that changes is X_j, $1 \leq j \leq n$, in order to compute the probability of the statement it suffices to set $n_j = 0$ in (4.4.1), so that there are no values of $X_{0,j}$ in the k-sized sample from μ_0 (hence no piece of information about μ_0 corresponding to the atom $X_{0,j}$ pass to μ_s). Hence the probability that one atom drops out is

$$
\begin{aligned}
&\sum_{j=1}^{n} \mathbb{P}((X_{0,1}, \ldots, X_{0,j-1}, X_{0,j+1}, \ldots, X_{0,n}) \in X_{s,\infty}, X_{0,j} \notin X_{s,\infty}) \\
&\quad = \sum_{j=1}^{n} \sum_{k \geq n-1} (\theta)_{n\uparrow 1} d_k(s) \\
&\qquad \times \sum_{n_1=1}^{k} \sum_{n_2=1}^{k-n_1} \cdots \sum_{n_{j-1}=1}^{k-\sum_{l=1}^{j-2} n_l} \cdots \sum_{n_{j+1}=1}^{k-\sum_{l=1}^{j-1} n_l} \sum_{n_n=1}^{k-\sum_{l\neq j}^{n-1} n_l} \binom{k}{n_1, \ldots, n_{j-1}, n_{j+1}, \ldots, n_n} \\
&\qquad \times \int_{(0,1)^n} \prod_{i\neq j}^{n} w_i^{n_i} (1-w_i)^{\theta+k-\sum_{h\neq j}^{i} n_h + n-i-1} (1-w_j)^{\theta+k-\sum_{h=1}^{j-1} n_h + n-j-1} dw_1 \ldots dw_n
\end{aligned}
$$

Proceeding as in Proposition 4.4.1, and simplifying with the multinomial coefficient the Gamma functions resulting from the integrals, yields

$$
\begin{aligned}
&\sum_{j=1}^{n} \sum_{k \geq n-1} (\theta)_{n\uparrow 1} \frac{\Gamma(k+1)}{\Gamma(\theta+k+n)} d_k(s) \\
&\qquad \times \sum_{n_1=1}^{k} \sum_{n_2=1}^{k-n_1} \cdots \sum_{n_{j-1}=1}^{k-\sum_{l=1}^{j-2} n_l} \sum_{n_{j+1}=1}^{k-\sum_{l=1}^{j-1} n_l} \cdots \sum_{n_n=1}^{k-\sum_{l\neq j}^{n-1} n_l} \frac{\Gamma(\theta + k - \sum_{h\neq j}^{n} n_h)}{\Gamma(k - \sum_{h\neq j}^{n} n_h + 1)}
\end{aligned}
$$

Applying $n-1$ times Lemma 4.4.3 we obtain

$$\begin{aligned}
&\sum_{j=1}^{n}\sum_{k\geq n-1}(\theta)_{n\uparrow 1}\frac{\Gamma(k+1)}{\Gamma(\theta+k+n)}d_k(s)\\
&\qquad\times\sum_{n_1=1}^{k}\sum_{n_2=1}^{k-n_1}\cdots\sum_{n_{j-1}=1}^{k-\sum_{l=1}^{j-2}n_l}\sum_{n_{j+1}=1}^{k-\sum_{l=1}^{j-1}n_l}\cdots\sum_{n_{n-1}=1}^{k-\sum_{l\neq j}^{n-2}n_l}\frac{\Gamma(\theta+k-\sum_{h\neq j}^{n-1}n_h)}{\theta\Gamma(k-\sum_{h\neq j}^{n-1}n_h)}\\
&=\sum_{j=1}^{n}\sum_{k\geq n-1}(\theta)_{n\uparrow 1}\frac{\Gamma(k+1)}{\Gamma(\theta+k+n)}d_k(s)\\
&\qquad\times\sum_{n_1=1}^{k}\sum_{n_2=1}^{k-n_1}\cdots\sum_{n_{j-1}=1}^{k-\sum_{l=1}^{j-2}n_l}\sum_{n_{j+1}=1}^{k-\sum_{l=1}^{j-1}n_l}\cdots\sum_{n_{n-2}=1}^{k-\sum_{l\neq j}^{n-3}n_l}\frac{\Gamma(\theta+k-\sum_{h\neq j}^{n-2}n_h)}{\theta(1+\theta)\Gamma(k-\sum_{h\neq j}^{n-2}n_h-1)}\\
&\;\vdots\\
&=\sum_{j=1}^{n}\sum_{k\geq n-1}(\theta)_{n\uparrow 1}\frac{\Gamma(k+1)}{\Gamma(\theta+k+n)}d_k(s)\frac{\Gamma(\theta+k)}{\theta(\theta+1)\dots(\theta+n-2)\Gamma(k-n+2)}\\
&=n(\theta+n-1)\sum_{k\geq n-1}\frac{(k)_{(n-1)\downarrow 1}}{(\theta+k)_{n\uparrow 1}}d_k(s)
\end{aligned}$$

from which, using (4.4.3) and the definition of λ_n, we get $\lambda_n(e^{-\lambda_{n-1}s}-e^{-\lambda_n s}/\lambda_n-\lambda_{n-1}) = \lambda_n s+o(s)$ which gives the result. □

Before stating the last proposition, we need the following technical result.

Lemma 4.4.4. *Let* $d_k(s)$ *be* (4.1.7)*. Then*

$$\sum_{k\geq n-2}\frac{(k)_{(n-2)\downarrow 1}}{(\theta+k)_{n\uparrow 1}}d_k(s)=\frac{(\lambda_{n-1}-\lambda_{n-2})e^{-\lambda_n s}-(\lambda_n-\lambda_{n-2})e^{-\lambda_{n-1}s}+(\lambda_n-\lambda_{n-1})e^{-\lambda_{n-2}s}}{4(\lambda_n-\lambda_{n-1})(\lambda_n-\lambda_{n-2})(\lambda_{n-1}-\lambda_{n-2})}.$$

Proof. Denote

$$G(t):=\sum_{k\geq n-2}\frac{(k)_{(n-2)\downarrow 1}}{(\theta+k)_{n\uparrow 1}}d_k(s);$$

from a result in Ethier and Griffiths [41], it follows that

$$\frac{dG(s)}{ds}+\lambda_n G(s)=\frac{1}{2}\sum_{k\geq n-2}\frac{(k)_{(n-2)\downarrow 1}}{(\theta+k)_{(n-1)\uparrow 1}}d_k(s)$$

and we know from Walker et al. [193] hat

$$\sum_{k\geq n-2}\frac{(k)_{(n-2)\downarrow 1}}{(\theta+k)_{(n-1)\uparrow 1}}d_k(s)=\frac{e^{-\lambda_{n-2}s}-e^{-\lambda_{n-1}s}}{2(\lambda_{n-1}-\lambda_{n-2})}.$$

The general solution of the differential equation is

$$G(s) = \frac{e^{-\lambda_{n-2}s}}{4(\lambda_n - \lambda_{n-2})(\lambda_{n-1} - \lambda_{n-2})} - \frac{e^{-\lambda_{n-1}s}}{4(\lambda_n - \lambda_{n-1})(\lambda_{n-1} - \lambda_{n-2})} + Ce^{-\lambda_n s}$$

and using the initial condition $G_0 = 0$ we obtain

$$C = \frac{1}{4(\lambda_n - \lambda_{n-1})(\lambda_n - \lambda_{n-2})}$$

from which the result follows □

The last proposition states that the probability of two atoms updates occurring in an infinitesimal time lag is negligible.

Proposition 4.4.3. *Let $\{\mu_t, t \geq 0\}$ be a neutral diffusion model with transition function (4.1.5), and suppose the time interval $[0, s]$ is of infinitesimal length. The probability that only $n-2$ particles of the vector $(X_{0,1}, \ldots, X_{0,n})$ are atoms in μ_s is $o(s)$.*

Proof. The event of two particles changing in $[0, s]$ means that $X_{0,j}, X_{0,h}$ for $1 \leq j \neq h \leq n$, are not selected in the m-sized sample from μ_0 and thus do not compare as atoms in μ_s. Similarly to Proposition 4.4.2, we set $n_j = n_h = 0$, and integrate out the indices, obtaining

$$\begin{aligned}
&\sum_{1\leq j\neq h\leq n} \mathbb{P}((X_{0,i},\ i \neq j, h) \in X_{s,\infty}, (X_{0,j}, X_{0,h}) \notin X_{s,\infty}) \\
&\quad = \sum_{1\leq j\neq h\leq n}^{n} \sum_{k\geq n-1} (\theta)_{n\uparrow 1} d_k(s) \sum_{(*)} \binom{k}{n_1, \ldots, n_{j-1}, n_{j+1}, \ldots, n_{h-1}, n_{h+1}, \ldots, n_n} \\
&\qquad \times \prod_{i\neq j,h}^{n} \int_0^1 w_i^{n_i} (1-w_i)^{\theta+k-\sum_{l\neq j,h}^{i} n_l + n-i-1} (1-w_j)^{\theta+k-\sum_{l\neq j,h}^{j-1} n_l+n-j-1} dw_i \\
&\qquad \times \int_0^1 (1-w_j)^{\theta+k-\sum_{l\neq h}^{j-1} n_l+n-j-1} dw_j \int_0^1 (1-w_h)^{\theta+k-\sum_{l\neq j}^{h-1} n_l+n-h-1} dw_h
\end{aligned}$$

where $(*)$ denotes the set of frequencies n_i for $1 \leq i \leq n$ and $i \neq h, n$, such that each n_i runs from 1 to $k - \sum_{l\neq j,h} n_l$. Proceeding as in Proposition 4.4.1 we obtain

$$n(n-1)(\theta+n-1)(\theta+n-2) \sum_{k\geq n-2} \frac{(k)_{(n-2)\downarrow 1}}{(\theta+k)_{n\uparrow 1}} d_k(s).$$

By Lemma 4.4.4 the previous equals, up to a multiplicative constant,

$$\begin{aligned}
&\lambda_n(e^{-\lambda_{n-2}s} - e^{-\lambda_{n-1}s}) + \lambda_{n-1}(e^{-\lambda_n s} - e^{-\lambda_{n-2}s}) + \lambda_{n-2}(e^{-\lambda_{n-1}s} - e^{-\lambda_n s}) \\
&\qquad = s\lambda_n(\lambda_{n-1} - \lambda_{n-2}) + s\lambda_{n-1}(\lambda_{n-2} - \lambda_n) + s\lambda_{n-2}(\lambda_n - \lambda_{n-1}) + o(s) = o(s)
\end{aligned}$$

which gives the result. □

Propositions 4.4.1, 4.4.2 and 4.4.3 imply that the interarrival times of the particle process are governed by a Poisson process with parameter λ_n, and that one particle at a time drops out of the n-dimensional time-dependent vector. Say that Y_i is such particle. Then, from Lemma 4.4.1 and the exchangeability of a sequence drawn according to (4.1.1), it follows that for any $n \in \mathbb{N}$, the incoming particle is a sample from

$$\frac{\theta}{\theta+n-1}\,\nu_0 + \frac{1}{\theta+n-1}\,\sum_{j\neq i}\delta_{X_j}. \tag{4.4.4}$$

This is due to the fact that conditionally on μ_t, the removed particle will be replaced by another variable in the infinite sequence from the Blackwell-MacQueen Pólya urn scheme that characterizes μ_t. Integrating out μ_t, the incoming variable will still be from the Blackwell-MacQueen Pólya urn scheme, but conditionally on the other $n-1$ particles, and its law will be the predictive distribution (4.4.4). This completes the proof of Theorem 4.4.1.

4.5 Discussion

We have constructed a particle process which is directly derived by the properties of neutral diffusion model. The key of the derivation is the representation of a Dirichlet process as μ_n in (4.4.1), as proved in Lemma 4.4.1. Then, given n atoms $(X_{0,1},\ldots,X_{0,n})$ of the starting state μ_0 of the neutral diffusion model, we can describe a particle process as follows. The state of the particle process remains constant until the first time t such that one of the particles is no longer an atom of μ_t. The computation of the probabilities that all n particles are still atoms of μ_t and that one of the n particles is no longer an atom of μ_t yields the distribution of the interarrival time of the particle process until the following renewal. When one of the particles is no longer an atom of the random measure, not having been sampled from the starting state, it is substituted with another atom of μ_0 which differs from the other $n-1$, and hence is another observation from the Blackwell-MacQueen Pólya urn scheme.

When the population size of the particle process grows to infinity, in Lemma 4.4.1 we have that the sum of weights $\sum_{1\leq i\leq n} p_i$ tends to one, and the second term in

$$P_n = \sum_{i=1}^{n} p_i\delta_{X_i} + \left(1-\sum_{i=1}^{n} p_i\right) P$$

vanishes. Then as $n\to\infty$,

$$P_n \Rightarrow P_\infty \qquad \text{a.s.-}\mathbb{P}$$

where P_∞ is still a Dirichlet process, but unlike for finite k, the particle process now fully characterises any instant state of the neutral diffusion model, as we have an infinite sequence of observations from μ_t, conditionally on μ_t, which provides full information on the distribution. From this setting it is now trivial to derive all usual infinite population results for the neutral diffusion model, like the weak convergence in the Skorohod space of the process of empirical measures of the particles.

II

Models beyond the Dirichlet process and applications to species sampling problems

5

On a class of random probability measures without Gibbs structure

One of the main research areas in Bayesian nonparametrics is the proposal and study of priors which generalize the Dirichlet process. In this chapter, by considering priors obtained by normalizing random measures with independent increments (NRMIs) we define a class on NRMIs, the so-called NRMIs with logarithmic singularity. This new class of random probability measures includes as particular case the celebrated Dirichlet process and on the other hand it does not include the normalized generalized Gamma process recently introduced in the context of mixture models and species sampling problems. In particular, our aim is to provide some developments for a random probability measures in the class of NRMIs with logarithmic singularity, termed generalized Dirichlet processes, which has been recently introduced in the literature. Such processes induce exchangeable sequences which are characterized by a more elaborated predictive structure that those one arising from Gibbs-type random probability measures. A natural area of application of these random probability measures is represented by species sampling problems and, in particular, prediction problems in genomics.

5.1 Introduction

Let $\{X_n, n \geq 1\}$ be an exchangeable sequence defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and such that each random variable (r.v.) X_n takes values in a Polish space $(\mathbb{X}, \mathcal{T})$ with associated Borel σ-field $\mathscr{X}$. Then, by de Finetti representation theorem, there exists a random probability measure (r.p.m.) $\tilde{P}$ on $\mathbb{X}$ with law Q such that given $\tilde{P}$, a sample $X_1, \ldots, X_n$ from the exchangeable sequence is independent and identically distributed (i.i.d.) with distribution $\tilde{P}$. That is, for any $n \geq 1$ and any $A_1, \ldots, A_n \in \mathscr{X}$

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \int_{\mathcal{P}_{\mathbb{X}}} \prod_{i=1}^{n} \tilde{P}(A_i) Q(d\tilde{P})$$

where $\mathcal{P}_{\mathbb{X}}$ denotes the space of probability measures on $(\mathbb{X}, \mathscr{X})$ endowed with its σ-field $\mathscr{P}_{\mathbb{X}}$ generated by the weak convergence topology $\mathcal{W}$ which makes $(\mathcal{P}_{\mathbb{X}}, \mathcal{W})$ a Polish space. Here, we focus on r.p.m.s which are almost surely discrete and with non-atomic prior guess at the shape $\alpha_0(\cdot) := \mathbb{E}[\tilde{P}(\cdot)]$; by the almost sure discreteness, we expect ties in the sample, namely that $X_1, \dots, X_n$ contain $k \leq n$ distinct observations $X_1^*, \dots, X_k^*$ with frequencies $n_1, \dots, n_k$) such that $\sum_{j=1}^k n_j = n$. Initial opinions about the sequence $\{X_n, n \geq 1\}$ are expressed through the probability distribution Q of $\tilde{P}$ and updated by means of Bayes theorem; all inferential questions related to the sequence $\{X_n, n \geq 1\}$ are the answered on the basis on the conditional probability distribution of $\tilde{P}$, given the observed sample $X_1, \dots, X_n$ from $\{X_n, n \geq 1\}$. This fact implies that a key problem in Bayesian nonparametric inference is the definition of a prior distribution Q on the space of all probability measures $(\mathcal{P}_{\mathbb{X}}, \mathscr{P}_{\mathbb{X}})$.

Starting from the papers by Freedman [72] and Ferguson [61], in which the celebrated Dirichlet process has been introduced various approaches for constructing r.p.m.s, whose distribution acts as a nonparametric prior, have been undertaken. They all aim at providing generalisations of the Dirichlet process. Among them we mention the neutral to the right r.p.m.s due to Doksum [25], which are obtained via an exponential transformation of an increasing process with independent increments, and the Pólya tree priors thoroughly studied by Mauldin et al. [135] and Lavine [116], which arise by considering suitable urn schemes on trees of nested partitions. In this chapter we focus on r.p.m.s derived by a suitable normalization procedure. To this end, it is worth recalling that the Dirichlet process can be defined by normalizing the increments of a Gamma process. Indeed, the idea of constructing r.p.m.s by means of a normalization procedure han been exploited an developed in a variety of contexts not closely related to Bayesian inference. See, as an early example, Kingman [109] where a random discrete distribution generated by the stable subordinator is considered in connection with optimal storage problems. Further examples can be found in Ewens and Tavaré [46] and Grote and Speed [79] for population genetics; in Donnelly and Grimett [27] and Pitman [157] for combinatorics and number theory; Pitman [153] and Pitman and Yor [154] for excursion theory.

In particular, Kingman [109] suggested that one can construct r.p.m.s as follows. First take the ranked points of a homogeneous Poisson process on $\mathbb{R}^+$, say $\{\Delta_i, i \geq 1\}$ such that their sum $\sum_{i\geq 1} \Delta_i$ is finite and positive a.s. Use these points to construct a sequence of probabilities $Q_i = \Delta_i / \sum_{i\geq 1} \Delta_i$ for $i \geq 1$. Independent of this sequence choose a sequence $\{Z_i, i \geq 1\}$ to be an i.i.d. sequence of random elements of a Polish space with common distribution, say H. A r.p.m. is then formed by $\sum_{i\geq 1} Q_i \delta_{Z_i}$. With the exception of the Dirichlet process and those models based on the stable law, the processes in class described by Kingman [109] have yet to yield tractable results suitable for practical implementation in Bayesian nonparametric problems. Thus, the main question of interest to a Bayesian statistician is whether there are any other processes in this class which are tractable. In Regazzini et al. [165] the class of normalized random measures with independent increments (NRMIs) on $\mathbb{R}$ is formally introduced as normalization of suitably time-changed independent increment processses and distributional results for their means are derived. Further developments related to the class of NRMIs can be found in James [96] and Nieto-Barajas et al. [145]. Recently, in order to both understand better the structural properties of the NRMIs on $\mathbb{X}$ and go beyond the specific processes dealt with in the above mentioned papers, James et al. [100] provided a complete and implementable description of the posterior distribution of a NRMI.

In Lijoi et al [119] attention is focused on a spacial case of NRMI, namely the normalized inverse Gaussian process: the quantities relevan for its implementation in the context of mixture models are derived and it is shown that such prior exhibits an interesting and useful clustering behaviour, quite different from that of the Dirichlet process. In Lijoi et al. [120] the normalized inverse Gaussian process is then embedded in a larger subclass of NRMIs, namely the normalized generalized Gamma process, thus allowing for an additional parameter which greatly influences the clustering structure. In particular, the Dirichlet process is shown to be a limit process in distribution of the normalized generalized Gamma process. By close inspection of these tractable processes, one can observe that they all generate samples $X_1, \ldots, X_n$, for $n \geq 1$, which are characterized by a system of predictive distributions of the type

$$\mathbb{P}(X_{n+1} \in \cdot | X_1, \ldots, X_n) = g_0(n,k)\alpha_0(\cdot) + g_1(n,k) \sum_{j=1}^{k} (n_j - \sigma)\delta_{X_j^*}(\cdot), \tag{5.1.1}$$

where $\sigma \in [0,1)$ and for some weights g_0 and g_1. An almost surely discrete r.p.m. generating a sample as the above is termed Gibbs-type r.p.m. (see Chapter 6 for a more formal definition of the class of Gibbs-type r.p.m.s). The class of Gibbs-type r.p.m.s has been recently introduced and studied by Gnedin and Pitman [74], where also a characterization of its members is provided: indeed, Gibbs-type r.p.m. are Dirichlet process mixtures when $\sigma = 0$ and Poisson-Kingman models based on the stable subordinators when $\sigma \in (0,1)$ (see Gnedin and Pitman [74], Theorem 12). Further investigations related to Bayesian nonparametrics can be found in Ho et. al. [86] and Lijoi et al. [124]. It is to be noted that the weights g_0 and g_1 in (5.1.1) depend on the distinct observed species k but not on their frequencies $n_1, \ldots, n_k$, whose conveyed information can be incorporated into the parameters of the model. In principle one would like priors which lead to richer predictive structures, in which the probability of sampling a new species depends explicitly on both k and $n_1, \ldots, n_k$. However, by dropping the Gibbs structure assumption, serious issues of mathematical tractability arise.

In this chapter, moving from the definition of NRMI on $\mathbb{X}$ and from a recent paper by Von Renesse et al. [186] we define a new class of NRMIs, the so-called NRMI with logarithmic singularity. In particular, the class of NRMIs with logarithmic singularity includes as particular case the Dirichlet process and on the other hand it does not include the normalized generalized Gamma process. An interesting example of NRMI with logarithmic singularity has been recently introduced in by Regazzini et al. [165] and further investigated by Lijoi et al. [118], the so-called generalized Dirichlet process. In particular, the generalized Dirichlet process was originally introduced in Regazzini et al. [165] by the normalization of suitably time-changed independent increment process on $\mathbb{R}$ characterized by the Lévy measure

$$\nu(dv) = \frac{(1-\mathrm{e}^{-v\gamma})\mathrm{e}^{-v}}{v(1-\mathrm{e}^{-v})}dv \qquad v \geq 0 \tag{5.1.2}$$

with $\gamma > 0$. It can be easily checked that under the constraint $\gamma \in \mathbb{N}$, the generalized Dirichlet process corresponds to a r.p.m. obtained by the normalization of suitably time-changed superposition of independent Gamma processes on $\mathbb{R}$ with increasing integer-valued scale parameter and gives rise to a system of predictive distributions of the type

$$\mathbb{P}(X_{n+1} \in \cdot | X_1, \dots, X_n) = w_0(n,k,\mathbf{n})\alpha_0(\,\cdot\,) + \sum_{j=1}^{k} n_j w_j(n,k,\mathbf{n})\delta_{X_j^*}(\cdot) \tag{5.1.3}$$

where the weights $w_0(n,k,\mathbf{n})$ and $w_i(n,k,\mathbf{n})$, for $j = 1,\dots,k$ now explicitly depend on $\mathbf{n} := (n_1,\dots,n_k)$ thus conveying the additional information provided by the frequencies $\mathbf{n}$ directly into the prediction mechanism. To our knowledge, the generalized Dirichlet process represents the first example in the literature of almost surely discrete r.p.m. which is not of Gibbs-type and still leads to a closed form predictive structure. In this chapter, using the characterization of the generalized Dirichlet process in terms of normalized superposition of independent Gamma processes, we consider a simple way to extend the class of generalized Dirichlet processes. In particular, our aim is to define a more flexible class of r.p.m.s without Gibbs structure which mantains the same mathematical tractability of generalized Dirichlet process.

Equation (5.1.2) corresponds to the Lévy measure of the negative logarithm transform of a Beta distribution function with parameter $(1,\gamma)$ which is an infinite divisible distribution function belonging to the class of generalized convolutions of mixture of Exponential distributions introduced by Bondesson [12]). In particular, it is known that the negative logarithm transform of a Beta distribution function with parameter $(1,\gamma)$ is characterized into the class of generalized convolutions of mixture of Exponential distributions by a measure Q_γ on $(\mathbb{R}^+,\mathscr{R}^+)$, the so-called Thorin measure, defined by

$$Q_\gamma(dt) = \sum_{n\geq 1} \mathbb{1}_{\{n,n+\gamma\}}(dt). \tag{5.1.4}$$

In this chapter, using the characterization of the generalized Dirichlet process in terms of normalized superposition of independent Gamma processes, we consider a simple way to extend the class the class of generalized Dirichlet processes. In particular, our aim is to define a more flexible example of NRMI with logarithmic singularity which mantains the same mathematical tractability of the generalized Dirichlet process. Using the definition and the properties of the the measure Q_γ on $(\mathbb{R}^+, \mathscr{R}^+)$ we define a more general class of measures on $(\mathbb{R}^+, \mathscr{R}^+)$. In particular, given $\gamma > 0$ and two further parameters θ and β such that $0 \leq \theta \leq \gamma$ and $\gamma - \theta - \beta > 0$, we define a measure $Q_{(\gamma,\beta,\theta)}$ on $(\mathbb{R}^+, \mathscr{R}^+)$ by

$$Q_{(\gamma,\beta,\theta)}(dt) = \sum_{n\geq 1}(\mathbb{1}_{\{n,n+\gamma-\theta\}}(t) + \mathbb{1}_{\{n+\gamma-\theta-\beta,n+\gamma-\beta\}}(dt)). \tag{5.1.5}$$

We provide a detailed analysis for the measure $Q_{(\gamma,\beta,\theta)}$ and we show that it preserves all the properties of the measure Q_γ. In particular, we show that $Q_{(\gamma,\beta,\theta)}$ is still a Thorin measure which characterizes into the class of generalized convolutions of mixture of Exponential distributions a new infinite divisible distribution function on $\mathbb{R}^+$, the so-called Gauss-Exponential distribution function, having Lévy measure

$$\nu(dv) = \frac{\mathrm{e}^{-v}}{v(1-\mathrm{e}^{-v})}\left(1 + \frac{1 - \mathrm{e}^{-v\beta} - \mathrm{e}^{-v\theta}}{\mathrm{e}^{v(\gamma-\theta-\beta)}}\right) dv \qquad v \geq 0. \tag{5.1.6}$$

Thus, using the Lévy measure (5.1.6) we provide some developments of the generalized Dirichlet process. In particular, we define a new NRMI on $\mathbb{X}$ characterized by the Lévy measure (5.1.6) and we show that it represents a further example of NRMI with logarithmic singularity which makes more flexible the generalized Dirichlet process. By considering the particular choice of the parameter $\gamma \in \mathbb{N}$ and $\theta \in \mathbb{N}_0$ in (5.1.6) we then provide a comprehensive treatment of this new NRMI with logarithmic singularity in terms of its finite dimensional distributions, moments, predictive distributions and posterior distributions. As for the Gibbs-type r.p.m., a natural area of application of this new class of r.p.m. is represented by species sampling problems and, in particular, prediction problems in genomics (see Chapter 6 for a formal introduction to the species sampling problems and related problems in genomics).

The Lévy measure (5.1.2) first appeared in Bayesian nonparametric in Ferguson [62] where it was used to prove some properties related to the Dirichlet process as a neutral to the right r.p.m. In the same context of neutral to right r.p.m.s, the Lévy measure (5.1.2) was used in Walker and Muliere [194] to prove the existence of a neutral to the right r.p.m., the beta-Stacy process. These observations suggested to investigate the possibility of defining a generalization of the beta-Stacy process via superposition of beta-Stacy processes.

The chapter is structured as follows. In Section 5.2, we introduce the class of NRMIs on a general Polish space and we remind the posterior analysis recently developed by James et al. [100]. In Section 5.3 we define the subclass of NRMIs with logarithmic singularity and we show that some NRMIs known in the literature belong to this class. In Section 5.4 we provide some developments of the generalized Dirichlet process while in Secion 5.5 we define a generalization of the beta-Stacy process via superposition of beta-Stacy processes.

5.2 Normalized random measures with independent increments (NRMIs)

This section review the class of NRMIs. In particular, our aim is to introduce the class of NRMIs on $\mathbb{X}$ and to remind some results recently obtained by James et al. [100] related to the posterior analysis.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be some probability space. Having set $\mathbb{S} := \mathbb{R}^+ \times \mathbb{X}$, let $\mathscr{S}$ be the smallest σ-field containing all open sest of $\mathbb{S}$ and introduce a mapping $\tilde{N}$ from Ω to the set of non-negative counting measures on $(\mathbb{S}, \mathscr{S})$, i.e. $\tilde{N}(\omega, C) \in \mathbb{N}_0 \cup \{\infty\}$ for any $\omega \in \Omega$ and $C \in \mathscr{S}$. Assume that $\omega \mapsto \tilde{N}(\omega; C)$ is $\mathscr{F}/\mathscr{B}(\mathbb{N}_0 \cup \{\infty\})$-measurable for any $C \in \mathscr{S}$ where $\mathscr{B}(\mathbb{N}_0 \cup \{\infty\})$ is the smallest σ-field containing all open sest of $\mathbb{N}_0 \cup \{\infty\}$. We denote by ν a measure on $(\mathbb{S}, \mathscr{S})$ and by $\tilde{N}$ a Poisson random measure with Poisson intensity measure ν, i.e.

i) for any $C \in \mathscr{S}$ such that $\nu(C) = \mathbb{E}[\tilde{N}(C)] < +\infty$

$$\mathbb{P}(\tilde{N}(C) = k) = \frac{\mathrm{e}^{-\nu(C)}(\nu(C))^k}{k!} \mathbb{1}_{\mathbb{N}_0}(k)$$

ii) for any finite collection of disjoint sets, $A_1, \ldots, A_k$ in $\mathscr{S}$, the r.v.s $\tilde{N}(A_1), \ldots, \tilde{N}(A_k)$ are mutually independent

Moreover, the measure ν must satisfy the following conditions

$$\int_{(0,1)} s\nu(ds, \mathbb{X}) < +\infty \qquad \nu([1, +\infty)) < +\infty.$$

See Daley and Vere-Jones [22] for an exhaustive account on the theory of Poisson random measures.

Let $\mathcal{M}_{\mathbb{X}}$ is the space of boundedly measures on $(\mathbb{X}, \mathscr{X})$ endowed with its σ-field $\mathscr{M}_{\mathbb{X}}$ generated by the weak convergence topology $\mathcal{W}$ which makes $(\mathcal{M}_{\mathbb{X}}, \mathcal{W})$ a Polish space. Let $\tilde{\mu}$ be a random element defined on $(\Omega, \mathscr{F}, \mathbb{P})$ and with values in $(\mathcal{M}_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$ which can be represented as a linear functional of the Poisson random measure $\tilde{N}$, with Poisson intensity measure ν, as follows

$$\tilde{\mu}(B) = \int_{\mathbb{R}^+ \times B} s\tilde{N}(ds, dx) \qquad \forall B \in \mathscr{X}. \tag{5.2.1}$$

It can be easily seen from the properties of $\tilde{N}$ that $\tilde{\mu}$ is, in the terminology of Kingman [108], a completely random measure (CRM) on $\mathbb{X}$, i.e.

i) $\tilde{\mu}(\emptyset) = 0$ a.s.-$\mathbb{P}$;

ii) for any collection of disjoint sets in $\mathscr{X}$, $A_1, A_2, \ldots$, the r.v.s $\tilde{\mu}(A_1), \tilde{\mu}(A_2), \ldots$ are mutually independent and $\tilde{\mu}(\cup_{j\geq 1} A_j) = \sum_{j\geq 1} \tilde{\mu}(A_j)$ holds true a.s.-$\mathbb{P}$.

Let, now, $\mathscr{H}_\nu$ be the space of functions $h : \mathbb{X} \to \mathbb{R}^+$ such that $\int_{\mathbb{S}}(1 - \mathrm{e}^{-sh(x)})\nu(ds, dx) < +\infty$. Then, $\tilde{\mu}$ is uniquely characterized by its Laplace functional which, for any $h \in \mathscr{H}_\nu$, is given by

$$\mathbb{E}[\mathrm{e}^{-\int_{\mathbb{X}} h(x)\tilde{\mu}(dx)}] = \mathrm{e}^{-\int_{\mathbb{S}}(1-\mathrm{e}^{-sh(x)})\nu(ds,dx)}. \tag{5.2.2}$$

For a proof of such representation, see Theorem 2 in Kingman [108]. For details and further references on CRMs see Kingman [114].

It is apparent that both the Poisson random measure $\tilde{N}$ and the completely random measure $\tilde{\mu}$ are identified by the corresponding Poisson intensity measure ν. This suggests a simple and useful distinction of the random measures we deal with according to the decomposition of ν. Letting H be a non-atomic σ-finite measure on $\mathbb{X}$, we have

i) if $\nu(ds, dx) = \rho(ds)H(dx)$, for some measure ρ on $\mathbb{R}^+$, we say that the corresponding $\tilde{N}$ and $\tilde{\mu}$ are homogeneous;

ii) if $\nu(ds, dx) = \rho(ds|x)H(dx)$, where $\rho : \mathscr{R}^+ \times \mathbb{X} \to \mathbb{R}^+$ is a kernel i.e. $x \mapsto \rho(C|x)$ is $\mathscr{X}$-measurable for any $C \in \mathscr{R}^+$ and $\rho(\cdot|x)$ is a σ-finite measure on $\mathscr{R}^+$ for any $x \in \mathbb{X}$, we say that the corresponding $\tilde{N}$ and $\tilde{\mu}$ are non-homogeneous.

In this framework ν always admits a disintegration as in ii); this follows e.g. from Theorem 15.3.3 in Kellenberg [104]. In the sequel we suppose that H is representable as $H = a\alpha_0$ where α_0 is a probability distribution on $\mathbb{X}$

Since the aim is to define r.p.m.s by means of normalization of completely random measures, the total mass $T := \tilde{\mu}(\mathbb{X})$ needs to be finite and positive, almost surely. This happens if $\nu(\mathbb{S}) = +\infty$ and the Laplace exponent

$$\Psi(\lambda) := \int_{\mathbb{S}}(1 - \mathrm{e}^{-\lambda s})\nu(ds, dx)$$

is finite for any positive λ. A proof of this fact can be found, e.g., in Regazzini et al. [165] p. 563 and Proposition 1, respectively. Where these conditions hold true, a NRMI on $\mathbb{X}$ is given by

$$\tilde{P}(\cdot) = \frac{\tilde{\mu}(\cdot)}{T}. \tag{5.2.3}$$

Note that, when $\mathbb{X} = \mathbb{R}$, this definition coincides with the one given in Regazzini et al. [165] in terms of increasing additive processes. Indeed, it is worth remarking that an increasing additive process can always be seen as the càdlàg distribution function induced by a completely random measure on $\mathbb{R}$. Moreover, as shown in James [97], NRMI select almost surely discrete distributions. In order to avoid some technical difficulties, we assume T to be a r.v. whose distribution is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ and denote its density as f_T (see Section 3 of Pitman [156] for further details).

It is worth noting that some priors that are used in Bayesian nonparametric inference can be defined as in (5.2.3). For instance, consider the Dirichlet process with parameter measure $H = a\alpha_0$. Then, as already noted by Ferguson [61], such a prior can be recovered by considering a Gamma random measure. Other examples are, e.g., the normalized stable process introduced by Kingman [109], the normalized inverse-Gaussian process introduced by Lijoi et al. [119], the normalized generalized Gamma process introduced by Lijoi et al. [120] (see also James [96]) and the generalized Dirichlet process introduced by Regazzini et al and further investigated by Lijoi et al. [118]. It is interesting to note that the normalized inverse Gaussian process and the normalized generalized Gamma process are derivable from a stable subordinator by a change of measure (see Pitman [156]).

Under the usual assumption of exchangeability of the observation process, James et al. [100] derived a representation for the posterior distribution of $\tilde{P}$ in terms of a mixture with respect to the distribution of a suitable latent variable. In particular, let $\{X_n, n \geq 1\}$ be a sequence of exchangeable r.v.s defined on $(\Omega, \mathscr{F}, \mathbb{P})$ and with values in $\mathbb{X}$ in such way, given $\tilde{P}$, the X_i's are i.i.d. with distribution $\tilde{P}$, i.e.

$$\mathbb{P}(X_1 \in C_1, \dots, X_n \in C_n | \tilde{P}) = \prod_{i=1}^{n} \tilde{P}(C_i). \tag{5.2.4}$$

It is clear that one can always represent $X_1, \dots, X_n$ as $(X_1^*, \dots, X_k^*, \pi)$, where $X_1^*, \dots, X_k^*$ denotes the distinct observations within the sample and π stands for a partition of $[n] := \{1, \dots, n\}$ of size k recording which observations within the sample are equal. The number of elements in the j-th set of the partition is indicated by n_j, fro $j = 1, \dots, k$, so that $\sum_{1 \leq j \leq k} n_j = n$.

We define a positive r.v. U_n as follows. Let Γ_n be a r.v. distributed according to a Gamma distribution function with scale parameter 1 and shape parameter n which is independent from the total mass T. Then, set $U_n = \Gamma_n / T$. It is immediate to show that, for any $n \geq 1$, the density function of U_n is given by

$$f_{U_n} = \frac{u^{n-1}}{\Gamma(n)} \int_{\mathbb{R}^+} t^n \mathrm{e}^{-ut} f_T(t) dt \tag{5.2.5}$$

where f_t is the density function of T.

Proposition 5.2.1. *(cfr. James et al. [100]) Let $\tilde{P}$ be a NRMI. Then, the conditional distribution of U_n, given $X_1, \ldots, X_n$, admits a density function coinciding with*

$$f_{U_n}^{(X_1,\ldots,X_n)}(u) \propto u^{n-1} \prod_{i=1}^{k} \tau_{n_i}(u|X_i^*) e^{-\Psi(u)} \tag{5.2.6}$$

where $\tau_{n_i}(u|X_i^*) = \int_{\mathbb{R}^+} s^{n_i} e^{-us} \rho(ds|X_i^*)$, for $i = 1, \ldots, k$.

In what follows, for any pair of random elements Z and W defined on $(\Omega, \mathscr{F}, \mathbb{P})$, we use the symbol $Z^{(W)}$ to denote a random element on $(\Omega, \mathscr{F}, \mathbb{P})$ whose distribution coincides with a regular conditional distibution of Z, given W. James et al. [100] provided the following result concerning a posterior characterization of the completely random measure itself.

Theorem 5.2.1. *(cfr. James et al. [100]) Let $\tilde{P}$ be a NRMI with Poisson intensity measure $\nu(ds, dx) = \rho(ds|x)H(dx)$. Then*

$$\tilde{\mu}^{(U_n,X_1,\ldots,X_n)} \stackrel{d}{=} \tilde{\mu}^{(U_n)} + \sum_{i=1}^{k} J_i^{(U_n,X_1,\ldots,X_n)} \delta_{X_i^*}$$

where

i) $\tilde{\mu}^{(U_n)}$ is a completely random measure with Poisson intensity measure

$$\nu^{(U_n)}(ds, dx) = e^{-U_n s} \rho(ds|x) H(dx)$$

ii) X_i^*, for $i = 1, \ldots, k$, are the fixed points of discontinuity and the r.v.s $J_i^{(U_n,X_1,\ldots,X_n)}$'s are the corresponding jumps whose density is proportional to $s^{n_i} e^{-U_n s} \rho(ds|X_i^*)$;

iii) $\tilde{\mu}^{(U_n)}$ and $J_i^{(U_n,X_1,\ldots,X_n)}$ for $i = 1, \ldots, k$ are independent.

Theorem 5.2.1 shows that, given some latent variable, a posteriori $\tilde{\mu}$ is still a completely random measure with fixed points of discontinuity corresponding to the locations of the observations. The previous result is essential for deriving the posterior distribution for the class of NRMIs. In the following, by posterior distriution of $\tilde{P}$, given U_n, we refer to the distribution of $\tilde{P}$ given the data $X_1, \ldots, X_n$ and U_n.

Theorem 5.2.2. *(cfr. James et al. [100]) If $\tilde{\mu}$ is a NRMI with Poisson intensity measure $\nu(ds, dx) = \rho(ds|x)H(dx)$, then the posterior distribution of $\tilde{P}$, given U_n, is again a NRMI (with fixed points of discontinuity). In particular, it coincides in distribution with the random measure*

$$w \frac{\tilde{\mu}^{(U_n)}}{T^{(U_n)}} + (1-w) \frac{\sum_{i=1}^{k} J_i^{(U_n,X_1,\ldots,X_n)} \delta_{Y_i}}{\sum_{i=1}^{k} J_i^{(U_n,X_1,\ldots,X_n)}}$$

where $T^{(U_n)} = \tilde{\mu}^{(U_n)}(\mathbb{X})$*,* $w = T^{(U_n)}(T^{(U_n)} + \sum_{i=1}^{k} J_i^{(U_n,X_1,\ldots,X_n)})^{-1}$*. The distribution of* $\tilde{\mu}^{(U_n)}$ *and* $J_i^{(U_n,X_1,\ldots,X_n)}$ *for* $i = 1,\ldots,k$ *and the distribution of* U_n*, given* $X_1,\ldots,X_n$*, are those specified in Theorem 5.2.1.*

Apart from the posterior distribution, a Bayesian can be also interested in a rule for predicting future values of the observation, given those already observed, and a sampling scheme for generating observations governed by a NRMI.

Proposition 5.2.2. *(cfr. James et al. [100]) Let* $\tilde{P}$ *be a NRMI with Poisson intensity measure* $\nu(ds, dx) = \rho(ds|x)H(dx)$*. Then the predictive distribution for* X_{n+1} *given* $X_1,\ldots,X_n$ *coincides with*

$$\mathbb{P}(X_{n+1} \in dx|X_1,\ldots,X_n) = w^{(n)}H(dx) + \frac{1}{n}\sum_{j=1}^{k} w_j^{(n)}\delta_{X_j^*}(dx) \qquad (5.2.7)$$

where, for $j = 1,\ldots,k$

$$w^{(n)} = \frac{1}{n}\int_{\mathbb{R}^+} u\tau_1(u|x) f_{U_n}^{(X_1,\ldots,X_n)}(u)du$$

and

$$w_j^{(n)} = \int_{\mathbb{R}^+} u\frac{\tau_{n_j+1}(u|X_j^*)}{\tau_{n_j}(u|X_j^*)} f_{U_n}^{(X_1,\ldots,X_n)}(u)du.$$

These predictive distribution have quite intuitive forms, since they consist of a linear combination of H and o a weighted version of the empirical distribution. Note that the prediction rule reduces to the one provided by Pitman [156] in the homogeneous case.

From the previous results on the posterior and the predictive distributions, it is apparent that the use of partitions is of great help. The same can be said when facing the issue of characterizing the marginal distribution of the vector of (exchangeable) observations $X_1,\ldots,X_n$, for any $n \geq 1$. The marginal distribution of $X_1,\ldots,X_n$ can be described in terms of the distribution of $(X_1^*,\ldots,X_k^*,\pi)$, where, as before, π is a partition of $[n]$ into $k \in \{n\}$ since, as was mentioned before, NRMI select discrete distributions on $(\mathbb{X}, \mathscr{X})$ with probability 1. Before describing the distribution of $X_1,\ldots,X_n$ we introduce the following quantity

$$\kappa_{n_j}(u) = \int_{\mathbb{X}} \tau_{n_j}(u|x)H(dx)$$

which is the cumulant of order n_j of the conditional distribution of the total mass T, given $U_n = u$.

Proposition 5.2.3. *(cfr. James et al. [100]) Let* $\tilde{P}$ *be a NRMI. Then the distribution of* $(X_1^*,\ldots,X_k^*,\pi)$ *coincides with*

$$\frac{1}{\Gamma(n)}\left(\int_{\mathbb{R}^+} u^{n-1}e^{-\Psi(u)}\left(\prod_{j=1}^{k}\tau_{n_j}(u|X_j^*)\right)du\right)\prod_{j=1}^{k} H(dX_i^*). \qquad (5.2.8)$$

Moreover, the marginal distribution of π yields the exchangeable partition probability function (EPPF) and it is given by

$$p_k^{(n)}(n_1,\dots,n_k) = \frac{1}{\Gamma(n)}\int_{\mathbb{R}^+} u^{n-1}e^{-\Psi(u)}\left(\prod_{j=1}^{k}\kappa_{n_j}(u)\right)du. \tag{5.2.9}$$

The EPPF given by (5.2.9) was first obtained by Pitman [156]. For a concrete use of the marginal distribution of the X_i's, we generally need a simpler description of the distribution of $X_1,\dots,X_n$ and of the corresponding EPPF. This can be achieved by working conditionally on the latent variable U_n. As for the EPPF, a tractable form we wish to obtain is of the kind

$$p_k^{(n)}(n_1,\dots,n_k) = V_{n,k}\prod_{i=1}^{k} W_{n_i} \tag{5.2.10}$$

where $V_{n,k}$ is a positive quantity not depending on the specific $(n_1,\dots,n_k)$ and each W_{n_i} is a positive number depending solely on the corresponding n_i. A random partition having such an EPPF is said to be of a Gibbs-type. See Pitman [157] for the notion of infinite and finite Gibbs partition. However it is worth recalling that the only infinite EPPF admitting such a representation are the EPPFs derived from a Dirichlet process and those derived from a stable law of index $\sigma \in (0,1)$ (see Pitman [157]). Among them, we mention the two parameter Poisson-Dirichlet process and the generalized Gamma class of processes.

By examining (5.2.8) an augmentation and an application of Bayes rule makes it apparent that, for fixed $u > 0$ and π

$$\mathbb{P}(X_i^* \in dy|U_n = u,\pi) = \frac{\tau_{n_i}(u|y)H(dy)}{\kappa_{n_i}(u)} =: H_{i,n}(dy|u) \tag{5.2.11}$$

for any $i = 1,\dots,k$. At this point we can provide a characterization of the distribution of X, conditional on U_n.

Proposition 5.2.4. *(cfr. James et al. [100]) Let $\tilde{P}$ be a NRMI. Conditional on U_n and on the partition π, the k distinct values $X_1^*,\dots,X_k^*$ among the X_i's are independent and the distribution of X_i^*'s is given by* (5.2.11)*, for any $i = 1,\dots,k$. Moreover, the conditional distribution of the random partition π, given $U_n = u$, coincides with*

$$p_k^{(n)}(n_1,\dots,n_\pi) = \frac{e^{-\Psi(u)}\prod_{i=1}^{k}\kappa_{n_i}(u)}{\int_{\mathbb{R}^+} t^n e^{-ut} f_T(t)dt}. \tag{5.2.12}$$

Hence, conditional on U_n, π is a finite Gibbs partition.

Note that in the homogeneous case the distinct observations are i.i.d. with common distribution α_0.

5.3 NRMIs with logarithmic singularity

In this section, moving from the definition of NRMI given in Section 5.2 and from a recent paper by Von Renesse et al. [186], our aim is to introduce the class of NRMIs with logarithmic singularity. In particular, we show how some homogeneous NRMIs known in the literature belong to the class of NRMIs with logarithmic singularity.

Let us start by defining a positive infinite divisible r.v. $\tilde{\xi}_1$ characterized by the Lévy measure $\nu(dv) = g(v)dv$ where $g : (0, +\infty) \to \mathbb{R}^+$ is any measurable function such that

i) $g > 0$ such that the following condition holds true

$$\int_1^{+\infty} g(v)dv < +\infty$$

ii) there exists $g_0 \geq 0$ and a measurable function $\zeta : [0, 1] \to \mathbb{R}$ such that the following conditions hold true

$$g(v) = \frac{g_0}{v} + \zeta(v) \qquad \forall v \in (0, 1]$$

and

$$\int_0^1 |\zeta(v)|dv < +\infty.$$

Relying on the infinite divisible r.v. $\tilde{\xi}_1$ characterized by the Lévy measure $\nu(dv) = g(v)dv$ where g is any measurable function satisfying i) and ii), we define a CRM $\tilde{\mu}$ on $\mathbb{X}$ by its Poisson intensity measure

$$\nu(ds, dx) = g(s)ds\alpha(dx) \tag{5.3.1}$$

where α is a finite measure on $(\mathbb{X}, \mathscr{X})$ with $a := \alpha(\mathbb{X}) > 0$. Note that, if $\mathbb{X} = \mathbb{R}^+$ and $\alpha(dx) = dx$, the corresponding subordinator is in the class of subordinators with logarithmic singularity deeply investigated in Von Renesse et al. [186]. We are now in a position to define the class of NRMIs with logarithmic singularity.

Definition 5.3.1. *Given the CRM identified by the Poisson intensity measure* (5.3.1)*, a NRMI with logarithmic singularity on* $\mathbb{X}$ *with parameter* g *is defined as*

$$\tilde{P}_g(\cdot) \stackrel{d}{=} \frac{\tilde{\mu}(\cdot)}{\tilde{\mu}(\mathbb{X})}.$$

By Definition 5.3.1, the class of NRMIs with logarithmic singularity is a subclass of NRMIs characterized by a particular class of Lévy measure $\nu(dv) = g(v)dv$ where g is any measurable function satisfying i) and ii). It can be easily checked that some homogeneous NRMIs known in the literature belong to the class of NRMIs with logarithmic singularity. In particular, if we consider a measurable function $g : (0, +\infty) \to \mathbb{R}^+$ defined by

$$s \mapsto \frac{e^{-s}}{s} \tag{5.3.2}$$

then the Poisson intensity measure in (5.3.1) reduces to the Poisson intensity measure of a Gamma CRM and, hence, $\tilde{P}_g$ becomes the Dirichlet process on $\mathbb{X}$ introduced by Ferguson [61]. Another remarkable example can be obtained by considering a measurable function $g : (0, +\infty) \to \mathbb{R}^+$ defined by

$$s \mapsto \frac{(1 - \mathrm{e}^{-\gamma s})\mathrm{e}^{-s}}{(1 - \mathrm{e}^{-s})s} \tag{5.3.3}$$

with $\gamma > 0$. Then, the Poisson intensity measure in (5.3.1) reduces to the Poisson intensity measure of a generalized Gamma CRM and $\tilde{P}_g$ becomes the generalized Dirichlet process on $\mathbb{X}$ with parameter γ originally introduced by Regazzini et al. [165] assuming $\mathbb{X} = \mathbb{R}$. We observe that, if $\gamma = 1$ in (5.3.3), then $\tilde{\mu}$ reduces to the Gamma CRM. Moreover, if and only if $\gamma \in \mathbb{N}$, then $\tilde{\mu}$ can be seen as arising from the superposition of γ independent Gamma CRMs with increasing integer-valued scale parameter and shape parameter α. In particular, $\tilde{\mu}(A)$, for some $A \in \mathscr{X}$, is then distributed as the convolution of γ independent r.v.s with parameters $(j, \alpha(A))$, for $j = 1, \dots, \gamma$, i.e.

$$\mathbb{E}[\mathrm{e}^{-\lambda\tilde{\mu}(A)}] = \prod_{j=1}^{\gamma} \left(\frac{j}{j+\lambda} \right)^{\alpha(A)} \qquad \lambda \geq 0.$$

A first treatment of the generalized Dirichlet process on $\mathbb{R}$ in this setup was provided by Lijoi et al. [118].

On the othere hand, the normalized stable process with index $\sigma \in (0, 1)$ introduced by Kingman [109] does not belong to the class of NRMIs with logarithmic singularity, the measurable function $g : (0, +\infty) \to \mathbb{R}^+$ is defined by

$$s \mapsto cs^{-1-\sigma} \tag{5.3.4}$$

where c is a constant. Consequently, it can be easily checked that the normalized inverse-Gaussian process introduced by Lijoi et al. [119] and the normalized generalized Gamma process introduced by Lijoi et al. [120] (see also James [96]) does not belong to the class of NRMIs with logarithmic singularity. In particular, the fact that the NRMIs with logarithmic singularity are not of Gibbs-type follows immediately from Gnedin and Pitman [74] and Lijoi et al. [124]: for $\sigma = 0$, the only Gibbs-type NRMI is the Dirichlet pocess, whereas for $\sigma > 0$ the only NRMIs of Gibbs-type are normalized generalized gamma processes.

We conclude this section by providing some observations on the Lévy measure $\nu(dv) = g(v)dv$ when g is specified by the measurable function (5.3.3). By this specification of the measurable function g, the Lévy measure $\nu(dv)$ corresponds to the Lévy measure of the negative logarithm transform of a Beta distribution function with parameter $(1, \gamma)$ which is a member of the class of generalized convolutions of mixtures of Exponential distribution functions introduced by Bondesson [11]. In particular, if and only if $\gamma \in \mathbb{N}$, the negative logarithm transform of a Beta distribution function with parameter $(1, \gamma)$ reduces to a distribution function belonging to the class of the generalized gamma convolutions due to Thorin [180]. See [101] for an interesting account on the connections between generalized Gamma convolutions and Bayesian nonparametrics.

The class of generalized Gamma convolutions is a large class of infinite divisible self-decomposable distribution functions introduced by Thorin [180] and used by Thorin [181] to prove the infinite divisibility of the log-normal distribution function. The class of generalized Gamma convolutions, which is often called the $\mathscr{T}$-class in honor of Thorin, is the smallest class of distribution functions on $\mathbb{R}^+$ that contains the Gamma distribution function and is closed with respect to convolution and weak limits. The $\mathscr{T}$-class is extensively studied in the book by Bondesson [12] where moreover some extended versions of the $\mathscr{T}$-class are presented. Among these extended versions of the $\mathscr{T}$-class, we focus on a particular one, the so-called class of generalized convolutions of mixtures of Exponential distributions, which is often called the $\mathscr{T}_2$-class.

The $\mathscr{T}_2$-class was first introduced by Bondesson[11] as an extension of the $\mathscr{T}$-class. For $s \geq 0$, the Laplace-Stieltjes transform ϕ for a distribution function in the $\mathscr{T}$-class satisfies $\phi(0) = 1$ and the relation

$$\frac{d\phi(\lambda)}{d\lambda}\frac{1}{\phi(\lambda)} = b + \int_{\mathbb{R}^+} \frac{1}{t+\lambda} U(dt)$$

where $b \geq 0$ and U a non-negative measure on $(\mathbb{R}^+, \mathscr{R}^+)$ such that the following conditions holds true: $\int_{[0,1)} |\log(t)| U(dt) < +\infty$ and $\int_{[1,+\infty)} t^{-1} U(dt) < +\infty$. From the absolute monotonicity of ϕ and $d\log(\phi(\lambda))/ds$, it follows that ϕ is the Laplace-Stieltjes transform of an infinite divisible distribution function The same conclusion holds for

$$\frac{d\phi(\lambda)}{d\lambda}\frac{1}{\phi(\lambda)} = b + \int_{\mathbb{R}^+} \frac{1}{(t+\lambda)^2} Q(dt). \tag{5.3.5}$$

for $b \geq 0$ and Q a non-negative measure on $(\mathbb{R}^+, \mathscr{R}^+)$ satisfies the condition $\int_{\mathbb{R}^+} 1/t(1+t)Q(dt)$. A distribution function on $\mathbb{R}^+$ with Laplace-Stieltjes transform ϕ satisfying equation (5.3.5) is said to belong to the $\mathscr{T}_2$-class.

Definition 5.3.2. *A distribution function belonging to the $\mathscr{T}_2$-class is a probability distribution on $(\mathbb{R}^+, \mathscr{R}^+)$ with Laplace-Stieltjes transform of the form*

$$\phi(\lambda) = e^{-b\lambda + \int_{\mathbb{R}^+}(1/t+\lambda - 1/t)Q(dt)} \qquad \lambda \geq 0$$

where $b \geq 0$ and the non-negative Thorin measure Q on $(\mathbb{R}^+, \mathscr{R}^+)$ satisfies the condition $\int_{\mathbb{R}^+} 1/t(1+t)Q(dt)$.

The $\mathscr{T}_2$-class is closed with respect to positive translation and change of scale, convolution and convolution roots. It is also closed with respect to weak limits. A distribution function belonging to the $\mathscr{T}_2$-class can be characterized in terms of the Lévy meaure and completely monotone functions, that is real-valued functions on $\mathbb{R}^+$ which are non-negative and have derivatives of all orders, alternating in sign.

Theorem 5.3.1. *(cfr. Bondesson [12]) A distribution function on $\mathbb{R}^+$ belongs to the $\mathscr{T}_2$-class if and only if it is an infinite divisible distribution function and the Lévy measure has a completely monotone density ℓ. In fact $\ell(y) = \int_{\mathbb{R}^+} e^{-ty} Q(dt)$.*

Since $\ell(y)$ is completely monotone when $y\ell(y)$ is, it follows that $\mathscr{T} \subset \mathscr{T}_2$. Another characterization of the $\mathscr{T}_2$-class is in terms of Pick functions (see Donoghue [31]).

Theorem 5.3.2. *(cfr. Bondesson [12]) A distribution function on $\mathbb{R}^+$ belongs to the $\mathscr{T}_2$-class if and only if, for $\lambda \leq 0$, its Laplace-Stieltjes transform ϕ is anlytic and zero-free in $\mathbb{C} \cap (\mathbb{R}^+)^c$ and*

$$\Im(\log(\phi(\lambda))) = \arg(\phi(\lambda)) \geq 0 \qquad \Im(\lambda) > 0$$

or

$$\Im\left(\frac{1}{\lambda}\log(\phi(\lambda))\right) \geq 0 \qquad \Im(\lambda) > 0.$$

A characterization of the $\mathscr{T}_2$-class is available in terms of the Thorin measure. This characterization provides more information about the relation between the $\mathscr{T}$-class and the $\mathscr{T}_2$-class and the relations between the class of mixture of Exponential distributions and the $\mathscr{T}_2$-class.

Theorem 5.3.3. *(cfr. Bondesson [12]) The $\mathscr{T}$-class is formed by those distributions in the $\mathscr{T}_2$-class for which the measure Q on $(\mathbb{R}^+, \mathscr{R}^+)$ has an increasing density $q(\cdot)$. In fact, if $U(dt)$ denotes the Thorin measure, then $q(t) = U(t) = \int_{(0,t]} U(dt)$.*

Theorem 5.3.4. *(cfr. Bondesson [12]) The class of mixture of exponential distribution is formed by those distributions in the $\mathscr{T}_2$-class for which the left extremity b is zero and Q is a measure on $(\mathbb{R}^+, \mathscr{R}^+)$ with density $q(t)$ such that $q(t) \leq 1$.*

In Chapter 9 and Chapter 10 on Bondesson [12] several examples of distribution functions belonging to the $\mathscr{T}_2$-class are provided. In particular, the compound Poisson distribution function, the non-central Chi-square distribution function, the negative logarithm transform of a Beta distribution function and the first passage time distribution function for random walks in continuous time. Further examples are related to the context of shot-noise processes.

In the next proposition we define a non-negative measure $Q_{(\gamma,\varrho,\beta,\theta)}$ on $(\mathbb{R}^+, \mathscr{R}^+)$ by means of the sum of two Thorin measures characterizing the Laplace-Stieltjes transform of the negative logarithm transform of the Beta distribution function for a specific choice of its parameter (see Bondesson [12]). In particular, we prove that $Q_{(\gamma,\varrho,\beta,\theta)}$ is still a Thorin measure and it characterizes a Laplace-Stieltjes transform having Lévy-Khintchine representation.

Proposition 5.3.1. *Let $Q_{(\gamma,\varrho,\beta,\theta)}$ be a positive measure on $(\mathbb{R}^+,\mathscr{R}^+)$ defined by*

$$Q_{(\gamma,\varrho\beta,\theta)}(dt)=\sum_{n\geq 1}(\mathbb{1}_{\{n-1+\varrho,n-1+\gamma+\varrho-\theta\}}(dt)+\mathbb{1}_{\{n-1+\gamma+\varrho-\theta-\beta,n-1+\gamma+\varrho-\beta\}}(dt))$$

with $\gamma>0$, $\varrho>0$, $0\leq\theta\leq\gamma$ and $\gamma-\theta-\beta>0$. Then $Q_{(\gamma,\varrho,\beta,\theta)}$ is the Thorin measure characterizing the Laplace-Stieltjes transform

$$\phi(\lambda)=\frac{\Gamma(\gamma+\varrho-\theta)\Gamma(\gamma+\varrho-\beta)\Gamma(\varrho+\lambda)\Gamma(\gamma+\varrho-\lambda-\theta-\beta)}{\Gamma(\varrho)\Gamma(\gamma+\varrho-\theta-\beta)\Gamma(\gamma-\theta+\varrho-\lambda)\Gamma(\gamma-\beta+\varrho-\lambda)} \tag{5.3.6}$$

having Lévy-Khintchine representation

$$\log(\phi(\lambda))=\int_{\mathbb{R}^+}(e^{-\lambda v}-1)\nu(dv)\qquad \lambda\geq 0, v\geq 0 \tag{5.3.7}$$

where

$$\nu(dv)=\frac{e^{-v\varrho}}{v(1-e^{-v})}\left(1+\frac{1-e^{-v\beta}-e^{-v\theta}}{e^{v(\gamma-\theta-\beta)}}\right)dv. \tag{5.3.8}$$

Proof. First of all we prove that $Q_{(\gamma,\varrho,\beta,\theta)}$ is the Thorin measure charachterizing a particular Laplace-Stieltjes transform. It can be easly cheked that $\int_{\mathbb{R}^+}1/t(1+t)Q(dt)<+\infty$. From equation (5.3.5), setting $b=0$ and using the series representation of the first derivative of the Digamma function, i.e. $\psi(x)=d/dx\log(\Gamma(x))=-C+\sum_{n\geq 1}(1/n-1/n-1-x)$, where C is the Eulero's constant, we have for $\lambda\geq 0$

$$\begin{aligned}\frac{d\phi(\lambda)}{d\lambda}\frac{1}{\phi(\lambda)}&=\psi(\varrho+\lambda)-\psi(\gamma+\varrho-\theta+\lambda)+\psi(\gamma+\varrho-\theta-\beta+\lambda)-\psi(\gamma+\varrho-\beta+\lambda)\\&=\sum_{n\geq 1}\left(\int_{n-1+\varrho}^{n-1+\gamma+\varrho-\theta}\frac{1}{(t+\lambda)^2}dt+\int_{n-1+\gamma+\varrho-\theta-\beta}^{n-1+\gamma+\varrho-\beta}\frac{1}{(t+\lambda)^2}dt\right)\\&=\int_{\mathbb{R}^+}\frac{1}{(t+\lambda)^2}(Q^{(1)}(dt)+Q^{(2)}(dt))\end{aligned}$$

where

$$Q^{(1)}(dt):=\sum_{n\geq 1}\mathbb{1}_{\{n-1+\varrho,n-1+\gamma+\varrho-\theta\}}(dt)$$

and

$$Q^{(2)}(dt):=\sum_{n\geq 1}\mathbb{1}_{\{n-1+\gamma+\varrho-\theta-\beta,n-1+\gamma+\varrho-\beta\}}(dt).$$

Under the condition $\gamma>0$, $\varrho>0$, $0\leq\theta\leq\gamma$ and $\gamma-\theta-\beta>0$, it can be easly checked that ϕ is a Laplace-Stieltjes transform, then we proved that $Q_{(\gamma,\varrho,\beta,\theta)}$ in is a

Thorin measure characterizing the Laplace-Stieltjes transform ϕ. We now prove the Lévy-Khintchine representation for ϕ. In particular, using the ralations for the Gamma function $\Gamma(x) = (x-1)\Gamma(x)$ and $\Gamma(x) = x^{-1}\Gamma(x+1)$ and the approximation given by Stirling formula $\Gamma(x) \sim (2\pi x)^{1/2}(x/e)^x$ to approximate the Gamma function, we have the following

$$\begin{aligned}
\phi(\lambda) &= \prod_{i\geq 0} \frac{(\varrho+i)(\gamma+\varrho-\theta-\beta+i)(\gamma-\theta+\varrho+\lambda+i)(\gamma-\beta+\varrho+\lambda+i)}{(\gamma+\varrho-\theta+i)(\gamma+\varrho-\beta+i)(\varrho+\lambda+i)(\gamma+\varrho+\lambda-\theta-\beta+i)} \\
&= \prod_{i\geq 0} \exp\left\{ \int_0^{+\infty} \frac{(\mathrm{e}^{-\lambda v}-1)}{v} (\mathrm{e}^{-v(\varrho+i)}(1+\mathrm{e}^{-v(\gamma-\theta-\beta)}-\mathrm{e}^{-v(\gamma-\theta)}-\mathrm{e}^{-v(\gamma-\beta)}))dv \right\} \\
&= \exp\left\{ \int_0^{+\infty} \frac{(\mathrm{e}^{-\lambda v}-1)}{v} \sum_{i\geq 0} \mathrm{e}^{-v(\varrho+i)}(1+\mathrm{e}^{-v(\gamma-\theta-\beta)}-\mathrm{e}^{-v(\gamma-\theta)}-\mathrm{e}^{-v(\gamma-\beta)})dv \right\} \\
&= \exp\left\{ \int_0^{+\infty} \frac{(\mathrm{e}^{-\lambda v}-1)\mathrm{e}^{-v\varrho}}{v(1-\mathrm{e}^{-v})} \left(1+\frac{1-\mathrm{e}^{-v\beta}-\mathrm{e}^{-v\theta}}{\mathrm{e}^{v(\gamma-\theta-\beta)}}\right) dv \right\}.
\end{aligned}$$

□

We provide a brief discuss of the Proposition 5.3.1. The measure $Q_{(\gamma,\varrho,\beta,\theta)}$ defined in (5.3.5) is the sum of two known Thorin measure characterizing the Laplace-Stieltjes trasform of the negative logarithm transform of the Beta distribution function for a specific choice of its parameter (see Bondesson [12]). The first one

$$Q^{(1)}(dt) = \sum_{n\geq 1} \mathbb{1}_{\{n-1+\varrho, n-1+\gamma+\varrho-\theta\}}(dt)$$

is the Thorin measure characterizing the Laplace-Stieltjes transform

$$\phi^{(1)}(\lambda) = \frac{\Gamma(\gamma+\varrho-\theta)\Gamma(\varrho+\lambda)}{\Gamma(\varrho)\Gamma(\gamma-\theta+\varrho+\lambda)}$$

of the distribution function of the negative logarithm transform of a Beta distribution function with parameter $(\varrho, \gamma-\theta)$, i.e. the r.v. $Y^{(1)} = -\log(1-V^{(1)})$ where $V^{(1)}$ is a r.v. distributed according to a Beta distribution function with parameter $(\gamma-\theta, \varrho)$. The second one, corresponding to

$$Q^{(2)}(dt) = \sum_{n\geq 1} \mathbb{1}_{\{n-1+\gamma+\varrho-\theta-\beta, n-1+\gamma+\varrho-\beta\}}(dt)$$

is the Thorin measure characterizing the Laplace-Stieltjes transform

$$\phi^{(2)}(\lambda) = \frac{\Gamma(\gamma+\varrho-\beta)\Gamma(\gamma+\varrho+\lambda-\theta-\beta)}{\Gamma(\gamma+\varrho-\theta-\beta)\Gamma(\gamma-\beta+\varrho+\lambda)}$$

of the distribution function of the negative logarithm transform of a Beta distribution function with parameter $(\gamma+\varrho-\theta-\beta, \theta)$, i.e. the r.v. $Y^{(2)} = -\log(1-V^{(2)})$ where $V^{(2)}$ is a r.v. distributed according to a Beta distribution function with parameter $(\theta, \gamma+\varrho-\theta-\beta)$. In particular, by mean of the summation of $Q^{(1)}$ and $Q^{(2)}$, we obtain the Thorin measure $Q_{(\gamma,\varrho,\beta,\theta)}$ which characterizes the Laplace-Stieltjes transform ϕ in (5.3.6). As we can see ϕ in (5.3.6) is the product of the Laplace-Stieltjes transforms $\phi^{(1)}$ and $\phi^{(2)}$, i.e. ϕ is the Laplace-Stieltjes transform of the convolution of the distribution functions of $Y^{(1)}$ and $Y^{(2)}$ where $Y^{(1)}$ is independent of $Y^{(2)}$. Reminding that the $\mathscr{T}_2$-class is convolution closed the last observation could be another way to see that $Q_{(\gamma,\varrho,\beta,\theta)}$ is a Thorin measure.

In the next corollary we provide some further characterization for the Thorin measure $Q_{(\gamma,\varrho,\beta,\theta)}$ introduced in Proposition 5.3.1. In particular, we prove a relation with the $\mathscr{T}$.

Corollary 5.3.1. *The following facts hold true*

i) ν *in* (5.3.8) *is a Lévy measure generalizing the Lévy measures* $g(v)dv$ *with* g *the measurable function defined by* (5.3.2)*;*

ii) ϕ *in* (5.3.6) *is the Laplace-Stieltjes transform of a mixture of Exponential distribution function if and only if* $\gamma \leq 1$*;*

iii) ϕ *in* (5.3.6) *is the Laplace-Stieltjes transform belonging to the* $\mathscr{T}$*-class if and only if* $\gamma \in \mathbb{N}$ *and* $\theta \in \mathbb{N}_0$.

Proof. As regard the point i), the proof follows from Proposition 5.3.1, especially from the measure ν identified by (5.3.8). The measure ν is a non-negative measure, then we need just to verify that

$$\int_{\mathbb{R}^+} (1 \wedge v)\nu(dv) < +\infty.$$

In particular, we can factorize the above condition as follows

$$\int_0^{+\infty} (1 \wedge v) \frac{\mathrm{e}^{-v\varrho}}{v(1-\mathrm{e}^{-v})} \left(1 + \frac{1 - \mathrm{e}^{-v\beta} - \mathrm{e}^{-v\theta}}{\mathrm{e}^{v(\gamma-\theta-\beta)}}\right) dv$$

$$= \int_0^{+\infty} (1 \wedge v) \frac{\mathrm{e}^{-v\varrho}}{v(1-\mathrm{e}^{-v})} ((1-\mathrm{e}^{-v(\gamma-\theta)}) + (1-\mathrm{e}^{-v(\gamma-\beta)}) - (1-\mathrm{e}^{-v(\gamma-\theta-\beta)}))dv$$

and we know that for any $a > 0$ and $b > 0$

$$\int_0^{+\infty} (1 \wedge v) \frac{\mathrm{e}^{-vb}}{v(1-\mathrm{e}^{-v})}(1-\mathrm{e}^{-va})dv < +\infty.$$

because $\mathrm{e}^{-vb}(1-\mathrm{e}^{-va})/v(1-\mathrm{e}^{-v})dv$ is a Lévy measure. It can be easly seen than the Lévy measure ν in (5.3.8) generalizes the Lévy measure $g(v)dv$ with g defined by (5.3.2) and (5.3.3), respectively. Actually we can obtain the Lévy measure $g(v)dv$ with g defined by (5.3.3), if we set in ν one of the following conditions: $\theta = 0$, $\beta = 0$ or $\beta = 0$ and $\theta = 0$. Moreover, we can obtain the Lévy measure $g(v)dv$ with g defined by (5.3.2), setting the further condition $\gamma = 1$. The point ii), given the equation (5.3.5), is a straightforward applications of Theorem 5.3.4. As regard the point iii), using the relation for the generalized factorials $\Gamma(a+n)/\Gamma(a) = (a)_{n\uparrow 1} = a(a+1)\cdots(a+n-1)$ (in particular, $(x)_{y\uparrow 1}$ stands for the Pochhammer symbol for the ascending factorial of x of order y (see Appendix A)) for any $a \in \mathbb{C}$ and $n \in \mathbb{N}_0$, it can be seen that if and only if $\gamma \in \mathbb{N}$ and $\theta \in \mathbb{N}_0$, ϕ is the Laplace-Stieltjes transform of a convolution of γ independent Exponential distribution function. Actually, if and only if $\gamma \in \mathbb{N}$ and $\theta \in \mathbb{N}_0$ we can write

$$\phi(\lambda) = \frac{(\varrho)_{(\gamma-\theta)\uparrow 1}(\gamma+\varrho-\theta-\beta)_{\theta\uparrow 1}}{(\varrho+\lambda)_{(\gamma-\theta)\uparrow 1}(\gamma+\varrho-\theta-\beta+\lambda)_{\theta\uparrow 1}} = \prod_{j=0}^{\gamma-1} \frac{j+\varrho-\beta\mathbb{1}_{\{j\geq\gamma-\theta\}}}{j+\varrho-\beta\mathbb{1}_{\{j\geq\gamma-\theta\}}+\lambda}$$

which is the convolution of γ independent Exponential distribution function, with parameters $(j+\varrho-\beta\mathbb{1}_{\{j\geq\gamma-\theta\}}+\varrho)$. □

As we proved, the Thorin measure $Q_{(\gamma,\varrho,\beta,\theta)}$ characterizes the Laplace-Stieltjes transform ϕ which is the Laplace-Stieltjes transform of the convolution of the distribution functions of the independent r.v.s $Y^{(1)}$ and $Y^{(2)}$ distributed according to a Beta distribution function with parameter $(\varrho, \gamma-\theta)$ and $(\gamma+\varrho-\theta-\beta, \theta)$, respectively. We now define a new distribution function, the so-called Gauss-Exponential distribution function corresponding to this convolution and we provide several properties of this new distribution function in terms of cumulative distribution function, moments and particular cases. The definition of the Gauss-Exponential distribution function is given in terms of the Gauss hypergeometric function ${}_2F_1$ (see Appendix C).

Definition 5.3.3. *A Gauss-Exponential distribution function with parameter* $(\gamma, \varrho, \beta, \theta)$ *is a distribution function having density function absolutley continuous with respect to Lebesgue measure on* $\mathbb{R}^+$

$$f(v) = \frac{\Gamma(\gamma+\varrho-\theta)\Gamma(\gamma+\varrho-\beta)}{\Gamma(\gamma)\Gamma(\varrho)\Gamma(\gamma+\varrho-\theta-\beta)}(1-e^{-v})^{\gamma-1}(e^{-v})^{\varrho}{}_2F_1(\theta,\beta;\gamma;1-e^{-v})\mathbb{1}_{\mathbb{R}^+}(v) \quad (5.3.9)$$

where $\gamma > 0$, $\varrho > 0$, $0 \leq \theta \leq \gamma$ *and* $\gamma - \theta - \beta > 0$.

Using Equation 7.527.3 in Gradshteyn and Ryzhik [77] it can be proved that (5.3.9) is a density function on $\mathbb{R}^+$. In particular, Figure 5.1 and Figure 5.2 show respectively the density functions and the corresponding numerical approximation of the cumulative distribution functions for the Gauss-Exponential distribution function for some fixed parameter γ and ϱ and variyng parameters θ and β.

It can be seen that the Gauss-Exponential distribution function includes as particular cases some known distribution functions

- the Exponential distribution function with parameter ϱ can be recovered from (5.3.9) by setting one of the following conditions: $\gamma = 1$ and $\theta = 0$, $\gamma = 1$ and $\beta = 0$ or $\gamma = 1$ and $\beta = 0$ and $\theta = 0$;
- the distribution function of the negative logarithm transform of a Beta distribution function with parameter (γ, ϱ) cab be recovered from (5.3.9) by setting one of the following conditions: $\theta = 0$, $\beta = 0$ or $\beta = 0$ and $\theta = 0$.

Alternatively we can recover the Exponential distribution function and the distribution function of the negative logarithm trasform of the Beta distribution function as the limit in distribution of the Gauss-Exponential distribution function for $\beta \to -\infty$. In particular, the distribution function of the negative logarithm trasform of the Beta distribution function with parameter $(\gamma - \theta, \varrho)$ can be recovered from the Gauss-Exponential distribution function with parameter taking the limit for $\beta \to -\infty$; the Exponential distribution function with parameter ϱ can be recovered from the Gauss-Exponential distribution function with parameter $(\gamma, \varrho, \beta, \gamma - 1)$ and taking the limit for $\beta \to -\infty$.

Proposition 5.3.2. *The Laplace-Stieltjes transform of a Gauss-Exponential distribution function with parameter* $(\gamma, \varrho, \beta, \theta)$ *correspond to* (5.3.6) *and*

$$\frac{d}{d\lambda}\phi(\lambda) = \phi(\lambda)(\psi(\varrho+\lambda)-\psi(\gamma+\varrho-\beta+\lambda)+\psi(\gamma+\varrho-\theta-\beta+\lambda)-\psi(\gamma+\varrho-\theta+\lambda)) \quad (5.3.10)$$

and for $r > 1$

$$\frac{d^r}{d\lambda^r}\phi(\lambda) = \phi(\lambda)\left(\frac{d/d\lambda\phi(\lambda)d^{r-1}/d\lambda^{r-1}\phi(\lambda)}{(\phi(\lambda)^2)} + \frac{d}{d\lambda}\left(\frac{d^{r-1}/d\lambda^{r-1}\phi(\lambda)}{\phi(\lambda)}\right)\right). \quad (5.3.11)$$

Proof. The constraints on the parameter of the Gauss-Exponential distribution function implies that $\varrho + s > \theta + \beta - \gamma$. Then, the Laplace-Sieltjes transform (5.3.6) can be obtained by a straightforward application of Equation 7.527.3 in Gradshteyn and Ryzhik [77]. Equation (5.3.10) and equation (5.3.11) can be obtained using the definition of Digamma function. □

Equation (5.3.10) and equation (5.3.11) can be used to derive the r-th moment for a r.v. distributed according to a Gauss-Exponential distribution function. In particular, given V is a r.v. distributed according to a Gauss-Exponential distribution function with parameter $(\gamma, \varrho, \beta, \theta)$, then

$$\mathbb{E}[V] = \psi(\gamma + \varrho - \beta) - \psi(\varrho) + \psi(\gamma + \varrho - \theta) - \psi(\gamma + \varrho - \theta - \beta)$$

and

$$Var(V) = \frac{d}{d\lambda}\left(\psi(\varrho + \lambda) - \psi(\gamma + \varrho - \beta + \lambda)\ \psi(\gamma + \varrho - \theta - \beta + \lambda) - \psi(\gamma + \varrho - \theta + \lambda)\right|_{\lambda=0}.$$

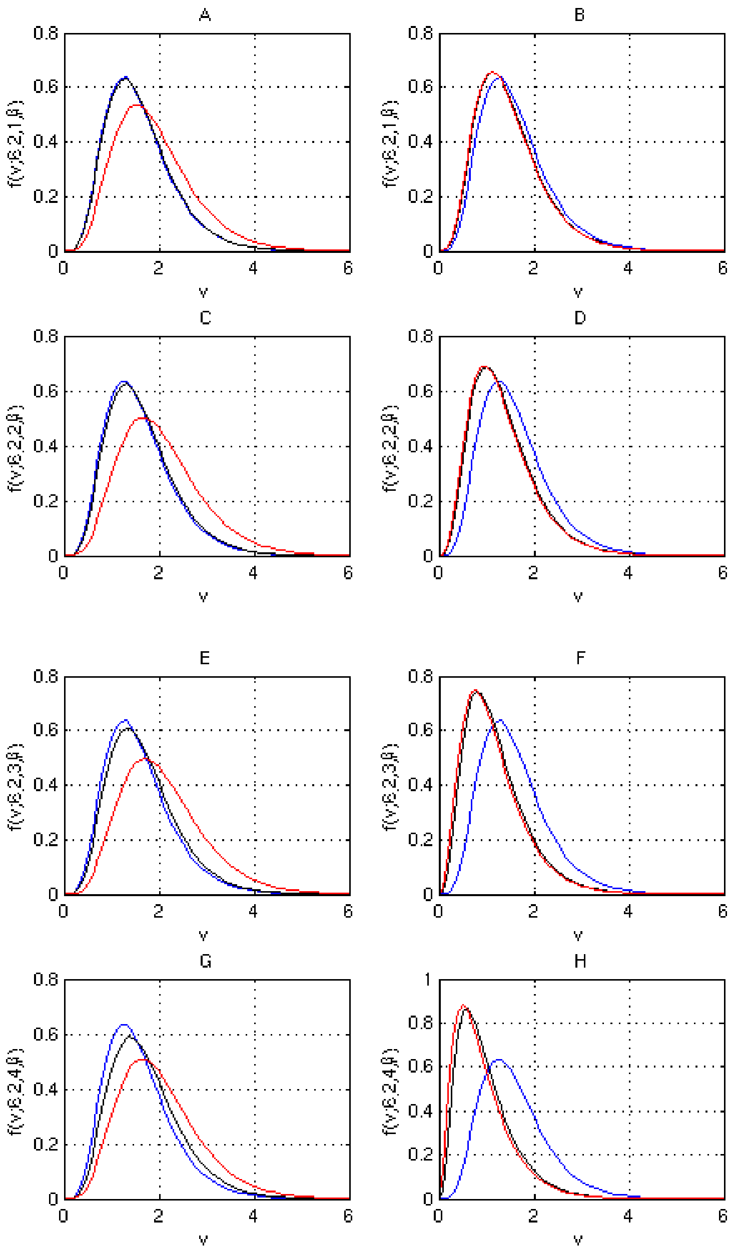

Figure 5.1: Gauss-Exponential density functions on $\mathbb{R}^+$ with $\gamma = 6$, $\varrho = 2$ and A) $\theta = 1$, $\beta = 0.8$ (red); $\theta = 1$, $\beta = 4.8$ (black) B) $\theta = 1$, $\beta = -20$ (red); $\theta = 1$, $\beta = -40$ (black) C) $\theta = 2$, $\beta = 0.8$ (red); $\theta = 2$, $\beta = 3.8$ (black) D) $\theta = 2$, $\beta = -20$ (red); $\theta = 2$, $\beta = -40$ (black) E) $\theta = 3$, $\beta = 0.8$ (red); $\theta = 3$, $\beta = 2.8$ (black) F) $\theta = 3$, $\beta = -20$ (red); $\theta = 3$, $\beta = -40$ (black) G) $\theta = 4$, $\beta = 0.8$ (red); $\theta = 4$, $\beta = 2.8$ (black) H) $\theta = 4$ $\beta = -20$ (red); $\theta = 4$, $\beta = -40$ (black).

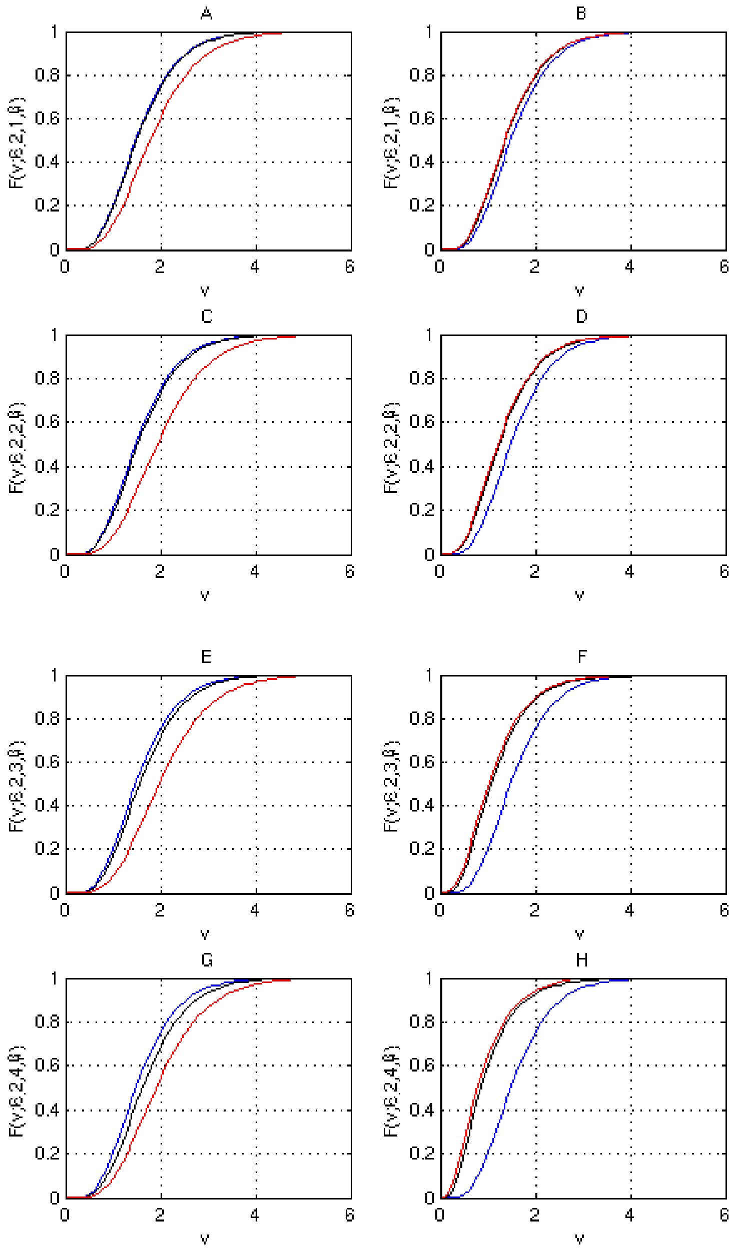

Figure 5.2: Gauss-Exponential cumulative distribution functions on $\mathbb{R}^+$ with $\gamma = 6$, $\varrho = 2$ and A) $\theta = 1$, $\beta = 0.8$ (red); $\theta = 1$, $\beta = 4.8$ (black) B) $\theta = 1$, $\beta = -20$ (red); $\theta = 1$, $\beta = -40$ (black) C) $\theta = 2$, $\beta = 0.8$ (red); $\theta = 2$, $\beta = 3.8$ (black) D) $\theta = 2$, $\beta = -20$ (red); $\theta = 2$, $\beta = -40$ (black) E) $\theta = 3$, $\beta = 0.8$ (red); $\theta = 3$, $\beta = 2.8$ (black) F) $\theta = 3$, $\beta = -20$ (red); $\theta = 3$, $\beta = -40$ (black) G) $\theta = 4$, $\beta = 0.8$ (red); $\theta = 4$, $\beta = 2.8$ (black) H) $\theta = 4$ $\beta = -20$ (red); $\theta = 4$, $\beta = -40$ (black).

Explicit formulae for the skewness coefficient and the kurtosis coefficient can be obtained using equation (5.3.10) and applying recursively Equation (5.3.11). Rather than providing an explicit formula for the skewness coefficient and the kurtosis coefficient, we provide in Figure 5.3 some graphical results which are useful to understand the behavior for the mean, variance, skewness and kurtosis coefficient of the Gauss-Exponential for some fixed parameter γ and ϱ and variyng parameter θ and β.

5.4 Some developments of a generalized Dirichlet process

In this section we use the results obtained in Section 5.3 for the Gauss-Exponential distribution function in order to provide some developments of the generalized Dirichlet process introduced by Regazzini et al. [165] and further investigated by Lijoi et al. [118].

Let us start by defining a r.v. $\tilde{\xi}_1$ distributed according to the Gauss-Exponential distribution function with parameter $(1, \varrho, \theta, \beta)$. Thus, $\tilde{\xi}_1$ is a positive infinite divisible r.v. characterized by the Lévy measure

$$\nu(dv) = \frac{\mathrm{e}^{-v}}{v(1-\mathrm{e}^{-v})}\left(1 + \frac{1-\mathrm{e}^{-v\beta}-\mathrm{e}^{-v\theta}}{\mathrm{e}^{v(\gamma-\theta-\beta)}}\right) \qquad v \geq 0$$

with $\varrho > 0$, $0 \leq \theta \leq \gamma$ and $\gamma - \theta - \beta > 0$. Relying on $\tilde{\xi}_1$, define now a CRM $\tilde{\mu}$ by its Poisson intensity measure

$$\nu(ds, dx) = \frac{\mathrm{e}^{-s}}{s(1-\mathrm{e}^{-s})}\left(1 + \frac{1-\mathrm{e}^{-s\beta}-\mathrm{e}^{-s\theta}}{\mathrm{e}^{s(\gamma-\theta-\beta)}}\right) ds\alpha(dx) \tag{5.4.1}$$

where α is a finite measure on $\mathbb{X}$ with $a := \alpha(\mathbb{X}) > 0$.

Definition 5.4.1. *Given the CRM $\tilde{\mu}$ on $\mathbb{X}$ identified by the Poisson intensity measure* (5.4.1)*, the generalized Dirichlet process on $\mathbb{X}$ with parameter (γ, β, θ) is defined as*

$$\tilde{P}_{(\gamma,\beta,\theta)}(\cdot) \stackrel{d}{=} \frac{\tilde{\mu}(\cdot)}{\tilde{\mu}(\mathbb{X})}.$$

It can be easily checked that the generalized Dirichlet process on $\mathbb{X}$ with parameter (γ, β, θ) is a NRMI with logarithmic singularity with parameter g where $g : (0, +\infty) \to \mathbb{R}^+$ is defined by $s \mapsto \mathrm{e}^{-s}/s(1-\mathrm{e}^{-s})(1 + (1-\mathrm{e}^{-s\beta}-\mathrm{e}^{-s\theta}/\mathrm{e}^{s(\gamma-\theta-\beta)}))$. In particular the generalized Dirichlet process on $\mathbb{X}$ with parameter γ can be recovered by setting one of the following conditions: $\theta = 0$, $\beta = 0$ or $\theta = 0$ and $\beta = 0$. If $\mathbb{X} = \mathbb{R}^+$ and $\alpha(dx) = dx$, the corresponding subordinator represents a special case of the class of subordinators with logarithmic singularity deeply investigated in Von Renesse et al.[186].

As proposed by Lijoi et al [118] for the generalized Dirichlet process with parameter γ with $\gamma \in \mathbb{N}$, also for the generalized Dirichlet process with parameter (γ, β, θ) we can

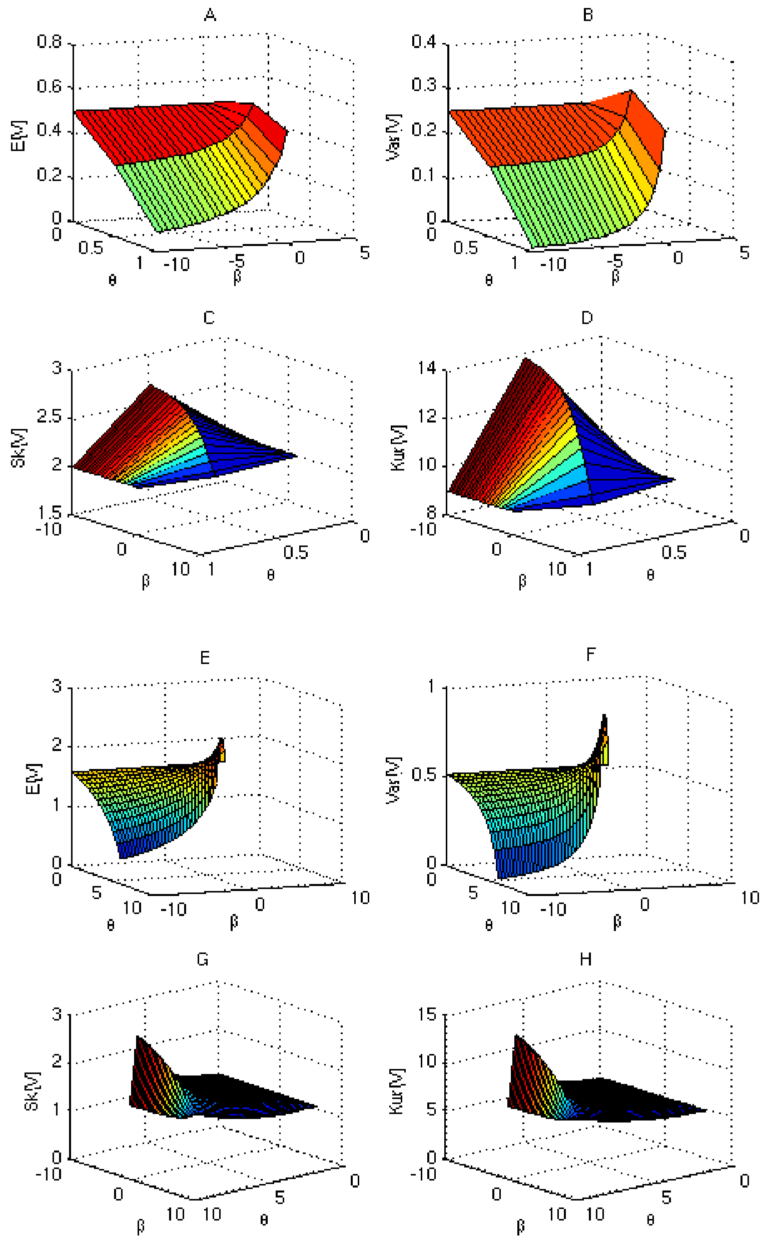

Figure 5.3: Figures A), B), C), D) respectively the mean, variance, skewness and kurtosis coefficient for the Gauss-Exponential distribution functions with $\gamma = 1$, $\varrho = 2$ and variyng θ and β. Figures E), F), G), H) respectively the mean, variance, skewness and kurtosis coefficient for the Gauss-Exponential distribution functions with $\gamma = 6$, $\varrho = 2$ and variyng θ and β.

restrict the attention to the particular case corresponding to $\gamma \in \mathbb{N}$ and $\theta \in \mathbb{N}_0$. Under these assumptions, for any $A \in \mathscr{X}$, the Laplace-Stieltjes transform of the distribution function of $\tilde{\mu}(A)$

$$\mathbb{E}[\mathrm{e}^{-\lambda\tilde{\mu}(A)}] = \prod_{j=1}^{\gamma}\left(\frac{j-\beta\mathbb{1}_{\{j>\gamma-\theta\}}}{j-\beta\mathbb{1}_{\{j>\gamma-\theta\}}+\lambda}\right)^{\alpha(A)} \qquad \lambda \geq 0. \tag{5.4.2}$$

Thus $\tilde{\mu}(A)$ is distributed as the convolution of two independent random elements. The first one is the convolution of $(\gamma-\theta)$ independent r.v.s distributed according to a Gamma distribution function with parameter $(j, \alpha(A))$ for $j = 1, \dots, (\gamma-\theta)$. The second one is the convolution of θ independent r.v.s distributed according to a Gamma distribution function with parameter $(\gamma-\theta-\beta+j, \alpha(A))$ for $j = 1, \dots, \theta$. Thus, distribution function of $\tilde{\mu}(A)$ is a member of $\mathscr{T}$ because is the convolution two members of $\mathscr{T}$.

In what follows for generalized Dirichlet process with parameter γ we mean the generalized Dirichlet process with parameter γ with $\gamma \in \mathbb{N}$ and for generalized Dirichlet process with parameter (γ, β, θ) we mean the generalized Dirichlet process with parameter (γ, β, θ) with $\gamma \in \mathbb{N}$ and $\theta \in \mathbb{N}_0$. From expression (5.4.2) we can clearly see the main difference between the generalized Dirichlet process investigated by Lijoi et al. [118] and the generalized Dirichlet process defined by Definition 5.4.1 under the assumpion $\gamma \in \mathbb{N}$ and $\theta \in \mathbb{N}_0$: the generalized Dirichlet process with parameter (γ, β, θ) is a r.p.m. on $\mathbb{X}$ obtained by normalization of superposed independent Gamma CRMs where the scale parameter is not necessarily increasing integer-valued. The scale parameter is increasing integer-valued for the first $\gamma-\theta$ Gamma CRMs superposed, then changes according to the parameter β.

By definition, the generalized Dirichlet process with parameter γ can be recovered setting one of the following condition: $\theta = 0$, $\beta = 0$ or $\theta = 0$ and $\beta = 0$. In particular, the Dirichlet process can be recoverd setting the further condition $\gamma = 1$. Another way to see the Dirichlet process and the generalized Dirichlet process with parameter γ as particular case of the generalized Dirichlet process with parameter γ, β, θ is to consider such processes as the limit process in distribution of the generalized Dirichlet process with parameter (γ, β, θ) for $\beta \to -\infty$. The generalized Dirichlet process with parameter $\gamma-\theta$ can be recovered from a generalized Dirichlet process with parameter (γ, β, θ) and taking the limit for $\beta \to -\infty$. In particular, the Dirichlet process can be recovered from the generalized Dirichlet process with parameter $(\gamma-1, \beta, \theta)$ and taking the limit for $\beta \to -\infty$.

5.4.1 Finite dimensional distributions

We now consider the finite dimensional distributions of the generalized Dirichlet process with parameter (γ, β, θ). As for the generalized Dirichlet process with parameter γ, the finite dimensional distributions can be derived in terms of Lauricella hypergeometric functions (see Appendix C).

First of all we establish some notation. For any vector $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, let $|x| := \sum_{1\leq i\leq n} x_i$ and for any measurable partition $B_1, \dots, B_n$ of $\mathbb{X}$ let $\alpha_i := \alpha(B_i)$ for $i = 1, \dots, n$. Finally, we set the following $(\gamma-1)$-dimensional vectors $\mathbf{1}_{\gamma-1} := (1, \dots, 1)$ and $J_{\gamma-1}^{(\theta,\beta)} := (1 + \beta(\mathbb{1}_{\{\gamma-1>\gamma-\theta\}} - \mathbb{1}_{\{\theta\neq 0\}}), \dots, \gamma - 1 + \beta(\mathbb{1}_{\{1>\gamma-\theta\}} - \mathbb{1}_{\{\theta\neq 0\}}))$. Thus, we have the following proposition.

Proposition 5.4.1. *Let $\tilde{P}_{(\gamma,\beta,\theta)}$ be a generalized Dirichlet process on $\mathbb{X}$ with parameter (γ, β, θ). For any measurable partition $B_1, \dots, B_n$ of $\mathbb{X}$, $(\tilde{P}_{(\gamma,\beta,\theta)}(B_1), \dots, \tilde{P}_{(\gamma,\beta,\theta)}(B_{n-1}))$ is a r.v. admitting density function with respect to Lebesgue measure on the simplex $\Delta^{(n-1)}$ given by*

$$\begin{aligned} &f_{(\tilde{P}_{(\gamma,\beta,\theta)}(B_1),\dots,\tilde{P}_{(\gamma,\beta,\theta)}(B_{n-1}))}(y_1, \dots, y_{n-1}) \qquad (5.4.3)\\ &\quad= \frac{\prod_{j=1}^{\gamma}(j - \beta\mathbb{1}_{\{j>\gamma-\theta\}})^a}{\prod_{i=1}^{n}\Gamma(\gamma\alpha_i)} \prod_{i=1}^{n-1} y_i^{\gamma\alpha_i - 1}\left(1 - \sum_{i=1}^{n-1} y_i\right)^{\gamma\alpha_n - 1} \int_0^{+\infty} e^{-y(\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}})} y^{\gamma a - 1}\\ &\quad\quad \times \prod_{i=1}^{n-1} \Phi_2^{(\gamma-1)}(\alpha_i \mathbf{1}_{\gamma-1}; \gamma\alpha_i; y y_i J_{\gamma-1}^{(\theta,\beta)}) \Phi_2^{(\gamma-1)}\left(\alpha_n \mathbf{1}_{\gamma-1}; \gamma\alpha_n; y\left(1 - \sum_{i=1}^{n-1} y_i\right) J_{\gamma-1}^{(\theta,\beta)}\right) dy\\ &\quad\quad \times \mathbb{1}_{\Delta^{(n-1)}}(y_1, \dots, y_{n-1}). \end{aligned}$$

In particular, the distribution of $\tilde{P}_{(\gamma,\beta,\theta)}(B)$, for any $B \in \mathscr{X}$ has density function with respect to the Lebesgue measure on $(0,1)$ given by

$$\begin{aligned} &f_{\tilde{P}_{(\gamma,\beta,\theta)}(B)}(y_1) \qquad (5.4.4)\\ &\quad= \frac{\prod_{j=1}^{\gamma}(j - \beta\mathbb{1}_{\{j>\gamma-\theta\}})^a}{\Gamma(\gamma\alpha(B))\Gamma(\gamma\alpha(B^c))} y_1^{\gamma\alpha(B)-1}(1 - y_1)^{\gamma\alpha(B^c)-1} \int_0^{+\infty} e^{-y(\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}})} y^{\gamma a - 1}\\ &\quad\quad \times \Phi_2^{(\gamma-1)}(\alpha(B)\mathbf{1}_{\gamma-1}; \gamma\alpha(B); y y_1 J_{\gamma-1}^{(\theta,\beta)}) \Phi_2^{(\gamma-1)}(\alpha(B^c)\mathbf{1}_{\gamma-1}; \gamma\alpha(B^c); y(1 - y_1) J_{\gamma-1}^{(\theta,\beta)}) dy\\ &\quad\quad \times \mathbb{1}_{\Delta^{(1)}}(y_1). \end{aligned}$$

Proof. Using the representation of the density function of the sum N of independent r.v.s distributed according to a Gamma distribution function in terms of the limiting form $\Phi_2^{(N)}$ of the second Lauricella hypergeometric function (see Chapter 7 in Exton [47]) we have for any $B \in \mathscr{X}$ the density function of $\tilde{\mu}(B)$

$$\begin{aligned} f_{\tilde{\mu}(B)}(x) = \frac{\prod_{j=1}^{\gamma}(j - \beta\mathbb{1}_{\{j>\gamma-\theta\}})^{\alpha(B)}}{\Gamma(\gamma\alpha(B))} &\mathrm{e}^{-x(\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}})} x^{\gamma\alpha(B)-1}\\ &\times \Phi_2^{(\gamma-1)}(\alpha(B)\mathbf{1}_{\gamma-1}; \gamma\alpha(B); x J_{\gamma-1}^{(\theta,\beta)}) \mathbb{1}_{\mathbb{R}^+}(x) \end{aligned}$$

and the joint density function of $(\tilde{\mu}(B_1),\dots,\tilde{\mu}(B_n))$ is given by

$$f_{(\tilde{\mu}(B_1),\dots,\tilde{\mu}(B_n))}(x_1,\dots,x_n)=\frac{\prod_{j=1}^{\gamma}(j-\beta\mathbb{1}_{\{j>\gamma-\theta\}})^a}{\prod_{i=1}^{n}\Gamma(\gamma\alpha_i)}$$
$$\times\,\mathrm{e}^{-|\mathbf{x}|(\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}})}\prod_{i=1}^{n}x_i^{\gamma\alpha_i-1}\prod_{i=1}^{n}\Phi_2^{(\gamma-1)}(\alpha_i\mathbf{1}_{\gamma-1};\gamma\alpha_i;x_iJ_{\gamma-1}^{(\theta,\beta)})\mathbb{1}_{(\mathbb{R}^+)^n}(x_1,\dots,x_n).$$

From the joint density function of $(\tilde{\mu}(B_1),\dots,\tilde{\mu}(B_n))$ it is possible to find the transformed normalized joint density function $(\tilde{P}_{(\gamma,\beta,\theta)}(B_1),\dots,\tilde{P}_{(\gamma,\beta,\theta)}(B_{n-1}))$. This density function is obtain first considering the following transformed joint probability distribution function

$$f_{(\tilde{P}_{(\gamma,\beta,\theta)}(B_1),\dots,\tilde{P}_{(\gamma,\beta,\theta)}(B_{n-1}),\tilde{\mu}(\mathbb{X}))}(y_1,\dots,y_{n-1},y)$$
$$=\frac{\prod_{j=1}^{\gamma}(j-\beta\mathbb{1}_{\{j>\gamma-\theta\}})^a}{\prod_{i=1}^{n}\Gamma(\gamma\alpha_i)}\mathrm{e}^{-y(\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}})}\prod_{i=1}^{n-1}y_i^{\gamma\alpha_i-1}y^{\gamma a-1}\left(1-\sum_{i=1}^{n-1}y_i\right)^{\gamma\alpha_n-1}$$
$$\times\prod_{i=1}^{n-1}\Phi_2^{(\gamma-1)}(\alpha_i\mathbf{1}_{\gamma-1};\gamma\alpha_i;yy_iJ_{\gamma-1}^{(\theta,\beta)})$$
$$\times\,\Phi_2^{(\gamma-1)}\left(\alpha_n\mathbf{1}_{\gamma-1};\gamma\alpha_n;y\left(1-\sum_{i=1}^{n-1}y_i\right)J_{\gamma-1}^{(\theta,\beta)}\right)\mathbb{1}_{\Delta^{(n-1)}}(y_1,\dots,y_{n-1})$$

and then integrating over y. Formula (5.4.4) follows from the same proof for the particular case of $n=2$. □

For a fixed γ, if we set $\theta=0$ or $\beta=0$, or both equal to zero, it can be sees in formula (5.4.3) that the finite dimensional distributions are of the generalized Dirichlet process with parameter measure γ (see Lijoi et al. [118]). In particular, if we fix $\gamma=1$, then if we set $\theta=0$ or $\beta=0$, or both equal to zero, we recover from (5.4.3) the finite dimensional distributions are of the Dirichlet process. Moreover, the finite dimensional distributions of the generalized Dirichlet process with parameter $\gamma-\theta$ can be recoverd by the finite dimensional distributions of the generalized Dirichlet process with parameter (γ,β,θ) and taking the limit in distribution for $\beta\to-\infty$; in the same way the finite dimensional distributions of the Dirichlet process can be be recoverd by the finite dimensional distributions of the generalized Dirichlet process with parameter $(\gamma,\beta,\gamma-1)$ and taking the limit in distribution for $\beta\to-\infty$.

The following corollary highlight the particular cases of the finite dimensional distributions (5.4.3) when $\gamma=2$. In such particular case the confluent form of the fourth Lauricella function reduces respectively to a confluent hypergeometric function ${}_1F_1$ (see Appendix C).

Corollary 5.4.1. *Let $\tilde{P}_{(\gamma,\beta,\theta)}$ be a generalized Dirichlet process on $\mathbb{X}$ with parameter $(2,\beta,\theta)$. For any measurable partition $B_1,\dots,B_n$ of $\mathbb{X}$, $(\tilde{P}_{(\gamma,\beta,\theta)}(B_1),\dots,\tilde{P}_{(\gamma,\beta,\theta)}(B_{n-1}))$*

is a r.v admitting density function with respect to Lebesgue measure on the simplex $\Delta^{(n-1)}$ *given by*

$$\begin{aligned}
&f_{(\tilde{P}_{(\gamma,\beta,\theta)}(B_1),\dots,\tilde{P}_{(\gamma,\beta,\theta)}(B_{n-1}))}(y_1,\dots,y_{n-1})\\
&\quad=\frac{\Gamma(2a)\prod_{j=1}^{2}(j-\beta\mathbb{1}_{\{j>2-\theta\}})^{a}}{(2-\beta\mathbb{1}_{\{\theta\neq 0\}})^{2a}\prod_{i=1}^{n}\Gamma(2\alpha_i)}\prod_{i=1}^{n-1}y_i^{2\alpha_i-1}\left(1-\sum_{i=1}^{n-1}y_i\right)^{2\alpha_n-1}\\
&\qquad\times F_A^{(n)}\left(2a,\alpha_1,\dots,\alpha_n;2(\alpha_1,\dots,\alpha_n);\frac{J_1^{(\theta,\beta)}y_1}{(2-\beta\mathbb{1}_{\{\theta\neq 0\}})},\dots,\frac{J_1^{(\theta,\beta)}(1-\sum_{i=1}^{n-1}y_i)}{(2-\beta\mathbb{1}_{\{\theta\neq 0\}})}\right)\\
&\qquad\times\mathbb{1}_{\Delta^{(n-1)}}(y_1,\dots,y_{n-1}).
\end{aligned}$$

Proof. When $\gamma = 2$ we have that $\Phi_2^{(N)}$ reduces to the confluent hypergeometric function ${}_1F_1$. So using the series representation of ${}_1F_1$ we have

$$\begin{aligned}
&f_{(\tilde{P}_{(\gamma,\beta,\theta)}(B_1),\dots,\tilde{P}_{(\gamma,\beta,\theta)}(B_{n-1}))}(y_1,\dots,y_{n-1})\\
&\quad=\frac{\prod_{j=1}^{2}(j-\beta\mathbb{1}_{\{j>2-\theta\}})^{a}}{\prod_{i=1}^{n}\Gamma(2\alpha_i)}\prod_{i=1}^{n-1}y_i^{2\alpha_i-1}(1-\sum_{i=1}^{n-1}y_i)^{2\alpha_n-1}\int_0^{+\infty}\mathrm{e}^{-y(2-\beta\mathbb{1}_{\{\theta\neq 0\}})}y^{2a-1}\\
&\qquad\times\prod_{i=1}^{n-1}\sum_{m_i\geq 0}\frac{(\alpha_i)_{m_i\uparrow 1}(yy_iJ_1^{(\theta,\beta)})^{m_i}}{m_i!(2\alpha_i)_{m_i\uparrow 1}}\sum_{m_n\geq 0}\frac{(\alpha_n)_{m_n\uparrow 1}(y(1-\sum_{i=1}^{n-1}y_i)J_1^{(\theta,\beta)})^{m_n}}{m_n!(2\alpha_n)_{m_n\uparrow 1}}dy\\
&\qquad\times\mathbb{1}_{\Delta^{(n-1)}}(y_1,\dots,y_{n-1})\\
&\quad=\frac{\prod_{j=1}^{2}(j-\beta\mathbb{1}_{\{j>2-\theta\}})^{a}}{\prod_{i=1}^{n}\Gamma(2\alpha_i)}\prod_{i=1}^{n-1}y_i^{2\alpha_i-1}\left(1-\sum_{i=1}^{n-1}y_i\right)^{2\alpha_n-1}\\
&\qquad\times\sum_{(m_1,\dots,m_n)\in(\mathbb{N}_0)^n}\frac{(\alpha_1)_{m_1\uparrow 1}\cdots(\alpha_n)_{m_n\uparrow 1}y_1^{m_1}\cdots y_{n-1}^{m_{n-1}}(1-\sum_{i=1}^{n-1}y_i)^{m_n}}{m_1!(2\alpha_i)_{m_i}}\\
&\qquad\times\int_0^{+\infty}\mathrm{e}^{-y(2-\beta\mathbb{1}_{\{\theta\neq 0\}})}(J_1^{(\theta,\beta)})^{|\mathbf{m}|}y^{2a-1+|\mathbf{m}|}dy\,\mathbb{1}_{\Delta^{(n-1)}}(y_1,\dots,y_{n-1})
\end{aligned}$$

and the proof is completed. □

5.4.2 Predictive distributions and posterior characterization

The greater flexibility of the nonparametric model has to be constrained in order to incorporate real qualitative prior knowledge into the model. This is usually done by tuning some moments according to one's prior opinion. Walker and Damien [190] suggest controlling the mean and the variance of $\tilde{P}_{(\gamma,\beta,\theta)}$. Walker et al. [192] provided a detailed discussion about on specification in nonparametric problems.

The expected value of $\tilde{P}_{(\gamma,\beta,\theta)}$ takes on the interpretetion of a prior guess at the shape of $\tilde{P}_{(\gamma,\beta,\theta)}$ and is a crucial quantity in terms of prior specification. Indeed, if $\tilde{P}_{(\gamma,\beta,\theta)}$ is a generalized Dirichlet process with parameter (γ,β,θ), then

$$\mathbb{E}[\tilde{P}_{(\gamma,\beta,\theta)}(B)] = \frac{\alpha(B)}{a}$$

for any measurable set B (see Pitman [156]). Once setting the prior guess at the shape $\tilde{\varphi}$ through the chice of α, one has still four degrees of freedom in order to complete the prior specification: the total mass a and the parameters γ, θ and β.

Proposition 5.4.2. *Let $\tilde{P}_{(\gamma,\beta,\theta)}$ be a generalized Dirichlet process on $\mathbb{X}$ with parameter (γ,β,θ). If α is a non-atomic measure on $(\mathbb{X},\mathscr{X})$, then for any $B\in\mathscr{X}$ and $B_1,B_2\in\mathscr{X}$ such that $C=B_1\cap B_2$*

$$Var(\tilde{P}_{(\gamma,\beta,\theta)}(B)) = \frac{\alpha(B)(a-\alpha(B))}{a^2}\mathcal{I}_{a,\gamma,\theta,\beta} \tag{5.4.5}$$

$$Cov(\tilde{P}_{(\gamma,\beta,\theta)}(B_1),\tilde{P}(B_2)) = \frac{a\alpha(C)-\alpha(B_1)\alpha(B_2)}{(a)^2}\mathcal{I}_{a,\gamma,\theta,\beta} \tag{5.4.6}$$

$$sk(\tilde{P}_{(\gamma,\beta,\theta)}(B)) = \frac{a-2\alpha(B)}{2(\alpha(B)(a-\alpha(B)))^{1/2}}\mathcal{K}_{a,\gamma,\theta,\beta} \tag{5.4.7}$$

where

$$\mathcal{I}_{a,\gamma,\theta,\beta} = \frac{\Gamma(\gamma a)a\prod_{j=1}^{\gamma}(j-\beta\mathbb{1}_{\{j>\gamma-\theta\}})^a}{\Gamma(\gamma a+2)(\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}})^{\gamma a}}\sum_{k=1}^{\gamma}F_D^{(\gamma-1)}\left(\gamma a,\alpha_k^*;\gamma a+2;\frac{J_{\gamma-1}^{(\theta,\beta)}}{\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}}}\right) \tag{5.4.8}$$

and

$$\mathcal{K}_{a,\gamma,\theta,\beta} = \frac{4(\prod_{j=1}^{\gamma}(j-\beta\mathbb{1}_{\{j>\gamma-\theta\}}))^{-a/2}\Gamma(\gamma a+2)^{1/2}}{a^{1/2}(\Gamma(\gamma a))^{1/2}(\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}})^{-\gamma a/2}(\gamma a+2)} \times \frac{\sum_{k=1}^{\gamma}F_D^{(\gamma-1)}(\gamma a,\alpha_k^{**};\gamma a+3;J_{\gamma-1}^{(\theta,\beta)}/\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}})}{(\sum_{k=1}^{\gamma}F_D^{(\gamma-1)}(\gamma a,\alpha_k^*;\gamma a+2;J_{\gamma-1}^{(\theta,\beta)}\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}}))^{3/2}} \tag{5.4.9}$$

where $\alpha_k^*=(a,\dots,a+2,\dots,a)\in\mathbb{R}^{\gamma-1}$ with $a+2$ being the $(\gamma-k)$-th element of the vector and $\alpha_k^{**}=(a,\dots,a+3,\dots,a)\in\mathbb{R}^{\gamma-1}$ with $a+3$ being the $(\gamma-k)$-th element of the vector.

Proof. As regard the variance, from Proposition 1 in James et al. [99] we have

$$\mathcal{I}_{a,\gamma,\theta,\beta} = a\int_0^{+\infty} u\mathrm{e}^{-a\Psi(u)}\int_{\mathbb{R}^+} v^2\mathrm{e}^{-uv}\nu(dv)du$$

where Ψ stands for the Laplace exponent of the random measure at issue and ν is the Lévy measure. Hence, in the case of a generalized Dirichlet process with parameter (γ, β, θ), one has

$$\begin{aligned}
\mathcal{I}_{a,\gamma,\theta,\beta} &= a \sum_{k=1}^{\gamma} \int_0^{+\infty} u \left(\prod_{j=1}^{\gamma} \frac{j - \beta \mathbb{1}_{\{j>\gamma-\theta\}}}{j - \beta \mathbb{1}_{\{j>\gamma-\theta\}} + u} \right)^a \int_0^{+\infty} v \mathrm{e}^{-v(u+k-\beta \mathbb{1}_{\{k>\gamma-\theta\}})} dv du \\
&= a \sum_{k=1}^{\gamma} \int_0^{+\infty} u \left(\prod_{j=1}^{\gamma} \frac{j - \beta \mathbb{1}_{\{j>\gamma-\theta\}}}{j - \beta \mathbb{1}_{\{j>\gamma-\theta\}} + u} \right)^a (u + k - \beta \mathbb{1}_{\{k>\gamma-\theta\}})^{-2} du \\
&= a \sum_{k=1}^{\gamma} (k - \beta \mathbb{1}_{\{k>\gamma-\theta\}})^{-2} \mathbb{E}[(\eta_k^{(\gamma,\theta,\beta)} + \zeta_a^{(\gamma,\theta,\beta)})^{-2}]
\end{aligned}$$

where $\eta_k^{(\gamma,\beta,\theta)}$ is r.v. distributed according to a Gamma distribution function with parameter $(k - \beta \mathbb{1}_{\{k>\gamma-\theta\}}, 2)$, for $k = 1, \ldots, \gamma$ and $\zeta_a^{(\gamma,\beta,\theta)}$ is the convolution of γ independent r.v.s distributed according to a Gamma distribution function with parameter $(j - \beta \mathbb{1}_{\{j>\gamma-\theta\}}, a)$, for $j = 1, \ldots, \gamma$. Then, the density function is

$$\begin{aligned}
f_{\eta_k^{(\gamma,\beta,\theta)} + \zeta_a^{(\gamma,\beta,\theta)}}(y) &= \frac{\prod_{j=1}^{\gamma} (j - \beta \mathbb{1}_{\{j>\gamma-\theta\}})^a (k - \beta \mathbb{1}_{\{k>\gamma-\theta\}})^2}{\Gamma(\gamma a)} \\
&\qquad \times \int_0^y (y - x) \mathrm{e}^{-(k-\beta \mathbb{1}_{\{k>\gamma-\theta\}})(y-x)} x^{\gamma a - 1} \mathrm{e}^{-x(\gamma - \beta \mathbb{1}_{\{\theta \neq 0\}})} \\
&\qquad \times \Phi_2^{(\gamma-1)}(\alpha(B) \mathbf{1}_{\gamma-1}; \gamma\alpha(B); x J_{\gamma-1}^{(\theta,\beta)}) \mathbb{1}_{\mathbb{R}^+}(y) dx \\
&= \frac{\prod_{j=1}^{\gamma} (j - \beta \mathbb{1}_{\{j>\gamma-\theta\}})^a (k - \beta \mathbb{1}_{\{k>\gamma-\theta\}})^2}{\Gamma(\gamma a)} y^{\gamma a + 1} \mathrm{e}^{-y(k-\beta \mathbb{1}_{\{k>\gamma-\theta\}})} \\
&\qquad \times \int_0^1 z^{\gamma a - 1} (1 - z) \mathrm{e}^{-zy(\gamma - k + \beta[\mathbb{1}_{\{k>\gamma-\theta\}} - \mathbb{1}_{\{\theta \neq 0\}}])} \\
&\qquad \times \Phi_2^{(\gamma-1)}(\alpha(B) \mathbf{1}_{\gamma-1}; \gamma\alpha(B); yz J_{\gamma-1}^{(\theta,\beta)}) \mathbb{1}_{\mathbb{R}^+}(y) dz.
\end{aligned}$$

Hence, we have

$$\begin{aligned}
\mathcal{I}_{a,\gamma,\theta,\beta} &= a\sum_{k=1}^{\gamma}(k-\beta\mathbb{1}_{\{k>\gamma-\theta\}})^{-2}\int_0^{+\infty} y^{-2} f_{\eta_k^{(\gamma,\beta,\theta)}+\zeta_a^{(\gamma,\beta,\theta)}}(y)dy \\
&= \frac{\prod_{j=1}^{\gamma}(j-\beta\mathbb{1}_{\{j>\gamma-\theta\}})^a}{\Gamma(\gamma a)} \\
&\quad\times\sum_{k=1}^{\gamma}\int_0^1 z^{\gamma a-1}(1-z)\int_0^{+\infty} y^{\gamma a-1}\mathrm{e}^{-[k-\beta\mathbb{1}_{\{k>\gamma-\theta\}}+(\gamma-k+\beta[\mathbb{1}_{\{k>\gamma-\theta\}}-\mathbb{1}_{\{\theta\neq 0\}}])z]y} \\
&\quad\times\Phi_2^{(\gamma-1)}(\alpha(B)\mathbf{1}_{\gamma-1};\gamma\alpha(B);yzJ_{\gamma-1}^{(\theta,\beta)})dydz \\
&= a\prod_{j=1}^{\gamma}(j-\beta\mathbb{1}_{\{j>\gamma-\theta\}})^a \\
&\quad\times\sum_{k=1}^{\gamma}\int_0^1\frac{z^{\gamma a-1}(1-z)}{[k-\beta\mathbb{1}_{\{k>\gamma-\theta\}}+(\gamma-k+\beta[\mathbb{1}_{\{k>\gamma-\theta\}}-\mathbb{1}_{\{\theta\neq 0\}}])z]^{\gamma a}} \\
&\quad\times F_D^{(\gamma-1)}\left(\gamma a, a\mathbf{1}_{\gamma-1};\gamma a;\frac{zJ_{\gamma-1}^{(\theta,\beta)}}{[k-\beta\mathbb{1}_{\{k>\gamma-\theta\}}+(\gamma-k+\beta[\mathbb{1}_{\{k>\gamma-\theta\}}-\mathbb{1}_{\{\theta\neq 0\}}])z]}\right)dz.
\end{aligned}$$

Because for any $x\in[0,1)^n$, $a>0$ and $b_i>0$ for $i=1,\dots,n$

$$F_D^{(n)}(a,b_1,\dots,b_n;a;x)=\prod_{i=1}^{n}(1-x_i)^{-b_i}$$

then, using a change the variable $z=v(k-\beta\mathbb{1}_{\{k>\gamma-\theta\}})/(\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}}-v(\gamma-k+\beta[\mathbb{1}_{\{k>\gamma-\theta\}}-\mathbb{1}_{\{\theta\neq 0\}}]))$ we obtain

$$\begin{aligned}
\mathcal{I}_{a,\gamma,\theta,\beta} &= \frac{a\prod_{j=1}^{\gamma}(j-\beta\mathbb{1}_{\{j>\gamma-\theta\}})^a}{(\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}})^{\gamma a}} \\
&\quad\times\sum_{k=1}^{\gamma}\int_0^1 v^{\gamma a-1}(1-v)\left(1-\frac{v(\gamma-k+\beta[\mathbb{1}_{\{k>\gamma-\theta\}}-\mathbb{1}_{\{\theta\neq 0\}}])}{\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}}}\right)^{-2} \\
&\quad\times\prod_{j=1}^{\gamma-1}\left(1-\frac{v(j-\beta[\mathbb{1}_{\{\gamma-j>\gamma-\theta\}}-\mathbb{1}_{\{\theta\neq 0\}}])}{\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}}}\right)^{-a}dv \\
&= \frac{a\prod_{j=1}^{\gamma}(j-\beta\mathbb{1}_{\{j>\gamma-\theta\}})^a}{(\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}})^{\gamma a}} \\
&\quad\times\sum_{k=1}^{\gamma}\int_0^1 v^{\gamma a-1}(1-v)\prod_{j=1}^{\gamma-1}\left(1-\frac{v(j-\beta[\mathbb{1}_{\{\gamma-j>\gamma-\theta\}}-\mathbb{1}_{\{\theta\neq 0\}}])}{\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}}}\right)^{-(a+2\mathbb{1}_{\{j=\gamma-k\}})}dv \\
&= \frac{\Gamma(\gamma a)a\prod_{j=1}^{\gamma}(j-\beta\mathbb{1}_{\{j>\gamma-\theta\}})^a}{\Gamma(\gamma a+2)(\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}})^{\gamma a}} \\
&\quad\times\sum_{k=1}^{\gamma}F_D^{(\gamma-1)}\left(\gamma a,\alpha_k^*;\gamma a+2;\frac{J_{\gamma-1}^{(\theta,\beta)}}{\gamma-\beta\mathbb{1}_{\{\theta\neq 0\}}}\right)
\end{aligned}$$

where $\alpha_k^* = (a, \ldots, a+2, \ldots, a) \in \mathbb{R}^{\gamma-1}$ with $a+2$ being the $(\gamma-k)$-th element of the vector. Using similar arguments to those employed for determining $\mathcal{I}_{a,\gamma,\theta,\beta}$ is it possible to prove

$$\mathcal{J}_{a,\gamma,\theta,\beta} = a \int_0^{+\infty} u^2 \mathrm{e}^{-a\Psi(u)} \int_{\mathbb{R}^+} v^3 \mathrm{e}^{-uv} \nu(dv) du = \frac{\Gamma(\gamma a) 4a \prod_{j=1}^{\gamma} (j - \beta \mathbb{1}_{\{j>\gamma-\theta\}})^a}{\Gamma(\gamma a + 3)(\gamma - \beta \mathbb{1}_{\{\theta \neq 0\}})^{\gamma a}}$$
$$\times \sum_{k=1}^{\gamma} F_D^{(\gamma-1)} \left(\gamma a, \alpha_k^{**}; \gamma a + 3; \frac{J_{\gamma-1}^{(\theta,\beta)}}{\gamma - \beta \mathbb{1}_{\{\theta \neq 0\}}} \right)$$

where $\alpha_k^{**} = (a, \ldots, a+3, \ldots, a) \in \mathbb{R}^{\gamma-1}$ with $a+3$ being the $(\gamma-k)$-th element of the vector. From the last equation one obtains $\mathcal{K}_{a,\gamma,\theta,\beta} = J_{a,\gamma,\theta,\beta}/\mathcal{I}_{a,\gamma,\theta,\beta}^{3/2}$. □

In Figure 5.4 and Figure 5.5 we show the values for $\mathcal{I}_{a,\gamma,\beta,\theta}$ and $\mathcal{K}_{a,\gamma,\beta,\theta}$ for the generalized Dirichlet process with parameter (γ, β, θ) respectivly with prior guess $\alpha(\cdot)$ and $\alpha(\cdot)/4$ for some fixed values of $\gamma \in \mathbb{N}$ and variyng $\theta \in \mathbb{N}_0$ and β. In particular, $\mathcal{I}_{a,\gamma,\beta,\theta}$ and $\mathcal{K}_{a,\gamma,\beta,\theta}$ are relevant in the context of the prior specification.

In the context of prior specification, the prior opinion on the unknown $\tilde{\varphi}$ is reflected by the choice of the prior guess α_0, while a, $\gamma \in \mathbb{N}$, $\theta \in \mathbb{N}$ and β represents degree of belief in α_0. In particular, we consider the prior variance and the prior skewness which can be tuned acting on $\mathcal{I}_{a,\gamma,\beta,\theta}$ and $\mathcal{K}_{a,\gamma,\beta,\theta}$. It is known from Lijoi et al. [118] that for the generalized Dirichlet process with parameter γ, the prior variance and the prior skewness decrease as a or $\gamma \in \mathbb{N}$ increase which means the bigger a or γ the greater is confidence in α_0. In other words, increasing a or $\gamma \in \mathbb{N}$ implies the constraint that more weight is given to the prior guess α_0. Such properties of the generalized Dirichlet process can be observed in Figure 5.4 and Figure 5.5 by looking the values corresponding to $\beta = 0$. The generalized Dirichlet process with parameter (γ, β, θ) extends such property of the generalized Dirichlet process with parameter γ. In particular, for any fixed a, it is possible to chose a parameter $\theta \in \mathbb{N}_0$ and β such that increasing $\gamma \in \mathbb{N}$ the prior variance and the prior skewness do not decrease. In other word, for generalized Dirichlet process with parameter (γ, β, θ) the weight given to the prior guess α_0 is not directly related to the choice of the parameter $\gamma \in \mathbb{N}$.

We now focus on another important feature of a nonparametric model which is related to the prediction of feature values of a random quantity based on its past outcomes. We set $\mathbf{n} := (n_1, \ldots, n_k)$, $\mathbf{n}^+ := (n_1, \ldots, n_k, 1)$ and $\mathbf{n}_j^+ := (n_1, \ldots, n_j + 1, \ldots, n_k)$. Under the hypothesis of non-atomicity of the measure α the predictive distribution can be found by the application of Proposition 5.2.2.

Proposition 5.4.3. *Let $\tilde{P}_{(\gamma,\beta,\theta)}$ be the generalized Dirichlet process on $\mathbb{X}$ with parameter (γ, β, θ). If α is a non-atomic measure on $(\mathbb{X}, \mathscr{X})$, then the predictive distribution, given*

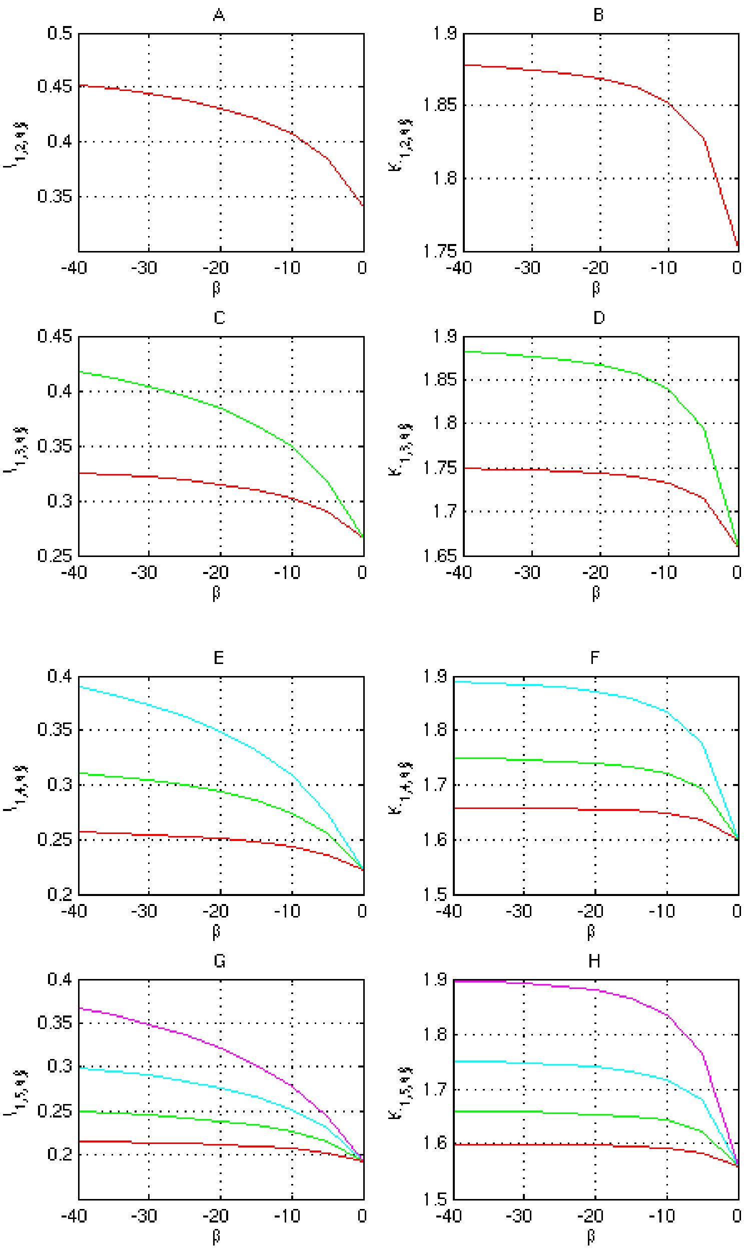

Figure 5.4: Values for $\mathcal{I}_{a,\gamma,\theta,\beta}$ and $\mathcal{K}_{a,\gamma,\theta,\beta}$ for a generalized Dirichlet process with parameter (γ, β, θ). Figures A), B) $\gamma = 2$, $\theta = 1$ (red). Figures C), D) $\gamma = 3$, $\theta = 1$ (red), $\theta = 2$ (green). Figures E), F) $\gamma = 4$, $\theta = 1$ (red), $\theta = 2$ (green), $\theta = 3$ (cian). Figures G), H) $\gamma = 5$, $\theta = 1$ (red), $\theta = 2$ (green), $\theta = 3$ (cian), $\theta = 4$ (magenta).

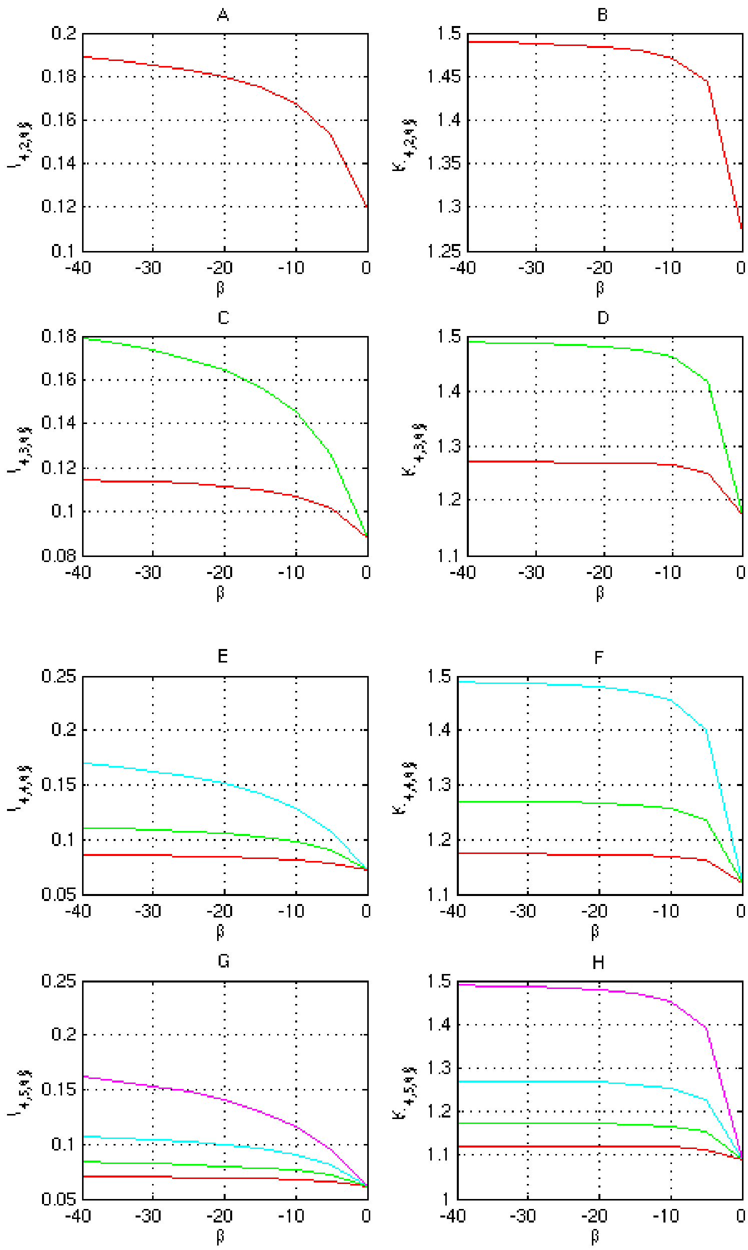

Figure 5.5: Values for $\mathcal{I}_{a,\gamma,\theta,\beta}$ and $\mathcal{K}_{a,\gamma,\theta,\beta}$ for a generalized Dirichlet process with parameter $(\alpha/4, \gamma, \beta, \theta)$. Figures A), B) $\gamma = 2$, $\theta = 1$ (red). Figures C), D) $\gamma = 3$, $\theta = 1$ (red), $\theta = 2$ (green). Figures E), F) $\gamma = 4$, $\theta = 1$ (red), $\theta = 2$ (green), $\theta = 3$ (cian). Figures G), H) $\gamma = 5$, $\theta = 1$ (red), $\theta = 2$ (green), $\theta = 3$ (cian), $\theta = 4$ (magenta).

$X_1, \dots, X_n$ *is of the form*

$$\mathbb{P}(X_{n+1} \in dx | X_1, \dots, X_n) = \frac{a}{\gamma a + n} \frac{\alpha(dx)}{a} w(\mathbf{n}^+) + \frac{n}{\gamma a + n} \frac{1}{n} \sum_{j=1}^{k} n_j w(\mathbf{n}_j^+) \delta_{X_j^*}(dx)$$

where

$$w(\mathbf{n}^+) = \frac{\sum_{\mathbf{r}^{k+1}} F_D^{(\gamma-1)}(\gamma a, \alpha^*(\mathbf{n}^+, \mathbf{r}^{k+1}); \gamma a + n + 1; J_{\gamma-1}^{(\theta,\beta)}/(\gamma - \beta \mathbb{1}_{\{\theta \neq 0\}}))}{\sum_{\mathbf{r}^k} F_D^{(\gamma-1)}(\gamma a, \alpha^*(\mathbf{n}, \mathbf{r}^k); \gamma a + n; J_{\gamma-1}^{(\theta,\beta)}/(\gamma - \beta \mathbb{1}_{\{\theta \neq 0\}}))} \tag{5.4.10}$$

and

$$w(\mathbf{n}_j^+) = \frac{\sum_{\mathbf{r}^k} F_D^{(\gamma-1)}(\gamma a, \alpha^*(\mathbf{n}^+{}_j, \mathbf{r}^k); \gamma a + n + 1; J_{\gamma-1}^{(\theta,\beta)}/(\gamma - \beta \mathbb{1}_{\{\theta \neq 0\}}))}{\sum_{\mathbf{r}^k} F_D^{(\gamma-1)}(\gamma a, \alpha^*(\mathbf{n}, \mathbf{r}^k); \gamma a + n; J_{\gamma-1}^{(\theta,\beta)}/(\gamma - \beta \mathbb{1}_{\{\theta \neq 0\}}))}. \tag{5.4.11}$$

where $\alpha^*(\mathbf{n}, \mathbf{r}^k) = (\alpha_1^*(\mathbf{n}, \mathbf{r}^k), \dots, \alpha_{\gamma-1}^*(\mathbf{n}, \mathbf{r}^k))$ *with*

$$\alpha_j^*(\mathbf{n}, \mathbf{r}^k) = a + \sum_{i=1}^{k} n_i \mathbb{1}_{\{j = r_i\}}$$

for $j = 1, \dots, \gamma - 1$.

Proof. Under the hypothesis of non-atomicity of α Pitman [156] shows (see also Proposition 5.2.2) that the predictive distribution corresponding to a NRMI is of the form

$$\begin{aligned}\mathbb{P}(X_{n+1} &\in \cdot | X_1, \dots, X_n) \\ &= \frac{p_{n+1}^{(k+1)}(n_1, \dots, n_k, 1)}{n p_n^{(k)}(n_1, \dots, n_k)} \alpha(\cdot) + \frac{1}{n} \sum_{j=1}^{n} \frac{p_{n+1}^{(k)}(n_1, \dots, n_j + 1, \dots, n_k)}{n p_n^{(k)}(n_1, \dots, n_k)} \delta_{X_j^*}(\cdot)\end{aligned}$$

where $p_n^{(k)}(n_1, \dots, n_k) = \int_0^{+\infty} u^{n-1} \mathrm{e}^{-a\Psi(u)} \mu_{n_j}(u) du$ and, for $j = 1, \dots, k$ we have $\mu_{n_j}(u) = \int_{\mathbb{R}^+} v^{n_j} \mathrm{e}^{-uv} \nu(dv)$. Thus, for the generalized Dirichlet process with parameter (γ, β, θ), we have

$$\begin{aligned}\mu_{n_j}(u) \int_0^{+\infty} v^{n_j} \mathrm{e}^{-uv} \sum_{r=1}^{\gamma} \frac{\mathrm{e}^{-v(r - \beta \mathbb{1}_{\{r > \gamma - \theta\}})}}{v} dv &= \sum_{r=1}^{\gamma} \int_0^{+\infty} s^{n_j - 1} \mathrm{e}^{-v(u + r - \beta \mathbb{1}_{\{r > \gamma - \theta\}})} dv \\ &= \Gamma(n_j) \sum_{r=1}^{\gamma} (u + r - \beta \mathbb{1}_{\{r > \gamma - \theta\}})^{-n_j}.\end{aligned}$$

Using $\mathbf{r}^k := (r_1, \ldots, r_k) \in [\gamma]^k$ we have

$$
\begin{aligned}
p_n^{(k)}(n_1, \ldots, n_k) &= \int_0^{+\infty} u^{n-1} \mathrm{e}^{-a\Psi(u)} \prod_{j=1}^{k} \int_0^{+\infty} v^{n_j} \mathrm{e}^{-uv} \sum_{r=1}^{\gamma} \frac{\mathrm{e}^{-v(r-\beta \mathbb{1}_{\{r>\gamma-\theta\}})}}{v} dv du \\
&= \prod_{j=1}^{k} \Gamma(n_j) \sum_{\mathbf{r}^k} \int_0^{+\infty} u^{n-1} \prod_{j=1}^{\gamma} \left(\frac{j - \beta \mathbb{1}_{\{j>\gamma-\theta\}} + s}{j - \beta \mathbb{1}_{\{j>\gamma-\theta\}}} \right)^{-a} \\
&\qquad \times \prod_{i=1}^{k} (u + r_i - \beta \mathbb{1}_{\{r_i>\gamma-\theta\}})^{-n_i} du \\
&= \prod_{j=1}^{k} \Gamma(n_j) \sum_{\mathbf{r}^k} \prod_{i=1}^{k} (r_i - \beta \mathbb{1}_{\{r_i>\gamma-\theta\}})^{-n_i} \int_0^{+\infty} u^{n-1} \mathbb{E}[\mathrm{e}^{-uV_{a,k}}] du \\
&= \Gamma(n) \prod_{j=1}^{k} \Gamma(n_j) \sum_{\mathbf{r}^k} \prod_{i=1}^{k} (r_i - \beta \mathbb{1}_{\{r_i>\gamma-\theta\}})^{-n_i} \mathbb{E}[V_{a,k}^{-n}]
\end{aligned}
$$

where $V_{a,k} = \sum_{j=1}^{\gamma} Y_j$ with the Y_j being independent r.v.s distributed according to a Gamma distribution function with parameter $(j - \beta \mathbb{1}_{\{j>\gamma-\theta\}}, a + \sum_{1 \le i \le k} n_i \mathbb{1}_{\{j=r_i\}})$ for $j = 1, \ldots, \gamma$. We first set $\alpha_j^*(\mathbf{n}, \mathbf{r}^k) := a + \sum_{1 \le i \le k} n_i \mathbb{1}_{\{j=r_i\}}$ and then we define $\alpha^*(\mathbf{n}, \mathbf{r}^k) := (\alpha_1^*(\mathbf{n}, \mathbf{r}^k), \ldots, \alpha_{\gamma-1}^*(\mathbf{n}, \mathbf{r}^k))$. We can express the convolution of independent r.v.s $V_{a,k}$ distributed according to a Gamma distribution funcitons in terms of the confluent form of the fourth Lauricella function. Then we have

$$
\begin{aligned}
p_n^{(k)}(n_1, \ldots, n_k) &= \int_0^{+\infty} u^{n-1} \mathrm{e}^{-a\Psi(u)} \prod_{j=1}^{k} \int_0^{+\infty} v^{n_j} \mathrm{e}^{-uv} \sum_{r=1}^{\gamma} \frac{\mathrm{e}^{-v(r-\beta \mathbb{1}_{\{r>\gamma-\theta\}})}}{v} dv du \\
&= \frac{\Gamma(n) \prod_{j=1}^{k} \Gamma(n_j) \prod_{j=1}^{\gamma} (j - \beta \mathbb{1}_{\{j>\gamma-\theta\}})^a}{\Gamma(\gamma a + n)} \\
&\qquad \times \sum_{\mathbf{r}^k} \int_0^{+\infty} v^{\gamma a - 1} \mathrm{e}^{-v(\gamma - \beta \mathbb{1}_{\{\theta \neq 0\}})} \Phi_2^{(\gamma-1)}(\alpha^*(\mathbf{n}, \mathbf{r}^k); \gamma a + n; v J_{\gamma-1}^{(\theta,\beta)}) dv \\
&= \frac{\Gamma(n) \prod_{j=1}^{k} \Gamma(n_j) \prod_{j=1}^{\gamma} (j - \beta \mathbb{1}_{\{j>\gamma-\theta\}})^a}{(\gamma - \beta \mathbb{1}_{\{\theta \neq 0\}})^{\gamma a} \Gamma(\gamma a + n)} \\
&\qquad \times \sum_{\mathbf{r}^k} \int_0^{+\infty} z^{\gamma a - 1} \mathrm{e}^{-z} \Phi_2^{(\gamma-1)} \left(\alpha^*(\mathbf{n}, \mathbf{r}^k); \gamma a + n; \frac{z J_{\gamma-1}^{(\theta,\beta)}}{(\gamma - \beta \mathbb{1}_{\{\theta \neq 0\}})} \right) dz \\
&= \frac{\Gamma(\gamma a) \Gamma(n) \prod_{j=1}^{k} \Gamma(n_j) \prod_{j=1}^{\gamma} (j - \beta \mathbb{1}_{\{j>\gamma-\theta\}})^a}{(\gamma - \beta \mathbb{1}_{\{\theta \neq 0\}})^{\gamma a} \Gamma(\gamma a + n)} \\
&\qquad \times \sum_{\mathbf{r}^k} F_D^{(\gamma-1)} \left(\gamma a, \alpha^*(\mathbf{n}, \mathbf{r}^k); \gamma a + n; \frac{J_{\gamma-1}^{(\theta,\beta)}}{(\gamma - \beta \mathbb{1}_{\{\theta \neq 0\}})} \right).
\end{aligned}
$$

Then the result follows immediately from the application of Proposition 5.2.2. □

It is useful to compare the predictive distributions of the generalized Dirichlet process with parameter (γ, β, θ) with the predictive distributions of other NRMIs known in the literature. In particular we are interested to compare the predictive distributions of the generalized Dirichlet process with parameter (γ, β, θ) with the predictive distributions of the generalized Dirichlet process with parameter γ, the predictive distributions of the Dirichlet process and the predictive distributions of the class of Gibbs-type r.p.m.s.

In terms of predictive mechanism the more simple case of Gibbs-type r.p.m. is the Dirichlet process. It is known that for the Dirichlet process the predictive distributions are a linear combination of the prior guess α_0 and a weighted empirical distribution: X_{n+1} is new with probability $a/(a+n)$ and it coincides with X_j^* with probability $n_j/(a+n)$, for $j = 1, \ldots, k$. Therefore for the Dirichlet process, the probability allocated to previous observations and the probability of a new observation do not depend on the number k of the distinct observations within the sample; in particular, the weight assigned to each X_j^* depends only on the number of observations equal to X_j^*. The predictive distributions of a Gibbs-type r.p.m. are still a linear combination of the prior guess α_0 and a weighted empirical distribution: X_{n+1} is new with probability $g_0(n,k)$ and and it coincides with X_j^* with probability $g_1(n,k)(n_j - \sigma)$, for $j = 1, \ldots, k$. Therefore for a Gibbs-type r.p.m. the weight assigned to each X_j^* depends on the number of distinct observation k and on the number of observations equal to X_j^*, while the weight assigned to a new observation depend on the number of distinct observation k. The balancing between new and old observations takes into account the number of distinct observation k.

As regard the generalized Dirichlet process with parameter (γ, β, θ), the predictive mechanism is more interesting and it exploits all available information in the sample $X_1, \ldots, X_n$. The predictive distributions of the generalized Dirichlet process with parameter (γ, β, θ) are still a linear combination of the prior guess α_0 and a weighted empirical distribution: X_{n+1} is new with probability $w(\mathbf{n}^+)a/(\gamma a + n)$ and and it coincides with X_j^* with probability $w(\mathbf{n}_j^+)n_j/(\gamma a + n)$, for $j = 1, \ldots, k$. Therefore, from equation (5.5.14) and equation (5.5.15) we observe that both weight assigned to each X_j^* and the weight assigned to a new observation depend on the number of distinct observation k and on the frequencies $(n_1, \ldots, n_k)$ of the k distinct observations X_j^*. We can conclude that differently from the Gibbs-type r.p.m.s, the predictive mechanism of the generalized Dirichlet process with parameter (γ, β, θ) is based on a balancing between new and old observations takes into account two elements: the number of distinct observation k and on the frequencies $(n_1, \ldots, n_k)$ of the k distinct observations X_j^*.

For construction, the generalized Dirichlet process with parameter γ proposed by Lijoi et al. [118] and the generalized Dirichlet process with parameter (γ, β, θ) have the same interesting predictive mechanism; the unique difference appears in the variable of the Lauricella function $F_D^{(\gamma-1)}$. In particular, the generalized Dirichlet process with parameter (γ, β, θ) is characterized by a more flexible definition of such variable which involves the two extra parameter $\theta \in \mathbb{N}_0$ and β. Such difference suggest that the parameter $\theta \in \mathbb{N}_0$ and β could determine a s ort of control in the reinforcement mechanism of the posterior probability allocated to X_j^*. It is known that a distinctive feauture of a Gibbs-type r.p.m. different from the Dirichlet process is that the posterior probability allocation to each X_j^* is more elaborated than for the Dirichlet case; instead of increasing the weight proportionally to the number ot ties, the probability assigned to X_j^* is reinforced more than proportionally each time a new tie has been recorded. This can be explained by the argument that the more often X_j^* is observed, the stronger is the statistical evidence in its favor and, thus it is sensible to reallocate mass toward it. In particular, it is known that for a Gibbs-type r.p.m. different from the Dirichlet process suth reinforcement mechanism is controlled by the parameter σ which greatly influence the clustering behaviour: a value of σ close to one generates a large number of clusters most of which with small size. Then, a reinforcement mechanism driven by σ acts on the mass allocation by penalizing clusters of small size and favouring those few groups containing a large number of elements. We conjecture that a reinforcement mechanism of same type above described for the Gibbs-type r.p.m.s follows for the generalized Dirichlet process with parameter (γ, β, θ); in particular, such reinforcement mechanism could be related to the choice of parameter $\gamma \in \mathbb{N}$, $\theta \in \mathbb{N}_0$ and β. A numerical example may clarify such conjecture.

To highlight the reinforcement mechanism, let the sample space be $[0, 1]$ and let the prior guess α_0 be the Uniform distribution with a specified is such way that all the process compared has the same mean and variance. Suppose that one has observed the samples $X^{(1)} = (X_1^{(1)}, \ldots, X_{10}^{(1)})$ and $X^{(2)} = (X_1^{(2)}, \ldots, X_{40}^{(2)})$ such that in each sample 40% of the observation are of type $X_1^* = 0.3$, 30% of the observations are of type $X_2^* = 0.1$, 20% of the observations are of type $X_3^* = 0.6$ and 10% of the observations are of type $X_4^* = 0.5$. Based on such asspumptions we compare for each sample the reinforcement mechanism for the predictive distributions of a Dirichlet process, a normalized inverse Gaussian process, a generalized Dirichlet process with parameter γ and a generalized Dirichlet process with parameter specified in Table 5.1. To emphasize the reinforcement mechanism, the numerical results are given in terms ratio between the posterior probability allocation and they are displayed in Figure 5.6 and Figure 5.7 for the sample $X^{(1)}$ and Figure 5.8 and Figure 5.9 for the sample $X^{(2)}$.

From Figures 5.6, 5.7, 5.8 and 5.9 we can observe for the generalized Dirichlet process with parameter γ a reinforcement mechanism is related to the parameter γ.

When γ increases the probability assigned to X_j^* is reinforced more than proportionally each time a new tie has been recorded. In particular, we observe that for $\gamma = 2$ the

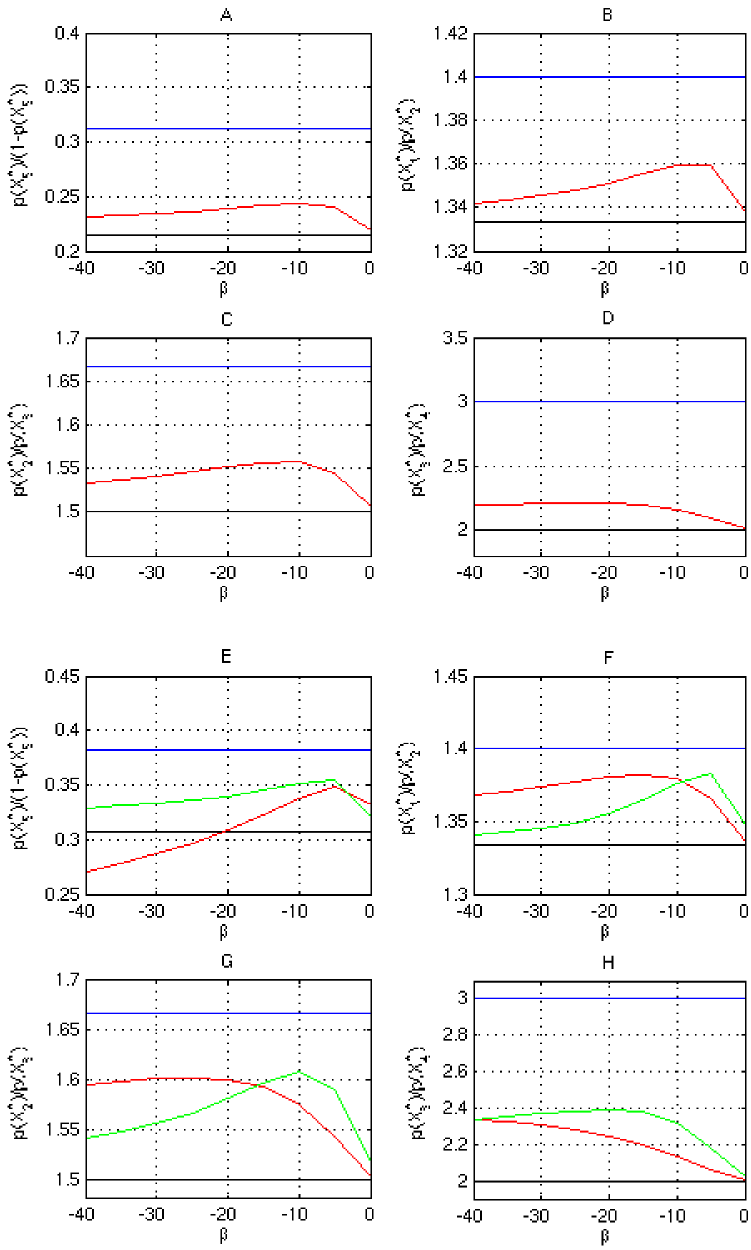

Figure 5.6: Ratio of posterior probability allocation for the Dirichlet process, normalized inverse-Gaussian process, generalized Dirichlet process with parameter γ and generalized Dirichlet process with parameter (γ, β, θ) using parameter specification of Table 5.1 and sample $X^{(1)}$. Figures A), B), C), D), ratio of posterior probability allocation for cases 1)-9) in Table 5.1. Figure E), F), G), H), ratio of posterior probability allocation for cases 10)-18) in Table 5.1.

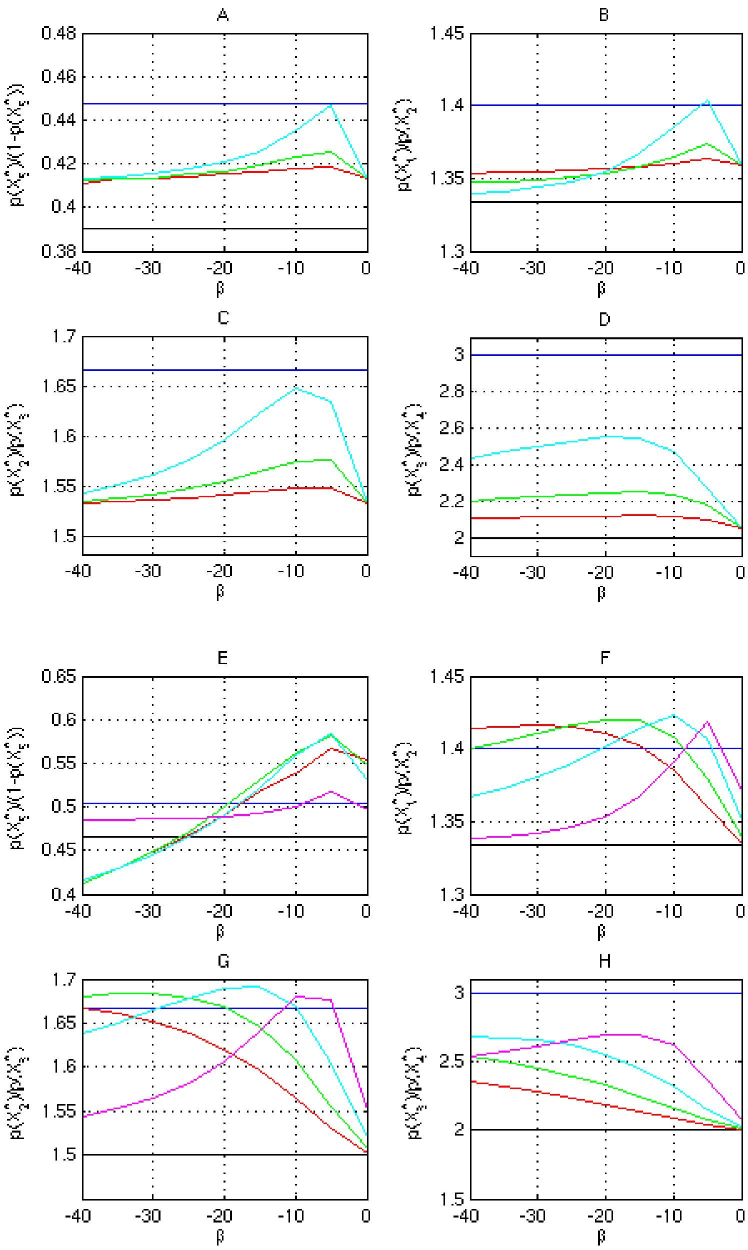

Figure 5.7: Ratio of posterior probability allocation for the Dirichlet process, normalized inverse-Gaussian process, generalized Dirichlet process with parameter γ and generalized Dirichlet process with parameter (γ, β, θ) using parameter specification of Table 5.1 and sample $X^{(1)}$. Figures A), B), C), D), ratio of posterior probability allocation for cases 19)-27) in Table 5.1. Figure E), F), G), H), ratio of posterior probability allocation for cases 28)-36) in Table 5.1.

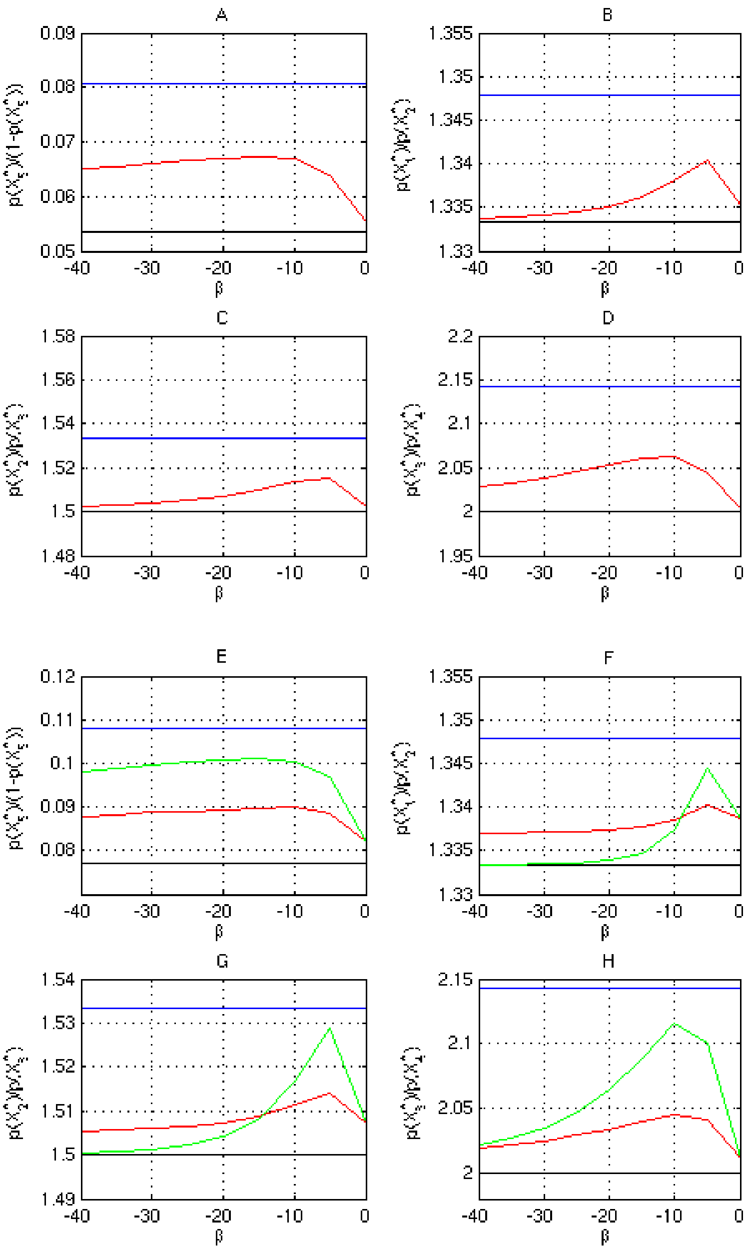

Figure 5.8: Ratio of posterior probability allocation for the Dirichlet process, normalized inverse-Gaussian process, generalized Dirichlet process with parameter γ and generalized Dirichlet process with parameter (γ, β, θ) using parameter specification of Table 5.1 and sample $X^{(2)}$. Figures A), B), C), D), ratio of posterior probability allocation for cases 1)-9) in Table 5.1. Figure E), F), G), H), ratio of posterior probability allocation for cases 10)-18) in Table 5.1.

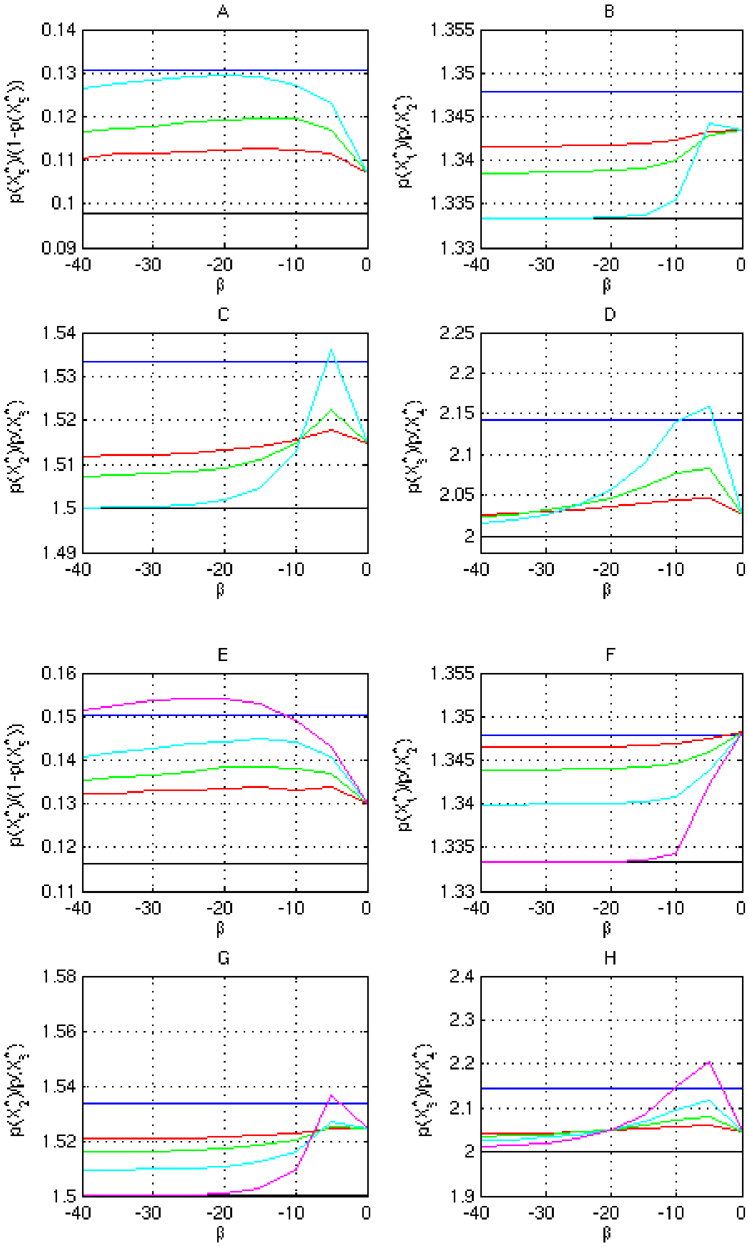

Figure 5.9: Ratio of posterior probability allocation for the Dirichlet process, normalized inverse-Gaussian process, generalized Dirichlet process with parameter γ and generalized Dirichlet process with parameter (γ, β, θ) using parameter specification of Table 5.1 and sample $X^{(2)}$. Figures A), B), C), D), ratio of posterior probability allocation for cases 19)-27) in Table 5.1. Figure E), F), G), H), ratio of posterior probability allocation for cases 28)-36) in Table 5.1.

Case	D	nIG	gD	GD	GD	GD	GD
	a	a	(a,γ)	(a,γ,θ,β)	(a,γ,θ,β)	(a,γ,θ,β)	(a,γ,θ,β)
1	1.93	0.65	(1,2)	(1.71,2,1,-40)	-	-	-
2	1.93	0.65	(1,2)	(1.68,2,1,-35)	-	-	-
3	1.93	0.65	(1,2)	(1.65,2,1,-30)	-	-	-
4	1.93	0.65	(1,2)	(1.62,2,1,-25)	-	-	-
5	1.93	0.65	(1,2)	(1.50,2,1,-20)	-	-	-
6	1.93	0.65	(1,2)	(1.50,2,1,-15)	-	-	-
7	1.93	0.65	(1,2)	(1.40,2,1,-10)	-	-	-
8	1.93	0.65	(1,2)	(1.24,2,1,-5)	-	-	-
9	1.93	0.65	(1,2)	(1,2,1,0)	-	-	-
10	2.77	1.36	(1,3)	(1.36,3,1,-40)	(2.30,3,2,-40)	-	-
11	2.77	1.36	(1,3)	(1.35,3,1,-35)	(2.25,3,2,-35)	-	-
12	2.77	1.36	(1,3)	(1.34,3,1,-30)	(2.19,3,2,-30)	-	-
13	2.77	1.36	(1,3)	1.32,3,1,-25)	(2.11,3,2,-25)	-	-
14	2.77	1.36	(1,3)	(1.30,3,1,-20)	(2,3,2,-20)	-	-
15	2.77	1.36	(1,3)	(1.27,3,1,-15)	(1.86,3,2,-15)	-	-
16	2.77	1.36	(1,3)	(1.23,3,1,-10)	(1.66,3,2,-10)	-	-
17	2.77	1.36	(1,3)	(1.15,3,1,-5)	(1.37,3,2,-5)	-	-
18	2.77	1.36	(1,3)	(1,3,1,0)	(1,3,2,0)	-	-
19	3.515	2.02	(1,4)	(1.23,4,1,-40)	(1.67,4,2,-40)	(2.78,4,3,-40)	-
20	3.515	2.02	(1,4)	(1.23,4,1,-35)	(1.65,4,2,-35)	(2.70,4,3,-35)	-
21	3.515	2.02	(1,4)	(1.22,4,1,-30)	(1.62,4,2,-30)	(2.60,4,3,-30)	-
22	3.515	2.02	(1,4)	(1.21,4,1,-25)	(1.59,4,2,-25)	(2.48,4,3,-25)	-
23	3.515	2.02	(1,4)	(1.20,4,1,-20)	(1.54,4,2,-20)	(2.32,4,3,-20)	-
24	3.515	2.02	(1,4)	(1.18,4,1,-15)	(1.48,4,2,-15)	(2.11,4,3,-15)	-
25	3.515	2.02	(1,4)	(1.15,4,1,-10)	(1.39,4,2,-10)	(1.83,4,3,-10)	-
26	3.515	2.02	(1,4)	(1.10,4,1,-5)	(1.24,4,2,-5)	(1.44,4,3,-5)	-
27	3.515	2.02	(1,4)	(1,4,1,0)	(1,4,2,0)	(1,4,3,0)	-
28	4.19	2.62	(1,5)	(1.17,5,1,-40)	(1.43,5,2,-40)	(1.92,5,3,-40)	(3.16,5,4,-40)
29	4.19	2.62	(1,5)	(1.16,5,1,-35)	(1.42,5,2,-35)	(1.89,5,3,-35)	(3.05,5,4,-35)
30	4.19	2.62	(1,5)	(1.16,5,1,-30)	(1.40,5,2,-30)	(1.85,5,3,-30)	(2.93,5,4,-30)
31	4.19	2.62	(1,5)	(1.15,5,1,-25)	(1.38,5,2,-25)	(1.80,5,3,-25)	(2.76,5,4,-25)
32	4.19	2.62	(1,5)	(1.14,5,1,-20)	(1.36,5,2,-20)	(1.73,5,3,-20)	(2.56,5,4,-20)
33	4.19	2.62	(1,5)	(1.13,5,1,-15)	(1.32,5,2,-15)	(1.64,5,3,-15)	(2.29,5,4,-15)
34	4.19	2.62	(1,5)	(1.10,5,1,-10)	(1.26,5,2,-10)	(1.41,5,3,-10)	(1.93,5,4,-10)
35	4.19	2.62	(1,5)	(1.08,5,1,-5)	(1.17,5,2,-5)	(1.30,5,3,-5)	(1.47,5,4,-5)
36	4.19	2.62	(1,5)	(1,5,1,0)	(1,5,2,0)	(1,5,3,0)	(1,5,4,0)

Table 5.1: Parameters specification for prior processes: Dirichlet (D), normalized inverse-Gausian (nIG) generalized Dirichlet with parameter γ (gD) and generalized Dirichlet process with parameter (γ, β, θ) (GD)

probability assigned to X_j^* is reinforced almost proportionally each time a new tie has been recorded. Increasing $\gamma \in \mathbb{N}$ a reinforced more than proportionally is obtained. For the generalized Dirichlet process with parameter (γ, β, θ) we can observe the same reinforcement mechanism related to the choice of the parameter $\gamma \in \mathbb{N}$. Nevertheless, for a fixed parameter $\gamma \in \mathbb{N}$ we also observe that for the generalized Dirichlet processs with parameter (γ, β, θ) the probability assigned to X_j^* is reinforced more than proportionally each time a new tie has been recorded and such proportionality is controlled through the parameter θ and β. Since it is difficult to specify a priori the reinforcement rate, as suggested by Lijoi et al. [120] it could be reasonable to specify for the generalized Dirichlet process with parameter (γ, β, θ) a prior for the parameter θ and β. Thus, the strength of the reinforcement mechanism could be controlled by the data.

These features of the generalized Dirichlet process with parameter (γ, β, θ) could make it a valid alternative to the Gibbs-type r.p.m.s in the context of Bayesian hierarchical mixture modelling. Actually, we have seen that the generalized Dirichlet process with parameter (γ, β, θ) is characterized by a predictive mechanism that uses more informations from the sample $X_1, \ldots, X_n$ with respect to the predictive mechanism that characterizes the Gibbs-type r.p.m.s. Moreover, the predictive distributions of the generalized Dirichlet process with parameter (γ, β, θ), as the predictive distributions of the Gibbs-type r.p.m., are characterized by a reinforcement which can be controlled by the data through a prior distribution on parameter θ and β.

In order to complete the description of the conditional structure of generalized Dirichlet processes with parameter (γ, β, θ) we now derive the posterior distribution that is the conditional distribution of $\tilde{P}_{(\gamma,\beta,\theta)}$ given a sample $X_1, \ldots, X_n$ featuring k distinct observations, denoted by $X_1^*, \ldots, X^{*_k}$, with frequencies $(n_1, \ldots, n_k)$. By specializing the general results for NRMIs of James et al. [100] reminded in Section 5.2, in the next proposition we provide the desired posterior characterization of both the un-normalized CRM $\tilde{\mu}$ with Poisson intensity measure (5.4.1) and the generalized Dirichlet process with parameter (γ, β, θ).

Proposition 5.4.4. *Let $\tilde{P}_{(\gamma,\beta,\theta)}$ be a generalized Dirichlet process on $\mathbb{X}$ with parameter (γ, β, θ). If α is a non-atomic measure on $(\mathbb{X}, \mathscr{X})$ then, the distribution of $\tilde{\mu}$, given the observations $X_1, \ldots, X_n$ and suitable latent variable U_n, coincides with*

$$\tilde{\xi}^{(U_n,X_1,\ldots,X_n)} \stackrel{d}{=} \tilde{\xi}^{(U_n)} + \sum_{j=1}^{k} J_i^{(U_n,X_1,\ldots,X_n)} \delta_{X_j^*}$$

where

i) *$\tilde{\xi}^{(U_n)}$ is a CRM with Poisson intensity measure*

$$\nu^{(U_n)}(dx, dv) = \sum_{j=1}^{\gamma} \frac{e^{-v(j-\beta \mathbb{1}_{\{j>\gamma-\theta\}}+U_n)}}{v} dv \alpha(dx) \tag{5.4.12}$$

ii) *X_j^* are fixed points of discontinuity, for $j = 1, \ldots, k$, and the r.v.s $J_j^{(U_n,X_1,\ldots,X_n)}$'s are the corresponding jumps which are absolutely continuous w.r.t. to the Lebesgue measure with density*

$$f_{J_j^{(U_n,X_1,\ldots,X_n)}}(v) \propto v^{n_i-1} \sum_{j=1}^{\gamma} e^{-v(j-\beta \mathbb{1}_{\{j>\gamma-\theta\}}+U_n)} \qquad j = 1, \ldots, k \tag{5.4.13}$$

iii) *the jumps $J_j^{(U_n,X_1,\ldots,X_n)}$, for $j = 1, \ldots, k$, are mutually independent and independent from $\tilde{\xi}^{(U_n)}$.*

Moreover, the latent variable U_n*, given* $X_1, \ldots, X_n$*, is absolutely continuous w.r.t. the Lebesgue measure with density*

$$f_{U_n}^{(X_1,\ldots,X_n)}(u) \propto \frac{u^{n-1}\prod_{i=1}^k \Gamma(n_i)\sum_{j=1}^{\gamma}(j-\beta\mathbb{1}_{\{i>\gamma-\theta\}}+u)^{-n_i}}{\prod_{i=1}^{\gamma}(i-\beta\mathbb{1}_{\{i>\gamma-\theta\}}+u)^{a}}. \tag{5.4.14}$$

Finally, the posterior distribution of $\tilde{P}_{(\gamma,\beta,\theta)}$*, given* $X_1, \ldots, X_n$ *and* U_n*, is again a NRMI (with fixed points of discontinuity) and coincides in distribution with*

$$w\frac{\tilde{\mu}^{(U_n)}}{\tilde{\mu}^{(U_n)}(\mathbb{X})} + (1-w)\frac{\sum_{j=1}^k J_j^{(U_n,X_1,\ldots,X_n)}\delta_{X_j^*}}{\sum_{j=1}^k J_j^{(U_n,X_1,\ldots,X_n)}} \tag{5.4.15}$$

where $w = \tilde{\xi}^{(U_n)}(\mathbb{X})(\tilde{\mu}^{(U_n)}(\mathbb{X}) + \sum_{1\le j\le k} J_j^{(U_n,X_1,\ldots,X_n)})^{-1}$.

Proof. Since $\gamma \in \mathbb{N}$ and $\theta \in \mathbb{N}_0$, the Poisson intensity measure (5.4.1) of $\tilde{\mu}$ reduces to the following Poisson intensity measure

$$\nu(ds,dx) = \sum_{j=1}^{\gamma} \frac{\mathrm{e}^{-s(j-\beta\mathbb{1}_{\{j>\gamma-\theta\}})}}{s} ds\alpha(dx).$$

Now since the generalized Dirichlet process with parameter (γ, β, θ) is a NRMI and, by hypothesis α is non-atomic, we can apply Theorem 5.2.1 we ensues the existence of a latent variable U_n, such that the distribution of $\tilde{\mu}$, given $X_1, \ldots, X_n$ and U_n coincides with the distribution of

$$\tilde{\mu}^{(U_n)} + \sum_{j=1}^k J_j^{(U_n,X_1,\ldots,X_n)}\delta_{X_j^*}$$

where $\tilde{\mu}^{(U_n)}$ is a suitably updated CRM and the $J_j^{(U_n,X_1,\ldots,X_n)}$'s are absolutely continuous with density expressed in the terms of the Poisson intensity measure of $\tilde{\mu}$. It is then straightforward to show that the Poisson intensity measure associated to $\tilde{\mu}^{(U_n)}$ is of the form (5.4.12) and that the density of $J_j^{(U_n,X_1,\ldots,X_n)}$ is given by (5.4.13). In order to derive the density function for conditional distribution of U_n, given $X_1, \ldots, X_n$ we resort to Proposition 5.2.1 and, after some algebra, we obtain (5.4.14). Given this, the characterization of the posterior distribution of $\tilde{P}$ in (5.4.15) follows from Theorem 5.2.2. □

Despite the fact that the previous result completes the theoretical analysis of the conditional structure induced by generalized Dirichlet processes with parameter (γ, β, θ), it is also useful for practical purposes. Indeed one can devise a simulation algorithm relying on the posterior characterization of Proposition 6.3.4. By combining an inverse Lévy measure algorithm, such as the Ferguson-Klass method (see Ferguson and Klass [64] and Walker and Damien [191]), for simulating trajectories of $\tilde{\xi}^{(U_n)}$ with a Metropolis-Hasting step for drawing samples from $U_n^{(X_1,\dots,X_n)}$, one easily obtains realizations of the posterior distribution of the generalized Dirichlet process with parameter (γ, β, θ).

Finally note that Proposition 6.3.4 is also important in the context of mixture modeling, where inference is necessarily simulation based given the complexity of the models: in fact, it allows to derive, in the terminology of Papaspiliopoulos and Roberts [148], conditional sampling schemes, which in the case of the generalized Dirichlet process with parameter (γ, β, θ), are simpler to implement than marginal ones.

5.5 Superposition of beta-Stacy processes

In this section we review and we provide some developments of the beta-Stacy process whose characterization and essential properties were presented by Walker and Muliere [194]. In particular, we use the idea of superposition of independent Gamma process in order to provide a constructive definition of a new class of neutral to the right r.p.m.s which includes the beta-Stacy process as particular case.

Let $\mathcal{F}_{\mathbb{R}^+}$ be the space of cumulative distribution functions on $\mathbb{R}^+$ endowed with the σ-field $\mathscr{F}_{\mathbb{R}^+}$ generated by the Skorohod topology $\mathcal{S}$ which makes $(\mathcal{F}_{\mathbb{R}^+}, \mathcal{S})$ a Polish space. In this section we place a probability distribution on $(\mathcal{F}_{\mathbb{R}^+}, \mathscr{F}_{\mathbb{R}^+})$ by defining a process $\{F_t, t \geq 0\}$ on $\mathbb{R}^+$, such that $F_0 = 0$ a.s., $\{F_t, t \geq 0\}$ is non-decreasing a.s., right continuous a.s. and $F_t \to 1$ a.s. as $t \to +\infty$ Thus, with probability 1, the sample paths of the process $\{F_t, t \geq 0\}$ are cumulative distribution functions. The paper is restricted to cumulative distribution functions on $\mathbb{R}^+$, although it is trivially extended to include cumulative distribution functions on $\mathbb{R}$.

First of all we recall the definition of Lévy process. Let $\{Z_t, t \geq 0\}$ be a Lévy process such that the following holds:

i) $\{Z_t, t \geq 0\}$ has non-negative independent increments;

ii) $\{Z_t, t \geq 0\}$ in non-decreasing a.s.;

iii) $\{Z_t, t \geq 0\}$ is right continuous a.s.;

iv) $Z_t \to +\infty$ a.s. as $t \to +\infty$;

v) $Z_0 = 0$ a.s.

For such a process there exist at most countably many fixed points of discontinuity at time points $\{t_k, k \geq 1\}$ with jumps $\{S_k, k \geq 1\}$, independent non-negative r.v.s. If for all $t \geq 0$ we set $X_t = Z_t - \sum_{t_k \leq t} S_k$, then $\{X_t, t \geq 0\}$ is a non-decreasing process with independent increments and with no fixed points of discontinuity such that for any $t \geq 0$, the Lévy-Khintchine representation for the Laplace-Stieltjes transform of the distribution function of X_t such that for any $t \geq 0$

$$\log(\mathbb{E}[\mathrm{e}^{-sX_t}]) = -sb(t) - \int_{\mathbb{R}^+} (1 - \mathrm{e}^{-sv})\nu_t(dv)$$

where b is non-decreasing and continuous, with $b(0) = 0$, and $\nu_t(\cdot)$ is a Lévy measure satisfying

i) for any Borel set B, $\nu_t(B)$ is continuous and non-decreasing;

ii) for any real $t > 0$, ν_t is a measure on the Borel sets of $\mathbb{R}^+$;

iii) $\int_{\mathbb{R}^+} v(1+v)^{-1}\nu_t(dv) < +\infty$;

iv) $\int_{\mathbb{R}^+} v(1+v)^{-1}\nu_t(dv) \to 0$ as $t \to +\infty$.

Since b represents a non-random component it is not considered and we assume it to be identically zero.

Let c be a positive function, let $G \in \mathcal{F}_{\mathbb{R}^+}$ be right continuous and let $\{t_k, k \geq 1\}$ be a countable set of points of discontinuity of G, that is, $G\{t_k\} = G(t_k) - G(t_k^-) > 0$ for all k. Now put $G_C = G(t) - \sum_{t_k \leq t} G\{t_k\}$ so that G_C is continuous.

Definition 5.5.1. *(cfr. Walker and Muliere [194]) $\{F_t, t \geq 0\}$ is a beta-Stacy process on $\mathbb{R}^+$ with parameters c and G, if, for all $t \geq 0$, $F_t = 1 - e^{-Z_t}$, where $\{Z_t, t \geq 0\}$ is a Lévy process with Lévy measure, for $v > 0$, given by*

$$\nu_t(dv) = \frac{1}{1 - e^{-v}} \int_{\mathbb{R}^+} e^{-vc(s)G(s,+\infty)} c(s) dG_C(s) dv \tag{5.5.1}$$

and Laplace-Stieltjes transform such that

$$\log(\mathbb{E}[e^{-sZ_t}]) = \sum_{t_k \leq t} \log(\mathbb{E}[e^{-sS_k}]) - \int_{\mathbb{R}^+} (1 - e^{-sv})\nu_t(dv) \tag{5.5.2}$$

where, for any $k \in \mathbb{N}$, $1 - e^{-S_k}$ is distributed according to a Beta distribution function with parameter $(c(t_k)G\{t_k\}, c(t_k)G[t_k, +\infty))$.

Since the function $G(t) \to 1$, as $t \to +\infty$, $F_t \leq 1$ a.s. and $\{F_t, t \geq 0\}$ is a.s non-decreasing, it follows that $F_t \to 1$ a.s. Then it follows from Lemma 1 in Walker and Muliere [194] that $\{F_t, t \geq 0\}$ belongs to $\mathcal{F}_{\mathbb{R}^+}$ a.s. and it is a neutral to the right r.p.m. In addition, as the Lévy process $\{Z_t, t \geq 0\}$ has non-random component, $\{Z_t, t \geq 0\}$ increase only in jumps a.s. and so $\{F_t, t \geq 0\}$ is with probability 1 a discrete member of $\mathcal{F}_{\mathbb{R}^+}$.

According to Ferguson and Phadia [65] the fundamental result in Doksum [25] on the posterior of a neutral to the right r.p.m. is that $\{F_t, t \geq 0\}$ conditionally given X, is also a neutral to the right r.p.m. for any observation of the type $X = x$ or $X > x$, where X is a sample from $\{F_t, t \geq 0\}$. In particular, in Walker and Muliere [194] it is shown that if $\{F_t, t \geq 0\}$ is a beta-Stacy process then $\{F_t, t \geq 0\}$ conditionally given X is a also a beta-Stacy process and hence the conjugacy property of the beta-Stacy process.

The following remarks emphasized the relation between the beta-Stacy process and the Dirichlet process whose characterization and properties were presented by Ferguson [61] and Ferguson [62] and the relation between the beta-Stacy process beta process introduced by Hjort [84].

Remark 5.5.1. *(cfr. Walker and Muliere [194]) The beta-Stacy process generalizes the Dirichlet process, which can be seen more easily if G is taken to be continuous, since, if $c(s) = c > 0$ for all $s \geq 0$, then* (5.5.1) *becomes*

$$\nu_t(dv) = \frac{1}{(1-e^{-v})}(e^{-vc(1-G(t))} - e^{-vc}) \tag{5.5.3}$$

the Lévy measure given in Ferguson [62] which represents the Lévy measure characterizing the Lévy process corresponding to the Dirichlet process when viewed as a neutral to the right r.p.m..

Remark 5.5.2. *(cfr. Walker and Muliere [194]) If $\{A_t, t \geq 0\}$ is the beta process introduce in Hjort [84] and for any $t \geq 0$, $dZ_t = -\log(1 - dA_t)$, then $F_t = 1 - e^{-Z_t}$ is the beta-Stacy process.*

It is important to use any available prior information to center the process $\{F_t, t \geq 0\}$ and express uncertainty in $\{F_t, t \geq 0\}$ about this centering, that is, to assign arbitrarily $\mathbb{E}[F_t]$ and $Var(F_t)$. As suggest in Walker and Muliere [194] this prior specification can be done by considering the first two moments of $S_t = 1 - F_t$ for any $t \geq 0$. Using the Lévy representation of a beta-Stacy process, it follows

$$\mu_t := -\log(\mathbb{E}[S_t]) = \int_0^{+\infty} \int_{(0,t)} \mathrm{e}^{-v\varrho_s} d\gamma_s dv$$

and

$$\lambda_t := -\log(\mathbb{E}[S_t^2]) = \int_0^{+\infty} \int_{(0,t)} \left(\frac{1-\mathrm{e}^{-2v}}{1-\mathrm{e}^{-v}}\right) \mathrm{e}^{-v\varrho_s} d\gamma_s dv$$

where we defined

$$\varrho_s := c(s)G[s, +\infty]$$

and

$$\gamma_s := \int_{(0,s)} c(t) dG_C(t) + \sum_{s_k \leq s} c(s_k) G\{s_k\}.$$

Note it is necessary that $0 < \mu_t < \lambda_t < 2\mu_t$, which corresponds to $\mathbb{E}[S_t])^2 < \mathbb{E}[S_t^2] < \mathbb{E}[S_t]$. The first of these condition is satisfied when

$$\mu_t = \int_{(0,t)} \frac{d\gamma_s}{\varrho_s}$$

that is, when

$$d\gamma_t = \varrho_t d\mu_t.$$

The second condition becomes, using transformation $u = 1 - \mathrm{e}^{-v}$

$$\lambda_t = \int_{(0,t)} \left(2 - \frac{1}{1+\varrho_s}\right) d\mu_s$$

leading to

$$\frac{d\lambda_t}{d\mu_t} = 2 - \frac{1}{1+\varrho_t}$$

and hence the solution.

In Walker and Muliere [194], the beta-Stacy process is derived as the limit in distribution of a sequence of discrete time neutral to the right r.p.m.s defined as follows. Let $G \in \mathcal{F}_{\mathbb{R}^+}$ be continuous, let c be a piecewise continuous positive function and for any fixed $n \geq 1$ define two sequences of positive real numbers $\gamma_\bullet := \{\gamma_{n,k}, k \geq 1\}$ and $\varrho_\bullet := \{\varrho_{n,k}, k \geq 1\}$ where $\gamma_{n,k} := c_{n,k} G[(k-1/n), k/n)$, $\varrho_{n,k} := c_{n,k} G[k/n, +\infty)$ and $c_{n,k} := c((k-1/2)/n)$. Finally, for any fixed $n \geq 1$ define the sequence of dependent r.v.s $\{Y_{n,k}, k \geq 1\}$ by $Y_{n,k} = V_{n,k} \prod_{1 \leq i \leq k-1} (1 - V_{n,i})$ where $\{V_{n,k}, k \geq 1\}$ is a sequence of independent r.v.s distributed according to a Beta distribution function with parameter $(\gamma_{n,k}, \varrho_{n,k})$. Then, if we set

$$F_{n,t} = \sum_{k/n \leq t} Y_{n,k} \qquad t \geq 0 \tag{5.5.4}$$

it can be proved (see Walker and Muliere [194]) that $\{F_{n,t}, t \geq 0\}$ is a discrete time neutral to the right r.p.m. indexed by n which converges in distribution to the beta-Stacy process $\{F_t, t \geq 0\}$, as $n \to \infty$. For any fixed $n \geq 1$, the process $\{F_{n,t}, n \geq 1\}$ is usually known as discrete time beta-Stacy process.

In general, given a sequence of time points $\{t_k, k \geq 1\}$ in $\mathbb{R}^+$ and given two sequence of positive real numbers $\gamma_\bullet$ and $\varrho_\bullet$ we have the following definition of discrete time beta-Stacy process with jumps at $\{t_k, k \geq 1\}$ and parameter $(\gamma_\bullet, \varrho_\bullet)$.

Definition 5.5.2. *Let $\{t_k, k \geq 1\}$ be sequence of time points in $\mathbb{R}^+$ and let $\gamma_\bullet$ and $\varrho_\bullet$ be two sequences of positive real numbers. The process $\{F_t, t \geq 0\}$ is a discrete time beta-Stacy process with jumps at $\{t_k, k \geq 1\}$ and parameter $(\gamma_\bullet, \varrho_\bullet)$ if*

$$F_t = \sum_{t_k \leq t} Y_k \qquad t \geq 0$$

and the random weights are defined by

$$Y_k = V_k \prod_{i=1}^{k-1}(1-V_i) \qquad k \geq 1$$

where $\{V_k, k \geq 1\}$ is a sequence of independent r.v.s distributed according to a Beta distribution function with parameter (γ_k, ϱ_k) for $k \geq 1$.

From Definiton 5.5.2 we observe that the sequence of random weights $\{Y_k, k \geq 1\}$ completely characterized the discrete time beta-Stacy process. In particular, using the theory of partially exchangeable random partitions introduced by Pitman [149] (see Appendix B), we can give an interesting characterization of the sequence $\{Y_k, k \geq 1\}$ characterizing a discrete time beta-Stacy process.

Theorem 5.5.1. *Let $\gamma_\bullet$, $\varrho_\bullet$ be two sequences of positive real numbers and let $\{Y_j, j \geq 1\}$ be the sequence of the random weights characterizing a discrete time beta-Stacy process with parameter $(\gamma_\bullet, \varrho_\bullet)$. Let $\{\Pi_n, n \geq 1\}$ be a sequence of partially exchangeable random partitions with frequncies $\{N_n, n \geq 1\}$. Suppose that $\{\Pi_n, n \geq 1\}$ is characterized by the prediction rules*

$$\mathbb{P}(\Pi_{n+1} = \mathbf{n}_j^+ | n_1, \dots, n_k) = \frac{\gamma_j + n_j - 1}{\gamma_1 + \varrho_1 + n - 1} \prod_{i=1}^{j-1} \frac{\varrho_i + \sum_{l=i+1}^k n_l}{\gamma_{i+1} + \varrho_{i+1} + \sum_{l=i+1}^k n_l - 1} \tag{5.5.5}$$

and

$$\mathbb{P}(\Pi_{n+1} = \mathbf{n}^+ | n_1, \dots, n_k) = \frac{\varrho_k}{\gamma_1 + \varrho_1 + n - 1} \prod_{i=1}^{k-1} \frac{\varrho_i + \sum_{l=i+1}^k n_l}{\gamma_{i+1} + \varrho_{i+1} + \sum_{l=i+1}^k n_l - 1} \tag{5.5.6}$$

where $\mathbf{n}_j^+ := (n_1, \dots, n_j + 1, \dots, n_k)$ and $\mathbf{n}^+ := (n_1, \dots, n_k, 1)$. Then, the vector of frequencies N_n/n converges in distribution to the sequence $\{Y_j, j \geq 1\}$.

Proof. By repeated applcation of the prediction rules (5.5.5) and (5.5.6) it can be easily checked that the probability distribution of any sample realization having $(n_1, \dots, n_k)$ as vector of frequencies is

$$p_k^{(n)}(n_1, \dots, n_k) = \prod_{j=1}^k \frac{(\gamma_j)_{(n_j-1)\uparrow 1} (\varrho_j)_{(\sum_{l=j+1}^k n_l)\uparrow 1}}{(\gamma_j + \varrho_j)_{(-1+\sum_{l=j}^k n_l)\uparrow 1}}.$$

Nevertheless, we can write the last equation as

$$p_k^{(n)}(n_1, \dots, n_k) = \mathbb{E}\left[\prod_{j=1}^k V_j^{n_j-1}(1-V_j)^{\sum_{l=j+1}^k n_l}\right] \tag{5.5.7}$$

where $\{V_j, j \geq 0\}$ a sequence of independent r.v.s such that $V_0 = 1$ a.s. and each V_j is distributed according to a Beta distribution function with parameter (γ_j, ϱ_j) for $j \geq 1$ If we define $Y_j := V_j \prod_{1 \leq i \leq j-1}(1 - V_i)$ for $j \geq 1$ we can see that (5.5.7) is equal in distribution to

$$p_k^{(n)}(n_1, \ldots, n_k) = \mathbb{E}\left[\prod_{j=1}^{k} Y_j^{n_j - 1}\left(1 - \sum_{i=1}^{j-1} Y_i\right)\right].$$

Then, by Theorem 6 in Pitman [149] we have that $\{Y_i, i \geq 1\}$ is the limit frequencie of N_n/n as $n \to +\infty$. □

A straightforward consequence of Theorem 5.5.1 is the following characterization for the sequence of random weights $\{Y_j, j \geq 1\}$ characterizing a discrete time beta-Stacy process.

Corollary 5.5.1. *Let $\gamma_\bullet$, $\varrho_\bullet$ be two sequences of positive real numbers and let $\{Y_j, j \geq 1\}$ be the sequence of the random weights characterizing a discrete time beta-Stacy process with parameter $(\gamma_\bullet, \varrho_\bullet)$. The predictive distribution of the partially exchangeable random partition induced by $\{Y_j, j \geq 1\}$ does not depend on the censored information if and only if, for any j*

$$\varrho_j = \gamma_{j+1} + \varrho_{j+1} - 1. \tag{5.5.8}$$

Proof. If (5.5.8) holds, then the term between brackets in equation (5.5.5) and equation (5.5.6) simplify to 1. Indeed equation (5.5.8) is the only condition for which the term between brackets reduce to a constant idependent of $\sum_{j+1 \leq i \leq k} n_i$, for $j = 1, \ldots, k-1$ and for each choice of $n_1, \ldots, n_k$. □

Let us consider

$$q_{(\gamma_\bullet, \varrho_\bullet)} := \prod_{j=1}^{k} \frac{(\gamma_j)_{(n_j - 1)\uparrow 1}(\varrho_j)_{(\sum_{l=j+1}^{k} n_l)\uparrow 1}}{(\gamma_j + \varrho_j)_{(-1 + \sum_{l=j}^{k} n_l)\uparrow 1}}$$

and observe that when (5.5.8) holds, then the predictive probabilities of the partially exchangeable random partition do not depend on the censored observations. This does not mean that the prediction of a multinomial (exchangeable) sample from a discrete time beta-Stacy process do not depend on the censored observation. In fact, the latter differs from $q_{(\gamma_\bullet, \varrho_\bullet)}$ because it is the expectation of a different functional

$$\mathbb{E}\left[\prod_{j \geq 1} Y_j^{n_j}\right] = \prod_{j \geq 1} \frac{(\gamma_j)_{(n_j - 1)\uparrow 1}(\varrho_j)_{(\sum_{l=j+1}^{k} n_l)\uparrow 1}}{(\gamma_j + \varrho_j)_{(-1 + \sum_{l=j}^{k} n_l)\uparrow 1}}$$

and in this case the only distribution not depending on the censored observations is, as we know, the Dirichlet process, for which $\varrho_j = \gamma_{j+1} + \varrho_{j+1}$. Of course, an example where the condition (5.5.8) is met is the two-parameter GEM distribution with parameter (α, θ) (see Pitman [149]), where $\varrho_j = \theta + \alpha j = \gamma_{j+1} + \varrho_{j+1} - 1$. In this (and only this) case, we have that $q_{(\gamma_\bullet, \varrho_\bullet)}$ is a symmetric function of its arguments, i.e. the partition Π is also exchangeable. If $\alpha = 0$ (one-parameter GEM distribution with parameter θ), the the predictive probabilities do not even depend on k.

It would be interesting to say more about this aspect. If we replace exchangeability condition with Pitman's partial exchangeability condition we see that all Beta-Stacy random discrete distribution satisfying equation (5.5.8) have a prediction rule such that the probability of next observing a certain category depends on how many times such a category has been observed in the past and on how many distinct categories were observed (k). This extend Zabell's considerations on Johnson's sufficientness postulate (see Zabell [198]) and the two parameter Poisson-Dirichlet distribution (see Pitman [149]). Let's call this property the k-Johnson's sufficientness postulate property. In particular, it would be interesting to know if the discrete time beta-Stacy process is the only class of random discrete distribtuions arising from partially exchangeable random partitions with the k-Johnson's sufficientness postulate property.

5.5.1 The discrete time superposed beta-Stacy process

We define and investigate a new neutral to the right r.p.m. termed superposed beta-Stacy process. As for the beta-Stacy process our definition is constructive and it follows along lines the construction of a neutral to the right r.p.m. proposed by Doksum [25].

For any $m \in \mathbb{N}$ let us consider m sequences of positive real numbers $(\alpha_{1,\bullet}, \beta_{1,\bullet}) := \{(\alpha_{1,k}, \beta_{1,k}), k \geq 1\}, \dots, (\alpha_{m,\bullet}, \beta_{m,\bullet}) := \{(\alpha_{m,k}, \beta_{m,k}), k \geq 1\}$ and m independent sequences of r.v.s $Y_{1,\bullet} := \{Y_{1,k}, k \geq 1\}, \dots, Y_{m,\bullet} := \{Y_{m,k}, k \geq 1\}$ such that $Y_{i,\bullet}$ is a sequence of independent r.v.s distributed according to a Beta distribution function with parameter $(\alpha_{i,k}, \beta_{i,k})$ for $i = 1, \dots, m$. Define the sequence of r.v.s $\{X_k | X_1, \dots, X_{k-1}, k \geq 1\}$ via the following construction

$$
\begin{aligned}
X_1 &\stackrel{d}{=} 1 - \prod_{i=1}^{m}(1 - Y_{i,1}) \\
X_2 | X_1 &\stackrel{d}{=} (1 - X_1)\left(1 - \prod_{i=1}^{m}(1 - Y_{i,2})\right) \\
&\vdots \\
X_k | X_1, \dots, X_{k-1} &\stackrel{d}{=} (1 - F_{k-1})\left(1 - \prod_{i=1}^{m}(1 - Y_{i,k})\right)
\end{aligned}
\tag{5.5.9}
$$

where

$$F_k := \sum_{j=1}^{k} X_j \tag{5.5.10}$$

with the proviso $X_1 := X_1|X_0$. For any $k \geq 1$, the distribution function of the r.v. $X_k|X_1,\ldots,X_{k-1}$ can be computed by using Theorem 7 in Springer and Thompson [176]; in particular, it can be checked that the r.v. $X_k|X_1,\ldots,X_{k-1}$ is distributed according to a distribution function on $(0,1)$ which admits probability density function absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$ given by

$$f_{X_k|X_1,\ldots,X_{k-1}}(x_k|x_1,\ldots,x_{k-1}) = \frac{1}{1-\sum_{j=1}^{k-1} x_j} \prod_{i=1}^{m} \frac{\Gamma(\alpha_{i,k}+\beta_{i,k})}{\Gamma(\beta_{i,k})}$$
$$\times G^{n,0}_{n,0}\left(\frac{1-x_k}{1-\sum_{j=1}^{k-1} x_j} \middle| \begin{array}{c} \alpha_{1,k}+\beta_{1,k}-1,\ldots,\alpha_{m,k}+\beta_{m,k}-1 \\ \beta_{1,k}-1,\ldots,\beta_{m,k}-1 \end{array} \right) \mathbb{1}_{(0,1)}(x_k)$$

where $G^{l,m}_{p,q}$ is the Meijer G-function (see Appendix C).

The construction (5.5.9) is similar to the construction proposed in Waker and Muliere [194] for the discrete time beta-Stacy process and it generalizes it by nesting for any $k > 0$ the product of independent r.v.s distributed according to a Beta distribution function. In particular, in Walker and Muliere [194] for any $k \geq 1$ the r.v. $X_k|X_1,\ldots,X_{k-1}$ is distributed according to a beta-Stacy distribution function with parameter $(\alpha_k, \beta_k, 1 - F_{k-1})$, i.e. for any $k \geq 1$ it follows that $X_k|X_1,\ldots,X_{k-1} \stackrel{d}{=} (1-F_{k-1})Y_k$ where Y_k belongs to the sequence of independent r.v.s distributed according to a Beta distribution function with parameter (α_k, β_k). Obviously, the construction proposed by Walker and Muliere [194] can be recovered from the construction (5.5.9) by setting $m = 1$.

It can be checked that under the construction (5.5.9), $X_k < 1 - F_{k-1}$ a.s.-$\mathbb{P}$, so that $F_k < 1$ a.s.-$\mathbb{P}$. In particular, we can consider the following technical result for the sequence $\{F_k, k \geq 1\}$ based on the construction (5.5.9).

Theorem 5.5.2. *Let $\{t_k, k \geq 0\}$ be a sequence of time points in $\mathbb{R}^+$ with $t_0 := 0$ and from construction* (5.5.9) *let $\{F_t, t \geq 0\}$ defined by $F_t := \sum_{t_k \leq t} X_k$ for any $t \geq 0$. If $F_0 = 0$ and $\prod_{k\geq 1}\prod_{1\leq i\leq m} \beta_{i,k}/\alpha_{i,k}+\beta_{i,k} = 0$, then the sample paths of $\{F_t, t \geq 0\}$ belong to $\mathcal{F}_{\mathbb{R}^+}$ a.s.-$\mathbb{P}$.*

Proof. Let $\{F_k, k \geq 1\}$ be the sequence of r.v.s defined by (5.5.10). In particular, from the construction (5.5.9) we have $\mathbb{E}[F_k] = 1 - \prod_{1\leq i\leq m} \beta_{i,k}/\alpha_{i,k} + \beta_{i,k} + (\prod_{1\leq i\leq m} \beta_{i,k}/\alpha_{i,k} + \beta_{i,k})\mathbb{E}[F_{k-1}]$ such that

$$\prod_{k\geq 1} \frac{1-\mathbb{E}[F_k]}{1-\mathbb{E}[F_{k-1}]} = 0$$

which implies that $\mathbb{E}[F_k] \to 1$. Moreover, we have that $\{F_k, k \geq 1\}$ is a sequence of non-negative r.v.s such that $F_k \geq 1$ a.s.-$\mathbb{P}$ for any $k \geq 1$ and $\{F_k, k \geq 0\}$ is non-decreasing a.s.-$\mathbb{P}$. This implies that $F_k \to 1$ a.s.-$\mathbb{P}$ and the proof in completed. □

Theorem 5.5.2 implies that the random process $\{F_t, t \geq 0\}$ is a discrete time neutral to the right r.p.m. according to the definition given by Doksum [25]. We term the random process $\{F_t, t \geq 0\}$ discrete time superposed beta-Stacy process. Here, the formal definition of discrete time superposed beta-Stacy process.

Definition 5.5.3. *Let $\{X_k | X_1, \ldots, X_{k-1}, k \geq 1\}$ be a sequence of r.v.s defined via construction* (5.5.9) *and let $\{t_k, k \geq 0\}$ be a sequence of time points in $\mathbb{R}^+$ with $t_0 := 0$. The random process $\{F_t, t \geq 0\}$ defined by $F_t := \sum_{t_k \leq t} X_k$ and satisfying conditions of Lemma 5.5.2 is a discrete time superposed beta-Stacy process with parameter $(m, (\alpha_{1,\bullet}, \beta_{1,\bullet}), \ldots, (\alpha_{m,\bullet}, \beta_{m,\bullet}))$ and jumps at $\{t_k, k \geq 0\}$.*

From Definition 5.5.3 it follows that the discrete time superposed beta-Stacy process includes as particular case the discrete time beta-Stacy process which can be recovered by setting $m = 1$. Moreover, using some known properties for the product of independent r.v.s distrubuted according to a Beta distribution function, further relations between the discrete time superposed beta-Stacy process and the discrete time beta-Stacy process can be considered. We provide two remarks: the first one provides conditions in order that a discrete time superposed beta-Stacy process is a discrete time beta-Stacy process while the second one provides a possible approximation of the discrete time superposed beta-Stacy process via a discrete time beta-Stacy process.

Remark 5.5.3. *From the construction* (5.5.9) *and by using Theorem 1 in Jambunathan [95] it is immediate to check that a discrete time superposed beta-Stacy process with parameter $(m, (\alpha_{1,\bullet}, \beta_{1,\bullet}), (\beta_{2,\bullet}, \alpha_{1,\bullet} + \beta_{1,\bullet}), \ldots, (\beta_{m,\bullet}, \alpha_{m-1,\bullet} + \beta_{m-1,\bullet}))$ is a discrete time beta-Stacy process with parameter $(\beta_{1,\bullet} + \cdots + \beta_{m,\bullet}, \alpha_{1,\bullet})$.*

Remark 5.5.4. *From the construction* (5.5.9) *and by using Theorem 1 in Fan [49] a discrete time superposed beta-Stacy process with parameter $(m, (\alpha_{1,\bullet}, \beta_{1,\bullet}), \ldots, (\alpha_{m,\bullet}, \beta_{m,\bullet}))$ can be approximated by a discrete time beta-Stacy process with parameter $((1 - S_\bullet)(S_\bullet - T_\bullet)/(T_\bullet - S_\bullet^2), S_\bullet/(T_\bullet - S_\bullet^2))$ where*

$$S_\bullet := \prod_{i=1}^{m} \frac{\beta_{i,\bullet}}{\beta_{i,\bullet} + \alpha_{i,\bullet}}$$

and

$$T := \prod_{i=1}^{m} \frac{\beta_{i,\bullet}(\beta_{i,\bullet} + 1)}{(\beta_{i,\bullet} + \alpha_{i,\bullet})(\beta_{i,\bullet} + \alpha_{i,\bullet} + 1)}.$$

By the construction (5.5.9) and by using the integral representation in Wilks [197] for the product of independent r.v.s distributed according to a Beta distribution functions, it can be checked that for any $s \in \mathbb{N}$ the r.v. $(X_1, \dots, X_s)$ is distributed according to a distribution function on the s-dimensional symplex $\Delta^{(s)}$ which admits a probability density function absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{s+1}$ given by

$$
\begin{aligned}
f_{(X_1,\dots,X_s)}(x_1,\dots,x_s) \propto \prod_{j=1}^{s} & \frac{x_j^{\delta_{0,j}-1}(1-\sum_{l=1}^{j} x_l)^{\beta_{m,j}-1}}{(1-\sum_{l=1}^{j-1} x_l)^{\delta_{0,j}+\beta_{m,j}-1}} \\
& \times \int_{(0,1)^{m-1}} \prod_{i=1}^{m-1} w_{i,j}^{\alpha_{i,j}}(1-w_{i,j})^{\delta_{i,j}-1}\left(1-\frac{x_j z_{i,j}}{1-\sum_{l=1}^{j-1} x_l}\right)^{c_{i,j}} dw_{i,j} \\
& \times \mathbb{1}_{\Delta^{(s)}}(x_1,\dots,x_s)
\end{aligned}
\tag{5.5.11}
$$

where $\delta_{i,j} := \sum_{i+1 \le l \le m} \alpha_{l,j}$, $c_{i,j} := -\beta_{i+1,j} - \alpha_{i+1,j} + \beta_{i,j}$ and $z_{i,j} := 1 - \prod_{1 \le l \le i}(1 - w_{l,j})$. In particular, from (5.5.11) it can be checked that for any $k \ge 1$ the r.v.s $X_1, X_2/(1 - F_1), \dots, X_k/(1-F_{k-1})$ are independent and such that $X_k/(1-F_{k-1}) \stackrel{d}{=} 1-\prod_{1 \le i \le m}(1-Y_{i,k})$ for $k = 1, \dots, s$.

We now consider the following inference framework. Let $\{t_k, k \ge 0\}$ be a sequence of time points in $\mathbb{R}^+$ with $t_0 := 0$ defining a partition of $\mathbb{R}^+$ and for any $n \in \mathbb{N}$ let $T_1, \dots, T_n$, with each $T_i \in \{t_k, k \ge 0\}$, be and random sample, possible with right censoring (with T_i being the censoring time if applicable), from a random cumulative distribution function F on $\mathbb{R}^+$ governed by a discrete time superposed beta-Stacy process with parameter $(m, (\alpha_{1,\bullet}, \beta_{1,\bullet}), \dots, (\alpha_{m,\bullet}, \beta_{m,\bullet}))$ and jumps at $\{t_k, k \ge 1\}$. The likelihood function, assuming that there are no censoring times or exact observation for $t > t_L$, is given by

$$
\mathscr{L}(t_1,\dots,t_L) \propto \prod_{i=1}^{L} t_i^{n_i}\left(1-\sum_{j=1}^{i} t_j\right)^{r_i} \mathbb{1}_{\Delta^{(L)}}(t_1,\dots,t_L)
$$

where n_k is the number of exact observation at t_k and r_k is the number of censoring times at t_k with $\sum_{1 \le i \le L} n_i + r_i = n$.

Proposition 5.5.1. *Let $\{t_k, k \ge 0\}$ be a sequence of time points in $\mathbb{R}^+$ with $t_0 := 0$ and for any $n \in \mathbb{N}$, let $T_1, \dots, T_n$ be random sample such that $T_i \in \{t_k, k \ge 0\}$ possibly with right censoring, from a random cumulative distribution function $\{F_t, t \ge 0\}$ on $\mathscr{F}_{\mathbb{R}^+}$. If $\{F_t, t \ge 0\}$ is a discrete time superposed beta-Stacy process with parameter $(m, (\alpha_{1,\bullet}, \beta_{1,\bullet}), \dots, (\alpha_{m,\bullet}, \beta_{m,\bullet}))$ and jumps at $\{t_k, k \ge 0\}$, then $\{F_t | T_1, \dots, T_n, t \ge 0\}$ is a discrete time neutral to the right r.p.m. with jumps at $\{t_k, k \ge 0\}$ defined by*

$F_t|T_1,\dots,T_n := \sum_{t_k\le t}\tilde X_k$ for any $k\ge 1$ and such that tor any $s\in\mathbb{N}$

$$f_{(\tilde X_1,\dots,\tilde X_s)}(\tilde x_1,\dots,\tilde x_s)\propto\prod_{j=1}^{s}\frac{\tilde x_j^{\delta^*_{0,j}-1}(1-\sum_{l=1}^{j}\tilde x_l)^{\beta^*_{m,j}-1}}{(1-\sum_{l=1}^{j-1}\tilde x_l)^{\delta^*_{0,j}+\beta^*_{m,j}-1}} \tag{5.5.12}$$
$$\times\int_{(0,1)^{n-1}}\prod_{i=1}^{m-1}w_{i,j}^{\alpha_{i,j}}(1-w_{i,j})^{\delta_{i,j}-1}\left(1-\frac{\tilde x_j z_{i,j}}{1-\sum_{l=1}^{j-1}\tilde x_l}\right)^{c_{i,j}}dw_{i,j}$$
$$\times\,\mathbb{1}_{\Delta^{(s)}}(\tilde x_1,\dots,\tilde x_s)$$

where, for $j=1,\dots,s$

$$\delta^*_{0,j}=\delta_{0,j}+n_j\qquad\beta^*_{m,j}=\beta_{m,j}+m_j$$

and n_j is the number of exact observations at t_j and m_j is the sum of the number of exact observation in $\{t_k,k>j\}$ and censored observation in $\{t_k,k\ge j\}$, i.e. $m_j=\sum_{k>j}n_k+\sum_{k\ge j}r_k$.

Proof. The fact that $\{F_t|T_1,\dots,T_n,t\ge 0\}$ is a discrete time neutral to the right r.p.m. derives from Theorem 3 in Ferguson and Phadia [65]. As regard the distribution function of the r.v. $(\tilde X_1,\dots,\tilde X_s)$ it is immediate by combining the likelihood functon with the distribution of the r.v. $(X_1,\dots,X_s)$ admitting a probability density function absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{s+1}$ given by (5.5.11). □

In order to complete the analysis of the discrete time superposed beta-Stacy process, we consider the predictive probability $\mathbb{P}(T_{n+1}=t_k|T_1,\dots,T_n)$.

Proposition 5.5.2. *Let $\{t_k,k\ge 0\}$ be a sequence of time points in $\mathbb{R}^+$ with $t_0:=0$ and for any $n\in\mathbb{N}$, let $T_1,\dots,T_n$ be random sample such that $T_i\in\{t_k,k\ge 1\}$, possibly with right censoring, from a discrete time superposed beta-Stacy process with parameter $(m,(\alpha_{1,\bullet},\beta_{1,\bullet}),\dots,(\alpha_{m,\bullet},\beta_{m,\bullet}))$ and jumps at $\{t_k,k\ge 0\}$. Then*

$$\mathbb{P}(T_{n+1}=t_k|T_1,\dots,T_n)=w_k(n_k,m_k)\prod_{j=1}^{k-1}w_j(n_j,m_j) \tag{5.5.13}$$

where

$$w_k(n_k,m_k):=(n_k+1)\frac{G^{m,1}_{m+1,1}\left(1\middle|\begin{array}{c}-m_k,\alpha_{1,k}+\beta_{1,k}-1,\dots,\alpha_{m,k}+\beta_{m,k}-1\\ \beta_{1,k}-1,\dots,\beta_{m,k}-1,-m_k-n_k-2\end{array}\right)}{G^{m,1}_{m+1,1}\left(1\middle|\begin{array}{c}-m_k,\alpha_{1,k}+\beta_{1,k}-1,\dots,\alpha_{m,k}+\beta_{m,k}-1\\ \beta_{1,k}-1,\dots,\beta_{m,k}-1,-m_k-n_k-1\end{array}\right)} \tag{5.5.14}$$

and for $j = 1, \dots, k-1$

$$w_k(n_j, m_j) := \frac{G^{m,1}_{m+1,1}\left(1 \middle| \begin{matrix} -m_j - 1, \alpha_{1,k} + \beta_{1,k} - 1, \dots, \alpha_{m,k} + \beta_{n,k} - 1 \\ \beta_{1,k} - 1, \dots, \beta_{m,k} - 1, -m_j - n_j - 2 \end{matrix}\right)}{G^{m,1}_{m+1,1}\left(1 \middle| \begin{matrix} -m_j, \alpha_{1,k} + \beta_{1,k} - 1, \dots, \alpha_{m,k} + \beta_{m,k} - 1 \\ \beta_{1,k} - 1, \dots, \beta_{m,k} - 1, -m_j - n_j - 1 \end{matrix}\right)} \tag{5.5.15}$$

and n_j *is the number of exact observations at* t_j *and* m_j *is the sum of the number of exact observation in* $\{t_k, k > j\}$ *and censored observation in* $\{t_k, k \geq j\}$*, i.e.* $m_j = \sum_{k>j} n_k + \sum_{k \geq j} r_k$.

Proof. From the construction (5.5.9) and Theorem 2 in Walker and Muliere [192] it follows that

$$\mathbb{P}(T_{n+1} = t_k | T_1, \dots, T_n) = w_k(n_k, m_k) \prod_{j=1}^{k-1} w_j(n_j, m_j)$$

where

$$w_k(n_k, m_k) := \frac{\mathbb{E}[V_k^{n_k+1}(1 - V_k)^{m_k}]}{\mathbb{E}[V_k^{n_k}(1 - V_k)^{m_k}]}$$

and for $j = 1, \dots, k-1$

$$w_j(n_j, m_j) := \frac{\mathbb{E}[V_j^{n_j}(1 - V_j)^{m_j+1}]}{\mathbb{E}[V_j^{n_j}(1 - V_j)^{m_j}]}$$

with $\{V_k, k \geq 1\}$ a sequence of independent r.v.s such that $V_k \overset{d}{=} X_k/(1 - F_{k-1})$. In particular, by using Theorem 7 in Springer and Thompson [176] we have that for any $p, q \in \mathbb{N}$ and for any $k \geq 1$

$$\begin{aligned}
\mathbb{E}[V_k^p(1 - V_k)^q] = \prod_{i=1}^{m} \frac{\Gamma(\alpha_{i,k} + \beta_{i,k})}{\Gamma(\beta_{i,k})} \int_0^1 & v_k^p (1 - v_k)^q \\
& \times G^{m,0}_{m,0}\left(1 - v_k \middle| \begin{matrix} \alpha_{1,k} + \beta_{1,k} - 1, \dots, \alpha_{m,k} + \beta_{m,k} - 1 \\ \beta_{1,k} - 1, \dots, \beta_{m,k} - 1 \end{matrix}\right) dv_k \\
= \Gamma(p+1) \prod_{i=1}^{m} & \frac{\Gamma(\alpha_{i,k} + \beta_{i,k})}{\Gamma(\beta_{i,k})} \\
& \times G^{m,1}_{m+1,1}\left(1 \middle| \begin{matrix} -q, \alpha_{1,k} + \beta_{1,k} - 1, \dots, \alpha_{m,k} + \beta_{m,k} - 1 \\ \beta_{1,k} - 1, \dots, \beta_{m,k} - 1, -q - p - 1 \end{matrix}\right)
\end{aligned}$$

where the last equation is obtained by applying equation 7.811.2 in Gradshteyn and Ryzhik [77]. The proof is complete by replacing p and q according to Theorem 2 in Walker and Muliere [192]. □

In Walker and Muliere [194], the discrete time beta-Stacy process is characterized by the introduction of a generalization of the Pólya urn scheme. We now provide a similar characterization for the discrete time superposed beta-Stacy process using the so-called neutral exchangeable urn scheme introduced by Walker and Muliere [194].

Let $\iota_1, \dots, \iota_m$ be m different colors in an urn and let c be a positive number. At the first draw a ball of color k is drawn with probability q_k. Following the Pólya urn scheme it is well-known that at the $(n+1)$-th draw the color k is taken with probability $p_n(k|k_1, \dots, k_n)$, where

$$p_n(k|k_1, \dots, k_n) = \frac{cq_k}{c+n} + \frac{\sum_{i=1}^n \mathbb{1}_{\{k_i = k\}}}{c+n}. \tag{5.5.16}$$

It is known that a sequence of r.v.s form the Pólya urn scheme is an exchangeable sequence. The connection between the Pólya urn scheme and the Dirichlet process is given by Blackwell and MacQueen [10]. In particular, let F be a random cumulative distribution function governed by a discrete time Dirichlet process with support the set $\{\iota_1, \dots, \iota_m\}$ and with total mass c and discrete location paramter $G \in \mathcal{F}_{\mathbb{R}^+}$ given by weights q_k at ι_k. Then

$$p_n(k|k_1, \dots, k_n) = \mathbb{E}[F(\iota_k)|k_1, \dots, k_n]$$

where $F(\iota_k)$ represents the random weight assigned to ι_k. Let $\gamma_k := cq_k$, for $k = 1, \dots, m$, and $\varrho_1 := c(1 - q_1)$, so that

$$p_n(k|k_1, \dots, k_n) = \frac{\gamma_k + n_k}{\gamma_1 + \varrho_1 + n} \tag{5.5.17}$$

where $n_k = \sum_{1 \le i \le n} \mathbb{1}_{\{k_i = k\}}$.

The generalization of the Pólya urn scheme proposed in Walker and Muliere [194] can be described as follows. Consider m Pólya urns: the first urn has the different colors $\iota_1, \dots, \iota_m$ and the parameters of the urn are $c_1 > 0$ and the weights for each color $q_1, \dots, q_m$. The second urn has the different colors $\iota_2, \dots, \iota_m$ and the parameters of the urn are $c_2 > 0$ and $q_2/(1 - q_1), \dots, q_m/(1 - q_1)$. Thi third urn has the different colors $\iota_3, \dots, \iota_m$ and parameters $c_3 > 0$ and $q_3/(1 - q_1 - q_2), \dots, q_m/(1 - q_1 - q_2)$. Continue in this fashion up to the m-th urn, which only has the color ι_m. The generalized Pólya urn scheme is now described

i) Start at urn $k = 1$;

ii) sample urn k once according to Pólya urn scheme;

iii) if the color sampled is ι_k then go to (iv), else $k = k + 1$ and go to ii);

iv) ι_k is a single sample from the generalized Pólya urn scheme.

In Walker and Muliere [194] is proved that a sequence of r.v.s from the genereralized Pólya urn scheme is still an exchangeable sequence. Following the generalized Pólya urn scheme it is well-known that at the $(n+1)$-th draw the color k is taken with probability $p_n(k|k_1,\dots,k_n)$, where

$$p_n(k|k_1,\dots,k_n) = \frac{c_k(q_k/(1-q_1-\cdots-q_{k-1})) + \sum_{i=1}^n \mathbb{1}_{\{k_i=k\}}}{c_k + \sum_{i=1}^n \mathbb{1}_{\{k_i=k\}} + \sum_{i=1}^n \mathbb{1}_{\{k_i>k\}}} \times \prod_{l=1}^{k-1}\left(1 - \frac{c_l(q_l/(1-q_1-\cdots-q_{l-1})) + \sum_{i=1}^n \mathbb{1}_{\{l_i>l\}}}{c_l + \sum_{i=1}^n \mathbb{1}_{\{l_i=l\}} + \sum_{i=1}^n \mathbb{1}_{\{l_i>l\}}}\right). \tag{5.5.18}$$

The connection between the generalized Pólya urn scheme and the discrete time beta-Stacy process is given in Walker and Muliere [194]. In particular, let F be a random cumulative distribution function governed by a discrete time beta-Stacy process with support the set $\{\iota_1,\dots,\iota_m\}$. Then

$$p_n(k|k_1,\dots,k_n) = \mathbb{E}[F(\iota_k)|k_1,\dots,k_n]$$

where $F(\iota_k)$ represents the random weight assigned to ι_k. Let $\gamma_k := c_k q_k/(1-q_1-\cdots-q_{k-1})$ and $\varrho_k := c_k(1-q_k/(1-q_1-\cdots-q_{k-1}))$ for $k=1,\dots,m-1$, so that

$$p_n(k|k_1,\dots,k_n) = \frac{\gamma_k + n_k}{\gamma_k + \varrho_k + n_k + m_k}\prod_{l=1}^{k-1}\frac{\varrho_l + m_l}{\gamma_l + \varrho_l + n_l + m_l}. \tag{5.5.19}$$

Note that, if $n_1 + m_1 = n$, $n_k + m_k = m_{k-1}$ and $\gamma_k + \varrho_k = \varrho_{k-1}$, for all $k = 2,\dots,m-1$, i.e. there are not censored observarion, then (5.5.19) reduces to (5.5.17).

Both the Pólya urn scheme and generalized Pólya urn scheme are particular case of the exchangeable neutral urn scheme. Briefly, the exchangeable neutral urn scheme replaces each of the Pólya urns in the generalized Pólya urn scheme with generalized Pólya urn themeselves. In particular, let $\{\iota_1,\dots,\iota_m\}$ be m different colors and let $\{\phi_1,\dots,\phi_{m-1}\}$ be a dummy space. Moreover, let $\{V_k, k\geq 1\}$ be a collection of independent r.v.s on $(0,1)$, such that, for all $\gamma,\varrho>0$ $\mathbb{E}[V_k^\gamma(1-V_k)^\varrho]$ exists. Then define

$$\lambda_k(\gamma,\varrho) := \Gamma(\gamma+1)\prod_{i=1}^m \frac{\Gamma(\alpha_{i,k}+\beta_{i,k})}{\Gamma(\beta_{i,k})} \times G^{m,1}_{m+1,1}\left(1\left|\begin{matrix} -\varrho, \alpha_{1,k}+\beta_{1,k}-1,\dots,\alpha_{m,k}+\beta_{m,k}-1 \\ \beta_{1,k}-1,\dots,\beta_{m,k}-1, -\varrho-\gamma-1\end{matrix}\right.\right)$$

for any $\gamma,\varrho>0$, $m\in\mathbb{N}$ and $(\alpha_{1,\bullet},\beta_{1,\bullet}),\dots,(\alpha_{m,\bullet},\beta_{m,\bullet})$ sequences of positive real numbers. Now, consider s urns indexed by $1,\dots,s$, respectively, and in the urn indexed by k, for $k=1,\dots,s-1$ puts the elements ι_k and ϕ_k in the ratio $\lambda_k(\gamma_k+1,\varrho_k)$ to $\lambda_k(\gamma_k,\varrho_k+1)$ where $\gamma_k>0$, $\varrho_k>0$ with $\varrho_k=\sum_{l>k}\gamma_l$. Obviously, the urn indexed by s has only the element ι_s. Then, generate a sequence $\{T_n, n\geq 1\}$ from the colours $\{\iota_1,\dots,\iota_s\}$ as follows

i) Start from urn indexed by $k=1$;

ii) sample from urn indexed by k;

iii) if the color sampled is ι_k then $\gamma_k=\gamma_k+1$ go to (iv), else $\varrho_k=\varrho_k+1$, $k=k+1$ and go to ii);

iv) ι_k is a single sample from the urn scheme.

Based on this urn scheme, at the $(n+1)$-th draw the color k is taken with the following probability

$$\mathbb{P}(T_{n+1}=\iota_k|T_1,\dots,T_n)=\frac{\lambda_k(\gamma_k+\sum_{i=1}^n \mathbb{1}_{\{T_i=\iota_k\}}+1,\varrho_k+\sum_{i=1}^n \mathbb{1}_{\{T_i>\iota_k\}})}{\lambda_k(\gamma_k+\sum_{i=1}^n \mathbb{1}_{\{T_i=\iota_k\}},\varrho_k+\sum_{i=1}^n \mathbb{1}_{\{k_i>k\}})} \times\prod_{j=1}^{k-1}\frac{\lambda_j(\gamma_j+\sum_{i=1}^n \mathbb{1}_{\{T_i=\iota_j\}},\varrho_l+\sum_{i=1}^n \mathbb{1}_{\{T_i>\iota_j\}}+1)}{\lambda_j(\gamma_j+\sum_{i=1}^n \mathbb{1}_{\{T_i=\iota_j\}},\varrho_j+\sum_{i=1}^n \mathbb{1}_{\{T_i>\iota_j\}})}.$$

which clearly corresponds to the predictive probability (5.5.13) by setting $\gamma_k=\varrho_k=0$ for any $k\geq 1$.

5.5.2 The continuous time superposed beta-Stacy process

So far, we provided a constructive definition of the discrete time superposed beta-Stacy process. In particular, a comprehensive treatment of the discrete time superposed beta-Stacy process has been given in terms of finite dimensional distributions, predictive distributions and posterior distributions. In the next results we focus on proving the existence of the continuous time superposed beta-Stacy process as infinitesimal weak limit of a sequence of discrete time superposed beta-Stacy processes.

Theorem 5.5.3. *For any $m\in\mathbb{N}$, let $G_1,\dots,G_m$ be a collection of continuous functions in $\mathcal{F}_{\mathbb{R}^+}$ and let $c_1,\dots,c_m$ be a collection of piecewise continuous positive functions. there exists a Lévy process $\{Z_t,t\geq 0\}$ such that*

$$\log(\mathbb{E}[e^{-\phi Z_t}])=\int_{(0,+\infty)}(e^{-\phi v}-1)\nu_t(dv) \tag{5.5.20}$$

where

$$\nu_t(dv)=\frac{dv}{1-e^{-v}}\sum_{i=1}^m\int_{(0,t)} e^{-vc_i(s)G_i[s,+\infty)}c_i(s)G_i(ds) \tag{5.5.21}$$

Proof. The proof is along lines similar to the proof of Theorem 2 in Walker and Muliere [194] which follows the idea in Hjort [84]. For any $n\in\mathbb{N}$ and for any $m\in\mathbb{N}$, let us consider m sequences of positive real numbers $(\alpha_{1,\bullet}^{(n)}, \beta_{1,\bullet}^{(n)}), \dots, (\alpha_{m,\bullet}^{(n)}, \beta_{m,\bullet}^{(n)})$ such that $\alpha_{i,k}^{(n)} := c_i\,(k/n - 1/2n)\,G_i\,[k-1/n, k/n)$ and $\beta_{i,k}^{(n)} := c_i\,(k/n - 1/2n)\,G_i\,[k/n, +\infty)$ for $i = 1, \dots, m$ and for any $k \geq 1$. Moreover, for any $n \in \mathbb{N}$ and for any $m \in \mathbb{N}$ let us consider m indepenent sequences of r.v.s $Y_{1,\bullet}^{(n)}, \dots, Y_{m,\bullet}^{(n)}$ such that $Y_{i,\bullet}^{(n)}$ is a sequence of independent r.v.s distributed according to a Beta distribution function with parameter $(\alpha_{i,k}^{(n)}, \beta_{i,k}^{(n)})$ for $i = 1, \dots, m$. Based on this setup of r.v.s, for any $n \in \mathbb{N}$ let us define the random process $Z_\bullet^{(n)} := \{Z_t^{(n)}, t \geq 0\}$ such that $Z_t^{(n)} := -\sum_{k/n \leq t} \log\left(1 - X_k^{(n)}/1 - F_{k-1}^{(n)}\right)$ with $Z_0^{(n)} := 0$ and where $\{X_k^{(n)} | X_1^{(n)}, \dots, X_{k-1}^{(n)}, k \geq 1\}$ is a sequence of r.v.s defined via (5.5.9). Our aim is to show that the sequence of random processes $\{Z_\bullet^{(n)}, n \geq 1\}$ converges weakly (as $n \to +\infty$) to a Lévy process $\{Z_t, t \geq 0\}$ having the required representation (5.5.20) and (5.5.21). By using the ralations for the Gamma function $\Gamma(x) = (x-1)\Gamma(x)$ and $\Gamma(x) = x^{-1}\Gamma(x+1)$ and the by Stirling formula $\Gamma(x) \sim (2\pi x)^{1/2}(x/e)^x$ to approximate the Gamma function, we have

$$
\begin{aligned}
\log(\mathbb{E}[\mathrm{e}^{-\phi Z_t^{(n)}}]) &= \log(\mathbb{E}[\mathrm{e}^{-\phi \sum_{k/n \leq t} \sum_{i=1}^m -\log(1 - Y_{i,k}^{(n)})}]) \\
&= \sum_{k/n \leq t} \sum_{i=1}^m \log \frac{\Gamma(\alpha_{i,k}^{(n)} + \beta_{i,k}^{(n)})\Gamma(\beta_{i,k}^{(n)} + \phi)}{\Gamma(\beta_{i,k}^{(n)})\Gamma(\alpha_{i,k}^{(n)} + \beta_{i,k}^{(n)} + \phi)} \\
&= \sum_{k/n \leq t} \sum_{i=1}^m \log \prod_{j=0}^{r-1} \frac{(\beta_{i,k}^{(n)} + j)(\alpha_{i,k}^{(n)} + \beta_{i,k}^{(n)} + \phi + j)}{(\alpha_{i,k}^{(n)} + \beta_{i,k}^{(n)} + j)(\beta_{i,k}^{(n)} + \phi + j)} \\
&\qquad \times \frac{\Gamma(\alpha_{i,k}^{(n)} + \beta_{i,k}^{(n)} + r)\Gamma(\beta_{i,k}^{(n)} + \phi + r)}{\Gamma(\beta_{i,k}^{(n)} + r)\Gamma(\alpha_{i,k}^{(n)} + \beta_{i,k}^{(n)} + \phi + r)} \\
&= \sum_{k/n \leq t} \sum_{i=1}^m \log \prod_{j \geq 0} \frac{(\beta_{i,k}^{(n)} + j)(\alpha_{i,k}^{(n)} + \beta_{i,k}^{(n)} + \phi + j)}{(\alpha_{i,k}^{(n)} + \beta_{i,k}^{(n)} + j)(\beta_{i,k}^{(n)} + \phi + j)} \\
&= \sum_{k/n \leq t} \sum_{i=1}^m \int_0^{+\infty} (\mathrm{e}^{-\phi v} - 1) \frac{\mathrm{e}^{-\beta_{i,k}^{(n)} v}(1 - \mathrm{e}^{-\alpha_{i,k}^{(n)} v})}{v(1 - \mathrm{e}^{-v})} dv \\
&= \int_0^{+\infty} \frac{\mathrm{e}^{-\phi v} - 1}{v(1 - \mathrm{e}^{-v})} \sum_{k/n \leq t} \sum_{i=1}^m \mathrm{e}^{-\beta_{i,k}^{(n)} v}(1 - \mathrm{e}^{-\alpha_{i,k}^{(n)} v}) dv.
\end{aligned}
$$

Since for $i = 1, \dots, m$, as $n \to +\infty$

$$
\sum_{k/n \leq t} \mathrm{e}^{-\beta_{i,k}^{(n)} v}(1 - \mathrm{e}^{-\alpha_{i,k}^{(n)} v}) \to v \int_{(0,t)} \mathrm{e}^{-v c_i(s) G_i[s, +\infty)} c_i(s) G_i(ds)
$$

then

$$
\log(\mathbb{E}[\mathrm{e}^{-\phi Z_t^{(n)}}]) \to \int_0^{+\infty} \frac{\mathrm{e}^{-\phi v} - 1}{v(1 - \mathrm{e}^{-v})} \sum_{i=1}^m \int_{(0,t)} \mathrm{e}^{-v c_i(s) G_i[s, +\infty)} c_i(s) G_i(ds) dv
$$

as $n \to +\infty$. By a similar argument it can be shown, for any $s \in \mathbb{N}$ and for any $0 = t_0 < t_1 < \cdots < t_s < +\infty$, that

$$\log(\mathbb{E}[\mathrm{e}^{-\sum_{j=1}^{s}\phi Z^{(n)}[t_{j-1},t_j)}]) \to \sum_{j=1}^{s}\int_{(0,+\infty)}(\mathrm{e}^{-v\phi_j}-1)\nu_{[t_{j-1},t_j)}(dv)$$

which ensures the convergence of the finite dimensional distributions of $\{Z_{\bullet}^{(n)}, n \geq 1\}$. The thightness of the sequence $\{Z_{\bullet}^{(n)}, n \geq 1\}$ follows by the same arguments used in Theorem 2 in Walker and Muliere [194]. □

Corollary 5.5.2. *For any $m \in \mathbb{N}$, let $G_1, \ldots, G_m$ be a collection of continuous functions in $\mathcal{F}_{\mathbb{R}^+}$ and $c_1, \ldots, c_m$ be a collection of piecewise continuous positive functions. Let $\{Y_{1,t}, t \geq 1\}, \ldots, \{Y_{m,t}, t \geq 1\}$ be m independent sequences of r.v.s such that $Y_{i,\bullet}$ is a sequence of independent r.v.s distributed according to a Beta distribution function with parameter $(d\alpha_{i,t}, \beta_{i,t})$ with $d\alpha_{i,t} = c_i(t)dG(t)$ and $\beta_{i,t} = c_i(t)G_i[t,+\infty)$ for $i = 1, \ldots, m$. If $\{F_t, t \geq 0\}$ is a random process defined by $F_t = 1 - e^{-Z_t}$ and with $F_0 := 0$, then at the infinitesimal level*

$$dF_t|F_t \stackrel{d}{=} (1-F_t)\left(1-\prod_{i=1}^{m}(1-Y_{i,t})\right) \tag{5.5.22}$$

and $\{F_t, t \geq 0\}$ belongs to $\mathscr{F}_{\mathbb{R}^+}$ a.s.-$\mathbb{P}$.

Proof. For any $n \in \mathbb{N}$ and for any $m \in \mathbb{N}$, let $(\alpha_{1,\bullet}^{(n)}, \beta_{1,\bullet}^{(n)}), \ldots, (\alpha_{m,\bullet}^{(n)}, \beta_{m,\bullet}^{(n)})$ be m sequences of positive real numbers and let $Y_{1,\bullet}^{(n)}, \ldots, Y_{m,\bullet}^{(n)}$ be m independent sequences of r.v.s such that $Y_{i,\bullet}^{(n)}$ is a sequence of independent r.v.s distributed according to a Beta distribution function with parameter $(\alpha_{i,k}^{(n)}, \beta_{i,k}^{(n)})$ for $i = 1, \ldots, m$. Based on this setup of r.v., for any $n \in \mathbb{N}$ let us define the discrete time superposed beta-Stacy process $F_{\bullet}^{(n)} := \{F_t^{(n)}, t \geq 0\}$ such that

$$F_t^{(n)} := \sum_{k/n \leq t} X_k^{(n)}$$

with $F_0^{(n)} := 0$ and where $\{X_k^{(n)}|X_1^{(n)}, \ldots, X_{k-1}^{(n)}, k \geq 1\}$ is a sequence of r.v.s defined via (5.5.9). Since we have

$$-\log(1-F_t^{(n)}) = -\sum_{k/n \leq t}\log\left(\prod_{i=1}^{m}(1-Y_{i,k})\right) = -\sum_{k/n \leq t}\log\left(1-\frac{X_k^{(n)}}{1-F_{k-1}^{(n)}}\right) = Z_t^{(n)}$$

then $F_t^{(n)} = 1 - \mathrm{e}^{-Z_t^{(n)}}$. Clearly

$$F^{(n)}\left[\frac{k-1}{n}, \frac{k}{n}\right) \mid F^{(n)}\left[0, \frac{k-1}{n}\right) \stackrel{d}{=} \left(1-F^{(n)}\left[0, \frac{k-1}{n}\right)\right)\left(1-\prod_{i=1}^{k}(1-Y_{i,k})\right)$$

and also $\{F_\bullet^{(n)}, n \geq 1\}$ converges weakly to $\{F_t, t \geq 0\}$. The fact that $\{F_t, t \geq 0\}$ belongs to $\mathscr{F}_{\mathbb{R}^+}$ a.s.-$\mathbb{P}$ follows from the fact that $\int_{(0,+\infty} G_i(ds)G_i[s,+\infty) = +\infty$ for $i = 1, \ldots, m$. □

Corollary 5.5.2 completes the proof of the existence of a continuous time version of the discrete time superposed beta-Stacy process. In particular, Corollary 5.5.2 implies that the random process $F_t, t \geq 0$ is a neutral to the right r.p.m. according to the definition given by Doksum [25]. We term the random process $\{F_t, t \geq 0\}$ superposed beta-Stacy process.

Let c be a positive function, $G \in \mathcal{F}_{\mathbb{R}^+}$ be right continuous and $\{t_k, k \geq 0\}$ be the countable set of points of discontinuity of G, i.e $G\{t_k\} = G(t_k) - G(t_k^-) > 0$ for any $k \geq 0$. We set $G^c := G(t) - \sum_{t_k \leq t} G\{t_k\}$ so that G^c is a continuous function. Here, the formal definition of superposed beta-Stacy process.

Definition 5.5.4. *The random process $\{F_t, t \geq 0\}$ is a superposed beta-Stacy process on $\mathbb{R}^+$ with parameter $(m, (c_1, G_1), \ldots, (c_m, G_m))$ if for all $t \geq 0$, $F_t = 1 - e^{-Z_t}$, where $\{Z_t, t \geq 0\}$ is a Lévy process with Lévy measure for $v > 0$*

$$\nu_t(dv) = \frac{dv}{1 - e^{-v}} \sum_{i=1}^{n} \int_{(0,t)} e^{-vc_i(s)G_i[s,+\infty)} c_i(s) G_i^c(ds) \tag{5.5.23}$$

and moment generating function given by

$$\log(\mathbb{E}[e^{-\phi Z_t}]) = \sum_{t_k \leq t} \log(\mathbb{E}[e^{\phi \sum_{i=1}^{m} \log(1-Y_{i,k})}]) - \int_{(0,+\infty)} (1 - e^{-\phi v}) \nu_t(dv) \tag{5.5.24}$$

where the $Y_{1,\bullet}, \ldots, Y_{m,\bullet}$ are m independent sequences of r.v.s such that $Y_{i,\bullet}$ is a sequence of independent r.v.s distributed according to a Beta distribution function with parameter $(c_i(t_k)G_i\{t_k\}, c_i(t_k)G_i[t_k, +\infty))$.

From Definition 5.5.4 it follows that the superposed beta-Stacy process includes as particular case the beta-Stacy process which can be recovered by setting $m = 1$. In particular, let us consider a superposed beta-Stacy process $\{F_t, t \geq 0\}$ on $\mathbb{R}^+$ with parameters $(m, (c_1, G_1), \ldots, (c_m, G_m))$ such that G_i is continuous and c_i is constant for $i = 1, \ldots, m$. Under these assumptions, (5.5.23) becomes

$$v_t(dv) = \frac{dv}{v(1 - \mathrm{e}^{-v})} \sum_{i=1}^{m} \mathrm{e}^{-vc_i}(\mathrm{e}^{vc_iG_i(t)} - 1)$$

which generalizes the Lévy measure of a Dirichlet process on $\mathbb{R}^+$ with parameter cG. See Ferguson [62] for more details.

We conclude this section by considering the posterior distribution of a superposed beta-Stacy process given a set of possibly right censored observations. Extending the result in Doksum [25] in the case of inclusively and exclusively right censored observations, Ferguson and Phadia [65] (see also Ferguson [62]) achieved a representation of the posterior distribution of a neutral to the right r.p.m. given a set of possibly right censored observations. We now apply results in Ferguson and Phadia [65] in order to obtain a characterization of the posterior distribution of a superposed beta-Stacy process given a set of possibly right censored observations. Let $\{F_t, t \geq 0\}$ be a superposed beta-Stacy process on $\mathbb{R}^+$ with parameter $(m, (c_1, G_1), \dots, (c_m, G_m))$. The prior distribution for the Lévy process $\{Z_t, t \geq 0\}$ such that for any $t \geq 0$, $F_t = 1 - \mathrm{e}^{-Z_t}$ is characterized by the set of fixed points of discontinuity $\{t_k, k \geq 0\}$ with $t_0 = 0$, the associated probability density functions $\{f_{t_k}, k \geq 1\}$ for the jumps and the Lévy measure (5.5.23). If T is a random sample possibly right censored from $\{F_t, t \geq 0\}$, then

i) given $T > t \in \{t_k, k \geq 0\}$, the posterior parameters are $\{\{t_k, k \geq 0\}$ and

$$f^*_{t_k}(v) = \begin{cases} \kappa \mathrm{e}^{-v} f_{t_k}(v) & \text{if } t_k \leq t \\ f_{t_k}(v) & \text{if } t_k > t \end{cases}$$

ii) given $T = t \in \{t_k, k \geq 0\}$, posterior parameters are $\{\{t_k, k \geq 0\}$ and

$$f^*_{t_k}(v) = \begin{cases} \kappa \mathrm{e}^{-v} f_{t_k}(v) & \text{if } t_k < t \\ \kappa(1 - \mathrm{e}^{-v}) f_{t_k}(v) & \text{if } t_k = t \\ f_{t_k}(v) & \text{if } t_k > t \end{cases}$$

iii) given $T = t \notin \{t_k, k \geq 0\}$, posterior parameters are $\{\{t_k, k \geq 0\} \cup \{t\}$ and

$$f^*_{t_k}(v) = \begin{cases} \kappa \mathrm{e}^{-v} f_{t_k}(v) & \text{if } t_k \leq t \\ f_{t_k}(v) & \text{if } t_k > t \end{cases}$$

where κ is the appropriate normalizing constant and the posterior Lévy measure is given by

$$v^*_t(dv) = \frac{dv}{1 - \mathrm{e}^{-v}} \sum_{i=1}^{n} \int_{(0,t)} \mathrm{e}^{-v(c_i(s) G_i[s, +\infty) + \mathbb{1}_{\{t \geq s\}})} c_i(s) G^c_i(ds)$$

6

On Bayesian nonparametric inference in species sampling problems

In this chapter we investigate a class of random probability measures, termed generalized Dirichlet processes, which has been recently introduced in the literature and further investigated in Chapter 5. Such processes induce exchangeable random partitions which are characterized by a more elaborate clustering structure than those arising from Gibbs-type random probability measures. A natural area of application of these random probability measures is represented by species sampling problems and, in particular, prediction problems in genomics. To this end we study both the distribution of the number of distinct species present in a sample and the distribution of the number of new species conditionally on an observed sample. Some developments on the conditional distributions and the corresponding Bayesian nonparametric estimators recently obtainied for the class of Gibbs-type random probability measures are also provided.

6.1 Introduction

Let $\{X_n, n \geq 1\}$ be an exchangeable sequence defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and such that each random variable (r.v.) X_n takes values in a Polish space $(\mathbb{X}, \mathcal{T})$ with associated Borel σ-field $\mathscr{X}$. Then, by de Finetti's representation theorem, there exists a random probability measure (r.p.m.) $\tilde{P}$ on $\mathbb{X}$ with law Q such that given $\tilde{P}$, a sample $X_1, \ldots, X_n$ from the exchangeable sequence is independent and identically distributed (i.i.d.) with distribution $\tilde{P}$. That is, for every $n \geq 1$ and any $A_1, \ldots, A_n \in \mathscr{X}$

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \int_{\mathcal{P}_{\mathbb{X}}} \prod_{i=1}^{n} \tilde{P}(A_i) Q(d\tilde{P}) \tag{6.1.1}$$

where $\mathcal{P}_{\mathbb{X}}$ denotes the space of all probability measures on $(\mathbb{X}, \mathscr{X})$ with associated Borel σ-field $\mathscr{P}_{\mathbb{X}}$, generated by the weak convergence topology.

In this chapter, we focus on r.p.m.s which are almost surely discrete and with non-atomic prior guess at the shape $\alpha_0(\cdot) := \mathbb{E}[\tilde{P}(\cdot)]$. By the almost sure discreteness, we expect ties in the sample, namely that $X_1, \ldots, X_n$ contain $K_n \leq n$ distinct observations $X_1^*, \ldots, X_{K_n}^*$ with frequencies $N_{K_n} = (N_1, \ldots, N_{K_n})$ such that $\sum_{1 \leq j \leq k} N_j = n$. By integrating (6.1.1) with respect to all samples of size n with $K_n = k$ distinct observations having frequencies $(n_1, \ldots, n_k)$, one obtains the joint distribution of K_n and N_{K_n}

$$\mathbb{P}(\{K_n = k\} \cap \{N_{K_n} = (n_1, \ldots, n_{K_n})\}) = p_k^{(n)}(n_1, \ldots, n_k) \qquad (6.1.2)$$

which is known as exchangeable partition probability function (EPPF), a fundamental concept introduced in Pitman [149] (see Appendix B), which uniquely determines the probability law of an exchangeable random partition. Almost sure discrete r.p.m.s and the exchangeable random partitions they induce have always played an important role in a variety of research areas such as population genetics, machine learning, Bayesian nonparametrics, combinatorics, excursion theory and statistical physics. In particular, in Bayesian nonparametric inference the use of random partitions dates back to the seminal work of Lo [128]: his approach consists in exploiting a discrete r.p.m. as a basic building block in hierarchical mixture models. In this way the discrete r.p.m. induces an exchangeable random partition for the latent variables providing an effective tool for inference on the clustering structure of the observations. See, e.g, Lo [129], James [96], Ishwaran and James [91] and Lijoi et al. [120] for extensions in various directions.

An early and well-known model which describes the grouping of n objects into k distinct classes is due to Ewens [43] and leads to the Ewens' sampling formula. The basic assumption is that individuals are sequentially sampled from an infinite set of different species and the proportion $\tilde{p}_i$ with which the i-th species is presented in the population is random. Then, if $\{W_k, k \geq 1\}$ is a sequence of i.i.d. r.v.s distributed according to a Beta distribution function with parameter $(1, \theta)$, the random proportion are defined as $\tilde{p}_1 = W_1$ and $\tilde{p}_j = W_j \prod_{1 \leq k \leq j-1}(1 - W_k)$ per $j \geq 2$. Now, if $X_1, \ldots, X_n$ is a sample of n individuals drawn from the population, set $M_n := (M_{1,n}, \ldots, M_{n,n})$ where $M_{j,n}$ is the number of species represented j times in the sample of size n. Hence, the distribution of M_n is supported by all those vectors $m_n = (m_{1,n}, \ldots, m_{n,n})$ for which $\sum_{1 \leq i \leq n} i m_i = n$. The Ewens sampling formula provides the probability distribution of the r.v. M_n and it coincides with

$$\mathbb{P}(M_n = m_n) = \frac{n!}{(\theta)_{n\uparrow 1}} \prod_{j=1}^{n} \frac{\theta^{m_{j,n}}}{j^{m_{j,n}} m_{j,n}!}$$

fro $\theta > 0$ and for any $(m_{1,n}, \ldots, m_{n,n})$ such that $\sum_{1 \leq i \leq n} i m_{i,n} = n$. Obviously, to the distribution of M_n there corresponds a distribution of the vector (K_n, N_{K_n}). Such a correspondence is one to one and, conditional on K_n, the distribution of N_{K_n} is supported on the set $\mathcal{D}_{K_n,n} := \{(n_1,\dots,n_{K_n}) \in [n]^{K_n} : \sum_{1\le j\le K_n} n_j = n\}$. In particular, to the Ewens sampling formula there corresponds the probability distribution

$$p_k^{(n)}(n_1,\dots,n_k) = \frac{\theta^k}{(\theta)_{n\uparrow 1}} \prod_{j=1}^{k} (n_j - 1)! \tag{6.1.3}$$

for any $k \in [n]$ and $(n_1,\dots,n_k) \in \mathcal{D}_{k,n}$. As described in Antoniak [2], equation (6.1.3) corresponds to the EPPF induced by a sample $X_1,\dots,X_n$ from a Dirichlet process (see Ferguson [61]) and it has found many interesting applications, for instance, in Bayesian nonparametrics and in population genetics. See Antoniak [2] and Pitman [157] for exhaustive accounts on the Ewens sampling formula.

Since the introduction of the Dirichlet process, other classes of almost surely discrete r.p.m.s have been proposed in the literature. Among them we mention species sampling models introduced by Pitman [151], stick-breaking r.p.m.s introduced by Ishwaran and James [90], normalized random measures with independent increments (NRMI) introduced by Regazzini et al. [165] and Poisson-Kingman models introduced by Pitman [156]. Within these classes, all specific r.p.m.s, which enjoy sufficient mathematical tractability, represent valid alternatives to the Dirichlet process: the most notable are the two parameter Poisson-Dirichlet process (see Pitman [149] and Pitman [151]) and the normalized generalized Gamma process (see Pitman [156], James [96] and Lijoi et al. [120]); both recover the normalized stable process introduced by Kingman [109] and the Dirichlet process as limiting cases and the latter also contains the normalized inverse Gaussian process introduced by Lijoi et al. [119]. By close inspection of these tractable processes, one can observe that they all generate samples $X_1,\dots,X_n$, for $n \ge 1$, which are characterized by a system of predictive distributions of the type

$$\mathbb{P}(X_{n+1} \in \cdot | X_1,\dots,X_n) = g_0(n,k)\alpha_0(\cdot) + g_1(n,k)\sum_{j=1}^{k}(n_j - \sigma)\delta_{X_j^*}(\cdot), \tag{6.1.4}$$

where $\sigma \in [0,1)$. An almost surely discrete r.p.m. generating a sample as the above is termed Gibbs-type r.p.m. The class of Gibbs-type r.p.m.s has been recently introduced and studied by Gnedin and Pitman [74], where also a characterization of its members is provided: indeed, Gibbs-type r.p.m. are Dirichlet process mixtures when $\sigma = 0$ and Poisson-Kingman models based on the stable subordinators when $\sigma \in (0,1)$ (see Gnedin and Pitman [74], Theorem 12). Further investigations related to Bayesian nonparametrics can be found in Ho et al. [86] and Lijoi et al. [124].

Motivated by species sampling problems and, in particular, by their renewed interest related to applications in genomics, in Lijoi et al. [123] and Lijoi et al. [125] properties of samples generated by Gibbs-type r.p.m.s have been analyzed. In particular, given a sample $X_1,\dots,X_n$ consisting in a collection of j distinct species with labels $X_1^*,\dots,X_j^*$ with frequencies $(n_1, \ldots, n_j)$, interest is in the distributional properties of an additional sample of size m and, especially, in the distribution of the new distinct species. The concrete motivation for this study is provided by the straightforward applicability of the results to inference in genetic experiments. As a matter of fact, an important setting of application is related to gene detection in expressed sequence tags (EST) experiments. ESTs are produced by sequencing randomly selected cDNA clones from a cDNA library. Given an initial EST dataset of size n, one is interested in the prediction of the outcomes of further sampling from the library. For instance, interest lies in the estimation of the number of new unique genes in a possible additional sample of size m: nonparametric frequentist estimators, however, yield completely unstable estimates when $m > 2n$. See Mao [133] for a discussion of this phenomenon. In contrast, for the corresponding Bayesian nonparametric estimators proposed in in Lijoi et al. [123] and Lijoi et al. [125], and based on Gibbs partitions, the relative dimension of m with respect to n is not an issue. Indeed, it is shown that the EPPF, whenever analytically available, yields straightforward and coherent answers to this and other related problems.

In Lijoi et al. [123] and Lijoi et al. [125] Bayesian estimators for species sampling problems have been derived under the hypothesis that the exchangeable sequence is governed by a Gibbs-type prior. It is to be noted that the number of distinct species in the given sample K_n turns out to be a sufficient statistic for prediction of the number of new distinct species (and other interesting quantities) to be observed in a future sample (see Lijoi et al. [125]). This implies that the information arising from the frequencies $(n_1, \ldots, n_j)$ has to be incorporated into the parameters of the model, since, otherwise, prediction of new species would not depend at all on $n_1, \ldots, n_j$. For instance, if the species are exchangeable with a two parameter Poisson-Dirichlet prior, then, given a sample of size n, the $(n+1)$-th observation is a new species with probability $(\theta+\sigma j)/(\theta+n)$, where $\theta > -\sigma$ and $\sigma \in [0, 1)$. Such a probability depends on the distinct observed species j but not on their frequencies $n_1, \ldots, n_j$, whose conveyed information can be summarized through the selection of θ and σ. In principle one would like priors which lead to richer predictive structures, in which the probability of sampling a new species depends explicitly on both K_n and N_{K_n}. However, by dropping the Gibbs structure assumption, serious issues of mathematical tractability arise.

In this chapter we provide some developments on the conditional distributions and the corresponding Bayesian nonparametric estimators recently obtained in Lijoi et al. [123] and Lijoi et al. [125] for discrete nonparametric priors which induce Gibbs-type random partitions. In particular, we focus on the two parameter Poisson-Dirichlet process and greatly simplify the expressions of relevant estimators in species sampling problems so that they can be easily evaluated for any sizes of n and m. Moreover, in order to associate a measure of uncertainty to the estimates, we study the asymptotic behaviour of the number of new species conditionally on the observed sample: such an asymptotic results, which is also of independent interest, allow to derive asymptotic highest posterior density intervals for the estimates of interest. In order to sample from the limiting r.v., we develop a suitable simulation scheme. Finally we illustrate the implementation of the proposed methodology by the analysis of 5 genomic dataset.

On the other hand, we consider a class of r.p.m.s which is not of Gibbs-type, and show that one can still derive analytic expressions for the quantities of interest. In pursuing this goal we will repeatedly encounter Lauricella multiple hypergeometric functions (see Appendix C) and partial Bell polynomials (see Appendix A), thus highlighting the interplay between Bayesian nonparametrics and exchangeable random partitions on one side and the theory of special functions on the other. Other examples of this close connection can be found in Regazzini [163], Lijoi and Regazzini [126] where functionals of the Dirichlet process are considered. As an interesting by-product of our analysis, we also obtain a generalization of the Chu-Vandermonde convolution formula involving the fourth Lauricella hypergeometric function. The specific class we consider is represented by the generalized Dirichlet process introduced in Regazzini et al. [165] and further investigated in Lijoi et al. [118]. In particular, the generalized Dirichlet process is a NRMI obtained by normalization of superposed independent Gamma processes with increasing integer-valued scale parameter and gives rise to a system of predictive distributions of the type

$$\mathbb{P}(X_{n+1} \in \cdot | X_1, \ldots, X_n) = w_0(n, k, \mathbf{n})\alpha_0(\cdot) + \sum_{j=1}^{k} n_j w_j(n, k, \mathbf{n})\delta_{X_j^*}(\cdot) \tag{6.1.5}$$

where $\mathbf{n} := (n_1, \ldots, n_k)$ and the weights $w_0(n, k, \mathbf{n})$ and $w_i(n, k, \mathbf{n})$, for $j = 1, \ldots, k$ now explicitly depend on $\mathbf{n}$ thus conveying the additional information provided by the frequencies $n_1, \ldots, n_k$ directly into the prediction mechanism. To our knowledge, the generalized Dirichlet process represents the first example in the literature of almost surely discrete r.p.m. which is not of Gibbs-type and still leads to a closed form predictive structure.

The chapter is structured as follows. In Section 6.2 we remind the definition of Gibbs-type r.p.m. and we provide some developments on the conditional distributions and the corresponding Bayesian nonparametric estimators obtainied in Lijoi et al. [123] and Lijoi et al. [125]. In Section 6.3 we provide distributional results related to the prior and posterior probability distribution of discovering a certain number of new species in a sample generated by a generalized Dirichlet process. In Section 6.4 we investigate some conditional structures that emerge when the observations are sampled from exchangeable sequences governed by a general homogeneous NRMIs.

6.2 Gibbs-type random probability measures

In this section we remind the definition of Gibbs-type r.p.m. and some results related to conditional distributions and Bayesian nonparametric estimators recently obtainied in Lijoi et al. [123] and Lijoi et al. [125]. Developments of these results are presented for the two parameter Poisson Dirichlet process and for the normalized generalized Gamma process.

A random partition of the set of natural numbers $\mathbb{N}$ is defined as a consistent sequence $\Pi := \{\Pi_n, n \geq 1\}$ of random elements, with Π_n taking values in the set of all the partitions of $[n] := \{1, \ldots, n\}$ into some number of disjoint blocks. Consistency implies that each Π_n is obtained from Π_{n+1} by discarding, from the latter, the integer $n + 1$. A random partition Π is exchangeable if, for each n, the probability distribution of Π_n is invariant under all permutations of $(1, \ldots, n)$. Let $\{p_k^{(n)}, n \geq 1\}$ be a sequence of function such that $p_k^{(n)} : \mathcal{D}_{k,n} \to \mathbb{R}^+$ satisfies the properties

i) $p_1^{(1)}(1) = 1$;

ii) for any $(n_1, \ldots, n_k) \in \mathcal{D}_{k,n}$ with $n \geq 1$ and $k \in [n]$

$$p_k^{(n)}(n_1, \ldots, n_k) = p_k^{(n)}(n_{\sigma(1)}, \ldots, n_{\sigma(k)})$$

where σ is an arbitrary permutation of the indices $(1, \ldots, k)$;

iii) for any $(n_1, \ldots, n_k) \in \mathcal{D}_{k,n}$ with $n \geq 1$ and $k \in [n]$ the following addition rule holds true

$$p_k^{(n)}(n_1, \ldots, n_k) = \sum_{j=1}^{k} p_k^{(n+1)}(n_1, \ldots, n_j + 1, \ldots, n_k) + p_{k+1}^{(n+1)}(n_1, \ldots, n_k, 1).$$

Then, $p_k^{(n)}$ is an EPPF. In particular, the EPPF uniquely determines the probability law of an exchangeable random partition according to the equality

$$\mathbb{P}(\Pi_n = \{A_1, \ldots, A_k\}) = p_k^{(n)}(|A_1|, \ldots, |A_k|) \qquad n \geq 1, k \leq n$$

where $|A|$ stands for the cardinality of the set A. As already seen in equation (6.1.2), for any given sample $X_1, \ldots, X_n$ from an exchangeable sequence of r.v.s $\{X_n, n \geq 1\}$ governed by an almost sure discrete r.p.m., marginalization with respect to the labels yields the EPPF of the associated exchangeable random partition. The EPPF (6.1.3) associated to the celebrated Ewens sampling formula, represents a particular case of a more geneneral EPPF which is associated to the so-called Pitman's sampling formula (see Pitman [149]). In particular, to the Pitman's sampling formula there corresponds the probability distribution

$$p_k^{(n)}(n_1,\dots,n_k) = \frac{\prod_{i=1}^{k-1}(\theta+i\sigma)}{(\theta+1)_{(n-1)\uparrow 1}} \prod_{j=1}^{k}(1-\sigma)_{(n_j-1)\uparrow 1} \tag{6.2.1}$$

where $\theta > -\sigma$ and $\sigma \in (0,1)$ or $\sigma < 0$ and $\theta = \nu|\sigma|$ for some positive ν. It can be easily checked that if we set $\sigma = 0$ in (6.2.1) we recover (6.1.3). In particular, (6.2.1) corresponds to the EPPF induced by a sample $X_1,\dots,X_n$ from a two parameter Poisson-Dirichlet process as described in Pitman [149] (see also Pitman and Yor [154]). A further interesting example of EPPF arises from the normalization of a generalized Gamma process, as defined in Brix [13], and leads to

$$p_k^{(n)}(n_1,\dots,n_k) = \frac{\sigma^{k-1}\mathrm{e}^{\beta}\prod_{j=1}^{k}(1-\sigma)_{(n_j-1)\uparrow 1}}{\Gamma(n)} \sum_{i=1}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\Gamma\left(k-\frac{i}{\sigma};\beta\right) \tag{6.2.2}$$

where $\beta > 0$ and $\Gamma(a;x) := \int_x^{+\infty} s^{a-1}\mathrm{e}^{-s}ds$ is, for any $x > 0$, the incomplete Gamma function. See James [96] and Pitman [156] and Lijoi et al. [120] for an applcation of the corresponding random discrete distribution in the context of mixture modelling.

The examples we have briefly illustrated so far share a common structure. Indeed, one may note that each EPPF in (6.1.3), (6.2.1) and (6.2.2) arises as a product of two factors: the first one depends only on (n,k) and the second one depends on the frequencies $(n_1,\dots,n_k)$ via the product $\prod_{1\le j\le k}(1-\sigma)_{(n_j-1)\uparrow 1}$. This structure is the main ingredient for defining a general family of exchangeable random partitions, namely Gibbs-type random partitions.

Definition 6.2.1. *(cfr. Gnedin and Pitman [74]) An exchangeable random partition* Π *of the set of natural numbers is said to be of Gibbs form if, for all* $1 \le k \le n$ *and for any* $(n_1,\dots,n_k)$ *in* $\mathcal{D}_{k,n}$*, the EPPF of* Π *can be represented as*

$$p_k^{(n)}(n_1,\dots,n_k) = V_{n,k}\prod_{j=1}^{k}(1-\sigma)_{(n_j-1)\uparrow 1} \tag{6.2.3}$$

for some $\sigma \in [0,1)$ *and some set of non-negative weights* $\{V_{n,k} : n \ge 1, 1 \le k \le n\}$ *satisfying the recursion* $V_{n,k} = V_{n+1,k+1} + (n-\sigma k)V_{n+1,k}$ *with* $V_{1,1} = 1$.

Recall that, according to Pitman [151], a species sampling model is an almost surely discrete r.p.m. $\tilde{P}(\cdot) = \sum_{i\ge 1}\tilde{w}_i\delta_{X_i}(\cdot)$ such that the masses $\tilde{w}_i$'s are independent from the locations X_i's, which are i.i.d. from a non-atomic distribution α_0. Then, one can define Gibbs-type r.p.m.s as the class of species sampling models which induces Gibbs-type random partition, i.e. the EPPF corresponding to a sample of size n generated by a

Gibbs-type r.p.m. is of the form (6.2.3). It then follows that the predictive distributions associated with a Gibbs-type r.p.m. are of the form (6.1.4) with weights of the form

$$g_0(n,k) := \frac{V_{n+1,k+1}}{V_{n,k}}, \qquad g_1(n,k) := \frac{V_{n+1,k}}{V_{n,k}}.$$

Before reminding the distributional results for samples drawn from Gibbs-type prior, we introduce some useful notation to be used throughout the paper. Consider a population which is composed of an (ideally) infinite number of species. Let $X_{K_n}^{(1:n)} := X_1, \dots, X_n$ be a sample of n individuals drawn from the population, where K_n is the number of distinct species detected among the n observations in the sample. We call this sample the "basic sample". We consider the vector (K_n, N_{K_n}) where $N_{K_n} = (N_{1,n}, \dots, N_{K_n,n})$ is the vector of frequencies with which each distinct species is observed; in particular, conditional on $K_n = j$, is supported by all vectors $(n_1, \dots, n_{K_n})$ of positive integers such that $\sum_{1 \le i \le K_n} n_i = s$. We denotes with $X^{(1:K_n)} := X_1^*, \dots, X_{K_n}^*$ the distinct observations within the "basic sample". We study distributional properties of the partition of the set of integers $\{n+1, \dots, n+m\}$, given $[n]$ has been partitioned into j classes with repective frequencies $(n_1, \dots, n_j)$. A few quantities, analogous to those describing the partition structure of $[n]$, need to be introduced in advance. We let $K_m^{(n)} = K_{n+m} - K_n$ stands for the number of new partition sets $C_1, \dots, C_{K_m^{(n)}}$ generated by the additional "second sample" $X_{K_m^{(n)}}^{(2:m)} := X_{n+1}, \dots, X_{n+m}$. We call this sample the "second sample". Furthermore, if $C := \cup_{1 \le i \le K_m^{(n)}} C_i$ whenever $K_m^{(n)} \ge 1$ and $C \equiv \emptyset$ if $K_m^{(n)} = 0$, we set $K_m^{(n)} := \sharp(\{X_{n+1}, \dots, X_{n+m}\} \cap C)$ as the number of observations belonging to the new clusters C_i. It is obvious that $L_m^{(n)} \in \{0, 1, \dots, m\}$ and that $m - L_m^{(n)}$ observations belong to the sets defining the partiton of the original n observations. According to this, if $S_{L_m^{(n)}} = (S_{1,L_m^{(n)}}, \dots, S_{K_m^{(n)},L_m^{(n)}})$ then the distribution of $S_{L_m^{(n)}}$, conditional on $L_m^{(n)} = s$, is supported by all vectors $\mathbf{s} := (s_1, \dots, s_{K_m^{(n)}})$ of positive integers such that $\sum_{1 \le i \le K_m^{(n)}} s_i = s$. The remaining $m - L_m^{(n)}$observations are allocated to the old K_n groups with the vector of non-negative frequencies $R_n = (R_1, \dots, R_{K_n})$ such that $\sum_{1 \le i \le K_n} R_i = m - L_m^{(n)}$. Throughout we also assume that all random quantities are defined on a common space $(\Omega, \mathscr{F}, \mathbb{P})$.

In a Bayesian framework, the probabilty distribution of K_n has been interpreted by Lijoi et al. [123] as the prior distribution on the number of species to be observed in the "basic sample". In particular, for a Gibbs-type r.p.m, the probability distribution of K_n was derived in Gnedin and Pitman [74] and it corresponds to

$$\mathbb{P}(K_n = k) = \frac{V_{n,k}}{\sigma^k} \mathscr{C}(n, k; \sigma). \tag{6.2.4}$$

where $\mathscr{C}(s, k; \sigma)$ is $\mathscr{C}(s, k; \sigma)(-1)^{s-k} C(s, k; \sigma)$ (see Appendix A). Next, a further sample of m individuals, the "second sample", is selected thus giving rise to the "enlarged sample" of size $n+m$. If one knows the number of sepecies observed in the "basic sample" and the frequency with which each species has been recorded, it would be interesting to determine

P1) the probability distribution, and the expected value, of the number of new species in the "second sample" conditionally on the "basic sample";

P2) the probability of discovering a new species at the $(n+m+1)$-th draw, without actually observing the "second sample".

Evaluating the probability in P1) is equivalent to determining $\mathbb{P}(K_m^{(n)} = k | X_{K_n}^{(1,n)})$ for any $j = 0, 1, \ldots, m$ and for any $k = 1, 2, \ldots, n$, which can be interpreted as the "posterior" probability distribution of the number of species to be observed in a sample of size m. The determination of the probability in P2) corresponds to estimating the probability $\mathbb{P}(K_1^{(n+m)} = 1 | X_{K_n}^{(1,n)}, X_{K_m^{(n)}}^{(2:m)})$ without observing the "second sample". This automatically provides a solution to the important problem of determining the sample size such that the probability of discovering a new species falls below a give threshold.

Proposition 6.2.1. *(cfr. Lijoi et al. [125]) Suppose that $\Pi = \Pi_n, n \geq 1$ is a Gibbs-type random partition with weights $V_{n,k}$ and parameter $[0,1)$. Then, the joint distribution of $K_m^{(n)}$, $K_m^{(n)}$ and $S_{L_m^{(n)}}$, given K_n and N_{K_n}, is of the form*

$$\begin{aligned} \mathbb{P}(K_m^{(n)} = k, L_m^{(n)} = s, S_{L_m^{(n)}} = \mathbf{s} | K_n = j, N_{K_n} = \mathbf{n}) \qquad (6.2.5) \\ = \mathbb{P}(K_m^{(n)} = k, L_m^{(n)} = s, S_{L_m^{(n)}} = \mathbf{s} | K_n = j) \\ = \frac{V_{n+m,j+k}}{V_{n,j}} \binom{m}{s} (n - j\sigma)_{(m-s)\uparrow 1} \prod_{i=1}^{k} (1-\sigma)_{(s_i - 1)\uparrow 1} \end{aligned}$$

Hence, the number K_n of partition sets in the basic n sample is sufficient for predicting: i) the number of sets into which $\{n+1, \ldots, n+m\}$ is partitioned, ii) the number of points from the subsequent m sample that belong to the new sets of the partition of $[n+m]$ and iii) the frequencies in each of these new groups.

By marginalizing the conditional distribution in (6.2.5) with respect to $S_{L_m^{(n)}}$ and, then, with respect to $K_m^{(n)}$ one obtains the conditional distribution for the number of new groups and the number of observations belonging to these new groups and the distribution of $L_m^{(n)}$, respectively.

Corollary 6.2.1. *(cfr. Lijoi et al. [125]) The joint distribution of $K_m^{(n)}$ and $L_m^{(n)}$, given K_n, can be expressed as*

$$\mathbb{P}(K_m^{(n)}, L_m^{(n)} = s | K_n = j) = \frac{V_{n+m,j+k}}{V_{n,j}} \binom{m}{s} (n - j\sigma)_{(m-s)\uparrow 1} \frac{\mathscr{C}(s, k; \sigma)}{\sigma^k} \qquad (6.2.6)$$

for $k \leq s = 0, \ldots, m$ and the conditional distribution of $L_m^{(n)}$ is of the form

$$\mathbb{P}(L_m^{(n)} = s | K_n = j) = \binom{m}{s} (n - j\sigma)_{(m-s)\uparrow 1} \sum_{k=0}^{s} \frac{V_{n+m,j+k}}{V_{n,j}} \frac{\mathscr{C}(s, k; \sigma)}{\sigma^k} \tag{6.2.7}$$

for $s = 0, \ldots, m$.

From (6.2.6) and (6.2.7) one can also deduce other explicit forms for conditional distributions of interests. For example, the distribution of the number of observations in the "second sample" which lie in new partition sets, given the number of groups present in the "basic sample" and the number of new clusters $K_m^{(n)}$, is of the form

$$\mathbb{P}(L_m^{(n)} = s | K_m^{(n)} = k, K_n = j) = \frac{\binom{m}{s} (n - j\sigma)_{(m-s)\uparrow 1} \mathscr{C}(s, k; \sigma)}{\mathscr{C}(m, k; \sigma, -n + j\sigma)} \tag{6.2.8}$$

for $s = k, \ldots, m$, where $\mathscr{C}(n, k; \sigma, \gamma)$ is $\mathscr{C}(n, k; \sigma, \gamma) := (-1)^{n-k} C(n, k; \sigma, \gamma)$ (see Appendix A). It is worth noting that the previous expression does not depend on the particular Gibbs prior it is derived from: interestingly, Gibbs-type random partitions share the same conditional structures once $K_m^{(n)}$ and K_n are fixed. This finding is reminiscent of a result in Gnedin and Pitman [74] where the authors show that K_n is sufficient for Gibbs-type random partition of the first n integers meaning that the conditional distribution of the partition on $[n]$ given K_n does not depend on the weights $V_{n,k}$. On the other hand, the conditional distribution of $K_m^{(n)}$, given $L_m^{(n)}$ and K_n, is of the form

$$\mathbb{P}(K_m^{(n)} = k | L_m^{(n)} = s, K_n = j) = \frac{V_{n+m,j+k} \mathscr{C}(s, k; \sigma)/\sigma^k}{\sum_{l=0}^{s} V_{n+m,j+l} \mathscr{C}(s, l; \sigma)/\sigma^l} \tag{6.2.9}$$

for any $k \in [s]$. Moreover, evaluating the probability in P1) for a Gibbs-type r.p.m. can be obtained by marginalizing the conditional distribution in (6.2.5) with respect to $L_m^{(n)}$ and $S_{L_m^{(n)}}$. In particular, the Bayes estimator, under quadratic loss function, for the expected number of new cluster, proposed by Lijoi et al. [123], is easily recovered from (6.2.5) as

$$\mathbb{E}[K_m^{(n)} | K_n = j] = \sum_{k=0}^{m} k \frac{V_{n+m,j+k}}{V_{n,j}} \frac{\mathscr{C}(m, k; \sigma, -n + j\sigma)}{\sigma^k} \tag{6.2.10}$$

Often iterest relies also in determining an estimator for the number of observations in the "second sample" that will belong to new species. For instance, in genomic applications this can be seen as a better measure of redundancy of a certain library. For this purpose, one can resort to (6.2.5) and the corresponding Bayes estimator is given by

$$\mathbb{E}[L_m^{(n)} | K_n = j] = \sum_{s=0}^{m} s \binom{m}{s} (n - j\sigma)(n - j\sigma)_{(m-s)\uparrow 1} \sum_{k=0}^{s} \frac{V_{n+m,j+k}}{V_{n,j}} \frac{\mathscr{C}(s, k; \sigma)}{\sigma^k} \tag{6.2.11}$$

Then $\mathbb{E}[L_m^{(n)}|K_n = j]/m$ is the expected proportion of genes in the new sample which do not coincide with previously observed ones. The expression in (6.2.11) admints a noteworthy simplification as outlined in the following proposition: indeed, the Bayes estimator is m times the probability that the $(n+1)$-th draw yields a new cluster, given that j distinct clusters are generated by the first n observations.

Proposition 6.2.2. *(cfr. Lijoi et al. [125]) For any $j \in [n]$ and $m \geq 1$ one has*

$$\mathbb{E}[L_m^{(n)}|K_n = j] = m\frac{V_{n+1,j+1}}{V_{n,j}} \tag{6.2.12}$$

Turning to the evaluation of the probability in P2) for a Gibbs-type r.p.m., a Bayesian nonparametric estimator for the probability of discovering a new species at the $(n+m+1)$-th draw, given the "basic sample" has been provided by Lijoi et al. [123]. If we suppose, for the moment, that we have observed both the "basic sample" and the "second sample", the discovery probability is given by

$$\mathbb{P}(K_1^{(n+m)} = 1|K_m^{(n)} = k, L_m^{(n)} = s, S_{L_m^{(n)}} = \mathbf{s}, K_n = j, N_{K_n} = \mathbf{n})$$

and by virtue of the highlited sufficiency of the number of distinct species, the discovery probability is also equal to $\mathbb{P}(K_1^{(n+m)} = 1|K_m^{(n)} = k, K_n = j)$. However, our estimate is obtained without observing the outcome of the "second sample" and, hence, we have to estimate the random probability

$$D_m^{(n,j)} := \mathbb{P}(K_1^{(n+m)} = 1|K_m^{(n)}, K_n = j) \tag{6.2.13}$$

where the randomness in the above expression is due to the randomness of $K_m^{(n)}$. Baysian inference on (6.2.13) is based on the posterior distribution provided in Corollary (6.2.1). Thus, the Bayesian estimator of (6.2.13), with respect to a quadratic loss function, is given by its expected value with respect to the posterior distribution of the number of species. This represents a Bayesian counterpart to the celebrated Good-Toulmin estimator. In other word we provide a Bayesian nonparametric estimator for

$$U_{n+m} = \sum_{i\geq 1} p_i \mathbb{1}_{\{0\}}(N_{i,n+m})$$

where $N_{i,n+m}$ represents the number of population units from the i-th species in the "enlarged sample" of size $n+m$. In particular, we have the following proposition.

Proposition 6.2.3. *(cfr. Lijoi et al. [123]) Suppose that $\Pi = \Pi_n, n \geq 1$ is a Gibbs-type random partition with weights $V_{n,k}$ and parameter $[0,1)$. Then the Bayes estimate, under* *squared loss function, of the probability of observing a new species at the* $(n+m+1)$*-th draw, conditional on the "basic sample" with* j *distinct species, is given by*

$$\hat{D}_m^{(n,j)} = \sum_{k=0}^{m} \frac{V_{n+m+1,j+k+1}}{V_{n,j}} \frac{\mathscr{C}(m,k;\sigma,-n+j\sigma)}{\sigma^k} \tag{6.2.14}$$

All the quantities described up to now, depend on the analysis of the conditional structure of a Gibbs-type random partition. Investigation of the conditional structure for the sequence of blocks $\{K_n, n \geq 1\}$ is pursued in Gnedin and Pitman [74] where the authors do consider the conditional distribution of the number of groups in the partition of $[n]$, given the number of blocks in which $[n+m]$ is partitioned.

Example 6.2.1. *The Dirichlet process is a Gibbs-type r.p.m. with* $\sigma = 0$*. In particular, let* α *be a non-atomic measure on* $(\mathbb{X}, \mathscr{X})$ *such that* $\theta := \alpha(\mathbb{X}) > 0$ *and let* $X_1, \dots, X_n$ *be a sample of size n from an exchangeable* $\{X_n, n \geq 1\}$ *governed by a Dirichlet process. The EPPF induced by the sample* $X_1, \dots, X_n$ *is known to be of the form* (6.1.3) *and the prior distribution of the number of distinct species within a sample of size n, due to Ewens [43] and Antoniak [2], is obtained by letting* $\sigma \to 0$ *in* (6.2.4)*, which yields*

$$\mathbb{P}(K_n = k) = \frac{\theta^k}{(\theta)_{n\uparrow 1}} |s(n,k)| \tag{6.2.15}$$

where $|s(n,k)| := \lim_{\sigma \to 0} \mathscr{C}(n,k,\sigma)/\sigma^k$ *stands for the signless or absolute Stirling number of the first kind (see Appendix A). Moreover, the posterior distribution of the number of distinct species to be observed in the additional sample becomes*

$$\mathbb{P}(K_m^{(n)} = j | K_n = j) = \frac{(\theta)^k (\theta)_{n\uparrow 1}}{(\theta)_{(n+m)\uparrow 1}} \sum_{l=j}^{m} \binom{m}{l} |s(l,k)| (n)_{(m-l)\uparrow 1} \tag{6.2.16}$$

for any $k = 0, 1, \dots, m$*. Finally, the Bayesian estimator of the discovery probability reduces to*

$$\hat{D}_m^{(n,j)} = \frac{\theta}{(\theta+n)_{(m+1)\uparrow 1}} \sum_{k=0}^{m} \theta^k \sum_{l=k}^{m} \binom{m}{l} |s(l,k)| (n)_{(m-l)\uparrow 1}. \tag{6.2.17}$$

Interestingly (6.2.16) *and* (6.2.15) *solely depend on the sample size: prediction does not depend on* K_n *and* N_{K_n} *and so all this information has to be summarized by the parameter a. This, which is a characterizing property of the Dirichlet process (see Zabell [198]), represents a severe limitation for predictive purposes.*

We are going to consider an important quantity which describes the partition structure of observations generating new groups in a further sampling procedure, conditional on the partition generated by the first n observations. In particular, there is a sort of reproducibility of the Gibbs structure as established by the following proposition.

Proposition 6.2.4. *(cfr. Lijoi et al. [125]) Let $\Pi = \{\Pi_n, n \geq 1\}$ be a Gibbs-type exchangeable random partition whose EPPF is characterized by the set of weights $\{V_{n,k} : k = 1, \dots, n; n \geq 1\}$ and by the parameter $\sigma \in (0,1)$. Then*

$$\mathbb{P}(K_m^{(n)} = k, S_{L_m^{(n)}} = \mathbf{s} | L_m^{(n)} = s, K_n = j, N_{K_n} = \mathbf{n}) \tag{6.2.18}$$

$$= \frac{V_{n+m,j+k}}{\sum_{i=0}^{s} V_{n+m,j+i} \mathscr{C}(s,i;\sigma)/\sigma^i} \prod_{i=1}^{k} (1-\sigma)_{(s_i-1)\uparrow 1}$$

for any $s \in [m]$, $k \in [s]$, $j \in [n]$, $(n_1, \dots, n_j) \in \mathcal{D}_{j,n}$ and $(s_1, \dots, s_k) \in \mathcal{D}_{k,s}$. Consequently the partition of the observations which belong to the new partition sets is, conditional on the "basic sample" of size n, a finite Gibbs-type random partition with weights $\{V_{n,k}(m,n,j) : s = 1, \dots, m; k = 1, \dots, s\}$ defined by

$$V_{s,k}(m,n,j) = \frac{V_{n+m,j+k}}{\sum_{i=0}^{s} V_{n+m,j+i} \mathscr{C}(s,i;\sigma)/\sigma^i} \tag{6.2.19}$$

and with parameter $\sigma \in [0,1)$.

Note from (6.2.18), again, that

$$\mathbb{P}(K_m^{(n)} = k, S_{L_m^{(n)}} = \mathbf{s} | L_m^{(n)} = s, K_n = j, N_{K_n} = \mathbf{n})$$
$$= \mathbb{P}(K_m^{(n)} = k, S_{L_m^{(n)}} = \mathbf{s} | L_m^{(n)} = s, K_n = j)$$

The finiteness of the random partition described by (6.2.18) is obvious, since it takes values on the space of all partitions of $[s]$, with $1 \leq s \leq m$. Moreover, the particular structure featured by the conditional distribution in (6.2.18) motivates the following definition.

Definition 6.2.2. *(cfr. Lijoi et al. [125]) The conditional probability distribution*

$$\tilde{p}_k^{(s)}(s_1, \dots, s_k; m, n, j) := \mathbb{P}(K_m^{(n)} = k, S_{L_m^{(n)}} = \mathbf{s} | L_m^{(n)} = s, K_n = j) \tag{6.2.20}$$

with $1 \leq s \leq m$ and $1 \leq k \leq s$, is termed conditional EPPF.

Hence, the probability distribution in (6.2.18) is a conditional EPPF giving rise to a finite Gibbs-type random partition. Even if the structure of $\tilde{p}_k^{(s)}(s_1, \dots, s_k; m, n, j)$ is quite general, one might wonder whether it is possible to provide more information about its $V_{s,k}(m,n,j)$ weights in some particular cases. For example, it would be interesting to ascertain when $V_{s,k}(m,n,j)$ does not depend on m and n, so that $\tilde{p}_k^{(s)}(s_1, \dots, s_k; m, n, j) = \tilde{p}_k^{(s)}(s_1, \dots, s_k; j)$, which means that the conditional EPPF is that corresponding to an infinite Gibbs partition.

Having the conditional EPPF $\tilde{p}_k^{(s)}$ at hand, one can compute some other interesting conditional distributions in a straightforward way. For example, if one combines the expression for $\tilde{p}_k^{(s)}$ with Corollary 6.2.1 it is immediate to check that

$$\mathbb{P}(S_{L_m^{(n)}} = \mathbf{s} | K_m^{(n)} = k, L_m^{(n)} = s, K_n = j) = \frac{\sigma^k}{\mathscr{C}(s, k; \sigma)} \prod_{i=1}^{k} (1-\sigma)_{(s_i - 1)\uparrow 1}$$

in an expression for the conditional distribution of detecting a particular configuration $(s_1, \ldots, s_k)$ for the observations belonging to the new partition sets, given the number of new sets, the number of observations falling into these sets and the "basic sample".

All the sampling formulae we have deduced so far have important applications in Bayesian nonparametrics and in population genetics. In Bayesian nonparametric, random discrete priors are commonly employed in order to define a clustering structure either at the level of the observations or at the level of a latent variables in a complex hierarchical model. In particular, any EPPF corresponds to some random discrete prior and it represents, together with all the expressions for the conditional distributions we have obtained, a useful tool for specifying prior opinions on the clustering of the data. In population genetics, the concept of conditional EPPF can be seen as follows. Given a sample of size n containing j distinct species with absolute frequencies $n_1, \ldots, n_j$, a new sample of size m is to be drawn. Given that s of the m observations contribute to generating newly observed species, i.e. they belong to new distinct clusters, one might be interested in evaluating the probability that the s observations are grouped into k clusters with respective frequencies $s_1, \ldots, s_k$. The answer of such a question is provided by a conditional EPPF. The other distributions, discussed previously, provided a wide range of sampling formulae which answer similar type of problems. In the following subsection we focus attention on some noteworthy particular cases of Gibbs-type r.p.m.s, namely the two parameter Poisson-Dirichlet process and the normalized generalized Gamma process.

6.2.1 The two parameter Poisson-Dirichlet process

We start this section by introducing the two parameter Poisson-Dirichlet process (see Pitman [149] and Pitman and Yor [154]). Among the various possible definitions, a simple and intuitive one follows from the so-called stick-breaking construction. For a pair of parameters (σ, θ) such that $\sigma \in (0,1)$ and $\theta > -\sigma$, let $\{V_k, k \geq 1\}$ denote a sequence of independent r.v.s, with V_k distributed according to a Beta distribution function with parameter $(\theta + k\sigma,\, 1-\sigma)$. Define the stick-breaking weights as $\tilde{p}_1 = V_1$, $\tilde{p}_j = V_j \prod_{1 \leq i \leq j-1} (1 - \tilde{p}_i)$ and suppose $\{Y_n, n \geq 1\}$ is a sequence of i.i.d. r.v.s, which are independent of the $\tilde{p}_i$'s and whose common probability distribution α_0 is non-atomic. If δ_a is the point mass at a, the discrete r.p.m. $\tilde{P}_{(\sigma,\theta)} = \sum_{j \geq 1} \tilde{p}_j \, \delta_{Y_j}$ is a Poisson-Dirichlet process with parameter (σ, θ). See Pitman [157] for a detailed account on general theoretical aspects and, e.g., Ishwaran and James [90], Navarrete et al. [144] , Jara et al. [102] for applications in Bayesian nonparametrics.

As shown in Pitman [149], the EPPF induced by a sample $X_1, \ldots, X_n$ from a Poisson-Dirichlet process with parameter (σ, θ) is given by (6.2.1). Then, basing upon Proposition (6.2.1), one has

$$\mathbb{P}(K_m^{(n)} = k, L_m^{(n)} = s, S_{L_m^{(n)}} = \mathbf{s} | K_n = j) \tag{6.2.21}$$
$$= \frac{\prod_{i=0}^{k-1}(\theta + j\sigma + i\sigma)}{(\theta + n)_{m\uparrow 1}} \binom{m}{s} (n - j\sigma)_{(m-s)\uparrow 1} \prod_{i=1}^{k} (1-\sigma)_{(s_i - 1)\uparrow 1}$$

and it is possible to derive explicit expressions for all the sampling formulae set forth. In particular, the evaluation of the probability in P1) for the Poisson-Dirichlet process with parameter (σ, θ), can be obtained by marginalizing the conditional distribution in (6.2.21) with respect to $L_m^{(n)}$ and $S_{L_m^{(n)}}$. Here we provide an alternative derivation of (6.2.22).

Proposition 6.2.5. *Under the Poisson-Dirichlet process with parameter (σ, θ), one has*

$$\mathbb{P}(K_m^{(n)} = k | K_n = j) = \frac{(\theta+1)_{(n-1)\uparrow 1}}{(\theta+1)_{(n+m-1)\uparrow 1}} \frac{\prod_{i=j}^{j+k-1}(\theta + i\sigma)}{\sigma^k} \mathscr{C}(m, k; \sigma, -n + j\sigma) \tag{6.2.22}$$

Proof. An important result proved in Pitman [151] concerns the representation of the posterior distribution of $\tilde{P}_{\sigma,\theta}$, given a sample $X_1, \ldots, X_n$ of data governed by $\tilde{P}_{(\sigma,\theta)}$. Indeed, if the observations X_i are, conditional on $\tilde{P}_{(\sigma,\theta)}$, i.i.d. from $\tilde{P}_{(\sigma,\theta)}$ and the sample $X_1, \ldots, X_n$ contains $j \leq n$ distinct values $X_1^*, \ldots, X_j^*$, then

$$\tilde{P}_{\sigma,\theta} | X_1, \ldots, X_n \stackrel{d}{=} \sum_{i=1}^{j} w_i \delta_{X_i^*} + w_{j+1} \tilde{P}_{\sigma, \theta + j\sigma} \tag{6.2.23}$$

where $(w_1, \ldots, w_j)$ is distributed according to a j-variate Dirichlet distribution with parameters $(n_1 - \sigma, \ldots, n_j - \sigma, \theta + j\sigma)$, $n_i = \sharp\{r : X_r = X_i^*\}$ is the frequency of X_i^* in the sample and $w_{j+1} = 1 - \sum_{1 \leq i \leq j} w_i$. In order to derive (6.2.22), we will make use of the posterior representation given in (6.2.23), and to the distributional properties of K_i, for any i. Indeed, from (6.2.23) one notes that, given w a r.v. distributed according to a Beta distribution function with parameter $(\theta + j\sigma, n - j\sigma)$, an observation X_{n+i}, with $i = 1, \ldots, m$, does not coincide with any of the $K_n = j$ distinct species observed in the "basic sample" with probability w. Consequently

$$\mathbb{P}(K_m^{(n)} = k | K_n = j)$$
$$= \frac{\Gamma(\theta + n)}{\Gamma(\theta + j\sigma)\Gamma(n - j\sigma)} \int_0^1 \mathbb{P}(K_m^{(n)} = k | K_n = j, w) w^{\theta + j\sigma - 1} (1-w)^{n - j\sigma - 1} dw$$

In order to have $K_m^{(n)} = k$, at least k of the m data $X_{n+1}, \dots, X_{n+m}$ must be allocated to the k new distinct species not observed among the $K_n = j$ species of the "basic sample". Hence we have

$$\mathbb{P}(K_m^{(n)} = k | K_n = j, w) = \sum_{i=k}^{m} \binom{m}{i} w^i (1-w)^{m-i} \mathbb{P}(K_i = k)$$

where it is to be noted that K_i is, now, the number of distinct species among the i observations generated by a Poisson-Dirichlet process with parameter $(\sigma, \theta + j\sigma)$. Such a probability distribution has been derived in Pitman [155] (see also Pitman, [157]) and in this case yields

$$\mathbb{P}(K_i = k) = \frac{\prod_{l=1}^{k-1} (\theta + j\sigma + l\sigma)}{\sigma^k \, (\theta + j\sigma + 1)_{(i-1)\uparrow 1}} \mathscr{C}(i, k; \sigma) \qquad i = k, \dots, m$$

Summing up the previous considerations we obtain (6.2.22) by noting that

$$\begin{aligned} \mathbb{P}(K_m^{(n)} = k | K_n = j) &= \frac{(\theta/\sigma + j)_{k\uparrow 1}}{(\theta + n)_{m\uparrow 1}} \sum_{i=k}^{m} \binom{m}{i} \mathscr{C}(i, k; \sigma)(n - j\sigma)_{i\uparrow 1} \\ &= \frac{(\theta/\sigma + j)_{k\uparrow 1}}{(\theta + n)_{m\uparrow 1}} \mathscr{C}(m, k; \sigma, -n + j\sigma) \end{aligned}$$

where the second equality follows from (2.56) in Charalambides [17]. □

Based on (6.2.22), the estimators of interest can be derived. The most significant one is the expected number of new species $\hat{E}_m^{(n,j)} := \mathbb{E}[K_m^{(n)} | K_n = j]$ which represents a Bayesian nonparametric analog of the Good-Toulmin estimator (see Good and Toulmin [76]). Moreover, evaluating the probability in P2) interpreted as the probability that the $(n+m+1)$-th observation will yield a new species, without observing the m intermediate records, is given by

$$\hat{D}_m^{(n,j)} = \frac{(\theta+1)_{(n-1)\uparrow 1}}{(\theta+1)_{(n+m)\uparrow 1}} \sum_{k=0}^{m} \frac{\prod_{i=j}^{j+k} (\theta + i\sigma)}{\sigma^k} \mathscr{C}(m, k; \sigma, -n + j\sigma). \qquad (6.2.24)$$

Finally, we consider the sample coverage which is the proportion of species represented in a "basic sample" of size n. In particular, for the Poisson-Dirichlet process with parameter (σ, θ), the sample coverage is given by

$$\hat{C}_1^{(n,j)} = 1 - \frac{\theta + j\sigma}{\theta + n}.$$

which represents an alternative to the frequentist Turing estimator (see Good, [75]). Hence, the estimated sample coverage after $n + m$ draws is given by $\hat{C}_m^{(n,j)} = 1 - \hat{D}_m^{(n,j)}$. The advantage of the formulae yielding $\hat{E}_m^{(n,j)}$ and $\hat{D}_m^{(n,j)}$ is that they are explicit and can be exactly evaluated. There are, however, situations of practical interest where the size of the additional sample of interest is very large and the computational burden for evaluating (6.2.22) and (6.2.24) becomes heavy. This happens, for instance, in genomic applications where one has to deal with relevant portions of cDNA libraries which typically consist of millions of genes. Our first aim is the achievement of a considerable simplification of the two above mentioned estimators. Moreover, since (6.2.22) is still required for determining the corresponding highest posterior density (HPD) intervals, we will study the asymptotics of $K_m^{(n)}$, given K_n, as $m \to +\infty$: this allows one to use the distribution of the limiting random quantity in order to approximate the HPD intervals.

The first important result concerns the moments of $K_m^{(n)}$, given K_n, which will be expressed in terms of non-central Stirling numbers of the second kind $S(n,k;r)$ (see Appendix A). Such moments allow to derive simplified expressions for the estimators of interest.

Proposition 6.2.6. *Under the Poisson-Dirichlet process with parameter (σ,θ), one has*

$$\mathbb{E}[(K_m^{(n)})^r | K_n = j] = \sum_{\nu=0}^{r} (-1)^{r-\nu} \left(j + \frac{\theta}{\sigma}\right)_{\nu\uparrow 1} S\left(r,\nu;\frac{\theta}{\sigma}+j\right) \frac{(\theta+n+\nu\sigma)_{m\uparrow 1}}{(\theta+n)_{m\uparrow 1}} \quad (6.2.25)$$

In particular, a Bayesian nonparametric estimator of $K_m^{(n)}$ coincides with

$$\hat{E}_m^{(n,j)} = \left(j+\frac{\theta}{\sigma}\right)\left(\frac{(\theta+n+\sigma)_{m\uparrow 1}}{(\theta+n)_{m\uparrow 1}} - 1\right), \quad (6.2.26)$$

the discovery probability is equal to

$$\hat{D}_m^{(n,j)} = \frac{\theta+j\sigma}{\theta+n} \frac{(\theta+n+\sigma)_{m\uparrow 1}}{(\theta+n+1)_{m\uparrow 1}} \quad (6.2.27)$$

and the sample coverage after $n+m$ draws is given by

$$\hat{C}_m^{(n,j)} = 1 - \frac{\theta+j\sigma}{\theta+n} \frac{(\theta+n+\sigma)_{m\uparrow 1}}{(\theta+n+1)_{m\uparrow 1}}. \quad (6.2.28)$$

Proof. Indeed, one has

$$\mathbb{E}[(K_m^{(n)})^r | K_n = j, w] = \sum_{i=0}^{m} \binom{m}{i} w^i (1-w)^{m-i} \mathbb{E}[K_i^r]$$

where the unconditional moment $\mathbb{E}[K_i^r]$ is evaluated with respect to the $\tilde{P}_{(\sigma,\theta+j\sigma)}$ prior. Such an expression is already available from Pitman [151] and Yamato and Sibuya [189] and it is given by

$$\mathbb{E}[K_i^r] = \sum_{\nu=0}^{r} (-1)^{r-\nu} \left(1+\frac{\theta+j\sigma}{\sigma}\right)_{\nu\uparrow 1} S\left(r,\nu;\frac{\theta+j\sigma}{\sigma}\right) \frac{(\theta+j\sigma+\nu\sigma+1)_{(i-1)\uparrow 1}}{(\theta+1)_{(i-1)\uparrow 1}}$$

Hence, one has

$$
\begin{aligned}
&\mathbb{E}[(K_m^{(n)})^r|K_n=j]\\
&=\frac{\Gamma(\theta+n)}{\Gamma(\theta+j\sigma)\Gamma(n-j\sigma)}\int_0^1 w^{\theta+j\sigma-1}(1-w)^{n-j\sigma-1}\mathbb{E}[(K_m^{(n)})^r|K_n=j,w]dw\\
&=\frac{\Gamma(\theta+n)}{\Gamma(\theta+j\sigma)\Gamma(n-j\sigma)}\sum_{\nu=0}^{r}(-1)^{r-\nu}\left(1+\frac{\theta+j\sigma}{\sigma}\right)_{\nu\uparrow 1}S\left(r,\nu;\frac{\theta+j\sigma}{\sigma}\right)\\
&\qquad\times\sum_{i=0}^{m}\binom{m}{i}\frac{(\theta+j\sigma+\nu\sigma+1)_{i-1}}{(\theta+1)_{(i-1)\uparrow 1}}\int_0^1 w^{\theta+j\sigma+i-1}(1-w)^{n-j\sigma+m-i-1}dw\\
&=\frac{1}{(\theta+n)_{m\uparrow 1}}\sum_{\nu=0}^{r}(-1)^{r-\nu}\left(1+\frac{\theta+j\sigma}{\sigma}\right)_{\nu\uparrow 1}S\left(r,\nu;\frac{\theta+j\sigma}{\sigma}\right)\frac{\theta+j\sigma}{\theta+j\sigma+\nu\sigma}\\
&\qquad\times\sum_{i=0}^{m}\binom{m}{i}(\theta+j\sigma+\nu\sigma)_{i\uparrow 1}(n-j\sigma)_{(m-i)\uparrow 1}\\
&=\frac{1}{(\theta+n)_{m\uparrow 1}}\sum_{\nu=0}^{r}(-1)^{r-\nu}\left(\frac{\theta}{\sigma}+j\right)_{\nu\uparrow 1}S\left(r,\nu;\frac{\theta+j\sigma}{\sigma}\right)(\theta+n+\nu\sigma)_{m\uparrow 1}
\end{aligned}
$$

where the last equality follows by an application of the Chu-Vandermonde formula. The expression for the discovery probability in (6.2.27) is obtained by inserting (6.2.26) into Equation 9 of Lijoi et al. [121] and some simple algebra. □

These formulae greatly simplify those employed in Lijoi et al. [123] and can be evaluated for any choice of n and m. Note also that the estimator in (6.2.26) admits an interesting probabilistic interpretation. Indeed, one has that

$$\hat{E}_m^{(n,j)}=\mathbb{P}(X_{n+1}=\text{new}|K_n=j)\mathbb{E}_{\sigma,\theta+n}[K_m]$$

where $\mathbb{E}_{\sigma,\theta+n}[K_m]$ stands for the unconditional expected number of distinct species, among m observations, with respect to the probability distribution of a Poisson-Dirichlet process with parameter $(\sigma,\theta+n)$. Moments of any order of the unconditional distribution, i.e. $E[K_n^r]$, have been determined by Pitman [151] and Yamato and Sibuya [189] and are recovered from (6.2.25) by setting $n=j=0$.

The formulae outlined in Proposition 6.2.6 provide point estimators for quantities of inteest in species sampling problems. Besides them, one would also like to determine HPD intervals since they provide a measure of uncertainty related to the point estimates. However, this can be by no means an easy task, especially for large values of m. In order to overcome this drawback, we analyze the asymptotic behaviour of $K_m^{(n)}$, for fixed n and as $m\to+\infty$, and use the appropriate quantiles of the limiting r.v. to obtain an HPD interval. Results of this type for the unconditional distribution are already known and have been determined by Pitman [152] and Pitman [155]. See also Pitman [157]. In order to recall Pitman's result, let f_σ be the density function of a positive σ-stable r.v. and Y_q be, for any $q \geq 0$, a positive r.v. with density function

$$f_{Y_q}(y) = \frac{\Gamma(q\sigma+1)}{\sigma\Gamma(q+1)} y^{q-1-1/\sigma} \, f_\sigma(y^{-1/\sigma}). \tag{6.2.29}$$

One, then, has that $K_n/n^\sigma \to Y_{\theta/\sigma}$ almost surely, as $n \to \infty$. As we shall now see, conditioning on the outcome of a "basic sample" leads to a different limiting result.

Proposition 6.2.7. *Under the Poisson-Dirichlet process with parameter (σ, θ), conditional on $K_n = j$ one has*

$$\frac{K_m^{(n)}}{m^\sigma} \to Z_{n,j} \qquad a.s. \tag{6.2.30}$$

and in the p-th mean, for any $p > 0$, where $Z_{n,j} \stackrel{d}{=} B_{j+\theta/\sigma,\, n/\sigma-j}\, Y_{(\theta+n)/\sigma}$, $B_{a,b}$ is a r.v. distributed according to a Beta distribution function with parameter (a, b) and the r.v.s $B_{j+\theta/\sigma,\, n/\sigma-j}$ and $Y_{(\theta+n)/\sigma}$ are independent. Moreover,

$$\mathbb{E}\left[(Z_{n,j})^r\right] = \left(j + \frac{\theta}{\sigma}\right)_{r\uparrow 1} \frac{\Gamma(\theta+n)}{\Gamma(\theta+n+r\sigma)} \tag{6.2.31}$$

Proof. The proof strategy is as follows: we first adopt a technique similar to the one suggested in Theorem 3.8 in Pitman [157] for the unconditional case in order to establish that $K_m^{(n)}/m^\sigma$ converges a.s. and in the p-th mean for any $p > 0$. Then, we determine the moments of the limiting r.v. and show that the limiting r.v. is characterized by its moments. Let us start by computing the likelihood ratio

$$M_{\sigma,\theta,m}^{(n)} := \frac{d\mathbb{P}_{\sigma,\theta}^{(n)}}{d\mathbb{P}_{\sigma,0}^{(n)}}\Bigg|_{\mathscr{F}_m^{(n)}} = \frac{q_{\sigma,\theta}^{(n)}(K_m^{(n)})}{q_{\sigma,0}^{(n)}(K_m^{(n)})}$$

where $\mathscr{F}_m^{(n)} = \sigma(X_{n+1}, \ldots, X_{n+m})$, $\mathbb{P}_{\sigma,\theta}^{(n)}$ is the conditional probability distribution of a Poisson-Dirichlet process with parameter (σ, θ) given K_n and, by virtue Proposition 1 in Lijoi, Prünster and Walker [125], $q_{\sigma,\theta}^{(n)}(k) = \sigma^{K_n}(\theta/\sigma + K_n)_{k\uparrow 1}/(\theta+n)_m$ for any integer $k \geq 1$ and $q_{\sigma,\theta}^{(n)}(0) := 1/(\theta+n)_{m\uparrow 1}$. Hence $\{(M_{\sigma,\theta,m}^{(n)}, \mathscr{F}_m^{(n)}), m \geq 1\}$ is a $\mathbb{P}_{\sigma,0}^{(n)}$-martingale. By a martingale convergence theorem (see Billingsley [8]), $M_{\sigma,\theta,m}^{(n)}$ has a $\mathbb{P}_{\sigma,0}^{(n)}$ almost sure limit, say $M_{\sigma,\theta}^{(n)}$, as $m \to +\infty$. Convergence holds in the p-th mean as well, for any $p > 0$. One clearly has that $\mathbb{E}_{\sigma,0}^{(n)}[M_{\sigma,\theta}^{(n)}] = 1$, where $\mathbb{E}_{\sigma,0}^{(n)}$ denotes the expected value with respect to $\mathbb{P}_{\sigma,0}^{(n)}$. It can be easily seen that

$$M_{\sigma,\theta,m}^{(n)} \sim \frac{\Gamma(\theta+n)\Gamma(K_n)}{\Gamma(n)\Gamma(\theta/\sigma+K_n)} \left(\frac{K_m^{(n)}}{m^\sigma}\right)^{\theta/\sigma}$$

as $m \to \infty$. Hence $(K_m^{(n)}/m^\sigma)^{\theta/\sigma}$ converges $\mathbb{P}_{\sigma,0}^{(n)}$-a.s. to a r.v., say $Z_{n,j}$ such that

$$\mathbb{E}_{\sigma,0}^{(n)}[Z_{n,j}^{\theta/\sigma}] = \frac{\Gamma(n)\Gamma(\theta/\sigma + K_n)}{\Gamma(\theta+n)\Gamma(K_n)}.$$

In order to identify the distribution of the limiting r.v. $Z_{n,j}$ with respect to the $\mathbb{P}_{\sigma,\theta}^{(n)}$, we consider the asymptotic behaviour of $\mathbb{E}[(K_m^{(n)})^r | K_n = j]$ as $m \to +\infty$, for any $r \geq 1$. Letting $m \to +\infty$ in (6.2.25) of Proposition 6.2.6, use the Stirling formula to obtain

$$\frac{1}{m^{r\sigma}}\mathbb{E}[(K_m^{(n)})^r | K_n] \to \left(K_n + \frac{\theta}{\sigma}\right)_{r\uparrow 1} \frac{\Gamma(\theta+n)}{\Gamma(\theta+n+r\sigma)} =: \mu_r^{(n)}. \tag{6.2.32}$$

Such a moment sequence clearly arises by taking $Z_{n,j} \stackrel{d}{=} B_{j+\theta/\sigma,\, n/\sigma - j}\, Y_{(\theta+n)/\sigma}$, with the r.v. $B_{j+\theta/\sigma, n/\sigma-j}$ is independent from the r.v. $Y_{(\theta+n)/\sigma}$, which has density (6.2.29). Hence, we are left with showing that the distribution of $Z_{n,j}$ is uniquely characterized by the moment sequence $\{\mu_r^{(n)}, r \geq 1\}$. In order to establish this, one can evaluate the characteristic function of $Z_{n,j}$ which, at any $t \in \mathbb{R}$, coincides with

$$\begin{aligned}
\Phi(t) &= \frac{\Gamma(\theta + n/\sigma)}{\Gamma(K_n + \theta/\sigma)\Gamma(n/\sigma - K_n)} \frac{\Gamma(\theta+n+1)}{\Gamma(\theta+n/\sigma+1)} \\
&\qquad \times \int_0^{+\infty} \mathrm{e}^{\mathrm{i}tz}\, z^{K_n+\theta/\sigma-1} \int_z^{+\infty} w(w-z)^{n/\sigma - K_n - 1}\, g_\sigma(w) dw dz \\
&= \frac{\sigma\,\Gamma(\theta+n)}{\Gamma(K_n+\theta/\sigma)\Gamma(n/\sigma - K_n)} \\
&\qquad \times \int_0^{+\infty} w\, g_\sigma(w) \int_0^w \mathrm{e}^{\mathrm{i}tz}\, z^{K_n+\theta/\sigma-1}(w-z)^{n/\sigma-K_n-1} dz dw \\
&= \frac{\Gamma(\theta+n+1)}{\Gamma(\theta+n/\sigma+1)} \sum_{r\geq 0} \frac{(\mathrm{i}t)^r}{r!} \frac{(K_n+\theta/\sigma)_{r\uparrow 1}}{(\theta+n/\sigma)_{r\uparrow 1}} \int_0^{+\infty} w^{\theta+n/\sigma+r} g_\sigma(w) dw \\
&= \sum_{r\geq 0} \frac{(\mathrm{i}t)^r}{r!} \frac{(K_n+\theta/\sigma)_{r\uparrow 1}}{(\theta+n/\sigma)_{r\uparrow 1}} \frac{\Gamma(\theta+n+1)}{\Gamma(\theta+n/\sigma+1)} \frac{\Gamma(\theta+n/\sigma+r+1)}{\Gamma(\theta+n+1+r\sigma)} \\
&= \sum_{r\geq 0} \frac{(\mathrm{i}t)^r}{r!}\, \mu_r^{(n)}
\end{aligned}$$

and the conclusion follows. □

It is worth stressing that the limiting r.v. in the conditional case is the same as in the unconditional case but with updated parameters and a rescaling induced by a r.v. distributed according to a Beta distribution function. The density of $Z_{n,j}$ in (6.2.30) can be formally represented as

$$f_{Z_{n,j}}(z) = \frac{\Gamma(\theta+n)}{\Gamma(\theta/\sigma+j)\Gamma(n/\sigma-j)} z^{\theta/\sigma+j-1} \int_z^{+\infty} v^{-1/\sigma}(v-z)^{n/\sigma-j-1} f_\sigma(v^{-1/\sigma}) dv.$$

When $\sigma = 1/2$, the density $f_{1/2}$ is known explicitly and the previous expression can be simplified to

$$f_{Z_{n,j}}(z) = \frac{\Gamma(\theta+n)4^{n+\theta-1}\, z^{\theta+k/2-1}}{\pi^{1/2}\Gamma(k+2\theta)\Gamma(2n-k)} \sum_{j=0}^{2n-k-1} \binom{2n-k-1}{j} (-z)^{j/2}\, \Gamma\left(n - \frac{k-1+j}{2};\, z\right).$$

Nonetheless, even in the latter case one cannot easily determine the quantiles of $Z_{n,j}$ we need to use in order to determine HPD intervals. Hence, we resort to a simulation algorithm for generating values of $Z_{n,j}$ and use the output to evaluate quantiles. The demanding part of this simulation is drawing samples from the probability distribution of Y_q. Note that the sampling strategy we are going to outline is also useful in the unconditional case, where the same tractability issue in deriving properties of Y_q is to be faced. The basic idea consists in setting $W_q = Y_q^{-1/\sigma}$ so that W_q has density function given by

$$f_{W_q}(w) = \frac{\sigma\Gamma(q\sigma)}{\Gamma(q)} w^{-q\sigma} f_\sigma(w) = \frac{\sigma}{\Gamma(q)} f_\sigma(w) \int_0^{+\infty} u^{q\sigma-1}\, \mathrm{e}^{-uw} du$$

Via augmentation, one then has

$$f_{(U_q,W_q)}(u,w) = \frac{\sigma}{\Gamma(q)}\, f_\sigma(w)\, u^{q\sigma-1}\, \mathrm{e}^{-uw} = f_{U_q}(u) f_\sigma(w|u)$$

where $f(u)$ is the density function of a r.v. U_q such that U_q^σ is distributed according to a Gamma distribution function with parameter $(q,1)$ and

$$f_\sigma(w|u) = f_\sigma(w)\, \mathrm{e}^{-uw+u^\sigma}.$$

This means that, conditional on U_q, W_q is a positive tempered-stable r.v. according to the terminology adopted in Rosinski [169]. In order to draw samples from a tempered stable r.v., a convenient strategy is to resort to the series representation derived in Rosinski [169], which, in our case, yields

$$W_q|U_q \stackrel{d}{=} \sum_{i=1}^{\infty} \min\{(a_i\Gamma(1-\sigma))^{-1/\sigma}, e_i\, v_i^{1/\sigma}\} \tag{6.2.33}$$

where $\{e_i, i \geq 1\}$ is a sequence of i.i.d. r.v.s such that e_i is distributed according to an Exponential distribution function with parameter U_q, $\{v_i, i \geq 1\}$ is a sequence of i.i.d. r.v.s such that v_i is distributed according to a Uniform distribution function on $(0,1)$ and $a_1 > a_2 > \cdots$ are the arrival times of a Poisson process with unit intensity. Other possibilities for simulating from a tempered stable r.v. are the inverse Lévy measure method as described in Ferguson and Klass [64] and a compound Poisson approximation scheme proposed in Cont and Tankov [20].

Summarizing the above considerations, an algorithm for simulating from the limiting r.v. $Z_{n,j} \stackrel{d}{=} B_{j+\theta/\sigma, n/\sigma-j}\, Y_{(\theta+n)/\sigma}$ is as follows:

1) Generate B from a Beta distribution with parameter $(j+\theta/\sigma, n/\sigma - j)$;

2) in order to sample from $Y_{(\theta+n)/\sigma}$:

   2.a) generate X from a Gamma distribution with parameter $((\theta+n)/\sigma, 1)$ and set $U = X^{1/\sigma}$;

   2.b) for a given truncation N and U sampled in step 2.a, generate the sequences $\{e_i, i \geq 1\}$ and $\{v_i, i \geq 1\}$ and take $a_i = \sum_{1\leq j \leq i} \xi_j$, for $i = 1, \ldots, N$;

   2.c) compute W according to (6.2.33) and set $Y = W^{-\sigma}$.

3) Take $Z = BY$.

Note that, in order to establish whether a chosen truncation threshold N for the series in step 2.b) is large enough, one can compare the sample moments with the simple exact moments of $Z_{n,j}$ given in (6.2.31).

We conclude the subsection devoted to the two parameter Poisson-Dirichlet process by showing some results for the factorial moments of the distribution of K_n and the distribution of $K_m^{(n)}|K_n$. In particular, the results for the factorial moments of K_n were provided by Yamato and Sibuya [189]. We extend these results to the distribution of $K_m^{(n)}|K_n$.

Proposition 6.2.8. *(cfr. Yamato and Sibuya [189]) Under the Poisson-Dirichlet process with parameter (σ, θ), the r-th descending factorial moment of K_n is*

$$\mathbb{E}[(K_n)_{r\downarrow 1}] = \left(\frac{\theta}{\sigma}\right)_{r\uparrow 1} \sum_{j=0}^{r} (-1)^{r-j} \binom{r}{j} \frac{(\theta + j\sigma)_{n\uparrow 1}}{\theta_{n\uparrow 1}}$$

Corollary 6.2.2. *(cfr. Yamato and Sibuya [189]) Under the Poisson-Dirichlet process with parameter (σ, θ), the r-th moment of K_n is*

$$\mathbb{E}[K_n^r] = \sum_{j=0}^{r} (-1)^{r-j} \left(1 + \frac{\theta}{\sigma}\right) R\left(r, j, \frac{\theta}{\sigma}\right) \frac{(\theta + j\sigma + 1)_{(n-1)\uparrow 1}}{(\theta+1)_{(n-1)\uparrow 1}}$$

where $R(r, j, \theta/\sigma)$ is the unique function satisfying

$$\sum_{j=0}^{r} (y)_{j\downarrow 1} R\left(r, j, \frac{\theta}{\sigma}\right) = \left(y + \frac{\theta}{\sigma}\right)^r$$

for any y.

Proposition 6.2.9. *(cfr. Yamato and Sibuya [189]) Under the Poisson-Dirichlet process with parameter (σ, θ), the r-th ascending factorial moment of K_n is*

$$\mathbb{E}[(K_n)_{r\uparrow 1}] = \sum_{j=0}^{r} (-1)^{r-j} \binom{r}{j} \left(\frac{\theta}{\alpha}\right)_{j\uparrow 1} \left(\frac{\theta}{\alpha}\right)_{(r-j)\uparrow 1} \frac{(\theta + j\alpha)_{n\uparrow 1}}{\theta_{n\uparrow 1}}$$

Before extending the results in Yamato and Sibuya [189] to the distribution of $K_m^{(n)}|K_n$, let us consider the following lemma.

Lemma 6.2.1. *For any $n \geq 0$, $k > 0$ the following identity holds true*

$$C(n+1,k;\sigma,r) = (\sigma k + r - n)C(n,k;\sigma,r) + \sigma C(n,k-1;\sigma,r) \tag{6.2.34}$$

Proof. By the Chu-Vandermonde equation we can express the non-central generalized factorial coefficient in terms of generalized factorial coefficient

$$C(n+1,k;\sigma,r) = \sum_{s=k}^{n+1} \binom{n+1}{s} C(s,k;\sigma)(r)_{(n+1-s)\downarrow 1}$$

and using the following important recurrence in the Pascal triangle (see Comtet [19])

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \qquad k,n \geq 1$$

we obtain

$$\sum_{s=k}^{n+1} \binom{n+1}{s} C(s,k;\sigma)(r)_{(n+1-s)\downarrow 1} \tag{6.2.35}$$
$$= \sum_{s=k}^{n+1} \binom{n}{s-1} C(s,k;\sigma)(r)_{(n+1-s)\downarrow 1} + \sum_{s=k}^{n+1} \binom{n}{s} C(s,k;\sigma)(r)_{(n+1-s)\downarrow 1}$$

As regard the first factor in the sum in (6.2.35) we have

$$\begin{aligned}
&\sum_{s=k}^{n+1} \binom{n}{s-1} C(s,k;\sigma)(r)_{(n+1-s)\downarrow 1} \\
&\qquad = \sum_{s=k-1}^{n} \binom{n}{s} C(s+1,k;\sigma)(r)_{(n-s)\downarrow 1} \\
&\qquad = \sum_{s=k-1}^{n} \binom{n}{s} ((\sigma k - s)C(s,k;\sigma) + \sigma C(s,k-1;\sigma))(r)_{(n-s)\downarrow 1} \\
&\qquad = -\sum_{s=k}^{n} \binom{n}{s} sC(s,k;\sigma)(r)_{(n-s)\downarrow 1} + \sigma k C(n,k;\sigma,r) + \sigma C(n,k-1;\sigma,r)
\end{aligned}$$

As regard the second factor in the sum in (6.2.35) we have

$$\begin{aligned}
\sum_{s=k}^{n+1} \binom{n}{s} C(s,k;\sigma)(r)_{(n+1-s)\downarrow 1} &= \sum_{s=k}^{n} (-r-n+s) \binom{n}{s} C(s,k;\sigma)(r)_{(n-s)\downarrow 1} \\
&= (-r-n)C(n,k;\sigma,r) + \sum_{s=k}^{n} s \binom{n}{s} C(s,k;\sigma)(r)_{(n-s)\downarrow 1}
\end{aligned}$$

Then, we have

$$\begin{aligned}C(n+1,k;\sigma,r) &= \sigma k C(n,k;\sigma,r) + \sigma C(n,k-1;\sigma,r) + (r-n)C(n,k;\sigma,r)\\ &= (\sigma k + r - n)C(n,k;\sigma,r) + \sigma C(n,k-1;\sigma,r)\end{aligned}$$

□

Using Lemma 6.2.10, the following results extend to the distribution of $K_m^{(n)}|K_n$ Proposition 6.2.8, Corollary 6.2.2 and Proposition 6.2.9, respectively. In particular, we provide an alternative proof for (6.2.25).

Proposition 6.2.10. *Under the Poisson-Dirichlet process with parameter (σ,θ), the r-th descending factorial moment of $K_m^{(n)}|K_n = j$ is*

$$\mathbb{E}[(K_m^{(n)})_{r\downarrow 1}|K_n = j] = \left(j + \frac{\theta}{\sigma}\right)_{r\uparrow 1} \sum_{l=0}^{r} (-1)^{r-l} \binom{r}{l} \frac{(\theta+n+l\sigma)_{m\uparrow 1}}{(\theta+n)_{m\uparrow 1}}$$

Proof. Using the definition of r-th descending factorial moment of $K_m||K_n = j$ we have

$$\mathbb{E}[(K_m^{(n)})_{r\downarrow 1}|K_n = j] = \sum_{k=0}^{m+1} (k)_{r\downarrow 1} \frac{(\theta+1)_{(n-1)\uparrow 1}\prod_{i=j}^{j+k-1}(\theta+i\sigma)}{\sigma^k(\theta+1)_{(n+m)\uparrow 1}} \mathscr{C}(m+1,k;\sigma,-n+j\sigma)$$

and by using Lemma 6.2.9 we obtain

$$\begin{aligned}&\sum_{k=0}^{m+1} (k)_{r\downarrow 1} \frac{(\theta+1)_{(n-1)\uparrow 1}(-1)^{(m+1-k)}\prod_{i=j}^{j+k-1}(\theta+i\sigma)}{\sigma^k(\theta+1)_{(n+m)\uparrow 1}} \\ &\qquad\times ((\sigma k - n + j\sigma - m)C(m,k;\sigma,-n+j\sigma) + \sigma C(m,k-1;\sigma,-n+j\sigma))\end{aligned} \tag{6.2.36}$$

As regard the first factor in (6.2.36) we have

$$\begin{aligned}&\sum_{k=0}^{m+1} (k)_{r\downarrow 1} \frac{(\theta+1)_{(n-1)\uparrow 1}(-1)^{(m+1-k)}\prod_{i=j}^{j+k-1}(\theta+i\sigma)}{\sigma^k(\theta+1)_{(n+m)\uparrow 1}} \\ &\qquad\qquad\times (\sigma k - n + j\sigma - m)C(m,k;\sigma,-n+j\sigma)\\ &= -\sum_{k=0}^{m+1} (k)_{r\downarrow 1} \frac{(\theta+1)_{(n-1)\uparrow 1}\prod_{i=j}^{j+k-1}(\theta+i\sigma)}{\sigma^k(\theta+1)_{(n+m)\uparrow 1}} (\sigma k - n + j\sigma - m)\mathscr{C}(m,k;\sigma,-n+j\sigma)\\ &= -\sum_{k=0}^{m} (k)_{r\downarrow 1} \frac{(\theta+1)_{(n-1)\uparrow 1}\prod_{i=j}^{j+k-1}(\theta+i\sigma)}{\sigma^k(\theta+1)_{(n+m)\uparrow 1}} (\sigma k - n + j\sigma - m)\mathscr{C}(m,k;\sigma,-n+j\sigma)\\ &= -\frac{1}{(\theta+m+n)} \sum_{k=0}^{m} (k)_{r\downarrow 1} \frac{(\theta+1)_{(n-1)\uparrow 1}\prod_{i=j}^{j+k-1}(\theta+i\sigma)}{\sigma^k(\theta+1)_{(n+m-1)\uparrow 1}}\\ &\qquad\qquad\times (\sigma k - n + j\sigma - m)\mathscr{C}(m,k;\sigma,-n+j\sigma)\end{aligned}$$

As regard the second factor in (6.2.36) we have

$$\sum_{k=0}^{m+1}(k)_{r\downarrow 1}\frac{(\theta+1)_{(n-1)\uparrow 1}(-1)^{(m+1-k)}\prod_{i=j}^{j+k-1}(\theta+i\sigma)}{\sigma^k(\theta+1)_{(n+m)\uparrow 1}}\sigma C(m,k-1;\sigma,-n+j\sigma)$$
$$=\sum_{k=1}^{m+1}(k)_{r\downarrow 1}\frac{(\theta+1)_{(n-1)\uparrow 1}\prod_{i=j}^{j+k-1}(\theta+i\sigma)}{\sigma^k(\theta+1)_{(n+m)\uparrow 1}}\sigma\mathscr{C}(m,k-1;\sigma,-n+j\sigma)$$
$$=\sum_{k=0}^{m}(k+1)_{r\downarrow 1}\frac{(\theta+1)_{(n-1)\uparrow 1}\prod_{i=j}^{j+k}(\theta+i\sigma)}{\sigma^k(\theta+1)_{(n+m)\uparrow 1}}\mathscr{C}(m,k;\sigma,-n+j\sigma)$$
$$=\frac{1}{(\theta+n+m)}\sum_{k=0}^{m}(k+1)_{r\downarrow 1}(\theta+(j+k)\sigma)$$
$$\times\frac{(\theta+1)_{(n-1)\uparrow 1}\prod_{i=j}^{j+k-1}(\theta+i\sigma)}{\sigma^k(\theta+1)_{(n+m-1)\uparrow 1}}\mathscr{C}(m,k;\sigma,-n+j\sigma)$$

Then, we have

$$\mathbb{E}[(K_m^{(n)})_{r\downarrow 1}|K_n=j]=\frac{1}{(\theta+n+m)}\sum_{k=0}^{m}\frac{(\theta+1)_{(n-1)\uparrow 1}\prod_{i=j}^{j+k-1}(\theta+i\sigma)}{\sigma^k(\theta+1)_{(n+m-1)\uparrow 1}}$$
$$\times(-(k)_{r\downarrow 1}(\sigma k-n+j\sigma-m)+(k+1)_{r\downarrow 1}(\theta+(j+k)\sigma))$$
$$=\frac{1}{(\theta+n+m)}\sum_{k=0}^{m}\frac{(\theta+1)_{(n-1)\uparrow 1}\prod_{i=j}^{j+k-1}(\theta+i\sigma)}{\sigma^k(\theta+1)_{(n+m-1)\uparrow 1}}$$
$$\times(k)_{(r-1)\downarrow 1}(-(k-r+1)(\sigma k-n+j\sigma-m)+(k+1)(\theta+(j+k)\sigma))$$
$$=\frac{1}{(\theta+n+m)}\sum_{k=0}^{m}\frac{(\theta+1)_{(n-1)\uparrow 1}\prod_{i=j}^{j+k-1}(\theta+i\sigma)}{\sigma^k(\theta+1)_{(n+m-1)\uparrow 1}}$$
$$\times(k)_{(r-1)\downarrow 1}((k-r+1)(m+n+r\sigma+\theta)+r(\sigma(-1+j+r)+\theta))$$

In particular, we obtain the following recurrence relation for the r-th descending factorial moment of $K_m^{(n)}|K_n=j$

$$\mathbb{E}[(K_{m+1}^{(n)})_{r\downarrow 1}|K_n=j] \tag{6.2.37}$$
$$=(1+\frac{r\sigma}{(\theta+n+m)})\mathbb{E}[(K_m^{(n)})_{r\downarrow 1}|K_n=j]+\frac{r(\sigma(-1+j+r)+\theta)}{(\theta+n+m)}\mathbb{E}[(K_m^{(n)})_{(r-1)\downarrow 1}|K_n=j]$$

Then, using the recurrence relation (6.2.37) the result follows by simple induction on m and on r. □

Corollary 6.2.3. *Under the Poisson-Dirichlet process with parameter (σ,θ), the r-th moment of $K_m^{(n)}|K_n=j$ is*

$$\mathbb{E}[(K_m^{(n)})^r|K_n=j]=\sum_{i=0}^{r}(-1)^{r-i}\left(j+\frac{\theta}{\sigma}\right)_{i\uparrow 1}R\left(r,i,\frac{\theta}{\sigma}+j\right)\frac{(\theta+n+i\sigma)_{m\uparrow 1}}{(\theta+n)_{m\uparrow 1}}$$

where $R(r,i,\theta/\sigma+j)$ is the unique function satisfying

$$\sum_{i=0}^{r}(y)_{j\downarrow 1}R\left(r,i,\frac{\theta}{\sigma}+j\right)=\left(y+\frac{\theta}{\sigma}+j\right)^{r}$$

for any y.

Proof. Using the (r,l)-th Stirling number of the second kind $S(r,l)$ we can write

$$\mathbb{E}[(K_m^{(n)})^r|K_n=j]=\sum_{l=0}^{r}S(r,l)\mathbb{E}[(K_m)_{l\downarrow 1}|K_n=j]$$

and by Proposition 6.2.10 we have

$$\begin{aligned}
&\mathbb{E}[(K_m^{(n)})^r|K_n=j]\\
&\quad=\sum_{l=0}^{r}S(r,l)\left(j+\frac{\theta}{\sigma}\right)_{l\uparrow 1}\sum_{i=0}^{l}(-1)^{l-i}\binom{l}{i}\frac{(\theta+n+i\sigma)_{m\uparrow 1}}{(\theta+n)_{m\uparrow 1}}\\
&\quad=\sum_{i=0}^{r}\left(j+\frac{\theta}{\sigma}\right)_{i\uparrow 1}(-1)^{r-i}\frac{(\theta+n+i\sigma)_{m\uparrow 1}}{(\theta+n)_{m\uparrow 1}}\sum_{l=i}^{r}\binom{l}{i}S(r,l)\left(j+i+\frac{\theta}{\sigma}\right)_{l-i\uparrow 1}(-1)^{r+l}\\
&\quad=\sum_{i=0}^{r}(-1)^{r-i}\left(j+\frac{\theta}{\sigma}\right)_{i\uparrow 1}T\left(r,i,\frac{\theta}{\sigma}+j\right)\frac{(\theta+n+i\sigma)_{m\uparrow 1}}{(\theta+n)_{m\uparrow 1}}
\end{aligned}$$

where

$$T\left(r,i,\frac{\theta}{\sigma}+j\right)=\sum_{l=i}^{r}\binom{l}{i}S(r,l)\left(j+i+\frac{\theta}{\sigma}\right)_{l-i\uparrow 1}(-1)^{r+l}$$

For any y

$$\begin{aligned}
\sum_{i=0}^{r}(y)_{i\downarrow 1}T\left(r,i,\frac{\theta}{\sigma}+j\right)&=\sum_{l=0}^{r}\left(\sum_{i=0}^{l}\binom{l}{i}(y)_{i\downarrow 1}\left(\frac{\theta}{\sigma}+j+l-1\right)_{(l-i)\downarrow 1}\right)S(r,l)(-1)^{r+l}\\
&=\sum_{l=0}^{r}\left(y+\frac{\theta}{\sigma}+j+l-1\right)_{l\downarrow 1}(-1)^{l}S(r,l)(-1)^{r}
\end{aligned}$$

which is equal to $\sum_{0\le l\le r}(-y-\theta/\sigma-j)_{l\downarrow 1}S(r,l)(-1)^r=(-y-\theta/\sigma-j)^r(-1)^r=(y+\theta/\sigma+j)^r$. Then, for any y

$$\sum_{i=0}^{r}(y)_{i\downarrow 1}T\left(r,i,\frac{\theta}{\sigma}+j\right)=\left(y+\frac{\theta}{\sigma}+j\right)^{r}$$

Then, $T(r,i,\theta/\sigma+j)$ corresponds to the function $R(r,i,\lambda)$ with $\lambda=\theta/\sigma+j$ introduced by Carlitz [14] (or non-central Stirling number of the second kind). The result follows by substitution. □

Proposition 6.2.11. *Under the Poisson-Dirichlet process with parameter* (σ, θ)*, the* r*-th ascending factorial moment of* $K_m^{(n)}|K_n = j$ *is*

$$\mathbb{E}[(K_m)_{r\uparrow 1}|K_n = j] = \sum_{i=0}^{r}(-1)^{r-i}\left(j+\frac{\theta}{\sigma}\right)_{i\uparrow 1}\binom{r}{i}\left(\frac{\theta}{\sigma}+j\right)_{(r-i)\downarrow 1}\frac{(\theta+n+i\sigma)_{m\uparrow 1}}{(\theta+n)_{m\uparrow 1}}$$

Proof. Using the (r,l)-th signless Stirling number of the first kind $|s(r,l)|$ we can write

$$\mathbb{E}[(K_m)_{r\uparrow 1}|K_n = j] = \sum_{l=0}^{r}|s(r,l)|\mathbb{E}[(K_m^{(n)})^l|K_n = j]$$

and by Corollary 6.2.3 we have

$$\begin{aligned}
\mathbb{E}[(K_m)_{r\uparrow 1}|K_n = j] &= \sum_{l=0}^{r}|s(r,l)|\sum_{i=0}^{l}(-1)^{l-i}\left(j+\frac{\theta}{\sigma}\right)_{i\uparrow 1}R\left(l,i,\frac{\theta}{\sigma}+j\right)\frac{(\theta+n+i\sigma)_{m\uparrow 1}}{(\theta+n)_{m\uparrow 1}}\\
&= \sum_{i=0}^{r}\sum_{l=i}^{r}|s(r,l)|(-1)^{l-i}\left(j+\frac{\theta}{\sigma}\right)_{i\uparrow 1}R\left(l,i,\frac{\theta}{\sigma}+j\right)\frac{(\theta+n+i\sigma)_{m\uparrow 1}}{(\theta+n)_{m\uparrow 1}}\\
&= \sum_{i=0}^{r}(-1)^{r-i}\left(j+\frac{\theta}{\sigma}\right)_{i\uparrow 1}I\left(r,i,\frac{\theta}{\sigma}+j\right)\frac{(\theta+n+i\sigma)_{m\uparrow 1}}{(\theta+n)_{m\uparrow 1}}
\end{aligned}$$

where

$$I\left(r,i,\frac{\theta}{\sigma}+j\right) = \sum_{l=i}^{r}s(r,l)R\left(l,i,\frac{\theta}{\sigma}+j\right)$$

For any μ

$$\sum_{i=0}^{r}I\left(r,i,\frac{\theta}{\sigma}+j\right)(\mu)_{i\downarrow 1} = \sum_{l=0}^{r}\left(\sum_{i=0}^{l}R(l,i,\frac{\theta}{\sigma}+j)(\mu)_{i\downarrow 1}\right)s(r,l)$$

which is equal to $\sum_{0\le l\le r}(\theta/\sigma+j+\mu)^l s(r,l) = (\theta/\sigma+j+\mu)_{r\downarrow 1}$. Then, for any μ

$$\sum_{i=0}^{r}I\left(r,i,\frac{\theta}{\sigma}+j\right)(\mu)_{i\downarrow 1} = \left(\frac{\theta}{\sigma}+j+\mu\right)_{r\downarrow 1}$$

Since then $\theta/\sigma + j \neq 0$, then $I(r,i,\theta/\sigma+j) = \binom{r}{i}(\theta/\sigma+j)_{(r-i)\downarrow 1}$. □

6.2.2 The normalized generalized Gamma process

We start this section by introducing the normalized generalized Gamma process (see Pitman [156], James [96] and Lijoi et al. [120]). Consider a generalized Gamma CRM (see Brix [13]) which is characterized by the Poisson intensity measure

$$\nu(ds,dx) = \frac{\sigma}{\Gamma(1-\sigma)}s^{-1-\sigma}\mathrm{e}^{-\tau s}\alpha(dx) \tag{6.2.38}$$

where $\sigma \in (0,1)$ and $\tau > 0$. Let us denote this CRM by $\tilde{\mu}_{(\sigma,\tau)}$. Note that if $\tau = 0$, then $\tilde{\mu}_{(\sigma,0)}$ coincides with the σ-stable CRM, whereas if $\sigma \to 0$ the Gamma CRM is obtained. If α in (6.2.38) is a non-atomic finite measure on $(\mathbb{X}, \mathscr{X})$, we have $0 < \tilde{\mu}_{(\sigma,\tau)}(\mathbb{X}) < +\infty$ a.s. and we define the normalized generalized Gamma process with parameter (σ, τ) as the NRMI $\tilde{P}_{(\sigma,\tau)}(\cdot) := \tilde{\mu}_{(\sigma,\tau)}(\cdot)/\tilde{\mu}_{(\sigma,\tau)}(\mathbb{X})$. See Pitman [156] for a discussion on its representation as Poisson-Kingman model. The special case of $\sigma = 1/2$, corresponding to the normalized inverse Gaussian process, has been recently examined in Lijoi et al [119] who also provided an expression for the family of the finite dimensional distributions of $\tilde{P}_{(\sigma,\tau)}$.

As shown in Lijoi et al. [120] (see also Pitman [156] and James [96]), the EPPF induced by a sample $X_1, \ldots, X_n$ from a normalized generalized Gamma process with parameter (σ, τ) is given by (6.2.2) where $\beta = \tau^\sigma/\sigma$. Then, basing upon Proposition (6.2.1), one has

$$\mathbb{P}(K_m^{(n)} = k, L_m^{(n)} = s, S_{L_m^{(n)}} = \mathbf{s} | K_n = j) = \frac{\sigma^k \sum_{i=1}^{n+m-1} \binom{n+m-1}{i} (-1)^i \beta^{i/\sigma} \Gamma(j+k-i/\sigma; \beta)}{(n)_{m\uparrow 1} \sum_{i=1}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\sigma} \Gamma(j-i/\sigma; \beta)} \tag{6.2.39}$$

$$\times \binom{m}{s} (n-j\sigma)_{(m-s)\uparrow 1} \prod_{i=1}^{k} (1-\sigma)_{(s_i-1)\uparrow 1}$$

and it is possible to derive explicit expressions for all the sampling formulae set forth. In particular, the evaluation of the probability in P1) for the normalized generalized Gamma process with parameter (σ, τ), can be obtained by marginalizing the conditional distribution in (6.2.39) with respect to $L_m^{(n)}$ and $S_{L_m^{(n)}}$

$$\mathbb{P}(K_m^{(n)} = k | K_n = j) = \frac{\sum_{i=0}^{n+m-1} \binom{n+m-1}{i} (-1)^i \beta^{i/\sigma} \Gamma(k+j-i/\sigma; \beta)}{(n)_{m\uparrow 1} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\sigma} \Gamma(j-i/\sigma; \beta)} \mathscr{C}(m, k; \sigma, -n+j\sigma) \tag{6.2.40}$$

Based on (6.2.39), the estimators of interest can be derived. The most significant one is the expected number of new species $\hat{E}_m^{(n,j)}$ which represents a Bayesian nonparametric analog of the Good-Toulmin estimator (see Good and Toulmin [76])

$$\hat{E}_m^{(n,j)} = \sum_{k=0}^{m} \frac{\sum_{i=0}^{n+m-1} \binom{n+m-1}{i} (-1)^i \beta^{i/\sigma} \Gamma(k+j-i/\sigma; \beta)}{(n)_{m\uparrow 1} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\sigma} \Gamma(j-i/\sigma; \beta)} \mathscr{C}(m, k; \sigma, -n+j\sigma) \tag{6.2.41}$$

Moreover, evaluating the probability in P2) interpreted as the probability that the $(n+m+1)$-th observation will yield a new species, without observing the m intermediate records, is given by

$$\hat{D}_m^{(n,j)} = \sum_{k=0}^{m} \frac{\sum_{i=0}^{n+m} \binom{n+m}{i} (-1)^i \beta^{i/\sigma} \Gamma(k+j+1-i/\sigma; \beta)}{(n+1)_{m\uparrow 1} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\sigma} \Gamma(j-i/\sigma; \beta)} \mathscr{C}(m, k; \sigma, -n+j\sigma) \tag{6.2.42}$$

And in particular the estimated sample coverage after $n+m$ draws is given by $\hat{C}_m^{(n,j)} = 1 - \hat{D}_m^{(n,j)}$. We conclude the subsection devoted to the normalized generalized Gamma process by showing some results for the factorial moments of the distribution of K_n and the distribution of $K_m^{(n)}|K_n$. In particular, using similar arguments to those used for the two parameter Poisson-Dirichlet process and using the theory of Fox H-functions and Meijer G-functions (see Appendix C), we show that is still possibile to obtain expression for the factorial moments of the distribution of K_n and the distribution of $K_m^{(n)}|K_n = j$ for the normalized generalized Gamma process.

Proposition 6.2.12. *Under the normalized generalized Gamma process with parameter (σ, τ), the r-th descending factorial moment of K_n is*

$$\mathbb{E}[(K_n)_{r\downarrow 1}] = \frac{e^{\beta}}{\sigma\Gamma(n)}\sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j} \times \sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}H_{2,3}^{3,0}\left[\beta\,\middle|\,\begin{matrix}(1,1),(\sigma j-i,\sigma)\\(0,1),(r-i/\sigma,1),(\sigma j-i+n,\sigma)\end{matrix}\right]$$

Proof. By the definition of r-th descending factorial moment of K_n we have

$$\mathbb{E}[(K_n)_{r\downarrow 1}] = \sum_{k=1}^{n}(k)_{r\downarrow 1}\frac{e^{\beta}\mathscr{C}(n,k;\sigma)}{\sigma\Gamma(n)}\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\Gamma\left(k-\frac{i}{\sigma};\beta\right)$$

and using the integral representation of Mellin-Barnes-type for the incomplete Gamma function $\Gamma(x,y) = 1/2\pi\mathrm{i}\oint_{\mathfrak{L}}\Gamma(t+x)\Gamma(t)/\Gamma(t+1)z^{-t}dt$ we obtain

$$\frac{e^{\beta}}{\sigma\Gamma(n)}\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\sum_{k=1}^{n}(k)_{r\downarrow 1}\mathscr{C}(n,k;\sigma)\frac{1}{2\pi\mathrm{i}}\oint_{\mathfrak{L}}\frac{\Gamma(t+k-i/\sigma)\Gamma(t)\beta^{-t}}{\Gamma(t+1)}dt$$

We now define $\Phi(r,n+1) := \sum_{1\leq k\leq n+1}(k)_{r\downarrow 1}\mathscr{C}(n+1,k;\sigma)\Gamma\left(t+k-i/\sigma\right)$ and using the triangolar recurrence equation (A.3.14) for the non-central generalized factorial coefficient with $r = 0$, i.e.

$$C(n+1,k;s) = (sk-n)C(n,k;s) + sC(n,k-1;s)$$

we obtain

$$\begin{aligned}
&\Phi(r,n+1)\\
&\quad=\sum_{k=1}^{n+1}(k)_{r\downarrow 1}(-1)^{n+1-k}((\sigma k-n)C(n,k;\sigma)+\sigma C(n,k-1;\sigma))\Gamma\left(t+k-\frac{i}{\sigma}\right)\\
&\quad-\sum_{k=1}^{n}(k)_{r\downarrow 1}(\sigma k-n)\mathscr{C}(n,k;\sigma)\Gamma\left(t+k-\frac{i}{\sigma}\right)\\
&\quad+\sum_{k=1}^{n+1}(k)_{r\downarrow 1}\sigma\mathscr{G}(n,k-1;\sigma)\Gamma\left(t+k-\frac{i}{\sigma}\right)
\end{aligned}$$

$$
\begin{aligned}
&= -\sum_{k=1}^{n}(k)_{r\downarrow 1}(\sigma k - n)\mathscr{C}(n,k;\sigma)\Gamma\left(t+k-\frac{i}{\sigma}\right) \\
&\quad + \sum_{k=1}^{n}(k+1)_{r\downarrow 1}\sigma\mathscr{C}(n,k;\sigma)\left(t+k-\frac{i}{\sigma}\right)\Gamma\left(t+k-\frac{i}{\sigma}\right) \\
&= \sum_{k=1}^{n}\left(-(k)_{r\downarrow 1}(\sigma k - n) + \sigma(k+1)_{r\downarrow 1}\left(t+k-\frac{i}{\sigma}\right)\right)\mathscr{C}(n,k;\sigma)\Gamma\left(t+k-\frac{i}{\sigma}\right) \\
&= \sum_{k=1}^{n}((k)_{(r-1)\downarrow 1}(-i(1+k) + (-r+1+k)n + \sigma t + \sigma k(r+t))) \\
&\qquad \times \mathscr{C}(n,k;\sigma)\Gamma\left(t+k-\frac{i}{\sigma}\right) \\
&= \sum_{k=1}^{n}(k)_{(r)\downarrow 1}(-i+n+\sigma(r+t))\mathscr{C}(n,k;\sigma)\Gamma\left(t+k-\frac{i}{\sigma}\right) \\
&\quad + \sum_{k=1}^{n}(k)_{(r-1)\downarrow 1}(-ir+\sigma t+\sigma(r-1)(r+t))\mathscr{C}(n,k;\sigma)\Gamma\left(t+k-\frac{i}{\sigma}\right)
\end{aligned}
$$

In particular, we obtain the following recurrence relation

$$
\Phi(r,n+1) = (-i+n+\sigma(r+t))\Phi(r,n) + (-ir+\sigma t+\sigma(r-1)(r+t))\Phi(r-1,n) \tag{6.2.43}
$$

Then, using the recurrence relation (6.2.39) the result follows by simple induction on n and on r. In particular, we obtain

$$
\Phi(r,n+1) = \Gamma\left(r-\frac{i}{\sigma}+t\right)\sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j}(-i+\sigma(j+t))_{(n+1)\uparrow 1}
$$

Then

$$
\begin{aligned}
\mathbb{E}[(K_n)_{r\downarrow 1}] &= \frac{e^{\beta}}{\sigma\Gamma(n)}\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma} \\
&\qquad \times \frac{1}{2\pi \mathrm{i}}\oint_{\mathfrak{L}}\frac{\Gamma(t)\Gamma(r-i/\sigma+t)\beta^{-t}}{\Gamma(t+1)}\sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j}(-i+\sigma(j+t))_{n\uparrow 1}dt \\
&= \frac{e^{\beta}}{\sigma\Gamma(n)}\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j} \\
&\qquad \times \frac{1}{2\pi \mathrm{i}}\oint_{\mathfrak{L}}\frac{\Gamma(t)\Gamma(r-i/\sigma+t)\Gamma(-i+\sigma(j+t)+n)}{\Gamma(t+1)\Gamma(-i+\sigma(j+t))}\beta^{-t}dt
\end{aligned}
$$

and the result follows by the definition of Fox-H function. □

Corollary 6.2.4. *Under the normalized generalized Gamma process with parameter* (σ, τ), *the* r*-th moment of* K_n *is*

$$\mathbb{E}[(K_n)^r] = \frac{e^{\beta}}{\Gamma(n)} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\sigma}$$
$$\times \sum_{j=0}^{r} (-1)^{r-j} \frac{1}{2\pi \mathrm{i}} \oint_{\mathfrak{L}} \frac{\Gamma(t)\Gamma(1+j-i/\sigma+t)\Gamma(-i+n+\sigma(j+t))}{\Gamma(t+1)\Gamma(1-i+\sigma(j+t))} \beta^{-t} R\left(r, j, t-\frac{i}{\sigma}\right) dt$$

where $R(r, j, t - i/j)$ *is the unique function satisfying*

$$\sum_{j=0}^{r} (y)_{j\downarrow 1} R\left(r, j, t-\frac{i}{\sigma}\right) = \left(y+t-\frac{i}{\sigma}\right)^r$$

for any y.

Proof. Using the (r, l)-th Stirling number of the second kind $S(r, l)$ we can write

$$\mathbb{E}[(K_n)^r] = \sum_{l=0}^{r} S(r,l) \mathbb{E}[(K_m)_{l\downarrow 1}]$$

and by Proposition 6.2.12 we have

$$\mathbb{E}[(K_n)^r] = \frac{\mathrm{e}^{\beta}}{\sigma\Gamma(n)} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\sigma}$$
$$\times \sum_{l=0}^{r} S(r,l) \sum_{j=0}^{l} (-1)^{l-j} \binom{l}{j} \frac{1}{2\pi \mathrm{i}} \oint_{\mathfrak{L}} \frac{\Gamma(t)\Gamma(l-i/\sigma+t)\Gamma(-i+\sigma(j+t)+n)}{\Gamma(t+1)\Gamma(-i+\sigma(j+t))} \beta^{-t} dt$$

We focus on

$$\sum_{l=0}^{r} S(r,l)\Gamma\left(l-\frac{i}{\sigma}+t\right) \sum_{j=0}^{l} (-1)^{l-j} \binom{l}{j} (-i+\sigma(j+t))_{(n)\uparrow 1}$$
$$= \Gamma\left(t-\frac{i}{\sigma}\right) (\sigma t - i)_{n\uparrow 1} \sum_{l=0}^{r} S(r,l) \frac{(\sigma t - i)_{l\uparrow\sigma}}{\sigma^l (\sigma t - i)_{n\uparrow 1}} \sum_{j=0}^{l} (-1)^{l-j} \binom{l}{j} (-i+\sigma(j+t))_{n\uparrow 1}$$
$$= \Gamma\left(t-\frac{i}{\sigma}\right) (\sigma t - i)_{n\uparrow 1} \sum_{l=0}^{r} S(r,l) \left(t-\frac{i}{\sigma}\right)_{l\uparrow 1} \sum_{j=0}^{l} (-1)^{l-j} \binom{l}{j} \frac{(-i+\sigma(j+t))_{n\uparrow 1}}{(\sigma t - i)_{n\uparrow 1}}$$
$$= \Gamma\left(t-\frac{i}{\sigma}\right) (\sigma t - i)_{n\uparrow 1} \sum_{j=0}^{r} (-1)^{r-j} \left(1+t-\frac{i}{\sigma}\right)_{j\uparrow 1}$$
$$\times T\left(r, j, t-\frac{i}{\sigma}\right) \frac{(1-i+\sigma(j+t))_{(n-1)\uparrow 1}}{(1+\sigma t - i)_{(n-1)\uparrow 1}}$$

where $T\left(r,j,t-i/\sigma\right)=\sum_{j\leq l\leq r}\left(t-i/\sigma+j\right)_{(l-j)\uparrow 1}S(r,l)(-1)^{r+l}\binom{l}{j}$. For any y

$$\begin{aligned}\sum_{j=0}^{r}(y)_{j\downarrow 1}T\left(r,j,t-\frac{i}{\sigma}\right) &= \sum_{l=0}^{r}\left(\sum_{j=0}^{l}\binom{l}{j}(y)_{j\downarrow 1}\left(t-\frac{i}{\sigma}+l-1\right)_{(l-j)\downarrow 1}\right)S(r,l)(-1)^{r+l}\\ &= \sum_{l=0}^{r}\left(y+t-\frac{i}{\sigma}+l-1\right)_{l\downarrow 1}(-1)^{l}S(r,l)(-1)^{r}\\ &= \sum_{l=0}^{r}\left(-y-t+\frac{i}{\sigma}\right)_{l\downarrow 1}S(r,l)(-1)^{r}\\ &= \left(-y-t+\frac{i}{\sigma}\right)^{r}(-1)^{r}=\left(y+t-\frac{i}{\sigma}\right)^{r}\end{aligned}$$

Then, we have

$$\sum_{j=0}^{r}(y)_{j\downarrow 1}T\left(r,j,t-\frac{i}{\sigma}\right)=\left(y+t-\frac{i}{\sigma}\right)^{r}$$

and $T(r,j,t-i/\sigma)$ corresponds to the function $R(r,j,\lambda)$ with $\lambda=t-i/\sigma$, introduced by Carlitz [14] (or non-central Stirling number of the second kind). The result follows by substitution. □

Proposition 6.2.13. *Under the normalized generalized Gamma process with parameter (σ,τ), the r-th ascending factorial moment of K_n is*

$$\begin{aligned}\mathbb{E}[(K_n)_{r\uparrow 1}] = \frac{e^{\beta}}{\sigma\Gamma(n)}\sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j}\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^{i}\beta^{i/\sigma}\\ \times H_{3,5}^{3,1}\left[\beta\,\middle|\,\begin{matrix}(1,1),(\sigma j-i,\sigma),(1-i/\sigma,1)\\(0,1),(j-i/\sigma,1),(\sigma j-i+n,\sigma),(1+i/\sigma-r+j,1)\end{matrix}\right]\end{aligned}$$

Proof. Using the (r,l)-th signless Stirling number of the first kind $|s(r,l)|$ we can write

$$\mathbb{E}[(K_n)_{r\uparrow 1}]=\sum_{l=0}^{r}|s(r,l)|\mathbb{E}[(K_n)^{l}]$$

and by Corollary 6.2.4 we have

$$\begin{aligned}
\mathbb{E}[(K_n)_{r\uparrow 1}] &= \frac{\mathrm{e}^{\beta}}{\sigma\Gamma(n)}\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\sum_{l=0}^{r}|s(r,l)|\sum_{j=0}^{l}(-1)^{l-j}\\
&\quad\times\frac{1}{2\pi\mathrm{i}}\oint_{\mathfrak{L}}\frac{\Gamma(t)\Gamma(1+j-i/\sigma+t)\Gamma(-i+n+\sigma(j+t))}{\Gamma(t+1)\Gamma(1-i+\sigma(j+t))}\beta^{-t}R\left(l,j,t-\frac{i}{\sigma}\right)dt\\
&= \frac{\mathrm{e}^{\beta}}{\sigma\Gamma(n)}\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\\
&\quad\times\frac{1}{2\pi\mathrm{i}}\oint_{\mathfrak{L}}\frac{\Gamma(t)\beta^{-t}}{\Gamma(t+1)}\sum_{l=0}^{r}|s(r,l)|\sum_{j=0}^{l}(-1)^{l-j}\Gamma\left(1+j-\frac{i}{\sigma}+t\right)\\
&\quad\times(1-i+\sigma(j+t))_{(n-1)\uparrow 1}R\left(l,j,t-\frac{i}{\sigma}\right)dt\\
&= \sum_{l=0}^{r}|s(r,l)|\sum_{j=0}^{l}(-1)^{l-j}\Gamma\left(1+j-\frac{i}{\sigma}+t\right)\\
&\quad\times(1-i+\sigma(j+t))_{(n-1)\uparrow 1}R\left(l,j,t-\frac{i}{\sigma}\right)\\
&= \Gamma\left(1-\frac{i}{\sigma}+t\right)(\sigma t-i+1)_{(n-1)\uparrow 1}\sum_{l=0}^{r}|s(r,l)|\sum_{j=0}^{l}(-1)^{l-j}\left(1-\frac{i}{\sigma}+t\right)_{j\uparrow 1}\\
&\quad\times\frac{(1-i+\sigma(j+t))_{(n-1)\uparrow 1}}{(1-i+\sigma t)_{(n-1)\uparrow 1}}R\left(l,j,t-\frac{i}{\sigma}\right)\\
&= \Gamma\left(1-\frac{i}{\sigma}+t\right)(\sigma t-i+1)_{(n-1)\uparrow 1}\\
&\quad\times\sum_{j=0}^{r}(-1)^{r-j}\left(t-\frac{i}{\sigma}\right)_{j\uparrow 1}I\left(r,j,t-\frac{i}{\sigma}\right)\frac{(\sigma t-i+j\sigma)_{n\uparrow 1}}{(\sigma t-i)_{n\uparrow 1}}
\end{aligned}$$

where $I\left(r,j,t-i/\sigma 7\right)=\sum_{j\leq l\leq r}s(r,l)R\left(l,j,t-\frac{i}{\sigma}\right)$. For any μ

$$\begin{aligned}
\sum_{j=0}^{r}I\left(r,j,t-\frac{i}{\sigma}\right)(\mu)_{j\downarrow 1} &= \sum_{l=0}^{r}\left(\sum_{j=0}^{l}R\left(l,j,t-\frac{i}{\sigma}\right)(\mu)_{j\downarrow 1}\right)s(r,l)\\
&= \sum_{l=0}^{r}\left(t-\frac{i}{\sigma}+\mu\right)^{l}s(r,l)=\left(t-\frac{i}{\sigma}+\mu\right)_{r\downarrow 1}
\end{aligned}$$

Then, for any μ

$$\sum_{j=0}^{r}I\left(r,j,t-\frac{i}{\sigma}\right)(\mu)_{j\downarrow 1}=\left(t-\frac{i}{\sigma}+\mu\right)_{r\downarrow 1}$$

Therefore, if $t-i/\sigma\neq 0$, then

$$I(r,j,t-i/\sigma)=\binom{r}{j}(t-i/\sigma)_{(r-j)\downarrow 1}$$

If $t - i/\sigma = 0$, $I(r, j, t - i/\sigma) = 1$ if $j = r$ and $I(r, j, t - i/\sigma) = 0$ se $j \neq r$. Then, we have

$$\begin{aligned}\mathbb{E}[(K_n)_{r\uparrow 1}] &= \frac{e^{\beta}}{\sigma\Gamma(n)} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\sigma} \sum_{j=0}^{r} (-1)^{r-j} \binom{r}{j} \\ &\quad \times \frac{1}{2\pi \mathrm{i}} \oint_{\mathfrak{L}} \frac{\Gamma(t)\Gamma(t - i/\sigma + j)(\sigma t - i + j\sigma)_{n\uparrow 1}(t - i/\sigma)_{(r-j)\downarrow 1}}{\Gamma(t+1)} \beta^{-t} dt \\ &= \frac{e^{\beta}}{\sigma\Gamma(n)} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\sigma} \sum_{j=0}^{r} \binom{r}{j} \\ &\quad \times \frac{1}{2\pi \mathrm{i}} \oint_{\mathfrak{L}} \frac{\Gamma(t)\Gamma(t - i/\sigma + j)(\sigma t - i + j\sigma)_{n\uparrow 1}(-t + i/\sigma)_{(r-j)\uparrow 1}}{\Gamma(t+1)} \beta^{-t} dt\end{aligned}$$

and the result follows by the definiton of Fox-H function. □

Proposition 6.2.14. *Under the normalized generalized Gamma process with parameter (σ, τ), the r-th descending factorial moment of $K_m^{(n)} | K_n = j$ is*

$$\begin{aligned}\mathbb{E}[(K_m^{(n)})_{r\downarrow 1} | K_n = j] &= \left(\Gamma(n+m) \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\sigma} \Gamma\left(j - \frac{i}{\sigma}; \beta\right) \right)^{-1} \\ &\quad \times \Gamma(n) \sum_{i=0}^{n+m-1} \binom{n+m-1}{i} (-1)^i \beta^{i/\sigma} \sum_{s=0}^{r} (-1)^{r-s} \binom{r}{s} \\ &\quad \times H_{2,3}^{3,0} \left[\beta \middle| \begin{matrix} (1,1), (\sigma j - i + n, \sigma) \\ (0,1), (r + j - i/\sigma, 1), (\sigma j - i + n + m, \sigma) \end{matrix} \right]\end{aligned}$$

Proof. By the definition of r-th descending factorial moment of $K_m^{(n)} | K_n = j$ we have

$$\begin{aligned}&\mathbb{E}[(K_m^{(n)})_{r\downarrow 1} | K_n = j] \\ &\quad = \sum_{k=0}^{m} (k)_{r\downarrow 1} \frac{\Gamma(n) \sum_{i=0}^{n+m-1} \binom{n+m-1}{i} (-1)^i \beta^{i/\sigma} \Gamma(k + j - i/\sigma; \beta)}{\Gamma(n+m) \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\sigma} \Gamma(j - i/\sigma; \beta)} \mathscr{C}(m, k; \sigma, -n + j\sigma)\end{aligned}$$

and using the integral representation of Mellin-Barnes-type for the incomplete Gamma function we obtain

$$\begin{aligned}&\left(\Gamma(n+m) \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{i/\sigma} \Gamma(j - \frac{i}{\sigma}; \beta) \right)^{-1} \Gamma(n) \sum_{i=0}^{n+m-1} \binom{n+m-1}{i} (-1)^i \beta^{i/\sigma} \\ &\quad \times \sum_{k=0}^{m} (k)_{r\downarrow 1} \mathscr{C}(m, k; \sigma, -n + j\sigma) \frac{1}{2\pi \mathrm{i}} \oint_{\mathfrak{L}} \frac{\Gamma(t + k + j - i/\sigma)\Gamma(t)\beta^{-t}}{\Gamma(t+1)} dt\end{aligned}$$

We focus on $\tilde{\Phi}(r, m+1) := \sum_{0 \le k \le m+1} (k)_{r\downarrow 1} \mathscr{C}(m+1, k; \sigma, -n + j\sigma) \Gamma\left(t + k + j - i/\sigma\right)$ and using the triangolar recurrence equation for the non-central generalized factorial coefficient we obtain

$$
\begin{aligned}
\tilde{\Phi}(r,m+1) &= -\sum_{k=0}^{m}(k)_{r\downarrow 1}(\sigma k - n + j\sigma - m)\mathscr{C}(m,k;\sigma,-n+j\sigma)\Gamma\left(t+k+j-\frac{i}{\sigma}\right) \\
&\quad + \sum_{k=0}^{m+1}(k)_{r\downarrow 1}\sigma\mathscr{C}(n,k-1;\sigma,-n+j\sigma)\Gamma\left(t+k+j-\frac{i}{\sigma}\right) \\
&= -\sum_{k=0}^{m}(k)_{r\downarrow 1}(\sigma k - n + j\sigma - m)\mathscr{C}(m,k;\sigma,-n+j\sigma)\Gamma\left(t+k+j-\frac{i}{\sigma}\right) \\
&\quad + \sum_{k=0}^{m}(k+1)_{r\downarrow 1}\sigma\mathscr{C}(m,k;\sigma,-n+j\sigma)\left(t+k+j-\frac{i}{\sigma}\right)\Gamma\left(t+k+j-\frac{i}{\sigma}\right) \\
&= \sum_{k=0}^{m}(-(k)_{r\downarrow 1}(\sigma k - n + j\sigma - m) + \sigma(k+1)_{r\downarrow 1}\left(t+k+j-\frac{i}{\sigma}\right)) \\
&\quad \times \mathscr{C}(m,k;\sigma,-n+j\sigma)\Gamma\left(t+k+j-\frac{i}{\sigma}\right) \\
&= \sum_{k=0}^{m}((k)_{(r)\downarrow 1}(-i+m+n+\sigma(r+t))\mathscr{C}(m,k;\sigma,-n+j\sigma)\Gamma\left(t+k+j-\frac{i}{\sigma}\right) \\
&\quad + \sum_{k=0}^{m}((k)_{(r-1)\downarrow 1}r(-i+\sigma(-1+j+r+t)) \\
&\quad\quad \times \mathscr{C}(m,k;\sigma,-n+j\sigma)\Gamma\left(t+k+j-\frac{i}{\sigma}\right)
\end{aligned}
$$

In particular, we obtain the following recurrence relation

$$
\tilde{\Phi}(r,m+1) = (-i+m+n+\sigma(r+t))\tilde{\Phi}(r,m)+r(-i+\sigma(-1+j+r+t))\tilde{\Phi}(r-1,m) \quad (6.2.44)
$$

Then, using recurrence relation (6.2.44) the result follows by induction on m and on r. In particular, we have

$$
\tilde{\Phi}(r,m+1) = \Gamma\left(r+j-\frac{i}{\sigma}+t\right)\sum_{s=0}^{r}(-1)^{r-s}\binom{r}{s}(-i+n+\sigma(s+t))_{(m+1)\uparrow 1}
$$

Then

$$\mathbb{E}[(K_m^{(n)})_{r\downarrow 1}|K_n=j]=\left(\Gamma(n+m)\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\Gamma\left(j-\frac{i}{\sigma};\beta\right)\right)^{-1}$$

$$\times\Gamma(n)\sum_{i=0}^{n+m-1}\binom{n+m-1}{i}(-1)^i\beta^{i/\sigma}$$

$$\times\sum_{s=0}^{r}(-1)^{r-s}\binom{r}{s}\frac{1}{2\pi\mathrm{i}}\oint_{\mathfrak{L}}\frac{\Gamma\left(r+j-i/\sigma+t\right)\Gamma(-i+n+\sigma(s+t)+m)\Gamma(t)\beta^{-t}}{\Gamma(t+1)\Gamma(-i+n+\sigma(s+t))}dt$$

$$=\left(\Gamma(n+m)\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\Gamma\left(j-\frac{i}{\sigma};\beta\right)\right)^{-1}$$

$$\times\Gamma(n)\sum_{i=0}^{n+m-1}\binom{n+m-1}{i}(-1)^i\beta^{i/\sigma}\sum_{s=0}^{r}(-1)^{r-s}\binom{r}{s}$$

$$\times\frac{1}{2\pi\mathrm{i}}\oint_{\mathfrak{L}}\frac{\Gamma\left(r+j-i/\sigma+t\right)\Gamma(-i+n+\sigma(s+t)+m)\Gamma(t)\beta^{-t}}{\Gamma(t+1)\Gamma(-i+n+\sigma(s+t))}dt$$

and the result follows by the definition of Fox-H function □

Corollary 6.2.5. *Under the normalized generalized Gamma process with parameter* (σ,τ), *the* r*-th moment of* $K_m^{(n)}|K_n=j$ *is*

$$\mathbb{E}[(K_m^{(n)})^r|K_n=j]=\left(\Gamma(n+m)\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\Gamma\left(j-\frac{i}{\sigma};\beta\right)\right)^{-1}$$

$$\times\Gamma(n)\sum_{i=0}^{n+m-1}\binom{n+m-1}{i}(-1)^i\beta^{i/\sigma}\frac{\sigma t-i+n}{j+t-i/\sigma}$$

$$\times\sum_{s=0}^{r}(-1)^{r-s}\frac{1}{2\pi\mathrm{i}}\oint_{\mathfrak{L}}\frac{\Gamma(-i+n+m+\sigma(s+t))\Gamma(1+j+t-i/\sigma+s)\Gamma(t)}{\Gamma(t+1)\Gamma(1-i+n+\sigma(s+t))}$$

$$\times\beta^{-t}R\left(r,s,j+t-\frac{i}{\sigma}\right)dt$$

where $R(r,s,j+t-i/\sigma)$ *is the unique function satisfying*

$$\sum_{s=0}^{r}(y)_{s\downarrow 1}R\left(r,s,j+t-\frac{i}{\sigma}\right)=\left(y+j+t-\frac{i}{\sigma}\right)^r$$

for any y.

Proof. Using the (r,l)-th Stirling number of the second kind $S(r,l)$ we can write

$$\mathbb{E}[(K_m^{(n)})^r|K_n=j]=\sum_{l=0}^{r}S(r,l)\mathbb{E}[(K_m^{(n)})_{l\downarrow 1}|K_n=j]$$

and by Proposition 6.2.14 we have

$$\begin{aligned}\mathbb{E}[(K_m^{(n)})^r|K_n=j] &= \left(\Gamma(n+m)\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\Gamma\left(j-\frac{i}{\sigma};\beta\right)\right)^{-1}\\ &\times\Gamma(n)\sum_{l=0}^{r}S(r,l)\sum_{i=0}^{n+m-1}\binom{n+m-1}{i}(-1)^i\beta^{i/\sigma}\sum_{s=0}^{l}(-1)^{l-s}\binom{l}{s}\\ &\times\frac{1}{2\pi\mathrm{i}}\oint_{\mathfrak{L}}\frac{\Gamma\left(l+j-i/\sigma+t\right)\Gamma(-i+n+\sigma(s+t)+m)\Gamma(t)\beta^{-t}}{\Gamma(t+1)\Gamma(-i+n+\sigma(s+t))}dt\end{aligned}$$

We focus on

$$\begin{aligned}&\sum_{l=0}^{r}S(r,l)\Gamma\left(l+j-\frac{i}{\sigma}+t\right)\sum_{s=0}^{l}(-1)^{l-s}\binom{l}{s}(-i+n+\sigma(s+t))_{m\uparrow1}\\ &=\Gamma\left(j+t-\frac{i}{\sigma}\right)(\sigma t-i+n)_{m\uparrow1}\sum_{l=0}^{r}S(r,l)\frac{(\sigma(t+j)-i)_{l\uparrow\sigma}}{\sigma^l(\sigma t-i+n)_{m\uparrow1}}\\ &\qquad\times\sum_{s=0}^{l}(-1)^{l-s}\binom{l}{s}(-i+n+\sigma(s+t))_{m\uparrow1}\\ &=\Gamma\left(j+t-\frac{i}{\sigma}\right)(\sigma t-i+n)_{m\uparrow1}\sum_{l=0}^{r}S(r,l)\left(j+t-\frac{i}{\sigma}\right)_{l\uparrow1}\\ &\qquad\times\sum_{s=0}^{l}(-1)^{l-s}\binom{l}{s}\frac{(-i+n+\sigma(s+t))_{m\uparrow1}}{(\sigma t-i+n)_{m\uparrow1}}\\ &=\Gamma\left(j+t-\frac{i}{\sigma}\right)(\sigma t-i+n)_{m\uparrow1}\sum_{s=0}^{r}(-1)^{r-s}\left(1+j+t-\frac{i}{\sigma}\right)_{s\uparrow1}\\ &\qquad\times T\left(r,s,j+t-\frac{i}{\sigma}\right)\frac{(1-i+n+\sigma(s+t))_{(m-1)\uparrow1}}{(1+\sigma t-i+n)_{(m-1)\uparrow1}}\end{aligned}$$

where $T\left(r,s,j+t-i/\sigma\right)=\sum_{s\le l\le r}(j+t-i/\sigma+s)_{(l-s)\uparrow1}S(r,l)(-1)^{r+l}\binom{l}{s}$. For any y

$$\begin{aligned}\sum_{s=0}^{r}(y)_{s\downarrow1}T\left(r,s,j+t-\frac{i}{\sigma}\right) &= \sum_{l=0}^{r}\left(\sum_{s=0}^{l}\binom{l}{s}(y)_{s\downarrow1}\left(j+t-\frac{i}{\sigma}+l-1\right)_{(l-s)\downarrow1}\right)S(r,l)(-1)^{r+l}\\ &=\sum_{l=0}^{r}\left(y+j+t-\frac{i}{\sigma}+l-1\right)_{l\downarrow1}(-1)^lS(r,l)(-1)^r\\ &=\sum_{l=0}^{r}\left(-y-j-t+\frac{i}{\sigma}\right)_{l\downarrow1}S(r,l)(-1)^r\\ &=\left(-y-j-t+\frac{i}{\sigma}\right)^r(-1)^r=\left(y+j+t-\frac{i}{\sigma}\right)^r\end{aligned}$$

Then, we have

$$\sum_{s=0}^{r}(y)_{s\downarrow 1}T\left(r,s,t-\frac{i}{\sigma}\right)=\left(y+t-\frac{i}{\sigma}\right)^{r}$$

and $T(r,s,j+t-i/\sigma)$ corresponds to the function $R(r,s,\lambda)$ with $\lambda=j+t-i/\sigma$, introduced by Carlitz [14] (or non-central Stirling number of the second kind). The result follows by substitution. □

Proposition 6.2.15. *Under the normalized generalized Gamma process with parameter (σ,τ), the r-th ascending factorial moment of $K_m^{(n)}|K_n=j$ is*

$$\begin{aligned}\mathbb{E}[(K_m^{(n)})_{r\uparrow 1}|K_n=j]&=\left(\Gamma(n+m)\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\Gamma\left(j-\frac{i}{\sigma};\beta\right)\right)^{-1}\\&\times\Gamma(n)\sum_{i=0}^{n+m-1}\binom{n+m-1}{i}(-1)^i\beta^{i/\sigma}\sum_{s=0}^{r}\binom{r}{s}\\&\times H_{3,5}^{3,1}\left[\beta\,\middle|\,\begin{matrix}(1,1),(\sigma s-i+n,\sigma),(1+j-i/\sigma,1)\\(0,1),(j+s-i/\sigma,1),(\sigma s-i+n+m,\sigma),(1+j+i/\sigma-r+s,1)\end{matrix}\right]\end{aligned}$$

Proof. Using the (r,l)-th signless Stirling number of the first kind $|s(r,l)|$ we can write

$$\mathbb{E}[(K_m^{(n)})_{r\uparrow 1}|K_n=j]=\sum_{l=0}^{r}|s(r,l)|\mathbb{E}[(K_m^{(n)})^l|K_n=j]$$

and by Corollary 6.2.5 we have

$$\begin{aligned}
\mathbb{E}[(K_m^{(n)})_{r\uparrow 1}|K_n = j] &= \left(\Gamma(n+m)\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\Gamma\left(j-\frac{i}{\sigma};\beta\right)\right)^{-1} \\
&\times \Gamma(n)\sum_{i=0}^{n+m-1}\binom{n+m-1}{i}(-1)^i\beta^{i/\sigma}\frac{\sigma t - i + n}{j+t-i/\sigma}\sum_{l=0}^{r}|s(r,l)| \\
&\times \sum_{s=0}^{l}(-1)^{l-s}\frac{1}{2\pi\mathrm{i}}\oint_{\mathfrak{L}}\frac{\Gamma(-i+n+m+\sigma(s+t))\Gamma(1+j+t-i/\sigma+s)\Gamma(t)}{\Gamma(t+1)\Gamma(1-i+n+\sigma(s+t))} \\
&\times \beta^{-t}R\left(l,s,j+t-\frac{i}{\sigma}\right)dt \\
&= \left(\Gamma(n+m)\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\Gamma\left(j-\frac{i}{\sigma};\beta\right)\right)^{-1} \\
&\times \Gamma(n)\sum_{i=0}^{n+m-1}\binom{n+m-1}{i}(-1)^i\beta^{i/\sigma}\frac{\sigma t - i + n}{j+t-i/\sigma}\sum_{l=0}^{r}|s(r,l)|\frac{1}{2\pi\mathrm{i}}\oint_{\mathfrak{L}}\frac{\Gamma(t)\beta^{-t}}{\Gamma(t+1)} \\
&\times \sum_{l=0}^{r}|s(r,l)|\sum_{s=0}^{l}(-1)^{l-s}\Gamma\left(1+j+t-\frac{i}{\sigma}+s\right) \\
&\times (1-i+n+\sigma(s+t))_{(m-1)\uparrow 1}R\left(l,s,j+t-\frac{i}{\sigma}\right)dt
\end{aligned}$$

and we focus on

$$\begin{aligned}
&\sum_{l=0}^{r}|s(r,l)|\sum_{s=0}^{l}(-1)^{l-s}\Gamma\left(1+j+t-\frac{i}{\sigma}+s\right)(1-i+n+\sigma(s+t))_{(m-1)\uparrow 1} \\
&\quad\times R\left(l,s,j+t-\frac{i}{\sigma}\right) \\
&= \Gamma\left(1+j+t-\frac{i}{\sigma}\right)(1-i+\sigma t+n)_{(m-1)\uparrow 1} \\
&\quad\times \sum_{l=0}^{r}|s(r,l)|\sum_{s=0}^{l}(-1)^{l-s}\left(1+j+t-\frac{i}{\sigma}\right)_{s\uparrow 1}\frac{(1-i+n+\sigma(s+t))_{(m-1)\uparrow 1}}{(1-i+\sigma(j+t))_{(m-1)\uparrow 1}} \\
&\quad\times R\left(l,s,j+t-\frac{i}{\sigma}\right) \\
&= \Gamma\left(1+j+t-\frac{i}{\sigma}\right)(1-i+\sigma t+n)_{(m-1)\uparrow 1} \\
&\quad\times \sum_{s=0}^{r}(-1)^{r-s}\left(j+t-\frac{i}{\sigma}\right)_{s\uparrow 1}I\left(r,s,j+t-\frac{i}{\sigma}\right)\frac{(-i+\sigma(t+s)+n)_{m\uparrow 1}}{(-i+\sigma t+n)_{m\uparrow 1}}
\end{aligned}$$

where $I\left(r,s,j+t-i/\sigma\right)=\sum_{s\leq l\leq r}s(r,l)R\left(l,s,j+t-i/\sigma\right)$. For any μ

$$\sum_{s=0}^{r} I\left(r,s,j+t-\frac{i}{\sigma}\right)(\mu)_{s\downarrow 1}=\sum_{l=0}^{r}\left(\sum_{s=0}^{l} R\left(l,s,j+t-\frac{i}{\sigma}\right)(\mu)_{s\downarrow 1}\right)s(r,l)$$
$$=\sum_{l=0}^{r}\left(j+t-\frac{i}{\sigma}+\mu\right)^{l} s(r,l)=\left(j+t-\frac{i}{\sigma}+\mu\right)_{r\downarrow 1}$$

Then, for any μ

$$\sum_{s=0}^{r} I\left(r,s,j+t-\frac{i}{\sigma}\right)(\mu)_{s\downarrow 1}=\left(j+t-\frac{i}{\sigma}+\mu\right)_{r\downarrow 1}$$

Therefore, if $j+t-i/\sigma\neq 0$, then

$$I(r,s,j+t-i/\sigma)=\binom{r}{s}(j+t-i/\sigma)_{(r-s)\downarrow 1}$$

If $j+t-i/\sigma=0$, $I(r,s,j+t-i/\sigma)=1$ if $s=r$ and $I(r,s,j+t-i/\sigma)=0$ se $s\neq r$. Then, we have

$$\mathbb{E}[(K_m^{(n)})_{r\uparrow 1}|K_n=j]=\left(\Gamma(n+m)\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\Gamma\left(j-\frac{i}{\sigma};\beta\right)\right)^{-1}$$
$$\times\Gamma(n)\sum_{i=0}^{n+m-1}\binom{n+m-1}{i}(-1)^i\beta^{i/\sigma}\sum_{s=0}^{r}(-1)^{r-s}\binom{r}{s}$$
$$\times\frac{1}{2\pi\mathrm{i}}\oint_{\mathfrak{L}}\frac{\Gamma(t)\Gamma(j+t+s-i/\sigma)(-i+\sigma(t+s)+n)_{m\uparrow 1}(j+t-i/\sigma)_{(r-s)\downarrow 1}\beta^{-t}}{\Gamma(t+1)}dt$$
$$=\left(\Gamma(n+m)\sum_{i=0}^{n-1}\binom{n-1}{i}(-1)^i\beta^{i/\sigma}\Gamma\left(j-\frac{i}{\sigma};\beta\right)\right)^{-1}$$
$$\times\Gamma(n)\sum_{i=0}^{n+m-1}\binom{n+m-1}{i}(-1)^i\beta^{i/\sigma}\sum_{s=0}^{r}\binom{r}{s}$$
$$\times\frac{1}{2\pi\mathrm{i}}\oint_{\mathfrak{L}}\frac{\Gamma(t)\Gamma(j+t+s-i/\sigma)(-i+\sigma(t+s)+n)_{m\uparrow 1}(-j-t+i/\sigma)_{(r-s)\downarrow 1}\beta^{-t}}{\Gamma(t+1)}dt$$

and the result follows by the definiton of Fox-H function. □

6.2.3 Applications to genomics

We now show how to use the results obtained in Subsection 6.2.1 for the two parametero Poisson-Dirichlet processs by applying them to 5 real EST datasets. As briefly mentioned in the Introduction, EST data arise by sequencing cDNA libraries consisting of millions of genes and one of the main quantities of interest is the number of distinct genes. Typically, due to cost constraints, only a small portion of the cDNA has been sequenced and, given this "basic sample", estimation of the number of new genes $K_m^{(n)}$ to appear in a hypothetical additional sample is required. Based on such estimates, geneticists have to decide whether it is worth to proceed with sequencing and, in the affirmative case, also the size of the additional sample. Here, we consider: (a) a tomato-flower cDNA library (Quackenbush et al. [159]), previously analyzed in Mao and Lindsay [134], Mao [132] and Lijoi et al. [123]; two cDNA libraries of the amitochondriate protist Mastigamoeba balamuthi (see Susko and Roger [177]): (b) the first is non-normalized, whereas (c) the second is normalized, i.e. it undergoes a normalization protocol which aims at making the frequencies of genes in the library more uniform so to increase the discovery rate; two Naegleria gruberi cDNA libraries prepared from cells grown under different culture conditions, (d) aerobic and (e) anaerobic (see Susko and Roger [177]).

In order to implement the Poisson-Dirichlet process with parameter (σ, θ), the first issue to face is represented by the specification of its parameters. The first possibility is to adopt an empirical Bayes approach. Since the "basic sample" consists of n observations featuring K_n distinct species with corresponding frequencies $(N_1, \dots, N_{K_n})$, the joint distribution of K_n and $(N_1, \dots, N_{K_n})$ is given by (6.2.1). See Quintana [160] for an investigation of the connection between random partition models and Bayesian nonparametrics. The empirical Bayes rule then suggests to fix (σ, θ) so to maximize (6.2.1) corresponding to the observed sample $(k, n_1, \dots, n_k)$, i.e.

$$(\hat{\sigma}, \hat{\theta}) = \arg\max_{(\sigma,\theta)} \frac{\prod_{i=1}^{k-1}(\theta + i\sigma)}{(\theta+1)_{(n-1)\uparrow 1}} \prod_{j=1}^{k} (1-\sigma)_{(n_j-1)\uparrow 1}. \tag{6.2.45}$$

An alternative way of eliciting (σ, θ) is by placing a prior distributions on it. Such an approach is useful when one is interested in testing the compatibility of clustering structures among different populations (Lijoi, Mena and Prünster [122]). However, in terms of estimates there are typically no relevant differences given the posterior distribution of (σ, θ) is highly concentrated. Hence, in order to keep the exposition as simple as possible, in the sequel we focus on the Poisson-Dirichlet process with parameter (σ, θ) with empirical Bayes prior specification. The extension to the case of priors on (σ, θ) is straightforward.

The computation of the estimators for the number of new genes (6.2.26), for the discovery probability (6.2.27) and for the sample coverage (6.2.28) is immediate. For each of the 5 EST datasets, the corresponding estimates for additional samples of size $m \in \{n, 10n, 100n\}$ are reported in Table 6.1 below together with the corresponding values n and j of the "basic sample" and the empirical Bayes specifications of (σ, θ).

The use of Proposition 6.2.7 is slightly more delicate. Here, we show it only for the estimator of the number of new genes; for the estimators of the discovery probability and

Library	n	j	$\hat{\sigma}$	$\hat{\theta}$	m	$\hat{E}_m^{(n,j)}$	$\hat{D}_m^{(n,j)}$	$\hat{C}_m^{(n,j)}$
					n	1281	0.447	0.553
(a)	2586	1825	0.612	741.0	$10n$	8432	0.240	0.760
					$100n$	40890	0.103	0.897
					n	346	0.452	0.548
(b)	715	460	0.770	46.0	$10n$	2634	0.307	0.693
					$100n$	16799	0.185	0.815
					n	180	0.456	0.544
(c)	363	248	0.700	57.0	$10n$	1280	0.278	0.722
					$100n$	7205	0.144	0.856
					n	307	0.290	0.710
(d)	959	473	0.670	46.3	$10n$	2085	0.166	0.834
					$100n$	11031	0.080	0.920
					n	440	0.412	0.588
(e)	969	631	0.660	155.5	$10n$	2994	0.236	0.764
					$100n$	15673	0.111	0.889

Table 6.1: Analysis of the 5 EST datasets. Size of the "basic sample" n, number of distinct genes j in the "basic sample" and empirical Bayes specifications for (σ, θ). Exact estimators for the number of new genes rounded to the nearest integer, for the discovery probability $\hat{D}_m^{(n,j)}$ and the coverage $\hat{C}_m^{(n,j)}$ for sizes of the additional sample $m \in \{n, 2n, 3n\}$.

the coverage one can proceed along the same lines. In order to combine the point estimate for $K_m^{(n)}$ with an asymptotic 95% HPD interval, one can simulate from the limiting r.v. $Z_{n,j}$ and determine the 95% HPD interval, (z_1, z_2), for $Z_{n,j}$. Then, given that the normalizing rate function for $K_m^{(n)}$ in Proposition 6.2.7, is m^σ, one obtains an asymptotic 95% HPD interval for $K_m^{(n)}$ as $(z_1\, m^\sigma, z_2\, m^\sigma)$. Table 6.2 below reports both exact and simulated mean and variance of the limiting r.v. $Z_{n,j}$ associated to each of the 5 EST datasets as well as the simulated 95% and 99% HPD intervals. The sampled values are obtained by generating 2000 random variates according to the algorithm devised in Subsection 6.2.1 with truncation of the series in (6.2.33) given by $N = 3 \times 10^7$. In fact, it is very important to get accurate samples from $Z_{n,j}$: a small bias could heavily affect the asymptotic HPD intervals for $K_m^{(n)}$, $(z_1 m^\sigma, z_2\, m^\sigma)$, since a large m^σ would amplify the bias. It should be emphasized that it is sufficient to run the simulation of $Z_{n,j}$ only once in order to obtain the HPD intervals for any choice of the additional sample size m. Hence, it seems definitely worth pursuing a high precision, which can be easily verified by comparing exact moments in (6.2.31) with the sampled ones.

Library	$\mathbb{E}[Z_{n,j}]$	$\mathrm{Var}(Z_{n,j})$	$\bar{Z}_{n,j}$	S^2	95% HPD	99% HPD
(a)	21.222	0.098	21.251	0.096	(20.62 , 21.83)	(20.46 , 22.02)
(b)	3.142	0.011	3.176	0.012	(2.95 , 3.37)	(2.89 , 3.44)
(c)	4.804	0.043	4.823	0.044	(4.43 , 5.24)	(4.28 , 5.36)
(d)	5.279	0.039	5.304	0.039	(4.93 , 5.69)	(4.78 , 5.82)
(e)	8.400	0.054	8.419	0.054	(7.97 , 8.88)	(7.80 , 8.98)

Table 6.2: Characteristics of the limiting r.v. $Z_{n,j}$ for the 5 cDNA libraries: exact mean $\mathrm{E}[Z_{n,j}]$, exact variance $\mathrm{Var}[Z_{n,j}]$, sample mean $\bar{Z}_{n,j}$, sample variance S^2, sample 95% and 99% HPD intervals.

Library	m	$\hat{E}_m^{(n,j)}$	rate m^σ		rate $r_{\sigma,\theta,n}(m)$	
			$m^\sigma \mathbb{E}[Z_{n,j}]$	Asym. 95% HPD	$r_{\sigma,\theta,n}\mathbb{E}[Z_{n,j}]$	Asym. 95% HPD
	n	1281	2602	(2528 , 2677)	1281	(1244 , 1318)
(a)	$10n$	8432	10649	(10347 , 10956)	8432	(8192 , 8675)
$n = 2586$	$100n$	40890	43583	(42345 , 44838)	40890	(39728 , 42067)
	n	346	495	(465 , 531)	346	(325 , 371)
(b)	$10n$	2634	2917	(2739 , 3129)	2634	(2473 , 2825)
$n = 715$	$100n$	16799	17179	(16130 , 18427)	16799	(15774 , 18020)
	n	180	298	(274 , 324)	180	(166 , 196)
(c)	$10n$	1280	1491	(1375 , 1625)	1280	(1181 , 1396)
$n = 363$	$100n$	7205	7474	(6893 , 8146)	7205	(6644 , 7852)
	n	307	525	(491 , 566)	307	(287 , 331)
(d)	$10n$	2085	2457	(2295 , 2648)	2085	(1947 , 2247)
$n = 959$	$100n$	11031	11492	(10735 ,12387)	11031	(10304 , 11889)
	n	440	786	(745 , 831)	440	(417 , 465)
(e)	$10n$	2994	3591	(3407 , 3797)	2994	(2841 , 3166)
$n = 969$	$100n$	15673	16414	(15572 , 17355)	15672	(14869 , 16571)

Table 6.3: Exact estimates of the number of new genes $K_m^{(n)}$ and its asymptotic approximation $f(m)\mathbb{E}[Z_{n,j}]$, with rate functions $f(m) = m^\sigma$ and $f(m) = r_{(\sigma,\theta,n)}$. The size m of the additional sample varies in $\{n, 10n, 100n\}$. The asymptotic 95% HPD intervals are evaluated for both rate functions, m^σ and $r_{(\sigma,\theta,n)}(m)$. All values are rounded to the nearest integer.

Having the asymptotic 95% HPD intervals for $Z_{n,j}$ at hand, the candidate approximate 95% HPD intervals for $K_m^{(n)}$ are $(z_1 m^\sigma, z_2\, m^\sigma)$. As apparent from Table 6.3, the HPD constructed through such a procedure is not centered on and, in most cases, does not even include the estimated number of new genes $\mathbb{E}[K_m^{(n)}|K_n = j]$. Indeed, if one looks at the exact estimator for $K_m^{(n)}$ given in (6.2.26), it is clearly much smaller than its asymptotic approximation $m^\sigma\mathbb{E}[Z_{n,j}]$. This is due to the fact that, when θ and n are moderately large and not overwhelmingly smaller than m, a finer normalization constant is to be used for approximating $K_m^{(n)}$: by close inspection of the derivation of the moments of the limiting r.v. $Z_{n,j}$ in (6.2.32), one sees that an equivalent, though less rough, normalization rate is given by

$$r_{(\sigma,\theta,n)}(m) := (\theta + n + m)^\sigma - (\theta + n)^\sigma.$$

Obviously, in terms of asymptotics, $r_{(\sigma,\theta,n)}(m)/m^\sigma \to 1$ as $m \to \infty$, but, importantly, as far as approximations of $K_m^{(n)}$ for finite m are concerned, it overcomes the above mentioned problems. In fact, we have that, for any m, $\mathbb{E}[K_m^{(n)}|K_n = j] \approx r_{(\sigma,\theta,n)}(m)\mathbb{E}[Z_{n,j}]$ and the asymptotic HPD interval $(r_{(\sigma,\theta,n)}(m)\, z_1, r_{\sigma,\theta,n}(m) z_2)$ is approximately centered on the following estimator $\mathbb{E}[K_m^{(n)}|K_n = j]$, as desired. Table 6.3 displays, for the 5 datasets, the exact estimator for $K_m^{(n)}$, its asymptotic approximation and the 95% asymptotic HPD intervals using both m^σ and $r_{(\sigma,\theta,n)}(m)$ as rate functions for sizes of the additional sample $m \in \{n, 10n, 100n\}$.

For the Tomato flower library we have that, even for $m = 100n = 258600$, the asymptotic approximation of the number of new genes with m^σ is about 6.6% larger than the asymptotic approximation with $r_{(\sigma,\theta,n)}(m)$, which coincides with the exactly es-

Library	m	$\hat{E}_m^{(n,j)}$	Exact 95% HPD	Asym. 95% HPD	Asym. 99% HPD
	n	1281	(1221 , 1341)	(1244 , 1318)	(1234 , 1329)
(a)	$2n$	2354	(2263 , 2449)	(2287 , 2422)	(2269 , 2442)
$n = 2586$	$3n$	3305	(3181 , 3434)	(3211 , 3400)	(3186 , 3430)
	n	346	(312 , 382)	(325 , 371)	(318 , 379)
(b)	$2n$	654	(599 , 711)	(614 , 701)	(601 , 716)
$n = 715$	$3n$	939	(865 , 1015)	(881 , 1007)	(863 , 1028)
	n	180	(156 , 206)	(166 , 196)	(160 , 201)
(c)	$2n$	336	(299 , 375)	(310 , 366)	(299 , 375)
$n = 363$	$3n$	477	(428 , 528)	(440 , 520)	(425 , 533)
	n	307	(271 , 345)	(287 , 331)	(278 , 338)
(d)	$2n$	566	(510 , 624)	(529 , 610)	(513 , 624)
$n = 959$	$3n$	798	(725 , 873)	(746 , 861)	(723 , 880)
	n	440	(402 , 478)	(417 , 465)	(408 , 470)
(e)	$2n$	812	(753 , 873)	(771 , 859)	(755 , 869)
$n = 969$	$3n$	1146	(1069 , 1225)	(1088 , 1212)	(1065 , 1226)

Table 6.4: Exact estimates of the number of new genes and corresponding exact 95% HPD intervals and asymptotic 95% and 99% HPD intervals with rate function $r_{(\sigma,\theta,n)}(m)$ for sizes of the additional sample $m \in \{n, 2n, 3n\}$. All values are rounded to the nearest integer.

timated number. Hence, for finite sample size approximation is definitely necessary to use $r_{(\sigma,\theta,n)}(m)$ as rate function.

We now move on to comparing the asymptotic HPD intervals obtained with the rate function $r_{(\sigma,\theta,n)}(m)$ with the exact HPD intervals determined using the probability distribution in (6.2.23). Hence, we consider $m \in \{n, 2n, 3n\}$, because otherwise the computational burden involved in (6.2.23) would become too heavy. Table 6.4 reports, for the 5 datasets, the exact estimator for $K_m^{(n)}$, the exact 95% HPD and both the 95% and 99% asymptotic HPD intervals. The table shows that the length of the asymptotic 95% HPD intervals is shorter than the exact one, although the difference is not big.

Indeed, such a finding is not surprising in the species sampling context. Obviously, the variability of $K_m^{(n)}$ increases as m increases. However, the variability of $K_m^{(n)}/r_{(\sigma,\theta,n)}(m)$, which can be interpreted as an average variability over the additional sample of size m, is necessarily decreasing as m increases, since the more distinct species are collected the lower the probability of detecting additional new ones will become. Hence, if we approximate $K_m^{(n)}/r_{(\sigma,\theta,n)}(m)$ by its asymptotic r.v. $Z_{n,j}$, we will necessarily underestimate its variability which is reflected on the length of the HPD intervals. Nonetheless the possibility of resorting to the asymptotic HPD intervals is extremely useful from a practical point of view: i) the HPD intervals of $Z_{n,j}$ automatically yield HPD intervals of $K_m^{(n)}$ for any choice of m, whereas the exact HPD intervals have to be recomputed for any m of interest and cannot even be calculated for large m; ii) the fact that the length of the asymptotic HPD intervals is always shorter than the exact length (and not oscillating), allows to interpret it a “lower bound” on the length of the exact ones and, moreover, the underestimation will decrease as m increases.

Given such a “lower bound”, it would be also of interest to have an “upper bound” on the length of the exact HPD. Indeed, if one considers the asymptotic 99% HPD intervals, by Proposition 6.2.7, there exists a $\bar{m}$ such that for any $m > \bar{m}$ the asymptotic 99% HPD interval for $K_m^{(n)}$ covers the exact 95% HPD interval. Hence, for sufficiently large m, the asymptotic 99% HPD interval acts as "upper bound" for the exact one. Although the determination of such a m suitable for any choice of parameters and "basic sample" is not possible, one can proceed empirically. From Table 6.4, where the 99% asymptotic HPD intervals are reported as well, one sees that for the Masingamoeba and Naegleria libraries the asymptotic 99% HPD interval covers the exact 95% HPD interval already starting from $m = 3n$. As for the Tomato flower library, whose distinctive feature with respect to the other libraries is represented by a larger "basic sample", such a covering is not yet achieved for $m = 3n$ but it is very close to happen. Hence, by the combination of the asymptotic 95% and 99% HPD intervals, we obtain a useful device for assessing uncertainty of species richness estimates. Figure 1 below shows, for the Naegleria anaerobic cDNA library, how the 95% and 99% asymptotic HPD intervals provide an envelope around the exact HPD interval from $m \approx 2500$ onwards. Given the two asymptotic HPD are quite close, we thus achieve a satisfactorily accurate estimate of the uncertainty.

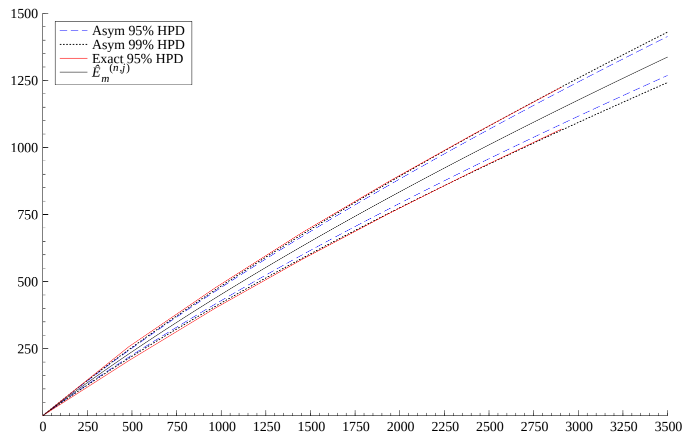

Figure 6.1: Exact estimator $\hat{E}_m^{(n,j)}$ and corresponding exact 95% HPD intervals and asymptotic 95% and 99% HPD intervals for the Naegleria anaerobic library.

In this subsection we applied the results obtained in Substection 6.2.1 whose greatly simplified the implementation of the two parameter Poisson-Dirichlet model for species sampling problems, which is now possible in a straightforward way for any sizes of the basic and the additional sample. This allows its full exploitation in genomics problems, where prediction over large unobserved portions of cDNA libraries is required. In particular, the estimators for the number of new genes, the discovery rate and the sample coverage are completely explicit. Moreover, the conditional asymptotic result concerning the number of new species yields also measures of uncertainty of the estimates in the form of asymptotic HPD intervals, which can be readily used as approximate HPD intervals. Given that the 95% asymptotic HPD interval is always included in the 95% exact HPD interval and that, for sufficiently large m, the 99% asymptotic HPD covers the exact 95% HPD interval, the combination of the 95% and 99% asymptotic HPD intervals provides a simple and valuable measure of uncertainty.

6.3 The generalized Dirichlet process

In this section our aim is to provide solutions to the species sampling problems P1) and P2) when the observed samples are from an exchangeable sequence governed by the generalized Dirichlet process.

The generalized Dirichlet process has been recently introduced in the literature by Regazzini et al. [165] and further investigated by Lijoi et al. [118]. See Chapter 5 for some developments of the generalized Dirichlet process and its characterization in the general class of r.p.m. with logarithmic singularity. In particular, consider a generalized Gamma CRM characterized by the Poisson intensity measure

$$\nu(ds, dx) = \frac{(1 - \mathrm{e}^{-\gamma s})\mathrm{e}^{-s}}{(1 - \mathrm{e}^{-s})s} ds\alpha(dx) \tag{6.3.1}$$

where $\gamma > 0$. Let us denote this CRM by $\tilde{\mu}_\gamma$. If α in (6.3.1) is a non-atomic nfinite measure on $(\mathbb{X}, \mathscr{X})$, we have $0 < \tilde{\mu}_\gamma(\mathbb{X}) < +\infty$ a.s. and the generalized Dirichlet process with parameter γ is defined as the NRMI $\tilde{P}_\gamma(\cdot) := \tilde{\mu}_\gamma(\cdot)/\tilde{\mu}(\mathbb{X})$. Note that if $\gamma = 1$, the intensity in (6.3.1) reduces to the intensity of a Gamma CRM and, hence, $\tilde{P}_\gamma$ becomes a simple Dirichlet process. The fact that the generalized Dirichlet process is not of Gibbs-type follows immediately from Gnedin and Pitman [74] and Lijoi et al. [124]: for $\sigma = 0$, the only Gibbs-type NRMI is the Dirichlet pocess, whereas for $\sigma > 0$ the only NRMIs of Gibbs-type are normalized generalized gamma processes.

As observed in Regazzini et al. [165] if and only if $\gamma \in \mathbb{N}$, $\tilde{\mu}_\gamma$ can be seen as arising from the superposition of γ independent Gamma CRMs with increasing integer-valued scale parameter and shape parameter α. In particular, $\tilde{\mu}_\gamma(A)$, for some $A \in \mathscr{X}$, is then distributed as the convolution of γ independent r.v.s with parameters $(l, \alpha(A))$, for $l = 1, \ldots, \gamma$, i.e.

$$\mathbb{E}[\mathrm{e}^{-\lambda \tilde{P}_\gamma(A)}] = \prod_{l=1}^{\gamma} \left(\frac{l}{l + \lambda} \right)^{\alpha(A)} \qquad \lambda \geq 0. \tag{6.3.2}$$

In the following we always assume $\gamma \in \mathbb{N}$, since this allows to establish the link with Lauricella multiple hypergeometric functions: it is well-known that convolutions of Gamma distribution functions can be represented in terms of Lauricella functions. Moreover, we assume throughout α to be a non-atomic measure, which is tantamount to requiring the prior guess at the shape $\alpha_0(\cdot) = \mathbb{E}[\tilde{P}_\gamma(\cdot)]$ to be non-atomic given that $\alpha_0(\cdot) = \alpha(\cdot)/\theta$. A first treatment of the generalized Dirichlet process in this setup was provided in Lijoi et al. [118], where its finite-dimensional distributions, moments and linear functionals were studied. Moreover, its EPPF, interpretable as the joint distribution of the number of species and their frequencies according to (6.1.2), is given by

$$p_k^{(n)}(n_1,\dots,n_k) = \frac{(\gamma!)^\theta(\theta)^k \prod_{j=1}^k \Gamma(n_j)}{\gamma^{\gamma\theta}(\gamma\theta)_{n\uparrow 1}} \mathfrak{F}(k,n,\mathbf{n},\theta,\gamma) \tag{6.3.3}$$

with

$$\mathfrak{F}(k,n,\mathbf{n},\theta,\gamma) := \sum_{\mathbf{r}^k} F_D^{(\gamma-1)}\left(\gamma\theta, \alpha^*(\mathbf{n},\mathbf{r}^k); \gamma\theta+n; \frac{J_{\gamma-1}}{\gamma}\right) \tag{6.3.4}$$

where the sum is over the set $\mathbf{r}^k := (r_1,\dots,r_k) \in [\gamma]^k$ and $F_D^{(\gamma-1)}$ is for the fourth Lauricella multiple hypergeometric function. The vectors appearing in the arguments of $F_D^{(\gamma-1)}$ are defined as $\alpha^*(\mathbf{n},\mathbf{r}^k) := (\alpha_1^*(\mathbf{n},\mathbf{r}^k),\dots,\alpha_{\gamma-1}^*(\mathbf{n},\mathbf{r}^k))$ with $\alpha_l^*(\mathbf{n},\mathbf{r}^k) := \theta + \sum_{1\le i\le k} n_i \mathbb{1}_{\{l=r_i\}}$ for $l = 1,\dots,\gamma-1$ and $J_{\gamma-1} := (1,\dots,\gamma-1)$. By setting $\gamma = 1$, from (6.3.3) one recovers (6.1.3). The predictive distributions associated with $\tilde{P}_\gamma$ are then of the form (6.1.5) with

$$w_0(n,k,\mathbf{n}) = \frac{\theta\mathfrak{F}(k+1,n+1,\mathbf{n}^+,\theta,\gamma)}{(\gamma\theta+n)\mathfrak{F}(k,n,\mathbf{n},\theta,\gamma)} \tag{6.3.5}$$

$$w_j(n,k,\mathbf{n}) = \frac{\mathfrak{F}(k,n+1,\mathbf{n}_j^+,\theta,\gamma)}{(\gamma\theta+n)\mathfrak{F}(k,n,\mathbf{n},\theta,\gamma)} \tag{6.3.6}$$

where we have set $\mathbf{n}^+ := (n_1,\dots,n_k,1)$ and $\mathbf{n}_i^+ := (n_1,\dots,n_i+1,\dots,n_k)$ for $i = 1,\dots,k$. Before proceeding, a comparison of the predictive structures of Gibbs-type r.p.m.s and the generalized Dirichlet process is in order. In the Gibbs case, the predictive distributions (6.1.4) are a linear combination of the prior guess α_0 and a weighted empirical distribution. So X_{n+1} is new with probability $g_0(n,k)$, whereas it coincides with X_j^* with probability $g_1(n,k)(n_j-\sigma)$, for $j = 1,\dots,k$. Therefore, the weight assigned to each X_j^* depends on the number of distinct observations k and on the number of observations equal to X_j^*, while the weight assigned to a new observation depends solely on the number of distinct observations k as well as the balancing between new and old observations. As already noted, in the limiting case of the Dirichlet process, the dependence on k disappears. The predictive distributions associated with a generalized Dirichlet process are characterized by a more elaborate structure, which exploits all available information in the sample $X_1,\dots,X_n$. Its predictive distributions are still a linear combination of the prior guess α_0 and a weighted empirical distribution, but now X_{n+1} is new with probability $w_0(n,k,\mathbf{n})$ and coincides with X_j^* with probability $n_j w_j(n,k,\mathbf{n})$, for $j = 1,\dots,k$. Therefore, from

(6.3.5) and (6.3.6), we observe that both the weight assigned to each X_j^* and the weight assigned to a new observation depend on the number of distinct observations k as well as on their frequencies $\mathbf{n}$. Moreover, the balance between new and old observations depends on k and $\mathbf{n}$. As pointed out in the Introduction, in principle one would like to work with r.p.m.s, whose predictive structure makes use of all the information contained in the sample as it happens for the generalized Dirichlet process. The fact that it represents, at least to the authors knowledge, the only r.p.m. not of Gibbs-type, which admits closed form expressions for the EPPF and the predictive distributions, makes it even more appealing. In light of the above considerations, it is worth looking for a solution to the problems P1) and P2) described in Section 6.2 for the generalized Dirichlet process case.

The first aim is to derive the distribution of the number of distinct species K_n. To this end we resort to the definition of the (n,k)-th partial Bell polynomial associated with a non-negative weight sequence $w_\bullet := \{w_i, i \geq 0\}$. This allows us to obtain the desired distribution in terms of an Eulero-type integral.

Proposition 6.3.1. *Let $\{X_n, n \geq 1\}$ be an exchangeable sequence governed by a generalized Dirichlet process with parameter $\gamma \in \mathbb{N}$. Then*

$$\mathbb{P}(K_n = k) = \frac{((\gamma)!)^\theta (\theta)^k}{\gamma^{\gamma\theta}\Gamma(n)} \int_0^1 \frac{z^{\gamma\theta-1}(1-z)^{n-1}}{\prod_{l=1}^{\gamma-1}(1-zl/\gamma)^\theta} B_{n,k}(w_\bullet(z,\gamma))dz \tag{6.3.7}$$

where by convention $\prod_{1\leq l\leq 0}(1-zl/\gamma)^a = 1$ and $B_{n,k}(w_\bullet(z,\gamma))$ is the (n,k)-th partial Bell polynomial with weight sequence

$$w_\bullet(z,\gamma) = (\bullet - 1)! \left(\sum_{l=1}^{\gamma-1}(1-zl/\gamma)^{-\bullet} + 1\right) \tag{6.3.8}$$

with the proviso $\sum_{1\leq l\leq 0}(\gamma - zl)^{-\bullet} = 0$ or, equivalently, with exponential generating function given by $w(t,z,\gamma) = -\log(1-t) - \sum_{1\leq l\leq \gamma-1}\log(1+\gamma t/(zl-\gamma))$.

Proof. For any $n \geq 1$ and for any $k = 1, \ldots, n$, we need to compute the following probability

$$\mathbb{P}(K_n = k) = \frac{1}{k!} \sum_{(n_1,\ldots,n_k)\in\mathcal{D}_{k,n}} \binom{n}{n_1,\ldots,n_k} \frac{(\gamma!)^\theta(\theta)^k \prod_{j=1}^k \Gamma(n_j)}{\gamma^{\gamma\theta}(\gamma\theta)_{n\uparrow 1}} \mathfrak{F}(k,n,\mathbf{n},\theta,\gamma),$$

where we recall that $\mathcal{D}_{k,n}\{(n_1,\ldots,n_k) \in [n]^k : \sum_{1\leq i\leq k} n_i = n\}$. Using a known Eulero-type integral representation for the function $F_D^{(\gamma-1)}$ (see Appendix C) we obtain

$$\frac{((\gamma)!)^{\theta}(\theta)^k}{\gamma^{\gamma\theta}\Gamma(n)}\int_0^1 \frac{z^{\gamma\theta-1}(1-z)^{n-1}}{\prod_{l=1}^{\gamma-1}(1-zl/\gamma)^{\theta}}\frac{1}{k!}\sum_{\mathbf{r}^k}\sum_{(n_1,\dots,n_k)\in\mathcal{D}_{k,n}}\binom{n}{n_1,\dots,n_k}$$
$$\times\prod_{j=1}^{k}\Gamma(n_j)\prod_{l=1}^{\gamma-1}\left(1-\frac{zl}{\gamma}\right)^{-\sum_{i=1}^k n_i\mathbb{1}_{\{l=r_i\}}}dz$$
$$=\frac{((\gamma)!)^{\theta}(\theta)^k}{\gamma^{\gamma\theta}\Gamma(n)}\int_0^1 \frac{z^{\gamma\theta-1}(1-z)^{n-1}}{\prod_{l=1}^{\gamma-1}(1-zl/\gamma)^{\theta}}\frac{1}{k!}\sum_{\mathbf{r}^k}\sum_{(n_1,\dots,n_k)\in\mathcal{D}_{k,n}}\binom{n}{n_1,\dots,n_k}$$
$$\times\prod_{j=1}^{k}\Gamma(n_j)\prod_{l=1}^{\gamma-1}\left(1-\frac{zl}{\gamma}\right)^{-\sum_{i=1}^k n_i\mathbb{1}_{\{l=r_i\}}}dz$$
$$=\frac{((\gamma)!)^{\theta}(\theta)^k}{\gamma^{\gamma\theta}\Gamma(n)}\int_0^1 \frac{z^{\gamma\theta-1}(1-z)^{n-1}}{\prod_{l=1}^{\gamma-1}(1-zl/\gamma)^{\theta}}\frac{1}{k!}\sum_{(n_1,\dots,n_k)\in\mathcal{D}_{k,n}}\binom{n}{n_1,\dots,n_k}$$
$$\times\prod_{j=1}^{k}\Gamma(n_j)\left(\sum_{l=1}^{\gamma-1}\left(1-\frac{zl}{\gamma}\right)^{-n_j}+1\right)dz$$
$$=\frac{((\gamma)!)^{\theta}(\theta)^k}{\gamma^{\gamma\theta}\Gamma(n)}\int_0^1 \frac{z^{\gamma\theta-1}(1-z)^{n-1}}{\prod_{l=1}^{\gamma-1}(1-zl/\gamma)^{\theta}}$$
$$\times\frac{n!}{k!}\sum_{(n_1,\dots,n_k)\in\mathcal{D}_{k,n}}\prod_{j=1}^{k}\frac{1}{n_j}\left(\sum_{l=1}^{\gamma-1}\left(1-\frac{zl}{\gamma}\right)^{-n_j}+1\right)dz.$$

Then, it can be easly checked that

$$\frac{n!}{k!}\sum_{(n_1,\dots,n_k)\in\mathcal{D}_{k,n}}\prod_{j=1}^{k}\frac{1}{n_j}\left(\sum_{l=1}^{\gamma-1}\left(1-\frac{zl}{\gamma}\right)^{-n_j}+1\right)$$

corresponds to the (n,k)-th partial Bell polynomial associated to the weight sequence $w_{\bullet}(z,\gamma)$ given by $w_{\bullet}(z,\gamma)=(\bullet-1)!\left(\sum_{1\leq l\leq\gamma-1}(1-zl/\gamma)^{-\bullet}+1\right)$, which is characterized by the exponential generating function $w(t,z,\gamma)=-\log(1-t)-\sum_{1\leq l\leq\gamma-1}\log(1+\gamma t/(zl-\gamma))$. □

Observe that the distribution of the number of distinct species K_n in (6.3.7) can be represented as

$$\mathbb{P}(K_n=k)=\int_0^1\mathbb{P}(K_n=k|z)\eta_n(z)dz$$

where $\eta_n(z)$ is a density on $(0,1)$ and $\mathbb{P}(K_n=k|Z_n)\propto\theta^k B_{n,k}(w_{\bullet}(Z_n,\gamma))$. It is interesting to see how to recover the distribution of K_n in the Dirichlet case. If $\gamma=1$, the weight sequence (6.3.8) does not depend on z anymore, since it simplifies to $w_{\bullet}=(\bullet-1)!$ which is characterized by the exponential generating function $w(t)=-\log(1-t)$. Now, it is well-known (see, e.g., Pitman [157]) that the (n,k)-th partial Bell polynomial with weight sequence $w_\bullet = (\bullet - 1)!$ reduces to the signless Stirling number of the first kind $|s(n,k)|$. Consequently, we have $\mathbb{P}(K_n = k|z) \propto \theta^k |s(n,k)|$ not depending on z and (6.3.7) becomes $\mathbb{P}(K_n = k) = (\theta)^k (\Gamma(n))^{-1} |s(n,k)| \int_0^1 z^{\theta-1}(1-z)^{n-1} dz$. Hence, the mixing variable Z_n is a r.v. distributed according to a Beta distribution function with parameter (θ, n) and, by noting that $(\Gamma(n))^{-1}\Gamma(\theta)\Gamma(n)/\Gamma(\theta+n) = (\theta)_{n\uparrow 1}$, one obtains the distribution of K_n when $X_1, \ldots, X_n$ are sampled from a Dirichlet process. Turning back to the general case (6.3.7), where the conditional distribution of K_n given Z_n is of the form $\mathbb{P}(K_n = k|Z_n) \propto \theta^k B_{n,k}(w_\bullet(Z_n, \gamma))$, we note that such a conditional distribution is induced by a finite Gibbs partition. By finite Gibbs partition we mean a random partition of the integers $[n]$ admitting EPPF of product form $p_k^{(n)}(n_1, \ldots, n_k) = V_{n,k}^* \prod_{1 \le j \le k} W_{n_j}$ but which does not satisfy the addition rule for random partiton. See Pitman [157] an exhaustive account. The fact that $\mathbb{P}(K_n = k|Z_n)$ arises from a finite Gibbs partition is easily seen in the proof of Proposition 6.3.1, which allows to write the conditional EPPF as

$$p_k^{(n)}(n_1, \ldots, n_k|Z_n) = V_{n,k}^* \prod_{j=1}^{k} \Gamma(n_j) \left(\sum_{l=1}^{\gamma-1} \left(1 - \frac{Z_n\, l}{\gamma}\right)^{-n_j} + 1 \right).$$

However, such an EPPF cannot correspond to an infinite exchangeable random partition and, a fortiori not to an infinite exchangeable random partition of Gibbs-type, since by Lemma 2 in Gnedin ane Pitman [74], the terms in the product would necessarily have to be of form $(1-\sigma)_{n_j\uparrow 1}$. This is tantamount of saying that an exchangeable sequence governed by a generalized Dirichlet process cannot be represented as mixture of exchangeable sequences directed by Gibbs-type priors. This is a desirable feature given we would like to have predictions more flexible than those arising from Gibbs-type priors, though, of course, it also implies that we loose the mathematical tractability of exchangeable Gibbs-type partitions.

As for the evaluation of (6.3.7), it is important to remark that, for fixed n and k, $B_{n,k}(w_\bullet)$ is a polynomial of degree n in the variable $(1 - zi/\gamma)$, for $i = 1, \ldots, \gamma - 1$, with a particular set of coefficients specified according to the coefficients of the (n,k)-th partial Bell polynomial $B_{n,k}(w_\bullet)$. Therefore, (6.3.7) can be easily evaluated using the Eulero-type integral representation of the fourth-type Lauricella multiple hypergeometric function $F_D^{(\gamma-1)}$. For instance, if $\gamma = 2$, $F_D^{(1)}$ corresponds to the Gauss hypergeometric function ${}_2F_1$ and (6.3.7) reduces to a weighted linear combination of Gauss hypergeometric functions.

Before passing to the study of distributional properties of conditional samples, we establish an auxiliary result, which is the key for the derivation of the following results and, moreover, is of independent interest. Indeed, it generalizes the well-known multivariate Chu-Vandermonde convolution formula (see Charalambides [17] and Charalambides [16]) and appears to be a new combinatorial identity involving fourth-type Lauricella multiple hypergeometric functions $F_D^{(\gamma-1)}$. Recall that $\mathcal{D}_{k,n} = \{(n_1,\ldots,n_k) \in [n]^k : \sum_{1\le i\le k} n_i = n\}$ and define $\mathcal{D}_{k,n}^{(0)} = \{(n_1,\ldots,n_k) \in (\{0\}\cup[n])^k : \sum_{1\le i\le k} n_i = n\}$.

Lemma 6.3.1. *For any $r \ge 1$, $k \ge 1$ and $a_i > 0$, with $i = 1,\ldots k$,*

$$\sum_{(r_1,\ldots,r_k)\in\mathcal{D}_{k,r}^{(0)}} \binom{r}{r_1,\ldots,r_k} \prod_{i=1}^k w_i^{r_i}(a_i)_{(n_i+r_i-1)\uparrow 1} \tag{6.3.9}$$

$$= w_k^r(n+\sum_{i=1}^k a_i - k)_{r\uparrow 1} \prod_{i=1}^k (a_i)_{n_i-1}\, F_D^{(k-1)}(-r,\mathbf{a}; n+\sum_{i=1}^k a_i - k; W)$$

where $(n_1,\ldots,n_k) \in \mathcal{D}_{k,n}$, $w_i \in \mathbb{R}^+$ for $i = 1,\ldots,k$, $\mathbf{a} := (n_1+a_1-1,\ldots,n_{k-1}-a_{k-1}-1)$ and $W := (w_k - w_1/w_k,\ldots,w_k - w_{k-1}/w_k)$.

Proof. First, we use the integral representation for the Gamma function, i.e. $\Gamma(x) := \int_0^{+\infty} \exp\{-t\}t^{x-1}dt$, together with the Multinomial theorem. In particular, we have

$$\sum_{(r_1,\ldots,r_k)\in\mathcal{D}_{k,r}^{(0)}} \binom{r}{r_1,\ldots,r_k} \prod_{i=1}^k w_i^{r_i}(a_i)_{(n_i+r_i-1),\uparrow 1}$$

$$= \sum_{(r_1,\ldots,r_k)\in\mathcal{D}_{k,r}^{(0)}} \binom{r}{r_1,\ldots,r_k} \frac{\prod_{i=1}^k w^{r_i}\Gamma(n_i+r_i+a_i-1)}{\prod_{i=1}^k \Gamma(a_i)}$$

$$= \frac{1}{\prod_{i=1}^k \Gamma(a_i)} \int_{(\mathbb{R}^+)^k} \mathrm{e}^{-\sum_{i=1}^k u_i} \left(\sum_{i=1}^k w_i u_i\right)^r \prod_{i=1}^k u_i^{n_i+a_i-2} du_1\cdots du_k.$$

By the change of variable $y_i = u_i$, for $i = 1,\ldots,k-1$, and $y_k = \sum_{1\le i\le k} u_i$, we obtain

$$\frac{1}{\prod_{i=1}^k \Gamma(a_i)} \int_0^{+\infty} \mathrm{e}^{-y_k} \int_{B(y_k)} \left(\sum_{i=1}^{k-1} w_i y_i + w_k\left(y_k - \sum_{i=1}^{k-1} y_i\right)\right)^r \times \prod_{i=1}^{k-1} y_i^{n_i+a_i-2}\left(y_k - \sum_{i=1}^{k-1} y_i\right)^{n_k+a_k-2} dy_1\cdots dy_k$$

where $B(y_k) = \{(y_1,\ldots,y_{k-1}) : y_i \ge 0, \sum_{1\le i\le k-1} y_i \le y_k\}$. A further change of variables to $(z_1, \dots, z_{k-1}, z_k) = (y_1/y_k, \dots, y_{k-1}/y_k, y_k)$ yields

$$\frac{1}{\prod_{i=1}^k \Gamma(a_i)} \int_0^{+\infty} \mathrm{e}^{-z_k} z_k^{r+n+\sum_{i=1}^k a_i - k - 1} \int_{\Delta^{(k-1)}} \left(\sum_{i=1}^{k-1} w_i z_i + w_k \left(1 - \sum_{i=1}^{k-1} z_i \right) \right)^r \times \prod_{i=1}^{k-1} z_i^{n_i + a_i - 2} \left(1 - \sum_{i=1}^{k-1} z_i \right)^{n_k + a_k - 2} dz_1 \cdots dz_k$$

$$= \frac{w_k^r \, \Gamma(n + r + \sum_{i=1}^k a_i - k)}{\prod_{i=1}^k \Gamma(a_i)} \int_{\Delta^{(k-1)}} \left(1 - \sum_{i=1}^{k-1} z_i \left(\frac{w_k - w_i}{w_k} \right) \right)^r \times \prod_{i=1}^{k-1} z_i^{n_i + a_i - 2} \left(1 - \sum_{i=1}^{k-1} z_i \right)^{n_k + a_k - 2} dz_1 \cdots dz_{k-1}.$$

Finally, the integral on $\Delta^{(k-1)}$ can be evaluated using an Eulero-type integral representation of the fourth Lauricella hypergeometric function $F_D^{(\gamma-1)}$ (see Appendix C). □

Note that by setting $w_i = 1$ and $a_i = 1$, for $i = 1, \dots, k$, in (6.3.9), one obtains the Chu-Vandermonde convolution formula

$$\sum_{(r_1,\dots,r_j) \in \mathcal{D}_{j,r}^{(0)}} \binom{n_1 + r_1 - 1}{r_1} \cdots \binom{n_j + r_j - 1}{r_j} = \binom{n + r - 1}{r}. \tag{6.3.10}$$

Moreover, if only $w_i = 1$, for $i = 1, \dots, k$, one recovers the extension provided in Lijoi et al. [125]. In the following corollary we consider (6.3.9) for $k = 2$ and we provide an alternative proof for it. For a general $k \geq 1$ the proof follows by simple induction.

Corollary 6.3.1. *For any $r \geq 1$ and $a_1 > 0$, $a_2 > 0$*

$$\sum_{r_1=0}^{r} \binom{r}{r_1} w_1^{r_1} w_2^{r-r_1} (a_1)_{r_1 \uparrow 1} (a_2)_{(r-r_1)\uparrow 1} = {}_2F_1 \left(-r, a_1; a_1 + a_2; w_2 - \frac{w_1}{w_2} \right) w_2^r (a_1 + a_2)_{r \uparrow 1} \tag{6.3.11}$$

Proof. Let us consider the following two known properties for the Gauss hypergeometric function ${}_2F_1$

$${}_2F_1(a, b; b - n; z) = (1 - z)^{-a-n} \sum_{k=0}^{n} \frac{(-n)_{k\uparrow 1} (b - a - n)_{k\uparrow 1} z^k}{(b - n)_{k\uparrow 1} k!} \qquad n \in \mathbb{N} \tag{6.3.12}$$

and

$${}_2F_1(a, b; b - n; z) = \frac{(-1)^n (a)_{n\uparrow 1}}{(1 - b)_{\uparrow 1}} (1 - z)^{-a-n} {}_2F_1(-n, b - a - n; 1 - a - n; 1 - z) \tag{6.3.13}$$

Using equation (6.3.12), we have

$$\begin{aligned} {}_2F_1(a,b;b-n;z) &= (1-z)^{-a-n}{}_2F_1(b-n-a,-n;b-n;z) \\ &= (1-z)^{-a-n}\sum_{k=0}^{n}\frac{(-n)_{k\uparrow 1}(b-a-n)_{k\uparrow 1}z^k}{(b-n)_{k\uparrow 1}k!} \end{aligned}$$

which implies

$${}_2F_1(b-n-a,-n;b-n;z) = \sum_{k=0}^{n}\frac{(-n)_{k\uparrow 1}(b-a-n)_{k\uparrow 1}z^k}{(b-n)_{k\uparrow 1}k!} \qquad n\in\mathbb{N}$$

Then, if we set $n=r$, $b=1-a_2$, $a=-a_2-r+1-a_1$, $k=r_1$ and $z=w_1/w_2$, we obtain the relations

$$\begin{aligned} {}_2F_1\left(a_1,-r;-a_2-r+1;\frac{w_1}{w_2}\right) &= \sum_{r_1=0}^{r}\frac{(-r)_{r_1\uparrow 1}(a_1)_{r_1\uparrow 1}\left(\frac{w_1}{w_2}\right)^{r_1}}{(1-a_2-r)_{r_1\uparrow 1}r_1!} \\ &= \sum_{r_1=0}^{r}\frac{(-r)_{r_1\uparrow 1}}{r_1!}\frac{(a_1)_{r_1\uparrow 1}}{(1-a_2-r)_{r_1\uparrow 1}}\left(\frac{w_1}{w_2}\right)^{r_1} \\ &= \sum_{r_1=0}^{r}\frac{(-1)^{r_1}(-r)_{r_1\uparrow 1}}{r_1!}\frac{(a_1)_{r_1\uparrow 1}}{(-1)^{r_1}(1-a_2-r)_{r_1\uparrow 1}}\left(\frac{w_1}{w_2}\right)^{r_1} \\ &= \sum_{r_1=0}^{r}\binom{r}{r_1}\frac{(a_1)_{r_1\uparrow 1}}{(a_2+r-r_1)_{r_1\uparrow 1}}\left(\frac{w_1}{w_2}\right)^{r_1} \end{aligned}$$

i.e.,

$$\sum_{r_1=0}^{r}\binom{r}{r_1}w_1^{r_1}w_2^{r-r_1}(a_1)_{r_1\uparrow 1}(a_2)_{(r-r_1)\uparrow 1} = {}_2F_1\left(a_1,-r;-a_2-r+1;\frac{w_1}{w_2}\right)w_2^r(a_2)_{r\uparrow 1}$$

Then, if we prove that

$${}_2F_1\left(a_1,-r;-a_2-r+1;\frac{w_1}{w_2}\right)(a_2)_{r\uparrow 1} = {}_2F_1\left(-r,a_1;a_1+a_2;w_2-\frac{w_1}{w_2}\right)(a_1+a_2)_{r\uparrow 1}$$

Using equation (6.3.13)

$$\begin{aligned} {}_2F_1(a,b;b-n;z) &= (1-z)^{-a-n}{}_2F_1(b-n-a,-n;b-n;z) \\ &= \frac{(-1)^n(a)_{n\uparrow 1}}{(1-b)_{\uparrow 1}}(1-z)^{-a-n}{}_2F_1(-n,b-a-n;1-a-n;1-z) \end{aligned}$$

which implies

$${}_2F_1(b-n-a,-n;b-n;z) = \frac{(-1)^n(a)_{n\uparrow 1}}{(1-b)_{\uparrow 1}}{}_2F_1(-n,b-a-n;1-a-n;1-z)$$

Then, if we set $n=r$, $b=1-a_2$, $a=-a_2-r+1-a_1$, $k=r_1$ and $z=w_1/w_2$, we obtain the relations

$$
\begin{aligned}
{}_2F_1(a_1,-r;1-a_2-r;w_1/w_2) &= \frac{(-1)^r(-a_2-a_1-r+1)_{r\uparrow 1}}{(a_2)_{\uparrow 1}}\,{}_2F_1\left(-r,a_1;a_1+a_2;1-\frac{w_1}{w_2}\right)\\
&= \frac{(a_1+a_2)_{r\uparrow 1}}{(a_2)_{\uparrow 1}}\,{}_2F_1\left(-r,a_1;a_1+a_2;w_2-\frac{w_1}{w_2}\right)
\end{aligned}
$$

and the proof is completed. □

The multivariate version of (6.3.11) can be obtained by inductive reasoning from (6.3.11). In particular, we have the following corollary.

Corollary 6.3.2. *For any $q\geq 1$, $j\geq 1$ let $\mathcal{D}_{j,q}:=\{(q_1,\ldots,q_j)\in\{1,\ldots,q\}^j:\sum_{1\leq i\leq j}q_i=q\}$ and let $w_1,\ldots,w_j\in\mathbb{R}^+$ and $a_1,\ldots,a_j>0$. Then*

$$
\begin{aligned}
&\sum_{(q_1,\ldots,q_j)\in\mathcal{D}_{j,q}}\binom{q}{q_1,\ldots,q_j}\prod_{i=1}^{j}w_i^{q_i}(a_i)_{q_i\uparrow 1} \qquad (6.3.14)\\
&\qquad = w_j^q(a)_{q\uparrow 1}F_D^{(j-1)}\left(-q,a_1,\ldots,a_{j-1},a;\frac{w_j-w_1}{w_j},\ldots,\frac{w_j-w_{j-1}}{w_j}\right)
\end{aligned}
$$

where $a:=\sum_{1\leq i\leq j}a_i$.

Proof. Using Equation (6.3.11), the proof follows by inductive reasoning. Suppose the identity holds true for $j-1$, i.e.

$$
\begin{aligned}
&\sum_{(q_1,\ldots,q_{j-1})\in\mathcal{D}_{j-1,q}}\binom{q}{q_1,\ldots,q_{j-1}}\prod_{i=1}^{j-1}w_i^{q_i}(a_i)_{q_i\uparrow 1}\\
&\quad=\sum_{(q_1,\ldots,q_{j-1})\in\mathcal{D}_{j,q}}\frac{q!}{q_1!\cdots q_{j-1}!}w_{j-1}^{q_{j-1}}(a_{j-1})_{q_{j-1}\uparrow 1}\prod_{i=1}^{j-2}w_i^{q_i}(a_i)_{q_i\uparrow 1}\\
&\quad=w_{j-1}^q(a-a_j)_{q\uparrow 1}F_D^{(j-2)}\left(-q,a_1,\ldots,a_{j-2},a-a_j;\frac{w_{j-1}-w_1}{w_{j-1}},\ldots,\frac{w_{j-1}-w_{j-2}}{w_{j-1}}\right)
\end{aligned}
$$

and we show it holds for j as well. Observe that

$$
\begin{aligned}
&\sum_{(q_1,\ldots,q_j)\in\mathcal{D}_{j,q}}\frac{q!}{q_1!\cdots q_j!}\prod_{i=1}^{j}w_i^{q_i}(a_i)_{q_i\uparrow 1}\\
&\quad=\sum_{q_j=0}^{q}\frac{q!}{q_j!(q-q_j)!}w_j^{q_j}(a_j)_{q_j\uparrow 1}\sum_{(q_1,\ldots,q_{j-1})\in\mathcal{D}_{j-1,q-q_j}}\frac{(q-q_j)!}{q_1!\cdots q_{j-1}!}\prod_{i=1}^{j-1}w_i^{q_i}(a_i)_{q_i\uparrow 1}.
\end{aligned}
$$

Then

$$\sum_{(q_1,\dots,q_j)\in\mathcal{D}_{j,q}} \frac{q!}{q_1!\cdots q_j!}\prod_{i=1}^{j} w_i^{q_i}(a_i)_{q_i\uparrow 1}$$
$$= \sum_{q_j=0}^{q} \frac{q!}{q_j!(q-q_j)!} w_j^{q_j}(a_{j-1})_{q_j\uparrow 1} w_{j-1}^{q-q_j}(a-a_j)_{(q-q_j)\uparrow 1}$$
$$\times F_D^{(j-2)}\left(-q+q_j, a_1,\dots,a_{j-2}, a-a_j; \frac{w_{j-1}-w_1}{w_{j-1}},\dots,\frac{w_{j-1}-w_{j-2}}{w_{j-1}}\right)$$
$$= \frac{\Gamma(a-a_j)}{\Gamma(a_1)\cdots\Gamma(a_{j-1})}\int_{\Delta^{(j-2)}}\prod_{i=1}^{j-2} z_i^{a_i-1}\left(1-\sum_{i=1}^{j-2} z_i\right)^{a_{j-1}-1}$$
$$\times \sum_{q_j=0}^{q} \frac{q!}{q_j!(q-q_j)!} w_j^{q_j}(a_j)_{q_j\uparrow 1} w_{j-1}^{q-q_j}(a-a_j)_{(q-q_j)\uparrow 1}$$
$$\times \left(1-\sum_{i=1}^{j-2} z_i \frac{w_{j-1}-w_i}{w_{j-1}}\right)^{q-q_j} dz_1\cdots dz_{j-2}$$
$$= \frac{\Gamma(a-a_j)}{\Gamma(a_1)\cdots\Gamma(a_{j-1})}\int_{\Delta^{(j-2)}}\prod_{i=1}^{j-2} z_i^{a_i-1}\left(1-\sum_{i=1}^{j-2} z_i\right)^{a_{j-1}-1}$$
$$\times \left(1-\sum_{i=1}^{j-2} z_i \frac{w_j-w_i}{w_j} - \left(1-\sum_{i=1}^{j-2} z_i\right)\frac{w_j-w_{j-1}}{w_j}\right)^{q}$$
$$\times \sum_{q_j=0}^{q} \frac{q!}{q_j!(q-q_j)!} w_j^{q-q_j}(a_j)_{q_j\uparrow 1}\left(\frac{-w_j}{\sum_{i=1}^{j-2} z_i \frac{w_j-w_i}{w_j} + \left(1-\sum_{i=1}^{j-2} z_i\right)\frac{w_j-w_{j-1}}{w_j} - 1}\right)^{q_j}$$
$$\times (a-a_j)_{(q-q_j)\uparrow 1} dz_1\cdots dz_{j-2}$$

Then, by applying (6.3.11), we obtain from the last equation

$$\frac{\Gamma(a-a_j)}{\Gamma(a_1)\cdots\Gamma(a_{j-1})} w_j^{q}(a)_{q\uparrow 1}$$
$$\times \int_{\Delta^{(j-2)}}\prod_{i=1}^{j-2} z_i^{a_i-1}\left(1-\sum_{i=1}^{j-2} z_i\right)^{a_{j-1}-1}$$
$$\times \left(1-\sum_{i=1}^{j-2} z_i \frac{w_j-w_i}{w_j} - \left(1-\sum_{i=1}^{j-2} z_i\right)\frac{w_j-w_{j-1}}{w_j}\right)^{q}$$
$$\times {}_2F_1\left(-q, a_j; a; \frac{\sum_{i=1}^{j-2} z_i \frac{w_j-w_i}{w_j} + \left(1-\sum_{i=1}^{j-2} z_i\right)\frac{w_j-w_{j-1}}{w_j}}{\sum_{i=1}^{j-2} z_i \frac{w_j-w_i}{w_j} + \left(1-\sum_{i=1}^{j-2} z_i\right)\frac{w_j-w_{j-1}}{w_j} - 1}\right) dz_1\cdots dz_{j-2}$$

$$
\begin{aligned}
&= \frac{\Gamma(a-a_j)}{\Gamma(a_1)\cdots\Gamma(a_{j-1})} w_j^q(a)_{q\uparrow 1} \\
&\times \int_{\Delta^{(j-2)}} \prod_{i=1}^{j-2} z_i^{a_i-1}\left(1-\sum_{i=1}^{j-2} z_i\right)^{a_{j-1}-1} \\
&\times {}_2F_1\left(-q, a-a_j; a; \sum_{i=1}^{j-2} z_i \frac{w_j-w_i}{w_j} + \left(1-\sum_{i=1}^{j-2} z_i\right)\frac{w_j-w_{j-1}}{w_j}\right) dz_1\cdots dz_{j-2}
\end{aligned}
$$

Since $a-a_j>0$ and

$$
1 > \max\left\{0, \Re\left(\sum_{i=1}^{j-2} z_i \frac{w_j-w_i}{w_j} + (w_j-w_{j-1})\left(1-\sum_{i=1}^{j-2} z_i\right)\right)\right\}
$$

then we can apply equation 7.621.4 in Gradshteyn and Ryzhik [77] and we obtain

$$
\begin{aligned}
&\frac{1}{\Gamma(a_1)\cdots\Gamma(a_{j-1})} w_j^q(a)_{q\uparrow 1} \int_0^{+\infty} \mathrm{e}^{-z_{j-1}} z_{j-1}^{a-a_j-1} \\
&\times \int_{\Delta^{(j-2)}} \prod_{i=1}^{j-2} z_i^{a_i-1}\left(1-\sum_{i=1}^{j-2} z_i\right)^{a_{j-1}-1} \\
&\times {}_1F_1\left(-q; a; z_{j-1}\left(\sum_{i=1}^{j-2} z_i \frac{w_j-w_i}{w_j} + \left(1-\sum_{i=1}^{j-2} z_i\right)\frac{w_j-w_{j-1}}{w_j}\right)\right) dz_1\cdots dz_{j-2}dz_{j-1}
\end{aligned}
$$

using the change of variable $y_i = z_i z_{j-1}$ per $i=1,\ldots,j-2$ and $y_{j-1}=z_{j-1}$ we have

$$
\begin{aligned}
&\frac{1}{\Gamma(a_1)\cdots\Gamma(a_{j-1})} w_j^q(a)_{q\uparrow 1} \int_0^{+\infty} \mathrm{e}^{-y_{j-1}} \\
&\times \int_{B(y_j)} \prod_{i=1}^{j-2} y_i^{a_i-1}\left(y_{j-1}-\sum_{i=1}^{j-2} y_i\right)^{a_{j-1}-1} \\
&\times {}_1F_1\left(-q; a; \sum_{i=1}^{j-2} y_i \frac{w_j-w_i}{w_j} + \left(y_{j-1}-\sum_{i=1}^{j-2} y_i\right)\frac{w_j-w_{j-1}}{w_j}\right) dy_1\cdots dy_{j-1}
\end{aligned}
$$

where

$$
B(y_j) = \left\{(y_1,\ldots,y_{j-1}) : y_i \geq 0, \sum_{i=1}^{j-1} y_i \leq y_j\right\}
$$

and using the change of variable $u_i = y_i$ per $i=1,\ldots,j-2$ e $u_{j-1} = y_{j-1} - \sum_{1\leq i\leq j-2} y_i$ we have

$$
\begin{aligned}
&\frac{1}{\Gamma(a_1)\cdots\Gamma(a_{j-1})} w_j^q(a)_{q\uparrow 1} \int_{(\mathbb{R}^+)^{j-1}} \mathrm{e}^{-\sum_{i=1}^{j-1} u_i} \prod_{i=1}^{j-1} u_i^{a_i-1} \\
&\qquad\times {}_1F_1\left(-q; a; \sum_{i=1}^{j-1} u_i \frac{w_j-w_i}{w_j}\right) du_1\cdots du_{j-1}
\end{aligned}
$$

□

Now, relying on Proposition 6.3.1 and Lemma 6.3.1, we are able to provide a solution to the problem P1) and to the problem P2) for the generalized Dirichlet process.

Proposition 6.3.2. *Let* $\{X_n, n \geq 1\}$ *be an exchangeable sequence governed by a generalized Dirichlet process with parameter* $\gamma \in \mathbb{N}$*. Then*

$$\mathbb{P}(K_m^{(n)} = j | X_{K_n}^{(1,n)}) = \frac{\theta^j (\gamma\theta)_{n\uparrow 1}}{\Gamma(n+m)\mathfrak{F}(k,n,\mathbf{n},\theta,\gamma)} \sum_{s=j}^{m} \binom{m}{s} (n)_{(m-s)\uparrow 1} \tag{6.3.15}$$
$$\times \sum_{\mathbf{r}^k} \int_0^1 \frac{z^{\gamma\theta-1}(1-z)^{n+m-1} B_{s,j}(w_\bullet(z,\gamma)) F_D^{(k-1)}(-(m-s), \mathbf{n}_{k-1}; n; W)}{\prod_{l=1}^{\gamma-1}(1-zl/\gamma)^{\theta+\sum_{i=1}^k n_i \mathbb{1}_{\{l=r_i\}} + (m-s)\mathbb{1}_{\{l=r_k\}}}} dz$$

where $\mathbf{n}_{k-1} := (n_1, \ldots n_{k-1})$ *and* $W := (w_k - w_1/w_k, \ldots, w_k - w_{k-1}/w_k)$ *with* $w_i := \prod_{1\leq l\leq \gamma-1}(1-zl/\gamma)^{-\mathbb{1}_{\{l=r_i\}}}$*. This also implies that* (K_n, N_{K_n}) *is sufficient for predicting the number of new distinct species* $K_m^{(n)}$*.*

Proof. Given the sample $X_{K_n}^{(1,n)}$ with k distinct species and frequencies $\mathbf{n}$, we need to compute

$$\mathbb{P}(K_m^{(n)} = j | X_{K_n}^{(1,n)}) \tag{6.3.16}$$
$$= \sum_{\pi\in\mathscr{P}_{m,k+j}} \frac{p_{k+j}^{(n+m)}(n_1+m_1(\pi), \ldots, n_k+m_k(\pi), m_{k+1}(\pi), \ldots, m_{k+j}(\pi))}{p_k^{(n)}(n_1, \ldots, n_k)},$$

where $\mathscr{P}_{m,k+j}$ denotes the set of all partitions on m observations into $q \leq m$ classes, with $q \in \{j, \ldots, j+k\}$ i.e. j observations are new species and $q - j \leq k$ coincide with some of those already observed in $X_{K_n}^{(1,n)}$. We first focus on the numerator of (6.3.16). In particular, we have

$$\sum_{\pi\in\mathscr{P}_{m,k+j}} p_{k+j}^{(n+m)}(n_1+m_1(\pi), \ldots, n_k+m_k(\pi), m_{k+1}(\pi), \ldots, m_{k+j}(\pi))$$
$$= \sum_{\pi\in\mathscr{P}_{m,k+j}} \sum_{\mathbf{r}^{k+j}} \frac{(\gamma!)^\theta \theta^{k+j} \prod_{i=1}^{k+j} \Gamma((n_i+m_i(\pi))\mathbb{1}_{\{i\leq k\}} + m_i(\pi)\mathbb{1}_{\{k<i\leq k+j\}})}{\gamma^{\gamma\theta}(\gamma\theta)_{(m+n)\uparrow 1}}$$
$$\times F_D^{(\gamma-1)}\left(\gamma\theta, \alpha^*(\mathbf{n}+\mathbf{m}, \mathbf{r}^{k+j}); \gamma\theta+n+m; \frac{J_{\gamma-1}}{\gamma}\right)$$

where $\alpha^*(\mathbf{n}+\mathbf{m}, \mathbf{r}^{k+j}) := (\alpha_1^*(\mathbf{n}+\mathbf{m}, \mathbf{r}^{k+j}), \ldots, \alpha_{\gamma-1}^*(\mathbf{n}+\mathbf{m}, \mathbf{r}^{k+j}))$ having set, for $l = 1, \ldots, \gamma-1$, $\alpha_l^*(\mathbf{n}+\mathbf{m}, \mathbf{r}^{k+j}) := \theta + \sum_{1\leq i\leq k+j}((n_i+m_i(\pi))\mathbb{1}_{\{i\leq k\}} + m_i(\pi)\mathbb{1}_{\{k<i\leq k+j\}})\mathbb{1}_{\{l=r_i\}}$.

Then, using an Eulero-type integral representation for the function $F_D^{(\gamma-1)}$, we get

$$\sum_{\pi\in\mathscr{P}_{m,k+j}} p_{k+j}^{(n+m)}(n_1+m_1(\pi),\dots,n_k+m_k(\pi),m_{k+1}(\pi),\dots,m_{k+j}(\pi))$$

$$= \frac{(\gamma!)^\theta\theta^{k+j}}{\gamma^{\gamma\theta}\Gamma(n+m)}\sum_{\mathbf{r}^{k+j}}\sum_{s=j}^{m}\binom{m}{s}\frac{1}{j!}\sum_{(m_{k+1},\dots,m_{k+j})\in\mathcal{D}_{j,s}}\binom{s}{m_{k+1},\dots,m_{k+j}}\prod_{i=1}^{j}\Gamma(m_{k+i})$$

$$\times\sum_{(m_1,\dots,m_k)\in\mathcal{D}_{k,m-s}^{(0)}}\binom{m-s}{m_1,\dots,m_k}\prod_{i=1}^{k}\Gamma(n_i+m_i)\int_0^1\frac{z^{\gamma\theta-1}(1-z)^{n+m-1}}{\prod_{l=1}^{\gamma-1}(1-zl/\gamma)^{\alpha_l^*(\mathbf{n}+\mathbf{m},\mathbf{r}^{k+j})}}dz \tag{6.3.17}$$

$$= \frac{(\gamma!)^\theta\theta^{k+j}}{\gamma^{\gamma\theta}\Gamma(n+m)}\sum_{\mathbf{r}^{k+j}}\sum_{s=j}^{m}\binom{m}{s}\frac{1}{j!}\sum_{(m_{k+1},\dots,m_{k+j})\in\mathcal{D}_{j,s}}\binom{s}{m_{k+1},\dots,m_{k+j}}\prod_{i=1}^{j}\Gamma(m_{k+i})$$

$$\times\int_0^1\frac{z^{\gamma\theta-1}(1-z)^{n+m-1}}{\prod_{l=1}^{\gamma-1}(1-zl/\gamma)^{\theta+\sum_{i=1}^{k+j}(n_i\mathbb{1}_{\{i\le k\}}+m_i\mathbb{1}_{\{k<i\le k+j\}})\mathbb{1}_{\{l=r_i\}}}}$$

$$\times\sum_{(m_1,\dots,m_k)\in\mathcal{D}_{k,m-s}^{(0)}}\binom{m-s}{m_1,\dots,m_k}\prod_{i=1}^{k}\prod_{l=1}^{\gamma-1}\left(1-\frac{zl}{\gamma}\right)^{-m_i\mathbb{1}_{\{l=r_i\}}}\Gamma(n_i+m_i)dz$$

and by Lemma 6.3.1 we obtain

$$\frac{(\gamma!)^\theta\theta^{k+j}}{\gamma^{\gamma\theta}\Gamma(n+m)}\sum_{\mathbf{r}^{k+j}}\sum_{s=j}^{m}\binom{m}{s}\frac{1}{j!}\sum_{(m_{k+1},\dots,m_{k+j})\in\mathcal{D}_{j,s}}\binom{s}{m_{k+1},\dots,m_{k+j}}\prod_{i=1}^{j}\Gamma(m_{k+i})$$

$$\times\int_0^1\frac{z^{\gamma\theta-1}(1-z)^{n+m-1}}{\prod_{l=1}^{\gamma-1}(1-zl/\gamma)^{\theta+\sum_{i=1}^{k+j}(n_i\mathbb{1}_{\{i\le k\}}+m_i\mathbb{1}_{\{k<i\le k+j\}})\mathbb{1}_{\{l=r_i\}}}}$$

$$\times\prod_{l=1}^{\gamma-1}\left(1-\frac{zl}{\gamma}\right)^{-(m-s)\mathbb{1}_{\{l=r_k\}}}(n)_{(m-s)\uparrow1}\prod_{i=1}^{k}\Gamma(n_i)F_D^{(\gamma-1)}(-(m-s),\mathbf{n}_{k-1};n;W)dz$$

where $\mathbf{n}_{k-1}=(n_1,\dots,n_{k-1})$ and $W=(w_k-w_1/w_k,\dots,w_k-w_{k-1}/w_k)$ with $w_i=\prod_{1\le i\le\gamma-1}(1-zl/\gamma)^{\mathbb{1}_{\{l=r_i\}}}$. The sum with respect the set $\mathcal{D}_{j,s}$ in (6.3.17) can now be computed using an approach similar to the one used for proving Proposition 6.3.1. This allows to write (6.3.17) as

$$\frac{(\gamma!)^\theta\theta^{k+j}}{\gamma^{\gamma\theta}\Gamma(n+m)}\sum_{\mathbf{r}^{k}}\sum_{s=j}^{m}\binom{m}{s}\int_0^1\frac{z^{\gamma\theta-1}(1-z)^{n+m-1}}{\prod_{l=1}^{\gamma-1}(1-zl/\gamma)^{\theta+\sum_{i=1}^{k}n_i\mathbb{1}_{\{l=r_i\}}}}$$

$$\times\frac{s!}{j!}\sum_{(m_{k+1},\dots,m_{k+j})\in\mathcal{D}_{j,s}}\prod_{i=1}^{j}\frac{1}{m_{k+i}}\left(\sum_{l=1}^{\gamma-1}\left(1-\frac{zl}{\gamma}\right)^{-m_{k+i}}+1\right)$$

$$\times\prod_{l=1}^{\gamma-1}\left(1-\frac{zl}{\gamma}\right)^{-(m-s)\mathbb{1}_{\{l=r_k\}}}(n)_{(m-s)\uparrow1}\prod_{i=1}^{k}\Gamma(n_i)F_D^{(\gamma-1)}(-(m-s),\mathbf{n}_{k-1};n;W)dz$$

where

$$\frac{s!}{j!}\sum_{(m_{k+1},\ldots,m_{k+j})\in\mathcal{D}_{j,s}}\prod_{i=1}^{j}\frac{1}{m_{k+i}}\left(\sum_{l=1}^{\gamma-1}\left(1-\frac{zl}{\gamma}\right)^{-m_{k+i}}+1\right)$$

corresponds to the (s, j)-th partial Bell polynomial with weight sequence $w_\bullet(z,\gamma)$ given by (6.3.8). Then, we obtain

$$\frac{((\gamma)!)^{\theta}\theta^{k+j}}{\gamma^{\gamma\theta}\Gamma(n+m)}\sum_{\mathbf{r}^k}\sum_{s=j}^{m}\binom{m}{s}\int_0^1\frac{z^{\gamma\theta-1}(1-z)^{n+m-1}B_{s,j}(w_\bullet(z,\gamma))}{\prod_{l=1}^{\gamma-1}(1-zl/\gamma)^{a+\sum_{i=1}^{k}n_i\mathbb{1}_{\{l=r_i\}}}}$$
$$\times\prod_{l=1}^{\gamma-1}\left(1-\frac{zl}{\gamma}\right)^{-(m-s)\mathbb{1}_{\{l=r_k\}}}(n)_{(m-s)\uparrow 1}\prod_{i=1}^{k}\Gamma(n_i)F_D^{(\gamma-1)}(-(m-s),\mathbf{n}_{k-1};n;W)dz$$

and the proof is completed by dividing the expression by the EPPF of the generalized Dirichlet process with parameter $\gamma\in\mathbb{N}$, which corresponds to the denominator in (6.3.16).

□

It is important to remark that, for the generalized Dirichlet process, the conditional distribution of the number of new distinct species exhibits the desired dependence on both K_n and N_{K_n}. This is in contrast to Gibbs-type r.p.m.s, where we have dependence solely on K_n and not even on K_n in the Dirichlet case. Hence, although the distribution in (6.3.15) is quite complicated, such a property makes the generalized Dirichlet process appealing for practical purposes. Moreover, having (6.3.15) at hand, the computation of the Bayes estimate of $K_m^{(n)}$, given the "basic sample", namely $\mathbb{E}[K_m^{(n)}|K_n=k]$, is straightforward.

With reference to problem P2), we now derive a Bayesian estimator of the probability of discovering a new species at the $(n+m+1)$-th draw, given an initial observed sample of size n with k distinct species and frequencies $\mathbf{n}$. The unobserved sample of size m will feature $K_m^{(n)}\in\{0,1,\ldots,m\}$ new species. Among the m observations $L_m^{(n)}\in\{K_m^{(n)},\ldots,m\}$ will belong to the $K_m^{(n)}$ new species with frequencies $M_{K_m^{(n)}}=(M_{K_n+1},\ldots,M_{K_n+K_m^{(n)}})$ such that $M_i\geq 1$, for $i=K_n+1,\ldots,K_n+K_m^{(n)}$, and $\sum_{K_n+1\leq i\leq K_n+K_m^{(n)}}M_i=L_m^{(n)}$; clearly, if $K_m^{(n)}=0$ also $L_m^{(n)}=0$. Moreover, $m-L_m^{(n)}$ observations will belong to species already observed in the "basic sample" and their frequencies can be characterized by the vector $M_{K_n}=(M_1,\ldots,M_{K_n})$, where $\sum_{1\leq i\leq K_n}M_i=m-L_m^{(n)}$ and $M_i\geq 0$, for $i=1,\ldots,K_n$. Note that $M_i=0$ means that the i-th species of the "basic sample" has not been observed and clearly, if $L_m^{(n)}=m$, then $M_{K_n}=(0,\ldots,0)$. Our aim is to estimate

$$\mathbb{P}(K_1^{n+m}=1|X_{K_n}^{(1,n)},X_{K_m^{(n)}}^{(2,m)}),\tag{6.3.18}$$

where $X_{K_n}^{(1,n)}$ is observed and $X_{K_m^{(n)}}^{(2,m)}$ not. Considering the sufficiency of the r.v. (K_n,N_{K_n}), (6.3.18) reduces to estimating the random probability

$$\mathbb{P}(K_1^{n+m}=1|K_n=k,N_k=\mathbf{n},K_m^{(n)},M_{K_n},M_{K_m^{(n)}})\tag{6.3.19}$$

where the randomness is due to the unobserved $(K_m^{(n)}, M_{K_n}, M_{K_m^{(n)}})$. The next result provides a solution to this problem.

Proposition 6.3.3. *Let $\{X_n, n \geq 1\}$ be an exchangeable sequence governed by a generalized Dirichlet process with parameter $\gamma \in \mathbb{N}$. Then, the Bayes estimate, with respect to a squared loss function, of the probability of observing a new species at the $(n+m+1)$-th draw, conditional on an initial sample of size n with k distinct species and frequencies $\mathbf{n}$, is given by*

$$\hat{D}_m^{(n,k,\mathbf{n})} = \sum_{j=0}^{m} \frac{\theta^{j+1}(\gamma\theta+n)^{-1}(\gamma\theta)_{n\uparrow 1}(\gamma\theta)_{(n+m+1)\uparrow 1}}{\mathfrak{F}(k,n,\mathbf{n},a,\gamma)\Gamma(n+m+1)(\gamma\theta)_{(n+m)\uparrow 1}} \sum_{s=j}^{m} \binom{m}{s} (n)_{(m-s)\uparrow 1} \tag{6.3.20}$$
$$\times \sum_{\mathbf{r}^{k+1}} \int_0^1 \frac{z^{\gamma\theta-1}(1-z)^{n+m-1} B_{s,j}(w_\bullet(z,\gamma)) F_D^{(k-1)}(-(m-s),\mathbf{n}_{k-1};n;W)}{\prod_{l=1}^{\gamma-1}(1-zl/\gamma)^{\theta+\sum_{i=1}^{k+1}(n_i \mathbb{1}_{\{i\leq k\}}+\mathbb{1}_{\{i>k\}})+(m-s)\mathbb{1}_{\{l=r_k\}}}} dz$$

where $\mathbf{n}_{k-1} := (n_1, \dots n_{k-1})$ and $W := (w_k - w_1/w_k, \dots, w_k - w_{k-1}/w_k)$ with $w_i := \prod_{1\leq l\leq \gamma-1}(1-zl/\gamma)^{-\mathbb{1}_{\{l=r_i\}}}$.

Proof. The Bayes estimator of the random probability (6.3.19), with respect to a squared loss function, is given by its expected value with respect to the distribution of the r.v. $(K_m^{(n)}, M_{K_n}, M_{K_m^{(n)}})$ given $X_{K_n}^{(1,n)}$. In particular, we have

$$\hat{D}_m^{(n,k,\mathbf{n})} = \sum_{j=0}^{m}\sum_{s=j}^{m} \binom{m}{s} \frac{1}{j!} \sum_{(m_{k+1},\dots,m_{k+j})\in\mathcal{D}_{j,s}} \binom{s}{m_{k+1},\dots,m_{k+j}}$$
$$\times \sum_{(m_1,\dots,m_k)\in\mathcal{D}_{k,m-s}^{(0)}} \binom{m-s}{m_1,\dots,m_k}$$
$$\times \mathbb{P}(K_1^{n+m}=1|K_n=k, N_{K_n}=\mathbf{n}, K_m^{(n)}=j, M_{K_n}=\mathbf{m}_k, M_{K_m^{(n)}}=\mathbf{m}_j)$$
$$\times \mathbb{P}(K_m^{(n)}=j, M_{K_n}=\mathbf{m}_k, M_{K_m^{(n)}}=\mathbf{m}_j|K_n=k, N_{K_n}=\mathbf{n}).$$

Note that in each summand the first factor is the one step prediction and the second factor is $\mathbb{P}(K_m^{(n)}=j, M_{K_n}=\mathbf{m}_k, M_{K_m^{(n)}}=\mathbf{m}_j|K_n=k, N_{K_n}=\mathbf{n})$, a distribution we have derived in the proof of Proposition 6.3.2. Hence, we obtain

$$\begin{aligned}
\hat{D}_m^{(n,k,\mathbf{n})} &= \sum_{j=0}^{m} \frac{\theta^j (\gamma\theta)_{n\uparrow 1}}{(\gamma\theta)_{n+m\uparrow 1} \prod_{i=1}^{k} \Gamma(n_i)} \sum_{s=j}^{m} \binom{m}{s} \frac{1}{j!} \\
&\qquad \times \sum_{(m_{k+1},\dots,m_{k+j})\in\mathcal{D}_{j,s}} \binom{s}{m_{k+1},\dots,m_{k+j}} \\
&\qquad \times \sum_{(m_1,\dots,m_k)\in\mathcal{D}_{k,m-s}^{(0)}} \binom{m-s}{m_1,\dots,m_k} w_0(n+m,k+j,\mathbf{n}+\mathbf{m}) \\
&\qquad \times \frac{\mathfrak{F}(k+j,n+m,\mathbf{n}+\mathbf{m},a,\gamma)\prod_{i=1}^{j}\Gamma(m_{k+i})\prod_{i=1}^{k}\Gamma(n_i+m_i)}{\mathfrak{F}(k,n,\mathbf{n},\theta,\gamma)} \\
&= \sum_{j=0}^{m} \frac{\theta^{j+1}(\gamma\theta+n)^{-1}(\gamma\theta)_{n\uparrow 1}}{\mathfrak{F}(k,n,\mathbf{n},\theta,\gamma)(\gamma\theta)_{n+m\uparrow 1}\prod_{i=1}^{k}\Gamma(n_i)} \\
&\qquad \times \sum_{s=j}^{m} \binom{m}{s} \frac{1}{j!} \sum_{(m_{k+1},\dots,m_{k+j})\in\mathcal{D}_{j,s}} \binom{s}{m_{k+1},\dots,m_{k+j}} \prod_{i=1}^{j}\Gamma(m_{k+i}) \\
&\qquad \times \sum_{(m_1,\dots,m_k)\in\mathcal{D}_{k,m-s}^{(0)}} \binom{m-s}{m_1,\dots,m_k} \prod_{i=1}^{k}\Gamma(n_i+m_i) \\
&\qquad \times \mathfrak{F}(k+j+1,n+m+1,(\mathbf{n}+\mathbf{m})^+,\theta,\gamma).
\end{aligned}$$

Using the same arguments exploited in the proof of Proposition 6.3.2 we obtain

$$\begin{aligned}
\hat{D}_m^{(n,k,\mathbf{n})} &= \sum_{j=0}^{m} \frac{\theta^{j+1}(\gamma\theta+n)^{-1}(\gamma\theta)_{n\uparrow 1}(\gamma\theta)_{(n+m+1)\uparrow 1}}{\mathfrak{F}(k,n,\mathbf{n},\theta,\gamma)\Gamma(n+m+1)(\gamma\theta)_{(n+m)\uparrow 1}\prod_{i=1}^{k}\Gamma(n_i)} \\
&\times \sum_{\mathbf{r}^{k+1}} \sum_{s=j}^{m} \binom{m}{s} \int_0^1 \frac{z^{\gamma\theta-1}(1-z)^{n+m} B_{s,j}(w_\bullet(z,\gamma))}{\prod_{l=1}^{\gamma-1}(1-zl/\gamma)^{\theta+\sum_{i=1}^{k+1}(n_i\mathbb{1}_{\{i\le k\}}+\mathbb{1}_{\{i>k\}})\mathbb{1}_{\{l=r_i\}}}} \\
&\times \prod_{l=1}^{\gamma-1}\left(1-\frac{zl}{\gamma}\right)^{-(m-s)\mathbb{1}_{\{l=r_k\}}} (n)_{(m-s)\uparrow 1} \prod_{i=1}^{k}\Gamma(n_i) F_D^{(\gamma-1)}(-(m-s),\mathbf{n}_{k-1};n;W)dz.
\end{aligned}$$

Suitable simplifications yield then the estimator in (6.3.20). □

The Bayes estimator in (6.3.20), together with $\mathbb{E}[K_m^{(n)}|(K_n, N_{K_n})]$, represent the new Bayesian counterparts to the celebrated Good-Toulmin estimator (see Good and Toulmin[76]) and represent alternatives to Bayesian estimators derived from Gibbs-type r.p.m.s. With respect to the latter, these estimators have the advantage of incorporating all the information conveyed by the sample at the cost of a higher computational complexity.

In order to complete the description of the conditional structure of generalized Dirichlet processes we now derive the posterior distribution that is the conditional distribution of $\tilde{P}_\gamma$

given a sample $X_1, \dots, X_n$ featuring K_n distinct observations, denoted by $X_1^*, \dots, X_{K_n}^*$, with frequencies N_{K_n}. For any pair of random elements Z and W defined on $(\Omega, \mathscr{F}, \mathbb{P})$, we use the symbol $Z^{(W)}$ to denote a random element on $(\Omega, \mathscr{F}, \mathbb{P})$ whose distribution coincides with a regular conditional distribution of Z, given W. By specializing the general results for NRMIs of James et al. [100], in the next proposition we provide the desired posterior characterization of both the un-normalized CRM $\tilde{\mu}$ with intensity (6.3.1) and the generalized Dirichlet process $\tilde{P}_\gamma(\cdot) := \tilde{\mu}(\cdot)/\tilde{\mu}(\mathbb{X})$. In particular, the result is a special case of Proposition 6.3.4.

Proposition 6.3.4. *Let $\tilde{P}_\gamma$ be a generalized Dirichlet process with parameter $\gamma \in \mathbb{N}$. Then, the distribution of $\tilde{\xi}$, given the observations $X_1, \dots, X_n$ and suitable latent variable U_n, coincides with*

$$\tilde{\mu}^{(U_n, X_1, \dots, X_n)} \stackrel{d}{=} \tilde{\mu}^{(U_n)} + \sum_{j=1}^{K_n} J_i^{(U_n, X_1, \dots, X_n)} \delta_{X_j^*}$$

where

i) $\tilde{\mu}^{(U_n)}$ is a CRM with intensity measure

$$\nu^{(U_n)}(dx, dv) = \sum_{l=1}^{\gamma} \frac{e^{-v(l+U_n)}}{v} dv \alpha(dx); \tag{6.3.21}$$

ii) X_j^* are fixed points of discontinuity, for $j = 1, \dots, K_n$, and the r.v.s $J_j^{(U_n, X_1, \dots, X_n)}$'s are the corresponding jumps which are absolutely continuous with respect to the Lebesgue measure with density

$$f_{J_j^{(U_n, X_1, \dots, X_n)}}(v) \propto v^{n_i - 1} \sum_{l=1}^{\gamma} e^{-v(l+U_n)} \qquad j = 1, \dots, K_n; \tag{6.3.22}$$

iii) the jumps $J_j^{(U_n, X_1, \dots, X_n)}$, for $j = 1, \dots, K_n$, are mutually independent and independent from $\tilde{\mu}(U_n)$.

Moreover, the latent variable U_n, given $X_1, \dots, X_n$, is absolutely continuous with respect to the Lebesgue measure with density

$$f_{U_n^{(X_1, \dots, X_n)}}(u) \propto u^{n-1} \prod_{l=1}^{\gamma} (l+u)^{-a} \prod_{j=1}^{K_n} \Gamma(n_j)(\zeta(n_j, 1+u) - \zeta(n_j, 1+\gamma+u)) \tag{6.3.23}$$

where $\zeta(x, y)$ stands for the generalized Riemann Zeta function (or Herwitz function) with parameters x and y.

Finally, the posterior distribution of $\tilde{P}_\gamma$, given $X_1, \ldots, X_n$ and U_n, is again a NRMI (with fixed points of discontinuity) and coincides in distribution with

$$w \frac{\tilde{\mu}^{(U_n)}}{\tilde{\mu}^{(U_n)}(\mathbb{X})} + (1-w) \frac{\sum_{j=1}^{K_n} J_j^{(U_n, X_1, \ldots, X_n)} \delta_{X_j^*}}{\sum_{j=1}^{K_n} J_j^{(U_n, X_1, \ldots, X_n)}} \tag{6.3.24}$$

where $w = \tilde{\mu}^{(U_n)}(\mathbb{X})[\tilde{\mu}^{(U_n)}(\mathbb{X}) + \sum_{1 \le j \le K_n} J_j^{(U_n, X_1, \ldots, X_n)}]^{-1}$.

Proof. Since $\gamma \in \mathbb{N}$ the intensity measure (6.3.1) of $\tilde{\mu}$ reduces to

$$\nu(dv, dx) = \sum_{l=1}^{\gamma} \frac{\mathrm{e}^{-vl}}{v} dv \alpha(dx).$$

Now since the generalized Dirichlet process is a NRMI and, by hypothesis α is non-atomic, we can apply Theorem 5.2.1 we ensues the existence of a latent variable U_n, such that the distribution of $\tilde{\mu}$, given $X_1, \ldots, X_n$ and U_n coincides with the distribution of $\tilde{\mu}^{(U_n)} + \sum_{1 \le j \le K_n} J_j^{(U_n, X_1, \ldots, X_n)} \delta_{X_j^*}$ where $\tilde{\mu}^{(U_n)}$ is a suitably updated CRM and the $J_j^{(U_n, X_1, \ldots, X_n)}$'s are absolutely continuous with density expressed in the terms of the intensity measure of $\tilde{\mu}$. It is then straightforward to show that the intensity measure associated to $\tilde{\mu}^{(U_n)}$ is of the form (6.3.21) and that the density of $J_j^{(U_n, X_1, \ldots, X_n)}$ is given by (6.3.22). In order to derive the density function for conditional distribution of U_n, given $X_1, \ldots, X_n$ we resort to to Proposition 5.2.1 and, after some algebra, we obtain (6.3.23). Given this, the characterization of the posterior distribution of $\tilde{P}_\gamma$ in (6.3.24) follows from from Theorem 5.2.2. □

Despite the fact that the previous result completes the theoretical analysis of the conditional structure induced by generalized Dirichlet processes, it is also useful for practical purposes. Indeed, large values of the parameter γ combined with large additional samples m, make the numerical computation of the distributions and estimators derived in Propositions cumbersome. If this is the case, then one can devise a simulation algorithm relying on the posterior characterization of Proposition 6.3.4. By combining an inverse Lévy measure algorithm, such as the Ferguson-Klass method (see Ferguson and Klass [64] and Walker and Damien [190]), for simulating trajectories of $\tilde{\mu}^{(U_n)}$ with a Metropolis-Hasting step for drawing samples from $U_n^{X_1, \ldots, X_n}$, one easily obtains realizations of the posterior distribution of the generalized Dirichlet process. Then one can sample a new value X_{n+1}, update the posterior according to Proposition 6.3.4 and sample a realization of the posterior given (X, X_{n+1}). Proceeding along these lines up to step m one obtains a realization of the additional sample $X_{n+1}, \ldots, X_{n+m}$. By repeating the procedure N times, one obtains a collection of future scenarios $\{(X_{n+1}^{(i)}, \ldots, X_{n+m}^{(i)}), i = 1, \ldots, N\}$ which can be used in order to evaluate the quantities of interest. For instance, if $j^{(i)}$ is the number of new distinct species observed in $X_{n+1}^{(i)}, \ldots, X_{n+m}^{(i)}$, $\mathbb{E}[K_m^{(n)}|K_n]$ can be evaluated as $1/N \sum_{1\le i\le N} j_m^{(i)}$. Finally note that Proposition 6.3.4 is also important in the context of mixture modeling, where inference is necessarily simulation based given the complexity of the models: in fact, it allows to derive, in the terminology of Papaspiliopoulos and Roberts [148], conditional sampling schemes, which in the case of the generalized Dirichlet process, are simpler to implement than marginal ones.

6.4 NRMIs in species sampling problems

We conclude this chapter by investigating some conditional structures that emerge when the observations are sampled from exchangeable sequences governed by a general homogeneous NRMIs.

The EPPF of the associated exchangeable random partition is known to be

$$p_k^{(n)}(n_1, \ldots, n_k) = \frac{1}{\Gamma(n)} \int_{\mathbb{R}^+} u^{n-1} \mathrm{e}^{-\Psi(u)} \prod_{j=1}^{k} \kappa_{n_j}(u) du \tag{6.4.1}$$

where

$$\kappa_{n_j}(u) = \int_{\mathbb{X}\times\mathbb{R}^+} v^{n_j} \mathrm{e}^{-uv} \rho(dv)\alpha(dx) \qquad j = 1, \ldots, k$$

The issue we address consists in evaluating, conditionally on an exchangeable random partition with EPPF (6.4.1), the probability of sampling in m draws a certain number of observations yielding new partition groups with specified frequencies.

Proposition 6.4.1. *Let $\{X_n, n \ge 1\}$ be an exchangeable sequence of r.v.s governed by an homogeneous NRMI on $\mathbb{X}$ with Poisson intensity measure $\nu(ds, dx) = \rho(ds)\alpha(dx)$. If $X_{K_n}^{(1:n)}$ and $X_{K_m^{(n)}}^{(2:m)}$ are two sample from $\{X_n, n \ge 1\}$, then the joint distribution of $K_m^{(n)}$, $L_m^{(n)}$ and $S_{L_m^{(n)}}$, given K_n and N_{K_n}, is of the form*

$$\mathbb{P}(K_m^{(n)} = k, L_m^{(n)} = s, S_{L_m^{(n)}} = \mathbf{s}|K_n = j, N_{K_n} = \mathbf{n}) \tag{6.4.2}$$
$$= \binom{m}{s} \frac{\Gamma(n) \int_{\mathbb{R}^+} u^{n+m-1} e^{-\Psi(u)} \iota_{m-s}(u) \prod_{i=1}^{k} \kappa_{s_i}(u) du}{\Gamma(n+m) \int_{\mathbb{R}^+} u^{n-1} e^{-\Psi(u)} \prod_{i=1}^{j} \kappa_{n_i}(u) du}$$

where we defined

$$\iota_{m-s}(u) := \int_{(\mathbb{X}\times\mathbb{R}^+)^j} \left(\sum_{i=1}^{j} v_i\right)^{m-s} \prod_{i=1}^{j} v_i^{n_i} e^{-uv_i} \rho(dv_i)\alpha(dx)$$

Proof. Let $\Lambda_{m-L_m^{(n)}} = (\Lambda_{1,m-L_m^{(n)}}, \dots, \Lambda_{K_n,m-L_m^{(n)}})$ be the vector of non-negative integers denoting the number of new observations in each of the j groups into which the first n observations are partitioned. If we define the set

$$D_{j,m-s}^{(0)} := \left\{ (\lambda_1, \dots, \lambda_j) \in \{0, \dots, m-s\}^j, \sum_{i=1}^{j} \lambda_i = m-s \right\}$$

then,

$$\begin{aligned}
&\mathbb{P}(K_m^{(n)} = k, L_m^{(n)} = s, S_{L_m^{(n)}} = \mathbf{s} | K_n = j, N_{K_n} = \mathbf{n}) \\
&\quad = \binom{m}{s} \sum_{(\lambda_1,\dots,\lambda_j)\in D_{j,m-s}^{(0)}} \binom{m-s}{\lambda_1 \cdots \lambda_j} \frac{1}{k!} \binom{s}{s_1 \cdots s_k} \\
&\qquad\qquad \times \frac{\Gamma(n) \int_{\mathbb{R}^+} u^{n+m-1} \mathrm{e}^{\Psi(u)} \prod_{i=1}^{j} \kappa_{n_i+\lambda_i(u)} \prod_{i=1}^{k} \kappa_{s_i}(u) du}{\Gamma(n+m) \int_{\mathbb{R}^+} u^{n-1} \mathrm{e}^{\Psi(u)} \prod_{i=1}^{j} \kappa_{n_i}(u) du}
\end{aligned}$$

In particular, using the expression for $\kappa_{n_i+\lambda_i}(u)$ at the numerator of the last equation, we obtain

$$\begin{aligned}
&\sum_{(\lambda_1,\dots,\lambda_j)\in D_{j,m-s}^{(0)}} \binom{m-s}{\lambda_1 \cdots \lambda_j} \prod_{i=1}^{j} \kappa_{n_i+\lambda_i}(u) \\
&\qquad = \sum_{(\lambda_1,\dots,\lambda_j)\in D_{j,m-s}^{(0)}} \binom{m-s}{\lambda_1 \cdots \lambda_j} \prod_{i=1}^{j} \int_{\mathbb{X}\times\mathbb{R}^+} v^{n_i+\lambda_i} \mathrm{e}^{-uv} \rho(dv) \alpha(dx)
\end{aligned}$$

For any integral in the product, we consider the change of variable $v = s_i$ for $i = 1, \dots, j$, then the last equation can be written as

$$\sum_{(\lambda_1,\dots,\lambda_j)\in D_{j,m-s}^{(0)}} \binom{m-s}{\lambda_1 \cdots \lambda_j} \int_{(\mathbb{X}\times\mathbb{R}^+)^j} \prod_{i=1}^{j} v_i^{n_i+\lambda_i} \mathrm{e}^{-uv_i} \rho(dv_i) \alpha(dx)$$

and for the Multinomial theorem we obtain

$$\int_{(\mathbb{X}\times\mathbb{R}^+)^j} \left(\sum_{i=1}^{j} v_i \right)^{m-s} \prod_{i=1}^{j} v_i^{n_i} \mathrm{e}^{-uv_i} \rho(dv_i) \alpha(dx)$$

Then, the result can be obtained by substitution. □

By marginalizing the conditional distribution in (6.4.2) with respect to $S_{L_m^{(n)}}$ and, then, with respect to $K_m^{(n)}$ one obtains the conditional distribution for the number of new groups and the number of observations belonging to these new groups and the distribution of $L_m^{(n)}$, respectively. In particular, the solution to the problem P1) can be obtained by marginalizing the conditional distribution in (6.4.2) with respect to $(L_m^{(n)}, S_{L_m^{(n)}})$.

Corollary 6.4.1. *Let $\{X_n, n \geq 1\}$ be an exchangeable sequence of r.v.s governed by an homogeneous NRMI on $\mathbb{X}$ with Poisson intensity measure $\nu(ds, dx) = \rho(ds)\alpha(dx)$. If $X^{(1:n)}_{K_n}$ and $X^{(2:m)}_{K^{(n)}_m}$ are two sample from $\{X_n, n \geq 1\}$, then the joint distribution of $K^{(n)}_m$ and $L^{(n)}_m$, given K_n and N_{K_n}, is of the form*

$$\mathbb{P}(K^{(n)}_m = k, L^{(n)}_m = s | K_n = j, N_{K_n} = \mathbf{n}) \tag{6.4.3}$$
$$= \binom{m}{s} \frac{\Gamma(n) \int_{\mathbb{R}^+} u^{n+m-1} e^{-\Psi(u)} B_{s,k}(\kappa_\bullet(u)) \iota_{m-s}(u) du}{\Gamma(n+m) \int_{\mathbb{R}^+} u^{n-1} e^{-\Psi(u)} \prod_{i=1}^{j} \kappa_{n_i}(u) du}$$

for $k \leq s = 0, \ldots, m$ where $B_{s,k}(w_\bullet(u))$ is the (s,k)-th partial Bell polynomial with weight sequence $w_\bullet(u) = \kappa_\bullet(u)$. The distribution of $K^{(n)}_m$, conditionally given K_n and N_{K_n}, and the distribution of $L^{(n)}_m$, given K_n and N_{K_n}, is of the form

$$\mathbb{P}(K^{(n)}_m = k | K_n = j, N_{K_n} = \mathbf{n}) \tag{6.4.4}$$
$$= \sum_{s=k}^{m} \binom{m}{s} \frac{\Gamma(n) \int_{\mathbb{R}^+} u^{n+m-1} e^{-\Psi(u)} B_{s,k}(w_\bullet(u)) \iota_{m-s}(u) du}{\Gamma(n+m) \int_{\mathbb{R}^+} u^{n-1} e^{-\Psi(u)} \prod_{i=1}^{j} \kappa_{n_i}(u) du}$$

and

$$\mathbb{P}(L^{(n)}_m = s | K_n = j, N_{K_n} = \mathbf{n}) = \binom{m}{s} \frac{\Gamma(n) \int_{\mathbb{R}^+} u^{n+m-1} e^{-\Psi(u)} B_s(w_\bullet(u)) \iota_{m-s}(u) du}{\Gamma(n+m) \int_{\mathbb{R}^+} u^{n-1} e^{-\Psi(u)} (\prod_{i=1}^{j} \kappa_{n_i}(u)) du} \tag{6.4.5}$$

where $B_s(w_\bullet(u)) = \sum_{0 \leq k \leq s} B_{s,k}(w_\bullet(u))$.

Proof. If we define

$$D^{(1)}_{k,s} := \left\{ (s_1, \ldots, s_k) \in [s]^k, \sum_{i=1}^{k} s_i = s \right\}$$

then we obtain

$$\mathbb{P}(K^{(n)}_m = k, L^{(n)}_m = s | K_n = j, N_{K_n} = \mathbf{n})$$
$$= \sum_{(s_1, \ldots, s_k) \in D^{(1)}_{k,s}} \binom{m}{s} \frac{1}{k!} \binom{s}{s_1 \cdots s_k} \frac{\Gamma(n) \int_{\mathbb{R}^+} u^{n+m-1} \mathrm{e}^{-\Psi(u)} \iota_{m-s}(u) \prod_{i=1}^{k} \kappa_{s_i}(u) du}{\Gamma(n+m) \int_{\mathbb{R}^+} u^{n-1} \mathrm{e}^{-\Psi(u)} \prod_{i=1}^{j} \kappa_{n_i}(u) du}$$
$$= \binom{m}{s} \frac{\Gamma(n) \int_{\mathbb{R}^+} u^{n+m-1} \mathrm{e}^{-\Psi(u)} B_{s,k}(w_\bullet(u)) \iota_{m-s}(u) du}{\Gamma(n+m) \int_{\mathbb{R}^+} u^{n-1} \mathrm{e}^{-\Psi(u)} \prod_{i=1}^{j} \kappa_{n_i}(u) du}$$

where

$$B_{s,k}(w_\bullet(u)) = \frac{s!}{k!} \sum_{(s_1, \ldots, s_k) \in D^{(1)}_{k,s}} \prod_{i=1}^{k} \frac{\kappa_{s_i}(u)}{s_i!}$$

is the (s, k)-th partial Bell polynomial with weight sequence $w_\bullet(u) = \kappa_\bullet(u)$. Equation (6.4.4) and equation (6.4.5) can be easily obtained from equation (6.4.3) summing over s and k, respectively. $\square$

Bayes estimator, under quadratic loss function, for the expected number of new clusters can be easily recoverd from equation (6.4.4) and equation (6.4.5). In particular, we have

$$\mathbb{E}[K_m^{(n)} | K_n = j, N_{K_n} = \mathbf{n}] \tag{6.4.6}$$
$$= \sum_{k=0}^{m} k \sum_{s=k}^{m} \binom{m}{s} \frac{\Gamma(n) \int_{\mathbb{R}^+} u^{n+m-1} \mathrm{e}^{-\Psi(u)} B_{s,k}(w_\bullet(u)) \iota_{m-s}(u) du}{\Gamma(n+m) \int_{\mathbb{R}^+} u^{n-1} \mathrm{e}^{-\Psi(u)} \prod_{i=1}^{j} \kappa_{n_i}(u) du}$$

Often, interest relies also in determining an estimator for the number of observations in the subsequent m sample that will belong to the new species. For this purpose, one can resort ro (6.4.5) and the corresponding Bayes estimator is given by

$$\mathbb{E}[L_m^{(n)} | K_n = j, N_{K_n} = \mathbf{n}] = \sum_{s=0}^{m} s \binom{m}{s} \frac{\Gamma(n) \int_{\mathbb{R}^+} u^{n+m-1} \mathrm{e}^{-\Psi(u)} B_{s}(w_\bullet(u)) \iota_{m-s}(u) du}{\Gamma(n+m) \int_{\mathbb{R}^+} u^{n-1} \mathrm{e}^{-\Psi(u)} \prod_{i=1}^{j} \kappa_{n_i}(u) du} \tag{6.4.7}$$

In particular, $\mathbb{E}[L_m^{(n)} | K_n = j]/m$ is the expected proportion of genes in the new sample which do not coincide with previously observed ones.

Turning to the problem ii), we now derive a Bayesian nonparametric estimator for the probability of discovering a new species at the $(n+m+1)$-th draw, given the "basic sample" $X_{K_n}^{(1:n)}$. If we suppose, for the moment, that we have observed both the basic sample and the second sample, the discovery probability is given by

$$\mathbb{P}(K_1^{(n+m)} = 1 | K_m^{(n)} = k, L_m^{(n)} = s, S_{L_m^{(n)}} = \mathbf{s}, K_n = j, N_{K_n} = \mathbf{n})$$

However, our estimate is obtained without observing the outcome of the "second sample" and, hence, we have to estimate the random probability

$$D_m^{(n:j)} := \mathbb{P}(K_1^{(n+m)} = 1 | K_m^{(n)}, L_m^{(n)}, S_{L_m^{(n)}}, K_n = j, N_{K_n} = \mathbf{n}) \tag{6.4.8}$$

where the randomness in the above expression is due to the randomness of $(K_m^{(n)}, L_m^{(n)}, S_{L_m^{(n)}})$. Baysian inference on (6.4.8) is based on the posterior distribution provided in Corollary (6.4.1). Thus, the bayesian estimator of (6.4.8), with respect to a quadratic loss function, is given by its expected value with respect to the posterior distribution of the number of species. This represents a Bayesian counterpart to the celebrated Good-Toulmin estimator. In other word we provide a Bayesian nonparametric estimator for

$$U_{n+m} = \sum_{i \geq 1} p_i \mathbb{1}_{\{0\}}(N_{i,n+m})$$

Proposition 6.4.2. *Let $\{X_n, n \geq 1\}$ be an exchangeable sequence of r.v.s governed by an homogeneous NRMI on $\mathbb{X}$ with intensity measure $\nu(ds,dx) = \rho(ds)\alpha(dx)$. The Bayes estimate, under quadratic loss function, of the probability of observing a new species at the $(n+m+1)$-th draw, given the $X_{K_n}^{(1,n)}$, is given by*

$$\hat{D}_m^{(n:j)} = \left((n+m)\Gamma(n+m)\int_{\mathbb{R}^+} u_2^{n-1} e^{-\Psi(u_2)} (\prod_{i=1}^{j} \kappa_{n_i}(u_2)) du_2\right)^{-1} \tag{6.4.9}$$
$$\times \Gamma(n) \int_{(\mathbb{R}^+)^2} \tau_1(u_1)(u_1^2 u_2)^{n+m-1} e^{-\Psi(u_2)-(j+k)\Psi(u_1)} \iota_{m-s}^*(u_1,u_2) B_{s,k}(w_\bullet(u_1,u_2)) du_2 du_1$$

where

$$\iota_{m-s}^*(u_1,u_2) := \int_{(\mathbb{X}\times(\mathbb{R}^+)^2)^j} \left(\sum_{i=1}^{j} v_{2,i} v_{1,i}\right)^{m-s}$$
$$\times \prod_{i=1}^{j} (v_{2,i}v_{1,i})^{n_i} e^{-u_{2,i}v_{2,i}-u_{1,i}v_{1,i}} \rho(dv_{2,i})\rho(dv_{1,i})\alpha(dx)$$

and where $B_{s,k}(w_\bullet(u_1,u_2))$ is the (s,k)-th partial Bell polynomial with weight sequence $w_\bullet(u_1,u_2) = \tau_\bullet(u_1)\kappa_\bullet(u_2)$.

Proof. Let us first considere $\tilde{P}$ an homogeneous NRMI. Undere this assumption, to find the Bayes estimate, under quadratic loss function, of the probability of observing a new species at the $(n+m+1)$-th draw, conditional on the $X_{K_n}^{(1,n)}$ we can use arguments similar to those one used in Proposition 6.4.1. In particular, let $\Lambda_{m-L_m^{(n)}} = (\Lambda_{1,m-L_m^{(n)}}, \ldots, \Lambda_{K_n,m-L_m^{(n)}})$ be the vector of non-negative integers denoting the number of new observations in each of the j groups into which the first n observations are partitioned. Let $D_{j,m-s}^{(0)}$ and $D_{k,s}^{(1)}$ be the sets defined in Proposition 6.4.1 and Corollary 6.4.1, respectively. Then

$$\hat{D}_m^{(n:j)} = \sum_{k=0}^{m}\sum_{s=k}^{m}\binom{m}{s} \sum_{(\lambda_1,\ldots,\lambda_j)\in D_{j,m-s}^{(0)}} \binom{m-s}{\lambda_1\cdots\lambda_j}\frac{1}{k!} \sum_{(s_1,\ldots,s_k)\in A_{k,s}^{(1)}} \binom{s}{s_1\cdots s_k}$$
$$\times \mathbb{P}(K_1^{(n+m)} = 1 | K_m^{(n)} = k, L_m^{(n)} = s, S_{L_m^{(n)}} = \mathbf{s}, K_n = j, N_{K_n} = \mathbf{n})$$
$$\times \mathbb{P}(K_m^{(n)} = k, L_m^{(n)} = s, S_{L_m^{(n)}} = \mathbf{s} | K_n = j, N_{K_n} = \mathbf{n})$$

where in each summand the first factor is the one step prediction which characterized a NRMI and which corresponds to (see James et. al [100])

$$\frac{1}{n+m}\int_{\mathbb{R}^+} u\tau_1(u) f_{U_n}^{(X_{K_n}^{(1,n)}, X_{K_n^{(m)}}^{(2,m)})}(u) du$$

where

$$f_{U_n}^{(X_{K_n}^{(1,n)}, X_{K_n^{(m)}}^{(2,m)})}(u) \propto u^{n+m-1} \prod_{i=1}^{K_n} \tau_{n_i+\lambda_i}(u) \prod_{i=1}^{K_n^{(m)}} \tau_{s_i}(u) \mathrm{e}^{-\Psi(u)}$$

and the second factor is given by equation (6.4.2). Then, we obtain

$$\sum_{k=0}^{m}\sum_{s=k}^{m}\binom{m}{s}\sum_{(\lambda_1,\dots,\lambda_j)\in D^{(0)}_{j,m-s}}\binom{m-s}{\lambda_1\cdots\lambda_j}\frac{1}{k!}\sum_{(s_1,\dots,s_k)\in A^{(1)}_{k,s}}\binom{s}{s_1\cdots s_k}$$
$$\times\frac{\Gamma(n)\int_{\mathbb{R}^+}u^{n+m-1}\mathrm{e}^{-\Psi(u)}(\prod_{i=1}^{j}\kappa_{n_i+\lambda_i(u)}\prod_{i=1}^{k}\kappa_{s_i}(u))du}{\Gamma(n+m)\int_{\mathbb{R}^+}u^{n-1}\mathrm{e}^{-\Psi(u)}(\prod_{i=1}^{j}\kappa_{n_i}(u))du}$$
$$\times\frac{1}{n+m}\int_{\mathbb{R}^+}u\tau_1(u)f_{U_n}^{(X^{(1,n)}_{K_n},X^{(2,m)}_{K_n^{(m)}})}(u)du$$

which is proportional to

$$\sum_{k=0}^{m}\sum_{s=k}^{m}\binom{m}{s}\sum_{(\lambda_1,\dots,\lambda_j)\in D^{(0)}_{j,m-s}}\binom{m-s}{\lambda_1\cdots\lambda_j}\frac{1}{k!}\sum_{(s_1,\dots,s_k)\in A^{(1)}_{k,s}}\binom{s}{s_1\cdots s_k}$$
$$\times\frac{\Gamma(n)\int_{\mathbb{R}^+}u_2^{n+m-1}\mathrm{e}^{-\Psi(u_2)}(\prod_{i=1}^{j}\kappa_{n_i+\lambda_i}(u_2)\prod_{i=1}^{k}\kappa_{s_i}(u_2))du_2}{(n+m)\Gamma(n+m)\int_{\mathbb{R}^+}u_2^{n-1}\mathrm{e}^{-\Psi(u_2)}(\prod_{i=1}^{j}\kappa_{n_i}(u_2))du_2}$$
$$\times\int_{\mathbb{R}^+}u_1\tau_1(u_1)u_1^{n+m-1}\prod_{i=1}^{j}\tau_{n_i+\lambda_i}(u_1)\prod_{i=1}^{k}\tau_{s_i}(u_1)\mathrm{e}^{-\Psi(u_1)}du_1$$

where for the first integral we used the change of variable $u = u_2$ and for the second integral we used the change of variable $u = u1$. As regard the sum on $D^{(0)}_{j,m-s}$ we can solve it by using the definition of $\tau_{n_i+\lambda_i}(u_1)$ and $\kappa_{n_i+\lambda_i}(u_2)$. In particular we have

$$\sum_{(\lambda_1,\dots,\lambda_j)\in D^{(0)}_{j,m-s}}\binom{m-s}{\lambda_1\cdots\lambda_j}\prod_{i=1}^{j}\tau_{n_i+\lambda_i}(u_1)\kappa_{n_i+\lambda_i}(u_2)$$
$$=\sum_{(\lambda_1,\dots,\lambda_j)\in D^{(0)}_{j,m-s}}\binom{m-s}{\lambda_1\cdots\lambda_j}\prod_{i=1}^{j}\int_{\mathbb{X}\times(\mathbb{R}^+)^2}(v_2v_1)^{n_i+\lambda_i}\mathrm{e}^{-u_2v_2-u_1v_1}\rho(dv_2)\rho(dv_1)\alpha(dx)$$

For any integral in the product, we consider the change of variable $v_1 = v_{1,i}$ and $v_2 = v_{2,i}$, for $i = 1,\dots,j$, then the last equation can be written as

$$\sum_{(\lambda_1,\dots,\lambda_j)\in D^{(0)}_{j,m-s}}\binom{m-s}{\lambda_1\cdots\lambda_j}\int_{(\mathbb{X}\times(\mathbb{R}^+)^2)^j}\prod_{i=1}^{j}(v_{2,i}v_{1,i})^{n_i+\lambda_i}\mathrm{e}^{-u_{2,i}v_{2,i}-u_{1,i}v_{1,i}}\rho(dv_{2,i})\rho(dv_{1,i})\alpha(dx)$$

and for the Multinomial theorem, we obtain

$$\int_{(\mathbb{X}\times(\mathbb{R}^+)^2)^j}\left(\sum_{i=1}^{j}v_{2,i}v_{1,i}\right)^{m-s}\prod_{i=1}^{j}(v_{2,i}v_{1,i})^{n_i}\mathrm{e}^{-u_{2,i}v_{2,i}-u_{1,i}v_{1,i}}\rho(dv_{2,i})\rho(dv_{1,i})\alpha(dx)$$

As regard the sum on $D_{k,s}^{(1)}$ we have

$$\frac{1}{k!}\sum_{(s_1,\dots,s_k)\in D_{k,s}^{(1)}}\binom{s}{s_1\cdots s_k}\prod_{i=1}^{k}\tau_{s_i}(u_1)\kappa_{s_i}(u_2)$$

which corresponds to the (s,k)-th partial Bell polynomial with weight sequence $w_\bullet(u_1,u_2) = \tau_\bullet(u_1)\kappa_\bullet(u_2)$. □

We are now going to consider an importan quantity which describes the partition structure of the observations generating new groups in a further sampling procedure, conditional on the partition generated by the first n observations.

Proposition 6.4.3. *Let $\{X_n, n\geq 1\}$ be an exchangeable sequence of r.v.s governed by a NRMI on $\mathbb{X}$ with intensity measure $\nu(ds,dx)=\rho(ds)\alpha(dx)$. If $X_{K_n}^{(1:n)}$ and $X_{K_m^{(n)}}^{(2:m)}$ are two samples from $\{X_n, n\geq 1\}$, then the joint distribution of $K_m^{(n)}$ and $S_{L_m^{(n)}}$, given $L_m^{(n)}$, K_n and N_{K_n}, is of the form*

$$\begin{aligned}\mathbb{P}(K_m^{(n)}=k, S_{L_m^{(n)}}=\mathbf{s}|L_m^{(n)}=s, K_n=j, N_{K_n}=\mathbf{n}) \\ = \frac{1}{k!}\binom{s}{s_1\cdots s_k}\frac{\int_{\mathbb{R}^+}u^{n+m-1}e^{-\Psi(u)}\prod_{i=1}^{k}\kappa_{s_i}(u)\iota_{m-s}(u)du}{\int_{\mathbb{R}^+}u^{n+m-1}e^{-\Psi(u)}B_s(w_\bullet(u))\iota_{m-s}(u)du}\end{aligned} \tag{6.4.10}$$

for any $s\in[m]$, $k\in[s]$, $j\in[n]$, $(n_1,\dots,n_j)\in\Delta_{n,j}$ and $(s_1,\dots,s_k)\in\Delta_{s,k}$. Consequentely the partition of the observations which belongs to the new partition set is, conditional on the basic sample of size n and conditional on U_{n+m} a finite Gibbs-type random partition

$$\begin{aligned}\mathbb{P}(K_m^{(n)}=k, S_{L_m^{(n)}}=\mathbf{s}|L_m^{(n)}=s, K_n=j, N_{K_n}=\mathbf{n}, U_{n+m}=u) \\ = \frac{1}{k!}\binom{s}{s_1\cdots s_k}\frac{\Gamma(n+m)e^{-\Psi(u)}\iota_{m-s}(u)\prod_{i=1}^{k}\kappa_{s_i}(u)}{\int_{\mathbb{R}^+}t^{n+m}e^{ut}f_T(t)dt\int_{\mathbb{R}^+}u^{n+m-1}e^{-\Psi(u)}B_s(w_\bullet(u))\iota_{m-s}(u)du}\end{aligned} \tag{6.4.11}$$

Proof. This is straightforward and follows from taking the ratio between (6.4.2) and (6.4.5). □

The finiteness of the random partition described by (6.4.11) is obvious, since it takes values on the space of all partitions of $[s]$ with $1\leq s\leq m$. Moreover, the partition structure featured by the conditional distribution in (6.4.11) motivates the following definition.

Definition 6.4.1. *The conditional probability distribution*

$$\tilde{p}_k^{(s)}(s_1,\dots,s_k;m,n,j,\mathbf{n}) := \mathbb{P}(K_m^{(n)}=k, S_{L_m^{(n)}}=\mathbf{s}|L_m^{(n)}=s, K_n=j, N_{K_n}=\mathbf{n}) \tag{6.4.12}$$

with $1\leq s\leq m$ and $1\leq k\leq s$, is termed conditional EPPF.

A

Appendix

This appendix reviews some definitions and results in the field of the enumerative combinatorics. In particular, after introducing the preliminary notations, we remind the definition of composite structure and the definition of central and non-central Stirling number and central and non-central generalized factorial coefficient.

A.1 Notation

For any $n \in \mathbb{N}$ and any arbitrary x let $(x)_{n\downarrow 1}$ be the factorial of x of order n, that is

$$(x)_{n\downarrow 1} := x(x-1)\cdots(x-n+1) = \prod_{i=0}^{n-1}(x-i)$$

with $(x)_{0\downarrow 1} := 1$ and for $x \neq -r$, $r = 1, \ldots, n$

$$(x)_{-n\downarrow 1} = \frac{1}{(x+n)(x+n-1)\cdots(x+1)}.$$

In addition to the notation $(x)_{n\downarrow 1}$ of the factorial of x of order n, we consider the notation $(x)_{n\uparrow 1}$ which denotes the relation of the factorials to the powers. The factorial $(x)_{n\downarrow 1}$ is called descending (or falling) in distinction to the ascending (or rising) factorial of x of order n, which for $n \in \mathbb{N}$ is defined by

$$(x)_{n\uparrow 1} := x(x+1)\cdots(x+n-1) = \prod_{i=0}^{n-1}(x+i).$$

A distinct notation is unnecessary since both factorials can be expressed by the same notation, only the argument being different. In particular, this product, with the adopted notation, equals

$$(x)_{n\uparrow 1} = (-1)^n(-x)_{n\downarrow 1}.$$

The generalized factorial of x of order n and increment α, denoted by $(x)_{n\downarrow\alpha}$, is defined for any $n \in \mathbb{N}$ and any arbitrary real x and α by

$$(x)_{n\downarrow\alpha} := x(x-\alpha)\cdots(x-\alpha n+\alpha) = \prod_{i=0}^{n-1}(x-\alpha i)$$

with $(x)_{0\downarrow\alpha} := 1$ and for $x \neq -r\alpha$, $r = 1, \ldots, n$

$$(x)_{-n\downarrow\alpha} = \frac{1}{(x+n\alpha)(x+n\alpha-\alpha)\cdots(x+\alpha)}.$$

In additional to the notation $(x)_{n\downarrow\alpha}$ of the generalized factorial of x of order n and increment α, we consider the notation $(x)_{n\uparrow\alpha}$ which denotes the relation of the factorials to the powers. The generalized factorial $(x)_{n\downarrow\alpha}$ is called descending (or falling) in distinction to the ascending (or rising) generalized factorial of x of order n and increment α, which for any $n \in \mathbb{N}$ is defined by

$$(x)_{n\uparrow\alpha} := x(x+\alpha)\cdots(x+n\alpha-\alpha) = \prod_{i=0}^{n-1}(x+i\alpha).$$

Note that the generalized factorial of x of order n and increment α, $(x)_{n\downarrow\alpha}$, may be expressed as a factorial of x of order n and scale parameter s

$$(x)_{n\downarrow\alpha} = s^{-n}(sx)_{n\downarrow 1}$$

with $s = 1/\alpha$.

Another equally important function is the binomial coefficient of x of order n which is defined for any $n \in \mathbb{N}$ and any arbitrary x by

$$\binom{x}{n} := \frac{(x)_{n\downarrow 1}}{n!}$$

with $\binom{x}{0} := 1$. The generalized binomial coefficient of x of order n and increment a which is defined for any $n \in \mathbb{N}$ and any arbitrary x and a by

$$\binom{x}{n}_a := \frac{(x)_{n\downarrow a}}{n!}$$

with $\binom{x}{0}_a := 1$.

A.2 Composite structures

Let $v_\bullet := \{v_i, i \geq 1\}$ and $w_\bullet := \{w_i, i \geq 1\}$ be two sequences of non-negative integers. Let V be some species of combinatorial structures (see Bergeron et al. [6] and Bergeron et al.

[5]), so for each finite set F_n with $|F_n| = n$ elements there is some construction of a set $V(F_n)$ of V-structures on F_n, such that the number of V-structures on a set of n elements is $|V(F_n)| = v_n$. For instance $V(F_n)$ might be $F_n \times F_n$, or $F_n^{F_n}$, or permutations from F_n to F_n, or rooted trees labeled F_n, corresponding to the sequences $v_n = n^2$, or n^n, of $n!$, or n^{n-1} respectively.

Let W be another species of combinatorial structures, such that the number of W-structures on a set of j elements is w_j. Let $(V \circ W)(F_n)$ denote the composite structure on F_n defined as the set of all ways to partition F_n into blocks $\{A_1, \dots, A_k\}$ for some $1 \le k \le n$, assign this collection of blocks a V-structure, and assign each block A_i a W-structure. Then, for each set F_n with n elements, the number of such composite structures is evidently

$$|(V \circ W)(F_n)| = B_n(v_\bullet, w_\bullet) := \sum_{k=1}^{n} v_k B_{n,k}$$

where, denoting by $\mathcal{P}_{[n]}^k$ the set of partition s of the set $[n] := \{1, \dots, n\}$

$$B_{n,k}(w_\bullet) := \sum_{\{A_1,\dots,A_k\} \in \mathcal{P}_{[n]}^k} \prod_{i=1}^{k} w_{|A_i|}$$

is the number of ways to partition F_n into k blocks and assign each block a W-structure.

The sum $B_{n,k}(w_\bullet)$ is a polynomial in variables $w_1, \dots, w_{n-k+1}$, known as the (n,k)-th partial Bell polynomial (see Comtet [19]). For a partition π_n of n into k parts with m_j parts equal to j for $1 \le j \le n$, the coefficient of $\prod_j w_j^{m_j}$ in $B_{n,k}(w_\bullet)$ is the number of partitions Π_n of the set $[n]$ corresponding to π_n. That is to say,

$$\left[\prod_j w_j^{m_j}\right] B_{n,k}(w_\bullet) = \frac{n!}{\prod_j (j!)^{m_j} m_j!}$$

with $\sum_j j m_j = n$ and $\sum_j m_j = k$.

Here, we provide an alternative and more formal definition of a partial Bell polynomial in terms of its associated exponential generating function. Given a sequence of real numbers $w_\bullet := \{w_i, i \ge 1\}$ we define the exponential generating function of this sequence as the formal power series

$$w(t) = \sum_{k \ge 0} w_k \frac{t^k}{k!}. \tag{A.2.1}$$

The power series is said to be formal since questions concerning convergence are never asked. We only wish to evaluate it, and its derivatives, at $t = 0$, i.e.,

$$w_k = \frac{d^k}{dt^k} w(t)\bigg|_{t=0}. \tag{A.2.2}$$

Since (A.2.2) is the defining relationship for coefficients of a Taylor series, (A.2.1) is sometimes referred to as a formal Taylor series and w_k is the k-th Taylor coefficient of $w(t)$.

Exponential generating functions are used extensively in combinatorics and closesly associated with them are the so-called partial Bell polynomials. These are polynomials in an infinite number of variables, these variables are actually the coefficients $w_\bullet$ of an undetermined exponential generating function

$$w(t) = \sum_{k \geq 1} w_k \frac{t^k}{k!} \tag{A.2.3}$$

where we assume $w_0 = w(0) = 0$. In particular, we have the following formal definitio of partial Bell polynomials.

Definition A.2.1. *Let $w_\bullet$ be a sequence of real numbers. Then the (n,k)-th partial Bell polynomial $B_{n,k}(w_\bullet)$ is defined by the expansion*

$$e^{xw(t)} = \sum_{n \geq 0} \sum_{k \geq 0} B_{n,k}(w_\bullet) x^k \frac{t^n}{n!}$$

where $w(t)$ is the exponential generating function of the sequence $w_\bullet$ and $w_0 = w(0) = 0$.

From Definition A.2.1 it is possible to isolate $B_{n,k}(w_\bullet)$ by differentiating the appropriate number of times and then setting $x = t = 0$, i.e.

$$B_{n,k}(w_\bullet) = \frac{\partial^n}{\partial t^n} \frac{1}{k!} \frac{\partial^k}{\partial x^k} \mathrm{e}^{xw(t)} \bigg|_{x=0,t=0}$$

or

$$B_{n,k}(w_\bullet) = \frac{1}{k!} \frac{d^n}{dt^n} w^k(t) \bigg|_{t=0}$$

for any $n \in \mathbb{N}_0$ and $k \in \mathbb{N}_0$. This shows that $B_{n,k}(w_\bullet)$ corresponds to the n-th Taylor coefficient of $(1/k!)w^k(t)$ or, more precisely, $w^k(t)/k! = \sum_{n \geq 0} B_{n,k}(w_\bullet) t^n/n!$. By setting $k = 0$ one gets $B_{0,0} = 1$ and $B_{n,0} = 0$, for any $n \in \mathbb{N}$, whereas for $k = 1$ one has $B_{n,1} = w_n$ for all $n \in \mathbb{N}_0$. Also, since $w_0 = 0$, one has

$$\frac{1}{k!} w^k(t) = \frac{1}{k!} \left(w_1 t + w_2 \frac{t^2}{2!} + \cdots \right)^k = w_1^k \frac{t^k}{k!} + \cdots \tag{A.2.4}$$

so that $B_{n,k}(w_\bullet) = 0$ whenever $k > n$ and $B_{n,n}(w_\bullet) = w_1^n$ for all $n \in \mathbb{N}_0$. By expanding (A.2.4) and examining the coefficient of $t^n/n!$, one obtains the following explicit expression for $B_{n,k}(w_\bullet)$

$$B_{n,k}(w_\bullet) = \sum_{\substack{i_1, i_2, \ldots \geq 0 \\ i_1 + i_2 + \cdots = k \\ i_1 + 2i_2 + 3i_3 + \cdots = n}} \frac{n!}{i_1! i_2! \cdots (1!)^{i_1} (2!)^{i_2} \cdots} w_1^{i_1} w_2^{i_2} \cdots . \tag{A.2.5}$$

Note that, if the variable w_s occurs in $B_{n,k}(w_\bullet)$, then the summation conditions imply that, for some $i_1 \geq 0, i_2 \geq 0, \ldots, i_s \geq 1, \ldots$, we have $s-1 \leq i_2+2i_3+\cdots+(s-1)i_s+\cdots = n-k$, giving $s \leq n-k+1$. Thus $B_{n,k}(w_\bullet) = B_{n,k}(w_1, \ldots, w_{n-k+1})$ meaning $B_{n,k}(w_\bullet)$ depends at most on the variable $w_1, \ldots, w_{n-k+1}$ and no others. Moreover, it is easy to see that $B_{n,k}(w_\bullet)$ is homogeneous of degree k and it can be shown by a combinatorial argument that all of the coefficients are actually integers. In Table A.1 we provide some partial Bell polynomials for $1 \leq k \leq n \leq 5$. In general, partial Bell polynomials can be computed

n	$B_{n,1}(w_\bullet)$	$B_{n,2}(w_\bullet)$	$B_{n,3}(w_\bullet)$	$B_{n,4}(w_\bullet)$	$B_{n,5}(w_\bullet)$
1	w_1	-	-	-	-
2	w_2	w_1^2	-	-	-
3	w_3	$3w_1w_2$	w_1^3	-	-
4	w_4	$4w_1w_3+3w_1^2$	$6w_1^2w_2$	w_1^4	-
5	w_5	$5w_1w_4+10w_2w_3$	$10w_1^2w_3+15w_1w_2^2$	$10w_1^3w_2$	w_1^5

Table A.1: Some partial Bell polynomilas

using (A.2.5). Alternatively, one can resort to the recurrence relation

$$B_{n,k}(w_\bullet) = \sum_{m=1}^{n-k+1} \binom{n}{m} w_m B_{n-m,k-1}(w_\bullet).$$

Finally, we give the definition of Bell polynomial which is a particular type of partition polynomial. The partition polynomials have been introduced by Bell [3] and they are multivariable polynomials that are defined by a sum extended over all partitons of their index. These polynomials have found many applications in combinatorics, probability theory and statistics, as well as in number theory.

Definition A.2.2. *Let $w_\bullet$ be a sequence of real numbers. Then the Bell polynomial $B_n(x, w_\bullet)$ is a polynomial in x defined by*

$$B_n(x, w_\bullet) = \sum_{k=0}^{n} x^k B_{n,k}(w_\bullet).$$

The Bell polynomials can also be thounght of as generated by the expansion

$$\mathrm{e}^{xw(t)} = \sum_{n\geq 0} B_n(x)\frac{t^n}{n!}.$$

Note that since $B_{n,n}(w_\bullet) = w_1^n$ we have deg $B_n(x) = n$ for all n if $w_1 \neq 0$ and deg $B_n(x) < n$ for all $n \in \mathbb{N}$ if $w_1 = 0$. It is always true that deg $B_0(x) = 0$ since $B_0(x) = 1$ for every $w(t)$.

Before giving some specific examples, we shall introduce a little more notation. It is sometimes convenient to write $B_{n,k}(w(t))$ for $B_{n,k}(w_\bullet)$ and $B_n(x, w(t))$ for $B_n(x, w_\bullet)$, where as always $w(t)$ is defined by (A.2.3). When this is done it is important to remember that $B_{n,k}(w(t))$ is not a function of t but of $w_\bullet$. Moving from the definition of partial Bell polynomial in term of its exponential generating function, we now consider some of the most well known instances of partial Bell polynomials $B_{n,k}(w(t))$ and their associated Bell polynomials $B_n(x, w(t))$:

i) the standard polynomials $x^n = B_n(x, t)$;

ii) the falling factorials $(x)_{n\downarrow 1} = B_n(x, \log(1+t))$

$$(x)_{n\downarrow 1} = x(x-1)\cdots(x-n+1) = \sum_{k=0}^{n} s(n,k)x^k$$

where the numbers $s(n,k) := B_{n,k}(\log(1+t))$ are the Stirling numbers of the first kind;

iii) the rising factorials $(x)_{n\uparrow 1} = B_n(x, -\log(1-t))$

$$(x)_{n\uparrow 1} = x(x+1)\cdots(x+n-1) = \sum_{k=0}^{n} |s(n,k)|x^k$$

so that $|s(n,k)| = B_{n,k}(-\log(1-t))$. In particular $|s(n,k)|$ are called the signless Stirling numbers of the first kind and

$$|s(n,k)| = \sharp\{\text{permutations of the set } [n] \text{ with } k \text{ cycles}\}$$

where the last equality corresponds to the representation of a permutation of the set $[n]$ as the product of cycles permutations acting on the blocks of some partition of the set $[n]$;

n	$\lvert s_{n,1}\rvert$	$\lvert s_{n,2}\rvert$	$\lvert s_{n,3}\rvert$	$\lvert s_{n,4}\rvert$	$\lvert s_{n,5}\rvert$
1	1	-	-	-	-
2	1	1	-	-	-
3	2	3	1	-	-
4	6	11	6	1	-
5	24	50	35	10	1

Table A.2: Some signless Stirling numbers of the first kind

iv) the exponential polynomials $\phi_n(x) := B_n(x, \mathrm{e}^t - 1)$

$$\phi_n(x) = \sum_{k=0}^{n} S(n,k)x^k$$

where the numbers $S(n,k) := B_{n,k}(\mathrm{e}^t - 1)$ are called the Stirling numbers of the second kind and

$$S(n,k) = \sharp\{\text{partitions of the set } [n] \text{ into } k \text{ blocks}\}$$

n	$S_{n,1}$	$S_{n,2}$	$S_{n,3}$	$S_{n,4}$	$S_{n,5}$
1	1	-	-	-	-
2	1	1	-	-	-
3	1	3	1	-	-
4	1	7	6	1	-
5	1	15	25	10	1

Table A.3: Some Stirling numbers of the second kind

In particular, related to the Stirling numbers of the second kind are the Bell numbers which are defined by $B_n := \phi_n(1) = \sum_{0\le k\le n} S(n,k)$ and they count the number of distinct partitions of a set with n elements;

v) the Laguerre polynomials $L_n(x) := B_n(x, t/(t-1))$

$$L_n(x) = \sum_{k=0}^{n} L(n,k)x^k$$

where the numbers $L(n,k) = B_{n,k}(t/(t-1))$ are called the Lah numbers which are given by

$$L(n,k) = (-1)^n \frac{n!}{k!}\binom{n-1}{k-1}$$

vi) the Abel polynomials $A_n(x) := B_n(x, 1, -2a, (3a)^2, \dots, (-ka)^{k-1}, \dots)$

$$A_n(x) = x(x-na)^{n-1}$$

where a is a fixed constant.

A.3 Stirling numbers and generalized factorial coefficients

We provide exact definitions of central generalized factorial coefficients and non-central generalized factorial coefficients. For further details and pointers to the literature see Charalambides and Singh [15], Charalambides [16] and Charalambides [17].

Consider the falling factorial of x of order n

$$(x)_{n\downarrow 1} = x(x-1)\cdots(x-n+1) \qquad n \in \mathbb{N} \tag{A.3.1}$$

with $(x)_{0\downarrow 1} = 1$. Clearly, this is a polynomial of x of degree n. Executing the multiplications and arranging the terms in ascending order of powers of x we get

$$(x)_{n\downarrow 1} = \sum_{k=0}^{n} s(n,k)x^k \qquad n \in \mathbb{N}_0. \tag{A.3.2}$$

Inversely, the n-th power of x may be expressed in the form of a polynomial of the factorials of x of degree n. In particular, using (A.3.1), we get successively the expressions

$$x^0 = (x)_{0\downarrow 1}$$

$$x^1 = (x)_{1\downarrow 1}$$

$$x^2 = (x)_{1\downarrow 1} + (x)_{2\downarrow 1}$$

$$x^3 = (x)_{1\downarrow 1} + 3(x)_{2\downarrow 1} + (x)_{3\downarrow 1}$$

and generally

$$x^n = \sum_{k=0}^{n} S(n,k)(x)_{n\downarrow 1} \qquad n \in \mathbb{N}_0. \tag{A.3.3}$$

Clearly, $s(0,0) = S(0,0) = 1$, $s(n,0) = S(n,0) = 0$, for any $n \in \mathbb{N}$ and $s(n,k) = S(n,k) = 0$ for $k > n$. Further, replacing in (A.3.2) x by $-x$, and since $(x)_{n\uparrow 1} = (-1)^n(-x)_{n\downarrow 1}$, we deduce the expression

$$(x)_{n\uparrow 1} = \sum_{k=0}^{n} |s(n,k)| x^k \qquad n \in \mathbb{N}_0 \tag{A.3.4}$$

where

$$|s(n,k)| = (-1)^{n-k} s(n,k).$$

Note that $|s(n,k)|$, according to (A.3.4), as a sum of products of $n-k$ positive integers from the set $[n-1]$, is a positive integer. Based on expansion (A.3.2), (A.3.3) and (A.3.4), the following definition is introduced.

Definition A.3.1. *(cfr. Charalambides [17]) The coefficients $s(n,k)$ and $S(n,k)$ in the expansions of factorials into powers and of powers into factorials are called Stirling numbers of the first and second kind, respectively. The coefficient $|s(n,k)|$ in the expansion of rising factorials into powers in the signless or absolute Stirling number of the first kind.*

An interesting and useful extension of the Stirling numbers, in combinatorics and discrete probability, is provided by the coefficients of the expansions of non-central factorials into powers and of powers into non-central factorials. Specifically, let

$$(x-r)_{n\downarrow 1} = \sum_{k=0}^{n} s(n,k;r) x^k \qquad n \in \mathbb{N}_0 \tag{A.3.5}$$

and

$$x^n = \sum_{k=0}^{n} S(n,k;r)(x-r)_{k\downarrow 1} \qquad n \in \mathbb{N}_0.$$

Clearly, $s(0,0;r) = S(0,0;r) = 1$, $s(n,0;r) = (-r)_{n\downarrow 1}$, $S(n,0;r) = r^n$ for any $n \in \mathbb{N}$, and $s(n,k;r) = S(n,k;r) = 0$ for $k > n$. Further, replacing x by $-x$ in (A.3.5) and since $(x+r)_{n\uparrow 1} = (-1)^n(-x+r)_{n\downarrow 1}$, we deduce the expression

$$(x+r)_{n\uparrow 1} = \sum_{k=0}^{n} |s(n,k;r))| x^k \qquad n \in \mathbb{N}_0 \tag{A.3.6}$$

where the coefficient

$$|s(n,k;r)| = (-1)^{n-k} s(n,k;r) \qquad k \in \mathbb{N}_0, n \in \mathbb{N}_0$$

for $r > 0$, as a sum of products of positive numbers, is positive. Then, the following definition is introduced.

Definition A.3.2. *(cfr. Charalambides [17]) The coefficients $s(n,k;r)$ and $S(n,k;r)$ in the expansions of non-central factorials into powers and of powers into non-central factorials are called non-central Stirling numbers of the first and second kind, respectively. The coefficient $|s(n,k;r)|$ for $r > 0$, in the expansion of non-central rising factorials into powers in the non-central signless or absolute Stirling number of the first kind.*

Notice that for $r = 0$, the non-central Stirling numbers reduce to the corresponding (central) Stirling numbers. For $r \neq 0$, these numbers may be expressed in terms of the corresponding central Stirling numbers. Specifically, expanding the non-central ascending factorial of x of order n, $(x+r)_{n\uparrow 1}$ into powers of $u = x + r$, using (A.3.4), and then expanding the powers of $u = x + r$ into powers of x, using the Binomial theorem, we deduce the expansion

$$\begin{aligned}(x+r)_{n\uparrow 1} &= \sum_{j=0}^{n} |s(n,j)|(x+r)^j = \sum_{j=0}^{n} |s(n,j)| \sum_{k=0}^{j} \binom{j}{k} x^k r^{j-k} \\ &= \sum_{k=0}^{n} \left(\sum_{j=k}^{n} \binom{j}{k} r^{j-k} |s(n,j)| \right) x^k\end{aligned}$$

which, compared to (A.3.6), yields the expression

$$|s(n,k;r)| = \sum_{j=k}^{n} \binom{j}{k} r^{j-k} |s(n,j)|$$

Also, expanding the non-central ascending factorial of x of order n, $(x+r)_{n\uparrow 1}$, into ascending factorials of x, using Vardermonde formula

$$(x+t)_{n\downarrow 1} = \sum_{r=0}^{n} (t)_{r\downarrow 1} (x)_{(n-r)\downarrow 1} \qquad t \in \mathbb{R}, x \in \mathbb{R} \tag{A.3.7}$$

and then expanding the ascending factorials of x into powers of x, using (A.3.4), we get the expansion

$$(x+r)_{n\uparrow 1} = \sum_{j=0}^{n} \binom{n}{j} (x)_{j\uparrow 1} (r)_{(n-j-1)\uparrow 1}$$
$$= \sum_{j=0}^{n} \binom{n}{j} (r)_{(n-j-1)\uparrow 1} \sum_{k=0}^{j} |s(j,k)| x^k = \sum_{k=0}^{n} \left(\sum_{j=k}^{n} \binom{n}{j} (r)_{(n-j-1)\uparrow 1} |s(j,k)| \right) x^k$$

which, compared to (A.3.6), implies the expression

$$|s(n,k;r)| = \sum_{j=k}^{n} \binom{n}{j} (r)_{(n-j-1)\uparrow 1} |s(j,k)|.$$

Similarly,

$$S(n,k;r) = \sum_{j=k}^{n} \binom{j}{k} (r)_{(j-k)\downarrow 1} S(n,j)$$

and

$$S(n,k;r) = \sum_{j=k}^{n} \binom{j}{k} r^{n-j} S(n,j).$$

The non-central Stirling numbers retain almost all the properties of the (central) Stirling numbers. In particular, the properties of the (central) Stirling numbers, whenever needed, are deduced by setting te non-centrality parameter equal to zero.

Consider the generalized factorial of x of order n and scale parameter s

$$(sx)_{n\downarrow 1} = sx(sx-1)\cdots(sx-n+1) \qquad n \in \mathbb{N}$$

with $(sx)_{0\downarrow 1} = 1$, where s a real number. It can be expressed as a polynomial of factorials of x of degree n. Specifically, we successively get the expressions

$$(sx)_{0\downarrow 1} = 1$$

$$(sx)_{1\downarrow 1} = s(x)_{1\downarrow 1}$$

$$(sx)_{2\downarrow 1} = s^2 (x)_{2\downarrow 1} + (s)_{2\downarrow 1} (x)_{1\downarrow 1}$$

$$(sx)_{3\downarrow 1} = s^3 (x)_{3\downarrow 1} + 3s(s)_{2\downarrow 1}(x)_{2\downarrow 1} + (s)_{3\downarrow 1}(x)_{1\downarrow 1}$$

and generally

$$(sx)_{n\downarrow 1} = \sum_{k=0}^{n} C(n,k;s)(x)_{k\downarrow 1} \qquad n \in \mathbb{N}_0. \qquad \text{(A.3.8)}$$

In particular, for $s = -1$ and introducing the coefficient $L(n, k) = C(n, k; -1)$ we deduce the expression

$$(-x)_{n\downarrow 1} = \sum_{k=0}^{n} L(n, k)(x)_{k\downarrow 1} \qquad n \in \mathbb{N}_0. \tag{A.3.9}$$

Further, since $(-x)_{n\downarrow 1} = (-1)^n (x)_{n\uparrow 1}$ and setting $|L(n, k)| = (-1)^n L(n, k)$, we get

$$(x)_{n\uparrow 1} = \sum_{k=0}^{n} |L(n, k)|(x)_{k\downarrow 1} \qquad n \in \mathbb{N}_0. \tag{A.3.10}$$

Based on expansions (A.3.8), (A.3.9) and (A.3.10), the following definition is introduced.

Definition A.3.3. *(cfr. Charalambides [17]) The coefficient $C(n, k; s)$ of the k-th order factorial of x in the expansion of the n-th order generalized factorial of x, with scale parameter s, is called the generalized factorial coefficient. In particular, the coefficient $L(n, k) = C(n, k; -1)$ and $|L(n, k)| = (-1)^n L(n, k)$ are called Lah and signless or absoluted Lah numbers, respectively*

n	$C(n,1;s)$	$C(n,2;s)$	$C(n,3;s)$	$C(n,4;s)$	$C(n,5;s)$
1	$(s)_{1\downarrow 1}$	-	-	-	-
2	$(s)_{2\downarrow 1}$	s^2	-	-	-
3	$(s)_{3\downarrow 1}$	$3(s)_{2\downarrow 1}s$	s^3	-	-
4	$(s)_{4\downarrow 1}$	$7(s)_{3\downarrow 1}s + 3(s)_{2\downarrow 1}s$	$6(s)_{2\downarrow 1}s^2$	s^4	-
5	$(s)_{5\downarrow 1}$	$15(s)_{4\downarrow 1}s + 20(s)_{3\downarrow 1}s$	$25(s)_{3\downarrow 1}s^2 + 15(s)_{2\downarrow 1}s^2$	$10(s)_{2\downarrow 1}s^3$	s^5

Table A.4: Some generalized factorial coefficients $C(n, k; s)$

A useful extension of the generalized factorial coefficients is provided by the coefficients of the expansion of the non-central generalized factorials into (usual) factorials. Specifically, let

$$(sx + r)_{n\downarrow 1} = \sum_{k=0}^{n} C(n, k; s, r)(x)_{k\downarrow 1} \qquad n \in \mathbb{N}_0. \tag{A.3.11}$$

In particular, for $s = -1$ and introducing $L(n, k; r) = C(n, k; -1, r)$, we get

$$(-x + r)_{n\downarrow 1} = \sum_{k=0}^{n} L(n, k; r)(x)_{k\downarrow 1} \qquad n \in \mathbb{N}_0.$$

Since $(-t + r)_{n\downarrow 1} = (-1)^n (t - r + n - 1)_{n\downarrow 1}$ this expression may be written as

$$(x - r)_{n\uparrow 1} = \sum_{k=0}^{n} |L(n, k; r)|(x)_{k\downarrow 1} \qquad n \in \mathbb{N}_0$$

where $|L(n, k; r)| = (-1)^n L(n, k; r)$. Then the following definition is introduced

Definition A.3.4. *(cfr. Charalambides [17]) The coefficient $C(n,k;s,r)$ of the k-th order factorial of x in the expansion of the n-th order non-central generalized factorial of x, with scale parameter s and non-centrality parameter r, is called the non-central generalized factorial coefficient. In particular, the coefficient $L(n,k;r) = C(n,k;-1,r)$ and $|L(n,k;r)| = (-1)^n L(n,k;r)$ are called non-central Lah and non-central signless or absolute Lah numbers, respectively.*

Clearly, this definition implies $C(0,0;s,r) = 1$, $C(n,0;s,r) = (r)_{n\downarrow 1}$ for $n \in \mathbb{N}$ and $C(n,k;s,r) = 0$ for $k > n$. Further, expansion (A.3.11) entails that the non-central generalized factorial coefficients are differences of non-central generalized factorials. Finally, note that for $r = 0$, the non-central generalized factorial coefficients reduces to the central generalized factorial coefficients. For $r \neq 0$, these coefficients may be expressed in terms of the central generalized factorial coefficients. Specifically, expanding the non-central generalized factorial $(sx + s\rho)_{n\downarrow 1} = (s(x+\rho))_{n\downarrow 1}$ into factorials of $u = x + \rho$, using (A.3.8), and then expanding the factorials of $u = x + \rho$ into factorials of t, using Vandermonde's formula we get the expression

$$
\begin{aligned}
(sx + s\rho)_{n\downarrow 1} &= \sum_{j=0}^{n} C(n,j;s)(x+\rho)_{j\downarrow 1} = \sum_{j=0}^{n} C(n,j;s) \sum_{k=0}^{j} \binom{j}{k} (x)_{k\downarrow 1} (\rho)_{(j-k)\downarrow 1} \\
&= \sum_{k=0}^{n} \left(\sum_{j=k}^{n} \binom{j}{k} (\rho)_{(j-k)\downarrow 1} C(n,j;s) \right) (x)_{k\downarrow 1}
\end{aligned}
$$

which, compared to (A.3.11), yelds the expression

$$
C(n,k;s,s\rho) = \sum_{j=k}^{n} \binom{j}{k} (\rho)_{(j-k)\downarrow 1} C(n,j;s).
$$

Also, expanding the non-central generalized factorials $(sx + r)_{n\downarrow 1}$ into generalized factorials $(sx)_{j\downarrow 1}$, $j = 0, \ldots, n$, using Vandermonde's formula (A.3.7), and then expanding the generalized factorials $(sx)_{j\downarrow 1}$, $j = 0, \ldots, n$, into factorials of x, using (A.3.8), we deduce the expansion

$$
\begin{aligned}
(sx + r)_{n\downarrow 1} &= \sum_{j=0}^{n} \binom{n}{j} (sx)_{j\downarrow 1} (r)_{(n-j)\downarrow 1} = \sum_{j=0}^{n} \binom{n}{j} (r)_{(n-j)\downarrow 1} \sum_{k=0}^{j} C(j,k;s)(x)_{k\downarrow 1} \\
&= \sum_{k=0}^{n} \left(\sum_{j=k}^{n} \binom{n}{j} (r)_{(n-j)\downarrow 1} C(j,k;s) \right) (x)_{j\downarrow 1}
\end{aligned}
$$

which, compared to (A.3.11), implies the expression

$$
C(n,k;s,r) = \sum_{j=k}^{n} \binom{n}{j} (r)_{(n-j)\downarrow 1} C(j,k;s).
$$

The non-central generalized factorial coefficient $C(n, k; s, \rho s - r)$ is a polynomial in s of degree n. Specifically, the following theorem is derived

Theorem A.3.1. *(cfr. Charalambides [17]) The non-central generalized factorial coefficient $C(n, k; s, \rho s - r)$ is a polynomial in s of degree n, the coefficient of the general term of which being a product of non-central Stirling numbers of the first and second kind*

$$C(n, k; s, \rho s - r) = \sum_{j=k}^{n} s(n, j; r) S(j, k; \rho) s^j. \tag{A.3.12}$$

In general, an explicit expression of the non-central genealized factorial coefficient can be given. The non-central generalized factorial coefficient $C(n, k; s, r)$, $k = 0, \ldots, n$, $n \in \mathbb{N}_0$, is given by the sum

$$C(n, k; s, r) = \frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} (sj + r)_{n\downarrow 1}.$$

Expression of the generalized factorial coefficient $C(n, k; s)$ as multiple sums over all compositions as well as over all partitions of n into k parts are given in the following theorem.

Theorem A.3.2. *(cfr. Charalambides [17]) The generalized factorial coefficient $C(n, k; s)$, $k = 0, \ldots, n$, $n \in \mathbb{N}_0$, is given by*

$$C(n, k; s) = \frac{n!}{k!} \sum \binom{s}{j_1} \binom{s}{j_2} \cdots \binom{s}{j_k}$$

where the summation is extended over all compositions on n into k parts, that is over all positive integer solutions of the equation $j_1 + j_2 + \cdots + j_k = n$. Alternatively,

$$C(n, k; s) = \sum \frac{n!}{k_1! k_2! \cdots k_n!} \binom{s}{1}^{k_1} \binom{s}{2}^{k_2} \cdots \binom{s}{n}^{k_n}$$

where the summation is extended over all partitions of n into k parts, that is over all non-negative integer solutions of the equations

$$k_1 + 2k_2 + \cdots + nk_n = n, \qquad k_1 + k_2 + \cdots + k_n = k.$$

Limiting expressions as $s \to 0$ and $s \to +\infty$ and an orthogonality relation for the non-central generalized factorial coefficient are deduced in the following theorems.

Theorem A.3.3. *(cfr. Charalambides [17]) Let $C(n, k; s, r)$ be the non-central generalized factorial coefficient. Then*

$$\lim_{s \to 0} s^{-k} C(n, k; s, r) = s(n, k; -r) \tag{A.3.13}$$

and

$$\lim_{s\to+\infty} s^{-n}C(n,k;s,\rho) = S(n,k;\rho)$$

where $s(n,k;r)$ *and* $S(n,k;\rho)$ *are the non-central Stirling numbers of the first and the second kind, respectively.*

Theorem A.3.4. *(cfr. Charalambides [17]) The non-central generalized factorial coefficients* $C(n,k;s,r)$*,* $k=0,\dots,n$*,* $n\in\mathbb{N}_0$*, satisfy the relation*

$$\sum_{j=k}^{n} C(n,j;s_1,r_1)C(j,k;s_2,r_2) = C(n,k;s_1s_2,s_1r_2+r_1).$$

In particular, they satisfy the orthogonality relation

$$\sum_{j=k}^{n} C(n,j;s,r)C(j,k;s^{-1},-rs^{-1}) = \delta_{n,k}$$

where $\delta_{n,k}=1$*, if* $k=n$ *and* $\delta_{n,k}=0$*, if* $k\neq n$ *is the Knonecker delta.*

A triangular recurrence relation for the non-central generalized factorial coefficients is derived in the following theorem.

Theorem A.3.5. *(cfr. Charalambides [17]) The non-central generalized factorial coefficients* $C(n,k;s,r)$*,* $k=0,\dots,n$*,* $n\in\mathbb{N}_0$*, satisfy the triangular recurrence relation*

$$C(n+1,k;s,r) = (sk+r-n)C(n,k;s,r) + sC(n,k-1;s,r) \tag{A.3.14}$$

for $k=1,\dots,n+1$*,* $n\in\mathbb{N}_0$*, with initial conditions*

$$C(0,0;s,r)=1, \qquad C(n,0;s,r)=(r)_{\downarrow 1},\ n\in\mathbb{N}, \qquad C(n,k;s,r)=0,\ k>n.$$

In the following theorem a vertical recurrence relation for the generalized factorial coefficients is deduced.

Theorem A.3.6. *(cfr. Charalambides [17]) The generalized factorial coefficients* $C(n,k;s)$*,* $k=0,\dots,n$*,* $n\in\mathbb{N}_0$*,with* $C(0,0;s)=1$*, satisfy the vertical recurrence relation*

$$C(n+1,k+1;s) = \sum_{j=k}^{n}\binom{n}{j}(s)_{(n-j+1)\downarrow 1}C(j,k;s).$$

In general, the generalized factorial coefficients $C(n,k;s)$ can be tabulated by using recurrence relation (A.3.14) for $r=0$, and its initial conditions. The non-central generalized factorial coefficients $C(n,k;s,r)$, $k=0,\dots,n$, $n\in\mathbb{N}_0$, are defined as the coefficients of the factorials in the expansion of the non-central generalized factorial, with scale parameter s and non-centrality parameter r.

B

Appendix

This appendix reviews some definitions and some results on the field of the random partitons. In particular, after introducing some preliminar concepts we remind the definition of partially exchangeable random partition.

B.1 Partially exchangeable random partitions

For any $n \in \mathbb{N}$, a partiton of $[n]$ is an unordered collection of disjoint non-empty subsets of $[n]$, say $\{A_1, \ldots, A_k\}$, with $\cup_{1 \leq i \leq k} A_i = [n]$. The A_i are called classes of the partition. Given a partition $\{A_1, \ldots, A_k\}$ on $[n]$ for $m < n$ the restriction of $\{A_1, \ldots, A_k\}$ to $[m]$ is the partiton of $[m]$ whose classes are the non-empty members of $\{A_1 \cap [m], \ldots, A_k \cap [m]\}$. Let $\mathcal{P}^k_{[n]}$ denote the set of partitions of the set $[n]$ into k blocks, and let $\mathcal{P}_{[n]} := \cup_{k=1}^n \mathcal{P}^k_{[n]}$, the set of all partitions of $[n]$. To be definite, the blocks A_i of a partition of $[n]$ are assumed to be listed in order of appearance, meaning the order of their least elements, except if otherwise specified. The sequence $(|A_1|, \ldots, |A_k|)$ of sizes of blocks of a partition of $[n]$ defines a composition of $[n]$, that is a sequence of positive integers with sum n. Let $\mathcal{C}_n$ denote the set of all compositions of n. An integer composition is an element of $\cup_{n \geq 1} \mathcal{C}_n$.

A random partition of $[n]$ is a r.v. Π_n taking values in $\mathcal{P}_{[n]}$. A random partiton of $\mathbb{N}$ is a sequence $\Pi := \{\Pi_n, n \geq 1\}$ of random partiton of $[n]$ defined on a common probability space, such that for $m < n$, the restriction of Π_n to $[m]$ is Π_m. Permutation of $[n]$ act in a natural way on partions of $[n]$ and on distributions of a random partiton of of $[n]$. Following Kingman [113] and Aldous [1], Π_n is exchangeable if the distribution of Π_n is invariant under the action of all such permutations. And Π is exchangeable if Π_n is exchangeable for every n.

The multiset $\{|A_1|, \ldots, |A_k|\}$ of unordered sizes of blocks of a partition Π_n of $[n]$ defines a partition of n, customarily encoded by one of the following:

i) the composition of n defined by the decreasing arrangement of block sizes of Π_n, say $(n_1^\downarrow, \ldots, n_k^\downarrow)$ where $n_i^\downarrow$ is the size of the i-th largest block Π_n and k is the number of blocks of the partition of $[n]$;

ii) the infinite decreasing sequence of non-negative integers $\{n_i^{\downarrow}, i \geq 1\}$ defined by appending an infinite string of zeros to $(n_1^{\downarrow}, \ldots, n_k^{\downarrow})$, so $n_i^{\downarrow}$ is the size of the i-th largest block of the partiton of $[n]$ if $k \geq i$, and 0 otherwise;

iii) the non-negative integer counts $(m_j, 1 \leq j \leq n)$, where $m_j := \{i : n_i^{\downarrow} = j\}$ for $j = 1, \ldots, n$, i.e. the number of blocks of the partiton of [n] of size j, with

$$\sum_j m_j = k, \qquad \sum_j j m_j = n.$$

Thus the set $\mathcal{P}_n$ of all partitions of n is bijectively identified with each of the following three sets of sequences of non-negative integers

$$\bigcup_{k=1}^{n} \left\{ (n_j)_{1 \leq j \leq k} : n_1 \geq n_2 \geq \cdots \geq n_k \geq 1, \sum_j n_j = n \right\}$$

or

$$\left\{ (n_j)_{1 \leq j \leq \infty} : n_1 \geq n_2 \geq \cdots \geq 0, \sum_j n_j = n \right\}$$

or

$$\left\{ (m_i)_{1 \leq i \leq n} : \sum_i i m_i = n \right\}$$

where $m_i = \sum_j \mathbb{1}_{\{n_j = i\}}$. Arandom partition of n is a r.v. π_n taking values in the set of all partitions of n. In Kingman [111] and in Kingman [112], moving from application in genetics, the concept of partiton structure has been developed. In particular, a partition stucture is a sequence of distributions $\{P_i, i \geq 1\}$ for $\{\pi_i, i \geq 1\}$ which is consistent in the following sense: if n objects are partitioned into classes with sizes given by π_n, and an object is delated uniformly at random, independently of π_n, the partition of the $n-1$ remaning objects has class sizes distributed according to P_{n-1}. As shown by Kingman [112], $\{P_n, n \geq 1\}$ is a partiton structure if and only if there exists an exchangeable random partiton Π of $\mathbb{N}$, such that $\{P_n, n \geq 1\}$ is the distribution of the partiton of n induced by the class sizes of Π_n.

For a sequence of r.v.s $\{X_i, i \geq 1\}$, let $\Pi(X_1, X_2, \ldots)$ be the random partition of $\mathbb{N}$ defined by the eqivalence classes for the random equivalence relation $i \sim j \Leftrightarrow X_i = X_j$. According to Kingman's representation theorem every exchangeable random partiton Π of $\mathbb{N}$ has the same distribution as $\Pi(X_1, X_2, \ldots)$ where $\{X_i, i \geq 1\}$ are conditionally i.i.d. according to P_∞ given some random probability distribution P_∞. The distribution P_n of the class sizes of Π_n is determined by the joint distribution of the sizes of the ranked atoms of P_∞, denoted

$$P_1^\downarrow \geq P_2^\downarrow \geq \cdots \geq 0$$

where $P_i^\downarrow = 0$ if P_∞ has fewer thatn i atoms. Moreover such $P_i^\downarrow$ can be recovered from Π as

$$P_i^\downarrow = \lim_{n\to+\infty} \frac{N_i^\downarrow}{n} \qquad \text{a.s.}, i \geq 1 \tag{B.1.1}$$

where $N_i^\downarrow$ is the size of the i-th largest class in Π_n.

The joint distribution of the limiting ranked proportions $P_i^\downarrow$ turns out to be rather complicated, even for the simplest partition structures, such as those corresponding to the Ewens's sampling formula, when the joint distribution of the $P_i^\downarrow$ is the Poisson-Dirichlet distribution (see Ignatov [89], Watterson [195] and Kingman [110]). The expression for the distribution P_n on the partiton of $[n]$ in terms of the joint distribution of the $P_i^\downarrow$, given by formulae (2.10) and (5.1) in Kingman [112], involves infinite sums of exptectations of products of the $P_i^\downarrow$, which is not easy to evaluate. It is known (see Donelly [26], Ewens [44], Hoppe [87] and Hoppe [88]) that in the case corresponding to the Ewens' sampling formula, there is a much simpler description of the joint distribution of the sequence $\{P_i, i \geq 1\}$, obtained by presenting the ranked sequence $\{P_i^\downarrow, i \geq 1\}$ in the random order in which the corresponding classes appear in the random partiton Π. In other word, write

$$\Pi = \{\mathscr{A}_i, i \geq 1\}$$

where $\mathscr{A}_i$ is the random subset of $\mathbb{N}$ defined as the i-th class of of Π to appear.That is to say $\mathscr{A}_1$ is the class containing 1, $\mathscr{A}_2$ is the class containing the first element of $\mathbb{N} \cap \mathscr{A}_1^c$, and so on. For convenience, let $\mathscr{A}_i = \emptyset$ if Π has fewer than i classes. Then, P_i is defined to be the long run relative frequency of $\mathscr{A}_i$

$$P_i = \lim_{n\to+\infty} \frac{\sharp(\mathscr{A}_i \cap [n])}{n} \qquad \text{a.s.}, i \geq 1. \tag{B.1.2}$$

The $P_i^\downarrow$ are obtained by ranking the P_i, and the existence of either collection of limits (B.1.1) or (B.1.2) follows easily from the other. See Lemma 11.8 in Aldous [1], which implies also that if $\sum_i P_i^\downarrow = 1$, then $P_i, i \geq 1$ is a sized biased random permutation of the ranked sequence $P_i^\downarrow, i \geq 1$ as studied by Donelly and Joyce [28] and by Pitman [151].

Pitman [149] provide a generalization of the Kingman's representation theorem. In particular the starting point is the definition of partially exchangeable random partition.

Definition B.1.1. *(cfr. Pitman [149]) Let $\mathbb{N}^* := \cup_{k\geq 1}\mathbb{N}^k$, the set of finite sequences of positive integers. Call a random partition Π_n of $[n]$ partially exchangeable if for every* *partition* $\{A_1, \dots, A_k\}$ *of* $[n]$ *where* $A_1, \dots, A_k$ *are in order of appearance, i.e.* $1 \in A_1$*, and for each* $2 \le i \le k$ *the first element of* $N_n \cap (A_1 \cup \dots, \cup A_{i-1})^c$ *belongs to* A_i

$$\mathbb{P}(\Pi_n = \{A_1, \dots, A_k\}) = p_k^{(n)}(|A_1|, \dots, |A_k|) \tag{B.1.3}$$

for some function $p_k^{(n)}(n_1, \dots, n_k)$ *defined for* $(n_1, \dots, n_k) \in \mathbb{N}^*$ *with* $\sum_{1 \le i \le k} n_i = n$*. Then call* $p_k^{(n)}(n_1, \dots, n_k)$ *a partially exchangeable partition probability function (PEPPF).*

As particular case of Definition B.1.1 we recover the definition of exchangeable random partition. In particular a random partition of $[n]$ (or of $\mathbb{N}$) is exchangeable if and only if it is partially exchangeable with PEPPF $p_k^{(n)}(n_1, \dots, n_k)$ which is a symmetric function of its arguments, i.e.

$$p_k^{(n)}(n_1, \dots, n_k) = p_k^{(n)}(n_{\sigma(1)}, \dots, n_{\sigma(k)})$$

for every permutation σ of $[k]$, $k = 2, 3, \dots$. When Π is exchangeable, call the symmetric PEPF derived from Π an exchangeable partition probability function (EPPF). The main results given by Pitman [149] is the following representation theorem which generalizes to the partially exchangeable random partitions the representation the Kingman's representation theorem. (see also[1]).

Theorem B.1.1. *(cfr. Pitman [149]) Let* $\Pi = \{\mathscr{A}_i, i \ge 1\}$ *be a random partition of* $\mathbb{N}$ *with* $\mathscr{A}_i$ *the i-th class of* Π *to appear. Let* Π_n *be the restriction of* Π *to* $[n]$*. The following conditions are equivalent*

i) Π *is partially exchangeable;*

ii) there is a sequence of r.v.'s $\{P_i, i \ge 1\}$ *with* $P_i \ge 0$ *and* $\sum_i P_i \le 1$ *such that the conditional distribution of* Π *given the whole sequence* $\{P_i, i \ge 1\}$ *is as follows: for any* $n \in \mathbb{N}$*, conditionally given* $\{P_i, i \ge 1\}$ *and* $\Pi_n = \{A_1, \dots, A_k\}$*, where the* A_i *are in order of appearence,* Π_{n+1} *is an extension of* Π_n *in which* $n+1$ *attaches to class* A_i *with probability* P_i *for* $1 \le i \le k$ *and forms a new class with probability* $1 - \sum_{1 \le j \le k} P_j$*.*

If Π *is partially exchangeable then* P_i *in (ii) are a.s. unique and equal to almost sure limiting relative frequencies of the classes* $\mathscr{A}_i$ *as in* (B.1.2)*.*

Theorem B.1.1 provides a characterization for partially exchangeable random partitions. In particular, a random partition of $\mathbb{N}$ is partially exchangeable if and only if it admits such relative frequencies $\{P_i, i \ge 1\}$ and shares this conditional distribution given the $\{P_i, i \ge 1\}$. The following corollary is an immediate consequence of Theorem B.1.1

Corollary B.1.1. *(cfr. Pitman [149]) The formula*

$$p_k^{(n)}(n_1,\dots,n_k) = \mathbb{E}\left[\left(\prod_{i=1}^{k} P_i^{n_i-1}\right)\prod_{i=1}^{k-1}\left(1-\sum_{j=1}^{i} P_j\right)\right] \tag{B.1.4}$$

sets up a one to one correspondence between PEPPF $p_k^{(n)} : \mathbb{N}^* \to [0,1]$ and the joint distributions for a sequence of r.v.'s $\{P_i, i \geq 1\}$ with $P_i \geq 0$ and $\sum_i P_i \leq 1$.

From Theorem B.1.1 all exchangeable random partitions Π of N share a common conditional distribution given $\{P_i, i \geq 1\}$ defined in ii) of Theorem B.1.1; moreover a random partition of N is partially exchangeable if and only if it admits relative frequencies $\{P_i, i \geq 1\}$ and shares this conditional distribution given the $\{P_i, i \geq 1\}$. Corollary B.1.1 emphasizes the correspondence between $\{P_i, i \geq 1\}$ and $p_k^{(n)}(n_1,\dots,n_k)$: given the PEPPF $p(n_1,\dots,n_k)$ of a partially exchangeable random partition Π of $\mathbb{N}$, the P_i are recovered as the limiting relative frequencies of the classes of Π in order of appearance. And given a distribution for $\{P_i, i \geq 1\}$, a partially exchangeable random partition is created by ii) of Theorem B.1.1. The following corollary underlines the connection between partially exchangeable random partitions and exchangeable random partitions. See also Pitman [151] for a version of the following corollary based on the theory of random discrete distribution $\{P_i, i \geq 1\}$ invariant under size-biased random permutation.

Corollary B.1.2. *(cfr. Pitman [149]) Let $\{P_i, i \geq 1\}$ be a sequence of r.v.s such that $P_i \geq 0$ and $\sum_i P_i \leq 1$ a.s. The following statements are equivalent:*

- *i) there exists an exchangeable random partition Π of $\mathbb{N}$ whose sequence of limiting relative frequencies of classes, in order of appearance, has the same distribution as $\{P_i, i \geq 1\}$;*
- *ii) for each $k \geq 2$ the function $p_k^{(n)} : \mathbb{N}^k \to [0,1]$ defined by (B.1.4) is a symmetric function of $(n_1,\dots,n_k)$;*
- *iii) fore each $k \geq 2$ the measure G_k on $\mathbb{R}^k$ defined by*

$$G_k(dp_1,\dots,dp_k) = \mathbb{P}(P_1 \in dp_1,\dots,P_k \in dp_k)\prod_{i=1}^{k-1}\left(1-\sum_{j=1}^{i} p_j\right)$$

is symmetric with respect to permutation of the coordinates in $\mathbb{R}^k$.

Then $p_k^{(n)}(n_1,\dots,n_k)$ defined by (B.1.4) is the EPPF of Π.

Let $(M_1, \dots, M_n)$ be a r.v. where M_j is the number of classes of Π_n of size j. Then, for any vector of non-negative integer counts $(m_1, \dots, m_n)$

$$P_n(m_1, \dots, m_n) := \mathbb{P}((M_1, \dots, M_n) = (m_1, \dots, m_n)) = \sharp(m_1, \dots, m_n)\tilde{p}_k^{(n)}(m_1, \dots, m_n) \tag{B.1.5}$$

where

$$\sharp(m_1, \dots, m_n) := \frac{n!}{\prod_{j=1}^{n}(j!)^{m_j} m_j!}$$

where $\tilde{p}_k^{(n)}(m_1, \dots, m_n)$ is the common value of the symmetric function $p_k^{(n)}(n_1, \dots, n_k)$ for all $n_1, \dots, n_k$ with $\sum_{1 \le i \le k} n_i = n$ and $m_j = \sharp\{i : n_i = j\}$ for $1 \le j \le n$. Remind that $\mathbb{N}^* := \cup_{k \ge 1} \mathbb{N}^k$ and define the set

$$\mathcal{D}_{k,n} := \left\{ (n_1, \dots, n_k) \in \mathbb{N}^* : \sum_{i=1}^{k} = n \right\}.$$

It follows from Definition B.1.1 and from (B.1.3) a one to one correspondence between the distribution of a partially exchangeable random partition Π_n of $[n]$, and non-negative functions $p_k^{(n)} : \mathcal{D}_{k,n} \to [0, 1]$ such that

$$\sum_{(n_1, \dots, n_k) \in \mathcal{D}_{k,n}} \sharp(n_1, \dots, n_k) p_k^{(n)}(n_1, \dots, n_k) = 1$$

where

$$\sharp(n_1, \dots, n_k) := \frac{n!}{n_k(n_k - n_{k-1}) \cdots (n_k + \cdots + n_1) \prod_{i=1}^{k}(n_i - 1)!}$$

is the number of partitions of $[n]$ whose class sizes in order of appearance are given by $(n_1, \dots, n_k$. Let $(N_1, \dots, N_k)$ be the random element of $\mathcal{D}_{k,n}$ representing the class sizes of Π_n in order of appearance. Then $(N_1, \dots, N_k)$ is sufficient statisitc for distributions of partially exchangeable random partition Π_n. In other words, Π_n is partially exchangeable if and only if, given $(N_1, \dots, N_k) = (n_1, \dots, n_k)$ for every $(n_1, \dots, n_k) \in \mathcal{D}_{k,n}$ the partition Π_n is uniformly distributed over the number of partitions of $[n]$ whose class sizes in order of appearance given by $(n_1, \dots, n_k$. The corresponding description of exchangeable random partitions on $[n]$ with $(N_1, \dots, N_k)$ replaced by the decresing rearrangement of $(N_1, \dots, N_k)$ which encodes the induced partiton of n is given by Aldous [1]. In particular, the distribution of $(N_1, \dots, N_k)$ for a partially exchangeable random partiton Π_n is related to the PEPPF of Π_n by

$$\mathbb{P}((N_1, \dots, N_k) = (n_1, \dots, n_k)) = \sharp(n_1, \dots, n_k) p_k^{(n)}(n_1, \dots, n_k).$$

Assuming Π_n is a partially exchangeable random partition, it can be seen that Π_n is echangeable if and only if $(N_1, \dots, N_k)$ is a size-biased random ordering of the partition of n (see Pitman [151]).

Proposition B.1.1. *(cfr. Pitman [149]) For* $1 \leq m \leq n$ *let* Π_m *be the restriction to* $[m]$ *of a partially exchangeable partition of* $[n]$ *with PEPPF* $p_k^{(n)}(n_1, \dots, n_k)$ *defined for* $(n_1, \dots, n_k) \in \mathcal{D}_{k,n}$*. Then*

i) Π_m *is partially exchangeable, with PEPPF* $p_k^{(m)}(n_1, \dots, n_k)$ *defined for* $(n_1, \dots, n_k) \in \mathcal{D}_{k,n}$ *by repeated application for* $m = n-1, n-2, \dots, 1$ *of the consistency relation:*

$$p_k^{(m)}(n_1, \dots, n_k) = \sum_{j=1}^{k+1} p(\mathbf{n}^{j+}) \tag{B.1.6}$$

where

$$p(\mathbf{n}^{j+}) := \begin{cases} p_k^{(m+1)}(n_1, \dots, n_{j-1}, n_j + 1, n_{j+1}, \dots, n_k) & 1 \leq j \leq k \\ \\ p_{k+1}^{(m+1)}(n_1, \dots, n_k, 1) & j = k+1 \end{cases}$$

i.e. $\mathbf{n}^{j+}$ *is derived from* $(n_1, \dots, n_k) \in \mathcal{D}_{k,m}$ *by incrementing* n_j *by* 1 *if* $1 \leq j \leq k$*, and by appending a* 1 *to* $(n_1, \dots, n_k)$ *ad place* $k+1$ *if* $j = k+1$*;*

ii) $(N_1, \dots, N_n)$ *is a Markov chain with transition probabilities*

$$\mathbb{P}(N_{m+1} = \mathbf{n}^{j+} | (N_1, \dots, N_k) = (n_1, \dots, n_k)) = \frac{p(\mathbf{n}^{j+})}{p_k^{(m)}(n_1, \dots, n_k)} \qquad j = 1, \dots, k+1$$

for $p : \bigcup_{m=1}^{n} \mathcal{D}_{k,m} \to [0, 1]$ *defined as in i).*

In the exchangeable case, the EPPF and the distribution of the correspondig partition of n are related by (B.1.5). The above consistency relation then becomes the expression in terms of EPPF of Kingman's notion of consistency of partition of n.

Corollary B.1.3. *(cfr. Pitman [149]) A function* $p : \mathcal{D}_{k,n} \to [0, +\infty)$ *is a PEPPF if and only if* $p : \cup_{1 \leq m \leq n} \mathcal{D}_{k,m} \to [0, +\infty)$ *defined by repeated application of* (B.1.6) *is such that* $p(1) = 1$.

To conclude the review on partially exchangeable random partition we remind two simple constructions that is easily seen to yield the most general partially exchangeable partition of $[n]$ and of $\mathbb{N}$, respectively. Let $\mathscr{A}_1, \dots, \mathscr{A}_{K_n}$ denote the random subset of $[n]$ defined by the classes of Π_n in order of appearance. Let N_1, the size of $\mathscr{A}_1$, have distribution

$$\mathbb{P}(N_1 = n_1) = P(n_1) \qquad 1 \leq n_1 \leq n$$

where P is some arbitrary probability distribution on $[n]$. Given $N_1 = n_1$, let $\mathscr{A}_1$ consist of 1 and a uniformly distributed random subset of $n_1 - 1$ elements of $\{2, \dots, n\}$. Inductively:

given that $\mathscr{A}_1, \dots, \mathscr{A}_i$ have been defined, with $N_j = n_j$ for $1 \le j \le i$, such that $\sum_{1 \le j \le i} n_j < n$, let N_{i+1} have distribution

$$\mathbb{P}(N_{i+1} = n_{i+1} | \mathscr{A}_1, \dots, \mathscr{A}_i) = P(n_{i+1} | n_1, \dots, n_i)$$

where P is some arbitrary probability distribution on $[1 - \sum_{1 \le j \le i} n_j]$. And given $\mathscr{A}_1, \dots, \mathscr{A}_i$ and $N_{i+1} = n_{i+1}$, let $\mathscr{A}_{i+1}$ comprise the first element of $[n] \cap (\cup_{1 \le j \le i} \mathscr{A}_j)^c$ together with a uniformly distributed random subset of $n_{i+1} - 1$ elements of the remaning $n - \sum_{1 \le j \le i} n_j - 1$ elements of $[n]$. The random partition Π_n so constructed is partially exchangeable, with PEPPF

$$p_k^{(n)}(n_1, \dots, n_k) = (\sharp(n_1, \dots, n_k))^{-1} P(n_1) P(n_2 | n_1) \cdots P(n_k | n_1, \dots, n_{k-1}).$$

The corresponding construction yielding to the most general partially exchangeable partition of $\mathbb{N}$ is the following. Given an arbitrary joint distribution for a sequence of r.v.s $\{W_i, i \ge 1\}$ with values $W_i \in [0,1]$, define a random partition Π of $\mathbb{N}$ into random subsets $\mathscr{A}_1, \mathscr{A}_2, \dots$ as follows. Let

$$\{X_{n,i}, n \ge 1, i \ge 1\}$$

be indicator variables that are conditionally independent given $\{W_i, i \ge 1\}$ with

$$\mathbb{P}(X_{n,i} = 1 | W_1, W_2, \dots) = W_i.$$

Let $\mathscr{A}_1 = \{1\} \cup n \in \mathbb{N} : X_{n,1} = 1$. Inductively: for $i \ge 1$ let $\mathscr{C}_i = \mathbb{N} \cap (\mathscr{A}_1 \cup \dots \cup \mathscr{A}_i)^c$. Given $\mathscr{C}_i$ is non-empty (or, what is the same, $\prod_{1 \le j \le i} (1 - W_j) > 0$), let

$$\mathscr{A}_{i+1} = \{\min\{\mathscr{C}_i\}\} \cup \{n \in \mathscr{C}_i : X_{n,i+1} = 1\}.$$

It is easily seen directly that Π is partially exchangeable. By construction the $\mathscr{A}_i$ are in order of appearance with limiting frequencies $P_i = W_i \prod_{1 \le j \le i-1} (1 - W_j)$, by repeated application of the law of large numbers. It can be also be seen directly that the conditional distribution of Π given $\{P_i, i \ge 1\}$ is as in ii) of Theorem B.1.1. So the most general possible distribution for a partially exchangeable partition of $\mathbb{N}$ can be obtained by the above construction. As an easy consequence of this construction, there is the followin corollary of Theorem B.1.1

Corollary B.1.4. *(cfr. Pitman [149]) Let* $\Pi = \{\mathscr{A}_i, i \ge 1\}$ *be a partially exchangeable partition on* $\mathbb{N}$, P_i *the almost sure limit as* $n \to +\infty$ *of* $\sharp(\mathscr{A}_i \cap [n])/n$. *For any* $i \in \mathbb{N}_0$, *given* $\{P_i, i \ge 1\}$ *and* $\{(\sharp(\mathscr{A}_i \cap [1]), \dots, \sharp(\mathscr{A}_i \cap [n])), i \ge 1\}$ *with* $\sum_{1 \le j \le i} \sharp(\mathscr{A}_j \cap [n]) < n$, *the r.v.* $\sharp(\mathscr{A}_{i+1} \cap [n]) - 1$ *is distributed according to a Binomial distribution with parameter* $(n - \sum_{1 \le j \le i} (\sharp(\mathscr{A}_j \cap [n]) - 1, W_{i+1})$ *where* $W_{i+1} = P_{i+1}/(1 - \sum_{1 \le j \le i} P_j)$.

C

Appendix

This appendix reviews some definitions and some results in the field of the special functions. In particular, we remind the definition of Lauricella hypergeometric function and the definition of Fox H-function and Meijer G-function.

C.1 Multiple hypergeometric function

The topic of multiple hypergeometric functions was first approached, in a systematic way, by Lauricella [117] at the end of the 19th century. See for example, Exton [47]. He proceeded to define and study four important functions which bear his name and have both multiple series and integral representations. The Lauricella hypergeometric functions have the following multiple series representations

$$F_A^{(n)}(a, b_1, \ldots, b_n; c_1, \ldots, c_n; x_1, \ldots, x_n) \qquad \text{(C.1.1)}$$
$$:= \sum_{(m_1,\ldots,m_n)\in(\mathbb{N}_0)^n} \frac{(a)_{(m_1+\cdots+m_n)\uparrow 1}(b_1)_{m_1\uparrow 1}\cdots(b_n)_{m_n\uparrow 1}x_1^{m_1}\cdots x_n^{m_n}}{(c_1)_{m_1\uparrow 1}\cdots(c_n)_{m_n\uparrow 1}m_1!\cdots m_n!}$$

$$F_B^{(n)}(a_1, \ldots, a_n, b_1, \ldots, b_n; c; x_1, \ldots, x_n) \qquad \text{(C.1.2)}$$
$$:= \sum_{(m_1,\ldots,m_n)\in(\mathbb{N}_0)^n} \frac{(a_1)_{m_1\uparrow 1}\cdots(a_n)_{m_n\uparrow 1}(b_1)_{m_1\uparrow 1}\cdots(b_n)_{m_n\uparrow 1}x_1^{m_1}\cdots x_n^{m_n}}{(c)_{(m_1+\cdots+m_n)\uparrow 1}m_1!\cdots m_n!}$$

$$F_C^{(n)}(a, b; c_1, \ldots, c_n; x_1, \ldots, x_n) \qquad \text{(C.1.3)}$$
$$:= \sum_{(m_1,\ldots,m_n)\in(\mathbb{N}_0)^n} \frac{(a)_{(m_1+\cdots+m_n)\uparrow 1}(b)_{(m_1+\cdots+m_n)\uparrow 1}x_1^{m_1}\cdots x_n^{m_n}}{(c_1)_{m_1\uparrow 1}\cdots(c_n)_{m_n\uparrow 1}m_1!\cdots m_n!}$$

$$F_D^{(n)}(a, b_1, \ldots, b_n; c; x_1, \ldots, x_n) \tag{C.1.4}$$
$$:= \sum_{(m_1,\ldots,m_n)\in(\mathbb{N}_0)^n} \frac{(a)_{(m_1+\cdots+m_n)\uparrow 1}(b_1)_{m_1\uparrow 1}\cdots(b_n)_{m_n\uparrow 1}x_1^{m_1}\cdots x_n^{m_n}}{(c)_{(m_1+\cdots+m_n)\uparrow 1}m_1!\cdots m_n!}.$$

If n, the number of variables, is made equal to two, these four function reduce to the Appell hypergeometric functions F_2, F_3, F_4 and F_1 respectively. If n is made equal to one, all four functions become the Gauss hypergeometric function ${}_2F_1$ which has been the starting point in the definition of $F_A^{(n)}$, $F_B^{(n)}$, $F_C^{(n)}$ and $F_D^{(n)}$. In particular, if we consider the product on n Gauss hypergeometric function

$$\prod_{i=1}^{n} {}_2F_1(a_i, b_i; c_i, x_i) = \prod_{i=1}^{n} \sum_{m_i \geq 0} \frac{(a_i)_{m_i\uparrow 1}(b_i)_{m_i\uparrow 1}}{(c_i)_{m_i\uparrow 1}(1)_{m_i\uparrow 1}} x_i^{m_i}$$

Then $F_A^{(n)}$ can be obtained under the condition $a_1 = a$ and

$$a_i = a + \sum_{j=1}^{i-1} m_j \qquad i = 2, \ldots, n$$

$F_B^{(n)}$ can be obtained under the condition $c_1 = c$ and

$$c_i = c + \sum_{j=1}^{i-1} m_j \qquad i = 2, \ldots, n$$

with $m_0 = 0$, $F_C^{(n)}$ can be obtained under the condition $a_1 = a$, $b_1 = b$ and

$$a_i = a + \sum_{j=1}^{i-1} m_j \qquad i = 2, \ldots, n$$

and

$$b_i = b + \sum_{j=1}^{i-1} m_j \qquad i = 2, \ldots, n$$

$F_D^{(n)}$ can be obtained under the conditions $a_1 = a$, $c_1 = c$ and

$$a_i = a + \sum_{j=1}^{i-1} m_j \qquad i = 2, \ldots, n$$

and

$$b_i = b + \sum_{j=1}^{i-1} m_j \qquad i = 2, \ldots, n.$$

Many other multiple hypergeometric functions exist, as Lauricella [117] himself indicated. He in fact conjectured the existence of fourteen triple hypergeometric functions including $F_A^{(3)}$, $F_B^{(3)}$, $F_C^{(3)}$ and $F_D^{(3)}$, all of which are complete and of the second order.

A number of limiting forms of $F_A^{(n)}$, $F_B^{(n)}$, $F_C^{(n)}$ and $F_D^{(n)}$ exists which may be obtained. We thus have the following limiting functions

$$\begin{aligned}\Psi_2^{(n)}(a; c_1, \ldots, c_n; x_1, \ldots, x_n) &:= \lim_{\varepsilon \to 0} F_A^{(n)}\left(a, \frac{1}{\varepsilon}, \ldots, \frac{1}{\varepsilon}; c_1, \ldots, c_n; \varepsilon x_1, \ldots, \varepsilon x_n\right) \\ &= \sum_{(m_1, \ldots, m_n) \in (\mathbb{N}_0)^n} \frac{(a)_{(m_1+\cdots+m_n)\uparrow 1} x_1^{m_1} \cdots x_n^{m_n}}{(c_1)_{m_1\uparrow 1} \cdots (c_n)_{m_n\uparrow 1} m_1! \cdots m_n!}\end{aligned}$$

$$\begin{aligned}\Phi_2^{(n)}(b_1, \ldots, b_n; c; x_1, \ldots, x_n) &:= \lim_{\varepsilon \to 0} F_D^{(n)}\left(\frac{1}{\varepsilon}, b_1, \ldots, b_n; c; \varepsilon x_1, \ldots, \varepsilon x_n\right) \\ &= \sum_{(m_1, \ldots, m_n) \in (\mathbb{N}_0)^n} \frac{(b_1)_{m_1\uparrow 1} \cdots (b_n)_{m_n\uparrow 1} x_1^{m_1} \cdots x_n^{m_n}}{(c)_{(m_1+\cdots+m_n)\uparrow 1} m_1! \cdots m_n!}\end{aligned}$$

We now remind some integral representations of the Lauricella hypergeometric functions and in particular we consider integral representations of Eulero-type and integral representations of Laplace-type. The single and multiple integrals of Eulero-type

$$\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} = \int_0^1 u^{p-1}(1-u)^{q-1} \tag{C.1.5}$$

and

$$\frac{\Gamma(p_1)\cdots\Gamma(p_n)\Gamma(r)}{\Gamma(p_1+\cdots+p_n+r)} = \int_{\Delta^{(n)}} u_1^{p_1-1}\cdots u_n^{p_n-1}(1-u_1-\cdots-u_n)^{r-1} du_1\cdots du_n \tag{C.1.6}$$

where $\Delta^{(n)}$ is the n-dimensional simplex. Integral (C.1.5) and integral (C.1.6), together with the binomial expansion, readily yield multiple integral representations of the functions $F_A^{(n)}$, $F_B^{(n)}$ and $F_D^{(n)}$. For the function $F_A^{(n)}$

$$\begin{aligned}&F_A^{(n)}(a, b_1, \ldots, b_n; c_1, \ldots, c_n; x_1, \ldots, x_n) \\ &\quad= \frac{\Gamma(c_1)\cdots\Gamma(c_n)}{\Gamma(b_1)\cdots\Gamma(b_n)\Gamma(c_1-b_1)\cdots\Gamma(c_n-b_n)} \\ &\qquad\qquad \times \int_{(0,1)^n} \prod_{i=1}^n u_i^{b_i-1}(1-u_i)^{c_i-b_i-1}\left(1-\sum_{i=1}^n u_i x_i\right)^{-a} du_1\cdots du_n\end{aligned}$$

where $\Re(b_1), \ldots, \Re(b_n)$ are positive and $\Re(c_1-b_1), \ldots, \Re(c_n-b_n)$ are positvie. For the function $F_B^{(n)}$

$$F_B^{(n)}(a_1,\dots,a_n,b_1,\dots,b_n;c;x_1,\dots,x_n)$$
$$= \frac{\Gamma(c)}{\Gamma(a_1)\cdots\Gamma(a_n)\Gamma(c-a_1-\cdots-a_n)}$$
$$\times \int_{\Delta^{(n)}} \prod_{i=1}^{n} u_i^{a_i-1}(1-u_i x_i)^{-b_i}\left(1-\sum_{i=1}^{n} u_i\right)^{c-a_1-\cdots-a_n-1} du_1\cdots du_n$$

where $\Re(b_1),\dots,\Re(b_n)$ are positive and $\Re(c-a_1-\cdots-a_n)$ is positive. A similar multiple integral for the function $F_D^{(n)}$ is now given in which the range of integration is still $\Delta^{(n)}$

$$F_D^{(n)}(a,b_1,\dots,b_n;c;x_1,\dots,x_n)$$
$$= \frac{\Gamma(c)}{\Gamma(b_1)\cdots\Gamma(b_n)\Gamma(c-b_1-\cdots-b_n)}$$
$$\times \int_{\Delta^{(n)}} \prod_{i=1}^{n} u_i^{b_i-1}\left(1-\sum_{i=1}^{n} u_i\right)^{c-b_1-\cdots-b_n-1}\left(1-\sum_{i=1}^{n} u_i x_i\right)^{-a} du_1\cdots du_n$$

where $\Re(b_1),\dots,\Re(b_n)$ are positive and $\Re(c-b_1-\cdots-b_n)$ is positive. Furthermore, the function $F_D^{(n)}$ may also be represented by means of a single integral

$$F_D^{(n)}(a,b_1,\dots,b_n;c;x_1,\dots,x_n) = \frac{\Gamma(a)}{\Gamma(a)\Gamma(c-a)}\int_0^1 u^{a-1}(1-u)^{c-a-1}\prod_{i=1}^{n}(1-ux_i)^{-b_i}du$$

where $\Re(a)$ is positive and $\Re(c-a)$ is positive. No simple integral formula of Eulero-type with an elementary integrand has been obtained for the function $F_C^{(n)}$ which, in many ways, is an exceptional function.

If we consider the integral representation of the Gamma function we have

$$(a)_{m\uparrow 1} = \frac{1}{\Gamma(a)}\int_0^{+\infty} \mathrm{e}^{-t} t^{a+m-1} dt \tag{C.1.7}$$

where $\Re(a)$ is positive and $m \in \mathbb{N}_0$. In particular, if we apply (C.1.7) to the series representations of the Lauricella hypergeometric functions (C.1.1) to (C.1.4), many different examples of integrals of Laplace-type may be obtained. These are of use, for example, in deducing recurrence relations for the Lauricella hypergeometric functions in a simple faschion. We now consider a few of more interesting examples. These results may all readily be established by the simple techniques of interchanging the order of summation and integration. In particular, we have

$$F_A^{(n)}(a,b_1,\dots,b_n;c_1,\dots,c_n;x_1,\dots,x_n) = \frac{1}{\Gamma(a)}\int_0^{+\infty} \mathrm{e}^{-t} t^{a-1}\prod_{i=1}^{n} {}_1F_1(b_i;c_i;x_i t)dt$$

where $\Re(a)$ is positive,

$$F_A^{(n)}(a, b_1, \ldots, b_n; c_1, \ldots, c_n; x_1, \ldots, x_n) \\ = \frac{1}{\Gamma(b_1)\cdots\Gamma(b_n)} \int_{(\mathbb{R}^+)^n} \mathrm{e}^{-\sum_{i=1}^n t_i} \prod_{i=1}^n t_i^{b_i-1} \Psi_2^{(n)}(a; c_1, \ldots, c_n; x_1 t_1, \ldots, x_n t_n) dt_1 \cdots dt_n$$

where $\Re(b_1), \ldots, \Re(b_n)$ are positive,

$$F_B^{(n)}(a_1, \ldots, a_n, b_1, \ldots, b_n; c; x_1, \ldots, x_n) \\ = \frac{1}{\Gamma(a_1)\cdots\Gamma(a_n)} \int_{(\mathbb{R}^+)^n} \mathrm{e}^{-\sum_{i=1}^n t_i} \prod_{i=1}^n t_i^{a_i-1} \Phi_2^{(n)}(b_1, \ldots, b_n; c; x_1 t_1, \ldots, x_n t_n) dt_1 \cdots dt_n$$

where $\Re(a_1), \ldots, \Re(a_n)$ are positive,

$$F_B^{(n)}(a_1, \ldots, a_n, b_1, \ldots, b_n; c; x_1, \ldots, x_n) \\ = \frac{1}{\Gamma(a_1)\cdots\Gamma(a_n)\Gamma(b_1)\cdots\Gamma(b_n)} \int_{(\mathbb{R}^+)^{2n}} \mathrm{e}^{-\sum_{i=1}^n s_i - t_i} \prod_{i=1}^n s_i^{a_i} t_i^{b_i-1} \\ \times {}_0F_1\left(-; c; \sum_{i=1}^n x_i s_i t_i\right) ds_1 \cdots ds_n dt_1 \cdots dt_n$$

where $\Re(a_1), \ldots, \Re(a_n)$ are positive and $\Re(b_1), \ldots, \Re(b_n)$ are positive. In particular, we have the following interesting special cases

$$F_B^{(n)}\left(\frac{a_1}{2}, \ldots, \frac{a_n}{2}, a_1 + \frac{1}{2}, \ldots, a_n + \frac{1}{2}; c; x_1, \ldots, x_n\right) \\ = \frac{1}{\Gamma(a_1)\cdots\Gamma(a_n)} \int_{(\mathbb{R}^+)^n} \mathrm{e}^{-\sum_{i=1}^n t_i} \prod_{i=1}^n t_i^{a_i-1} {}_0F_1\left(-; c; \frac{x_1 t_1^2}{4} + \cdots + \frac{x_n t_n}{4}\right) dt_1 \cdots dt_n$$

where $\Re(a_1), \ldots, \Re(a_n)$ are positive,

$$F_C^{(n)}(a, b; c_1, \ldots, c_n; x_1, \ldots, x_n) \\ = \frac{1}{\Gamma(a)\Gamma(b)} \int_0^{+\infty} \int_0^{+\infty} \mathrm{e}^{-s-t} s^{a-1} t^{b-1} \prod_{i=1}^n {}_0F_1(-; c_1; x_i st) ds dt \tag{C.1.8}$$

where $\Re(a)$ is positive and $\Re(b)$ is positive. In particular, Equation (C.1.8) also has a special case worth mentioning, namely

$$F_C^{(n)}\left(\frac{a}{2}, a + \frac{1}{2}; c_1, \ldots, c_n; x_1, \ldots, x_n\right) = \frac{1}{\Gamma(a)} \int_0^{+\infty} \mathrm{e}^s s^{a-1} \prod_{i=1}^n {}_0F_1\left(-; c_i; \frac{s^2 x_i}{4}\right) ds$$

where $\Re(a)$ is positive. Finally the function $F_D^{(n)}$ has the interesting multiple Laplace integral representation

$$F_D^{(n)}(a, b_1, \ldots, b_n; c; x_1, \ldots, x_n)$$
$$= \frac{1}{\Gamma(b_1)\cdots\Gamma(b_n)} \int_{(\mathbb{R}^+)^n} \mathrm{e}^{-\sum_{i=1}^n t_i} \prod_{i=1}^n t_i^{b_i - 1} {}_1F_1\left(a; c; \sum_{i=1}^n x_i t_i\right) dt_1 \cdots dt_n$$

where $\Re(b_1), \ldots, \Re(b_n)$ are positive, and the single integral representation

$$F_D^{(n)}(a, b_1, \ldots, b_n; c; x_1, \ldots, x_n) = \frac{1}{\Gamma(a)} \int_0^{+\infty} \mathrm{e}^{-t} t^{a-1} \Phi_2^{(n)}(b_1, \ldots, b_n; c; x_1 t_1, \ldots, x_n t_n) dt$$

where $\Re(a)$ is positive. Integral representation of Laplace-type for the Lauricella hypergeometric function can be employed in the discussion of special Fourier transforms and Laplace transforms. Furthermre, they arise frequently in applications involving multiple hypergeometric functions.

Various interesting relations may be obtained as simple consequences of the integral representaion (C.1.5) and (C.1.6). In particular

$$F_A^{(n)}(a, b_1, \ldots, b_n; c_1, \ldots, c_n; x_1, \ldots, x_n)$$
$$= \frac{\Gamma(d)}{\Gamma(a)\Gamma(a-d)} \int_0^1 u^{a-1}(1-u)^{d-a-1} F_A^{(n)}(d, b_1, \ldots, b_n; c_1, \ldots, c_n; ux_1, \ldots, ux_n) du$$

where $\Re(a)$ and $\Re(d-a)$ are positive, and $|x_1| + \cdots + |x_n| < 1$,

$$F_A^{(n)}(a, b_1, \ldots, b_n; c_1, \ldots, c_n; x_1, \ldots, x_n)$$
$$= \frac{\Gamma(c)}{\Gamma(d)\Gamma(c-d)} \int_0^1 u^{d-1}(1-u)^{c-d-1} F_D^{(n)}(a, b_1, \ldots, b_n; d; ux_1, \ldots, ux_n) du$$

where $\Re(d)$ and $\Re(c-d)$ are positive and the moduli of the variables $x_1, \ldots, x_n$ all less than unity. Moreover, we have

$$F_A^{(n)}(a, b_1, \ldots, b_n; c_1, \ldots, c_n; x_1, \ldots, x_n)$$
$$= \frac{\Gamma(d_1)\cdots\Gamma(d_n)}{\Gamma(b_1)\cdots\Gamma(b_n)\Gamma(d_1 - b_1)\cdots\Gamma(d_n - b_n)}$$
$$\times \int_{(0,1)^n} \prod_{i=1}^n u_i^{b_i - 1}(1-u_i)^{d_i - b_i - 1} F_A^{(n)}(a, d_1, \ldots, d_n; c_1, \ldots, c_n; u_1 x_1, \ldots, u_n x_n) du_1 \cdots du_n$$

where $\Re(b_1), \ldots, \Re(b_n)$ are positive and $\Re(d_1 - b_1), \ldots, \Re(d_n - b_n)$ are positive and the sum of the moduli of the variables is less than unity,

$$F_A^{(n)}(a, b_1, \ldots, b_n; c_1, \ldots, c_n; x_1, \ldots, x_n)$$
$$= \frac{\Gamma(c_1)\cdots\Gamma(c_n)}{\Gamma(d_1)\cdots\Gamma(d_n)\Gamma(c_1 - d_1)\cdots\Gamma(c_n - d_n)}$$
$$\times \int_{(0,1)^n} \prod_{i=1}^n u_i^{d_i - 1}(1-u_i)^{c_i - d_i - 1} F_A^{(n)}(a, b_1, \ldots, b_n; d_1, \ldots, d_n; u_1 x_1, \ldots, u_n x_n) du_1 \cdots du_n$$

where $\Re(d_1),\ldots,\Re(d_n)$ are positive and $\Re(c_1-d_1),\ldots,\Re(c_n-d_n)$ are positive and the sum of the moduli of the variables is less than unity. A multiple integral relation involving the function $F_D^{(n)}$ is

$$F_D^{(n)}(a,b_1,\ldots,b_n;c;x_1,\ldots,x_n) = \frac{\Gamma(d_1)\cdots\Gamma(d_n)}{\Gamma(b_1)\cdots\Gamma(b_n)\Gamma(d_1-b_1)\cdots\Gamma(d_n-b_n)} \times \int_{(0,1)^n}\prod_{i=1}^{n}u_i^{b_i-1}(1-u_i)^{d_i-b_i-1}F_D^{(n)}(a,d_1,\ldots,d_n;c;u_1x_1,\ldots,u_nx_n)du_1\cdots du_n$$

where $\Re(b_1),\ldots,\Re(b_n)$ are positive and $\Re(d_1-b_1),\ldots,\Re(d_n-b_n)$ are positive and the moduli of the variables are less than unity. Moreover, we give four integral relations which result from (C.1.6)

$$F_A^{(n)}(a,b_1,\ldots,b_n;c_1,\ldots,c_n;x_1,\ldots,x_n) = \frac{\Gamma(d)}{\Gamma(b_1)\cdots\Gamma(b_n)\Gamma(d-b_1-\cdots-b_n)} \times \int_{\Delta^{(n)}}\prod_{i=1}^{n}u_i^{b_i-1}\left(1-\sum_{i=1}^{n}u_i\right)^{d-b_1-\cdots-b_n-1} F_C^{(n)}(a,d;c_1,\ldots,c_n;u_1x_1,\ldots,u_nx_n)du_1\cdots du_n$$

where $\Re(b_1),\ldots,\Re(b_n)$ are positive and $\Re(d-b_1-\cdots-b_n)$ is positive and the sum of the moduli of the square roots of the variables is less than unity,

$$F_B^{(n)}(a_1,\ldots,a_n,b_1,\ldots,b_n;c;x_1,\ldots,x_n) = \frac{\Gamma(c)}{\Gamma(d_1)\cdots\Gamma(d_n)\Gamma(c-d_1-\cdots-d_n)} \times \int_{\Delta^{(n)}}\prod_{i=1}^{n}u_i^{d_i-1}{}_2F_1(a_i,b_i;d_i;u_ix_i)\left(1-\sum_{i=1}^{n}u_i\right)^{c-d_1-\cdots-d_n-1} du_1\cdots du_n$$

where $\Re(d_1),\ldots,\Re(d_n)$ are positive and $\Re(c-d_1-\cdots-d_n)$ is positive and the moduli of the variables are less than unity,

$$F_B^{(n)}(a_1,\ldots,a_n,b_1,\ldots,b_n;c;x_1,\ldots,x_n) = \frac{\Gamma(d)}{\Gamma(b_1)\cdots\Gamma(b_n)\Gamma(d-b_1-\cdots-b_n-1)} \times \int_{\Delta^{(n)}}\prod_{i=1}^{n}u_i^{b_i-1}\left(1-\sum_{i=1}^{n}u_i\right)^{d-b_1-\cdots-b_n-1} F_D^{(n)}(d,a_1,\ldots,a_n;c;u_1x_1,\ldots,u_nx_n)du_1\cdots du_n$$

where $\Re(b_1),\ldots,\Re(b_n)$ are positive and $\Re(d-b_1-\cdots-b_n)$ is positive and the moduli of the variables are less than unity,

$$F_D^{(n)}(a, b_1, \dots, b_n; c; x_1, \dots, x_n)$$
$$= \frac{\Gamma(c)}{\Gamma(d_1)\cdots\Gamma(d_n)\Gamma(c - d_1 - \cdots - d_n)}$$
$$\times \int_{\Delta^{(n)}} \prod_{i=1}^{n} u_i^{d_i} \left(1 - \sum_{i=1}^{n} u_i\right)^{c-d_1-\cdots-d_n-1} F_A^{(n)}(a, b_1, \dots, b_n; d_1, \dots, d_n; u_1x_1, \dots, u_nx_n) du_1 \cdots du_n.$$

C.2 Fox H-function and Meijer G-function

For integers m, n, p, q such that $0 \le m \le q$, $0 \le n \le p$ for $a_i, b_j \in \mathbb{C}$ and for $\alpha_i, \beta_j \in \mathbb{R}^+$ ($i = 1, \dots, p$ and $j = 1, \dots, q$), the H-function $H_{p,q}^{m,n}(z)$ (see Fox [71]) is defined via a Mellin-Barnes-type integral in the form

$$H_{p,q}^{m,n}(z) = H_{p,q}^{m,n}\left[z \left| \begin{matrix} (a_1, \alpha_1), \dots, (a_p, \alpha_p) \\ (b_1, \beta_1), \dots, (b_n, \beta_n) \end{matrix} \right.\right] := \frac{1}{2\pi \mathrm{i}} \oint_{\mathfrak{L}} \mathcal{H}_{p,q}^{m,n}(s) z^{-s} ds \qquad \text{(C.2.1)}$$

where

$$\mathcal{H}_{p,q}^{m,n}(s) := \frac{1}{2\pi \mathrm{i}} \oint_{\mathfrak{L}} \frac{\prod_{j=1}^{m} \Gamma(b_j - \beta_j s) \prod_{i=1}^{n} \Gamma(1 - a_i - a_i s)}{\prod_{i=n+1}^{p} \Gamma(a_i + \alpha_i s) \prod_{j=m+1}^{q} \Gamma(1 - b_j - \beta_j s)}.$$

Here

$$z^{-s} = \mathrm{e}^{-s(\log(|z|) + i \arg z)} \qquad z \neq 0, \ \mathrm{i} = \sqrt{-1}$$

where $\log(|z|)$ represents the natural logarithm of $|z|$ and $\arg z$ is not necessarily the principar value. An empty product in (C.2.1), if it occours, is taken to be one, and the poles

$$b_{j,l} = \frac{-b_j - l}{\beta_j} \qquad j = 1, \dots, m, \ l \ge 0 \qquad \text{(C.2.2)}$$

of the Gamma functions $\Gamma(b_j + \beta_j s)$ and the poles

$$a_{i,k} = \frac{1 - a_i + k}{\alpha_i} \qquad i = 1, \dots, n, \ k \ge 0 \qquad \text{(C.2.3)}$$

of the Gamma functions $\Gamma(1 - a_i - \alpha_i s)$ do not coincide

$$\alpha_i(b_j + l) \neq \beta_j(a_i - k - 1) \qquad i = 1, \dots, n, \ j = 1, \dots, m, \ k, l \ge 0.$$

$\mathfrak{L}$ in (C.2.1) is the infinite contour which separates all the poles $b_{j,l}$ in (C.2.2) to the left and all the poles $a_{i,k}$ in (C.2.3) to the right of $\mathfrak{L}$, an has one of the following forms:

i) $\mathfrak{L} = \mathfrak{L}_{-\infty}$ is a left loop situated in a horizontal strip starting at the point $-\infty + \mathrm{i}\varphi_1$ and terminating at the point $-\infty + \mathrm{i}\varphi_2$ with $-\infty < \varphi_1 < \varphi_2 < +\infty$;

ii) $\mathfrak{L} = \mathfrak{L}_{+\infty}$ is a right loop situated in a horizontal strip starting at the point $+\infty + \mathrm{i}\varphi_1$ and terminating at the point $+\infty + \mathrm{i}\varphi_2$ with $-\infty < \varphi_1 < \varphi_2 < +\infty$;

iii) $\mathfrak{L} = \mathfrak{L}_{\mathrm{i}\gamma\infty}$ is a contour starting at the point $\gamma - \mathrm{i}\infty$ and terminating at the poing $\gamma + \mathrm{i}\infty$, where $\gamma \in \mathbb{R}$.

In particular, the properties of the H-function $H^{m.n}_{p,q}(z)$ depend on the numbers a^*, Δ, δ, μ, a_1^* and a_2^* which are expressed via m, n, p, q, a_i, α_i $(i = 1, \ldots, p)$ and b_j, β_j $(j = 1, \ldots, q)$ by the following relations

$$a^* := \sum_{i=1}^{n} \alpha_i - \sum_{i=n+1}^{p} + \sum_{j=1}^{m} \beta_j - \sum_{j=m+1}^{q} \beta_j \tag{C.2.4}$$

$$\Delta := \sum_{j=1}^{q} \beta_j - \sum_{i=1}^{p} \alpha_i \tag{C.2.5}$$

$$\delta := \prod_{i=1}^{q} \alpha_i^{-\alpha_i} \prod_{j=1}^{q} \beta_j^{\beta_j} \tag{C.2.6}$$

$$\mu := \sum_{j=1}^{q} b_j - \sum_{i=1}^{p} a_i + \frac{p-q}{2} \tag{C.2.7}$$

$$a_1^* := \sum_{j=1}^{m} \beta_j - \sum_{i=n+1}^{p} \alpha_i \tag{C.2.8}$$

$$a_2^* := \sum_{i=1^n} \alpha_i - \sum_{j=m+1}^{q} \beta_j \tag{C.2.9}$$

$$a_1^* + a_2^* = a^* \tag{C.2.10}$$

$$a_1^* - a_2^* = \Delta \tag{C.2.11}$$

$$\xi := \sum_{j=1}^{m} b_j - \sum_{j=m+1}^{q} b_j + \sum_{i=1}^{n} a_i - \sum_{i=n+1}^{p} a_i \tag{C.2.12}$$

$$c^* := m + n - \frac{p+q}{2}. \tag{C.2.13}$$

An empty sum in (C.2.4), (C.2.5), (C.2.7), (C.2.8), (C.2.9), (C.2.12) and an empty product in (C.2.6), if they occour, are taker to be zero and one, respectively. The conditions for the existence of the H-function follow by virtue ot these relations.

Theorem C.2.1. *(crf. Kilbas and Saigo [107]) Let* a^**,* Δ*,* δ *and* μ *be given in* (C.2.4), (C.2.5), (C.2.6) *and* (C.2.7). *Then the H-function* $H_{p,q}^{m,n}(z)$ *defined by* (C.2.1) *makes sense in the following cases*

$$\mathfrak{L} = \mathfrak{L}_{-\infty}, \ \Delta > 0, \ z \neq 0 \tag{C.2.14}$$

$$\mathfrak{L} = \mathfrak{L}_{-\infty}, \ \Delta = 0, \ 0 < |z| \leq \delta \tag{C.2.15}$$

$$\mathfrak{L} = \mathfrak{L}_{-\infty}, \ \Delta = 0, \ z = \delta, \ \Re(\mu) < -1 \tag{C.2.16}$$

$$\mathfrak{L} = \mathfrak{L}_{+\infty}, \ \Delta < 0, \ z \neq 0 \tag{C.2.17}$$

$$\mathfrak{L} = \mathfrak{L}_{+\infty}, \ \Delta = 0, \ |z| > 0 \tag{C.2.18}$$

$$\mathfrak{L} = \mathfrak{L}_{+\infty}, \ \Delta = 0, \ |z| = \delta, \ \Re(\mu) < -1 \tag{C.2.19}$$

$$\mathfrak{L} = \mathfrak{L}_{\mathrm{i}\gamma\infty}, \ a^* > 0, \ |\arg z| < \frac{a^*\pi}{2}, \ z \neq 0 \tag{C.2.20}$$

$$\mathfrak{L} = \mathfrak{L}_{\mathrm{i}\gamma\infty}, \ a^* = 0, \ \Delta\gamma + \Re(\mu) < -1, \ \arg z = 0, \ z \neq 0. \tag{C.2.21}$$

Refer to Kilbas and Saigo [107] and to Yakubovich and Luchko [187] for a detailed discussion about H-function and in particular for further conditions of existence of the H-functions.

For integers m, n, p, q such that $0 \leq m \leq q$, $0 \leq n \leq p$ for $a_i, b_j \in \mathbb{C}$ and for $\alpha_i, \beta_j \in \mathbb{R}^+$ ($i = 1, \dots, p$ and $j = 1, \dots, q$), the G-function $G_{p,q}^{m,n}(z)$ (see Meijer [136], Meijer [137] and Meijer [138]) is defined via a Mellin-Barnes-type integral in the form

$$G_{p,q}^{m,n}(z) = G_{p,q}^{m,n}\left[z \,\middle|\, \begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix}\right] := \frac{1}{2\pi \mathrm{i}} \oint_{\mathfrak{L}} \mathcal{G}_{p,q}^{m,n} z^{-s} ds$$

where

$$\mathcal{G}_{p,q}^{m,n} := \frac{1}{2\pi \mathrm{i}} \oint_{\mathfrak{L}} \frac{\prod_{j=1}^{m} \Gamma(b_j + s) \prod_{j=1}^{n} \Gamma(1 - a_j - s)}{\prod_{j=n+1}^{p} \Gamma(a_j + s) \prod_{j=m+1}^{q} \Gamma(1 - b_j - s)}$$

where the contour of integration $\mathfrak{L}$ is set up to lie between the poles of $\Gamma(a_i + s)$ and the poles of $\Gamma(b_j + s)$. The G-function is defined uner the following hypothesis:

i) $0 \leq m \leq q$, $0 \leq n \leq p$ and $p \leq q - 1$

ii) $z \neq 0$

iii) no couple of b_j for $j = 1, \dots, m$ differs by an integer or a zero;

iv) the parameter $a_i \in \mathbb{C}$ and $b_j \in \mathbb{C}$ are so that no pole of $\Gamma(b_j + s)$ for $j = 1, \dots, m$ coincide with any pole of $\Gamma(a_i + s)$ for $i = 1, \dots, n$;

v) $a_i - b_j \neq 1, \dots$ for $i = 1, \dots, n$ and $j = 1, \dots, m$;

vi) if $p = q$, then the definition makes sense only for $|z| > 1$.

Refer to Erdélyi et al. [37] for a more thorough discussion of the G-functions and in particular for further conditions of existence of the G-functions

Both H-functions and G-functions are very general functions whose special cases cover most of the mathematical functions such as the trigonometric functions, Bessel functions and generalized hypergeometric functions. Comparison between the definition of H-function and the definition of G-function, reveals that any G-function is an H-function, but not vive versa. In particular, when $\alpha_i = \beta_j$ for $i = 1, \ldots, p$ and $j = 1, \ldots, q$, then

$$H^{m,n}_{p,q}\left[z \middle| \begin{matrix}(a_1,1),\ldots,(a_p,1)\\(b_1,1),\ldots,(b_n,1)\end{matrix}\right] = G^{m,n}_{p,q}\left[z \middle| \begin{matrix}a_1,\ldots,a_p\\b_1,\ldots,b_q\end{matrix}\right].$$

The following list shows how it is possible to express some mathametical functions in terms of the H-function

$$\mathrm{e}^z = H^{1,0}_{0,1}\left[-z \middle| \begin{matrix}\cdot\\(0,1)\end{matrix}\right] \qquad \forall z \tag{C.2.22}$$

$$\cos(z) = \frac{1}{\sqrt{\pi}} H^{1,0}_{0,2}\left[\frac{z^2}{4} \middle| \begin{matrix}\cdot\\(0,1),(1/2,1)\end{matrix}\right] \qquad \forall z \tag{C.2.23}$$

$$\sin(z) = \frac{2}{\sqrt{\pi}} H^{1,0}_{0,2}\left[\frac{z^2}{4} \middle| \begin{matrix}\cdot\\(0,1),(-1/2,1)\end{matrix}\right] \qquad z \geq 0 \tag{C.2.24}$$

$$\log(1+z) = H^{1,0}_{1,2}\left[\frac{z^2}{4} \middle| \begin{matrix}(1,1),(1,1)\\(1,1),(0,1)\end{matrix}\right] \qquad |z| < 1 \tag{C.2.25}$$

$$\Gamma(x,z) = H^{2,0}_{1,2}\left[z \middle| \begin{matrix}(1,1)\\(0,1),(x,1)\end{matrix}\right] \qquad \forall z \tag{C.2.26}$$

$$K_\eta(z) = \frac{1}{2}\left(\frac{x}{2}\right)^{-a} = H^{1,0}_{0,2}\left[\frac{z^2}{2} \middle| \begin{matrix}\cdot\\(a-\eta/2,1),(a+\eta/2,1)\end{matrix}\right] \qquad \forall z \tag{C.2.27}$$

$${}_pF_q(a_1,\ldots,a_p;b_1,\ldots,b_1;z) = \frac{\prod_{i=1}^p \Gamma(a_i)}{\prod_{j=1}^q \Gamma(b_j)} \times \tag{C.2.28}$$

$$\times H^{1,p}_{p,q+1}\left[-z \middle| \begin{matrix}(1-a_1,1),\ldots,(1-a_p,1)\\(0,1),(1-b_1,1),\ldots,(1-b_q,1)\end{matrix}\right] \qquad p \leq q, 0 \leq |z| \leq 1$$

where the last two equations corresponds to the modified Bessel function of the third kind and to the generalized hypergeometric function, respectively Refer to Kilbas and Saigo [107] for a more exhaustive list of relationships between the H-functions and some other mathematical functions.

Bibliography

[1] Aldous, D. J. (1985). Exchangeability and related topics. In P. Hennequin, editor. *Ecole d'ete de probabilites de Saint-Flour, XIII - 1983.* Springer, Berlin. Lecture notes in Mathematics 1117.

[2] Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2**, 1152–1174.

[3] Bell, E. T. (1927). Partition polynomials. *Ann. Math.* **29**, 38–46.

[6] Bergeron, F., Labelle, G. and Leroux, P. (1994) *Théorie des espèces et combinatorie des structures arborescentes*, volume 19 of *Publ. du Lab. de Combinatorie et d'Informatique Mathématique.* Univ de Québex, Montrèal.

[5] Bergeron, F., Labelle, G. and Leroux, P. (1998) *Combinatorial species and tree-like structures.* Cambridge University Press, Cambridge. Translated from 1994 French original by Margaret Readdy, with a foreword by Gian Carlo Rota.

[6] Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory.* Wiley, Chichester.

[7] Berti, P., Pratelli, L., Rigo, P. (2006). Almost sure weak convergence of random probability measures. *Stochastics* **78** 91–97.

[8] Billingsley, P. (1992). *Convergence of probability measures.* Wiley-Interscience.

[9] Blackwell, D. (1973). Discretness of Ferguson selections. *Ann. Statist.* **1**, 356–358.

[10] Blackwell, D., MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1**, 353–355.

[11] Bondesson, L. (1981). Classes of infinitely divisible distributions and densities. *Probab. Theory Related Fields.* **57**, 39–71.

[12] Bondesson, L. (1992). *Generalized gamma convolutions and related classes of distributions and densities*, Lecture Notes in Statistics. Springer-Verlag.

[13] Brix, A. (1999). Generalized Gamma measures and shot-noise Cox processes. *Adv. Appl. Probab.* **31**, 929–953.

[14] Carlitz, L. (1980). Weighted Stirling numbers of the first and the second kind-I. *Fibonacci Quart.* **18**, 147–162.

[15] Charalambides, C. A. (1988). A review of the Stirling numbers, their generalizations and statistical applications. *Commun. Statist. A.* **17**, 2533–1595.

[16] Charalambides, C. A. (2002). *Enumerative combinatorics*, Chapman and Hall CRC.

[17] Charalambides, C. A. (2005). *Combinatorial methods in discrete distributions*, Hoboken, NJ: Wiley.

[18] Cifarelli, D. M., Regazzini, E. (1990). Distribution functions of means of a Dirichlet process. *Ann. Statist.* **18**, 429–442.

[19] Comtet L. (1974). *Advanced Combinatorics*, D. Reidel Publishing Company, Boston.

[20] Cont, R. and Tankov, P. (2004). *Financial modelling with jump processes.* Chapman and Hall/CRC, Boca Raton.

[21] Connnor, R. J. and Mosimann, I. E. (1969). Concepts of independence for proportions with a generalization of the Dirichlet distribution. *J. Amer. Statist. Assoc.* **64**, 194-206.

[22] Daley D. J. and Vere-Jones, D. (1988). *An introduction to the theory of point processes.* Springer, New York.

[23] Dawson, D. A. and Hochberg, K, J. (1982). Wandering random measures in the Fleming-Viot model. *Ann. Probab.* **10**, 554-580.

[24] Diaconis, P. and Kemperman, J. B. K. (1995). Some new tools for Dirichlet priors. In: Bernardo, J., Berger, J. O., Dawid, A. and Smith A. F. M. (Eds.), Bayesian Statistics, Oxford University Press, pp. 95-104.

[25] Doksum, K. A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2**, 183–201.

[26] Donelly, P. (1986). Partition structures, Pólya urns, the Ewens sampling formula, and the ages of alleles. *Theor. Popul. Biol.* **30**, 271–288.

[27] Donnelly, P. and Grimmet, G. (1993). On the asymptotic distribution of large prime factors. *J. Lond. Math. Soc.* **47**, 395–404.

[28] Donelly, P., Joyce, P. (1989). Continuity and weak convergence of ranked and size-biased permutations on the infinite simplex.*Stochastic Process. Appl.* **31**, 89–103.

[29] Donelly, P. and Kurtz, T. G. (1996). A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Probab.* **24**, 743-760.

[30] Donelly, P. and Kurtz, T. G. (1999). Genealogical processes for Fleming-Viot models with selection and recombination. *Ann. Appl. Probab.* **9**, 1091-1148.

[31] Donoghue, W. (1974). *Monotone matrix functions and analytic continuation.* Springer-Verlag, Berlin-Heidelberg-New York.

[32] Doss, H. (1991). Bayesian nonparametric estimation for incomplete data via successive substitution sampling. Technical Report, Florida State University. **850**.

[33] Doss, H. and Sellke, T. (1982). The talis probabilities chosen from a Dirichlet prior. *Ann. Statist.* **10**, 1302–1305.

[34] Duan, J. A., Gelfand, E. G. and Guindani, M. (2005). Generalized spatial Dirichlet process models. Technical Report, Duke University.

[35] Dubins, L. and Freedman, D. (1963). Random distribution functions. *Bull. Amer. Math. Soc.* **69**, 548–551.

[36] Dubins, K. and Freedman, D. (1966). Random distribution functions. *Proc. Fifth Berkley Symp. Math. Statist. Probab.* **2**, 183–214. University of California Press.

[37] Erdélyi, A., Magnus, W., Oberhettinger, F., and Tricomi, F. G. (1976). *Higher trascendental functions Vol. I.* McGraw-Hill, New York.

[38] Erhardsson, T. (2008). Nonparametric Bayesian inference for integrals with respect to an unknown finite Mmeasure. *Scand. J. Statist.* **35**, 354–368.

[41] Ethier, S. N. and Griffiths, R. C. (1993). The transition function of a Fleming-Viot process. *Ann. Probab.* **21**, 1571–1590.

[40] Ethier, S.N. and Kurtz, T. G. (1993). The infinitely-many-alleles model with selection as a measure-valued diffusion. *Lecture notes in Biomatemathics.* **70**, Springer-Verlag, Berlin.

[41] Ethier, S. N. and Kurtz, T. G. (1993). Fleming-Viot processes in population genetics. *SIAM J. Control Optim.* **31**, 345–386.

[42] Ethier, S. N. and Kurtz, T. G. (1994). Convergence to Fleming-Viot processes in the weak atomic topology. *Stochastic Process. Appl.* **54**, 1–27.

[43] Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3**, 87–112.

[44] Ewens, W. J. (1986). Population genetics theory - the past and the future. In: Lessard, S. (ed.), Mathematical and statistical problems in evolution. Montreal: University of Montreal Press 1988.

[45] Ewens, W.J. (2004). *Mathematical population genetics.* Springer.

[46] Ewens, W. J. and Tavaré, S. (1995). The Ewens sampling formula. In *Multivariate discrete distributions* (Johnson, N.S., Kots, S. and Balakrishnan, N., eds.). Wiley, New York.

[47] Exton, H. (1976). *Multiple hypergeometric functions and application.* Ellis Horwood, Chinchester.

[48] Fabius, J. (1963). On the asymtotic behavior of Bayes' estimates. *Ann. Math. Statist.* **35**, 856–856.

[49] Fan, D.Y. (1991). The distribution of the product of independent beta variables. *Ann. Math. Stat.* **20**, 4043–4052.

[50] Favaro, S. and Walker, S. G. (2008). A generalized constructive definition for the Dirichlet process. *Submitted.*

[51] Favaro, S. and Walker, S. G. (2008). Some developments on a class of random probability measures without Gibbs structure . *Submitted.*

[52] Favaro, S. Ruggiero, M. and Walker, S. G. (2008). On the stationary distribution of finite approximations to the neutral diffusion model. *Submitted.*

[53] Favaro, S. Ruggiero, M. Spano, D. and Walker, S. G. (2008). The neutral population model and Bayesian nonparametrics. *Submitted.*

[54] Favaro, S. Prünster, I. and Walker, S. G. (2008). On a class of random probability measures with general predictive structure. *Submitted.*

[55] Favaro, S., Guglielmi, A. and Walker, S. G. (2008). A generalized Feigin-Tweedie Markov chain. *Submitted.*

[56] Favaro, S., Lijoi, A., Mena, R. H. and Prünster, I. (2008). Bayesian nonparametric inference for species variety with a two parameter Poisson-Dirichlet process prior. *Submitted.*

[57] Favaro S. and Walker, S. G. (2008). On the Wilks' integral representation for products of beta random variables. *Submitted.*

[58] Feigin, P. D., Tweedie, R. L. (1989). Linear functionals and Markov chains associated with Dirichlet process. *Math. Proc. Cambridge Philos. Soc.* **105**, 579–585.

[59] Feller, W. (1964). *An introduction to probability theory and its applications.* Volume I. John Wiley and Sons, Inc.

[60] Feller, W. (1971). *An introduction to probability theory and its applications.* Volume II. John Wiley and Sons, Inc.

[61] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.

[62] Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 615–629.

[63] Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In: Rizvi, M. H., Rustagi J. and Siegmund D. (Eds.), Recent Advances in Statistics: Papers in Honor of Herman Chernoff on his Sixtieth Birthday, Academic Press, pp. 287–302.

[64] Ferguson, T. S. and Klass, M. J. (1972). A representation of independent increments processes without Gaussian components. *Ann. Math. Statist.* **43**, 1634–1643.

[65] Ferguson, T. S., Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. *Ann. Statist.* **7**, 163–186.

[66] Ferguson, T. S., Phadia, E. G., Tiwari, R. C. (1992). Bayesian nonparametric inference. In: Ghosh, M. and Pathak, P. K. (Eds.), Current Issues in Statistical Inference: Essays in Honor of D. Basu, IMS Lecture Notes, Monograph Series. **17**, 127–150.

[67] De Finetti, B. (1930). La funzione caratteristica di un fenomeno aleatorio. *Atti Reale Accademia Nazionale dei Lincei, Serie VI, Rend.*, **12**, 367-373.

[68] De Finetti, B. (1935). Il problema della perequazione. In: Atti della Societa Italiana per il Progresso delle Scienze (XXIII Riunione, Napoli, 1934). SIPS, Roma. **2**, 227–228.

[69] De Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Ann. Instit. H. Poincaré*, **7**, 1-68.

[70] Fleming, W. H. and Viot, M. (1979). Some measure-valued Markov processes in population genetics theory. *Indiana J. Math.* **28**, 817–843.

[71] Fox, C. (1893). The G and H functions as symmetrical Fourier kernels. *Trans. Amer. Math. Soc.* **98**, 395–429.

[72] Freedman, D. A. (1963). On the asymtotic behavior of Bayes' estimates in the discrete case. *Ann. Math. Statist.* **34**, 1386–1403.

[73] Gelfand, A. E., Kottas, A. and MacEachern, S. (2005). Bayesian nonparametric spatial modelling with Dirichlet processes mixing. *J. Amer. Statist. Assoc.* In press.

[74] Gnedin, A. and Pitman, J. (2005). Exchangeable Gibbs partitions and Stirling triangles. *Zap. Nauchn. Sem. S. Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **325**, 83–102.

[75] Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika.* **40**, 237–264.

[76] Good, I. J. and Toulmin, G. H. (1956). The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika.* **43**, 45–63.

[77] Gradshteyn, L. S., Ryzhik, L. M (2000). *Table of integrals, series, and products*, Academic Press.

[78] Griffin, J. and Steel, M. (2006). Order-based dependent Dirichlet processes. *J. Amer. Statist. Assoc.* **101**, 179–194.

[79] Grote, M. and Speed, T. P. (2002). Approximate Ewens formulae for symmetric over-dominance selection. *Ann. Appl. Probab.* **12**, 637–663.

[80] Guglielmi, A. (1998). A simple procedure calculating the generalized Stieltjes transform of the mean of a Dirichlet process. *Statist. Probab. Lett.* **38**, 299–303.

[81] Guglielmi, A. and Tweedie, R. (2001). MCMC estimation of the law of the mean of a Dirichlet process. *Bernoulli.* **4**, 573–592.

[82] Halmos, P. R. (1944). Random alms. *Ann. Math. Statist.* **15**, 182–189.

[83] Hannum, R. C., Hollander, M., Langberg, N. A. (1981). Distributional results for random functionals of a Dirichlet process. *Ann. Probab.* **9**, 665–670.

[84] Hjort, N. L. (1990). Nonparametric Bayes estimators based on Beta processes in models for life history data. *Ann. Statist.* **18**, 1259–1294.

[85] Hiort N., Ongaro, A. (2005). Exact inference for random Dirichlet means. *Stat. Inf. Stoch. Proc.* **8**, 227–254.

[86] Ho, M.–W., James, L.F. and Lau, J.W. (2007). Gibbs Partitions (EPPF's) Derived From a Stable Subordinator are Fox H and Meijer G Transforms. Preprint. *MatharXiv* arXiv:0708.0619v2.

[87] Hoppe, F. M. (1984). Pólya-like urns and the Ewens sampling formula.*J. Math. Biol.* **20**, 91–94.

[88] Hoppe, F. M. (1987). The sampling theory of neutral alleles and an urn model in population genetics.*J. Math. Biol.* **25**, 123–159.

[89] Ignatov, T. (1982). On a constant arising in the theory of symmetric groups and on Poisson-Dirichlet measures.*Theory Probab. Appl.* **27**, 146–147.

[90] Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Stat. Assoc.* **96**, 161–173.

[91] Ishwaran, H. and James, L. F. (2003). Generalized weighted Chinese restaurant processes for species sampling mixture models. *Statist. Sinica.* **13**, 1211–1235.

[92] Ishwaran, H. and Zarepour, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two parameter process hierarchical models. *Biometrika.* **87**, 371–390.

[93] Ishwaran, H. and Zarepour, M. (2002). Dirichlet prior sieves in finite normal mixtures. *Statist. Sinica.* **12**, 941–963.

[94] Ishwaran, H. and Zarepour, M. (2002). Exact and approximate sum-representation for the Dirichlet process. *Canad. J. Statist.* **30**, 269–283.

[95] Jambunathan, M. V. (1954). Some properties of beta and gamma distributions. *Ann. Math. Statist.* **25**, 401–405.

[96] James, L. F. (2002). Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics. MatharXive, arXiv:math/0205093v1.

[97] James, L. F. (2003). A simple proof of the almost sure discreteness of a class of random measures. *Statist. Probab. Lett.* **65**, 363–368.

[98] James, L. F. (2005). Functionals of Dirichlet processes, the Cifarelli-Regazzini identity and beta-gamma processes. *Ann. Statist.* **33**, 647–660.

[99] James, L. F., Lijoi, A. and Prünster, I. (2005). Conjugacy as a distinctive feature of the Dirichlet process.*Scand. J. Statist.* **20**, 1–16

[100] James, L. F., Lijoi, A. and Prünster, I. (2006). Posterior analysis of normalized random measure with independent increments. *Scand. J. Statist.* In press.

[101] James, L. F., Roynette, B. and Yor, M. (2008). Generalized Gamma Convolutions, Dirichlet means, Thorin measures, with explicit examples. Preprint. *MatharXiv* arXiv:0708.3932v1.

[102] Jara, A., Lesaffre, E., De Iorio, M. and Quintana F. (2008). Bayesian Semiparametric Inference for Multivariate Doubly-Interval-Censored Data. Technical Report, Katholieke Universiteit Leuven.

[103] Jarner, S. F., Tweedie, R. L. (2002). Convergence rates and moments of Markov chains associated with the mean of Dirichlet processes. *Stochastic Process. Appl.* **101**, 257–271.

[104] Kallenberg, O. (1983). *Random measures*, Akademie-Verlag. Berlin.

[105] Karlin. S., and McGregor, J. (1967). The number of mutant forms maintained in a population. “Proceedings of Fifth Berkley Symposium on Math. Statist. and Prob.”, Vol. IV, University of California Press, Berkley.

[106] Kelly, F. P. (1976). On stochastic population models in genetics. *J. Appl. Prob.* **13**, 127–131.

[107] Kilbas, A. A., Saigo, M. (2004). *H-Transforms: Theory and Applications (Analytical Methods and Special Functions, 9).* CRC.

[108] Kingman, J. F. C. (1967). Completely random measures. *Pacific. J. Math.* **21**, 59–78.

[109] Kingman, J. F. C (1975). Random discrete distributions. *J. Roy. Statist. Soc. Ser. B.* **37**, 1–22.

[110] Kingman, J. F. C. (1977). The population structure associated with the Ewens sampling formula. *Theor. Popul. Biol.* **11**, 274–283.

[111] Kingman, J. F. C. (1978). Random partitions in population genetics. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **361**, 1–20.

[112] Kingman, J. F. C. (1978). The representation of partition structures. *J. Lond. Math. Soc.* **18**, 374–380.

[113] Kingman, J. F. C. (1982). The coalescent.*Stochastic Process. Appl.* **13**, 235–248.

[114] Kingman, J. F. C. (1993). *Poisson processes.* Oxford University Press, Oxford.

[115] Kumar, S. and Tiwari, R. C. (1898). Bayes estimation of reliability under random environment governed by a Dirichlet prior. *IEEE Trans. Reliab.* **38**, 218–223.

[116] Lavine, M. (1992). Some aspects of Pólya tree distributions for statistical modelling.*Ann. Statist.* **20**, 1222–1235.

[117] Lauricella, G. (1893). Sulle funzioni ipergeometriche a più variabili. *Rend. Circ. Mat. Palermo.* **7**, 111–158.

[118] Lijoi, A., Mena, R. H. and Prünster, I. (2005). Bayesian Nonparametric analysis for a generalized Dirichlet process prior. *Stat. Inference Stoch. Process.* **8**, 283–309.

[119] Lijoi, A., Mena, R. H. and Prünster, I. (2005). Hierarchical mixture modelling with normalized inverse-Gaussian priors. *J. Amer. Stat. Assoc.* **100**, 1278–1291.

[120] Lijoi, A., Mena, R. H. and Prünster, I. (2007). Controlling the reinforcement in Bayesian nonparametric mixture models. *J. Roy. Statist. Soc. Ser. B.* In press.

[121] Lijoi A., Mena, R. H., Prünster, I. (2007). A Bayesian nonparametric method for prediction in EST analysis. *Bioinformatics.* **8**, 339.

[122] Lijoi A., Mena, R. H., Prünster, I. (2008). A Bayesian Nonparametric approach for comparing clustering structures in EST libraries. Technical Report, IIMAS, Mexico.

[123] Lijoi, A., Mena, R. H., Prünster, I. (2007). Bayesian nonparametric estimation of the probability of discovering a new species *Biometrika.* **94**, 769–786.

[124] Lijoi, A., Prünster, I., Walker, S. G. (2008). Investigating nonparametric priors with Gibbs structure. *Statist. Sinica.* In press.

[125] Lijoi, A., Prünster, I., Walker, S.G. (2008). Bayesian nonparametric estimators derived from conditional Gibbs structures. *Ann. Appl. Probab.* In press.

[126] Lijoi, A., Regazzini, E. (2004). Means of a Dirichlet process and multiple hypergeometric functions. *Ann. Probab.* **32**, 1469–1495.

[127] Liu, J. S. (1996). Nonparametric hierarchical bayes via sequential imputations. *Ann. Statist.* **24**, 911–930.

[128] Lo, A. (1984). On a class of Bayesian nonparametric estimates I. Density estimates. *Ann. Statist.* **12**, 351–357.

[129] Lo, A.I. and Weng C. S. (1984). On a class of Bayesian nonparametric estimates II. Hazard rate estimates *Ann. Inst.Statist. Math.* **41**, 227–245.

[130] MacEachern, S. N. (1999). Dependent nonparametric processes. In: ASA Proceedings of the section on Bayesian statistical science. Alexandria, VA: American Statistical Association.

[131] MacEachern, S. N. (2000). Dependent Dirichlet Processes. Technical Report, Department of Statistics, Ohio State University.

[132] Mao, C. X. (2004). Prediction of the conditional probability of discovering a new class. *J. Amer. Statist. Assoc.* **99**, 1108–1118.

[133] Mao, C. X. (2006). Estimating species accumulation curves and diversity indices. Preprint. *MatharXiv* arXiv:0609577.

[134] Mao, C. X. and Lindsay, B. G. (2002). A Poisson model for the coverage problem with a genomic application. *Biometrika.* **89**, 669–682.

[135] Maudlin, R. D., Sudderth, W. D. and Williams, S. C. (1992). Pólya trees and random distributions.*Ann. Statist.* **20**, 1203–1221.

[136] Meijer, G. S. (1936). Über Whittakersche bezw. Besselsche funktionen und deren Produkte. *Nieuw. Arch. Wiskunde* **18**, 10–39.

[137] Meijer, G. S. (1941). Multiplikationstheoreme fr di Funktion $G_{p,q}^{m,n}(z)$. *Proc. Nederl. Akad. Wetensch.* **44**, 1062–1070.

[138] Meijer, G. S. (1946). On the G function, I-VIII, *Proc. Nederl. Akad. Wetensch.* **49**, 227–237, 344–356, 457–469, 632–641, 765–772, 936–943, 1063–1072, 1165–1175.

[139] Muliere, P. and Secchi, P. (1995). A note on a proper Bayesian bootstrap. Technical report n. 18, Dipartimento di Economia Politica e Metodi Quantitativi, Universita degli Studi di Pavia.

[140] Muliere, P. and Tardella, L. (1998). Approximating distributions of random functional of Ferguson-Dirichlet priors. *Canad. J. Statist.* **26**, 283–297.

[141] Meyn, S. P., Tweedie, R. L. (1993). *Markov chains and stochastic stability*, Springer, New York.

[142] Nadarajah S. L (2008). A review of results on sum of random variables. *Acta Appl. Math.* 10.1007/s10440-008-9224-4.

[143] Nandi, S. B. (1980). On the exact distribution of a normalized ratio of the weighted geometric mean to the unweighted arithmetic mean in samples from gamma distributions. *J. Amer. Statist. Assoc.* **75**, 283–297.

[144] Navarrete, C., Quintana, F. A., and Müller, P.. Some issues on Nonparametric Bayesian modeling using species sampling models. *Stat. Modell.* **41** 3-21.

[145] Nieto-Barajas, L. E., Prünster, I. and Walker, S. G. (2004). Normalized random measures driven by increasing additive processes. *Ann. Statist.* **32**, 2343–2360.

[146] Nummelin E. (1984). *General irriducible Markov chain and nonnegative operators*, Cambridge University Press.

[147] Otha, T. and Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophorerically detectable alleles in a finite population. *Genet. Res. Camb.* **22**, 201–204.

[148] Papaspiliopoulos, O. and Roberts, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika.* **95**, 169–186.

[149] Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields.* **102**, 145–158.

[151] Pitman, J. (1996). Random discrete distributions invariant under size-biased permutations. *Adv. in Appl. Probab.* **28**, 525–539.

[151] Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory. Papers in honor of David Blackwell* (Eds. Ferguson, T. S. et al.). Lecture Notes, Monograph Series, **30** 245–267. Institute of Mathematical Statistics, Hayward.

[152] Pitman, J. (1996b). Notes on the two parameter generalization of Ewens random partition structure. *Unpublished manuscript.*

[153] Pitman, J. (1997). Partition structures derived from Brownian motion and stable subordinators. *Bernoulli.* **3**, 79–96.

[154] Pitman, J. and Yor, M. (1997). The two parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **27**, 1870–1902.

[155] Pitman, J. (1999). Brownian motion, bridge, excursion and meander characterized by sampling at independent uniform times *Electron. J. Probab.* **4**, 1–33.

[156] Pitman, J. (2003). Poisson-Kingman partitions. In: Goldstein, D. R. (ed.) *Science and Statistics: a festschrift for Terry Speed.* Institute of Mathematical Statistics, Hayward, Lecture Notes, Monograph Series. **40**, 1–35.

[157] Pitman, J. (2006). *Combinatorial stochastic processes.* Ecole d'Eté de Probabilités de Saint-Flour XXXII. Lecture Notes in Mathematics N. 1875, Springer, New York.

[158] Prokhorov, Y. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.* **1**, 157-214.

[159] Quackenbush, J., Cho, J., Lee, D., Liang, F., Holt, I., Karamycheva, S., Parvizi, B., Pertea, G., Sultana, R. and White, J. (2000). The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic Acids Research.* **29**, 159–164.

[160] Quintana F. A. (2006). A predictive view of Bayesian clustering. *J. Statist. Plann. Inference* **136**, 2407–29.

[161] Regazzini, E. (1978). Intorno ad alcune questioni relative alla definizione del premio secondo la teoria della credibilità. *G. Ist. Ital. Att.* **41**, 77–89.

[162] Regazzini, E. and Petris, G. (1992). Some critical aspects on the use of exchangeability in statistics. *J. Ital. Statist. Soc.* **1**, 103–130.

[163] Regazzini, E. (1998). An example of the interplay between statistics and special functions. In *Tricomi's idea and contemporary applied mathematics* Acc. Naz. Lincei e Acc. Scienze Torino, Roma, Atti dei Convegni Lincei Vol. 147, 1998 pp. 303-320.

[164] Regazzini, E. (2001). *Foundation of Bayesian statistics and some theory of Bayesian nonparametric methods.* Lecture notes, Stanford University.

[165] Regazzini, E., Lijoi, A. and Prünster, I. (2003). Distributional results for means of random measures with independent increments. *Ann. Statist.* **31**, 560–585.

[166] Roberts, G. O. and Tweedie, R. L. (1999). Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Process. Appl.* **80**, 211–229.

[167] Roberts, G. O. and Tweedie, R. L. (2000). Rates of convergence of stochastically monotone stochastic processes. *J. Appl. Probab.* **37**, 359–373.

[168] Rolph, J. E. (1968). Bayesian estimation of mixing distributions. *Ann. Math. Statis.* **39**, 1289–1302.

[169] Rosiński, J. (2007). Tempering stable processes. *Stochastic Process. Appl.* **117**, 677–707.

[170] Rubin, D. B. (1981). The Bayesian bootstrap. *Ann. Statist.* **9**, 130–134.

[171] Ruggero, M. and Walker, S. G. (2008). Bayesian nonparametric construction of the Fleming-Viot process with fertility selection. *Statist. Sinica.* To appear.

[172] Serfling, R. J. (1981). *Approximation theorems of mathematical statistics*, New York, Wiley.

[173] Sethuraman, J. and Tiwari, T. C. (1982). Convergence of Dirichlet measures and the interpretation of their parameter. In: Gupta, S. S. and Berger, J. O. (Eds.), Statistical Decision Theory and Related Topics III, pp. 305-315.

[174] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica.* **4**, 639–650.

[175] Shervish, M. J. (1995). *Theory of statistics.* Springer, New York.

[176] Springer, M. D., Thompson, W. E. (1970). The distribution of products of beta, gamma and Gaussian random variables. *SIAM J. Appl. Math.* **18**, 721–737.

[177] Susko, E. and Roger, A. J. (2004). Estimating and comparing the rates of gene discovery and expressed sequence tag (EST) frequencies in EST surveys. *Bioinformatics.* **20**, 2279–2287.

[178] Tang, J., Gupta, A. K. (1984). On the distribution of the product of independent beta random variables. *Statist. Probab. Lett.* **2**, 165–168.

[179] Tavarè, S. (1984). Line-of-descent and genealogical processes, and their applications in population genetic models. *Theor. Popul. Biol.* **26**, 119–164.

[180] Thorin, O. (1977). On the infinite divisibility of the Pareto distribution. *Scand. Actuar. J.* 31–40.

[181] Thorin, O. (1977). On the infinite divisibility of the log-normal distribution *Scand. Actuar. J.* 121-148.

[182] Tweedie, R. L. (1975). Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stochastic Process. Appl.* **3**, 385–403.

[183] Tweedie, R. L. (1975). Criteria for classifying general Markov chains. *Adv. Appl. Probab.* **8**, 737–771.

[184] Tweedie, R. L. (1983). Criteria for rates of convergence of Markov chains, with application to queueing and storage theory. In *Probability, Statistics and Analysis.* Cambridge University Press. **8**, 737–771.

[185] Vervaat, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Adv. Appl. Probab.* **11**, 750–783.

[186] Von Renesse, M. K., Yor, M and Zambotti, L. (2008). Quasi-invariance properties of a class of subordinators. *Stochastic Process. Appl.* In press.

[187] Yakubovich, S. B., Luchko Y (1994). *The Hypergeometric Approach to Integral Transforms and Convolutions (Mathematics and Its Applications)*. Springer.

[188] Yamato, H. (1884). Characteristic functions of means of distributions chosen from a Dirichlet process. *Ann. Probab.* **12**, 262–267.

[189] Yamato, H. and Sibuya, M. (2000). Moments of some statistics of Pitman sampling formula. *Bull. Inform. Cybernet.*, **32** 1–10.

[190] Walker, S. G., Damien, P. (1998). A full Bayesian nonparametric analysis involving a neutral to the right process. *Scand. J. Statist.* 669–680.

[191] Walker, S. G. and Damien, P. (2000). Representation of Lévy processes without Gaussian components. *Biometrika.* **87**, 477–483.

[192] Walker, S. G., Damien, P., Laud, P. and Smith, A. F. M. (1999). Bayesian nonparamteric inference for random distributions and related functions. *J. Roy. Statist. Soc. Ser. B.* **61**, 485–527.

[193] Walker, S. G., Hatjispyros, S. J. and Nicoleris, T. (2007). A Fleming-Viot process and Bayesian nonparametrics. *Ann. Appl. Probab.* **17**, 67–80.

[194] Walker, S. G., Muliere, P. (1997). Beta-Stacy processes and a generalization of the Polya-urn scheme. *Ann. Statist.* **25**, 1762–1780.

[195] Watterson, G. A. (1976). The stationary distribution of the infinitely-many neutral alleles diffusion model. *J. Appl. Probab.* **13**, 639–651.

[196] WILKS, S. S (1962). *Mathematical Statistics.* John Wiley, New York.

[197] WILKS, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika.* **24**, 471–494.

[198] ZABELL, S.L. (1982). W. E. Johnson's "sufficientness" postulate. *Ann. Statist.* **10**, 1090–1099.