THESIS DECLARATION

The undersigned

SURNAME	Ventz
FIRST NAME	Steffen
PhD Registration Number	1369685

Thesis title:

Some Reinforced Stochastic Processes in Bayesian Statistics

PhD in	Statistics
Cycle	24
Candidate's tutor	Prof. Pietro Muliere
Year of discussion	2013

DECLARES

Under his responsibility:

1) that, according to the President's decree of 28.12.2000, No. 445, mendacious declarations, falsifying records and the use of false records are punishable under the penal code and special laws, should any of these hypotheses prove true, all benefits included in this declaration and those of the temporary embargo are automatically forfeited from the beginning;

2) that the University has the obligation, according to art. 6, par. 11, Ministerial Decree of 30th April 1999 protocol no. 224/1999, to keep copy of the thesis on deposit at the Biblioteche Nazionali Centrali di Roma e Firenze, where consultation is permitted, unless there is a temporary embargo in order to protect the rights of external bodies and industrial/commercial exploitation of the thesis;

3) that the Servizio Biblioteca Bocconi will file the thesis in its 'Archivio istituzionale ad accesso aperto' and will permit on-line consultation of the complete text (except in cases of a

temporary embargo);

4) that in order keep the thesis on file at Biblioteca Bocconi, the University requires that the thesis be delivered by the candidate to Società NORMADEC (acting on behalf of the University) by online procedure the contents of which must be unalterable and that NORMADEC will indicate in each footnote the following information:

- thesis (thesis title)
  ……
  ……
  ……
  …… ;
- by (candidate's surname and first name)
  …… ;
- discussed at Università Commerciale Luigi Bocconi – Milano in (year of discussion) ……;
- the thesis is protected by the regulations governing copyright (law of 22 April 1941, no. 633 and successive modifications). The exception is the right of Università Commerciale Luigi Bocconi to reproduce the same for research and teaching purposes, quoting the source;
- **only in cases where another declaration has been undersigned requesting a temporary embargo:** the thesis is subject to a temporary embargo for (indicate duration of the embargo) …… months;

5) that the copy of the thesis deposited with NORMADEC by online procedure is identical to those handed in/sent to the Examiners and to any other copy deposited in the University offices on paper or electronic copy and, as a consequence, the University is absolved from any responsibility regarding errors, inaccuracy or omissions in the contents of the thesis;

6) that the contents and organization of the thesis is an original work carried out by the undersigned and does not in any way compromise the rights of third parties (law of 22 April 1941, no. 633 and successive integrations and modifications), including those regarding security of personal details; therefore the University is in any case absolved from any responsibility whatsoever, civil, administrative or penal and shall be exempt from any requests or claims from third parties;

7) **choose hypothesis 7a or 7b indicated below:**

7ª) that the PhD thesis is not the result of work included in the regulations governing industrial property, it was not produced as part of projects financed by public or private bodies with restrictions on the diffusion of the results; it is not subject to patent or protection registrations, and therefore not subject to an embargo;

Or

7[b]) that the thesis meets one of the temporary embargo hypotheses included in the declaration **"TEMPORARY EMBARGO REQUEST OF THE PhD THESIS"** undersigned elsewhere.

Date 31.10.2012

SURNAME	Ventz	NAME	Steffen

Summary

In the present thesis we consider the application of stochastic processes with reinforcement and Bayesian methods to the analysis of discrete-time inhomogeneous Markov chains and continuous-times queuing processes. The aim is to use Pólya urn models and generalizations of such models to analyze the above processes from a predictive perspective. Pólya urn processes in discrete time were introduced under the name Reinforced urn processes (RUP) in [117]. In [118] the same authors extended to analysis to continuous time problems. These models provide simple tools for the analysis multiple-states processes and Semi-Markov processes, although the later approach in [118] is rather difficult to implement in practice. The thesis is divided into two parts and a preliminary chapter. The first preliminary chapter reviews some theory on Bayesian non-parametric inference and reinforced stochastic processes which possess a mixture representation. Both of these tools will be used in the next two parts. We de not intend to provide an exhaust review of the complete literature on Bayesian nonparametrics and reinforced stochastic processes. Especial the literature on reinforced stochastic processes developed extensively in a varying of different directions. Most of these developments are not necessarily directly related to Bayesian statistics, meaning that such processes do not have a mixture representation.

In part one of the thesis we introduce some predictive analysis of discrete-time inhomogeneous Markov chains using Pólya urn schemes. We focus on processes that appear in multiple-states models and system reliability. In chapter 3, the first chapter of part one, we define a particular RUP for the analysis of multiple-state processes. The multiple-states process is assumed to have a time-inhomogeneous dependence structure and a finite state-space.

In the remaining two sections of part one, we extend this analysis to bivariate and multivariate state-spaces. All chapters are papers currently under preparation for submission to a journal. This unfortunately causes extensive repetition of definitions and concepts. Also, since chapter 4 and chapter 5 are extensions of chapter 3, the structure of all three chapters will be identical. We first introduce a particulate Pólya urn process, study the mixture representation and the mixing measure. In a second step we use a sequence of hitting times of the process to construct an exchangeable sequences of random elements, where each random element is itself a mixture of a finite-length discrete-time inhomogeneous Markov chain. These random elements are used to derive predictive estimator for summary measures of time-inhomogeneous Markov chains.

In the second part of the thesis we analyze one-counter queuing models with Markov arrival processes using reinforced stochastic processes in continuous time and Bayesian nonparametric prior processes. In chapter 7, the first chapter of part two, we construct a continuous-time jump process with reinforcement in a predictive approach. The process has, under recurrence, a representation as a mixture of Semi-Markov processes. The random transition matrix of the embedded Markov chain can be expresses as a function of a discrete Beta-Stacy random probability. The random holding-time distribution function can be expresses in terms of two independent Beta-Stacy random distribution functions. We use the process to to approximate the embedded Semi-Markov process of the $M/G/1$ queue. This approach enables us to derive predictive estimator for some performance measures of queue. In the chapter 8 we apply Bayesian method to a semi-parametric analysis of the $M^X/G/1$ bulk queue where items enter a system in bathes. For the arrival rate, the bath size distribution and the service time density we chose a conjugated Gamma prior, a Beta-Stacy random probability and the non-conjugated Poisson-Dirichlet mixture models with a gamma-kernel. The posterior factorizes into three independent probabilities where the posterior service-time density and the posterior bath-size probability are approximated by MCMC simulations and a Bootstrap scheme. We predict basic performance measures of the $M/G/1$ queue in the transient- and steady-phase.

Contents

Acknowledgements

This is the end of a long journey, which leads me from a small rural island in the Baltic ocean, via Rostock (Germany) and Paris to Milan. Many people joined me on this journey on several occasions. First and foremost, I would like to thank my advisor Prof. Pietro Muliere for introducing me to the Italian school of Bayesian statistics and supporting me over the last four years. I am very fortunate to have an advisor who is always available and devotes some much effort toward his students.

I also want to thank Prof. James Vaupel and Prof. Gabriele Doblhammer-Reiter who facilitated the second part of my journey from Rostock (Germany) to Paris and who made me become interested in statistics and probability.

Special thanks go to Max and Polina for joining me on many hiking trips through the Alps. Your mental support helped me over many difficult moments and made me finishing this journey. I also want to thank Giuseppe, Maria Anna and Giulia for valuable discussions on statistical philosophy and for introducing me to the Italian way of life.

Most of all I want to thank my parents, Regina and Uwe Ventz, for their unconditional support and love during every part of my life, for just being there for me whenever I need them, and for making me the person I am today.

Finally I want to thank my sister, Sandra, and my grandparents, Charlotte and Werner Klitz for your warm friendship and for believing in me. I hope you will join me on my new journey, which now starts in Boston, and stay with me as long as God will.

Chapter 1

Some Preliminaries for Part I and Part II

1.1 Introduction and purpose

In the following chapter we introduce some key concepts and main results on Bayesian nonparametric statistics, mixtures of stochastic processes and a particular class of reinforced stochastic processes which we require in the later chapters. Hence we do not claim to give an objective and exhaust overview over all three topics, but focus on concepts important for this thesis. For an overview with a wider range of scope for each of the topics we refer to [75, 85, 125].

1.2 Bayesian non-parametric inference

In the Bayesian context the process of statistical inference may be summarized by two steps. Suppose we observe a random element $Z = (Z_v, v \in \mathcal{V})$ defined on a suitable probability space $(\Omega, \mathcal{F}, P)$ and taking values in $(\tilde{\mathcal{Z}} = \mathcal{Z}^{\mathcal{V}}, \mathbb{B}(\tilde{\mathcal{Z}}))$. In a first step, we describe Z by a sampling model $P(Z \in B|\theta)$ for measurable sets $B \in \mathbb{B}(\tilde{\mathcal{Z}})$, where the sampling model depends on a parameter $\theta \in \Theta \subset \mathbb{R}^d$ for $d \in \mathbb{N}$. In a Bayesian framework the model must now be completed by adding in a further step a prior distribution over the

parameter space, which yield the formulation

$$[Z|\theta] \sim P(dZ|\theta) \tag{1.1}$$

$$[\theta|\vartheta] \sim \Pi(d\theta|\vartheta). \tag{1.2}$$

In a hierarchical model one may add a further step and assumes $\vartheta \sim \Pi(d\vartheta)$. The object of a Bayesian inference will then simply be to estimate functions of θ and Z by using Bayes theorem

$$P(h(\theta, Z) \in A|Z) = \frac{\int_{h^{-1}(A)\cap(\Theta\times\tilde{\mathcal{Z}})} P(Z|\theta)\Pi(d\theta|\vartheta)\Pi(\vartheta)}{\int_{\Theta\times\tilde{\mathcal{Z}}} P(Z|\theta)\Pi(d\theta|\vartheta)\Pi(\vartheta)}. \tag{1.3}$$

Modern statistical problems can become very complex and it is not always possible to assume a parametric (finite dimensional) model for (1.1) or (1.2).

A Bayesian non-parametric (or infinite dimensional) model can relax the parametric assumptions by assuming that some subset of the parameter of θ and ϑ will be infinite dimensional. For example

$$P(dZ|\theta) = P(d(Z_v, v \in \mathcal{V}_1)|\theta_1, (Z_v, v \in \mathcal{V}_2)) \times \theta_2(d(Z_v, v \in \mathcal{V}_2)) \tag{1.4}$$

for a partition $\mathcal{V}_1$ and $\mathcal{V}_2$ of the index set $\mathcal{V}$ or

$$P(d\theta|\vartheta) = P(d\theta_1|\vartheta_1, \theta_2) \times \vartheta_2(d\theta_2) \tag{1.5}$$

for a partition of Θ into Θ_1 and Θ_2. In both cases θ_2 and ϑ_2 are probability measures on some subspaces of $\tilde{\mathcal{Z}}$ and Θ. The first case describes a situation where we are reluctant to specify a single parametric family of sampling models. The second case represents a situation where we do not trust a single parametric family of prior distributions. In both cases our parameters θ_2 and ϑ_2 belong to infinite-dimensional spaces. If we want to treat infinite dimensional objects such as probabilities as random elements, i.e. introducing a prior distribution on a space of functions, we need to treat such objects as stochastic processes.

1.2.1 Random probability measures

In the following I will focus on the case where the infinite dimensional random element will be a random probability, since I will only use such quantities in the later chapters. Let $\mathcal{X}$ denote a generic Polish space (a complete and separable metric set) with σ-algebra $\mathbb{B}$ and denote with $M(\mathcal{X})$ the set of all probability measures on $\mathcal{X}$. $M(\mathcal{X})$ can be made into a topological space through the topology of weak convergence generated by a neighborhood base of the from $B_\epsilon(P) = \{Q \in M(\mathcal{X}) : |Q(f_i) - P(f_i)| < \epsilon, f_i \in C(\mathbb{X} \text{ for } 1 \leq i \leq k\}$ for $\epsilon > 0$ and $P \in M(\mathcal{X})$ [70]. One of the cornerstones of Bayesian non-parametric statistics was the paper of Doksum [46] which was strongly influence by the work of [62, 47, 96, 31]. According to [46]

Definition 1 ([46])**.** *A $M(\mathcal{X})$-values stochastic process $P = (P(A),\ A \in \mathbb{B})$ is called a finite additive random probability if*

(i) $P(A) \in [0,1]$ *a.s. for all* $A \in \mathbb{B}$*;*

(ii) $P(\mathcal{X}) = 1$ *a.s. and*

(iii) for any $n \geq 1$ *and any finite class of disjointed sets* $(A_{i,j}, 1 \leq j \leq m_i)$ *for* $1 \leq i \leq n$

$$\begin{aligned} &\Big(P(\bigcap_{1\leq j\leq m_1} A_{1,j}), \cdots, P(\bigcap_{1\leq j\leq m_n} A_{n,j})\Big) \\ \stackrel{\mathcal{L}}{=} &\Big(\sum_{1\leq j\leq m_1} P(A_{1,j}), \cdots, \sum_{1\leq j\leq m_n} P(A_{n,j})\Big) \end{aligned}$$

where $\stackrel{\mathcal{L}}{=}$ *denoted equality in distribution.*

If in addition the stochastic process also satisfies

(iv) $P(A_k) \Rightarrow 0$ *whenever* $A_k \downarrow \emptyset$ *('$\Rightarrow$' denotes convergence in distribution),*

then the process P is called a random probability.

Ferguson [54], Lemma 1, showed that if one defines a finite additive random probability trough a system of finite dimensional laws for the process

$(P(A_1), \cdots, P(A_k))$, $k \geq 1$ for each measurable partition $(A_1, \cdots, A_k)$ of $\mathcal{X}$ such that the conditions (i)-(iii) are satisfied, then there exists a probability measure $\mathbb{P}$ on $M(\mathcal{X})$ [1] with the same system of finite dimensional law.

For random probability measures on ordered spaces, like the real line, Doksum clarified the relation between a random probability P and the corresponding random cumulative distribution function (cdf) $F = P((-\infty, \cdot])$ [2]. He showed ([46], Proposition 3.2) that, if one starts with a random cdf F on the real line and defines a random probability by an extension of $P_0((a, b]) = F(b) - F(a)$, then one obtains a probability $\mathbb{P}_0$ on $M(\mathbb{R})$, such that $P_0 \sim \mathbb{P}_0$. Now, if $F = P((-\infty, \cdot])$ then $\mathbb{P}_0$ agrees with $\mathbb{P}$, where $P \sim \mathbb{P}$. Hence I will sometimes switch between a property of a random probability measure and corresponding random cdf even those both random elements take values on a different space.

Neutral-to-the-Right random distribution functions

In the same paper Doksum [46] introduce a rich class of random probabilities on the real line $\mathcal{X} = \mathbb{R}$ which he called Neutral-to-the-Right (NTR) random probabilities. Roughly speaking a random probability measure P is NTR if the corresponding random cdf F has independent normalized increments. In a formal way this means that

Definition 2 ([46])**.** *The associated non-decreasing and right-continuous process $F = (F(t) = P(-\infty, t], t \in \mathbb{R})$ of a random probability P is NTR if for every n and every $t_1 < t_2 < \cdots < t_n < \infty$ there exist independent random variables $(V_j)_{1 \leq j \leq n}$ with values in $[0, 1]$ such that*

$$\Big(F(t_1), F(t_2), \cdots, F(t_n)\Big) \stackrel{\mathcal{L}}{=} \Big(V_1, 1 - (1 - V_1)(1 - V_2), \cdots, 1 - \prod_{j \leq k}(1 - V_j)\Big). \tag{1.6}$$

[1] More formal $\mathbb{P}$ is defied on $[0, 1]^{\mathbb{B}}$ with σ-fields generated by the cylinder sets.

[2] Any random element $X = (X(t), t \in \mathbb{R}^d)$ with sample paths that are a.s. (i) non-decreasing, (ii) right-continuous and (iii) with limits 0 (and 1) if $t_i \to -\infty$ for some $1 \leq i \leq d$ (and $t_i \to +$ for all $1 \leq i \leq d$) will be called a random cdf on $\mathbb{R}^d$)

If we forget for the moment a possible division by zero, then one can informally state the NTR property as

$$\left(F(t_1), \frac{F(t_2)-F(t_1)}{1-F(t_1)}, \cdots, \frac{F(t_n)-F(t_{n-1})}{1-F(t_{n-1})}\right) \stackrel{\mathcal{L}}{=} (V_1, \cdots, V_n), \tag{1.7}$$

which means that the associated hazard process $H(dt) = F(dt)/(1-F(t))$ has independent increments. In fact Doksum showed that

Theorem 1 ([46], Theorem 3.1)**.** *A random cdf F (and hence P) is NTR if and only if $F \stackrel{\mathcal{L}}{=} 1-\exp\{-Z\}$, where Z is a non-decreasing, independent increment process (IIP) with right-continuous sample paths, which starts at zero a.s. and converges to $+\infty$ with probability one as $t \to +\infty$*[3]*.*

The definition of a NTR processe is very intuitive, but hard to work with directly. Instead theorem 1 gives a very suitable equivalent definition of NTR processes which is simpler to use in the context of Bayesian nonparametrics. [4] Any IIP can be decomposed into a deterministic part and two pure jump processes [5] (see for example [57, 55, 149])

$$Z \stackrel{\mathcal{L}}{=} D + Z_d + Z_c = D(\cdot) + \sum_{t_j \in \mathcal{J}} Z_{d,j}\delta_{t_j} + \sum_j Z_{c,j}\delta_{T_j}. \tag{1.8}$$

The jump process Z_d has jumps at a countable number of fixed locations $\mathcal{J} = \{t_j\}$ of random size $Z_{j,d} \sim f_j(\cdot)$. Whereas the jump process Z_c has jumps only at random locations T_j of random size $Z_{c,j}$ and is characterized by the Laplace transform

$$\mathbb{E}[\exp\{-vZ_c(t)\}] = \exp\{-\int_0^\infty (1-e^{-vs})v(ds,t)\} \quad t \in \mathbb{R} \tag{1.9}$$

with Levy measure $v(ds,dt)$ such that $\int_{[0,\infty)} s(1-s)^{-1}v(ds,t) < \infty$ for all

[3] Note that Z cannot be interpreted as a random cumulative hazard process since F (and Z) has a.s. discrete sample paths.

[4] Although the previous theorem was originally stated for NTR processes on $\mathcal{X} = \mathbb{R}$ it is usually quoted for $\mathcal{X} = [0,\infty)$ since NTR processes are mainly used in survival analysis where $[0,\infty)$ is the natural space for failure times.

[5] The Gaussian part needs to be zero since an IIP non-decreasing

$t > 0$ (see [55]). Since F is supposed to be a random cdf the deterministic part D is not very interesting and is usually set to zero.

For the case $\mathcal{X} = [0, \infty)$ and $v([0,t], \mathbb{R}_+) < \infty$ for every $t > 0$ it is possible to write Z_c directly as a function of a Poisson random measure (PRM), say N, on $\mathbb{R}_+^2$ with intensity $v(ds, dt)$

$$Z_c(t) = \int_{[0,t]} \int_{\mathbb{R}_+} N(ds, dt) = \sum_{s \leq t : \Delta T(s) = 1} J(s) \tag{1.10}$$

where T is a inhomogeneous Poisson process on $[0, \infty)$ with rate $\lambda(t)dt = v(\mathbb{R}_+, dt)$ and compounding variables

$$[J(t) | \Delta T(t) = 1] \overset{ind.}{\sim} \frac{v(ds, dt)}{\lambda(t)} \tag{1.11}$$

(see for example [92, 100]).

If $v([0,t], \mathbb{R}_+) = \infty$ this relation does not hold. But in some cases the former relation holds in a limit-sense. This means that it is possible to find a PRM, say N_n, with Levy intensity $v_n(\cdot, \cdot)$ satisfying $v_n((0,t], \mathbb{R}_+) < \infty$ for all t such that $Z_{c,n} = \int_{[0,\cdot]} \int_{\mathbb{R}_+} N_n(ds, d\cdot)$ converges weakly to Z_c in the Skorohod topology [92, 100].

The representation of NTR processes as exponential of an IIP gives a simple way to defined prior probabilities on the spaces $M([0, +\infty))$. James [81] extended this approach via PRM's to prior probabilities on $M(\mathbb{R}_+ \times \mathcal{X})$ for arbitrary Polish spaces $\mathbb{X}$. All one needs is to ensure that there exists an IIP, Z say, or a PRM, N say, such that every sample path of F is a.s. a cdf.

Example 1.2.1 (Dirichlet Process, [54])**.** *The first prior probability on a space of probability measures $M(\mathcal{X})$ introduced in Bayesian analysis was the Dirichlet process, with can be define on more general spaces then $M(\mathbb{R})$. For $\mathcal{X} = \mathbb{R}$ choose an absolutely continuous probability P_0 on $\mathbb{R}$ and define the diffuse measure $\alpha = MP_0$ with total mass $M = \alpha(\mathbb{R})$. The random cdf of a Dirichlet process $F = 1 - \exp\{-Z_c\}$ with parameters (M, α) is NTR, where*

Z_c has Levy measure

$$v(ds, dt) = \frac{e^{-s\alpha(t,\infty)}}{1 - e^{-s}} ds\alpha(dt). \tag{1.12}$$

Example 1.2.2 (Beta-Stacy Process,[150])**.** *Let $\mathcal{X} = [0, \infty)$ and suppose α is a finite Borel measure on $[0, \infty)$ which can be decomposed into continuous part α_c and discrete part α_d with countable discontinuity set $\mathcal{J}$. Let β a positive piecewise continuous function on $[0, \infty)$. Then a random cdf $F = 1 - \exp\{-Z_d - Z_c\}$ is called a beta-Stacy NTR process with parameters $(\alpha, \beta\mathcal{J})$ if the jump process of $Z_d = \sum_{t_j \in \mathcal{J}} J_{j,d}\delta_{t_j}$ has jump size*

$$1 - e^{-J_{j,d}} \sim Beta(\alpha_d\{t_j\}, \beta t_j)) \tag{1.13}$$

at $t_j \in \mathcal{J}$ and the IIP process Z_c has Levy measure

$$v(ds, dt) = \frac{e^{-s\beta t)}}{1 - e^{-s}} \alpha_c(dt). \tag{1.14}$$

If $\beta t) = MP_0(t, \infty)$ and $\alpha_c = MP_0$, then the beta-Stacy process reduces to a Dirichlet process on $[0, \infty)$. Hence on the positive real line the beta-Stacy process generalizes the Dirichlet process. Note that this is generally true. A Dirichlet random measure can be defined on any Polish space, but the beta-Stacy random cdf can only be defined on the positive real line.

Now suppose that $D_n = (Y_1, \cdots, Y_n)$ constitutes a potential right-censored sample from F where F is NTR with Levy measure v, discontinuity set $\mathcal{J} = \{t_j\}$ and jump size distribution f_j for $t_j \in \mathcal{J}$ [6]. We want to learn from the data D_n about the functional form of random cdf F, therefore we are interested in the posterior law $\mathcal{L}(F|D_n)$.

Theorem 2 ([46, 55, 58])**.** *Suppose that D_n is a potentially right censored sample from a NTR random cdf F with Levy measure v, fixed discontinuity set $\mathcal{J}$ and jump size density f_j at $t_j \in \mathcal{J}$, short NTR$(v, \mathcal{J}, f_j)$. Then $F_n =$*

[6] A right censored sample means that $Y_i = \min(X_i, C_i)$ where $X_i \sim F$ and C_i is a censoring time independent of F with censoring indicator $\delta_i = I(X_i \leq C_i)$.

$[F|D_n]$ is again NTR with parameters $(v_n, \mathcal{J}_n, f_{j,n})$ where

$$v_n(ds, dt) = e^{-sY_n(t)} v(ds, dt); \tag{1.15}$$

with discontinuity set $\mathcal{J}_n = \mathcal{J} \cup \{Y_i : \delta_i = 1\}$ and jump size distribution

$$f_{j,n}(s) \propto \begin{cases} e^{-s[Y_n(t_j) - \Delta N_n(t_j)]}(1 - e^{-s})^{\Delta N_n(t_j)} f_j(s) & \text{if } t_j \in \mathcal{J} \\ e^{-s[Y_n(t_j) - \Delta N_n(t_j)]}(1 - e^{-s})^{\Delta N_n(t_j)} v(ds, dt_j) & \text{if } t_j \in \mathcal{J}_n \setminus \mathcal{J} \end{cases} \tag{1.16}$$

where $Y_n(\cdot) = \sum_{1 \le j \le n} I(Y_j \ge \cdot)$ and $N_n(\cdot) = \sum_{1 \le j \le n} I(Y_j \le \cdot, \delta_j = 1)$.

Example 1.2.3 (Posterior of a beta-Stacy process)**.** *For the beta-Stacy process F defined in example 1.2.2 the previous theorem shows that F_n has Levy measure*

$$v_n(ds, dt) = \frac{e^{-s(Y_n(t) + \beta(t))}}{1 - e^{-s}} ds \alpha_c(dt). \tag{1.17}$$

Furthermore the jump size density is give by

$$f_{j,n} \propto (1 - e^{-s})^{\Delta N_n(t_j) + \alpha\{t_j\} - 1} e^{-s(\beta(t_j) + Y_n(t_j) - \Delta N_n(t_j))} \tag{1.18}$$

for $t_j \in \mathcal{J}_n$ or similar

$$1 - e^{-Z_{n,d,j}} \sim beta(\Delta N(t_j) + \alpha\{t_j\}, \beta(t_j) + Y_n(t_j) - \Delta N(t_j)). \tag{1.19}$$

Hence F_n is again a beta-Stacy process with parameters $\beta + Y_n$ and $\alpha_n = \alpha_{d,n} + \alpha_c$ where $\alpha_{d,n} = \alpha_d + \sum_{j:\delta_j = 1} \delta_{Y_j}$.

Example 1.2.4 (Posterior of a Dirichlet process)**.** *For the Dirichlet process, give D_n, the Levy measure is given by*

$$v_n(ds, dt) = \frac{e^{-s(Y_n(t) + \alpha(t, \infty))}}{1 - e^{-s}} ds \alpha(dt) \tag{1.20}$$

and the jump sizes of $Z_{n,d}$ have density

$$1 - e^{-Z_{n,d,j}} \sim beta(\Delta N(t_j), Y_n(t_j) - \Delta N_n(t_j) + \alpha(t_j, +\infty)). \tag{1.21}$$

Without censoring the counting process equals $\Delta N(t_i) = \sum_{1 \leq j \leq n} \delta_{Y_j}(t_i)$ and the measure α is simply updated to $\alpha_n = \alpha + \sum_j \delta_{Y_j} = (M+n)(P_0 + \hat{P}_{empirical})$ and F_n is again a Dirichlet process with parameter $(M+n, \alpha_n)$.

If there are censored observations $\Delta N(t_i) = \sum_{1 \leq j \leq n: \delta_j = 1} \delta_{Y_j}(t_i)$, hence the random cdf F_n is not a Dirichlet process any more. But since theorem 2 shows that F_n is NTR, the posterior process F_n must belong to some class of NTR processes. The previous example showed that F_n is a beta-Stacy process with parameters $\beta_n(t) = Y_n + \alpha(t, \infty)$ and $\alpha_n = \alpha_{n,d} + \alpha_c = N_n + \alpha$.

The last example shows that conjugated NTR prior can be separated into 'parametric' and 'structural' conjugated processes [101]. For a 'parametric' conjugated prior, the posterior can be found by updating the prior parameters by some summary statistic of the data. A 'structural' conjugated prior yield a posterior which belongs again to the class of all NTR random cdf's, but does not necessarily belong to the same sub-class of NTR processes as the prior did [101].

Stick-breaking prior and related random probabilities

The concept of NTR random cdf does not carry over very nicely to more general spaces $\mathcal{X}$ as the real line. An exception is the case $\mathcal{X} = [0, \infty) \times \mathcal{Z}$ where $\mathcal{Z}$ is an arbitrary Polish space [81]). Ferguson [54] defined a Dirichlet process $P = (P(A), A \in \mathbb{B}_{\mathcal{X}})$ on more abstract spaces in the following way.

Definition 3 ([54])**.** *Let $\alpha = MP_0$ be a finite measure on a Polish space $\mathcal{X}$ where P_0 is a fixed probability measure. A $M(\mathcal{X})$-valued stochastic process is called a Dirichlet random probability on $\mathcal{X}$ if for every measurable partition $A_1, \cdots, A_n$ of $\mathcal{X}$ the random vector $(P(A_1), \cdots, P(A_n))$ has a Dirichlet distribution with parameters $(\alpha(A_1), \cdots, \alpha(A_k))$.*

Ferguson [54] showed that, given a sample of uncensored observations $D_n = (X_1, \cdots, X_n)$, the posterior process $P_n := (P|D_n)$ is again a Dirichlet

process with $\alpha_n = (M+n)(P_0 + \sum_{1\leq j\leq n} \delta_{X_j})$. Clearly this result agrees with Theorem 2 if $\mathcal{X} = \mathbb{R}$.

A constructive equivalent definition of the Dirichlet process was given by Sethuraman [142], he established a representation of the Dirichlet process as

$$P(\cdot) = \sum_{j\geq 1} w_j \delta_{X_j}(\cdot) \tag{1.22}$$

where $(X_j)_{j\geq 1}$ is an iid sample from P_0 and $w_j = \beta_j \prod_{l<j}(1-\beta_l)$ with β_j iid beta$(1, M)$ distributed for all $j \geq 1$. The representation (1.22) is usually called the stick-breaking representation of the Dirichlet process. The representation of the Dirichlet process inspired [78] to introduce a quite general class of discrete random probability measures, called stick-breaking prior.

Definition 4 ([78])**.** *Let P_0 be a non-atomic probability measure on a Polish space $\mathcal{X}$ and $a = (a_k)_{k\geq 1}$ and $b = (b_k)_{k\geq 1}$ be two sequences of non-negative real numbers. Then a $M(\mathcal{X})$-valued stochastic process P is called a stick breaking random probability measure if*

(i) $P(\cdot) = \sum_{j\geq 1} p_j \delta_{X_j}(\cdot)$ where $(X_j)_{j\geq 1}$ is an iid sample from P_0;

(ii) the sequence of random weights $(p_j)_{j\geq 1}$ has a stick-breaking representation

$$p_j = \xi_j \prod_{l<j}(1-\xi_l) \tag{1.23}$$

where ξ_j are independent beta(a_j, b_j) random variables and

a and b are chosen such that $\sum_{j\geq 1} p_j = 1$ a.s..

Iswaran and James [78] supplied a simple condition

$$\sum_{j\geq 1} \log(1 + a_k/b_k) = +\infty \tag{1.24}$$

which ensures that the random weights sum to one almost surely.

Example 1.2.5 (Dirichlet process)**.** *For example the Dirichlet process can be obtained as a special case for $a_k = 1$ and $b_k = M$ for all $k \geq 1$.*

Example 1.2.6 (Two parameter Poisson-Dirichlet process)**.** *Another famous random probability, called the 2-parameter Poisson-Dirichlet process [128] with parameter $a \in [0, 1)$ and $b > -a$ arises if $a_k = 1 - a$ and $b_k = b + ka$.*

1.3 Markov-/Exchangeable stochastic processes

1.3.1 Exchangeable processes

A more foundational concept of classical Bayesian statistics is the concept of exchangeability which date back to work of de Finetti in the 1920s and 1930s [37]. Again assume that $X = (X_n)_{n\geq 0}$ is a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values on a Polish space $(\mathcal{X}, \mathbb{B})$. For simplicity we use a coordinate representation of X by taking $\Omega = \mathcal{X}^\infty$ where $\mathcal{F}$ is generated by the cylinder sets and let $X(\omega)_n = \omega_n$ for $\omega = (\omega)_{n\geq 0} \in \Omega$ (see [22] p.21-22). The following overview follows closely the work of [95, 5].

An exchangeable sequence of random elements can be seen as a sequence of random variables such that the joined probability law of the process is invariant under a finite permutations of coordinates. The invariance property can be studied for finite and infinite sequences of random elements.

Definition 5. *(i) A finite sequence of random elements $(X_0, \cdots, X_n)$ is called (n+1)-exchangeable if*

$$(X_0, X_1, \cdots, X_n) \overset{\mathcal{L}}{=} (X_{\sigma(0)}, X_{\sigma(1)}, \cdots, X_{\sigma(n)}) \tag{1.25}$$

for every permutations σ of $\{0, \cdots, n\}$.

(ii) An infinite sequence of random elements $(X_j)_{j\geq 0}$ is called exchangeable if

$$(X_0, X_1, \cdots) \overset{\mathcal{L}}{=} (X_{\sigma(0)}, X_{\sigma(1)}, \cdots) \tag{1.26}$$

for all finite permutations σ of the non-negative integers $\mathbb{N}_0$.

An exchangeable structure can be generated in a variety of ways. A few simple ways to generate such processes are summarized in the following

Example 1.3.1. *(i) Let $(X_i)_{i\geq 0}$ be an infinite iid sequence of random variables on $\mathcal{X} = \{1, \cdots, K\}$ and assume $P(X_i = k) = p_k > 0$ for all $i \geq 0$ such that $\sum_{1\leq k\leq K} p_k = 1$. This gives an urn model with sample with replacement and produces trivially an infinite exchangeable sequence.*

(ii) Let $(X_i)_{0\leq i\leq n}$ be a sequence of random variables on $\mathcal{X} = \{0, 1\}$ and assume $c_0 = m, c_1 = v$ for some positive integers such that $m + v = n+1$. Now draw successively and without replacement labels $x_i \in \{0, 1\}$ form an urn with an initial composition of c_0 balls of color 0; c_1 balls of color 1 and set $X_i = x_i$. Then $(X_i)_{0\leq i\leq n}$ is a finite $(n+1)$-exchangeable but not an iid sequence.

(iii) Let $\mathcal{X} = M_P(\mathbb{R}_+)$ be the space of all Radon point measure on the positive real line (see [133] for a construction of a topology on $M_P(\mathbb{R}_+)$ and the induced Borel σ-field). Choose a random probability P_0 on $(\mathbb{R}_+, \mathbb{B}(\mathbb{R}_+))$ such that $P_0(\{0\}) = 0$. For $n \geq 0$ let

$$\lambda_n \overset{iid}{\sim} P_0 \tag{1.27}$$

$$[(Y_{i,n})_{i\geq 1}|\lambda_n] \overset{iid}{\sim} Exp(\lambda_n) \tag{1.28}$$

$$X_n = \sum_{i\geq 1} \delta_{S_{i,n}}(\cdot) \tag{1.29}$$

where $S_{i,n} = \sum_{v\leq i} Y_{v,n}$. Then $(X_n)_{n\geq 0}$ is an exchangeable sequence of Point processes which are not independent but identical distributed.

Remark 1. *Exchangeability itself is a theoretical concept which is rather hard to check in practice. In probability theory this concept is well established and deeply studied [85, 5]. But in the context of statistical modeling a 'true' sampling model is usually not known. In this context one needs to be careful in judging a sequence of random objects to be exchangeable or not. For example, it would be silly to assume that the propensity to budget deficit of all EU states is exchangeable. Also, judging all US states to be exchangeable may not be a*

good idea either [68]. On the other side exchangeability is a very important concept when no further background information are available to distinguish (X_n) into subgroups. In such cases a coherent statistical model must assume probabilistic symmetry among random elements (see [69], pp.121).

The importance of Exchangeability in statistics come from the fact that an infinite exchangeable sequence can be related via de Finetti's theorem to Bayesian inference for an iid sampling model with unknown distribution function. To make this statement more precise, following Aldous [5], we define the concept of conditional independence give a sub-sigma field as

Definition 6 (Mixtures of iid)**.** *(i) An infinite sequence of random elements $X = (X_n)_{n\geq 0}$, is called a mixture of iid elements with directing random probability measure P if the power probability P^∞ is a regular conditional probability for X given $\mathcal{H} = \sigma(P)$. That is for all $n \geq 0$ and all $A_i \in \mathbb{B}$*

$$\mathbb{P}(\cap_{0\leq i\leq n}\{X_i \in A_i\}|\mathcal{H}) = \prod_{0\leq i\leq n} P(A_i) \quad a.s.. \tag{1.30}$$

(ii) Let $\mathcal{G}$ be a sub-sigma file of $\mathcal{F}$, then $X = (X_n)_{n\geq 0}$ are conditional iid given $\mathcal{G}$ if $\mathbb{P}(X_i \in A|\mathcal{G}) = \mathbb{P}(X_j \in A|\mathcal{G})$ a.s. for all $i \neq j$ and all $A \in \mathbb{B}$, and for all $n \geq 1$ and all $A_i \in \mathbb{B}$

$$\mathbb{P}(\cap_{0\leq i\leq n}\{X_i \in A_i\}|\mathcal{G}) = \prod_{0\leq i\leq n} \mathbb{P}(X_i \in A_i|\mathcal{G}). \tag{1.31}$$

The first part of the definitions is equivalent to requiring that, given $\mathcal{H}$, $(X_0, \cdots, X_n)$ are independent for each $n \geq 1$ and each $[X_i|\mathcal{H}] \sim P$ a.s.. The second definition does not specify the directing random probability explicitly, but the random probability can (at least in theory) be recovered as seen in the next lemma. For de Finetti's theorem (see below) it is very useful to know where the directing random probability comes from.

Lemma 1 ([5])**.** *Let X be a conditional iid sequence of random elements with directing random probability P. Then the directing random measure P is a.s. equivalent to the following random probability measures*

(i) $\hat{P}_\infty = \lim_n n^{-1} \sum_{0 \le i \le n-1} \delta_{X_i}$

(ii) $\mathbb{P}(X_0|\mathcal{F}_\infty)$

(iii) $\mathbb{P}(X_0|\mathcal{F}_n)$

where $\mathcal{F}_n = \sigma(\{X_j\}_{j>n})$, $\mathcal{F}_\infty = \cap_{n\ge 1}\mathcal{F}_n$ *denotes the tail* σ*-field of* X *and the limit in* (i) *is assumed to be the weak limit of the empirical probability measure.*

A remarkable result first proved in its most generality by Hewitt and Savage [73] states that every exchangeable sequence of random variables can be represented as a mixture of an iid sample directed by some random probability P.

Theorem 3 (De Finetti's theorem)**.** *An infinite exchangeable sequence* $X = (X_n)_{n\ge 0}$ *with values on a Polish space* $(\mathcal{X}, \mathbb{B})$ *is a mixture of iid random variables.*

Remark 2. *de Finetti's theorem can be interpreted by lemma 1 in several ways. The simplest way of restating the theorem is that, given the tail* σ*-field of* X*,* $\mathcal{F}_\infty$*, all random elements* $X_0, X_1, \cdots$ *are independent and identical distributed. Or equally stated: all information contained in* $\mathcal{F}_\infty$ *are sufficient and necessary to make the random elements independent. Hence, given* $\mathcal{F}_\infty$*, we cannot learn anything about* X_n *from* $X_i, i \ne n$*, since all relevant information from* X_i *are already summarized in the sufficient statistics* $\mathcal{F}_\infty$*. The directing random probability (sometimes called de Finetti measure) is simple the conditional probability* $P(\cdot|\mathcal{F}_\infty)$*.*

Remark 3. *de Finetti's theorem holds on much more general spaces than Polish spaces. A simple change of variable argument can be used to show that de Finetti's theorem holds as well on any standard space* [7] *by using the fact that such spaces are isomorphic to some Borel-subset of the real line.*

[7] A standard space $(S, \mathcal{S})$ is a space where $\mathcal{S} = \sigma(\mathcal{A})$ can be generated by a standard filed $\mathcal{A} = \bigcup_n \mathcal{A}_n$ from an increasing sequence of finite fields $\mathcal{A}_n$, such that any decreasing sequence of atoms $V_n \in \mathcal{A}_n$, $V_{n+1} \in V_n$ has non-empty intersection $\cap_n V_n$ (see [71]), Chapter 2.2.

Diaconis and Freedman [43] showed a more general version of de Finetti's theorem for compact Hausdorf spaces which are not necessarily standard.

Remark 4. *de Finetti's theorem does not hold on every space. Durbins and Freedman [48] gave an example of an exchangeable sequence on a separable metric space where de Finetti's theorem does not hold.*

Remark 5. *De Finetti's theorem justifies in an abstract sense the use of prior probabilities on quite general spaces since one can always write down a mixture representation of the finite dimensional law*

$$P(\cap_{0\leq i\leq n}\{X_i \in A_i\}) = \mathbb{E}_\mu\Big[\prod_{1\leq i\leq n} P(A_i)\Big]. \tag{1.32}$$

where μ is a 'prior' probability measure on $M(\mathcal{X})$. Still, we remark that the sequence X must actually be of infinite length which cannot be assumes for every statistical phenomena. There are for example only 50 US states and not an infinite number of states. Furthermore in order to check exchangeability one would need to know the exact form of the finite dimensional law on the left hand side of (1.32). But this probability is usually rather assumed (via the right hand side) than known.

One point mentioned in the last remark can be relaxed in an approximate way. A quite general result due to Diaconis and Freedman [43] states that the finite dimensional law of a finite but large sequence of exchangeable elements can be approximate by some mixture probability.

In particular for a $P \in M(\mathcal{X})$ let $P^k \in M(\mathcal{X}^k)$ be the k-power probability of P and for a probability μ on $M(\mathcal{X})$ define as in [43] a mixture probability $P_{k,\mu} \in M(\mathcal{X}^k)$ by

$$P_{k,\mu}(\cdot) = \int_{M(\mathcal{X})} P^k(\cdot)\mu(dP). \tag{1.33}$$

If P_n denotes a probability on $\mathcal{X}^n$ denote with $P_{n(k)} = P_n(\cdot \times \mathcal{X}^{n-k})$ the projection of P_n onto $\mathcal{X}^k$.

Theorem 4 ([43])**.** *Let $(X_i)_{0\leq i\leq n-1}$ be an n-exchangeable sequence of random elements on $\mathcal{X}$ with joined probability P_n. Then there exists a probability μ*

on $M(\mathcal{X})$ *such that for any* $k \leq n-1$

$$||P_{n(k)} - P_{k,\mu}||_{TV} \leq \frac{k(k-1)}{n} \tag{1.34}$$

where $||P_1 - P_2|| = \sup_{A\in\mathbb{B}} |P_1(A) - P_2(A)|$ *denotes the total-variation metric.*

In terms of statistical modeling the last theorem says that, if you have observations $X_0, X_1, \cdots, X_k$ from a finite, but large exchangeable total population $X_0, \cdots, X_k, \cdots X_n$, then the finite dimensional law $P_{n+1(k+1)}$ of the sub-sequence $(X_i)_{0\leq i\leq k}$ can be model quite accurately by some mixture probability with an approximation error of less than $k(k+1)/(n+1)$.

1.3.2 Partial exchangeable processes

Suppose the stochastic elements $(X_i)_{i\geq 0}$ represent the evolution of a random phenomenon over time, then the judgement of exchangeability will most likely be not sufficient to catch most of the statistical properties of this phenomena. This is due to the fact that, conditional on the tail sigma field of the process, the evolution in time would be completely independent. Next to an iid assumption the most elementary dependence structure in time would be Markov. A Markov dependence structure will most likely be still unrealistic but preferable to an iid assumption. A generalization of exchangeability from mixtures of iid sequences to mixtures of Markov sequences was introduced by Freedman [63] and studied in deep in [44]. In the following the stochastic process X takes values in a complete countable metric space $\mathcal{X} = (x_j)_{j\geq 1}$. For the concept of mixtures of Markov sequences on general Polish spaces see Fortini el al. [60] and Freedman [65, 66].

The main idea of characterizing mixtures of Markov processes is motivated by the following simple fact. Suppose a stochastic process X with values in $\{x_j\}$ is Markov with stochastic matrix Π and initial distribution λ. Then for any n and any tuple $v_{0:n} = (v_j)_{0\leq j\leq n}$ of states in $\{x_j\}$

$$P_{\Pi,\lambda}(v_{0:n}) := \mathbb{P}(\cap_{0\leq j\leq n}\{X_j = v_j\}|\Pi,\lambda) = \lambda_{x_0}\prod_{i,j\in\mathcal{X}} \pi_{i,j}^{n(i,j)} \tag{1.35}$$

where $n(i,j) = \sum_{1 \le t \le n} I(v_{t-1} = i, v_t = j)$ denote the number of transitions from i to j. Hence any two sequences of states τ and η with the same number of transitions from i to j for all $i, j \in \mathcal{X}$ and the same initial states will be equal likely. Such sequences are called equivalent, denoted by $\tau \sim \eta$. Clearly the same property still holds if we integrate over (1.35) with respect to μ, where μ represents a probability on the space of all transition matrices on $\{x_j\}$, say $M_M(\mathcal{X})$, i.e.

$$P_{\mu,\lambda}(v_{0:n}) = \int_{M_M(\mathcal{X})} P_{\Pi,\lambda}(v_{0:n}) \mu(d\Pi). \tag{1.36}$$

Freedman [63] expressed this invariance property with respect to equivalent sequences $\tau \sim \eta$ in terms the sufficient statistics $T_n(\tau_{0:n}) = (\tau_0, n_\tau(i,j) : i, j \in \mathcal{X})$. He calls a probability $\mathbb{P}$ of the process X "summarized by (T_n)" if $T_n(\tau_{0:n}) = T_n(\eta_{0:n})$ implies that the projection of $\mathbb{P}$ satisfies

$$\mathbb{P}(\cap_{0 \le j \le n} \{X_j = \tau_j\}) = \mathbb{P}(\cap_{0 \le j \le n} \{X_j = \eta_j\}). \tag{1.37}$$

Since $T_n(\tau_{0:n}) = T_n(\eta_{0:n})$ implies by definition $\tau_{0:n} \sim \eta_{0:n}$ the following definition of Diaconis and Freedman describes the same idea.

Definition 7 (Partial Exchangeable [44])**.** *A stochastic process X defined on a Polish space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a countable space $\mathcal{X} = \{x_i\}$ is called partial exchangeable if*

$$\mathbb{P}(\cap_{0 \le j \le |\tau|} \{X_j = \tau_j\}) = \mathbb{P}(\cap_{0 \le j \le |\eta|} \{X_j = \eta_j\}) \tag{1.38}$$

for all equivalent sequences $\tau \sim \eta$ of state in $\mathcal{X}$.

Clearly, by the previous discussion any discrete time, discrete state Markov process (we will call such a process a Markov chain) is partial exchangeable. Also any process with finite dimensional law expresses as a mixture of Markov processes as defined in (1.36) is partial exchangeable. Writing $\mu = \delta_\Pi(\cdot)$ for $\Pi \in M_M(\mathcal{X})$ the first case can be seen as a special case of the second case.

In general not every partial exchangeable process can be expresses as a mixture of Markov chains with finite dimensional law given by (1.36). But if

one adds the assumption of ('weak') recurrence, i.e.

$$\mathbb{P}(X_n = X_0 \ i.o.[n]) = 1, \tag{1.39}$$

then only processes with finite dimensional laws of the form (1.36) can be partial exchangeable.

Theorem 5 (Representation theorem,[44])**.** *Suppose that a stochastic process X satisfies $\mathbb{P}(X_n = X_0 \ i.o.[n]) = 1$. Then X is a mixture of Markov chains if and only if X is partial exchangeable.*

Remark 6. *Freedman [63] proved essentially the same result, but since his proof was based on ergotic limit theory applied to ergodic statistics he requires the stochastic process to be stationary. Unless a Markov chains starts at $n = 0$ with its invariant distribution, a Markov chain will in not generally be stationary. The recurrence condition in [44] relaxes this assumption. The proof in [44] shows the existence of the directing random transition matrix by using the exchangeability of the blocks of states between successive returns to the initial state X_0. As long as $X_n = X_0$ infinitely often, this is sufficient to ensure the existence of some directing random measure for such blocks, which yields by some smart logical considerations the existence of the random directing transition matrix. This means, that the tail sigma-field of this blocks make the process a Markov chain.*

Mixtures of Markov chains - Alternative Characterization

An alternative access to mixture of Markov chains was given by Fortini et al. [60] in terms of an infinite sequence of successor states for all states x in $\mathcal{X} = \{x_i\}$. The idea is again intuitive to understand. Any Markov chain with stochastic matrix Π can write as $[X_{n+1}|X_n] \stackrel{\mathcal{L}}{=} f(U_{n+1}, X_n)$ where $(U_n) \stackrel{iid}{\sim} \text{Uniform}(0,1)$ and

$$f(u, x_i) = \sum_{j \in \mathcal{X}} x_j I\Big(\sum_{k \leq j-1} \Pi_{i,k} < u \leq \sum_{k \leq j} \Pi_{i,k} \Big) \tag{1.40}$$

(see for example [132]). Therefore the successor sequence of a recurrence state i is an infinite iid sequence which is equal in distribution to $\{f(U_n, i)\}_n$. Intuitively, integrating over the finite dimensional law of a Markov chain with a probability on $M_M(\mathcal{X})$ should therefore produce a sequences of successor states that are mixture of iid, i.e. exchangeable successors to state i.

Extent the state-space to $\mathcal{X} = \{x_i\} \cup \{\Upsilon\}$ where Υ will be used as a dummy for successors of transitive states and denote with $\mathcal{P}(\mathcal{X})$ the power sigma algebra of $\mathcal{X}$. Let $\mathcal{K} \subset M_M(\mathcal{X})$ be the subset of all stochastic matrices on $\mathcal{X}$ such that, if $\Pi \in \mathcal{K}$, then there exists a recurrence set of states A_Π, depending on Π, containing x_0 but not Υ, with $\Pi_{y,\Upsilon} = 1, \Pi_{x,y} = 0$ for all $x \in A_\Pi$ and $y \notin A_\Pi$.

Furthermore, for every $x \in \mathcal{X}$ define as usual the sequience of hitting time to x by $\tau_x(n) = \inf\{n > \tau_x(n-1) : X_n = x\}$ where $\inf \emptyset = +\infty$ and define the matrix of successor states $V = (V_x(n); x \in \mathcal{X}, n \geq 1)$ as

$$V_x(n) = X_{\tau_x(n)+1} I(\tau_x(n) \in \mathbb{N})) + \Upsilon I(\tau_x(n) = +\infty) \quad x \in \mathcal{X}. \tag{1.41}$$

Definition 8. *the matrix of successor states $V = (V_x(n); n \geq 1, x \in \mathcal{X})$ is row-wise partial exchangeable if for every n and any $x_1, \cdots, x_n$ in $\mathcal{X}$*

$$\begin{pmatrix} V_{x_1}(1), & V_{x_1}(2), & V_{x_1}(3), & \cdots \\ V_{x_2}(1), & V_{x_2}(2), & V_{x_2}(3), & \cdots \\ & \cdots & & \cdots \\ V_{x_n}(1), & V_{x_n}(2), & V_{x_n}(3), & \cdots \end{pmatrix} \stackrel{\mathcal{L}}{=} \begin{pmatrix} V_{x_1}(\sigma_1(1)), & V_{x_1}(\sigma_1(2)), & V_{x_1}(\sigma_1(3)), & \cdots \\ V_{x_2}(\sigma_2(1)), & V_{x_2}(\sigma_2(2)), & V_{x_2}(\sigma_2(3)), & \cdots \\ & \cdots & \cdots & \\ V_{x_n}(\sigma_n(1)), & V_{x_n}(\sigma_n(2)), & V_{x_n}(\sigma_n(3)), & \cdots \end{pmatrix} \tag{1.42}$$

where for $1 \leq i \leq n$ σ_i permutes a finite number of coordinates of $\mathbb{N}$.

Fortini et al. [60] showed that one can characterize the class of all mixtures of Markov chains alternatively to Diaconis and Freedman [44] as

Theorem 6 ([60])**.** *The matrix of successor states $V = (V_x(n))_{x \in \mathcal{X}, n \geq 1}$ is row-wise partial exchangeable if and only if there exists a $M_M(\mathcal{X})$-valued random element $\tilde{\Pi}$ such that (a) X, given $\tilde{\Pi}$, is a.s. Markov$(\delta_{x_0}, \tilde{\Pi})$ and (b) $\tilde{\Pi}$ belongs to $\mathcal{K}$ a.s..*

Since partial exchangeability and row-wise exchangeability of the successor state both lead to characterizations of mixture of Markov chains it is natural to require a relation between the Diaconis and Freedman [44] and Fortini et al. [60] condition. Fortini et al. showed that

Theorem 7 ([60]). *The matrix V of successor states of X is row-wise partial exchangeable if and only if X is recurrent* [8] *and partial exchangeable.*

From a statistical point, where the object is estimation and not probability theory, one usually wants to use such processes for prediction or estimation rather than studying probabilistic properties. In this case both characterizations of mixture of Markov chains can be equally useful. Dependent on the context and definition of the actual process, it might be easier to check Diaconis and Freedman conditions in some cases. In other cases the Fortini et al. condition might be simpler to check.

1.3.3 Extension to mixture of continuous-time semi/-Markov processes

We shortly mention an extension of the concept of mixtures of Markov chains to the case of continuous time and countable state-spaces. For continuous time processes $X = (X(t), t \geq 0)$ the concept of exchangeability does not make much sense, since such processes would have either constant or non-measureable sample paths. Instead the concept of inter-exchangeable processes can be used [23, 65, 64, 5]. For such a process all non-overlapping increments of equal length are exchangeable. Under regularity conditions also intra-exchangeable processes have a mixture representation, but the mathematics behind such processes becomes much more involved [9].

An alternative approach base on the row-wise exchangeable matrix $V = (V_x; x \in \mathcal{X}, n \geq 1)$, where $V_x(n) = (X_{\tau_x(n)+1}, T_{\tau_x(n)})$ was extended by the n-th holding time [10] in state $x \in \mathcal{X}$ if $\tau_x(n)$ is finite (and $V_x(n) = (\Upsilon, +\infty)$

[8] Recurrence, means again "weak" recurrence, i.e. $P(X_n = x_0 \ i.o.[n]) = 1$.

[9] Freedman [66] for example re-proofed in a simplify way the same results he showed in his 1963 paper [65].

[10] For a definition see below.

otherwise) was proposed by Epifani et al. [51].

Restrict X to be a measurable map from some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ into the space of all right-continuous step functions on $\mathcal{X}$, say $D([0,\infty), \mathcal{X})$. Then the sample path $t \to X(t)$ can be summarized by the initial state $X_0 = X(0)$ together with the jump chain and holding times of X, say $\{(X_n, T_n)\}_{n\geq 0}$, where $T_n = (S_{n+1} - S_n) I_{\mathbb{R}_+}(S_n) + \infty I_\infty(S_n)$, for $S_n = \inf\{t > S_{n-1} : X(t) \neq X(S_n)\}$ and $X_n = X(S_n)$ (see for example [121], p.67-70). Furthermore assume X to be minimal, i.e. set $X(t) = \Upsilon$ if $t \geq \mathcal{E} = \sup_n S_n$ [121]. The X can be written as

$$X(\cdot) = \sum_{n\geq 0} X_n I[S_n, S_{n+1})(\cdot) \tag{1.43}$$

where $S_0 = 0$. Assume for simplicity $X_0 = x_0$ a.s.. Then for V as defined above one can characterize mixtures of continuous time Semi-Markov chains[11] as

Theorem 8 ([51]). *Let X be a minimal $D([0,\infty), \mathcal{X})$-valued random elements with matrix of successor states and holding times of states in $\mathcal{X}$, denoted by V, as defined above. Then V is row-wise partial exchangeable if and only if there exists a $M(\mathcal{X} \times [0,\infty])^\infty$-values random elements Q such that (a) X, given Q, is a recurrent Semi-Markov process with initial state δ_{x_0} and transition kernel Q and (b) Q belongs to $\mathcal{K}$ a.s..*

1.4 Pólya urn schemes

In this section I will shortly summarize some basic facts about Pólya urn schemes. I focus on schemes that will be used in the later chapters. After some general constructions we focus on Pólya urn schemes, which generate Markov-/Exchangeable processes. Hence no generality is intended. For an exhaust treatment of Pólya urn schemes we refer to [106]. For a general treatment of urn schemes in connection with discrete probability theory see

[11] A process is semi-Markov if $\{(X_n, T_n)\}_{n\geq 0}$ is a Markov process on $\mathcal{X} \times [0,\infty]$. Hence the holding times are not required to be exponential distributed any more.

[83].

We use the following particular probability space. Let $\Omega = \Sigma^{\infty}$ where $\Sigma = [0,1]^{K+1}$ with Borel sigma-field ($\mathcal{F} = \mathbb{B}(\Omega), \mathbb{P} = \mathbb{LEB}^{\infty}_{K+1}$) for some fixed $K \in \mathbb{N} \cup \{+\infty\}$, where $\mathbb{LEB}_{K+1}$ denotes the Lebesgue measure on $[0,1]^{K+1}$. Denote with $\mathcal{X} = \{c_i\}_{1 \leq i \leq K}$ the state space on which the Pólya process takes his values. We think of c_i as color i of some urn. The initial number of balls of color c_i is denoted by $m_i \in \mathbb{R}_+$ and I assume that the total mass $m = \sum_{1 \leq i \leq K} m_i < +\infty$ is finite. Furthermore for every c_i let F_i be a proper cdf on $\mathbb{R}^K$.

Definition 9. *A Pólya urn process is a process of the form* $(X, W) = \{(X_n, W_n)\}_{1 \leq n \leq V}$ *where* X *denotes the label process and* W *the weight process. For any* $\omega = \{(U_n, (V_{n,i})_{1 \leq i \leq k})_{n \geq 1}\} \in \Omega$ *and* $W_0 = (m_i/m, i \geq 1)$ *the map* (X, W) *is defined by the following rule: Set* $n = 0$.
(A) While $W_n > \mathbf{0}$ *iterate as follows,*
(A.i) Set

$$X(\omega)_{n+1} = \sum_{1 \leq i \leq K} c_i I\left(\sum_{l \leq i-1} W(\omega)_{n,l} < U_{n+1} \leq \sum_{l \leq i} W(\omega)_{n,l} \right). \tag{1.44}$$

(A.ii) Increase n *by one (i.e.* $n \leftarrow n+1$*) and set*

$$W(\omega)_{n,i} = \frac{m_i + \sum_{t=1}^{n} \sum_{l=1}^{K} A(\omega)_{l,i}^{(t)} I(X(\omega)_t = c_l)}{m + \sum_{t=1}^{n} \sum_{l=1}^{K} \sum_{j=1}^{K} A(\omega)_{l,j}^{(t)} I(X(\omega)_t = c_l)} \tag{1.45}$$

where for $1 \leq i \leq K$ *the row* $A(\omega)_i^{(n)} = (A^{(n)}(\omega)_{i,j})_{1 \leq j \leq K} = F_i^{-1}(V_{n,i})$.
(B) If n *in (A) is finite set* $V = n$ *otherwise set* $V = \infty$.

The interpretation is straight forward. At any point in discrete time, say n, check if the measure $\tilde{W}_{n-1} = \sum_i \tilde{W}_{n-1,i} \delta_{c_i}$, with $\tilde{W}_{n-1,i} = W_{n-1,i} \sum_j W_{n-1,j}$, has only non-negative weights. If the statement is false set $n - 1 = V$ and terminate the process at V. Otherwise draw a random label $X_n \in \mathcal{X}$ from W_{n-1}, say $X_n = c_i$. Now add a random mass $A_{i,j}^{(n)}$ to $\tilde{W}_{n-1}$ for every atom c_j, i.e. $\tilde{W}_{n,j} = \tilde{W}_{n-1,j} + A_{i,j}^{(n)}$ for $1 \leq j \leq K$ and normalize $\tilde{W}_n$ to obtain W_n.

The matrix A (where $A \stackrel{\mathcal{L}}{=} \left(A_{i,j}^{(n)}, 1 \leq i, j \leq K\right)$) is called the "re-

reinforcement" matrix or "generator" of the label process X (see [106], p.45-50). The check for non-negativity of the sampling weights W_n is necessary in order to ensure that the process is well defined. But in many cases, especially when A is deterministic or every row of A has support $\mathbb{R}_+^K$, one needs to check "tenability" only once before starting the algorithm [106]. In the following we give some examples to clarify the scheme.

Example 1.4.1 (Randomized play-the-winner algorithm)**.** *Consider the problem of adaptive allocation of treatments in clinical trials [137]. The aim is to place future patients to treatments which turn out to be more effective than other treatments. For $K = 2$ let $\mathcal{X} = \{c_1, c_2\}$ where c_i denotes treatment $i = 1, 2$. The initial composition $m = (m_1, m_2)$ determines the initial probability of allocation to treatment c_1 or c_2. Denote with $X_n \in \{c_1, c_2\}$ the assignment of the n-th patient to one of the two treatments and let $S_n \in \{0, 1\}$ denote the failure or success (recovery from the disease or not) after the patient receivedone of both treatments. Assume*

$$P(S_n = 1|X_n = c_i) = P(A_i^{(n)} = 1) = p_i \quad i \in \{0, 1\}. \tag{1.46}$$

Rosenberger [137] proposed a Pólya urn scheme, which adaptively assigns patient according to the superiority or inferiority of the treatment. If the treatment was successful, i.e. $[S_n|X_n = c_i] = 1$ add βballs of color $c_{\ i}$ to the urn. If on the other hand $[S_n|X_n = c_i] = 0$ add βballs of color $c_{\ 1-i}$ to the urn. The reinforcement matrix can be expressed in terms of the previous notation as

$$A = \beta \begin{pmatrix} A_1 & 1 - A_1 \\ 1 - A_2 & A_2 \end{pmatrix} \tag{1.47}$$

where $\beta > 0$ is a fixed constant and $A_{\ i} \stackrel{\mathcal{L}}{=} A_i^{(n)} \sim$ Bernoulli(p_j) for $i \in \{0, 1\}$. The predictive allocation of the $(n+1)$th patient given the past n experiments

is defined as

$$P(X_{n+1}=j|\mathcal{F}_n)=\frac{m_j+\beta\sum_{t=1}^{n}[I(X_t=j,A_j^{(t)}=1)+I(X_t=1-j,A_{1-j}^{(t)}=0)]}{m_1+m_2+\beta n} \tag{1.48}$$

for $j\in\{1,2\}$*, where* $\mathcal{F}_n=\sigma(\{X_t,A^{(t)}\}_{1\leq t\leq n})$.

An other example of a Pólya urn scheme of the above form is a scheme which is called a "randomly reinforced urn scheme" [114, 113, 6, 108, 19] and has wide applications in adaptive clinical trials.

Example 1.4.2 (Randomly reinforced urn scheme)**.** *The context is similar to the previous example. Suppose we have* K *treatment arms instead of two and we sequentially assign patients to one of the treatment arms in an adaptive fashion. Hence* $\mathcal{X}=\{c_1,\cdots,c_k\}$ *and the initial urn composition is give by* $\{m_k,1\leq k\leq K\}$*. In the previous example all outcomes after the treatment where Bernoulli distributed. More abstractly, assume now that, given an allocation to arm* $1\leq k\leq K$*, the outcome of the n-th patient* $[S_n|X_n=k]$ *is distributed according to some general probability measure* μ_k *with support contained in* $0<s_1<s_2<\infty$*. In terms of the general Pólya urn scheme described above let* $A_{k,k}^{(n)}:=[S_n|X_n=k]\sim\mu_k$ *and* $A_{k,j}^{(n)}\sim\delta_0$ *for* $j\neq k$*. Then a randomly reinforced urn scheme is a Polya urn scheme as defined in definition 9 with generator matrix given by*

$$A=\begin{pmatrix} A_{1,1} & 0 & \cdots & 0 & \\ 0 & A_{2,2} & \cdots & 0 & \\ \cdots & \cdots & \cdots & \cdots & \\ 0 & 0 & \cdots & 0 & A_{K,K}\end{pmatrix} \tag{1.49}$$

where $A_{k,k}\overset{d}{=}A_{k,k}^{(n)}$ *for* $1\leq k\leq K$*. Suppose that higher values of* $A_{k,k}$ *denote a positive treatment effect, like disability free person years or a reduction on tumor size. Hence* $M_k=\int x\mu_k(dx)>\int x\mu_j(dx)=M_j$ *indicates that treatment* k *is on average superior to treatment* j*. [19] showed that if without loss of generality* $M_1>M_k$ *for all* $k>1$ *and the support of all probability*

measures is bounded away from zero, then the proportion of c_1 balls in the Pólya urn converges almost surely to one. This means that asymptotically all patents are treated with the best treatment arm. Without the assumption on the first moment of the reinforcement distributions, [114] showed that the proportion of balls of each color converges almost surely to a random limit on the K dimensional simplex. This means the sequence of draws from the urn is asymptotically exchangeable (but not exchangeable) and for n very large $X_{n+k}, k \geq 0$ is iid Multinomial given limit of the proportion of balls of each color in the urn.

We stress that the urn scheme in the last example does not produce a mixture representation for the complete sequence of draws of labels. The goal of the scheme is to asymptotic allocate each patient to the superior treatment. Hence the scheme clearly does not intend, and does not need, to produce a mixture representation or an exchangeable sequence. More recent papers study the form of the limiting law of the reinforced urn scheme by a characterization of a unique continuous solution of a functional equation [6]. Other recent work on the randomly reinforced urn scheme for $K = 2$ study the asymptotic distribution of the estimators of the mean respond to the treatment [108].

Example 1.4.3 (Finite Pólya Urn, [50])**.** *A finite Pólya urn scheme for $K \in \mathbb{N}$ and for some fixed, deterministic $\beta > 0$ has generator given by $A = \beta I_K$, where I_K denotes the $K \times K$ identity matrix. The reinforcement scheme is non-random and simply suggest that, after the n-th ball X_n was drawn, we add the same label together with βadditional labels to the urn. The predictive distribution of X_{n+1} given the past is n observations is given by*

$$P(X_{n+1} = j|\mathcal{F}_n) = \frac{m_j + \beta \sum_{t=1}^{n} I(X_t = j)}{\beta n + m} \tag{1.50}$$

for $j \in \{1, \cdots, K\}$. It is easily seen that the joined law of the first n draws

is given by

$$\mathbb{P}(\cap_{1\le t\le n}\{X_t = j_t\}) = \frac{\prod_{1\le j\le K}(m_j/\beta)^{[n(j)]}}{(m/\beta)^{[n]}}, \tag{1.51}$$

where $n(i) = \sum_{1\le t\le n} I(X_t = i)$. *Since for any permutation* σ *of* $\{1, \cdots, n\}$ *the sequence* $(X_{\sigma(1)}, \cdots, X_{\sigma(n)})$ *will produce the same transition counts, the label process* X *must be exchangeable. Hence the finite dimensional joined law can be written uniquely as a mixture probability of some random multinomial probability* $P = \sum_1^K p_j \delta_{c_j}$, *i.e.* $\mathbb{P}(\cap_{1\le t\le n}\{X_t = j_t\}) = \mathbb{E}[\prod_{1\le j\le K} p_j^{n(j)}]$. *All joined* $(k_1, \cdots, k_K)$*-moments of* $(p_1, \cdots, p_{K-1}, 1 - \sum_1^{k-1} p_j)$ *can be recovered from (1.51) by*

$$\mathbb{E}[\prod_{1\le j\le K} p_j^{k_j}] = \frac{\prod_{1\le j\le K}(m_j/\beta)^{[k_j]}}{(m/\beta)^{[n]}}. \tag{1.52}$$

But this are the $(k_1, \cdots, k_K)$*-moments of a Dirichlet vector on the* k*-dimensional simplex with parameters vector* $(m_1, \cdots, m_K)/\beta$ *Since the moments of all order of a random vector* P *with finite support determine uniquely the probability law of* P, *can conclude that* $P \sim Dir((m_1, \cdots, m_K)/\beta)$.

1.4.1 Pólya urn schemes generating mixture processes

There has been much research on urn scheme which generate mixture processes [50, 21, 76, 159, 34, 107, 117, 118, 119, 61]. We will focus on schemes that will be used in the later part. Mixture of finite multinomial random probabilities generated by urn schemes were already discussed in example 1.4.3.

A very general scheme that generates a mixture process was introduced by Diaconis and Copperfield [42] and is called "Random walk with reinforcement" (RWR). The process is actually not generated by an urn scheme. But some special cases of the RWR, like the "Reinforces Urn Process", can be described by urn schemes. Again let $(\Omega, \mathcal{F}, \mathbb{P})$ be defined as before and let $(\mathcal{X}, E)$ be an undirected graph with vertex set $\mathcal{X}$ and edge set E and some initial edge weights $W = \{w(e) > 0 : e \in E\}$. One or sev-

eral loops $e = \{x, x\}$ around the same vertex $x \in \mathcal{X}$ are permitted. The edge set is decomposed into loops and non-loops $E = E_{NL} \cup E_L$,where $E_L = \{e \in E : e = \{x, x\}\ \exists x \in \mathcal{X}\}$ and $E_{NL} = E \setminus E_L$. Assume all neighborhood sets $N(x) = \{y \in \mathcal{X} : \{x, y\} \in E\}$ are finite[12] and choose an initial starting vertex $X_0 = x_0 \in \mathcal{X}$ a.s..

Definition 10 (Random walks with reinforcement, [34])**.** *Let $X = (X_n)$ be a stochastic process with value on the graph $\mathcal{X}$ starting at $X_0 = x_0$ a.s and for $n \geq 1$, given the history $\mathcal{F}_{n-1} = \sigma(\{X_t\}_{0 \leq t \leq n-1})$, let*

$$P(X_n = x|\mathcal{F}_{n-1}) = \frac{w_n(\{X_{n-1}, x\})}{\sum_{y \in N(X_{n-1})} w_n(\{X_n, y\})} I(x \in N(X_{n-1})), \qquad (1.53)$$

where the set of edge weights W_n at time n is given, for $e \in E$, by

$$w_n(e) = w_{n-1}(e) + I(\{X_{n-1}, X_n\} = e) = w(e) + n(e),$$

where $n(e) = \sum_{1 \leq t \leq n} I(\{X_{t-1}, X_t\} = e)$.

The interpretation is as follows. Initially the process moves from vertex x to vertex $y \in N(x)$ via the undirected edge $\{y, x\}$ with probability proportional to the initial weight $w(\{x, y\})$. But every time the process cross the same edge (from x to y or y to x) the edge weight is increased by one. For simplicity, assume for the moment, that all initial weights are equal. Now, suppose that the edge $\{y, x\} \in E$ has been traversed more frequently in the past than any other edges $\{y, z\}, \{x, v\} \in E$. In this case, the next time the process moves to vertex y (or x) it is more likely that the process moves along edge $\{x, y\}$ than any other edge adjacent to the neighborhood $N(y)$ (or $N(x)$). The transition probability will be a mixture of the initial weights and the number of times an edge was traversed in the past.

For any admissible path $(x_0, \cdots, x_n)$ on the graph $(\mathcal{X}, E)$ the finite dimensional distribution of $(X_0, \cdots, X_n)$ given the weights W and initial state

[12]In the initial report [42] also the vertex set $\mathcal{X}$ was assumed to be finite, but this assumption is not necessary. As long as one can proof recurrence, all results hold also on infinite graphs, although the actual 'class' of unique mixing measure might be different (see for example [136, 110, 49])

x_0 is easily seen to be (see [87, 45])

$$\mathbb{P}(\cap_{0\leq t\leq n}\{X_t = x_t\}|W, x_0) = \frac{\prod_{e\in E} w(e)^{[n(e)]}}{\prod_{x\in\mathcal{X}} w(x)^{[n(x)]}}, \tag{1.54}$$

where $w(x) = \sum_{x\in e} w(e)$ and $n(x) = \sum_{1\leq t\leq n} I(X_t = x)$. Hence, from the previous section, the random walk with reinforcements is partial exchangeable and in the case of a finite graph, by definition of W, also recurrent (see [87]). Therefore by theorem 5 the Random walk with reinforcements is a mixture of Markov chains.

Theorem 9 ([42, 87, 45])**.** *Suppose the graph $(\mathcal{X}, E)$ is finite.*
(i) Then X is a mixture of reversible Markov chains, i.e. for any n and any admissible sequence $(x_0, \cdots, x_n)$ of elements in V

$$\mathbb{P}(\cap_{0\leq t\leq n}\{X_t = x_t\}|W, x_0) = \int_{\Delta(E)} \prod_{0\leq t\leq n} \frac{y(\{x_{t-1}, x_t\})}{y(x_{t-1})} f(\boldsymbol{y}|W, x_0) d\boldsymbol{y} \tag{1.55}$$

where $Y(x) = \sum_{e:y\in e} Y(e)$ and the random measure Y on the $|E|$-dimensional simplex has absolutely distribution function.
(ii) The density of Y is given by

$$f(\boldsymbol{y}|W, x_0) \propto \frac{\prod_{e\in E_{NL}} y(e)^{w(e)-.5} \prod_{e\in E_L} y(e)^{w(e)-1}}{y(x_0)^{.5w(x_0)} \prod_{x\neq x_0} y(x)^{\cdot 5(w(x)+1)}} (det(A(y)))^{.5} \tag{1.56}$$

where $A(y) = (A(y)_{i,j})$ is defined in terms of the additive cycle base $(c_i)_{1\leq i\leq l}$, $l := |E| - |V| - |E_L| + 1$, of the H_1 homology[13]

$$A(y)_{i,j} = \begin{cases} \sum_{e\in c_i} y(e)^{-1} & \text{if } i = j \\ \sum_{e\in c_i\cap c_j} g(e)_{i,j} y(e)^{-1} & \text{if } i \neq j \end{cases} \tag{1.57}$$

where $g(e)_{i,j}$ is one if e has the same orientation in the cycle c_i and c_j, and zero otherwise.

[13]For a fixed spanning tree $\mathcal{T}$ of $\mathcal{X}$, every non-loop edge e not in $\mathcal{T}$ forms, when added to $\mathcal{T}$, a cycle c_e [87, 45]. There are $l := |E| - |V| - |E_L| + 1$ such cycles $((c_i)_{1\leq i\leq l})$ which need to be oriented in a fixed, but arbitrary way [87, 45].

The importance of the "Random walk with reinforcements" comes from the fact that the mixing measure is supported on the space for reversible Markov chains.

Remark 7. *Rolles [135] argues that, as the natural conjugated prior for a non-reversible stochastic matrix* Π *are independent Dirichlet random probabilities for every row of* Π*, the "Random walk with reinforcements" (meaning the actual mixing probability* $f(\cdot|W, x_0)$ *) constitutes the natural conjugated prior for reversible Markov chains.*

Remark 8. *One special case of the "Random walk with reinforcements" arises when* $E = \{e_1, \cdots, e_n\}$ *where* $e_i = \{x_0, x_0\}$ *for all* i*. Looking at the formula for at the mixing measure (1.56),* $f(\cdot|W, x_0)$ *becomes a Dirichlet distribution on the* n*-dimensional simplex.*

Example 1.4.4 (Dirichlet Random Probability on $\mathbb{N}$, [42])**.** *Take first* $(\mathcal{X} = \{0, 1, \cdots, K\}$ $E = \{\{0, i\} : 1 \leq i \leq K\}, W = \{w(\{0, i\}) : 1 \leq i \leq K\})$*. This is the star graph (tree) described in [42]. A transition from* 0 *to* $i \leq K$*, leads with probability one back to zero at the next transition, i.e.* $\tau_0(n) = 2n$*. In this cases the mixing measure become a Dirichlet random vector on the* k*-dimensional simplex with parameters* $.5(w(\{0, 0\}), \cdots, w(\{0, k\}))$*. Division by two is necessary since a move* $0, i, 0$ *increases the edge weight of* $\{0, i\}$ *by two, i.e. a Pólya urn scheme with reinforcements matrix* $A = 2Diag(w)$*.*

Now, take $(\mathcal{X} = \mathbb{N}_0,$ $E = \{\{0, i\} : i \in \mathbb{N}\}, W = \{w(\{0, i\}) : i \in \mathbb{N}\})$ *such that the initial weights satisfies* $\sum_i w(\{0, i\}) < +\infty$*. This will be sufficient to ensure recurrence and produces a Markov exchangeable sequence such that* $W_n(\cdot)/n$ *converges to a Dirichlet random measure on* $\mathbb{N}_0$ *(to be precise the random stochastic matrix* Π*, conditional on which* X *is Markov, will have entries* $\pi_{i,j} = \delta_0(j)$ *for* $i \geq 1$ *and the first row* $\pi_{(0,\cdot)}$ *will be a Dirichlet random probability with base measure* $.5w(\{0, \cdot\})$ *)*

Pemantle [123] showed that, for a recurrent "Random Walk with reinforcement" on a tree, the directing random measure is given by independent Dirichlet random vectors for every vertex of the tree.

Reinforced Urn Processes

Reinforced urn processes (RUP's) introduced in [117] can be seen as a special case or an application of several formerly discussed concepts. A RUP can generate a variety of processes which are mixtures of non-reversible processes. In particular a RUP can be seen as a particular "Random Walk with reinforcement" of Coppersmith and Diaconis for a directed graph. RUP's can also be studied via an application of the theory of Fortini et al. [60] in terms of the matrix of successor states.

In particular let $\mathcal{X} = \{x_n, n \geq 0\}$ denote a countable state space and denote with $C = \{c_1, \cdots, c_K\}$ a set of finite labels. Now, connect to every state x_i a finite Pólya urn, with initial urn composition $m_i = (m_i(1), \cdots, m_i(K)) \in \mathbb{R}_+^K$ and reinforcements matrix I_K (see example 1.4.3). The idea of Muliere et al. [117] is to separate colors C and states $\mathcal{X}$ by an "urn function" which plays an equivalent role to the edge set and allows a sparse parametrization. The urn function is a map from $\mathcal{X} \times C$ to $\mathcal{X}$. Suppose $f(x, c) = y$ for some $x, y \in \mathcal{X}$ and $c \in C$. Now, the urn function can be interpreted as follows. Give that the process is in state x at time n, then the process moves to state y at time $n + 1$ if we draw color c from the Pólya urn connected to state x. For a consistent definition we clearly require that $f_x = f(x, \cdot)$ is injective, i.e for each state y there exists at most one color $c_{x,y}$ which leads to a transition from x to $y = f_x(c_{x,y})$ via color $c_{x,y}$. A RUP on $\mathcal{X}$ with parameters $(m = (m_i), C, f)$ is an algorithm defined as follows.

Definition 11 (Reinforced Urn process, [117])**.** *Let $X_0 = x_0$ be a fixed deterministic starting point. For $n \geq 1$ and given $\mathcal{F}_n = \sigma(\{X_i\}_{0 \leq i \leq j})$ sample a label $c \in C$ from the Pólya urn connected with state X_n. Set $X_{n+1} = f_{X_n}(c)$ and replace two colors of label c to the same urn.*

From example 1.4.3, we know that the sequence of draws of colors from a Pólya urn is exchangeable. Furthermore the sequence of successor states of $x_i \in \mathcal{C}$ can be mapped by the inverse of f_{x_i} to the sequence of draws of colors from the Pólya urn which is connected to state x_i. This means that, for $x_j \in f_{x_i}(C)$, if we hit state x_i at time n for the l-th time, the event $\{X_n = x_j | X_{n-1} = x_i\}$ occurs if and only if the l-th color drawn from the urn

connected to state x_i has color $c_{x_i,x_j} = f_{x_i}^{-1}(x_j)$. Therefore, if X is recurrent, the sequence of successor states to each $x \in \mathcal{X}$ is exchangeable. The result of Fortini et al. [60] tells use also that X a recurrent partial exchangeable process. Hence a recurrent RUP X is a mixture of Markov chains.

Furthermore let V_{x_0} denote the subset of recurrent states which are connected with the initial state x_0, i.e. for each $x \in V_{X_0}$ there exists a path from x_0 to x which has positive probability. For states $x_i \in V_{x_0}$ we can find the mixing measure by an application of example 1.4.3. Since $\{X_n = y | X_{n-1} = x_i\}$ occurs if and only if we sample color $c_{x_i,y} = f_{x_i}^{-1}(y)$ from the urn connected to state x, we know that the mixing measure can assign to at most K elements of the transition row $\Pi_{x_i} = \{\Pi_{x_i,y}, y \in \mathcal{X}\}$ non-zero mass, namely to $\Pi_{x_i,y}$ for $y \in f_{x_i}(C)$ (remember $|C| = K$). The remaining elements of Π_{x_i} are zero a.s.. In fact, for the states $f_{x_i}(c_1), \cdots, f_{x_i}(c_K)$ the random transition sub-row $(\Pi_{x_i,f_{x_i}(c_1)}, \cdots, \Pi_{x_i,f_{x_i}(c_1)})$ must be distributed according to a Dirichlet distribution with parameters $(m_i(1), \cdots, m_i(K))$ [14].

An important point to not that non-recurrent state may

The constructive specification of a RUP makes it fortunately relatively simple to deduce some properties of the process. Like the a version of the strong and ordinary Markov property extended to mixture of Markov processes. Suppose that we observe a RUP until time n, then the future process $\{X_{n+k}\}_{k\geq 0}$ is again a RUP, now starting from state X_N, with the same state-space and urn function, but with updated initial urn compositions. This implies that the process $\{X_{n+k}\}_{k\geq 0}$ is again a mixture of Markov chains, whenever the original process is recurrent. The new mixing measure is now the posterior, updated for the observed sequence $X_0, \cdots, X_n$. Updating means an update of the Pólya urn compositions, i.e. from $m_i(j)$ to $m_i(j) + \#\{t < n : X_t = x_i, X_{t+1} = f_{x_i}(c_j)\}$ for all $x_i \in \mathcal{X}$ and all $c_j \in C$. This is of course equivalent to the statement that, given $X_0, \cdots, X_n$, the new mixing measure is such that the random transition sub-row $(\Pi_{x_i,f_{x_i}(c_1)}, \cdots, \Pi_{x_i,f_{x_i}(c_1)})$ is distributed according to a Dirichlet distribution with parameters $(m_i(1) + \#\{t <$

[14] In the original paper the authors proofed a slightly different result. The authors showed that the transition rows should by independent Dirichlet processes. But these Dirichlet processes are in fact only Dirichlet vector since only K elements are non-negative and the elements are identified by the urn function f.

$n : X_t = x_i, X_{t+1} = f_{x_i}(c_1)\}, \cdots, m_i(K) + \#\{t < n : X_t = x_i, X_{t+1} = f_{x_i}(c_K)\})$.

The RUP has been used for wide range of Bayesian modelling purposes of discrete structures without covariates. [25, 26] used RUP's to estimate bivariate and multivariate survival functions. Whereas [111] used a RUP to determine the maximum tolerance doze in clinical trials.

All these applications do not consider covariate effects. Extensions to structures with covariates within the context of RUP's are non-trivial. Approaches to models of the form $\Pi_{x_i,x_j}(Z) = g(\Pi_{x_i,x_j}, \beta z)$, where g is a known link function, $\beta$is a finite or infinite dimensional effect function of the covariate process Z and Π_{x_i,x_j} is a random transition probability, usually do not have a closed form prior or posterior expressions. A RUP essentially models the marginal transition probability of a Markov chain with respect to the posterior law under a conjugated prior. Hence the only way to retain a closed form expression under covariates may be to extend the state space from $\mathcal{X}$ to $\mathcal{X} \times \mathcal{Z}$ where $\mathcal{Z}$ denotes the covariate space. In this case one may models the outcomes and discrete covariates as one random process itself. The effect of covariates may be obtained by marginalizing the probability of $X_{n+1}, Z_{n+1}|X_n, Z_n$ to obtain the probability of $X_{n+1}|X_n, Z_n$. This will be possible at least by numerical summation, even does no neat close form expression will exists. But this approach is clearly unable to provide simple and interpretable effects βas for example in Bayesian semi-parametric models like the Cox model [94, 93] or Bayesian parametric models based on multinomial-regressions [144, 18], which are all implemented by MCMC simulations.

Part I

Generalized Pólya-urn schemes for inhomogeneous Markov chains

Chapter 2

Some preliminaries on urn schemes for Markov chains

2.1 Introduction and relevant literature

A stochastic processes with reinforcement is a process with the following property: the future behavior of the process, given the presence and the past, should be affected by the complete past and not just by the presence. This type of dependence structure, usually means that the condition transition probability "$P(future|past)$" depends on a finite or infinite dimensional statistics which summarizes the past. This may sound very general and non-specific, but even literature surveys on random processes with reinforcement avoid to give a precise definition [125].

Processes with reinforcement have a long history in applied modeling and an urn model is a basic building block for such processes [125]. For example Pólya developed his famous Pólya urn to model the epidemic spread of a disease and the famous Stanford economist Brian Arthur developed an extension of the Pólya's urn to analyze social and industrial processes [50, 13].

Statistical applications of processes with reinforcement are less frequent. The most important process with reinforcement for Bayesian statistics is the random walk with reinforcement (RRW) [34, 42] which we already discussed in section 1.4.1. [45] used the RRW for a Bayesian analyses of reversible

Markov chains. The authors also considered model selection via Bayes factors, which can be done in closed form since the marginal distribution of the RRW is known in closed form (1.54). More recently, [17] extended the RRW to statistical analysis of reversible Markov chains with variable-order dependence. The RRW enables an implicit Bayesian analysis through the representation as a mixture of Markov chains, provided that the process is recurrent. The condition of recurrence usually forced the analyst, who wants to use the model for statistics analysis, to consider a finite or low dimensional state-space.

Hence much of the recent research on random walks with reinforcement focusses on recurrence for infinite state-spaces like $\mathbb{N}_0$, $\mathbb{Z} \times \{1, \cdots, d\}$ with $d \in \mathbb{N}$ or $\mathbb{Z}^k$ for $k > 1$ [36, 87, 110, 136]. For the state spaces $\mathbb{N}_0$ and $\mathbb{Z}$ this problem was solved in [36]. A partially positive answer for recurrence on a two-dimensional graph was given recently in [110] for the graph $V_r = \{(x, y) \in \mathbb{Z}^2 : x \in r\mathbb{Z}, y \in r\mathbb{Z}\}$ with edge set $E_r = \{\{x, y\} \subset V_r : |x-y|_2 = 1\}$ for $r \geq 130$. Recurrence results for reinforced random walks on $\mathbb{Z}^k, k \geq 3$ are still unknown.

A simple and flexible stochastic process with reinforcement, based on a system of finite-color Pólya urns, was introduced in [148, 117], the "reinforced urn process" (RUP) (see section 3.2). The RUP was used to characterize discrete-time Neutral-to-the-right processes and to determine the maximum tolerance dose in drug development [117, 111]. Further analysis bases on RUP's include the estimation of bivariate and multivariate survival functions and the estimation of the number of mixture components of Bernstein polynomial prior [25, 26, 147].

My task in this thesis, and hence the aims of part one, is to study urn processes for the analysis of inhomogeneous Markov chains and in particular multi-state models. This restricts the attention to discrete time problems and in general agreement with the literature we call a discrete-time discrete-state Markov process a Markov chain [121].

In general a multi-state process is a discrete or continuous time process $X = \{X(t), 0 \leq t \leq T\}$, with $T \leq +\infty$ known and values in a finite state space $\mathcal{X} = \{1, \cdots, K\}$. Classical areas of application are demography, actu-

arial science, biostatistics and more recently economics and social science. A transition into a state $i \in \mathcal{X}$ represents in most application the occurrence of an event during a life-time, i.e. emigration into a new geographic region during the physical life-time, the transition into AIDS after infection with the HIV virus or the time to death after surgery. A common feature to all the situations is the fact that all this phenomena is observed for several replicas $X_1, X_2, \cdots, X_n$ and the object of interest is not just if, but also when an event happens, i.e. some residence of a geographic region may leave the region at a certain point during their individual life-time and possible return back later on.

Frequently observations share the same initial state, say 1, for example the event of surgery or the infection with HIV. Usually one is not interested in all states of $\mathcal{X}$, but rather focuses on a particular state, say K, (or several states $K-m, \cdots, K$) in the presence of all other states, for example the death of an observation or the transition to AIDS. These states are called absorbing since either no transition to an other state is possible or one is not interested in the life-time after such an event. A transition into an absorbing state will terminate each life-time X_i. Sometimes a distinction between bi- and uni-directional models is important to simplify computations. In a bi-directional model transitions from i to j permit also transitions from j to i at least for some states. This is not possible for a uni-directional model.

Part of my task in part I is to analysis such multi-state models using urn processes, but there are several other, tailored towards the problem approaches. We briefly mention some of these approaches. If one wants to evaluate the effect of covariates on the transition rate from state i to state j, one approach is to model the hazard rate as a Cox's proportional hazard model

$$P(T_i \in [t, t+dt) | T_i \geq t, Z) = \lambda_{i,j}(t)dt \exp\{\beta_{i,j}(t)Z(t)\} \tag{2.1}$$

for $i, j \in \mathcal{X}$, where T_i is the holding time in state i if the successor state to i is state j. In a frequentist approach one estimates the effect $\beta_{i,j}(t)$ through the partial likelihood, which leaves $\lambda_{i,j}(t)$ unspecified [131, 1]. An

estimator for $P(T_i \in [t, t+dt)|T_i \geq t, Z)$ is obtained by estimating the base-line hazard $\lambda_{i,j}(t)$ by a modified Breslow or Kaplan-Meier method [131, 1] and then use a plug-in estimator for (2.1). In a Bayesian context [93] introduced very recently the semi-proportional intensity model for Markov processes. The model is based on a multivariate sub-ordinator prior for the random cumulative hazard $\Lambda_i = \{\Lambda_{i,j}, j \neq i\}$ for $i \in \mathcal{X}$ of an continuous-time inhomogeneous Markov chain. The model assume a link between the transition probability and covariates of the form

$$P(X(t+dt) = j|X_t = i, Z) = \frac{\Lambda_{i,j}(dt) \exp \beta'_{i,j}(t)Z(t)}{\sum_{h \neq i} \Lambda_{i,h}(dt) \exp \beta'_{i,h}(t)Z(t)}. \tag{2.2}$$

for $j \neq i$ and reduces to the proportional hazard model above if $\beta_{i,j}(t)$ is independent of j, i.e. $\beta_{i,j}(t) = \beta_{i,v}(t)$ for all j and v distinct from i. The model is estimated by MCMC simulation. Discrete time multi-state processes can be models in a several ways as well. A classical approach in mathematical demography and life-insurance mathematics is a model based on stable-population theory. These models assume that the transition rates change during life-time (i.e. ageing) but are constant across birth-cohorts [90, 91, 140, 134]. An alternative to stable-population models are multi-state life-tables. Life-tables are extensively studied in biostatistics and life-insurance mathematics [30, 29, 90, 91, 141]. A life-table describes the life-time of a birth-cohort or a hypothetical period-population of initial size l_0 from "birth" (start at state state 1) to "death" (termination into state K). In the classical approach of Chaing [30, 29] one needs to distinguish between period and cohort life-tables since both concepts have very different interpretations. Recent data sources for life-tables shift from census data to survey data. Furthermore modern life-table estimation is implemented via Poisson or hazard regression models and micro-simulation [28, 104, 105]. A discrete time "version" of (2.1) based on a Multi-logit regression model was proposed in a Bayesian framework by [144]. The model assumes the form

$$P(X_{t+1} = j|X_t = i, Z) = \frac{\exp\{\beta_{i,j} Z_t\}}{\sum_{h \in \mathcal{X}} \exp\{\beta_{i,h} Z_t\}} \tag{2.3}$$

where the baseline hazard rate gets absorbed into the intercept of the regression parameter. The prior for the time-constant effect parameter is assumed to be Gaussian and the estimator are obtained by MCMC simulation.

2.2 Outline of Part I

In the following we introduced some particular urn processes for the analysis of some discrete-time inhomogeneous Markov chains. This problem was analyzed before with RUP's in [148] and [117]. In [117] the authors introduced the general theory of RUP's for a very general class of generic Pólya urn processes first and considered some examples in the remaining sections. The last example [117], section 5, introduced a RUP for a multi-state process which was motivated by the work of [148]. These particular example of a RUP was aimed towards a very particular kind of multi-state processes. The authors considered a survival problems, with death as absorbing state and several non-absorbing states, say $1, 2, 3, \cdots$. The transition rule among states during the lifetime takes the form $i \to i$ or $i \to i + 1$ which restricts real life applications. Clearly the aim of the particular example of a RUP was very specific. We think is is worth to look at other kind of RUP's, and Pólya urn models in general, with different transition-structures and a focus on time-inhomogeneous Markov chains with finite state-spaces. A shortcomings of the Pólya urn model approach is the disregard of the effect of covariates on the transition probability operator. This clearly restricts the use of such models in real-life applications.

The outline of part I is as follows. In chapter 3 we define a particular RUP for the analysis of multiple-state processes. These multiple-states processes are assumed to have a time-inhomogeneous dependence structure and a finite state-space. We will explore results for RUP's from [117] to provide Bayes estimator for summary measure, like the t-step prediction probability, the occupation probability for a given state and the expected lifetime.

This analysis is extended in the remaining two chapters of part one to bivariate and multivariate state-spaces and towards exchangeable sequences of multivariate time inhomogeneous Markov chains with distinct and random

initial states. Each chapter will have the same structure. We first introduce a particulate Pólya urn process, study the mixture representation and the mixing measure. Then, we use a sequence of hitting times of the process to construct exchangeable sequences of mixtures of time-inhomogeneous Markov chains. These sequences are used to derive predictive estimator for summary measures of time-inhomogeneous Markov chain. All chapters are papers currently under preparation for submission to a journal. This unfortunately causes extensive repetition of definitions and concepts.

Chapter 3

A Reinforced Urn Process for the Analysis of Multiple-State Processes

3.1 Introduction

The reinforced random walk (RRW) was introduced in [34, 42]. A reinforced random walk describes the random movement of a walker on a countable state space. Initially the walker moves at random on the state space. But as time goes on regions that have been traversed frequently in the past become more likely to be revisited.

One may wonder whether the this property seems reasonable. The RRW model can be seen as a sequential leaning process for a generic discrete-state and discrete-time process. In the initial stage of the learning process the probability of transition among stages is unknown and often no prior knowledge is available. Hence it is natural to treat transitions from a certain state x to all neighbour states of x, say $y_1, \cdots, y_n$, to be uniformly distributes (i.e. $x \to y_i$ with probability $1/n$). If prior information are available one may can change the edge weights. With every unite of time the process is observed one learns more about the process. If one observes a transition $x \to y_i$ more often than $x \to y_j$ it is natural to treat the first event as more likely than the

second and this is what a RRW does. In practice, one observes a discrete-state and discrete-time process at once or sequential and tries to estimate the transition probability by keeping track of the number of crossings of every edge and then use the estimator (1.53). The predictive rule of the RRW is then use to estimate the likelihood of a path and to estimate the most likely paths. In theory also other reinforcement schemes could be used, even random reinforcement schemes. But if y_i and y_j are not linked to each other in some interpretable meaning it may not be a good idea to increase the likelihood of moving from x to y_i whenever we observe a transition from x to y_i.

Since introduced by Coppersmith and Diaconis, RRW's have been analyzed and further developed in a variety of different ways [36, 123, 124, 87, 135, 136, 110] and became an important tool for statistical modeling and inference [148, 45, 111, 25, 24, 26, 147]. An important reinforced random walk in Bayesian statistics is the reinforced urn process (RUP) [117]. A RUP is a random walk on a space of Pólya urns and is a frequently used tool in Bayesian non-parametrics [111, 25, 24, 26, 147].

In the following we introduce a particular RUP for the analysis of multi-state processes which extends the work in [117], section 5, to a general class of multi-state processes. In [117], the authors introduced a generic RUP and studies some general properties of the reinforced random walk. As an example, the authors introduced also a RUP for the analysis of a multi-state process for a very specific problem. The described process has the property that, conditional on the current state, the process may remain in the same state, moves to the next higher ordered state or terminates. This property restricts the application of the particular RUP. For instance, in biostatistics the state-space of a multi-state process may correspond to health-states and non-monotone transitions along both improvement and deterioration are possible. Such problems require a process that handles arbitrary transition rules. In the following, we introduce another particular RUP for the analysis of a general class of multi-state processes. We apply the process to Bayesian inference for multi-state models with a discrete-time inhomogeneous Markov structure.

The paper is organized in the following way. In Section 2 we review some basic facts about RUP's used in the sequel. In Section 3 we define a RUP for the analysis of inhomogeneous Markov chains. In Section 4 we provide Bayes estimates for the transition array and summary measures. In Section 5 we shortly discuss missing data problems and right censoring. In Section 6 we apply the constructed process to inference in heart-transplantation monitoring. Final comments and further developments are discussed in the last section.

3.2 A Review of Reinforced Urn Processes

The reinforced urn process (RUP) was introduced in [117]. RUP is a reinforced stochastic process which combines two fundamental probabilistic concepts: a Pólya urn scheme and a random walk. A RUP can be seen as a random walk on a countable space of Pólya urns. Initially all trajectories of the RUP are determined by a prior guess only, but the transition probabilities are sequentially updated after each transition step. We will shortly recall some basic facts of RUP's which will be used later. Each RUP consists of four elements,

(i) a countable state space $\mathcal{S}$;

(ii) a finite set of colors $\mathcal{C} = \{c_j\}_{1 \leq j \leq k}$ of cardinality $1 \leq k < \infty$;

(iii) an urn composition function, $U : \mathcal{S} \to [0, +\infty)^k$, which assigns to every state s a Pólya urn $U(s)$ with initial composition $\{n_s(c_1), \cdots, n_s(c_K)\}$ of $n_s(c_j)$ balls of color $c_j \in \mathcal{C}$ and strictly positive total mass $m_s = \sum_1^k m_s(c_j)$;

(iv) a law of motion $q : \mathcal{S} \times \mathcal{C} \to \mathcal{S}$, which leads from a state s to state $s' = q(s, c)$ through color c.

It is assumed that for all states s, s' in $\mathcal{S}$, there exists at most one color $c \in \mathcal{C}$ such that $q(s, c) = s'$. A RUP can now be defined by the following sequentially sampling scheme.

Definition 12. *Fix an initial state $X_0 = s_0 \in S$ and sample a ball from the Pólya urn associated with s_0. Let $c \in \mathcal{C}$ denotes the color of the ball sampled from urn $U(s_0)$. Set $X_1 = q(s_0, c)$ and return two balls of color c to the urn $U(s_0)$. Now, conditional on a sample of $(n+1)$ observations $\{X_i = s_i\}_{i=0,\cdots,n}$, sample a ball from the Pólya urn associated with s_n. If c^* denotes the color of the sampled ball, then set $X_{n+1} = q(s_n, c^*)$ and return two balls of color c^* to urn $U(s_n)$.*

Now, recall that two finite sequences σ and ϕ of elements in $\mathcal{S}$ are called equivalent if (i) they start with the same initial state and (ii) for every pair of states s, s^* in $\mathcal{S}$ the number of transitions from s to s^* is the same in both sequences. An important property for a Bayesian analysis of stochastic processes is the following

Definition 13 ([44])**.** *A stochastic process $\{Y_n\}$ with values in a countable set $\mathcal{I}$ is said to be partial exchangeable if for all equivalent sequences $\sigma = (w_0, \cdots, w_n)$ and $\phi = (v_0, \cdots, v_n)$ of states in $\mathcal{I}$*

$$P(Y_0 = w_0, \cdots, Y_n = w_n) = P(Y_0 = v_0, \cdots, Y_n = v_n). \tag{3.1}$$

A RUP has the this property. Roughly speaking, a process is partial exchangeable if the finite dimensional joined distribution of the process depends only on the number of transitions between states but not on the order of these transitions. An important result, dating back to [44], states that any recurrent partial exchangeable process can be represented as a mixture of Markov chains with unique mixing measure.

Now, let $\mathcal{H}_s = \{y \in \mathcal{S} : \exists c \in \mathcal{C}\ y = q(s, c) \text{ and } n_s(c) > 0\}$ be the set of all states attainable from state s in one transition step. Furthermore set $\mathcal{R}^{(0)} = \{s_0\}$ and define inductively $\mathcal{R}^{(n)} = \bigcup_{s \in \mathcal{R}^{(n-1)}} \mathcal{H}_s$. $\mathcal{R}^{(n)}$ denotes the set of all states that the RUP can reach with positive probability within n steps. Let $\mathcal{R} = \bigcup_{n=0}^{\infty} \mathcal{R}^{(n)}$ denote the set of states attainable within a finite number of transitions. As a consequence of de Finetti's theorem for partial exchangeable processes [44], a RUP has the following representation

Theorem 10 ([117], Theorem 2.16)**.** *Suppose the RUP* $\{X_n\}_{n\geq 0}$ *is recurrent, i.e.* $P(X_n = s_0 \text{ } i.o.) = 1$.
(i) Then there exists a unique probability measure μ *on the set* $\mathcal{P}$ *of stochastic matrices on* $\mathcal{R} \times \mathcal{R}$ *such that for all* $n \geq 1$ *and all sequences* $(s_0, \cdots, s_n)$ *of elements in* $\mathcal{R}$

$$P(X_0 = s_0, \cdots, X_n = s_n) = \int_{\mathcal{P}} \prod_{j=0}^{n-1} \Pi(s_j, s_{j+1}) \mu(s_0, d\Pi). \tag{3.2}$$

(ii) The rows of Π *are mutually independent random probability distributions on* $\mathcal{S}$ *and for all* $s \in \mathcal{R}$ *the row* $\Pi(s) = \{\Pi(s, s^*), s^* \in \mathcal{S}\}$ *is a Dirichlet process with base measure* α_s *given by*

$$\alpha_s(s^*) = \begin{cases} n_s(c) & \text{if } s^* \in \mathcal{R} \text{ and } \exists\ c \in \mathcal{C} : s^* = q(s, c) \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

Theorem 10 states that, for almost all realization of the random transition matrix $\Pi \in \mathcal{P}$, $\{X_n\}$ is a Markov chain with transition matrix Π . Where the transition rows of Π are independent Dirichlet random vectors. This results will be used in the next section to construct a process for a Bayesian analysis of multiple-state processes.

3.3 A RUP for the analysis of multiple-state processes

In this section we define a RUP for inference for multi-state processes that arise in survival analysis, reliability theory or event-history models. Typically in such situations we observe several individuals, $\{\mathbf{Y}_m\}_{1\leq m\leq M}$, repeatedly over time. Individuals may be judged to be exchangeable. Furthermore, for a fixed individual m and conditional on a stochastic matrix Π, the sample path $\mathbf{Y}_m = \{Y_1^{(m)}, Y_2^{(m)}, \cdots, Y_{\tau_m}^{(m)}\}$ is assumed to follow a time-inhomogeneous Markov chain with values in a countable state space $\mathcal{G}$ and survival time τ_m such that $P(\tau_m < \infty) = 1$.

We will construct such a sequence based on the RUP defined below. From a Bayesian perspective every element $\mathbf{Y}_m$ will be a mixture of time-inhomogeneous Markov chains. Conditional on a random stochastic matrix $\Pi = \{\Pi(t)_{i,j}; t \geq 0, i, j \in \mathcal{G}\}$, any $\mathbf{Y}_m$ is a Markov chain with values in $\mathcal{G}$ such that for any i, j in $\mathcal{G}$ and any $0 \leq t \leq \tau_m$

$$P(Y_{t+1}^{(m)} = j | Y_t^{(m)} = i, \Pi) = \Pi(t)_{i,j} \quad a.s. \tag{3.4}$$

We specify the following state-space, color set and law of motion.

(i) The state space is $\mathcal{S} = \{0, 0\} \bigcup (\mathcal{G} \times \mathbb{N})$ where $\mathcal{G} = \{1, 2, \cdots, k\}$ for some fixed $k \in \mathbb{N}$, where $\mathbb{N}$ denotes the set of positive integers;

(ii) the color set is $\mathcal{C} = \{0, 1, \cdots, k\}$;

(iii) the set of urn functions will be denoted as $U_{(i,t)} = \{n_{(i,t)}(0), \cdots, n_{(i,t)}(k)\}$, $(i, t) \in \mathcal{S}$.

(iv) Finally, the law of motion is defined for all $(i, t) \in \mathcal{S}$ by

$$q((i,t), c) = \begin{cases} (0,0) & \text{if } c = 0 \\ (j, t+1) & \text{if } c = j \in \mathcal{G}; \end{cases} \tag{3.5}$$

Now, define the process $\{X_n\}_{n \geq 0}$ inductively as follows

Definition 14. *Let $\{X_n\}$ be a stochastic process with values in $\mathcal{S}$ defined inductively by $P[X_0 = (0,0)] = 1$ and for any $m \geq 1$*

$$\begin{aligned} &P[X_{m+1} = (j,s) | X_m = (i,t), \mathcal{F}_m] \\ &= \begin{cases} \dfrac{n_{(i,t)}(j) + N_m[(i,t);(j,s)]}{\sum_{v=0}^{k} n_{(i,t)}(v) + N_m[(i,t)]} & (j,s) \in \{(0,0)\} \cup (\mathcal{G} \times \{t+1\}) \\ 0 & \textit{otherwise} \end{cases} \end{aligned} \tag{3.6}$$

where $\mathcal{F}_m = \sigma(\{X_i; 0 \leq i \leq m\})$ *and for* $(i,t),(j,s)$ *in* $\mathcal{S}$

$$N_m[(i,t);(j,s)] = \sum_{v=0}^{m-1} I(X_v = (i,t); X_{v+1} = (j,s)) \tag{3.7}$$

$$N_m[(i,t)] = \sum_{(j,s)\in\mathcal{S}} N_m[(i,t);(j,s)] \tag{3.8}$$

denotes the counting processes of transitions from (i,t) *to* (j,s) *and the number of visits to* (i,t) *until time* m.

The intuition behind the bivariate process $\{X_n\}$ is the following. The initial state of the process $(0,0)$ defines a reference event 0 at time zero. Then the process moves along the sequence of states $(i_t,t)_{t\geq 1}$ in $\mathcal{S}\setminus\{(0,0)\}$ where t denotes the time since the reference event occurred last and i_t denotes a state in $\mathcal{G}$ t-units of time after the reference event occurred. The reference event 0 occurs next at time ζ_1. At this point the time since event 0 occurred last will restart, i.e. $(0,0)$, and the process moves again along a sequence of states $(v_t,t)_{t\geq 1}$ in $\mathcal{S}\setminus\{(0,0)\}$ until the reference event 0 occurs next, say at ζ_2. The transition probabilities are updated after each transition step according to (3.6).

The RUP just defined is partial exchangeable, which implies that the process will be a mixture of Markov chains whenever the process is recurrence. In order to ensure a mixture representation, we now provide some conditions which are sufficient of the recurrence of the process. Let $\{\zeta_n\}$, denote the sequence of hitting times to the initial state of the process. That is $\zeta_0 = 0$ and $\zeta_{n+1} = \inf\{m > \zeta_n : X_m = (0,0)\}$ where $\inf\{\emptyset\} = \infty$. The process $\{X_n\}$ is recurrent if

$$P[X_n = (0,0) \text{ for infinity many n}] = P\left[\bigcap_{n=0}^{\infty}\{\zeta_n < \infty\}\right] = 1. \tag{3.9}$$

Furthermore define the sub-distribution $\Lambda^{(0)}(0)$ and the sequence of stochastic sub-matrices $\{\Lambda^{(0)}(t)\}_{t\geq 1}$ by

$$\Lambda^{(0)}(0) = \left(\frac{n_{(0,0)}(j)}{\sum_{v=0}^{k} n_{(0,0)}(v)} \right)_{j \in \mathcal{G}} \text{ and } \Lambda^{(0)}(t) = \left(\frac{n_{(i,t)}(j)}{\sum_{v=0}^{k} n_{(i,t)}(v)} \right)_{i,j \in \mathcal{G}} \tag{3.10}$$

Lemma 2. *Suppose the sequence of stochastic sub-matrices* $\{\Lambda^{(0)}(t)\}_{t \geq 1}$ *satisfies the following conditions*

(i) $\forall (i,t) \in \mathcal{G} \times \mathbb{N} \;\; \forall j \in \mathcal{G}$ *such that* $n_{(i,t)}(c_j) > 0$ *and*

(ii) $\lim_{n \to \infty} \prod_{t=1}^{n} \Pi^{(0)}(t) = \boldsymbol{0}$.

Then the RUP $\{X_n\}$ *is recurrent.*

The proof of this lemma and all further proof are given in the appendix. The recurrence condition essentially requires that after a certain time $t \geq 1$ we need to have sufficient positive weight $n_{i,t}(0)$ of balls of color 0 in the urn connected with state (i,t), for each $i \in \mathcal{G}$. As a consequence of the previous lemma and theorem 10 we obtain the following.

Corollary 1. *Suppose condition (i) and (ii) of lemma 2 hold, then*

(i) $\{X_n\}$ *will be a mixture of Markov chains on* $\mathcal{S}$ *with random transition array* $\Pi = \{\Pi_{(i,t)(j,v)}; (i,t),(j,v) \in \mathcal{S}\}$.

(ii) If, for $(i,t) \in \mathcal{S}$, $\Pi_{(i,t)}$ *denotes the transition row state* (i,t) *to states* $s \in \mathcal{S}$, *then the rows of* Π *are mutually independent random probability measures, where* $\Pi_{(i,t)}$ *is a Dirichlet process with base measure* $\alpha_{(i,t)}$ *equal to*

$$\alpha_{(i,t)}(j,s) = \begin{cases} n_{(i,t)}(j) & \text{if } (j,s) \in \{(0,0)\} \cup (\mathcal{G} \times \{t+1\} \\ 0 & \text{otherwise.} \end{cases} \tag{3.11}$$

Remark 9. *Let* Π *denote the random transition matrix of corollary 1. For a fix* $(i,t) \in \mathcal{G} \times \mathbb{N}$ *the base measure* $\alpha_{(i,t)}$ *assigns positive mass at most to the states* $G \times \{t+1\} \cup \{0,0\}$. *Hence the random weights of the transition row* $\Pi_{(i,t)}$ *are zero almost surely for all states different from the set* $\mathcal{G} \times \{t+1\} \cup \{(0,0)\}$.

This implies than Π *has the form of a block matrix with an additional first column vector*

$$\Pi \overset{a.s.}{=} \begin{pmatrix} \Pi(0)_0 & \Pi(0)_{-0} & \boldsymbol{0} & \boldsymbol{0} & \cdots \\ \Pi(1)_0 & \boldsymbol{0} & \Pi(1)_{-0} & \boldsymbol{0} & \cdots \\ \Pi(2)_0 & \boldsymbol{0} & \boldsymbol{0} & \Pi(1)_{-0} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}, \tag{3.12}$$

where the random elements of Π *are given by*

$$\Pi(0) = (\Pi_0(0), \Pi_{-0}(0)) = (\Pi_{(0,0)(0,0)}, \quad \Pi_{(0,0)(j,1)}; j \in \mathcal{G}) \quad \textit{and} \tag{3.13}$$

$$\Pi(t) = (\Pi_0(t), \Pi_{-0}(t)) = (\Pi_{(i,t)(0,0)}; i \in \mathcal{G}, \quad \Pi_{(i,t)(j,t+1)}; i, j \in \mathcal{G}), \tag{3.14}$$

where $\Pi_{-0}(0) = \left(\Pi_{(0,0)(j,1)}; j \in \mathcal{G}\right)$ *and* $\Pi_{-0}(t) = \left(\Pi_{(i,t)(j,t+1)}; i, j \in \mathcal{G}\right)$ *for* $t \geq 1$. *Therefore in the sequel we prefer to write* Π *as a transition array* $\Pi = \{\Pi(t)\}_{t\geq 0}$ *where the elements* $\Pi(t)$ *are defined above.*

For prediction purposes later we state the following lemma and corollary. Both lemmas together state that, given a sample of observations from the RUP, the process starts again as the same RUP, just with a different initial state and an updated initial urn composition function.

Lemma 3. *Suppose the RUP* $\{X_n\}$ *is recurrent. Then conditional on a sample of* $(n+1)$ *observation* $\{X_k = s_k\}_{0\leq k\leq n}$ *from the RUP,*
(i) $\{X_{n+k}\}_{k\geq 0}$ *is again a mixture of Markov chains with initial state* s_n *and random transition array* $\Pi^{(n)} = \{\Pi^{(n)}(t); t \geq 0\}$.
(ii) The random transition matrices $\{\Pi^{(n)}(t); t \geq 0\}$ *are mutually independent with respect to* t. *For a fixed* $t \geq 0$ *the rows of* $\Pi^{(n)}(t)$ *are independent random probabilities and the* i*-th row of* $\Pi^{(n)}(t)$ *is a Dirichlet vector with base measure*

$$\alpha_{n,(i,t)}(\cdot) = \alpha_{(i,t)}(\cdot) + \sum_{j=1}^{k} \delta_{(j,t+1)}(\cdot) N_n[(i,t)(\cdot)] + \delta_{(0,0)} N_n[(i,t)(\cdot)]. \tag{3.15}$$

Remark 10. *We emphasize that the random transition matrix* $\Pi^{(n)} = \{\Pi^{(n)}_{s,s^*}; s, s^* \in \mathcal{S}\}$ *in the previous lemma has the same block-form as the ran-*

dom transition matrix Π. *Hence, we wrote* $\Pi^{(n)}$ *again as a system of transition matrices defined by*

$$\Pi^{(n)}(0) = \left(\Pi_0^{(n)}(0),\ \Pi_{-0}^{(n)}(0)\right) = \left(\Pi^{(n)}_{(0,0)(0,0)},\qquad \Pi^{(n)}_{(0,0)(j,1)}; j \in \mathcal{G}\right)$$
$$\Pi^{(n)}(t) = \left(\Pi_0^{(n)}(t),\ \Pi_{-0}^{(n)}(t)\right) = \left(\Pi^{(n)}_{(i,t)(0,0)}; i \in \mathcal{G},\ \ \Pi^{(n)}_{(i,t)(j,t+1)}; i,j \in \mathcal{G}\right),\ t \geq 1,$$

where $\Pi_{-0}^{(n)}(0) = \left(\Pi^{(n)}_{(0,0)(j,1)}; j \in \mathcal{G}\right)$ *and* $\Pi_{-0}^{(n)}(t) = \left(\Pi^{(n)}_{(i,t)(j,t+1)}; i,j \in \mathcal{G}\right)$ *for* $t \geq 1$.

It is not hard to see that the previous result still holds if we condition on the history of the process until the n-th hitting time ζ_n.

Corollary 2. *Suppose the RUP* $\{X_n\}$ *is recurrent. Then conditional on the event* $\{\zeta_n < \infty\}$ *and the sample* $\{X_k\}_{0\leq k\leq \zeta_n}$ *from the RUP,*
(i) $\{X_{\zeta_n+k}\}_{k\geq 0}$ *is again a mixture of Markov chains, with initial state* $(0,0)$ *and random transition array* $\Pi^{(\zeta_n)} = \{\Pi^{(\zeta_n)}(t); t \geq 0\}$.
(ii) The random transition matrices of $\{\Pi^{(\zeta_n)}(t); t \geq 0\}$, *are mutually independent with respect to* t. *For a fixed* $t \geq 1$ *the rows of* $\Pi^{(\zeta_n)}(t)$ *are independent random probability distributions and the* i*-th row of* $\Pi^{(\zeta_n)}(t)$ *is a Dirichlet vector with base measure*

$$\alpha_{\zeta_n,(i,t)}(\cdot) = \alpha_{(i,t)}(\cdot) + \sum_{j=1}^{k} \delta_{(j,t+1)}(\cdot) N_{\zeta_n}[(i,t)(\cdot)] + \delta_{(0,0)} N_{\zeta_n}[(i,t)(\cdot)]. \quad (3.16)$$

Recall that the aim of this section was to construct an exchangeable sequence of random elements $\{\mathbf{Y}_m\}_{m\geq 1}$ such that, for every fix m, the random element $\mathbf{Y}_m$ is a mixture of inhomogeneous Markov chains on $\mathcal{G}$ stopped at a random point in time. We will use the sequence of trajectories between two successive returns to the initial state $(0,0)$ of the process $\{X_n\}_{n\geq 0}$ to construct such a sequence. Denote with $\mathcal{S}^*$ the space of all finite sequences of elements in $\mathcal{S}$ endowed with the discrete topology.

Definition 15. *Suppose the process* $\{X_n\}$ *is recurrent. Define the n-th* $(0,0)$*-block for the RUP to be the sequence of states between the* $(n-1)$*-th and* n*-th*

return to the initial state $(0,0)$*, that is*

$$\boldsymbol{B}_n = (X_{\zeta_{n-1}}, X_{\zeta_{n-1}+1}, \cdots X_{\zeta_n-1}). \tag{3.17}$$

Notice that, if $\{X_n\}$ is recurrent, $\{\mathbf{B}_n\}$ will be a well-defined sequence of random elements with values in $\mathcal{S}^*$. The next lemma state a useful property of the blocks $\{\mathbf{B}_n\}$ which will of frequent use in the sequel.

Lemma 4. *Suppose the RUP* $\{X_n\}$ *is recurrent, then the sequence of* $(0,0)$*-blocks* $\{\boldsymbol{B}_n\}$ *is exchangeable.*

Notice that if L is a measurable function from $\mathcal{S}^*$ into another measurable space, then $\{L(\mathbf{B}_n)\}_{n\geq 1}$ is exchangeable too. Now, let $L : \mathcal{S} \to \mathcal{G} \cup \{0\}$ denote the projection onto the first coordinate of $\mathcal{S}$, that is $L((i,t)) = i$ for all $(i,t) \in \mathcal{S}$ and let $\{\mathbf{Y}_n\}$ be the sequence of projections of $\{B_n\}$ onto the first coordinates, that is

$$\begin{aligned} \mathbf{Y}_n = L(\mathbf{B}_n) &= \left(L(X_{\zeta_{n-1}}), L(X_{\zeta_{n-1}+1}), \cdots, L(X_{\zeta_n-1})\right) \\ &= \left(Y_0^{(n)}, Y_1^{(n)}, \cdots, Y_{\tau_n}^{(n)}\right) \end{aligned} \tag{3.18}$$

where $\tau_n = \zeta_n - \zeta_{n-1} - 1$. By exchangeability of the $(0,0)$-blocks and measurably of the map L it follows that $\{\mathbf{Y}_n\}$ is exchangeable as well. It is worth stating the following obvious, but useful property of $\{\mathbf{Y}_n\}_{n\geq 1}$.

Lemma 5. *The sequence of random elements* $\{\boldsymbol{Y}_n\}_{n\geq 1}$ *is exchangeable. Furthermore, conditional on the random array* $\Pi = \{\Pi(t), t \geq 0\}$*, for any* n $\boldsymbol{Y}_n = \{Y_t^{(n)}, 0 \leq t \leq \tau_n\}$ *is a time-inhomogeneous Markov chain on* $\{0\} \cup \mathcal{G}$ *with initial distribution* δ_0 *and transition matrix* Π *independent of all* $\boldsymbol{Y}_m$*,* $m \neq n$*.*

Remark 11. *The sequence of stochastic processes* $\{\boldsymbol{Y}_m\}$ *is a simple tool for Bayesian inference for inhomogeneous Markov chains and in particular multi-state processes.*

Suppose information about a process arises from observing the same process for several, say M*, independent replicates, conditionally on an unknown*

transition array. In such a case we can take $\mathcal{G}$ as the state- space of the multi-state process and let the observed data be the first M blocks $\{\boldsymbol{Y}_m\}_{1\leq m\leq M}$. Bayesian inference for summary measures of the process $\phi(\boldsymbol{Y}_m)$, with respect to a square loss function, can now be reduce to computations which involve only the predictive rule of the RUP (see section 3.4 and 3.6 for a numerical illustration).

Remark 12. *We like to comment on the implication of the partial exchangeability property on the behaviour of the sequence $\{\boldsymbol{Y}_m\}$. Take the x_0-blocks of a generic partial exchangeable process on a general state space $\boldsymbol{X}$. Call them $\{\boldsymbol{B}_m\}$ as well. Take any two sequences of states $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$ in $\boldsymbol{X}$, which start with $x_1 = x_0 = y_1$, have the same transition counts, and do not contain x_0 except at the first coordinate. Then the event $\{B_n = (x_1, x_2, \cdots, x_n)\}$ and the event $\{B_n = (y_1, y_2, \cdots, y_n)\}$ will be equal likely. This clearly needs to be the case as the x_0-blocks are exchangeable. Hence it doesn't matter at which point in time each transition occurs as long as the overall transition counts are the same in both sequences. Therefore the transition probability of the partial exchangeable process and the x_0-blocks is stationary. This is an essential part of why a recurrent partial exchangeable processes is a mixture of Markov chains. The random transition matrix can be found as the weak limit of the empirical transition probabilities, hence transitions should be time inhomogeneous.*

In the case of modelling a multi-state process a stationary transition probability is undesirable. This is why we extended the state-space in the rather unintuitive way from $\mathcal{G}$ (which is $\boldsymbol{X}$ in the above notation) to the state-space $\mathcal{S}$. The RUP we defined above implies the same stationary properties of the transition probability for the x_0-blocks as just described. But the point is that two sequences of states in $\mathcal{S}$, which start with $x_1 = (0,0) = y_1$, have the same transition counts, and do not contain $(0,0)$ except at the first coordinate, now have the form $(0,0), (x_2, 1), \cdots, (x_n, n-1)$. Hence in order for $(0,0), (x_2, 1), \cdots, (x_n, n-1)$ and $(0,0), (y_2, 1), \cdots, (y_n, n-1)$ to contain the same transition counts in $\mathcal{S}$ (not in $\boldsymbol{X} = \boldsymbol{G}$) we require that the two sequences have the same number of transition from (i,t) to $(j, t+1)$ for each $i, j \in \mathcal{G}$ at each $t \geq 0$.

The implication for the sequence $\boldsymbol{Y}_n$, derived from the $(0,0,)$-blocks $\{\boldsymbol{B}_m\}$ is therefore as follows. The fact that two blocks $\boldsymbol{Y}_n$ and $\boldsymbol{Y}_m$ have the same number of transitions for each state $i \in \mathcal{G}$ will generally not be sufficient to have the same probability. Since both sequences start with the same initial state, we actually require that $Y_t^{(n)} = i_t = Y_t^{(m)}$ for some state $i_t \in \mathcal{G}$ for all $t \leq \tau_m = \tau_n$ in order for $\boldsymbol{Y}_n$ and $\boldsymbol{Y}_m$ to be equal likely. The only case for which two blocks $\boldsymbol{Y}_n$ and $\boldsymbol{Y}_m$ will have equal probability, whenever they contain the same number of transitions among states, is when $n_{i,t}(j) = n_{i,t+n}(j)$ for all $t, h \geq 0$ and all states $i, j \in \mathcal{G}$, i.e. when the masses $n_{i,t}(j)$ in every urn are independent of t.

The same fact holds as well for the processes we consider in chapter 4 and 5, since we use the same ideas of extending the actual state-space to the state-space time the natural numbers.

An other property we want to remark is the possibility of absorbing states $i \in G$ of a multi-state process. In some cases the elements of $\mathbf{Y}_m$ may remain constant after some point in time $Y_t^{(m)} = i \in G$ over the remaining lifetime τ_m. Such states are absorbing the scene of a multi-state process. But as long as the lifetime τ_m is finite, this does not effect the recurrence property of the RUP. The recurrence property of the RUP only requires and implies that the each element $Y^{(m)}$ is finite with probability one.

By definition of the $(0,0)$-blocks we may simplify the counting processes, defined in (3.7) and (3.8), to

$$N_{\zeta_m}(i,j,t) = \sum_{n=1}^{m} I(Y_t^{(n)} = i, Y_{t+1}^{(n)} = j) = N_{\zeta_m}[(i,t);(j,t+1)] \; j \in \mathcal{G}, \quad (3.19)$$

$$N_{\zeta_m}(i,t) = \sum_{j=0}^{k} N_{\zeta_m}(i,j,t) \; i \in \mathcal{G} \quad (3.20)$$

for every i in $\mathcal{G} \cup \{0\}$ and $t \geq 0$. Where $N_{\zeta_m}(i,j,t)$ denotes the number of times one of the m processes $\{\mathbf{Y}_m\}_{1 \leq m \leq M}$ moves from state i to j at time t. If we wants to predict survival times, τ_m, of the process $\mathbf{Y}_m$, we also need

the counting processes

$$N_{\zeta_m}(i,0,t) = \sum_{n=1}^{m} I(Y_t^{(n)} = i, \tau_n = t) \quad \text{for } i \in \mathcal{G} \cup \{0\}. \tag{3.21}$$

3.4 Predictive inference for multi-state processes

In this section we derive some general predictive estimator for summary measure of a multi-state process using the blocks $\{\mathbf{Y}_m\}_{m\geq 1}$.

In practice, one may observe M trajectories $\{\mathbf{Y}_m\}_{1\leq m\leq M}$. Each observation $\mathbf{Y}_m = \{Y_n^{(m)}\}_{0\leq n\leq \tau_m}$ follows a multi-state process on a countable state space $\mathcal{G}$, observed over a random length of time τ_m. Call τ_m the survival time of the m-the observed trajectories. We want to estimate basic characteristics of the underling process, like the one-step or n-step ahead transition probability or the expected time that an observation spends in a certain state during her lifetime.

As noted in remark 11, we use the exchangeable sequence of mixture of inhomogeneous Markov chains $\{\mathbf{Y}_m\}_{m\geq 1}$ of the RUP to model the observed data. A predictive estimator $\mathbb{E}[\phi(\mathbf{Y}_{M+1})|\mathbf{Y}_1, \cdots, \mathbf{Y}_M]$ can be derived through the predictive rule of the RUP. This estimator will be equal to a Bayesian estimator of $\mathbb{E}[\phi(\mathbf{Y}_{M+1})|\Pi]$ with respect to a square loss function, where $\Pi = \{\Pi(t)_{i,j}\}_{i,j\in\mathcal{G},t\geq 0}$ denotes the random transition matrix conditional on which all M observations are independent time-inhomogeneous Markov chains.

At first, we focus on the estimation of the one- and n-step transition probabilities and the expected time an observation spend in a fixed state during the lifetime. Define for an arbitrary $m \geq 1$ and for a state $j \in \mathcal{G}$ the occupation-time to state j during the lifetime $\mathbf{Y}_m$ by $T_j(\mathbf{Y}_m) =$

$\sum_{n\geq 0} I\{Y_n^{(m)} = j\}$. To keep expressions simple, let

$$\Lambda_M(0) = E[\Pi(0)_{-0}|(\mathbf{Y}_j)_1^M] = \left(\frac{n_{(0,0)}(j) + N_{\zeta_M}(0,j,0)}{\sum_{v=0}^k n_{(0,0)}(v) + M};\ j \in \mathcal{G}\right)', \tag{3.22}$$

$$\Lambda_M(t) = E[\Pi(t)_{-0}|(\mathbf{Y}_j)_1^M] = \left(\frac{n_{(i,t)}(j) + N_{\zeta_M}(i,j,t)}{\sum_{v=0}^k n_{(i,t)}(v) + N_{\zeta_M}(i,t)};\ i,j \in \mathcal{G}\right) \tag{3.23}$$

for $t \geq 1$ denote the updated version of initial sub-transition matrices $\{\Lambda^{(0)}(t)\}_{t\geq 0}$, which are updated according to (3.6) for the observed data $\{\mathbf{Y}_m\}_{1\leq m\leq M}$.

Lemma 6. *For state i and j in $\mathcal{G}$*
(i) the predicted one-step transition probability from i to j at time n is given by

$$\hat{P}(Y_{n+1} = j|Y_n = i) = \frac{n_{(i,n)}(j) + N_{\zeta_M}(i,j,n)}{\sum_{v\in\mathcal{G}\cup\{0\}} n_{(i,n)}(v) + N_{\zeta_M}(i,n)}; \tag{3.24}$$

(ii) the predicted n-step transition probabilities into state j is given by

$$\hat{P}(Y_n = j) = \left(\Lambda_M(0)\prod_{v=1}^{n-1}\Lambda_M(v)\right)_j; \tag{3.25}$$

(iii) the predicted average time spend in state $j \in \mathcal{G}$ during a lifetime is given by

$$\hat{E}[T_j(Y)] = \sum_{n\geq 1}\left(\Lambda_M(0)\prod_{v=1}^{n-1}\Lambda_M(v)\right)_j, \tag{3.26}$$

where we define $\prod_{n=0}^{-1}$ to be equal to the identity matrix I.

In many situations one likes to estimate the expected lifetime $\mathbb{E}[\tau_m|\Pi]$ or the survival function $\{P(\tau_m > n|\Pi), n \geq 0\}$ as well. The predictive estimator for both quantities is given in the following lemma

Lemma 7. *(i) The predictive probability of surviving more than n units of*

time is given by

$$\hat{P}(\tau > n) = \sum_{j \in \mathcal{G}} \Big[\Lambda_M(0) \prod_{v=1}^{n} \Lambda_M(v) \Big]_j ; \tag{3.27}$$

(ii) the predicted total life-time is given by

$$\hat{E}(\tau) = \sum_{n \geq 0} \sum_{j \in \mathcal{G}} \Big[\Lambda_M(0) \prod_{v=1}^{n} \Lambda_M(v), \Big]_j \tag{3.28}$$

where $\prod_{n=0}^{-1} = I$.

In this section we considered some predictive estimator for selected measures of a multiple-state process. Estimator for further functions may be expressed in terms of the predictive law of the RUP in a similar way.

3.5 Right censoring and missing at random

In this section we like to discuss how to deal with censoring when we use a RUP, but the data $\{\mathbf{Y}_m\}$ may not be observed entirely.

Each of the blocks $\mathbf{Y}_m = \{Y_n^{(m)}\}_{0 \leq n \leq \tau_n}$ terminates after a finite time, τ_m, where no assumption was made about the cause of termination. In practice, one may observes measurements $\mathbf{Y}_m = (Y_t^{(m)} = i_t)_{0 \leq t \leq \tau_m^*}$ under right censoring, where $\tau_m^* = \min(\tau_m, \gamma_m)$ and γ_m is the censoring time of observation m. In this case one can adapt to right censoring. Suppose observation m is censored at time $\tau_m^* = \min(\tau_m, \gamma_m)$ (hence $\gamma_m < \tau_m$). The updating mechanism (3.19) and (3.21) for such a sequence is given by

$$N_{\zeta_m^*}(i_t, i_{t+1}, t) = N_{\zeta_m^*-1}(i_t, i_{t+1}, t) + 1 \quad \text{for } 1 \leq t \leq \tau_m^* - 1 \text{ and} \tag{3.29}$$

$$N_{\zeta_m^*}(i_{\tau_m^*}, 0, \tau_m^*) = N_{\zeta_{m-1}^*}(i_{\tau_m^*}, 0, \tau_m^*) \tag{3.30}$$

As noted in [152], the $(0,0)$-blocks remain exchangeable provided that the censoring mechanism is independent of the multiple-state process.

Next to right censoring one may also needs to deal with missing at random (MAR) for some of the repeated measurements. This case corresponds to

observing some paths $\mathbf{Y}_m$ only partially, i.e.

$$\mathbf{Y}_m = \left(Y_0^{(m)}, \cdots, Y_{c1}^{(m)}, \text{missing}, Y_{c2}^{(m)}, \cdots, Y_{\tau_m}^{(m)}\right).$$

Where 'missing' corresponds to the unobserved string $\mathbf{J}_m = \left(Y_{c1+1}^{(m)}, \cdots, Y_{c2-1}^{(m)}\right)$ of $\mathbf{Y}_m$, which depends on the observed path of observation m and on all other observations, $\mathbf{Y}_n$ $n \neq m$, as well.

For such cases we may use the following algorithm. The idea is the following. Suppose w observations contain a 'missing' string and denote this observations with $(\mathbf{Y}_{M_1}, \mathbf{Y}_{M_2}, \cdots, \mathbf{Y}_{M_w})$ and let $\mathcal{W} = \{M_1, \cdots M_w\}$ denote the corresponding index set of such observations. By exchangeability and MAR for any permutation σ of $\{1, \cdots, w\}$

$$(\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_M) \stackrel{d}{=} (\mathbf{Y}_i \; i \notin \mathcal{W}, \mathbf{Y}_{M_\sigma(1)}, \mathbf{Y}_{M_\sigma(2)}, \cdots, \mathbf{Y}_{M_\sigma(w)}). \tag{3.31}$$

Hence we may draw first for all $j = 1, \cdots, w$ the missing string $(\mathbf{J}_{M_j} | (\mathbf{Y}_i)_1^M, \mathbf{J}_{M_i} \; i \neq j)$ and complete $\mathbf{Y}_{M_i}$ for ; secondly we compute all counting processes (3.19), (3.21); and third estimates the unknown quantities of interest as explained above. The final estimator will be obtained by repeatedly drawing the missing strings and then average over the estimator over the total number of draws (after some burn-in period). The actually draw of $\mathbf{J}_{M_j}$ will be implemented by sampling for $t = c_1 + 1, \cdots, c_2 - 1$ from

$$\begin{aligned}
&\mathbb{P}\big(Y_t^{(M_j)} = j | (\mathbf{Y}_i)_1^M, \mathbf{J}_{M_i} \; i \neq j, \mathbf{J}_{M_i, -t}\big) \\
&\propto \big(n_{(Y_{t-1}^{(M_j)}, t-1)}(j) + N^*_{\zeta_M}\big(Y_{t-1}^{(M_j)}, j, t-1\big)\big)\big(n_{(j,t)}\big(Y_{t+1}^{(M_j)}\big) + N^*_{\zeta_M}(j, Y_t^{(M_j)}, t+1)\big)
\end{aligned} \tag{3.32}$$

for $j \in \mathcal{G}$ where $\mathbf{J}_{M_i, -t} = \left(Y_{c1+1}^{(M_j)}, \cdots, Y_{t-1}^{(M_j)}, Y_{t+1}^{(M_j)}, \cdots, Y_{c2-1}^{(M_j)}\right)$ (and for each $t < c_2 - 1$ we use the required $Y_{t+1}^{(M_j)}$ from the previous iteration) and

$$N^*_{\zeta_M}(i, j, t) = \sum_{m \notin \mathcal{W}} I(Y_t^{(m)} = i, Y_{t+1}^{(m)} = j) + \sum_{M_l \in \mathcal{W}, l \neq j} I(Y_t^{(M_j)} = i, Y_{t+1}^{(M_j)} = j).$$

3.6 Numerical Illustration

In this section we numerical illustrate the statistical procedure outlined in the last section. Consider the problem of heart transplantation monitoring. Typically one wants to monitor the health status and the survival time of a patient after heart transplantation. In this context, cardiac allograft vasculophathy (CAV), which is a deterioration of the arterial walls, is a major cause of death among long term survivors of heart transplantation [143]. Therefore, the analysis of onset and progression of CAV is a major issue in heart transplantation monitoring.

We used heart transplantation monitoring data from Papworth Hospital, UK (available within the R package 'msm', [80]). A total of 622 patients are yearly re-examined after transplantation. Every re-examination provides for each patient a CAV measurements in form of three different states; state 1 represents no CAV, 2 denotes moderate CAV and 3 represents severe CAV. The reference point for each patient is the time of heart transplantation. This corresponds to state $(0, 0)$ in our previous notation.

We assume that, conditional on a transition array Π, the process of heart transplantation monitoring follows a time-inhomogeneous Markov chain. Our aim is to estimate the unknown transition array Π and functions of the process and not to assess the effects of covariates. Hence, we assume for simplicity that patients are exchangeable[1]. We use the exchangeable sequence of blocks $\{\mathbf{Y}_m\}$ defined in (3.18) of a RUP with state-space $\mathcal{S} = \{(0, 0)\} \cup (\{1, 2, 3\} \times \mathbb{N})$ to model the heart transplantation monitoring. $\mathcal{G} = \{1, 2, 3\}$ denotes the potential CAV states and 0 denotes the event of heath transplantation at time zero. The first 622 blocks $\{\mathbf{Y}_m\}_{1 \leq m \leq 622}$ of the RUP represent the observed trajectories of all 622 patients.

Figure 3.1 shows graphical the relation between individual trajectories of three patients $\mathbf{Y}_i, \mathbf{Y}_j, \mathbf{Y}_k$ and the resulting partial exchangeable RUP $\{X_n\}$. For example patient i stays at the first two re-examinations in state 1, moves

[1]The assumption of exchangeability might be questionable in real applications since patients have different ages at the time of transplantation. But previous analysis of the data showed that only CAV and sex have a significant effect on the survival status after transplantation. Hence one could run a separate analysis for each sex.

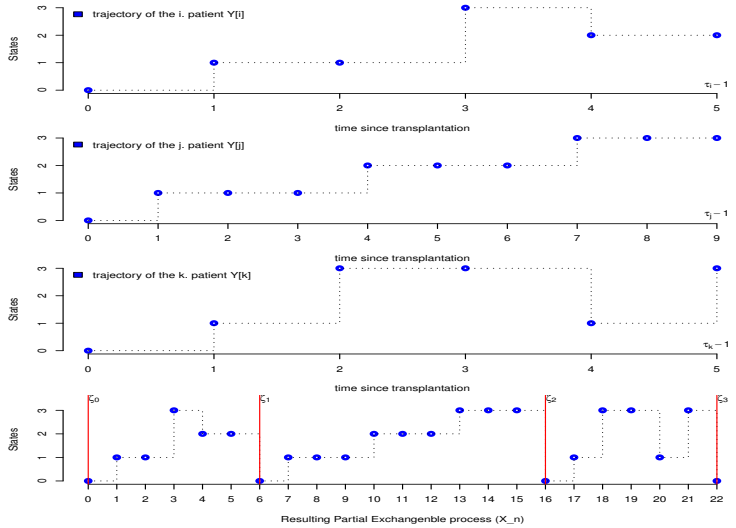


Figure 3.1: Structure of the Papworth Hospital data

to state 3 at the third re-examination. At the fourth and fifth re-examination he moves back to state 2 and dies after the fifth re-examination. Her lifetime is $\tau_i = 5$. Note that in this example transitions of the form $i \to i-1, i, i+1$ are possible for state 2. By exchangeability of the blocks $\{\mathbf{Y}_m\}_{m\leq 622}$, the finite dimensional joined likelihood is invariant of the order of observing the patients. Therefore the predictive distribution of $\mathbf{Y}_{623}$ given $\{\mathbf{Y}_m\}_{1\leq m\leq 622}$ is also invariant of the order of observing $\{\mathbf{Y}_m\}_{1\leq m\leq 622}$.

Figure 3.2(a) shows the estimated n-step transition probabilities for all three CAV states, while Figure 3.2(b) shows the estimated discrete survival function after heart transplantation. The expected time spend in CAV state 1, 2 and 3 are 6.75, 1.8 and 0.99 years respectively whereas the expected length of survival after transplantation equals $10.54 = 1 + 6.75 + 1.8 + .99$ years.

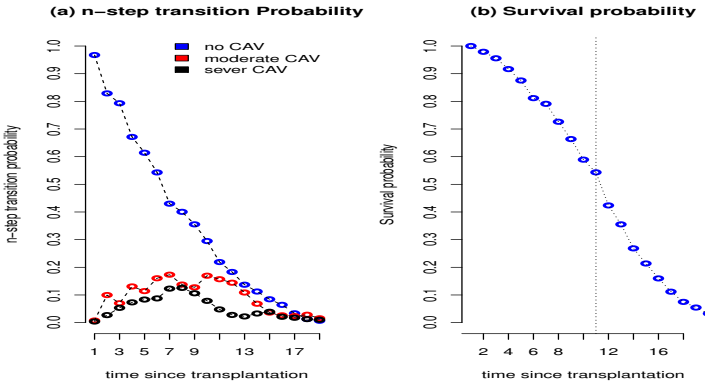


Figure 3.2: Predicted n-step transition probabilities of being in CAV state 1, 2 or 3 after transplantation and the predicted survival function. (State 1 denotes no CAV, state 2 denotes moderate CAV and state 3 denotes severe CAV)

3.7 Concluding Remarks

In the present paper we introduced a particular reinforce urn process (RUP) for the analysis of multi-state processes. We used the sequence of trajectories between successive hitting times to the initial state of the RUP to construct an exchangeable sequence of stochastic processes $\{\mathbf{Y}_m\}_{m\geq 1}$. Every element $\mathbf{Y}_m$ is a mixture of time-inhomogeneous Markov chains stopped at a random point in time τ_m. Inference for functions of the multiple-state process can be implemented through the predictive law of the RUP.

The process can be used for problems in survival analysis, event history modelling or system reliability where one observes the same underlining process for several independent replicas. As a numerical illustration we applied the RUP to inference in heart transplantation monitoring. A clear weakness of the proposed method is the missing adjustment for covariate effects, which reduces the usefulness of the method in practice substantially.

Potential extensions are manly directed towards the construction of mix-

ture of time-inhomogeneous Markov chains on multivariate state spaces of an arbitrary dimension. As an example, suppose that one wants to monitor K potentially correlated health indicators for M patients with varying initial conditions. The object of interest could be to monitor jointly all K indicators and the survival time with an implicit correlation structure. This can be done by considering hierarchical urn schemes and will be the object of future work.

3.8 Appendix: Proofs

Proof of lemma 2. We show (3.9) by induction on n. By definition of the RUP $\{X_n\}_{n\geq 0}$

$$\begin{aligned} P[\zeta_1 > n] &= P[X_0 = (0,0), X_1 \neq (0,0), \cdots, X_n \neq (0,0)] \\ &= \sum_{i_1,\cdots,i_n \in \mathcal{G}} P[X_0 = (0,0), X_1 = (i_1,1), \cdots, X_n = (i_n, n)] \\ &= \sum_{i_n \in \mathcal{G}} \left[\Lambda^{(0)}(0) \prod_{t=1}^{n-1} \Lambda^{(0)}(t) \right]_{i_n} \end{aligned}$$

where the subscript i_n stands for the i_n-th component of a vector. The second condition implies that

$$P(\zeta_1 < \infty) = 1 - \lim_{n\to\infty} P(\zeta_1 > n) = 1. \tag{3.33}$$

Assume that $P[\cap_{i=0}^{m}\{\zeta_i < \infty\}] = 1$ for a fixed $m \in \mathbb{N}$ and set $K_m = \max_{1\leq j\leq m}\{\zeta_j - \zeta_{j-1}\} + 1$. Notice that K_m is finite with probability one. Let $V(\zeta_1, \cdots, \zeta_m) = (s_0, \cdots, s_{\zeta_m})$ be a trajectory of the process until ζ_m such that $s_j = (0,0)$ if $j \in \{\zeta_0, \cdots, \zeta_m\}$ and $s_j \neq (0,0)$ if $j \notin \{\zeta_0, \cdots, \zeta_m\}$. Then for $n > K_m + \zeta_m$

$$P[\zeta_{m+1} > n | V(\zeta_1, \cdots, \zeta_m), K_m] = \sum_{i\in\mathcal{G}} \left[\Lambda^{(m)}(0) \prod_{t=1}^{K_m-1} \Lambda^{(m)}(t) \prod_{t=K_m}^{n-1} \Lambda^{(0)}(t) \right]_i$$

a.s., where $\Lambda^{(m)}(0)$ is a sub-distribution and $\Lambda^{(m)}(t)$ a stochastic sub-matrix, both defined analogous to $\Lambda^{(0)}(0)$ and $\Lambda^{(0)}(t)$ but the elements are updated according to (3.6) for the sequence $V(\zeta_1, \cdots, \zeta_m) = (s_0, \cdots, s_{\zeta_m})$. It follows by condition two that $\prod_{t=K_m}^{n-1} \Lambda^{(0)}(t) \to \mathbf{0}$. Hence

$$\begin{aligned} P[\zeta_{m+1} < \infty | V(\zeta_1, \cdots, \zeta_m), K_m] &= 1 - \lim_{n\to\infty} P[\zeta_{m+1} > n | V(\zeta_1, \cdots, \zeta_m), K_m] \\ &= 1 \ a.s. \end{aligned}$$

The induction step follows now from

$$P(\zeta_{m+1} < \infty) = E\Big[E\big[P(\zeta_{m+1} < \infty | V(\zeta_1, \cdots, \zeta_m), K_m) \ \big|\zeta_1, \cdots, \zeta_m\big]\Big] = 1.$$

□

Proof of Lemma 3. Suppose $\{X_n\}$ is recurrent, by theorem 2.20 in [117] conditional on a sample $\{X_j = s_k\}_{0\leq k\leq n}$ from the RUP, $\{X_{n+j}\}_{j\geq 0}$ is again a RUP on $\mathcal{S}$ with initial state s_n and initial urn compositions

$$\tilde{U}_{(i,t)} = \{n_{(i,t)}(0) + N_n[(i,t)(0,0)], n_{(i,t)}(j) + N_n[(i,t)(j,t+1)]; j \in \mathcal{G}\}. \quad (3.34)$$

Furthermore also $\{X_{n+j}\}_j$ is recurrent. Let $\{\zeta_i\}_{0\leq i\leq k_n}$ denotes all hitting times to $(0,0)$ of $\{X_k = s_k\}_{0\leq k\leq n}$ (where $\zeta_i \leq n$ for $1 \leq i \leq k_n < n$) and we define $\{\hat{\zeta}_n\}_{n\geq 0}$ to be the sequence of hitting times to $(0,0)$ for $\{X_{n+j}\}_{j\geq 0}$. By the same argument as in the induction step of lemma 29 it follows that $\hat{\zeta}_1$ is finite almost surely (where we condition on $\cap_{i=0,\cdots,k_n}\{\zeta_i < \infty\}$). Assume that $P(\cap_{i=1}^m\{\hat{\zeta}_i < \infty\}) = 1$, and use again the same argumentation as in the induction step of lemma 29 to show that $\hat{\zeta}_{m+1}$ is a.s. finite, but conditioning this time on $[\cap_{i=0,\cdots,k_n}\{\zeta_i < \infty\}] \cap [\cap_{i=0,\cdots,m}\{\hat{\zeta}_i < \infty\}]$.

Hence, $\{X_{n+j}\}_{j\geq 0}$ is a recurrent RUP with initial urn compositions given by (3.34) Now (i) and (ii) follow simply from theorem 10 apply to the RUP $\{\tilde{X}_j\}_{j\geq 0} := \{X_{n+j}\}_{j\geq 0}$, where $\tilde{X}_j := X_{n+j}$, with initial urn compositions (3.34). □

Proof of Corollary 2. The proof follows from the previous lemma by observing that, give $\{\zeta_n = l\}$, $\{X_k\}_{0\leq k\leq \zeta_n}$ is determined by $\{X_k\}_{0\leq k\leq l}$ and $P(\zeta_n < \infty) = 1$. □

Proof of Lemma 4. Fix $n > 1$ and consider the first n blocks $\{\mathbf{B}_m\}_{1\leq m\leq n}$ of the RUP. Let σ be a permutation of $\{1, \cdots, n\}$. We need to show

$$(\mathbf{B}_1, \cdots, \mathbf{B}_n) \stackrel{d}{=} (\mathbf{B}_{\sigma(1)}, \cdots, \mathbf{B}_{\sigma(n)}). \quad (3.35)$$

By definition of the $(0,0)$-blocks, we have

$$\begin{aligned}
&P\Big[\bigcap_{k=1}^{n}\Big\{\mathbf{B}_k=\big((0,0),(i_1^k,1)\cdots,(i_{w_k}^k,w_k)\big)\Big\}\Big]\\
=&P\Big[\bigcap_{k=1}^{n}\Big\{X_{\zeta_{k-1}}=(0,0),X_{\zeta_{k-1}+1}=(i_1^k,1),\cdots,X_{\zeta_k-1}=(i_{w_k}^k,w_k)\Big\}\Big]\\
=&\prod_{t=0}^{\infty}\prod_{j=0}^{k}\frac{n_{(j,t)}(0)^{[N_\phi((j,t);(0,0))]}\prod_{i=1}^{k}n_{(j,t)}(i)^{[N_\phi((j,t);(i,t+1))]}}{\Big(\sum_{v=0}^{k}n_{(j,t)}(v)\Big)^{[N_\phi(j,t)]}}. \qquad (3.36)
\end{aligned}$$

The last equality follows from the finite dimensional joined law of a RUP (equation (2.4) in [117]), where $a^{[n]}=a(a+1)\cdots(a+n-1)$, by convention $a^{[0]}=1$, and $N_\phi[(i,t),(j,s)]$ denotes the number of transitions from $(i,t)\in\mathcal{S}$ to $(j,s)\in\mathcal{S}$ within the string $\big((0,0),(i_1^1,1)\cdots,(i_{w_1}^1,w_1),(0,0),\cdots,(i_{w_m}^m,w_m)\big)$.

Within any block $\mathbf{B}_m$, a transition from (i,t) to $(j,t+1)$ occurs if and only if the $(t+1)$-th and $(t+2)$-th coordinate of $\mathbf{B}_m$ equals (i,t) and $(j,t+1)$. Now, for $0\le t\le\max_{1\le k\le n}(w_k)$ and $i\in\mathcal{G}\cup\{0\}$, $j\in\mathcal{G}$, we have

$$\begin{aligned}
N_\phi[(i,t);(j,t+1)]&=\sum_{k=1}^{n}I(i_t^k=i,\ i_{t+1}^k=j,w_k>t)\\
&=\sum_{k=1}^{n}I(i_t^{\sigma(k)}=i,\ i_{t+1}^{\sigma(k)}=j,w_{\sigma(k)}>t)\\
&=:N_\mu[(i,t);(j,t+1)]
\end{aligned}$$

where $N_\mu[(i,t);(j,t+1)]$ denotes the number of transitions from state (i,t) to $(j,t+1)$ within the string $\big((0,0),(i_1^{\sigma(1)},1),\cdots,(i_{w_1}^{\sigma(1)},w_{\sigma(1)}),(0,0),\cdots,(i_{w_{\sigma(m)}}^{\sigma(m)},w_{\sigma(m)})\big)$. Similarly for $(i,t)\in\mathcal{S}$

$$\begin{aligned}
N_\phi((i,t),(0,0))&=\sum_{k=1}^{n}I(i_t^{(k)}=i,w_k=t)\\
&=\sum_{m=1}^{n}I(i_t^{\sigma(k)}=i,w_{\sigma(k)}=t)=N_\mu((i,t),(0,0)).
\end{aligned}$$

For $t>\max_{1\le k\le n}w_k$ and all $(i,t),(j,s)$ in $\mathcal{S}$, clearly $N_\phi((i,t),(j,s))=0=$

$N_\mu((i,t),(j,s))$. Consequently also $N_\phi(i,t) = N_\mu(i,t)$ for $(i,t) \in \mathcal{S}$. Therefore (3.36) equals

$$\prod_{t=0}^{\infty}\prod_{j=0}^{k} \frac{n_{(j,t)}(c_0)^{[N_\mu((j,t);(0,0))]} \prod_{i=1}^{k} n_{(j,t)}(c_i)^{[N_\mu((j,t);(i,t+1))]}}{\left(\sum_{v=0}^{k} n_{(j,t)}(c_v)\right)^{[N_\mu(j,t)]}}$$
$$= P\Big[\bigcap_{k=1}^{n}\Big\{\mathbf{B}_k = \big((0,0),(i_1^{\sigma(k)},1)\cdots,(i_{w_{\sigma(k)}}^{\sigma(k)},w_{\sigma(k)})\big)\Big\}\Big].$$

Since σ and n are arbitrary the result follows. □

Proof of Lemma 5. For a fixed n and any sequence $(i_t)_{0\le t\le k}$ of states in $\mathcal{G}$

$$P(\cap_{t=0}^{k}\{Y_t^{(n)} = i_t\}|\Pi) = P(\cap_{t=0}^{k}\{X_{\zeta_{n-1}+t} = (i_t,t)\}|\Pi) \quad a.s$$
$$= \delta_0(i_0)\prod_{t=0}^{k-1}\Pi(t)_{i_t,i_{t+1}} \quad a.s,$$

which shows that, given Π, $\mathbf{Y}_n$ is an inhomogeneous Markov chain a.s.. Furthermore, $\mathbf{Y}_n$ is a measurable function of $\mathbf{B}_n$. Now, since given Π, $\{\mathbf{B}_n\}_{n\ge1}$ is an iid sequence, we deduce that, given Π, $\mathbf{Y}_n$ are mutually independent. □

Proof of Lemma 6. By definition of $\{\mathbf{Y}_m\}$

$$\hat{P}(Y_{n+1} = j|Y_{n+1} = i) = P(Y_{n+1}^{(M+1)} = j|Y_n^{(M+1)} = i, \mathbf{Y}_1\cdots,\mathbf{Y}_M)$$
$$= P(X_{\zeta_M+n+1} = j|X_0,\cdots X_{\zeta_M+n-1}, X_{\zeta_M+n} = i)$$
$$= E[\Pi^{(\zeta_M+n)}(n)_{i,j}]$$
$$= \frac{n_{(i,n)}(j) + N_{\zeta_M+n}(i,j,n)}{\sum_{v\in\mathcal{G}\cup\{0\}} n_{(i,n)}(v) + N_{\zeta_M+n}(i,n)},$$

where the last two equalities hold by lemma 3 and corollary 2. Since $N_{\zeta_M+n}(i,j,n) =$

$N_{\zeta_M}(i,j,n)$, this yields (i). Furthermore (ii) can be derived similar to (i),

$$\begin{aligned}
&\hat{P}[Y_n^{(M+1)} = j] \\
&\quad = P[Y_n^{(M+1)} = j|\mathbf{Y}_1, \cdots, \mathbf{Y}_M] \\
&\quad = P[X_{\zeta_M} = (0,0), \cap_{k=1}^{n-1}\{X_{\zeta_M+k} \in \mathcal{G} \times \{k\}\}, X_{\zeta_M+n} = (j,n)|X_0, \cdots, X_{\zeta_M}] \\
&\quad = E\left[\left(\Pi_{-0}^{(\zeta_M)}(0)\prod_{k=1}^{n-1}\Pi_{-0}^{(\zeta_M)}(k)\right)_j\right] = \left(\Lambda_M(0)\prod_{k=1}^{n-1}\Lambda_M(k)\right)_j,
\end{aligned}$$

where last two equalities hold by corollary 2.
Finally, using (ii), the predicted average time spend in state i for observation $M+1$ is given by

$$\begin{aligned}
\hat{\mathbb{E}}(T(Y)_j) &= \mathbb{E}\Big[\sum_{n\geq 0} I(Y_n^{(M+1)} = j)\Big|\mathbf{Y}_1, \cdots, \mathbf{Y}_M\Big] \\
&= \sum_{n\geq 1} P(Y_n^{(M+1)} = j|\mathbf{Y}_1, \cdots, \mathbf{Y}_M) \\
&= \sum_{n\geq 1}\left(\Lambda_M(0)\prod_{k=1}^{n-1}\Lambda_M(k)\right)_j,
\end{aligned}$$

where the second equality hold by the monotone convergence theorem for conditional expectations. □

Proof of Lemma 7. We use again the predictive rule of the RUP and the definition of $\{\mathbf{Y}_m\}$ to express τ_{M+1} as

$$\begin{aligned}
&\hat{P}(\tau > n) = P(\tau_{M+1} > n|\mathbf{Y}_1, \cdots \mathbf{Y}_M) \\
&= P(Y_0^{(M+1)} = 0, \cap_{k=1}^{n+1}\{Y_k^{(M+1)} \in \mathcal{G}\}|\mathbf{Y}_1, \cdots \mathbf{Y}_M) \\
&= \sum_{j\in\mathcal{G}}\Big[\Lambda_M(0)\prod_{v=1}^{n}\Lambda_M(v)\Big]_j
\end{aligned}$$

where the last two equalities holds again by Corollary 2. Finally (ii) follows

from (i) by

$$\begin{aligned}\hat{\mathbb{E}}[\tau] &= E(\tau_{M+1}|\mathbf{Y}_1, \cdots, \mathbf{Y}_M) \\ &= \sum_{n\geq 0} P(\tau_{M+1} > n|\mathbf{Y}_1, \cdots, \mathbf{Y}_M) \\ &= \sum_{n\geq 0}\sum_{j\in\mathcal{G}} \Big[\Lambda_M(0)\prod_{k=1}^{n}\Lambda_M(k)\Big]_j\end{aligned}$$

□

Chapter 4

Reinforced Bivariate Urn Processes

4.1 Introduction

Suppose we observe the random movement of a particle on a countable bivariate state space $G_1 \times G_2$ over a random length of time, say τ. The underlying stochastic process is supposed to be Markov, conditional on an unknown transition operator. The stochastic process will generally not be time-homogeneous. Suppose we obtain information about the underlying stochastic process by observing the same stochastic phenomena for a finite number of exchangeable particles. Frequently one wants to learn from the observed particle how future particle behave. In the following, we use a particular reinforced stochastic process to study such a problem and apply this process to a Bayesian analysis of bivariate multi-state processes.

Reinforced stochastic processes and in particular random walks with reinforcement, introduced by Coppersmith and Diaconis ([34] see also [123, 42, 36]), are a class of processes characterized, loosely speaking, by the property that trajectories frequently traversed by the process in the past become more likely to be revisited in the future than trajectories which have not yet been traversed. A statistical model based on a reinforced stochastic process can therefore be seen as a sequential learning process. Furthermore classical

random walks with reinforcement are partial exchangeable. Therefore a recurrence reinforced random walk can be represented as a mixture of Markov chains. This feature links a sequential learning process with a Bayesian analysis of Markov chains.

In the following paper we study a bivariate urn process, which can be seen as a random walk on a space of some hierarchical Pólya urns. The process has, under recurrence, a unique representation as a mixture of Markov chains too. In a second step we use a sequence of stopping times of the process and the trajectories between successive stopping times to construct an exchangeable sequence of stochastic processes $\{\mathbf{Y}_n\}_{n\geq 1}$, where for every $n \geq 1$ the string $\mathbf{Y}_n = \{(Y_m^{(n,1)}, Y_m^{(n,2)})\}_{0\leq m\leq \tau_n}$ is a mixtures of bivariate time-inhomogeneous Markov chains on $G_1 \times G_2 \subset \mathbb{N}$ stopped at a random point in times τ_n. The hierarchical structure of the hierarchical Pólya urns makes it possible to model dependence between values taken by the process in G_1 and in G_2.

The outline of the paper is the following. In section 4.2 we fist motivate the practical relevance of our process for stochastic modelling. In a next step we and introduce the actual process. Base of the process, we construct an exchangeable sequence of mixture of inhomogeneous bivariate Markov chains in section 4.3. In section 4.3.1 we use the constructed sequence for prediction for Markov chains. In section 4.3.2 we numerically illustrate the procedure for some simulated data for one of the motivating examples. In the final section we give some concluding remarks.

4.2 A Reinforced Bivariate Urn Process.

4.2.1 Motivating examples.

The aim of this subsection is to motivate the construction of the bivariate urn process (BUP). The first example will also be used in section 4.3.2 as numerical illustration.

Example 1. *Suppose that patients share the same initial event like the occurrence of a certain disease or a surgery. Each patient is followed on a*

regular base until death or until a predefined point. The follow-up may contain information on the level of some co-disease D1 and D2 as a consequence of the initial event (for instance D1 may be hypertension and D2 be diabetes), where D1 and D2 are potentially correlated. We may measure D1 on a scale of $G_1 = \{1, \cdots, N\}$ *and D2 on a scale of* $G_2 = \{1, \cdots, M\}$*. One could be interested in modelling jointly the evolution of D1, D2 during reaming lifetime of a patient. In this situation we have data of the form* $\{\boldsymbol{Y}_n\}_{1 \leq n \leq K}$ *where for every* $1 \leq n \leq K$ $\boldsymbol{Y}_n = (Y_1^{(n)}, \cdots, Y_{\tau_n - 1}^{(n)}, Y_{\tau_n}^{(n)}) \in S_{(G_1 \times G_2)}$*. Here* $S_{(G_1 \times G_2)}$ *denotes the space of all finite sequences of elements in* $G_1 \times G_2$ *and* τ_n *denoted the survival time or the end of the observation-period of the* n*-th patient.*

Example 2. *Suppose our interest lies in studying the decision of the first pregnancy for young women in relation to their current occupational status and marital status. Hence let* $G_1 = \{1 =$ *'in education'*$, 2 =$ *'employed'*$, 3 =$ *'self-employed'*$, 4 =$ *'unemployed'*$\}$ *and* $G_2 = \{1 =$ *'single'*$, 2 =$ *'in committed relationship'*$, 3 =$ *'engaged'*$, 4 =$ *'divorced'*$, 5 =$ *'widowed'*$\}$ *and let* $W = \{l \in \mathbb{N} : l =$ *current age* $- v\} \subset \mathbb{N}$*, where* v *is the age of entrance into the biological fertile lifespan (i.e. set* $v = 12$ *or* $v = 15$*). Denote with* τ *the age of delivery. Clearly the decision for a first child and the age at delivery depends on both the employment and marital history of a woman. Usually the employment and marital history are dependent as well. Typical data for such processes arise from longitudinal surveys or register data in a prospective or retrospective form. Therefore we have data of the form* $\{\boldsymbol{Y}_n\}_{1 \leq n \leq K}$ *where* $\boldsymbol{Y}_n = (Y_1^{(n)}, \cdots, Y_{\tau_n - 1}^{(n)}, Y_{\tau_n}^{(n)}) \in S_{(G_1 \times G_2)}$*.*

In both examples the simplest form of time dependence would be Markov, but clearly both processes are most likely not homogeneous in time. It should be noted

4.2.2 Construction of the BUP

To define a bivariate urn process we will take the following elements. As a state space we choose $S = \{(0, 0, 0)\} \cup \big(G_1 \times G_2 \times \mathbb{N}\big)$, where $\mathbb{N}$ indicates

the set of positive integers and $G_1 = \{1, \cdots, N\}$, $G_2 = \{1, \cdots, M\}$ for some finite integers $N, M < \infty$.

We also fix two sets of labels. $C_1 = \{a_0, a_1, \cdots, a_N\}$ will be a color set at the 'first level' and $C_2 = \{b_1, b_2, \cdots, b_M\}$ be a color set at the 'second level'.

Now, with every state (i, j, t) in S we connect a first level Pólya urn $U^{(1)}(i, j, t)$ with color set C_1 containing $N + 1$ colors of initial composition

$$\big(n_{(i,j,t)}(a_0), n_{(i,j,t)}(a_1), \cdots n_{(i,j,t)}(a_N)\big) \in \mathbb{R}_+^{N+1}, \tag{4.1}$$

where $n_{(i,j,t)}(a_v)$ denotes the number of balls of color $a_v \in C_1$ initially in urn $U^{(1)}(i, j, t)$.

Additionally, we connect to every state (i, j, t) and every $v \in G_1$ a further 'second level' Pólya urn $U^{(2)}(i, j, t|v)$ with color set C_2 containing M colors of initial composition

$$\big(m_{(i,j,t|v)}(b_1), m_{(i,j,t|c)}(b_2), \cdots, m_{(i,j,t|v)}(b_M)\big) \in \mathbb{R}_+^M. \tag{4.2}$$

of $m_{(i,j,t|v)}(b_l)$ balls of color b_l.

To ensure that the resulting process will be well defined we assume the following

Assumption 1. *For all (i, j, t) in S and all v in G_1*

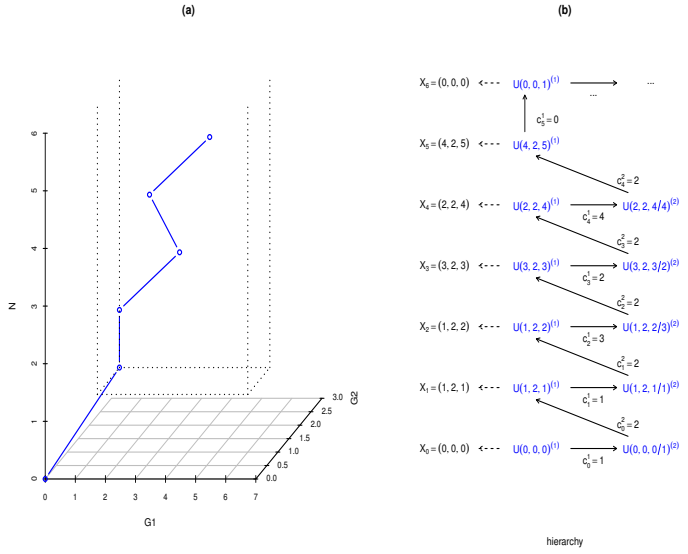
$$|U^{(1)}(i, j, t)| := \sum_{c=0}^{N} n_{(i,j,t)}(a_c) > 0 \tag{4.3}$$

$$|U^{(2)}(i, j, t|v)| := \sum_{c=1}^{M} m_{(i,j,t|v)}(b_c) > 0 \quad \textit{if } n_{(i,j,t)}(a_v) > 0. \tag{4.4}$$

This basically states that every first level Pólya urn is initially non-empty and if the number of balls of color a_v in the first level urn $U^{(1)}(i, j, t)$ is positive, then the second level urn $U^{(2)}(i, j, t|v)$ will be non-empty too. Informally speaking, a bivariate hierarchical urns process can be described by the following recursive sampling scheme.

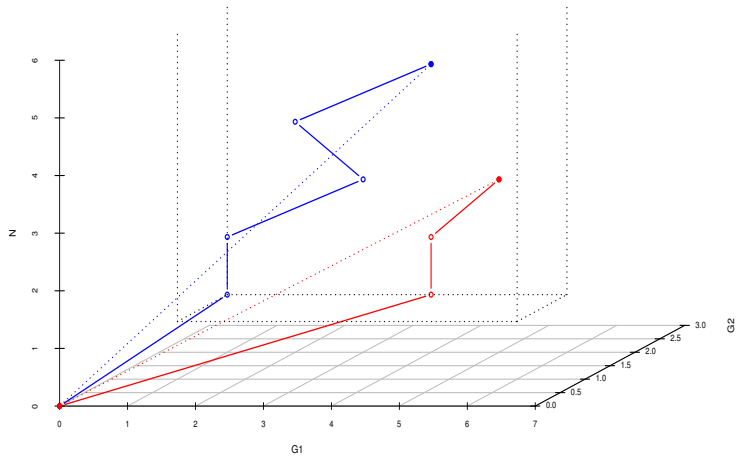
Procedure 1. *Let the initial state of the process be* $X_0 = (0,0,0)$ *and suppose we observe* $\{X_0 = (0,0,0), X_k = (i_k, j_k, t_k); 1 \leq k \leq n\}$ *(see figure 4.1(a)). Sample* X_{n+1} *given the first* $(n+1)$ *observations as follows. Sample at first a ball from the first level Pólya urn* $U^{(1)}(i_n, j_n, t_n)$*. If the color of the sampled ball equals* a_0 *set* $X_{n+1} = (0,0,0)$ *and add two balls of color* a_0 *to urn* $U^{(1)}(i_n, j_n, t_n)$*. If otherwise the color of the sampled ball equals* a_v *where* $v \in G_1$*, sample a second ball from the second level Pólya urn connected to* $((i_n, j_n, t_n), v)$*, that is from* $U^{(2)}(i_n, j_n, t_n|v)$ *(see figure 4.1(b)). If* b_w *(where* $w \in G_2$*) denotes the color of the ball drawn from urn* $U^{(2)}((i_n, j_n, t_n)|v)$ *set* $X_{n+1} = (v, w, t_n + 1)$ *and add two balls of color* a_v *to the first level urn* $U^{(1)}(i_n, j_n, t_n)$ *and add two balls of color* b_w *to the second level urn* $U^{(2)}((i_n, j_n, t_n)|v)$*.*

Figure 4.1: Sampling Scheme of an BUP on S with $|G_1| = 6$ and $|G_2| = 2$. (c_i^j denotes the index of the color drawn at time $i \geq 0$ at hierarchy $j \in \{1,2\}$.)



If the process is recurrence, meaning $P(X_n = (0,0,0)\ i.o.) = 1$, the pro-

Figure 4.2: Simulation of a trajectory of a BUP on S with $|G_1| = 6$ and $|G_2| = 2$ until the third hitting time to the initial state.



cess samples an infinite number of trajectories within the infinite high box $\{1, \cdots, N\} \times \{1, \cdots, N\} \times \mathbb{N}$, say $\mathbf{B} = \{\mathbf{B}_n\}_{n\geq 1}$. Where the n-th trajectory consists of the states visited by the process between the n-th and $(n+1)$-th hit to the initial state $(0, 0, 0)$. Figure 4.2 shows a simulated trajectory of a process (for $|G_1| = 6$ and $|G_2| = 2$) until the third hit to the initial state. We will consider this trajectories and related functions in more detail in section 4.3, where we use the trajectories to construct an exchangeable sequence of mixtures of time-inhomogeneous bivariate Markov chains. For the later analysis we translate the previous sampling scheme into the following equivalent definition

Definition 16. *The bivariate urn process $X = \{X_k\}_{k\geq 0}$ with values in S is defined recursively by $\mathbb{P}[X_0 = (0, 0, 0)] = 1$ and conditional on the history*

$\mathcal{F}_k = \sigma(\{X_m\}_{0\leq m\leq k})$ *until process time* $k \geq 1$ *where* $X_k = (w, v, l) \in S$

$$\mathbb{P}[X_{k+1} = (i, j, s)|\mathcal{F}_k] = \tag{4.5}$$

$$\begin{cases} \dfrac{n_{(w,v,l)}(a_0) + t_k(w, v, l)(0, 0, 0)}{\sum_{c=0}^{N} n_{(w,v,l)}(a_c) + t_k(w, v, l)} & \text{if } (i, j, s) = (0, 0, 0) \\ \left(\dfrac{n_{(w,v,l)}(a_i) + t_k(w, v, l)(i, \cdot, l + 1)]}{\sum_{c=0}^{N} n_{(w,v,l)}(a_c) + t_k(w, v, l)} \right. & \\ \left. \times \dfrac{m_{(w,v,l|i)}(b_j) + t_k(w, v, l)(i, j, l + 1)}{\sum_{c=1}^{M} m_{(w,v,l|i)}(b_c) + t_k(w, v, l)(i, \cdot, l + 1)} \right) & \text{if } (i, j, s) \in G_1 \times G_2 \times \{l + 1\}, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$t_k(w, v, l)(i, j, s) = \sum_{n=1}^{k} I(X_{n-1} = (w, v, l), X_n = (i, j, s)), \tag{4.6}$$

$$t_k(w, v, l)(i, \cdot, s) = \sum_{j=1}^{M} t_k(w, v, l)(i, j, s), \tag{4.7}$$

$$t_k(w, v, l) = \sum_{\boldsymbol{s}\in S} t_k(w, v, l)(\boldsymbol{s}) \tag{4.8}$$

denotes the counting processes of transitions between states and transitions into states until time k.

At first we are going to determine the finite dimensional law of the process. This will be helpful to explore some basic properties of the process. Since the transition rule of the process allows only some restricted transitions, we will call a sequence $\{(i_n, j_n, s_n)\}_{0\leq n\leq k}$ of $(k + 1)$ elements in S admissible if $(i_0, j_0, s_0) = (0, 0, 0)$ and, for $1 \leq n \leq k$, $(i_k, j_k, s_k) = (0, 0, 0)$ or $(i_k, j_k, s_k) \in G_1 \times G_2 \times \{s_{k-1} + 1\}$.

Proposition 1. *Let* $\{(i_n, j_n, s_n)\}_{0\leq n\leq k}$ *be a finite admissible sequence of*

states in S, then

$$\begin{aligned}
&\mathbb{P}[X_0 = (i_0, j_0, t_0), X_1 = (i_1, j_1, s_1), \cdots, X_k = (i_k, j_k, s_k)] \\
&= \prod_{\substack{l \geq 0 \\ i \in G_1 \cup \{0\} \\ j \in G_2 \cup \{0\}}} \prod_{v=1}^{N} \left[\frac{n_{(i,j,l)}(a_0)^{[t_k(i,j,l)(0,0,0)]} n_{(i,j,l)}(a_v)^{[t_k(i,j,l)(v,\cdot,l+1)]}}{\left(\sum_{c=0}^{N} n_{(i,j,l)}(a_c)\right)^{[t_k(i,j,l)]}} \right. \\
&\qquad \left. \times \frac{\prod_{w=1}^{M} m_{(i,j,l|v)}(b_w)^{[t_k(i,j,l)(v,w,l+1)]}}{\left(\sum_{c=0}^{M} m_{(i,j,l|v)}(b_c)\right)^{[t_k(i,j,l)(v,\cdot,l+1)]}} \right] \qquad (4.9)
\end{aligned}$$

where $a^{[n]} = a(a+1)\cdots(a+n-1)$ and $a^{[0]} = 1$. Furthermore every inadmissible sequence has zero probability.

The BUP is defined inductively through the system of predictive distributions (4.5) and in the previous proposition we obtained the system of finite dimensional distributions (4.9). Our object is now to ensure that there exists as stochastic process on some probability space having (4.9) as finite dimensional law.

Lemma 8. *The BUP exists. That is, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process $X = \{X_n\}$ define on the same space such that the finite dimensional distribution of X is give by (4.9).*

The remaining part to this section will be focused on representing the process X as a mixture of Markov chains, provided X is recurrent. That is we want to represent the finite dimensional probability of X as

$$\mathbb{P}(\cap_{k=0}^{n} \{X_k = (i_k, j_k, s_k)\}) = \mathbb{E}_{\mu, i_0}\left[\prod_{k=0}^{n-1} \Pi_{(i_k, j_k, s_k), (i_{k+1}, j_{k+1}, s_{k+1})}\right] \qquad (4.10)$$

for some random transition matrix Π is $S \times S$. Where Π is distributed according to μ, for some probability measure on the space of stochastic matrices on $S \times S$. We will proceed as usual and use the theory of partial exchangeable processes [63, 44]. Hence recall that two strings $\sigma = (s_0, s_1, \cdots, s_n)$ and $\mu = (v_0, v_1, \cdots, v_n)$ of elements in S are called equivalent if (i) $s_0 = v_0$ and (ii) $t_{n,\sigma}(s, s^*) = t_{n,\mu}(s, s^*)$ for all $s, s^* \in S$ where $t_{n,\sigma}(s, s^*)$ denotes the

number of transitions from s to s^* within σ. An important property for representing a process as a mixture of markov chains is partial exchangeability. A process is partial exchangeable if, under the probability law of the process, any two equivalent sequences are equal likely. Any recurrent partial exchangeable process can be represented as a mixture of Markov chains. We note the following property of the BUP which will be of frequency use in the sequel.

Lemma 9. *The BUP $\{X_n\}$ is partial exchangeable.*

All that remains is to ensure the recurrence of the bivariate urn process. Recurrence will be the last part to establish a mixture representation. We proceed as usual through 'regenerative' cycles. Define the sequence of hitting times $\{\zeta_n\}_n$ to the initial state by $\zeta_0 = 0$ and

$$\zeta_{n+1} = \inf\{k > \zeta_n : X_k = (0,0,0)\}, \tag{4.11}$$

which denote all time points where the process terminates the excursion from the box $G_1 \times G_1 \times \mathbb{N}$ back to the bottom (see figure 4.2). The process $\{X_n\}$ is recurrent if

$$\mathbb{P}(X_n = (0,0,0) \text{ for infinite many } n) = \mathbb{P}(\cap_{j=1}^{\infty}\{\zeta_k < \infty\}) = 1. \tag{4.12}$$

We will assume the following condition.

Assumption 2. *There exist an $\epsilon \in (0,1)$ and a $t_\epsilon \in \mathbb{N}$ such that*

$$\frac{n_{(i,j,t)}(a_0)}{\sum_{c=0}^{N} n_{(i,j,t)}(a_c)} > \epsilon \quad \forall\ t > t_\epsilon,\ (i,j,t) \in S. \tag{4.13}$$

The assumption is reasonable for our particular purpose. Assumption 2 states that a priory, for some $\epsilon > 0$, there exists some coordinate t_ϵ such that for every state in the space $G_1 \times G_2 \times \{t\}, t \geq t_\epsilon$ there is at least an ϵ-risk to return to the initial state $\{(0,0,0)\}$. In section 4.3 we will construct an exchangeable sequence of stochastic processes on $G_1 \times G_2$ for problems as described in example 1 and 2 of section 4.2.1 where $\tau_{n+1} = \zeta_{n+1} - \zeta_n - 1$

corresponds to the survival times of the n^{th} observed process. In such cases the condition merely states that at least at some point t_ϵ there has to be an arbitrary small but positive risk for a failure under any combination of covariate state (i, j) in $G_1 \times G_2$.

Lemma 10. *Suppose that the initial urn compositions of the BUP satisfies assumption 1 and 2, then $\{X_n\}$ is recurrent.*

As discussed before, by [44], a recurrent partial exchangeable process can be represented as a Mixture of Markov chains. Hence from lemma 4 and the last lemma we may state the following corollary.

Corollary 3. *Suppose that the initial urn compositions of the bivariate urn process satisfy assumption 1 and 2, then $\{X_n\}$ is a mixture of Markov chains with unique mixing measure.*

It only remains to determining the mixing measure in the representation (4.10). Clearly, since we constructed the process through draws from a system of Pólya urns, the mixing measure must be related to Dirichlet distributions. And we determine the mixing measure as usual through the weak limit of the empirical transition probabilities, whenever the limit exists. Hence we show the mixing measure only for rows of recurrent states.

For this purpose define for every state (i, j, t) in S the set of potential successor states to (i, j, t), $R_{(i,j,t)} = R^1_{(i,j,t)} \cup R^2_{(i,j,t)}$, where

$$R^2_{(i,j,t)}\{(v, w, t + 1) \in S : n_{(i,j,t)}(a_v)m_{(i,j,t|v)}(b_w) > 0\}$$

and $R^1_{(i,j,t)} = \{(0, 0, 0)\}$ if $n_{(i,j,t)}(a_0) > 0$ and $R^1_{(i,j,t)} = \emptyset$ otherwise. The set $R_{(i,j,t)}$ contains all states in S that are attainable from (i, j, t) with positive probability within one transition step. Now let $S^* = \bigcup_{n=0}^{\infty} S^*_n$ where $S^*_0 = \{(0, 0, 0)\}$ and $S^*_{n+1} = \bigcup_{(i,j,t)\in S^*_n} R_{(i,j,t)}$ for $n \geq 1$. Then S^* denote all states in S which are attainable within a finite number of transition steps with positive probability.

Notice that for every $s \in S$ the cardinality of R_s is at most $(N \times M) + 1$. Hence every row of Π will have at most $(N \times M) + 1$ positive elements $a.s.$.

The following proposition determines the probability law of the rows of the random transition matrix Π for every state s in S^*.

Proposition 2. *Suppose that the initial urn compositions satisfies assumption 1 and 2. Then for every state (i,j,t) in S^* the random transition row $\{\Pi_{(i,j,t),(v,w,s)}\}_{(v,w,s)}$, is given by*

$$\Pi_{(i,j,t),(v,w,s)} = \begin{cases} 1 - \sum_{v=1}^{N} p(i,j,t)_v & \text{if } (v,w,s) = (0,0,0) \\ p(i,j,t)_v p(i,j,t|v)_w & \text{if } (v,w,s) \in G_1 \times G_2 \times \{t+1\} \\ 0 & \text{otherwise,} \end{cases}$$

where for $v \in G_1$

$$p(i,j,t) \sim \textit{Dirichlet}\big(n_{(i,j,t)}(a_0), n_{(i,j,t)}(a_1), \cdots n_{(i,j,t)}(a_N)\big) \textit{ and}$$
$$p(i,j,t|v) \sim \textit{Dirichlet}\Big(m_{(i,j,t|v)}(b_1), \cdots m_{(i,j,t|v)}(b_M)\Big).$$

Furthermore, the vectors $\{p(i,j,t), p(i,j,t|v)\}_{v \in G_2}$ are mutually independent. Moreover the rows of Π are mutually independent.

As mentioned before every row of the random transition matrix Π under which $\{X_n\}$ is a mixture of Markov chains has at most $(N \times M) + 1$ positive elements *a.s.*. Proposition 2 leaves the probability law of all rows of Π for states $s \in S \setminus S^*$ undefined. If $s \in S \setminus S^*$, there does not exist any finite admissible sequence of states in S that lead to s with positive probability, that is

$$P(X_n \in S \setminus S^* \ \exists\ n \geq 0) = \mathbb{E}[P(X_n \in S \setminus S^* \ \exists\ n \geq 0 | \Pi, X_0))] = 0 \quad (4.14)$$

Hence $P(X_n \in S \setminus S^* \ \exists\ n \geq 0 | \Pi, X_0) = 0$ with probability one and any modification of the probability law of rows of Π for states in $S \setminus S^*$ leaves the properties of the process unchanged for the fixed $X_0 = (0,0,0)$. To ensure that every rows of Π sums to one, we make the convention that for $s \in S \setminus S^*$

$$\Pi_{s,s'} = \delta_{(0,0,0)}(s'). \quad (4.15)$$

Now Proposition 2 and the last convention shows that Π is a block matrix given by

$$\Pi \overset{a.s.}{=} \begin{pmatrix} \Pi_0(0) & \Pi_{-0}(0) & \mathbf{O} & \mathbf{O} & \cdots \\ \Pi_0(1) & \mathbf{O} & \Pi_{-0}(1) & \mathbf{O} & \cdots \\ \Pi_0(2) & \mathbf{O} & \mathbf{O} & \Pi_{-0}(2) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{4.16}$$

where, for $t = 0$, $\Pi_0(0) = p(0, 0, 0)$,

$$\Pi_{-0}(0) = \{\Pi_{(0,0,0),(v,w,1)}\}_{(v,w,1)\in S} \tag{4.17}$$

and, for $t \geq 1$, $\Pi_0(t) = \{\Pi_{(i,j,t)(0,0,0)}\}_{(i,j,t)\in S}$ and

$$\Pi_{-0}(t) = \{\Pi_{(i,j,t),(v,w,t+1)}\}_{(i,j,t),(v,w,t+1)\in S} \tag{4.18}$$

Notice that, by proposition 2, the sequence $\{(\Pi_0(t), \Pi_{-0}(t))\}_{t\geq 0}$ are mutually independent random stochastic matrices.

4.2.3 Properties given a sample of observations

We shortly discuss some properties of the process. In particular let us assume that $\{X_k\}_{k\leq n}$ denotes an observed sampled path of the BUP until time n. From the sampling scheme 1 and the definition through the predictive distributions 4.5, it is clear that the process does not restart from X_n, completely regenerative like Markov chains. The process will restart at X_n and behave like the original urn process but with initial urn compositions updated according to (4.6), (4.7) and (4.8) for the past $\{X_k\}_{k\leq n}$. That is, for all $(i, j, t) \in S$ and all $v \in G_1$ the "new" initial urn compositions will be

$$n^*_{(i,j,t)}(a_0) = n_{(i,j,t)}(a_0) + t_n(i, j, t)(0, 0, 0) \tag{4.19}$$

$$n^*_{(i,j,t)}(a_v) = n_{(i,j,t)}(a_v) + t_n(i, j, t)(v, \cdot, t + 1) \tag{4.20}$$

$$m^*_{(i,j,t|v)}(b_w) = m_{(i,j,t|v)}(b_w) + t_n(i, j, t)(v, w, t + 1) \tag{4.21}$$

for $w \in G_2$. Similarly, if the process is recurrent and we observe a trajectory until ζ_n, $\{X_k\}_{k\leq\zeta_k}$, then the future path $\{X_{\zeta_n+k}\}_{k\geq 0}$ will be again a bivariate urn process which starts at $(0,0,0)$. The new urn compositions are updated according the history $\{X_k\}_{k\leq\zeta_k}$ as before (the counting processes $t_n(\cdot)$ above are replaced by $t_{\zeta_n}(\cdot)$). The statement can be made formal by first conditioning on $\{\zeta_m = n\}$ and using the case above. Then we may sum over all events $\{\zeta_m = n\}$ and use the fact that $P(\zeta_m < \infty) = 1$. By the previous reasoning we can see that, conditional on the past, the future of a recurrent bivariate urn process will be again a mixture of Markov chains. This fact is stated in the following

Corollary 4. *Suppose that $\{X_n\}$ satisfies assumptions 1 and 2. Let $\{X_k\}_{0\leq k\leq v_n}$ be a sample until $v_n = n$ (or $v_n = \zeta_n$). Then the future $\{X_{v_n+k}\}_{k\geq 0}$ is again a mixture of Markov chains with random transition matrix $\Pi^{(v_n)}$. The rows of $\Pi^{(v_n)}$ are independent, where for $(i,j,t) \in S^*$*

$$\Pi^{(v_n)}_{(i,j,t),(v,w,s)} = \begin{cases} 1 - \sum_{v=1}^{N} p^{(v_n)}(i,j,t)_v & \text{if } (v,w,s) = (0,0,0) \\ p^{(v_n)}(i,j,t)_v p^{(v_n)}(i,j,t|v)_w & \text{if } (v,w,s) \in G_1 \times G_2 \times \{0\} \\ 0 & \text{otherwise} \end{cases}$$

where for $v \in G_1$

$$p^{(m)}(i,j,t) \sim \textit{Dirichlet}\big(n^*_{(i,j,t)}(a_0), \cdots, n^*_{(i,j,t)}(a_{N_1})\big) \tag{4.22}$$

$$p^{(m)}(i,j,t|v) \sim \textit{Dirichlet}\Big(m^*_{(i,j,t|v)}(b_1), \cdots, m^*_{(i,j,t|v)}(b_{N_2})\Big) \tag{4.23}$$

and the random vectors $\{p^{(v_n)}(i,j,t), p^{(v_n)}(i,j,t|v)\}_{v\in G_1}$ are mutually independent.

We remark that for $v_n = n$ (or $v_n = \zeta_n$) the random transitions matrix $\Pi^{(v_n)}$ conditional on which $\{X_{v_n+k}\}_{k\geq 0}$, given $\{X_k\}_{0\leq k\leq v_n}$, is Markov will have the same form as (4.16), where $\Pi_0(t)$ and $\Pi_{-0}(t)$ are replaced by $\Pi_0^{(v_n)}(t)$ and $\Pi_{-0}^{(v_n)}(t)$ for all $t \geq 0$.

4.3 Sequences of Mixtures of bivariate Markov chains.

In this section, we will use the bivariate urn process to analysis binary multi-state processes which are similar to the ones in example 1 and 2. We will focus on the blocks between successive hits to the initial state $(0,0,0)$. The sequence and the predictive definition for the BUP can be used to predict summary measure for a multi-state process. In the following we assume that $\{X_n\}$ is recurrent and satisfies assumption 1 and 2.

To fix some notation let $\tilde{S}$ be the space of all finite strings of elements of S endowed with the discrete topology. We define on $\tilde{S}$ the sequence of blocks of excursion from $(0,0,0)$. That is, for every $n \geq 1$, the element $\mathbf{B}_n$ represents the sequence of states between ζ_{n-1} and ζ_n

$$B_n := (X_{\zeta_{n-1}}, X_{\zeta_{n-1}+1}, \cdots, X_{\zeta_n-1}) \tag{4.24}$$

where $\tau_n = \zeta_n - \zeta_{n-1} - 1$ denotes the length of the n-th block.

Under recurrence $\{\mathbf{B}_n\}$ is will be a well defined random element on $\tilde{S}$. The next lemma states that the sequence $\{\mathbf{B}_n\}$ is invariant under permutation of coordinated.

Lemma 11. *The sequence $\{\boldsymbol{B}_n\}$ is exchangeable.*

Notice that if L is a measurable map from $\tilde{S}$ into another measurable space, than $\{L(\mathbf{B}_n)\}$ is exchangeable too. Therefore let

$$L : S \to \{(0,0)\} \cup (G_1 \times G_2) \quad L(i,j,t) = (i,j) \tag{4.25}$$

be the projection from S onto the first two coordinates of S and let

$$\begin{aligned} \mathbf{Y}_n :=& (Y_0^{(n)}, Y_1^{(n)}, \cdots, Y_{\tau_n}^{(n)}) = L(\mathbf{B}_n) \\ =& \Big(L(X_{\zeta_{n-1}}), L(X_{\zeta_{n-1}+1}), \cdots, L(X_{\zeta_n-1}) \Big), \end{aligned} \tag{4.26}$$

where $\tau_n = \zeta_n - \zeta_{n-1} - 1$. We removed the last coordinate since this is a deterministic part within any block $\mathbf{B}_n$. It follows that $\{Y_n\}$ is an exchange-

able sequence of random trajectories observed over a random length of times $\{\tau_n\}$. Notice that $\{\tau_n\}$ is exchangeable too. Furthermore for any sequence $\{(i_k, j_k)\}_{k\leq t}$ of states in $G_1 \times G_1$ such that $(i_0, j_0) = (0, 0)$ and any $n \geq 1$

$$\begin{aligned}
&\mathbb{P}\big[\bigcap_{k=0}^{t}\{Y_k^{(n)} = (i_k, j_k)\}, \tau_n = t|\Pi\big] \\
=&\mathbb{P}\big[\bigcap_{k=0}^{t}\{X_{\zeta_{n-1}+k}^{(n)} = (i_k, j_k, k)\}, X_{\zeta_{n-1}+t+1} = (0, 0, 0), \zeta_n - \zeta_{n-1} - 1 = t|\Pi\big] \\
=&\Pi_{-0}(k)_{(0,0),(i_1,j_1)}\Big(\prod_{k=1}^{t-1}\Pi_{-0}(k)_{(i_k,j_k),(i_{k+1},j_{k+1})}\Big)\Pi_0(t)_{(i_t,j_t),(0,0)}
\end{aligned}$$

where the elements Π_0 and Π_{-0} are given in (4.16). Hence, given the random matrices $\Pi = \{\Pi(t)\}_{t\geq 0}$, the sequence $\{\mathbf{Y}_n\}$ are iid replica of a time inhomogeneous Markov chain on $G_1 \times G_2$ with termination probability Π_0, transition sub-matrix Π_{-0} and initial distribution $\delta_{(0,0)}$. Similar to the former reasoning and from corollary 7 it should be clear that for a fixed $n \geq 1$ and given $\{\mathbf{Y}_k\}_{1\leq k\leq n}$ the process $\mathbf{Y}_{n+1}$ is a mixture of time-inhomogeneous Markov chains on $G_1 \times G_2$. This follows from

$$\begin{aligned}
&\mathbb{P}[\cap_{k=t}^{k}\{Y_k^{(n+1)} = (i_k, j_k)\}|\{\mathbf{Y}_k\}_{1\leq k\leq n}] \\
&\quad = \mathbb{P}[\cap_{k=0}^{t}\{X_{\zeta_n+k} = (i_k, j_k, k)\}|\{X_l\}_{0\leq l\leq \zeta_n}] \\
&\quad = \mathbb{E}\Big[\Pi_{-0}^{(\zeta_n)}(0)_{(0,0)(i_1,j_1)}\prod_{k=1}^{t-1}\Pi_{-0}^{(\zeta_n)}(k)_{(i_k,j_k)(i_{l+1},j_{l+1})}\Big].
\end{aligned}$$

The mixing measure is given by corollary 4 and the counting process can be rewritten as

$$\begin{aligned}
t_n[k;(i,j)(i,j)] &:= \sum_{l=1}^{n} I(Y_k^{(l)} = (i,j), Y_{k+1}^{(l)} = (v,w), \tau_l > k) = t_{\zeta_n}(i,j,k)(i,j,k+1) \\
t_n[k;(i,j)(0,0)] &:= \sum_{l=1}^{n} I(Y_k^{(l)} = (i,j), \tau_l = k) = t_{\zeta_n}(i,j,k)(0,0,0)
\end{aligned}$$

Remark 13. *Let* $\mathbb{P}_\Pi(\cdot) := \mathbb{P}(\cdot|\Pi)$ *denote the random transition probability*

induced by the random transition matrix Π *conditional on which elements* $\{\boldsymbol{Y}_n\}_n$ *are iid replica of a Markov chain. Then for any* $\boldsymbol{Y}_n = \{Y_t\}_{0\le t<\tau} = \{(Y_t^{(1)}, Y_t^{(2)})\}_{0\le t<\tau}$ *(where for simplicity we suppress the index* n*) the previous reasoning and 2 shows that*

$$\begin{aligned}&\mathbb{P}_\Pi\Big[(Y_{t+1}^{(1)}, Y_{t+1}^{(2)}) = (v,w)\Big|(Y_t^{(1)}, Y_t^{(2)}) = (i,j)\Big]\\ &=\mathbb{P}_\Pi\Big[Y_{t+1}^{(1)} = v\Big|(Y_t^{(1)}, Y_t^{(2)}) = (i,j)\Big]\mathbb{P}_\Pi\Big[Y_{t+1}^{(2)} = w\Big|(Y_t^{(1)}, Y_t^{(2)}) = (i,j), Y_{t+1}^{(1)} = v\Big]\end{aligned}$$

where

$$p(i,j,t)_v := \mathbb{P}_\Pi\Big[Y_{t+1}^{(1)} = v\Big|(Y_t^{(1)}, Y_t^{(2)}) = (i,j)\Big] \text{ and} \tag{4.27}$$

$$p(i,j,t|v)_w := \mathbb{P}_\Pi\Big[Y_{t+1}^{(2)} = w\Big|(Y_t^{(1)}, Y_t^{(2)}) = (i,j), Y_{t+1}^{(1)} = v\Big]. \tag{4.28}$$

Which means that under $\mathbb{P}_\Pi$ *(and* Π *respectively) the joined random transition probability of* $\{(Y_t^{(1)}, Y_t^{(2)})\}$ *is specified as the product of the transition probability of* $Y_{t+1}^{(1)}$ *given* Y_t *times the conditional transition probability of* $Y_{t+1}^{(2)}$ *given* $Y_t, Y_{t+1}^{(1)}$*. Which means that the two sequences* $\{Y_t^{(1)}\}$ *and* $\{Y_t^{(2)}\}$ *are clearly not independent. The way of expression the joined transition probability of* $\{Y_t^{(1)}\}$ *and* $\{Y_t^{(2)}\}$ *as a product of the marginal times the conditional transition probability is probably the most elementary way of expression a prior on a random bivariate transition probability.*

Remark 14. *The sequence* $\{Y_n\}_{n\ge1}$ *is a useful function for prediction of bivariate multi-state processes with Markov dependence. Typical in such situations one may observe longitudinal data of* n *observation* $\{\boldsymbol{Y}_k\}_{1\le k\le n} = \{(\boldsymbol{Y}_k^{(1)}, \boldsymbol{Y}_k^{(2)})\}_{1\le k\le n}$ *which are conditional independent given an unknown transition array* Π*. The object of interest might be to estimate functions of the unknown transition array and to predict longitudinal trajectories. The sequence* $\{\boldsymbol{Y}_n\}_{n\ge1}$ *derived from the BUP might be of some use for such problems.*

Remark 15. *We want to point out that modelling an observed phenomena, as described in the previous remark, with a random element* $\boldsymbol{Y}_n$ *does not lead to causal dependence assessment how one dimension affects the other. In principle we can predict marginal and conditional probabilities of one of the*

two dimensions give the other dimension but this is no assessment of causality. It is only an assessment of how likely the probability of one dimension is given the other dimension after observing both dimension jointedly for several trajectories $\boldsymbol{Y}_{n\leq k}$.

In the next section we will use the sequence $\{\mathbf{Y}_n\}_{n\geq 1}$ and their representation as an exchangeable sequence of mixtures of Markov chains for Bayesian prediction for time-inhomogeneous bivariate Markov chains.

4.3.1 Inference for inhomogeneous bivariate Markov chains

In this section we discuss prediction for bivariate Markov chains based in the BUP. We will use the sequence $\{\mathbf{Y}_n\}_{n\geq 1}$ defined in the previous section. To fix some ideas, suppose information about the underlying phenomenon are obtained by observing bivariate trajectories over a random length of time for K exchangeable particle where $\mathbf{Y}_k = \left(Y_l^{(k)} = (i_l^{(k)}, j_l^{(k)}); 0 \leq l \leq \tau_k\right)$ represents an observed trajectory of particle $1 \leq k \leq K$. Typically $Y_l^{(k)} \in G_1 \times G_2$ for all $l = 1, \cdots, \tau_k$ and all $k = 1, \cdots, K$ where $G_1 \times G_2$ is some bounded subset of $\mathbb{N}^2$. We may assume that a fixed trajectory has a time inhomogeneous Markov dependence structure conditional on an unknown transition array (a system of unknown time dependent transition matrices). Furthermore the particles are not necessarily independent but rather exchangeable, which implies independents given the unknown transition array.

The object of interest might be to estimate the unknown transition array and functions of the process like the k-step predictive probability, occupation probabilities or the expected state of the process at a given point in time. Furthermore in some case the estimation of the survival time τ or the survival function might be of some interest as well.

We model the trajectories jointly through the blocks $\{\mathbf{Y}_k\}_{1\leq k\leq K}$ defined in section 4.3, equation (4.26) for a BUP on $S = \{0, 0, 0\} \cup \left(G_1 \times G_2 \times \mathbb{N}\right)$. We assume that the initial urn compositions $\{U^{(1)}(s), U^{(2)}(s|v); s \in S, v \in G_1\}$ satisfy assumption 1 and 2.

Furthermore we will assume that the termination of every observed random trajectory $\mathbf{Y}_k$ is a random variable itself, but this assumption is generally not necessary.

For every $t \geq 0$ denote with $\hat{\Pi}_{-0}^{(M)}(t) = \mathbb{E}[\Pi_{-0}(t)|\{\mathbf{Y}_k\}_{1\leq k\leq K}]$ and $\hat{\Pi}_{0}^{(M)}(t) = \mathbb{E}[\Pi_{0}(t)|\{\mathbf{Y}_k\}_{1\leq k\leq K}]$, where for $(i,j),(v,w) \in G_1 \times G_2$ and for $t \geq 1$

$$\hat{\Pi}_{-0}^{(M)}(t)_{(i,j),(v,w)} \tag{4.29}$$

$$= \frac{n_{(i,j,t)}(a_v) + t_K[t;(i,j)(v,\cdot)]}{\sum_{c=0}^{N} n_{(i,j,t)}(a_c) + t_K[t;(i,j)]} \frac{m_{(i,j,t|v)}(b_w) + t_K[t;(i,j)(v,w)]}{\sum_{c=1}^{M} m_{(i,j,t|v)}(b_c) + t_K[t;(i,j)(v,\cdot)]} \tag{4.30}$$

and for $t = 0$

$$\hat{\Pi}_{0}^{(M)}(t)_{(v,w),(0,0)} = \frac{n_{(i,j,s)}(a_0) + t_K[s;(i,j)(0,0)]}{\sum_{c=0}^{N} n_{(i,j,t)}(a_c) + t_K[s;(i,j)]}. \tag{4.31}$$

Note that by exchangeability of $\{\mathbf{Y}_k\}$ a predictive estimator of an event of $\mathbf{Y}_k$ will be equal to a Bayes estimator with respect to a square loss function for a prior as states in proposition 2.

Lemma 12. *Let (i,j) and (v,w) be states in $G_1 \times G_2$. (i) The predictive transition matrices at time s is given by $\hat{\Pi}(t) = [\hat{\Pi}_{0}^{(M)}(t), \hat{\Pi}_{0}^{(M)}(t)]$*
(ii) The predictive marginal transition probabilities at time s is given by

$$\hat{P}[Y_{s+1}^{(1)} = v|Y_s = (i,j)] = \frac{n_{(i,j,s)}(a_v) + t_K[s;(i,j)(v,\cdot)]}{\sum_{c=0}^{N} n_{(i,j,t)}(a_c) + t_K[s;(i,j)]} \tag{4.32}$$

$$\hat{P}[Y_{s+1}^{(2)} = w|Y_s = (i,j)] = \frac{1}{\sum_{c=0}^{N} n_{(i,j,s)}(a_c) + t_K[s;(i,j)]}$$
$$\times \sum_{v=1}^{N} \frac{[n_{(i,j,s)}(a_v) + t_K[s;(i,j)(v,\cdot)]\ ][m_{(i,j,s|v)}(b_w) + t_K[s;(i,j)(v,w)]\ ]}{\sum_{c=1}^{M} m_{(i,j,s|v)}(b_c) + t_K[s;(i,j)(v,\cdot)]}. \tag{4.33}$$

(iii) The predictive s-step transition probability is given by

$$\hat{P}[Y_s = (v,w)] = \left(\hat{\Pi}_{-0}^{(M)}(0)\prod_{j=1}^{s-1}\hat{\Pi}_{-0}^{(M)}(j)\right)_{(v,w)} \tag{4.34}$$

$$\hat{P}[Y_s^{(1)} = v)] = \sum_{w=1}^{M}\left(\hat{\Pi}_{-0}^{(M)}(0)\prod_{j=1}^{s-1}\hat{\Pi}_{-0}^{(M)}(j)\right)_{(v,w)} \tag{4.35}$$

$$\hat{P}[Y_s^{(2)} = w)] = \sum_{v=1}^{N}\left(\hat{\Pi}_{-0}^{(M)}(0)\prod_{j=1}^{s-1}\hat{\Pi}_{-0}^{(M)}(j)\right)_{(v,w)}. \tag{4.36}$$

Next to the transition probabilities we may also predict some functions of the process. For example the occupation time in state $(i,j) \in G$ given by $T(\mathbf{Y}_k)_{(i,j)} = \sum_s I(Y_s^{(k)} = (i,j))$ might be of some interest. Similarly define $T(\mathbf{Y}_k)_{(i,\cdot)} = \sum_s I(Y_s^{(k)} \in \{i\} \times G_2)$ and $T(\mathbf{Y}_k)_{(\cdot,j)} = \sum_s I(Y_s^{(k)} \in G_1 \times \{j\})$ to be the time a process spend marginal in a state of G_v, $v = 1,2$ during the lifetime.

Lemma 13. *The predicted mean occupation time for a given state $(v,w) \in G_1 \times G_2$ is given by given by*

$$\hat{\mathbb{E}}[T(Y)_{(v,w)}] = \sum_{s\geq 1}\left(\hat{\Pi}_{-0}^{(K)}(0)\prod_{j=1}^{s-1}\hat{\Pi}_{-0}^{(K)}(j)\right)_{(v,w)}, \tag{4.37}$$

$$\hat{\mathbb{E}}[T(Y)_{(v,\cdot)}] = \sum_{s\geq 1}\sum_{j=1}^{M}\left(\hat{\Pi}_{-0}^{(K)}(0)\prod_{j=1}^{s-1}\hat{\Pi}_{-0}^{(K)}(j)\right)_{(v,w)} \tag{4.38}$$

$$\hat{\mathbb{E}}[T(Y)_{(\cdot,w)}] = \sum_{s\geq 1}\sum_{i=1}^{N}\left(\hat{\Pi}_{-0}^{(K)}(0)\prod_{j=1}^{s-1}\hat{\Pi}_{-0}^{(K)}(j)\right)_{(v,w)}. \tag{4.39}$$

(ii) The predicted conditional and marginal survival function and the prediced

survival time are given by

$$\hat{P}(\tau > s) = \sum_{v=1,w=1}^{N,M} \left(\hat{\Pi}_{-0}^{(K)}(0) \prod_{j=1}^{s} \hat{\Pi}_{-0}^{(K)}(j)\right)_{(v,w)} \tag{4.40}$$

$$\hat{P}[\tau > s | Y_k = (v,w)] = 1 - \frac{n_{(v,w,s)}(a_0) + t_K[s;(v,w)(0,0)]}{\sum_{c=0}^{N} n_{(v,w,s)}(a_c) + t_K[s;(v,w)]} \tag{4.41}$$

$$\hat{E}[\tau] = \sum_{s\geq 0} \sum_{(v,w)\in G_1\times G_2} \left(\hat{\Pi}_{-0}^{(K)}(0) \prod_{l=1}^{s} \hat{\Pi}_{-0}^{(K)}(l)\right)_{(v,w)}. \tag{4.42}$$

Finally we give an expression for the predicted covariance between the first and the second component of the process at a given time point. We stress two facts. First the expression if not easy to interpret and secondly the covariance as a measure of dependence if usually not very useful for categorical data.

Lemma 14. *The predicted covariance between the two joined processes* $Y^{(1)}$ *and* $Y^{(2)}$ *at time* $(s+1)$ *given* $Y_s = (i,j)$ *is given by*

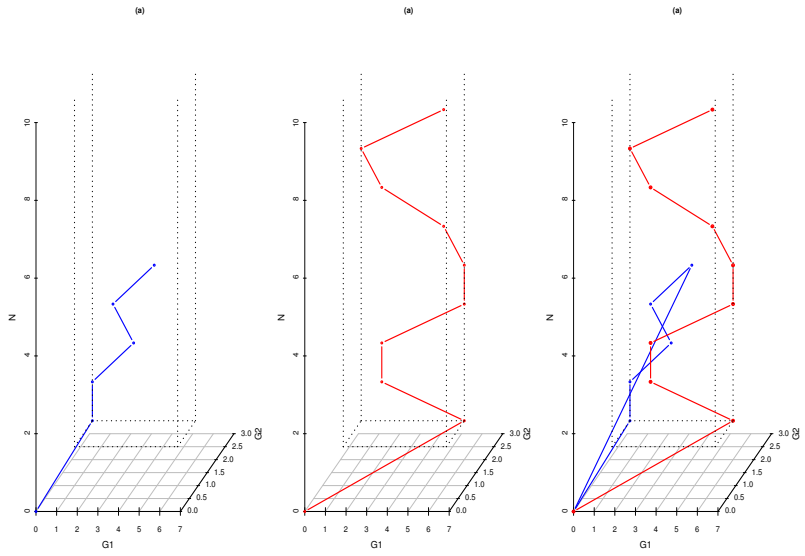
$$\begin{aligned}
\hat{Cov}[Y_{s+1}^{(1)}, Y_{s+1}^{(2)} | Y_s = (i,j)] &= \frac{1}{\sum_{c=0}^{N} n_{(i,j,s)}(a_c) + t_K[s;(i,j)]} \\
&\left[\sum_{v=1,w=1}^{N,M} \frac{(v\cdot w)\cdot [n_{(i,j,s)}(a_v) + t_K[s;(i,j)(v,w)]] \cdot [m_{(i,j,s|v)}(b_w) + t_K[s,(i,j)(v,w)]]}{\sum_{c=1}^{M} m_{(i,j,s|v)}(b_c) + t_K[s;(i,j)]} \right. \\
&\quad - \left(\frac{\sum_{v=1}^{N} v\cdot [n_{(i,j,s)}(a_v) + t_K[s;(i,j)(v,\cdot)]\ ]}{\sum_{c=0}^{N} n_{(i,j,s)}(a_c) + t_K[s;(i,j)]} \right. \\
&\left. \left. \sum_{w=1}^{M} w \sum_{v=1}^{N} \frac{[n_{(i,j,s)}(a_v) + t_K[s;(i,j)(v,\cdot)]\ ] \cdot [m_{(i,j,s|v)}(b_w) + t_K[s;(i,j)(v,w)]}{\sum_{c=1}^{M} m_{(i,j,s|v)}(b_c) + t_K[s;(i,j)(v,\cdot)]} \right) \right]
\end{aligned}$$

The last three lemmas gave Bayes estimator for some summary measures of bivariate inhomogeneous Markov chains which may or may not be use in practice. Other characteristic can be estimated through the predictive rule of the process in a similar manner.

4.3.2 Example and Simulation

In this section we numerical illustrate prediction for a bivariate multi-state process using a BUP. We simulated data for example 1 section 4.2.1 on the joined analysis of two disease and the survival time. Let D1 be the first disease classified on a scale from one to six (1 indicates absence of D1 and 6 denotes a strong form of D1) and let D2 be a second disease with binary classification (where 2 indicates the presence and 1 the absence of D2). We simulate lifetimes $\mathbf{Y}_k \in S_{\{1,\cdots,6\}\times\{1,2\}}$ for $k = 1, \cdots, K$ patients from a transition array $\Pi_0 = \{\Pi_0(t)\}$ and a maximum lifespan of 11 unites of time for $K = 500, 1000$ and 5000.

Figure 4.3: Simulation of an R-BHUP on S where $|G_1| = 6$ and $|G_2| = 2$. **(a)** shows the trajectories of the first observations; **(b)** shows the trajectories of the second observations; and **(c)** shows the resulting R-(B)HUP until the second hit to $(0, 0, 0)$.



To make the sample $\{\mathbf{Y}_k\}$ exchangeable we let the transition array $\Pi^0 =$

$\{\Pi^0(t)\}$ be random, sampled from the probability law described in proposition 2. Where at time $t \geq 1$ the (i,j)-th row of the random matrix $\Pi^0(t)$ has $E[\Pi_0(t)_{(i,j),(v,w)}] \propto n_{(i,j,t)} m_{(i,j,t|v)}$ for

$$n_{(i,j,t)}(a_v) = \begin{cases} c_1 & \text{if } 1 \leq v \leq i-1 \\ c_1 + c_2 & \text{if } v = 0 \\ c_1 + c_2 + (1 - .5I_{\{j=2\}}) \frac{(c_3(t)-c_2)}{(N-i+1)}(v-i) & \text{if } i \leq v \leq 6 \end{cases}$$

$$m_{(i,j,t|v)}(b_w) = \begin{cases} b4\Big(1 - \frac{b_1 i + b_2 t}{b_1 i + b_2 t + b_3}\Big) & \text{if } j = 1, w = 1 \\ b4\Big(\frac{b_1 i + b_2 t}{b_1 i + b_2 t + b_3}\Big) & \text{if } j = 1, w = 2 \\ 0 & \text{if } j = 2, w = 1 \\ b4 & \text{if } j = 2, w = 2 \end{cases}$$

where $c_3(t) = c_3 + (c_2 - c_3)(t-1)/10$ with $b_l > 0$ for $l = 1, \cdots, 4$ and $c_2 > c_3$, $c_l > 0$ for $l = 1, 2, 3$. For all states (i,j) the expected probability of death between t and t+1 is always proportional to $c_1 + c_2$. For states of D1, the expected probability of transition to a worse health condition is always higher than the expected probability of a transition into a better health condition. This effect is reduced with time and by the presence of the second disease. For D2 there is zero probability to recover from D2. Figure 4.3 (a) and (b) show two arbitrary lifetime trajectories with a survival time of 6 and 10 units.

Furthermore we define a BUP with $G_1 = \{1, 2, \cdots, 6\}$ and $G_2 = \{1, 2\}$, to estimates the 'unknown' transition array $\Pi^0 = \{\Pi^0(t)\}$, the k-step transition probabilities and the survival function via the estimator derived in section 4.3.1. We set the initial urn compositions non-informative, i.e. for $i, v \in \{0, \cdots, 6\}$ and $w \in \{1, 2\}$ we set $n_{(i,j,t)}(a_v) \equiv u_1$, $m_{(i,1,t|v)}(b_w) = m_{(i,2,t|v)}(b_2) \equiv u_2$ and $m_{(i,2,t|v)}(b_1) \equiv 0$ where u_1 and u_2 are positive constants. This means no incorporation of any prior information. The initial urn compositions must not be confused with the above base measures used to sample Π^0, Π^0 is unknown in practice. The observed sequence $\{\mathbf{Y}_k, 1 \leq k \leq K\}$ sampled from Π^0 is assumed to be the first K blocks of the BUP defined in (4.26). Figure 4.3(c) shows the resulting BUP until ζ_2, the third hit to

the initial state. Notice that by exchangeability, the joined probability law of the process is invariant of the order of observing the first K lifetimes and therefore invariant for the prediction of the $(K+1)^{th}$ lifetime as well.

We computed the predictive s-step ahead transition probabilities for 500, 1000 and 5000 observations and compare the estimates with the true value in Figure 4.4 where (a) shows the 1-step, (b) the 5-step and (c) the 10-step ahead transition probabilities for all states $(i,j) \in \{1,\cdots 6\} \times \{1,2\}$. We choose non-informative initial urn compositions, therefore expectedly we observe that the estimates improve with increasing sample size and are close the true s-step ahead transition probabilities for a sample size of 5000 observations.

Furthermore we computed the survival function for a patients according to (4.40) (see figure 4.5). The predicted survival functions are already quite close to the true function for a sample size of 500 and are practically identical to the true one for 5000 trajectories. Finally table 4.1 shows the true and predicted average time spend in a given state. Also in this case the predicted values come close to the true one as the sample size increases.

Table 4.1: Predicted expected time spend in a given level of disease D1 and of D2 within a lifetime

$E[T(Y)_{(i,\cdot)}]$	1	2	3	4	5	6	$E[T(Y)_{(\cdot,j)}]$	1	2
True	0.84	0.53	0.73	1.05	0.66	0.97	True	1.35	3.42
K=500	0.89	0.59	0.75	1.06	0.72	0.96	K=500	1.23	3.74
K=1000	0.85	0.57	0.76	1.08	0.70	0.92	K=1000	1.25	3.63
K=5000	0.83	0.55	0.75	1.05	0.68	0.97	K=5000	1.32	3.52

Notice that most bivariate multi-state processes will have the same structure as problem just studied. The statistical analysis based on the BUP will have always the same steps. Specify the set G_1 and G_2 and try to figure out a decent prior guess to specify the initial urn composition. If you don't have a concrete prior guess let the initial urn compositions be equal to some constants and estimate the functions of interest as derived above.

Figure 4.4: Estimated k-step ahead predictive probabilities for the bivariate multi-state disease monitoring model where $|G_1| = 6$ and $|G_2| = 2$

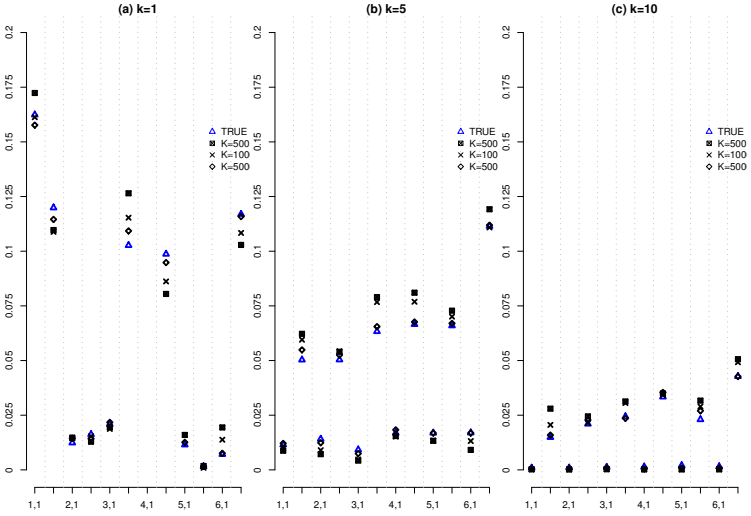
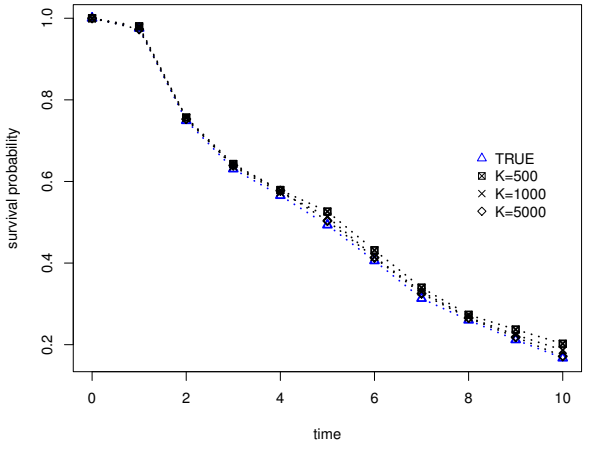


Figure 4.5: Estimated survival function for the bivariate multi-state disease monitoring model where $|G_1| = 6$ and $|G_2| = 2$



4.4 Conclusion

In this paper we introduced a bivariate hierarchical urn process through a sampling scheme from the space of hierarchical Pólya urns $\{U^{(1)}(i,j,t), U^{(2)}(i,j,t|v); v \in G_1\}_{(i,j,t)\in(G_1\times G_2\times\mathbb{N})\cup\{(0,0,0)\}}$. Under regularity conditions, the process can be represented as a mixture of Markov chains on $(G_1 \times G_2 \times \mathbb{N}) \cup \{(0,0,0)\}$. We used the sequence of hitting times to the initial state and the trajectories between successive excursion from the initial state of the process to construct an exchangeable sequence $\{\mathbf{Y}_n\}$. Where for every n the random element $\mathbf{Y}_n = \{Y_k^{(n)}, 0 \leq k \leq \tau_k\} = \{(Y_k^{(n,1)}, Y_k^{(n,1)}), 0 \leq k \leq \tau_n)$ is a mixtures of time inhomogeneous Markov chains starting at a dummy state $(0,0)$ and with values in $G_1 \times G_2$ between $1 \leq k \leq \tau_n$. The constructed sequence of exchangeable processes provides a simple and flexible tool for Bayesian inference for bivariate multi-state models.

Extensions of the current work will mainly focus on the construction of an exchangeable sequence $\{\mathbf{Y}_n\}$ of mixtures of non-homogeneous Markov chains on a general K dimensional state space $S = \times_{i=1}^{K} G_i$ where for $1 \leq i \leq K$ G_i is some subset of $\mathbb{N}$. In this case for any n and given the tail sigma-algebra of $\{\mathbf{Y}_n\}_{n\geq 1}$

$$\mathbf{Y}_n = \{(Y_s^{(n,1)}, \cdots, Y_s^{(n,K)}); 0 \leq s \leq \tau_n\} \overset{iid}{\sim} Markov(\mu, \Pi)$$

where μ is an initial random probability on S and the array $\Pi = \{\Pi(t)\}$ is a sequence of random stochastic matrices on $S \times S$. Such a process may or may not be of some use for for Bayesian analysis multivariate time-inhomogeneous Markov chains.

4.5 Appendix: Proofs

Proof of Proposition 4.9. Let $\big((i_1, j_1, s_1), \cdots, (i_k, j_v, s_k)\big)$ be a finite sequence of elements in S. If the sequence is not admissible the joined probability of the sequence is zero by construction of the process through the system of predictive distributions defined in (4.5). On the other hand suppose the sequence is admissible, then by definition of $\{X_n\}$ through (4.5) we may write

$$\begin{aligned}
&\mathbb{P}[X_0 = (0,0,0), X_1 = (i_1, j_1, s_1), \cdots, X_k = (i_k, j_k, s_k)] \\
&= \prod_{w=0}^{k-1} \left[\frac{n_{(i_w,j_w,s_w)}(a_{i_{w+1}}) + t_w(i_w, j_w, s_w)(i_{w+1}, \cdot, s_w + 1)}{\sum_{c=0}^{N} n_{(i_w,j_w,s_w)}(a_c) + t_w(i_w, j_w, s_w)} \right. \\
&\qquad \left. \times \frac{m_{(i_w,j_w,s_w|i_{w+1})}(b_{j_{w+1}}) + t_w(i_w, j_w, s_w)(i_{w+1}, j_{w+1}, s_w + 1)}{\sum_{c=1}^{M} m_{(i_w,j_w,s_w|i_{w+1})}(b_c) + t_k(i_w, j_w, s_w)(i_{w+1}, \cdot, s_w + 1)} \right]^{I_{\{s_{w+1} \neq 0\}}} \\
&\times \prod_{w=0}^{k-1} \left[\frac{n_{(i_w,j_w,s_w)}(a_0) + t_w(i_w, j_w, s_w)(0,0,0)]}{\sum_{c=0}^{N} n_{(i_w,j_w,s_w)}(a_c) + t_w(i_w, j_w, s_w)} \right]^{I_{\{s_{w+1}=0\}}}
\end{aligned} \tag{4.43}$$

Let $(i, j, t) \in S$ be an arbitrary state among $(i_w, j_w, s_w)_{1 \leq w \leq k}$ and suppose that (i, j, t) has at least one successor state. Denote with $(i^*_w, j^*_w, s^*_w)_{1 \leq w \leq t_k(i,j,t)}$ all successors to (i, j, t). Notice that by admissibility $s^*_w \in \{t + 1, 0\}$ and $s^*_w = 0$ if and only if $(i^*_w, j^*_w, s^*_w) = (0, 0, 0)$. The product component in the joint probability (4.43) associated with (i, j, t) can be rewritten as

$$\begin{aligned}
&\frac{\big[n_{(i,j,t)}(a_0)\big]^{I_{\{s^*_1=0\}}}}{\sum_{c=0}^{N} n_{(i,j,t)}(a_c)} \left[\frac{n_{(i,j,t)}(a_{i^*_1}) \times m_{(i,j,t|i^*_1)}(b_{j^*_1})}{\sum_{c=1}^{M} n_{(i,j,t|i^*_1)}(b_c)} \right]^{I_{\{s^*_1 \neq 0\}}} \\
&\times \frac{\big[n_{(i,j,t)}(a_0) + I_{\{s^*_1=0\}}\big]^{I_{\{s^*_2=0\}}}}{1 + \sum_{c=0}^{N} n_{(i,j,t)}(a_c)} \\
&\times \left[\frac{\big[n_{(i,j,t)}(a_{i^*_2}) + I_{\{i^*_1=i^*_2\}}\big]\big[m_{(i,j,t|i^*_2)}(b_{j^*_2}) + I_{\{i^*_1=i^*_2, j^*_1=j^*_2\}}\big]}{I_{\{i^*_1=i^*_2\}} + \sum_{c=1}^{M} n_{(i,j,t|i^*_2)}(b_c)} \right]^{I_{\{s^*_2 \neq 0\}}} \\
&\times \cdots
\end{aligned} \tag{4.44}$$

$$\times \frac{\left[n_{(i,j,t)}(a_0) + \sum_{w=1}^{v-1} I_{\{s_w^*=0\}}\right]^{I\{s_v^*=0\}}}{v-1+\sum_{c=0}^{N} n_{(i,j,t)}(a_c)}$$
$$\times \left[\frac{\left[n_{(i,j,t)}(a_{i_v^*}) + \sum_{w=1}^{v-1} I_{\{i_w^*=i_v^*\}}\right]\left[m_{(i,j,t|i_v^*)}(b_{j_v^*}) + \sum_{w=1}^{v-1} I_{\{i_w^*=i_v^*, j_w^*=j_v^*\}}\right]}{\sum_{w=1}^{v-1} I_{\{i_w^*=i_v^*\}} + \sum_{c=1}^{M} n_{(i,j,t|i_v^*)}(b_c)}\right]^{I_{\{s_v^*\neq 0\}}}$$

where $v = t_k(i,j,t)$. Therefore the product associated with (i,j,t) reduces to

$$\frac{n_{(i,j,t)}(a_0)^{[t_k(i,j,t)(0,0,0)]}}{\left[\sum_{c=0}^{N} n_{(i,j,t)}(a_c)\right]^{[t_k(i,j,t)]}} \prod_{l=1}^{N}\left[n_{(i,j,t)}(a_l)^{[t_k(i,j,t)(l,\cdot,t+1)]} \times \frac{\prod_{v=1}^{M} m_{(i,j,t|l)}(b_v)^{[t_k(i,j,t)(l,v,t+1)]}}{\left[\sum_{c=1}^{M} m_{(i,j,t|l)}(b_c)\right]^{[t_k(i,j,t)(l,\cdot,t+1)]}}\right].$$

If (i,j,t) has no successor states or is not contained in $\{i_v, j_v, s_v\}_{1\leq v\leq k}$ then $t_k(i,j,t) = 0$ and the produce component associated with (i,j,t) is one. The result now follows by taking the product over all states (i,j,t)in S. □

Proof of lemma 8. We may use the previous proposition to check that the system of finite dimensional distribution defined by (4.9) satisfies Kolmogorov's consistency conditions. But since we define the R-BHUP through the system of predictive distributions (4.5) we may rather use the Ionescu Tulcea extension theorem. Define for every $k \geq 1$, every singleton $s \in S$ and any path $s_{0:k} = \{(0,0,0), (X_j = s_j)\}_{1\leq j\leq k}$ in S^{k+1} the transition kernel $\mu_k(s_{0:k}, s) = P[X_{k+1} = s|\sigma(\{X_i = s_i\}_{i=0,\cdots,k})]$ as defined in (4.5) and for $A \in \mathcal{P}(S)$ let $\mu_k(s_{0:k}, A) = \sum_{s\in A} \mu_k(s_{0:k}, s)$.

By construction for a fixed k and $s_{0:k}$ in S^{k+1} the map $A \to \mu_k(s_{0:k}, A)$ is a probability on $\mathcal{P}(S)$. Furthermore the map $s_{0:k} \to \mu_k(s_{0:k}, A)$ is the ration of two linear functions of the counting processes $\{t_k(s_k)(s')\}_{s\in A}$ defined in (4.6) and hence is easily seen to be measurable. Which implies that the system $\{\mu_k\}_k$ are proper transition probabilities and hence the Ionescu Tulcea extension theorem ensures the existence of a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process $\{X_n\}_{n\geq 0}$ having (4.5) as system of transition probabilities. But by proposition 1 this process has (4.9) as its system of finite dimensional distributions. □

Proof of Lemma 4. Suppose $\sigma = \big((i_k, j_k, s_k)\big)_{0\le k\le n}$ and $\mu = \big((v_k, w_k, t_k)\big)_{0\le k\le n}$ are two equivalent sequences of elements in S. If σ is not admissible then either there exists $(i, j, t) \in S$ and $(v, w, l) \notin \{(0, 0, 0)\} \cup (G_1 \times G_2 \times \{t+1\})$ with $t_{n,\mu}(i, j, t)(v, w, l) = t_{n,\sigma}(i, j, t)(v, w, l) > 0$ or $(i_0, j_0, s_0) = (v_0, w_o, t_0) \neq (0, 0, 0)$. Which, in both cases, implies that μ is inadmissible too. Therefore both strings have zero probability.

On the other hand suppose that σ is admissible (and therefore μ too). By definition, both $t_k(i, j, t)(l, \cdot, s)$ and $t_k(i, j, t)$ are functions of $t_k(i, j, t)(l, v, s)$ and will be equal for both sequences. Since the joined likelihood of the process depends only on $t_k(i, j, t)(l, v, s)$, $t_k(i, j, t)(l, \cdot, s)$ and $t_k(i, j, t)$ for all $(i, j, t), (l, v, s) \in S$ it follows that

$$\mathbb{P}(\cap_{k=0}^{n}\{X_k = (i_k, j_k, s_k)\}) = \mathbb{P}(\cap_{k=0}^{n}\{X_k = (v_k, w_k, t_k)\}).$$

Either way both sequences have the same likelihood. □

Proof of Lemma 29. Without loss of generality assume that $t_\epsilon = 1$. We prove the claim by induction on n. By definition of the process

$$\begin{aligned}
\mathbb{P}(\zeta_1 > t) &= \mathbb{P}(X_0 = (0,0,0), X_j \in G_1 \times G_2 \times \{j\}; 1 \le j \le t) \\
&= \sum_{\substack{i_1,i_2,\cdots,i_t\in G_1 \\ j_1,j_2,\cdots,j_t\in G_2}} \prod_{k=0}^{t-1} \frac{n_{(i_k,j_k,k)}(a_{i_{k+1}})}{\sum_{c=0}^{N} n_{(i_k,j_k,k)}(a_c)} \frac{m_{(i_k,j_k,k|i_{k+1})}(b_{j_{k+1}})}{\sum_{c=1}^{M} m_{(i_k,j_k,k|i_{k+1})}(b_c)} \\
&\le (1-\epsilon)^{t-1} \sum_{\substack{i_1,i_2,\cdots,i_t\in G_1 \\ j_1,j_2,\cdots,j_{t-1}\in G_2}} \left(\frac{n_{(0,0,0)}(a_{i_1})}{\sum_{c=0}^{N} n_{(0,0,0)}(a_c)} \frac{m_{(0,0,0|i_1)}(b_{j_1})}{\sum_{c=1}^{M} m_{(0,0,0|i_1)}(b_c)} \right. \\
&\qquad \left. \times \prod_{m=1}^{t-2} \frac{m_{(i_k,j_k,k|i_{k+1})}(b_{j_{k+1}})}{\sum_{c=1}^{M} m_{(i_k,j_k,k|i_{k+1})}(b_c)} \right) \\
&= (1-\epsilon)^{t-1} \sum_{i_1,i_2,\cdots,i_t\in G_1} \frac{n_{(0,0,0)}(a_{i_1})}{\sum_{c=0}^{N} n_{(0,0,0)}(a_c)} \\
&= N \times (t-1) \times \exp\{-\beta(t-1)\} = o(t) \text{ as } t \to \infty
\end{aligned}$$

where $\beta = -\log(1-\epsilon) \in (0,\infty)$. Therefore we obtain

$$\mathbb{P}(\zeta_1 < \infty) = 1 - \lim_{t\to\infty} \mathbb{P}(\zeta_1 > t) = 0. \tag{4.45}$$

Now suppose for a fixed $n > 1$ that $\mathbb{P}(\cap_{j=1}^{n}\{\zeta_k < \infty\}) = 1$. Set $T = 1 + \max_{1\le k\le n}\{\zeta_k - \zeta_{k-1}\}$. Notice that T is finite with probability one. Let $\mathbb{P}^*(\cdot) = \mathbb{P}(\cdot \mid \cap_{j=1}^{n}\{\zeta_k < \infty\}, T)$ and $\mathbb{P}_v^*(\cdot) = \mathbb{P}(\cdot \mid \cap_{j=1}^{n}\{\zeta_k < \infty\}, T, \tau_n = v)$. Furthermore on the set $\{\zeta_n \le v\}$ for $t > T + v$ all counting processes of transition from (i,j,t) to (v,w,s) (with $(i,j,t),(v,w,s) \in S$) are equal to zero. Hence for $t > T + v$ it follows that (all relations hold in a.s. sense)

$$\begin{aligned}
&\mathbb{P}_v^*(\zeta_{n+1} > t)\\
&= \mathbb{P}_v^*(X_v = (0,0,0), \cap_{j=1}^{t-v}\{X_{v+j} \in G_1 \times G_2 \times \{j\}\})\\
&\le \mathbb{P}_v^*(\cap_{j=T}^{t-v}\{X_{v+j} \in G_1 \times G_2 \times \{j\}\})\\
&= \sum_{\substack{i_0,i_1,\cdots,i_{t-T-v}\in G_1\\ j_0,j_1,\cdots,j_{t-T-v}\in G_2}} \prod_{k=0}^{t-T-v-1} \frac{n_{(i_k,j_k,T+k)}(a_{i_{k+1}})}{\sum_{c=0}^{N} n_{(i_k,j_k,T+k)}(a_c)} \frac{m_{(i_k,j_k,T+k|i_{k+1})}(b_{j_{k+1}})}{\sum_{c=1}^{M} m_{(i_k,j_k,T+k|i_{k+1})}(b_c)}\\
&\le (1-\epsilon)^{t-T-v-1} \sum_{\substack{i_0,i_1,\cdots,i_{t-T-v}\in G_1\\ j_0\in G_2}} \frac{n_{(i_0,j_0,T)}(a_{i_1})}{\sum_{c=0}^{N} n_{(i_0,j_0,T)}(a_c)} \frac{m_{(i_0,j_0,T|i_1)}(b_{j_1})}{\sum_{c=1}^{M} m_{(i_0,j_0,1|i_1)}(b_c)}\\
&\le Nt \times \exp\{-\beta t - T - v)\} = o(t) \text{ as } t \to \infty.
\end{aligned}$$

which holds of every fixed $v \in \mathbb{N}$. Now fix a $\delta > 0$, since by assumption $P(\tau_n < \infty) = 1$, choose a $k > 1$ such that $P(\tau_n \le k) \ge 1 - \delta$. Therefore (all relations hold in a.s. sense)

$$\begin{aligned}
\mathbb{P}^*(\zeta_{n+1} > t) &= \mathbb{E}[\mathbb{P}^*_{\zeta_n}(\zeta_{n+1} > t) I_{(\zeta_n > k)}] + \mathbb{E}[\mathbb{P}^*_{\zeta_n}(\zeta_{n+1} > t) I_{(\zeta_n \le k)}]\\
&\le \delta + N \sum_{v=1}^{k} t \times \exp\{-\beta t - T - v)\}
\end{aligned}$$

Hence, take the limit as $t \to \infty$ we obtain $lim_{t\to\infty}\mathbb{P}^*(\zeta_{n+1} > t) < \delta$ a.s. and since δ was arbitrary $lim_{t\to\infty}\mathbb{P}^*(\zeta_{n+1} > t) = 0$ a.s.. Finally the induction

step follows from

$$\begin{aligned}\mathbb{P}(\zeta_{n+1} < \infty) &= \mathbb{E}\big[\mathbb{P}(\zeta_{n+1} < \infty | \cap_{j=1}^{n} \{\zeta_k < \infty\}, T)\big] \\ &= \mathbb{E}\big[1 - \lim_{t\to\infty} \mathbb{P}^*(\zeta_{n+1} > t)\big] = 1\end{aligned}$$

This completes the proof. □

Proposition 2 will follow directly from the following two lemmas.

Lemma 15. *Suppose that the initial urn composition satisfy assumption 1 and 2. Then for every $(i,j,t) \in S^*$ proposition 2 holds if $\mathbb{P}(X_n = (i,j,t)\ i.o.) = 1$.*

Proof. Let $(i,j,t) \in S^*$ and consider the set of possible successor states of (i,j,t), $R_{(i,j,t)} \subset \{(0,0,0)\} \bigcup G_1 \times G_2 \times \{t+1\}$.

If $n_{(i,j,t)}(a_0) > 0$ and $n_{(i,j,t)}(a_v) = 0$ for all $v \in G_1$, then $R_{(i,j,t)} = \{(0,0,0)\}$) and the process moves from (i,j,t) to $(0,0,0)$ with probability one. Hence we can write the degenerated point mass $\Pi_{(i,j,t)} = \delta_{(0,0,0)}$ in the required form where $p(i,j,t) \sim Dirichlet(n_{(i,j,t)}(a_0), 0, \cdots, 0)$. And since $p(i,j,t)_v = 0$ a.s. for $v \in G_1$ we can choose the random conditional distribution $p(i,j,t|v)$ arbitrary as a Dirichlet distribution with the required parameters.

If on the other hand there exists some $v \in G_1$ such that $n_{(i,j,t)}(a_v) > 0$, then by definition once the process moves to (i,j,t), say $X_n = (i,j,t)$, the first coordinate of a successor state equals $v \in \{0\} \cup G_1$ if the color of the ball sampled from the Pólya urn $U^{(1)}(i,j,t)$ equals a_v. That is, if $\phi(i,j,t)_{t_n(i,j,t)} = a_v$, where $\{\phi(i,j,t)_n\}_n$ is the sequence of colors draws from Pólya urn $U^{(1)}(i,j,t)$. Let $I_{n,v} = \sum_{m=1}^{n} I\{\phi(i,j,t)_m = a_v\}$ for $0 \le v \le N$. From [21]

$$p(i,j,t) \overset{a.s.}{=} \lim_n n^{-1}(I_{n,0}, \cdots, I_{n,N}) \sim \text{Dirichlet}\Big(n_{(i,j,t)}(a_0), \cdots, n_{(i,j,t)}(a_N)\Big)$$

$$\phi(i,j,t|v)_n \overset{iid}{\sim} \text{Multinomial}(1, p(i,j,t))$$

For $v \in G_1$ such that $n_{(i,j,t)}(v) > 0$, $p(i,j,t)_v > 0$ a.s., since $p(i,j,t)_v \sim \text{Beta}\Big(n_{(i,j,t)}(a_v),\ \sum_{c=0,c\neq v}^{N} n_{(i,j,t)}(a_c)\Big)$. Hence by the Borel-Cantelli lemma

$P[\phi(i,j,t|v)_n = a_v \; i.o. | p(i,j,t)] = 1$ a.s. and by Fubini theorem $P[\phi(i,j,t|v)_n = a_v \; i.o.] = 1$. Which implies that $\{X_n\}$ moves infinity often from (i,j,t) to a state in $\{v\} \times G_2 \times \{t+1\}$ with probability one.

Now, conditional on the event $\{X_n = (i,j,t), X_{n+1} \in \{v\} \times G_2 \times \{t+1\}\}$, X_{n+1} will be equal to $(v,w,t+1)$ if the color of the ball drawn from the second level Pólya urn $U^{(2)}(i,j,t|v)$ equals b_w. That is, if $\theta(i,j,t|v)_{t_n(i,j,t)(v,\cdot,t+1)} = b_w$ where $\{\theta(i,j,t|v)_n\}_n$ denotes the sequence of successive colors drawn from urn $U^{(2)}(i,j,t|v)$. Since we draw, with probability one, infinitely often from $U^{(2)}(i,j,t|v)$, let $J_{n,w} = \sum_{m=1}^{n} I\{\theta(i,j,t|v)_m = b_w\}$ for $1 \leq w \leq M$ and again

$$p(i,j,t|v) \stackrel{a.s.}{=} \lim_n n^{-1}(J_{n,1}, \cdots, J_{n,M}) \sim \text{Dirichlet}\Big(m_{(i,j,t|v)}(b_1), \cdots, m_{(i,j,t|v)}(b_M)\Big)$$

This argument holds for all $v \in G_1$ such that $n_{(i,j,t)}(v) > 0$. Conditional of the event of drawing infinitely often from each urn $U^{(2)}(i,j,t|v)$, such that $n_{(i,j,t)}(v) > 0$, the actual draws are independent and the limits as well.

Now, denote with $\{V^{(n)}\}_n$ the sequence of empirical transition matrices with elements

$$v_{s,s'}^{(n)} = \begin{cases} \dfrac{t_n(s)(s')}{t_n(s)} & \text{if } t_n(s) > 0 \\ 0 & \text{otherwise} \end{cases},$$

From [44], for every recurrent partial exchangeable process $V^{(n)} \to V_\infty \; a.s.$ in the topology of coordinate-wise convergence where the random element V_∞ belongs to the space of stochastic matrices on $S \times S$ and, given X_0, $V_\infty \stackrel{d}{=} \Pi$.

Hence, for each $(i,j,t) \in S^*$, coordinate-wise

$$\Pi_{(i,j,t),s} \stackrel{d}{=} \lim_{n\to\infty} \frac{t_n(i,j,t)(s)}{t_n(i,j,t)} \; s \in S. \tag{4.46}$$

For $s \notin G_1 \times G_2 \times \{t+1\} \cup \{0,0,0\}$ clearly $t_n(i,j,t)(s) = 0$ and hence

$$\Pi_{(i,j,t),s} \stackrel{d}{=} \lim_{n\to\infty} \frac{t_n(i,j,t)(s)}{t_n(i,j,t)} = 0$$

For $s = (v, w, t+1)$ such that $n_{(i,j,t)}(a_v) > 0$

$$\begin{aligned}\Pi_{(i,j,t),(v,w,t+1)} &\overset{d}{=} \lim_n \frac{t_n(i,j,t)(v,w,t+1)}{t_n(i,j,t)} \\ &\overset{a.s.}{=} \lim_n \frac{t_n(i,j,t)(v,\cdot,t+1)}{t_n(i,j,t)} \frac{t_n(i,j,t)(v,w,t+1)}{t_n(i,j,t)(v,\cdot,t+1)} \\ &\overset{a.s.}{=} \left(\lim_k \frac{I_{k,v}}{k}\right)\left(\lim_k \frac{J_{k,w}}{k}\right) \\ &\overset{a.s.}{=} p(i,j,t)_v p(i,j,t|v)_w.\end{aligned}$$

For $s = (v, w, t+1) \in S$ such that $n_{(i,j,t)}(a_v) = 0$ clearly $t_n(i,j,t)(v,\cdot,t+1) = 0$ for all n and since $0 \overset{a.s.}{=} p(i,j,t)_v \sim Beta(n_{(i,j,t)}(a_v),\ \sum_0^N n_{(i,j,t)}(a_u))$ we may write

$$\begin{aligned}\Pi_{(i,j,t),(v,w,t+1)} &\overset{d}{=} \lim_{n\to\infty} \frac{t_n(i,j,t)(v,w,t+1)}{t_n(i,j,t)} \\ &= 0 \overset{a.s.}{=} p(i,j,t)_v p(i,j,t|v)_w\end{aligned}$$

where we used again the fact that we can choose $p(i,j,t|v)$ arbitrary. Finally for the last case, $s = (0,0,0)$, we can write the limit as

$$\begin{aligned}\Pi_{(i,j,t),(0,0,0)} &\overset{d}{=} \lim_{n\to\infty} \frac{t_n(i,j,t)(0,0,0)}{t_n(i,j,t)} \\ &\overset{a.s.}{=} 1 - \lim_{n\to\infty} \sum_{v=1}^{N} \frac{t_n(i,j,t)(v,\cdot,t+1)}{t_n(i,j,t)} \\ &\overset{a.s.}{=} 1 - \sum_{v=1}^{N} \lim_{k\to\infty} \frac{I_{k,v}}{k} \overset{a.s.}{=} 1 - \sum_{1}^{N} p(i,j,t)_v.\end{aligned}$$

It remains to show that the transition rows of Π are mutually independent. Notice that for distinct states $(i_1,j_1,t_1),(i_2,j_2,t_2),\cdots,(i_k,j_k,t_k)$ in S^* the transition rows $\Pi_{(i_1,j_1,t_1)},\Pi_{(i_2,j_2,t_2)},\cdots,\Pi_{(i_k,j_k,t_k)}$ are functions of regular conditional distributions with respect to the tail σ-fields generated by the

independent sequences

$$\Big(\{\phi(i_1,j_1,t_1)_n\},\{\theta(i_2,j_2,t_2|v)_n\},v\in G_1\Big),$$
$$\dots$$
$$\Big(\{\phi(i_k,j_k,t_k)_n\},\{\theta(i_k,j_k,t_k|v)_n\},v\in G_1\Big)$$

hence $\Pi_{(i_1,j_1,t_1)},\Pi_{(i_2,j_2,t_2)},\cdots,\Pi_{(i_k,j_k,t_k)}$ are mutually independent. Since k and the states $(i_1,j_1,t_1),(i_2,j_2,t_2),\cdots,(i_k,j_k,t_k)$ in S^* are arbitrary the result follows. □

Lemma 16. *For every* $s\in S^*$ $\mathbb{P}(X_n=s\ i.o.)=1$

Proof. We show that $P(X_m=s\ i.o.)=1$ for every n and every $s\in S_n^*$. For $n=0$ $S_0^*=\{(0,0,0)\}$ this follows from Lemma 29. Assume now that for every $s\in S_n^*$ $P(X_m=s\ i.o.)=1$. For s' in S_{n+1}^* arbitrary there exists $s\in S_n^*$ such that $s'\in R_{(s)}$. Assume that $s'\neq(0,0,0)$, since otherwise the recurrence follows from Lemma 29. Hence write $s=(i,j,t)$ and $(v,w,t+1)$ for appropriate $v\in G_1,w\in G_2$ and $(i,j,t)\in S$.

Since by assumption $P(X_m=(i,j,t)\ i.o.)=\mathbb{E}[P(X_m=(i,j,t)\ i.o.|\Pi)]=1$, there exists a set $A\in\mathcal{F}$ such that $P(A)=0$ and for $\omega\in A^c$ $P(X_m=(i,j,t)\ i.o.|\Pi)(\omega)>0$. But by elementary properties for Markov chains for all $\omega\in A^c$ $P(X_m=(i,j,t)\ i.o.|\Pi)(\omega)\in\{0,1\}$. Hence $P(X_m=(i,j,t)\ i.o.|\Pi)=1$ on A^c.

Furthermore, by definition of $R_s=R_{(i,j,t)}$ we have that $n_{(i,j,t)}(a_v)>0,m_{(i,j,t|v)}(b_w)>0$. Now use the previous Lemma to see that $P(X_{m+1}=(v,w,t+1)|X_m=(i,j,t),\Pi)=p(i,j,t)_v p(i,j,t|v)_w>0$ on a set $B^c\in\mathcal{F}$ such that $P(B)=0$. Hence for all $\omega\notin A\cup B$ the state s is recurrent and communicates with s', which by elementary properties of Markov chains implies that $P(X_m=s'\ i.o.|\Pi)=1$ on the set $A^c\cap B^c$ of probability one. Now the result $P(X_m=s'\ i.o.)=1$ follows from Fubini theorem. Since $s'\in S_{n+1}$ was arbitrary the induction step follows. □

Proof of Corollary 4. By an arguments similar to the inductions step of lemma 29 one can show that $\{X_{m+n}\}_n$ is recurrent. Let $Z_k=X_{v_n+k}$. Then by the

reasoning above $\{Z_k\}$ is a recurrent BUP with initial state X_{v_n} ($(0,0,0)$ if $v_n = \zeta_n$) and initial urn compositions at the first and second level as stated in (4.19), (4.20) and (4.21) (for $v_n = \zeta_n$ the counting processes $t_n(\cdot)$ are replaced by $t_{\zeta_n}(\cdot)$). Now, by corollary 3 $\{Z_n\}_n = \{X_{m+n}\}_n$ is a mixture of Markov chains where the mixing distribution follows from proposition 2. □

Proof of Lemma 11. Fix $n \geq 1$ and consider the first n blocks $(\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_n)$. Let σ be an arbitrary permutation of $\{1, \cdots, n\}$. We want to show that

$$(\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_n) =^d (\mathbf{B}_{\sigma(1)}, \mathbf{B}_{\sigma(2)}, \cdots, \mathbf{B}_{\sigma(n)}) \tag{4.47}$$

Let (i,j), $(v,w) \in G_1 \times G_2$. For any $l \geq 1$ the number of transitions form (i,j,l) to $(v,w,l+1)$ in $\mathbf{B}_s$ is one if $\tau_s > l$ and the $(l+1)^{th}$ coordinate of $\mathbf{B}_s$ equals (i,j,l) and the $(l+2)^{th}$ coordinate equals $(v,w,l+1)$, whereas it is zero otherwise. Let $t_{B_1,\cdots,B_n}(i,j,l)(v,w,l+1)$ denote the number of transitions from (i,j,l) to $(v,w,l+1)$ in the ordered string $(\mathbf{B}_1, \cdots, \mathbf{B}_n)$ and define similarly $t_{B_{\sigma(1)},\cdots,B_{\sigma(n)}}$ for the ordered string $(\mathbf{B}_{\sigma(1)}, \cdots, \mathbf{B}_{\sigma(n)})$. It follows that

$$\begin{aligned}
t_{B_1,\cdots,B_n}(i,j,l)(v,w,l+1) &= \sum_{s=0}^{n-1} I(X_{\zeta_s+l} = (i,j,l), X_{\zeta_s+l+1} = (v,w,l+1), \tau_s > l) \\
&= \sum_{s=0}^{n-1} I(X_{\zeta_{\sigma(s)}+l} = (i,j,l), X_{\zeta_{\sigma(s)}+l+1} = (v,w,l+1), \tau_{\sigma(s)} > l) \\
&= t_{B_{\sigma(1)},\cdots,B_{\sigma(n)}}(i,j,l)(v,w,l+1); \\
t_{B_1,\cdots,B_n}(i,j,l)(0,0,0) &= \sum_{s=0}^{n-1} I_{\{X_{\zeta_s+l}=(i,j,l), \tau_s = l\}} = \sum_{s=0}^{n-1} I_{\{X_{\zeta_{\sigma(s)}+l}=(i,j,l), \tau_{\sigma(s)} = l\}} \\
&= t_{B_{\sigma(1)},\cdots,B_{\sigma(n)}}(i,j,l)(0,0,0).
\end{aligned}$$

Clearly this implies also that $t_{B_{1,\cdots,n}}(i,j,l)(v,\cdot,l+1) = t_{B_{\sigma(1),\cdots,\sigma(n)}}(i,j,l)(v,\cdot,l+1)$. But since all blocks start with the same initial state $(0,0,0)$ this implies that $\mu_1 = (\mathbf{B}_1, \cdots, \mathbf{B}_n)$ and $\mu_2 = (\mathbf{B}_{\sigma(1)}, \cdots, \mathbf{B}_{\sigma(n)})$ are equivalent sequences.

By partial exchangeability of $\{X_n\}$ we can deduce that

$$\begin{aligned}
&\mathbb{P}\big[\cap_{s=1}^{n}\{\mathbf{B}_s = \big((i_l^s, j_l^s); l = 1, ..., \tau_s\big)\}\big] \\
&= \mathbb{P}\big[\cap_{s=0}^{n-1}\{X_{\zeta_s} = (0,0,0), X_{\zeta_s+l} = (i_l^s, j_l^s); l = 1, ..., \tau_s\}\big] \\
&= \prod_{\substack{l\geq 0 \\ i\in G_1\cup\{0\} \\ j\in G_2\cup\{0\}}} \prod_{v=1}^{N} \Bigg[\frac{n_{(i,j,l)}(a_0)^{[t_{B_1,\cdots,B_n}(i,j,l)(0,0,0)]} n_{(i,j,l)}(a_v)^{[t_{B_1,\cdots,B_n}(i,j,l)(v,\cdot,l+1)]}}{\Big(\sum_{c=0}^{N} n_{(i,j,l)}(a_c)\Big)^{[t_{B_1,\cdots,B_n}(i,j,l)]}} \\
&\qquad \times \frac{\prod_{w=1}^{M} m_{(i,j,l|v)}(b_w)^{[t_{B_1,\cdots,B_n}(i,j,l)(v,w,l+1)]}}{\Big(\sum_{c=0}^{M} m_{(i,j,l|v)}(b_c)\Big)^{[t_{B_1,\cdots,B_n}(i,j,l)(v,\cdot,l+1)]}} \Bigg] \\
&= \mathbb{P}\big[\cap_{s=1}^{n}\{\mathbf{B}_s = \big((i_l^{\sigma(s)}, j_l^{\sigma(s)}); l = 1, ..., \tau_s\big)\}\big] \qquad (4.48)
\end{aligned}$$

Where the last equality holds by equality of all counting processes of $(\mathbf{B}_s)_{1\leq s\leq n}$ and $(\mathbf{B}_{\sigma(s)})_{1\leq s\leq n}$. □

Proof of Lemma 12. Notice first that $t_K[s;(i,j)(v,w)] = t_{\tau_K+s}(i,j,s)(v,w,s+1)$, hence

$$\begin{aligned}
&\hat{P}\big[Y_{s+1} = (v,w)|Y_s = (i,j)\big] \\
&= P\big[Y_{s+1}^{(K+1)} = (v,w)|Y_s^{K+1} = (i,j), \{\mathbf{Y}^{(l)}\}_{1\leq l\leq K}\big] \\
&= \frac{n_{(i,j,s)}(a_v) + t_K[s;(i,j)(v,\cdot)]}{\sum_{c=0}^{N} n_{(i,j,s)}(a_c) + t_K[s;(i,j)]} \frac{m_{(i,j,s|v)}(b_w) + t_K[s;(i,j)(v,w)]}{\sum_{c=1}^{M} m_{(i,j,s|v)}(b_c) + t_K[s;(i,j)(v,\cdot)]}
\end{aligned}$$

if $(v,w) \in G_1 \times G_2$. For $(v,w) = (0,0)$ the case is similar. (ii) follows from (i) by marginalizing. Finally the predicted s-step transition probability of $\{Y_t^{(2)}\}$ can be computed as

$$\begin{aligned}
&\hat{P}[Y_s^{(2)} = w] = P\big[Y_s^{(K+1)} \in G_1 \times \{w\}|\{\mathbf{Y}^{(l)}\}_{1\leq l\leq K}\big] \\
&= P\big[\cap_{j=1}^{s-1}\{Y_j^{(K+1)} \in G_1 \times G_2\}, Y_s^{(K+1)} \in G_1 \times \{w\}|\{\mathbf{Y}^{(l)}\}_{1\leq l\leq K}\big] \\
&= E\Bigg[\sum_{v=1}^{N} \big(\Pi_{-0}^{(\zeta_K)}(0) \prod_{j=1}^{s-1} \Pi_{-0}^{(\zeta_K)}(j)\big)_{(v,w))}\Bigg] \\
&= \sum_{v=1}^{N} \Big(\hat{\Pi}_{-0}^{(K)}(0) \prod_{j=1}^{s-1} \hat{\Pi}_{-0}^{(K)}(j)\Big)_{(v,w)},
\end{aligned}$$

The computation of $\hat{P}[Y_{s+1} = v]$ and $\hat{P}[Y^{(1)}_{s+1} = v]$ are similar. □

Proof of Lemma 13. We compute the expected time spend in a state $w \in G_2$

$$\begin{aligned}\hat{\mathbb{E}}[T(Y)_{(\cdot,w)}]) &= \mathbb{E}\big[T(\mathbf{Y}^{(K+1)})_{(\cdot,w)}]|\{\mathbf{Y}^{(k)}\}_{1\le k\le K}\big] \\ &= \sum_{s\ge 1}\mathbb{P}(Y^{(K+1)}_s \in G_1\times\{w\}|\{\mathbf{Y}^{(k)}\}_{1\le k\le K}\big] \\ &= \sum_{s\ge 1}\mathbb{E}\Big[\sum_{v=1}^{N}\big(\Pi^{(\zeta_K)}_{-0}(0)\prod_{t=1}^{s-1}\Pi^{(\zeta_K)}_{-0}(t)\big)_{(v,w)}\Big] \\ &= \sum_{s\ge 1}\sum_{v=1}^{N}\big(\hat{\Pi}^{(K)}_{-0}(0)\prod_{t=1}^{s-1}\hat{\Pi}^{(K)}_{-0}(t)\big)_{(v,w)}.\end{aligned}$$

The derivation of $\hat{\mathbb{E}}[T(Y)_{(i,j)}]$ and $\hat{\mathbb{E}}[T(Y)_{(i,\cdot)}]$ are similar. (ii) follows from

$$\begin{aligned}\hat{P}(\tau > s) &= P(\cap_{l=1}^{s+1}\{Y_l^{(K+1)} \in G_1\times G_2\}|\{\mathbf{Y}^{(k)}\}_{1\le k\le K}) \\ &= E\Big[\sum_{v,w=1}^{N,M}\big(\Pi^{(\zeta_K)}_{-0}(0)\prod_{l=1}^{s}\Pi^{(\zeta_K)}_{-0}(l)\big)_{(v,w)}\Big] \\ &= \sum_{v,w=1}^{N,M}\big(\hat{\Pi}^{(K)}_{-0}(0)\prod_{l=1}^{s}\hat{\Pi}^{(K)}_{-0}(l)\big)_{(v,w)}\end{aligned}$$

and

$$\begin{aligned}&\hat{P}[\tau > s|Y_s = (i,j)] \\ &\quad= P(Y^{(K+1)}_{s+1} \in G_1\times G_2|Y^{(K+1)}_s = (i,j), \{\mathbf{Y}^{(k)}\}_{1\le k\le K}) \\ &\quad= \mathbb{E}\Big[1 - \Pi^{(\zeta_K)}(s)_{(i,j),(0,0)}\Big] = 1 - \frac{n_{(i,j,s)}(a_0) + t_K[s;(i,j)(0,0)]}{\sum_{c=0}^{N} n_{(i,j,s)}(a_c) + t_K[s;(i,j)]}.\end{aligned}$$

Finally

$$\begin{aligned}\hat{E}[\tau] &= \mathbb{E}[\tau_{K+1}|\{Y_k\}_{1\leq k\leq K}] \\ &= \sum_{s\geq 0} P(\tau_{K+1} > s|\{Y_k\}_{1\leq k\leq K}) \\ &= \sum_{s\geq 0} \sum_{(v,w)\in G_1\times G_2} \left(\hat{\Pi}^{(K)}_{sub}(0)\prod_{l=1}^{s}\hat{\Pi}^{(K)}_{sub}(l)\right)_{(v,w)}.\end{aligned}$$

□

Chapter 5

Reinforced Multivariate Urn Processes

5.1 Introduction

In the following chapter we extend the analysis of inhomogeneous Markov chains of the previous sections to the case of multivariate state-spaces. We focus again on exchangeable trajectories $\{\mathbf{Y}_n\}$, where each element $\mathbf{Y}_n = \{Y_k^{(n)}\}_{0\leq k\leq \tau_n}$ is a mixtures of Markov chains of random length $\tau_n < +\infty$. Furthermore each vector $Y_k^{(n)}$ will take values on a finite product space $G \subset \mathbb{N}^k$. The construction will be based on the same idea as before. We will define first a Random walk on a space of Pólya urns. This time the space will have a hierarchical dimension as well. Then we will use again a sequence of hitting times and the sequence of states between successive hitting times to construct a sequence of random elements $\{\mathbf{Y}_n\}$. Predictive estimator for time-inhomogeneous Markov chains are implemented through the predictive law of $\mathbf{Y}_{M+1}$, given $\{\mathbf{Y}_n\}_{n\leq M}$.

The outline of the paper is as follows. In section 5.2 we introduce the hierarchical urn process. We study some basic characteristics of the process which will again lead to a mixture representation. In section 5.3 we will use the hierarchical urn process to construct an exchangeable sequence of Mixture of inhomogeneous Markov chains and derive predictive estimator

for some characteristics of the process. In the final section we conclude with some potential direction of future work.

5.2 A multivariate reinforced hierarchical urn process

The aim of this section is to construct a stochastic process which we use to derive exchangeable mixtures of Markov chains on a finite multivariate state-space. We use again a random walk on a space of hierarchical connected Pólya urns.

To simplify exposure we introduce some notation. In the sequel bold letters $\mathbf{i} = (i_1, \cdots, i_K)$ denote integer valued vectors. $i_{1:k} = (i_1, \cdots, i_k)$ denotes the sub-vector which consists of the first k-coordinates of $\mathbf{i}$. Furthermore, if V is a set of the form $V = (\times_{j=1}^{K} D_j) \times W$, we sometimes write elements of V as $\mathbf{v} = (\mathbf{d}, w)$ where $\mathbf{d} = (d_1, \cdots, d_K) \in \times_{j=1}^{K} D_j$ and $w \in W$. If necessary we also use $\mathbf{v} = \big((d_{1:k}, d_{(k+1):K}), w\big)$ with $d_{(k+1):K} = (d_{k+1}, \cdots, d_K)$.

To define the process we introduce a state space and a system of Pólya urn.

(i) The state space will have the form $S = \Big(\times_{k=1}^{K} G_k\Big) \times \mathbb{N}_0$, where the single set G_k has the form $G_k = \{1, \cdots, N_k\}$ for $N_k < \infty$ and $1 \leq k \leq K$.

(ii) We use K sets of labels/colors C_k, $k = 1, \cdots, K$ where $C_k = G_k$ for $k = 2, \cdots, K$ and $C_1 = \{0\} \cup G_1$.

(iii) We will use the following system of Pólya urns. Connect with every state $s \in S$ a Pólya urn $U^{(1)}(s)$ with color set C_1 and initial composition

$$\big(m_s(0), \cdots, m_s(N_1)\big) \in \mathbb{R}_{+}^{N_1+1}, \tag{5.1}$$

where $m_{(\mathbf{i},s)}(j)$ denotes the initial number of balls of color $0 \leq j \leq N_1$ and $m_{(\mathbf{i},s)} = \sum_{j=0}^{N_1} m_{(\mathbf{i},s)}(j) > 0$ denotes the total initial mass of urn $U^{(1)}(s)$.

Furthermore, for $k = 1, \cdots, K-1$ connect to the same state s and every sub-state $j_{1:k} = (j_1, \cdots, j_k)$ in $\times_{j=1}^{k-1} G_j$ a $(k+1)$-th level Pólya urn

$U^{(k+1)}((\mathbf{i},s)|j_{1:k})$ with color set C_{k+1} and initial composition

$$\big(m_{(\mathbf{i},s|j_{1:k})}(1),\cdots,m_{(\mathbf{i},s|j_{1:k})}(N_{k+1})\big)\in\mathbb{R}_+^{N_{k+1}}. \tag{5.2}$$

Where $m_{(\mathbf{i},s|j_{1:k})}(j)$ denotes the number of balls of color $1\leq j\leq N_k$ initially in urn $U^{(k+1)}((\mathbf{i},s)|j_{1:k})$ and $m_{(\mathbf{i},s|j_{1:k}}=\sum_{j=1}^{N_k}m_{(\mathbf{i},s|j_{1:k})}(j)>0$ indicated the total mass.

(iv) Finally, similar to (iii) we introduce a second system of urns which we call return urns. Let $U^{(*1)}$ be a first level return urn with color set $C_1\setminus\{0\}$ and initial composition

$$(\tilde{m}(1),\cdots,\tilde{m}(N_1))\in\mathbb{R}_+^{N_1}. \tag{5.3}$$

And for $2\leq k\leq K$ connect to every sub-state $j_{1:(k-1)}$ in $\times_{j=1}^{k-1}G_j$ a k-th level 'return' Pólya urn $U^{(*k)}(j_{1:(k-1)})$ with color set C_k and initial urn composition

$$\big(\tilde{m}_{(j_{1:k})}(1),\cdots,\tilde{m}_{(j_{1:k})}(N_{k+1})\big)\in\mathbb{R}_+^{N_{k+1}} \tag{5.4}$$

and initial total mass $\tilde{m}_{j_{1:k}}=\sum_{j=1}^{N_{k+1}}\tilde{m}_{(j_{1:k})}(j)>0$.

The idea is the following. Based on a system of hierarchical Pólya urns we want to construct a process $\{X_n\}_{n\geq 0}$, where each value of X_n is determined by a hierarchical sampling scheme. Write the state space S as $G\times\mathbb{N}_0$ where $G=\times_{k=1}^{K}G_k$. If the process visits a given state $(\mathbf{i},s)\in G\times\mathbb{N}_0$ the next state, say $(\mathbf{j},t)$, will be sampled in a hierarchical way: (forget for the moment about the component t)

(1) sample at first j_1 conditional on $\big(\mathbf{i},s\big)$,

(2) sample j_2 conditional on $\big((\mathbf{i},s),j_1\big)$, and continue until

(K) sample j_K conditional on $\big((\mathbf{i},s),j_{1:(K-1)}=(j_1,\cdots,j_{K-1})\big)$.

In the following we show two equivalent way of expressing this idea. The first one uses a sampling scheme, whereas the second way translates the sampling scheme into a more formal definition. To ensure that the process is well defined we will require the minor assumptions that all urns have strictly positive total mass. Figure 5.1 gives a definition of the process in terms of a sampling scheme. The same sampling scheme is represented graphically in

figure 5.3.

Remark 16. *The process samples a sequence of trajectories in* $G \times \mathbb{N}_0$*, say* $\{\boldsymbol{B}_n\}$*, where each element* $\boldsymbol{B}_n$ *starts in* $G \times \{0\}$ *and moves along* $G \times \{t\}, t \geq 1$*. The set* G *corresponds to a vector of states of* $\boldsymbol{B}_n$ *visited during the lifetime which is represented by the second component* $\mathbb{N}_0$ *of* S*. Figure 5.2 shows a sample path of the process until the first two blocks* B_1 *and* B_2 *terminate.*

If the process is recurrent, in the sense that $P(X_n = X_0 \ i.o.) = 1$*, then* $\{\boldsymbol{B}_n\}$ *will be an infinite sequence and each trajectory will be finite with probability one. We use this sequence of trajectory in section 5.3 to construct our target process* $\{\boldsymbol{Y}_n\}$ *of exchangeable mixtures of inhomogeneous Markov chains on the multivariate state space* G*.*

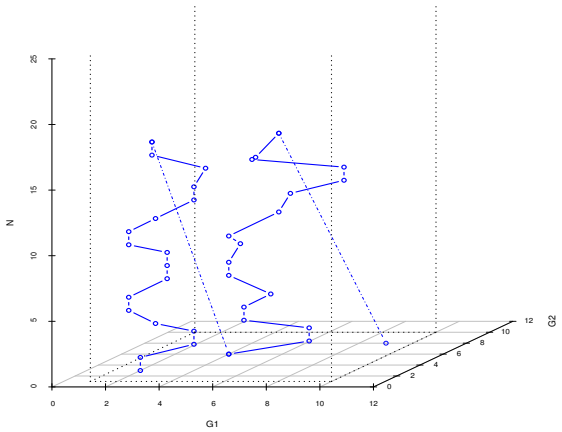
Figure 5.1: Sampling scheme of the process $\{X_n\}_{n\geq 0}$

(A) Sample X_0 as follows
(1) Draw a ball from urn $U^{(*1)}$ and denote the sampled color with i_1^0.
(2) Draw a ball from urn $U^{(*2)}(i_1^0)$ and denote the sampled color with i_2^0.
...
(K) Draw a ball from urn $U^{(*K)}(i_{1:(K-1)}^0)$ and denote the sampled color with i_K^0.
$\Rightarrow$ Set $X_0 = (\mathbf{i}^0, 0)$ and for $1 \leq k \leq K$ replace two balls of color i_k^0 to urn $U^{(*k)}(i_{1:(k-1)}^0)$.

(B) Sample X_{n+1} given $\{X_j = (\mathbf{i}^k, s_k), 0 \leq j \leq n\}$ as follows
(a) Draw a ball from urn $U^{(1)}(\mathbf{i}^n, s_n)$ and denote the sampled color with j^{n+1}.
(b.1) if $j^{n+1} = 0$
(1) Draw a ball from urn $U^{(*1)}$ and denote the sampled color with i_1^{n+1}.
(2) Draw a ball from urn $U^{(*2)}(i_1^{n+1})$ and denote the sampled color with i_2^{n+1}.
...
(K) Draw a ball from urn $U^{(*K)}(i_{1:(K-1)}^{n+1})$ and denote the sampled color with i_K^{n+1}.
$\Rightarrow$ Set $X_{n+1} = (\mathbf{i}^{n+1}, 0)$. Replace two balls of color 0 to urn $U^{(1)}(\mathbf{i}^n, s_n)$ and for $k = 1, \cdots, K$ replace two balls of the color i_k^{n+1} to urn $U^{(*k)}(i_{1:(k-1)}^{n+1})$.

(b.2) if $j^{n+1} \in G_1$ set $i_1^{n+1} = j^{n+1}$
(2) Draw a ball from urn $U^{(2)}(\mathbf{i}^n, s_n | i_1^{n+1})$ and denote the sampled color with i_2^{n+1}.
...
(K) Draw a ball from urn $U^{(K)}(\mathbf{i}^n, s_n | i_{1:(K-1)}^{n+1})$ and denote the sampled color with i_K^{n+1}.
$\Rightarrow$ Set $X_{n+1} = (\mathbf{i}^{n+1}, s_n + 1)$ and for $k = 1, \cdots, K$ replace two balls of color i_k^{n+1} to urn $U^{(k)}(\mathbf{i}_n, s_n | i_{1:(k-1)}^{n+1})$.

Figure 5.2: Example of the a sample path of the process $\{X_n\}$ on the cube $\{1, \cdots, 10\}^2 \times \mathbb{N}_0$ and the resulting trajectories

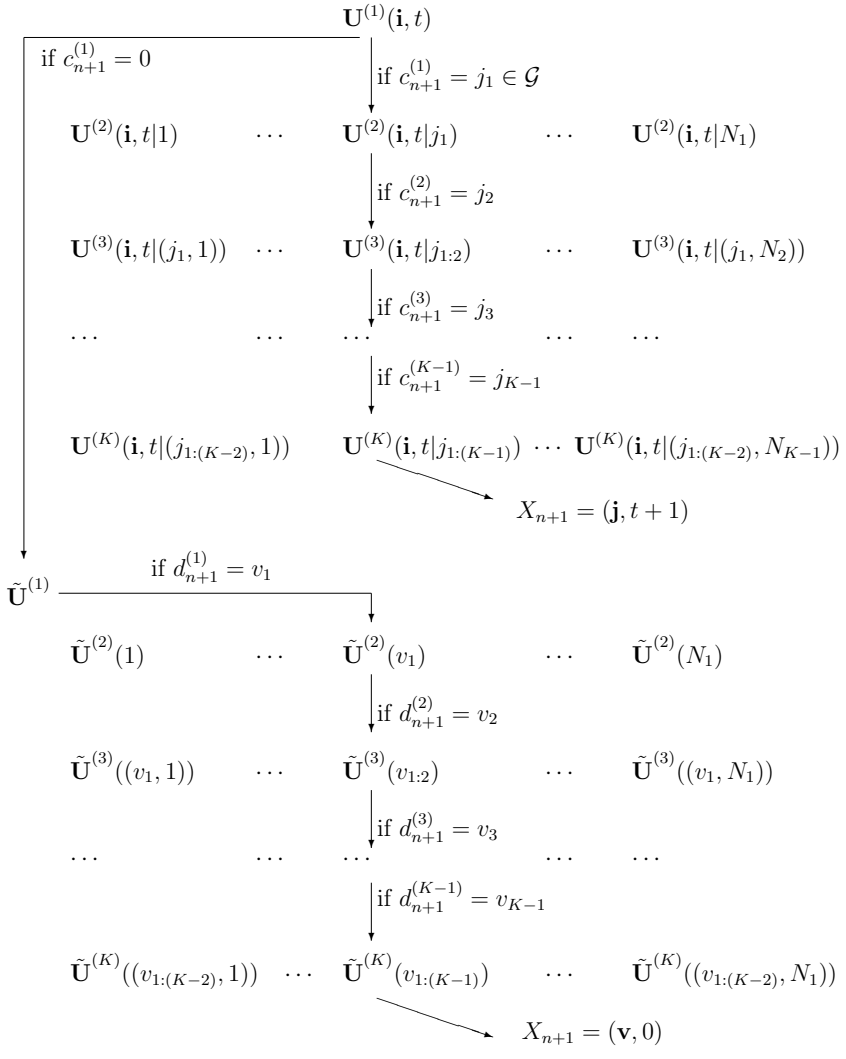


We formalize the sampling scheme for the further analysis by defining the process through a system of predictive distributions.

Definition 17. *The process $X = \{X_k\}_{k\geq 0}$, called a reinforced hierarchical urn process (RHUP), with values in S is defined recursively by*

$$P[X_0 = (\boldsymbol{j}, s)] = \begin{cases} \frac{\tilde{m}(j_1)}{\tilde{m}} \prod_{k=1}^{K-1} \frac{\tilde{m}_{(j_{1:k})}(j_{k+1})}{\tilde{m}_{(j_{1:k})}} & \text{if } (\boldsymbol{j}, s) \in G \times \{0\} \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

Figure 5.3: Sampling state X_{n+1} given $\{X_k\}_{0\leq k\leq n}$ when $X_n = (\mathbf{i}, t)$. $\mathbf{U}^{(m)}(\mathbf{i}, t|j_{1:(m-1)})$ denotes the urn connected to $(\mathbf{i}, t)$ and sub-state $j_{1:(m-1)}$ and $c_{n+1}^{(m)}$ is the color of the ball extracted from $\mathbf{U}^{(m)}(\mathbf{i}, t|j_{1:(m-1)})$. $\tilde{\mathbf{U}}^{(m)}(j_{1:(m-1)})$ denoted the return urn connected to sub-state $j_{1:(m-1)}$ and $d_{n+1}^{(m)}$ is the color of the ball extracted from $\tilde{\mathbf{U}}^{(m)}(j_{1:(m-1)})$.



and conditional on the history $\mathcal{F}_n = \sigma\big(\{X_l\}_{0\le l\le n}\big)$ *where* $X_n = (\boldsymbol{i}, s)$*, let*

$$P[X_{n+1} = (\boldsymbol{j}, v)|\mathcal{F}_n] \tag{5.6}$$
$$= \begin{cases} \dfrac{m_{(\boldsymbol{i},s)}(j_1) + t_n(\boldsymbol{i},s)(j_1, s+1)}{m_{(\boldsymbol{i},s)} + t_n(\boldsymbol{i},s)} & \\ \times \prod_{k=1}^{K-1} \dfrac{m_{(\boldsymbol{i},s|j_{1:k})}(j_{k+1}) + t_n(\boldsymbol{i},s)(j_{1:(k+1)}, s+1)}{m_{(\boldsymbol{i},s|j_{1:k})} + t_n(\boldsymbol{i},s)(j_{1:k}, s+1)} & \text{if } (\boldsymbol{j}, v) \in G \times \{s+1\} \\ \dfrac{m_{(\boldsymbol{i},s)}(0) + t_n(\boldsymbol{i},s)(\cdot, 0)}{m_{(\boldsymbol{i},s)} + t_n(\boldsymbol{i},s)} & \\ \times \dfrac{\tilde{m}(j_1) + t_n^*(j_1)}{\tilde{m} + t_n^*} \prod_{k=1}^{K-1} \dfrac{\tilde{m}_{(j_{1:k})}(j_{k+1}) + t_n^*(j_{1:(k+1)})}{\tilde{m}_{(j_{1:k})} + t_n^*(j_{1:k})} & \text{if } (\boldsymbol{j}, v) \in G \times \{0\} \\ 0 & \text{otherwise} \end{cases}$$

where for $(\boldsymbol{i}, s)$ *and* $(\boldsymbol{j}, v)$ *in* S

$$t_n(\boldsymbol{i},s)(\boldsymbol{j},v) = \sum_{l=1}^{n} I\{X_{l-1} = (\boldsymbol{i},s), X_l = (\boldsymbol{j},v)\}, \tag{5.7}$$

$$t_n(\boldsymbol{i},s)(\cdot,0) = \sum_{l=1}^{n} I\{X_{l-1} = (\boldsymbol{i},s), X_l \in G \times \{0\}\} \tag{5.8}$$

denotes the number of transitions form $(\boldsymbol{i}, s)$ *to* $(\boldsymbol{j}, v)$ *and from* $(\boldsymbol{i}, s)$ *to a state in* $G \times \{0\}$ *until time n. Furthermore*

$$t_n(\boldsymbol{i},s) = \sum_{\boldsymbol{j}\in G, v\in\{0,s+1\}} t_n(\boldsymbol{i},s)(\boldsymbol{j},v), \tag{5.9}$$

$$t_n(\boldsymbol{i},s)(j_{1:k},v) = \sum_{j_{(k+1):K}\in\times_{k+1}^{K} G_j} t_n(\boldsymbol{i},s)(\boldsymbol{j},v) \tag{5.10}$$

denote the number of visits to state $(\boldsymbol{i}, s)$ *and the number of transition from* $(\boldsymbol{i}, s)$ *into a state* $(\boldsymbol{j}, v)$ *that agrees on the first* k *components with* $j_{1:k} \in$

$\times_{v=1}^{k} G_v$. *Similarly*

$$t_n^*(\boldsymbol{i}) = \sum_{j=0}^{n} I\{X_j = (\boldsymbol{i}, 0)\}, \tag{5.11}$$

$$t_n^*(i_{1:k}) = \sum_{i_{(k+1):K} \in \times_{k+1}^{K} G_j} t_n^*(\boldsymbol{i}), \tag{5.12}$$

$$t_n^* = \sum_{i \in G} t_n^*(\boldsymbol{i}). \tag{5.13}$$

The definition of the predictive distributions may look a bit confusing. But is basically translates the sampling scheme described in figure 5.1 into a transition probability. By construction, if $X_n = (\mathbf{i}, t) \in S$ then $X_{n+1} \in G \times \{0, t+1\}$, hence we call a sequence $(\mathbf{i}_j, s_j)_{0 \le j \le n}$ of elements in S admissible if $(\mathbf{i}_0, s_0) \in G \times \{0\}$ and $(\mathbf{i}_j, s_j) \in G \times \{s_{j-1} + 1, 0\}$ for all $1 \le j \le n$. As usually we determine first the system of finite dimensional distributions of the process.

Proposition 3. *For any $n \ge 1$ and every admissible sequence $(\boldsymbol{i}_j, s_j)_{0 \le j \le n}$ of elements in S*

$$\begin{aligned} &P[X_0 = (\boldsymbol{i}^0, s_0), X_1 = (\boldsymbol{i}^1, s_1), \cdots, X_n = (\boldsymbol{i}^n, s_n)] \qquad (5.14)\\ &= \prod_{\substack{\boldsymbol{i} \in G, \\ s \ge 0}} \left[\left(\frac{\tilde{m}(i_1)^{[t_n^*(i_1)]}}{\tilde{m}^{[t_n^*]}} \prod_{k=1}^{K-1} \frac{\tilde{m}_{(i_{1:k})}(i_{k+1})^{[t_n^*(i_{1:k+1})]}}{\tilde{m}_{(i_{1:k})}^{[t_n^*(i_{1:k})]}} \right) \left(\frac{m_{(\boldsymbol{i},s)}(0)^{[t_n(\boldsymbol{i},s)(\cdot,0)]}}{m_{(\boldsymbol{i},s)}^{[t_n(\boldsymbol{i},s)]}} \right) \right.\\ &\qquad \left. \times \prod_{\boldsymbol{j} \in G} m_{(\boldsymbol{i},s)}(j_1)^{[t_n(\boldsymbol{i},s)(j_1, s+1)]} \prod_{k=1}^{K-1} \frac{m_{(\boldsymbol{i},s|j_{1:k})}(j_{k+1})^{[t_n(\boldsymbol{i},s)(j_{1:k+1}, s+1)]}}{m_{(j_{1:k})}^{[t_n(\boldsymbol{i},s)(j_{1:k}, s+1)]]}} \right] \end{aligned}$$

where $a^{[n]} = \prod_{j=1}^{n-1}(a + j)$ for $n > 1$ and $a^{[m]} = 1$ for all integers $m \le 1$. Whereas every inadmissible sequence of elements in S has zero probability.

We defined the process from a system of predictive distributions (5.5) and (5.6) and found the finite dimensional distributions define by (5.14). The following corollary states that there existence of a stochastic process defined on some probability space which has a finite law given by (5.14).

Corollary 5. *There exists a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process $\{X_n\}_{n\geq 0}$ defined on the same probability space having (5.14) as its system of finite dimensional distributions.*

5.2.1 Recurrence and mixture representation

In the remaining part of this section we explore some basic properties of the process. In principle we want to show that the process can be seen as a mixture of Markov chains on S, provided that the process is recurrent. Meaning we want to represent the finite dimensional law of the process as

$$\mathbb{P}\Big[\bigcap_{k=0}^{n}\{X_k = (\mathbf{i}_k, s_k)\Big] = \delta_0(s_0)\mathbb{E}_\mu\left[\Lambda_{(\mathbf{i}_0,s_0)}\prod_{k=0}^{n-1}\Pi_{(\mathbf{i}_k,s_k);(\mathbf{i}_{k+1},s_{k+1})}\right]. \tag{5.15}$$

where Λ is a random distribution on S and Π is a random transition matrix on S such that $\mu \times \Pi \sim \mu$.

Recall that two strings σ and μ of elements in S are called equivalent if (i) they start with the same initial element and (ii) the number of transitions from state $s \in S$ to $s' \in S$ is the same in σ and μ. Also recall that a stochastic process is called partial exchangeable according to [44], if the finite dimensional law of the process gives the same probability to any two equivalent strings.

Partial exchangeable processes have good theoretical properties. Most importantly a theorem due to Diaconis and Freedman [63, 44] states that every recurrent partial exchangeable process is mixture of Markov chains. We use this result and show first that the process X is partial exchangeable and under some assumptions recurrence as well. This gives the desired representation (5.15). The next proposition states the first property.

Proposition 4. *The process $\{X_n\}$ is partial exchangeable.*

In the following part we provide conditions for recurrence of $\{X_n\}$, which will be the last requirement to represent the process as a mixture of Markov chains. In the sequel the process $\{X_n\}$ is said to be recurrent if

$$P[X_n = X_0 \ i.o.] = 1. \tag{5.16}$$

To study recurrence we split the problem into two parts. First we require that the process visits infinitely often the set $G \times \{0\}$. Secondly we require, whenever the process hits $G \times \{0\}$, he visits every $(\mathbf{i}, 0) \in G \times \{0\}$ infinitely often. After conditioning on the initial state this shows (5.16).

At first, let us define the sequence of successive hitting times to $G \times \{0\}$ as usual by $\zeta_0 = 0$ and for $n \geq 0$

$$\zeta_{n+1} = \inf\{t > \zeta_n : \exists\ \mathbf{i} \in G\ X_t = (\mathbf{i}, 0)\} \tag{5.17}$$

where $\inf\{\emptyset\} = \infty$. Furthermore let $\{Z_n\}$ be the sequence of states in G at the hitting times, that is $\{Z_n = i\} = \{X_{\zeta_n} = (\mathbf{i}, 0)\}$ or equivalently $\{\zeta_n = l, Z_n = \mathbf{i}\} = \{X_l = (\mathbf{i}, 0)\}$.

Notice that from (5.6)

$$\begin{aligned} P(\zeta_n = l, Z_n = \mathbf{i}|\mathcal{F}_{l-1}) &= P(X_l = (\mathbf{i}, 0)|\mathcal{F}_{l-1}) \\ &= P(\zeta_n = l, |\mathcal{F}_{l-1}) \times P(Z_n = \mathbf{i}|\mathcal{G}_{n-1}) \end{aligned}$$

where

$$P(\zeta_n = l, |\mathcal{F}_{l-1}) = \left[\frac{m_{(X_{l-1})}(0) + t_{l-1}(X_{l-1})(\cdot, 0)}{m_{(X_{l-1})} + t_{l-1}(X_{l-1})}\right] \quad \text{and}$$

$$\begin{aligned} P(Z_n = \mathbf{i}|\mathcal{G}_{n-1}) = &\frac{\tilde{m}(i_1) + \sum_{v=1}^{n-1} I\{Z_v \in \{i_1\} \times (\times_{k=2}^{K} G_k)\}}{\tilde{m} + n} \\ &\times \prod_{k=1}^{K-1} \frac{\tilde{m}_{(i_{1:k})}(i_{k+1}) + \sum_{v=1}^{n-1} I\{Z_v \in \{i_{1:(k+1)}\} \times (\times_{j=k+2}^{K} G_k)\}}{\tilde{m}_{(i_{1:k})} + \sum_{v=1}^{n-1} I\{Z_v \in \{i_{1:k}\} \times (\times_{j=k+1}^{K} G_k)\}} \end{aligned}$$

where $\mathcal{G}_{n-1} = \sigma(\{X_{\zeta_j} = (Z_j, 0)\}_{j=1,\cdots,n-1})$. Which implies the independents of Z_n from ζ_n given $(Z_j)_{j=0,\cdots,n-1}$. We will assume in the following, that

Assumption 3. *(i) There exists an $\epsilon \in (0, 1]$ and a $t_\epsilon \geq 0$ such that*

$$\frac{m_{(\boldsymbol{i},s)}(0)}{m_{(\boldsymbol{i},s)}} > \epsilon \quad \forall(\boldsymbol{i}, s) \in S \text{ such that } s \geq t_\epsilon.$$

(ii) For all $\boldsymbol{j} \in G$ $\tilde{m}(j_1) \prod_{k=1}^{K-1} \tilde{m}_{(j_1,\cdots,j_k)(j_{k+1})} > 0$.

The meaning of both conditions become clear when we look at a typical sample path of the process as plotted in figure 5.2. The process starts in $G \times \{0\}$ and moves along $G \times \{t\}, t \geq 1$ until a 0 ball will be sampled from some urn $U^{(1)}(\mathbf{i}, t)$ for some $(\mathbf{i}, t) \in G \times \mathbb{N}$. Hence the first condition requires that there exists some ϵ risk to return from every state in $G \times \{t\}$ back to $G \times \{0\}$ for all $t \geq t_\epsilon$.

Whereas the second condition requires that the initial urn composition of every return urn contains a positive number of balls of every color. This condition leads a priory to a possibility of sampling any $\{\mathbf{i}, 0\} \in G \times \{0\}$, conditional on the event of returning to $G \times \{0\}$. Both conditions are sufficient to ensure the recurrence (5.16).

Proposition 5. *Suppose the initial urn compositions of $\{X_n\}$ satisfy assumption 3, then $\{X_n\}$ is recurrent.*

Combining both recurrence and partial exchangeability we can apply de Finetti's theorem for Markov chains and obtain a mixture representation of $\{X_n\}$.

Corollary 6. *Suppose that the initial urn compositions satisfy assumption 3, then $\{X_n\}$ is a mixture of Markov chains with unique mixing measure.*

It only remains to determine the mixing measure in the representation (5.15). This measure is most easily studied by looking at the weak limit of the transition counts and the empirical transition probabilities. Since recurrent to the initial state does not imply recurrence for all other states as well, the limits of the empirical transition probabilities may not exists for all transition rows of the process.

In fact, the probability law of the transition matrix Π, conditional on which $\{X_n\}$ is Markov, needs only to be studied for states $(\mathbf{i}, t) \in S$ for which there exists a finite sequence of states, say $\{(\mathbf{i}_k, s_k)\}_{0 \leq k \leq n}$ such that $(\mathbf{i}_n, s_n) = (\mathbf{i}, t)$, which lead with positive probability from the initial state to $(\mathbf{i}, t)$, i.e.

$$\mathbb{P}\Big(\bigcap_{k=0}^{n} \{X_k = (\mathbf{i}_k, s_k)\}\Big) > 0. \tag{5.18}$$

For an ordinary Markov chain, this would be the disjoint and closed class of recurrent states which contains the initial state. All other states are irrelevant, since the process will not visit any of this states with probability one.

We will restrict ourselves to the analysis of the transition rows of the set of recurrent states. For this purpose define for every state $(\mathbf{i}, s) \in S$ the set of potential successor states of $(\mathbf{i}, s)$ by $R_{(\mathbf{i},s)} = R^1_{(\mathbf{i},s)} \cup R^2_{(\mathbf{i},s)}$, where

$$R^1_{(\mathbf{i},s)} = \Big\{(\mathbf{j}, s+1) \in S : m_{(\mathbf{i},s)}(j_1) \prod_{k=1}^{K-1} m_{(\mathbf{i},s|_{j_{1:k}})}(j_{k+1}) > 0\Big\} \tag{5.19}$$

and $R^2_{(\mathbf{i},s)} = \emptyset$ if $m_{(\mathbf{i},s)}(0) = 0$ and $R^2_{(\mathbf{i},s)} = G \times \{0\}$ if $m_{(\mathbf{i},s)}(0) > 0$.

Furthermore define $\mathcal{H}_0 = G \times \{0\}$ and for every $n \geq 1$ let $\mathcal{H}_n = \cup_{(\mathbf{i},s)\in\mathcal{H}_{n-1}} R_{(\mathbf{i},s)}$. Hence $\mathcal{H}_n$ denotes the set of all states in S that are attainable by the process within n transition step and $\mathcal{H} = \cup_n \mathcal{H}_n$ denotes all states which are attainable within a finite number of transition steps. For states in $\mathcal{H}$ we find the mixing measure in the following proposition.

Proposition 6. *Suppose that the initial urn compositions satisfy assumption 3. Then for every $(\boldsymbol{i}, s) \in \mathcal{H}$ the transition row $\Pi_{(\boldsymbol{i},s)} = (\Pi_{(\boldsymbol{i},s),(\boldsymbol{j},v)})_{(\boldsymbol{j},s)\in S}$ of Π, is given by*

$$\Pi_{(\boldsymbol{i},s),(\boldsymbol{j},v)} = \begin{cases} \Lambda_{\boldsymbol{j}}\, p(\boldsymbol{i}, s)_0 & \text{if } (\boldsymbol{j}, v) \in G \times \{0\} \\ p(\boldsymbol{i}, s)_{j_1} \prod_{k=1}^{K-1} p(\boldsymbol{i}, s|_{j_{1:k}})_{j_{k+1}} & \text{if } (\boldsymbol{j}, v) \in G \times \{s+1\} \\ 0 & \text{otherwise} \end{cases}$$

where $\Lambda = (\Lambda_{\boldsymbol{j}})_{\boldsymbol{j}\in G}$ is a distribution on G with $\Lambda_{\boldsymbol{j}} = \lambda_{j_1} \prod_{k=1}^{K-1} \lambda(j_{1:k})_{j_{k+1}}$ for $\boldsymbol{j} \in G$; and

$$\begin{aligned} \lambda &\sim Dir(\tilde{m}(1), \cdots, \tilde{m}(N_1)), \\ p(\boldsymbol{i}, s) &\sim Dir(m_{(\boldsymbol{i},s)}(0), \cdots, m_{(\boldsymbol{i},s)}(N_1)), \\ \lambda(j_{1:k}) &\sim Dir(\tilde{m}_{(j_{1:k})}(1), \cdots, \tilde{m}_{(j_{1:k})}(N_{k+1})) && \text{for } 1 \leq k \leq K-1, \\ p(\boldsymbol{i}, s|j_{1:k}) &\sim Dir(m_{(\boldsymbol{i},s)|j_{1:k})}(1), \cdots, m_{(\boldsymbol{i},s|j_{1:k})}(N_{k+1})) && \text{for } 1 \leq k \leq K-1. \end{aligned}$$

Furthermore the random vectors $\{\lambda, p(\boldsymbol{i}, s), \lambda(j_{1:k}), p(\boldsymbol{i}, s|j_{1:k})\}_{\boldsymbol{j}\in G}$ *are mutually independent.*

Lemma 27 in the proof of the previous proposition shows that $P(X_n = (\mathbf{i}, s)\ i.o.) = 1$ for all $(\mathbf{i}, s) \in \mathcal{H}$. On the other hand, by definition of $\mathcal{H}$, if state $(\mathbf{i}, s)$ does not belong to $\mathcal{H}$ then

$$0 = P(X_n = (\mathbf{i}, s)\ \exists n) = \mathbb{E}[P(X_n = (\mathbf{i}, s)\ \exists n|\Pi)]. \tag{5.20}$$

and therefore $P(X_n = (\mathbf{i}, s)\ \exists n|\Pi) = 0$ with probability one. Furthermore by definition of $\mathcal{H}$ it is easily seen that $\mathcal{H}$ is a closed class and contains the initial state. Therefore

$$1 = P(X_n \in \mathcal{H}\ \forall n) = \mathbb{E}[P(X_n \in \mathcal{H}\ \forall n|\Pi)]. \tag{5.21}$$

Hence there exists a measurable set $N \in \mathcal{F}$ such that $P(X_n \in \mathcal{H}\ \forall n|\Pi(\omega)) = 1$ for all $\omega \in N^c$ and any modification of the probability law of rows of Π for states in $S \setminus \mathcal{H}$ leaves all characteristics of the process unchanged.

Proposition 6 determines the probability law of the rows of the random stochastic matrix Π only for states in $\mathcal{H}$, to ensure that all rows of Π are proper distributions we make the convention that $\Pi_{(\mathbf{i},s),(\mathbf{j},l)} = \delta_{(\mathbf{i},s+1)}((\mathbf{j}, l))$ for all states $(\mathbf{i}, s) \in S \setminus \mathcal{H}$. By the previous remark, this convention leases all characteristics of the mixture representation of the process unchanged. Now, proposition 6 and the previous convention shows that the random matrix Π, conditional on which $\{X_n\}$ is Markov, has the block form

$$\Pi \overset{a.s.}{=} \begin{pmatrix} \Pi_0(0) & \Pi_{-0}(0) & \mathbf{O} & \mathbf{O} & \mathbf{O} & \cdots \\ \Pi_0(1) & \mathbf{O} & \Pi_{-0}(1) & \mathbf{O} & \mathbf{O} & \cdots \\ \Pi_0(2) & \mathbf{O} & \mathbf{O} & \Pi_{-0}(2) & \mathbf{O} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{5.22}$$

where the elements of $\Pi_{-0}(t) = (\Pi_{(\mathbf{i},t),(\mathbf{j},t+1)})_{(\mathbf{i},t),(\mathbf{j},t+1)\in S}$ are given by

$$\Pi_{(\mathbf{i},t)(\mathbf{j},t+1)} = \begin{cases} p(\mathbf{i},t)_{j_1} \prod_{k=1}^{K-1} p(\mathbf{i},t|j_{1:k})_{j_{k+1}} & \text{if } (\mathbf{i},t) \in \mathcal{H} \\ \delta_{(\mathbf{i},t+1)}(\mathbf{j},t+1) & \text{if } (\mathbf{i},t) \in S \setminus \mathcal{H} \end{cases} \tag{5.23}$$

and the elements of $\Pi_0(t) = \{\Pi_{(\mathbf{i},t),(\mathbf{j},0)}\}_{(\mathbf{i},t),(\mathbf{j},0)\in S}$ are given by

$$\Pi_{(\mathbf{i},t)(\mathbf{j},0)} = \begin{cases} \Lambda_{\mathbf{j}}\ p(\mathbf{i},s)_0 & \text{if } (\mathbf{i},s)\mathcal{H} \\ 0 & \text{if } (\mathbf{i},t) \in S \setminus \mathcal{H}. \end{cases} \tag{5.24}$$

Since the elements on the first column of Π share the same random vector Λ, the rows of Π are not independent. But $\Pi_{-0} = \{\Pi_{-0}(t)\}_{t\geq 0}$ consists of independent sub-matrices with independent sub-transition rows.

5.2.2 Properties given a sample of observations

In this section we discuss some properties of the process conditional on the history of the process until a fixed or random point in time. These properties are neat facts for predicting summary measure of Markov chains. Clearly, from the sampling scheme 5.1, it is clear that once we observe the process until a certain point n, $\{X_k\}_{0\leq k\leq n}$, where $X_n = (\mathbf{i},t)$, then the process continues as an urn scheme from $(\mathbf{i},t)$, just with updated urn compositions. Even simpler is the case when the process is recurrent and we condition on the event ζ_n. In this case the process re-starts in $G \times \{0\}$ where the first component in G is draws as in step (A) of the scheme 5.1 with updated initial urn compositions.

The following lemma is an obvious fact, nevertheless we state it

Lemma 17. *Let $v_n = n$ (or ζ_n) and $\{X_k = (\boldsymbol{i}_k, t_k)\}_{k=0,\cdots,v_n}$ be a sample from a recurrent RHUP. Then $\{X_{v_n+k}\}_{k\geq 0}$ is again a RHUP with initial state $(\boldsymbol{i}_n, t_n)$ (or $(\boldsymbol{i}_{\zeta_n}, 0)$ if $v_n = \zeta_n$) and initial urn compositions*

$$\tilde{U}^{(*1)} = \{m^*(i); i \in G_1\}$$
$$\tilde{U}^{(1)}(s) = \{m_s^*(i); i \in G_1 \cup \{0\}\}$$

for $s \in S$ *and for* $\boldsymbol{j} \in G$ *and* $1 \leq k \leq K-1$

$$\tilde{U}^{(k+1)}(s|j_{1:k}) = \{m^*_{(s|j_{1:k})}(i); i \in G_{k+1}\}$$
$$\tilde{U}^{(*(k+1))}(j_{1:k}) = \{\tilde{m}^*_{(j_{1:k})}(i); i \in G_{k+1}\},$$

where for $s = (\boldsymbol{i}, t) \in S$ *and* $\boldsymbol{j} \in G$

$$m^*_{(\boldsymbol{i},t)}(j_1) = m_{(\boldsymbol{i},t)}(j_1) + t_{v(n)}(\boldsymbol{i}, t)(j_1, t+1), \tag{5.25}$$
$$\tilde{m}^*(j_i) = \tilde{m}(j_1) + t^*_{v(n)}(j_1) \tag{5.26}$$

and $k = 1, \cdots, K-1$

$$m^*_{(\boldsymbol{i},t|j_{1:k})}(j_{k+1}) = m_{(\boldsymbol{i},t|j_{1:k})}(j_{k+1}) + t_{v_n}(\boldsymbol{i}, t)(j_{1:(k+1)}, t+1), \tag{5.27}$$
$$\tilde{m}^*_{(j_{1:k})}(j_{k+1}) = \tilde{m}_{(j_{1:k})}(j_{k+1}) + t^*_{v_n}(j_{k+1}). \tag{5.28}$$

The next corollary states that given a sample of v_n observations from a RHUP, the process will be again a mixture of Markov chains. The mixing measure will have the same form as for the original process but with updated parameter. Since the statement holds for $v_n = n$ and similar for $v_n = \zeta_n$ we state the corollary for v_n, where v_n can be taken as n or ζ_n.

Corollary 7. *Suppose the RHUP is recurrent. Then for* $v_n = n$ *(or* $v_n = \zeta_n$*) and given a sample* $\{X_k\}_{0 \leq k \leq v_n}$, $\{X_{v_n+k}\}_{k \geq 0}$ *is again a mixture of Markov chains.*

For $(\boldsymbol{i}, t) \in \mathcal{H}$ *the transition row* $\Pi^{(v_n)}_{(\boldsymbol{i},t)}$ *of the stochastic matrix* $\Pi^{(v_n)}$*, conditional on which* $\{X_{v_n+k}\}_{k \geq 0}$ *is Markov, is given by*

$$\Pi^{(v_n)}_{(\boldsymbol{i},t),(\boldsymbol{j},v)} = \begin{cases} \Lambda^{(v_n)}_{\boldsymbol{j}}\ p^{(v_n)}(\boldsymbol{i}, t)_0 & \text{if } (\boldsymbol{j}, v) \in G \times \{0\} \\ p^{(v_n)}(\boldsymbol{i}, t)_{j_1} \prod_{k=1}^{K-1} p^{(v_n)}(\boldsymbol{i}, t|_{j_{1:k}})_{j_{k+1}} & \text{if } (\boldsymbol{j}, v) \in G \times \{t+1\} \\ 0 & \text{otherwise} \end{cases}$$

with $\Lambda^{(v_n)} = (\Lambda^{(v_n)}_{\boldsymbol{j}})_{\boldsymbol{j} \in G}$ *and* $\Lambda^{(v_n)}_{\boldsymbol{j}} = \lambda^{(v_n)}_{j_1} \prod_{k=1}^{K-1} \lambda^{(v_n)}(j_{1:k})_{j_{k+1}}$ *for* $\boldsymbol{j} \in G$.

Where for $\boldsymbol{j} \in G$

$$\begin{aligned}
\lambda^{(v_n)} &\sim Dir(\tilde{m}^*(1), \cdots, \tilde{m}^*(N_1)), \\
p^{(v_n)}(\boldsymbol{i}, t) &\sim Dir(m^*_{(\boldsymbol{i},t)}(0), \cdots, m^*_{(\boldsymbol{i},t)}(N_1)), \\
\lambda^{(v_n)}(j_{1:k}) &\sim Dir(\tilde{m}^*_{(j_{1:k})}(1), \cdots, \tilde{m}^*_{(j_{1:k})}(N_{k+1})) \qquad \text{for } 1 \le k \le K-1, \\
p^{(v_n)}(\boldsymbol{i}, t|j_{1:k}) &\sim Dir(m^*_{(\boldsymbol{i},t)|j_{1:k})}(1), \cdots, m^*_{(\boldsymbol{i},t|j_{1:k})}(N_{k+1})) \quad \text{for } 1 \le k \le K-1.
\end{aligned}$$

The random vectors $\{\lambda^{(v_n)}, p^{(v_n)}(\boldsymbol{i}, t), \lambda^{(v_n)}(j_{1:k}), p^{(v_n)}(\boldsymbol{i}, t|j_{1:k})\}_{j\in G, k=1,\cdots,K-1}$ *are mutually independent.*

5.3 Mixtures of inhomogeneous multivariate Markov chains

Recall that our aim was to construct an exchangeable sequence of random elements, such that each random element is a mixture of inhomogeneous Markov chain. We will use the RHUP to define such a sequence. For the remaining part we assume that the initial urn compositions of the RHUP satisfies assumption 3. Hence we assume recurrence. We take the sequences of states between successive hits to the set $G \times \{0\}$. That is, for every $n \ge 1$, let $\mathbf{B}_n$ be define by

$$\mathbf{B}_n = (X_{\zeta_{n-1}}, X_{\zeta_{n-1}+1}, \cdots, X_{\zeta_n - 1}) \tag{5.29}$$

where ζ_n was define in (5.17). We may note the following neat fact.

Lemma 18. *The sequence* $\{\boldsymbol{B}_n\}$ *is exchangeable.*

Note that every $\mathbf{B}_n$ is a sequence of elements in $S = G \times \mathbb{N}_0$ such that the last coordinate of every element of $\mathbf{B}_n$ increases deterministically. Hence we will remove the last coordinate from each element $\mathbf{B}_n$. Therefore, let $L : S \to G$, be the projection from $S = G \times \mathbb{N}_0$ onto G, $L((\mathbf{i}, n)) = \mathbf{i}$, and

define the sequence $\{\mathbf{Y}_n\}$ by

$$\begin{aligned}\mathbf{Y}_n := L(\mathbf{B}_n) &= \big(L(X_{\zeta_{n-1}}), L(X_{\zeta_{n-1}+1}), \cdots, L(X_{\zeta_n-1})\big) \\ &= \big(Y_0^{(n)}, Y_1^{(n)}, \cdots, Y_{\tau_n}^{(n)}\big) \end{aligned} \tag{5.30}$$

where $\tau_n = \zeta_n - \zeta_{n-1} - 1$. Clearly $\{\mathbf{Y}_n\}$ is exchangeable too. Since $\{X_n\}$, given Π, is Markov we find that $\{\mathbf{Y}_n\}$, given Π, is Markov as well. The following lemma state that the sequence $\{\mathbf{Y}_n\}_{n\geq 1}$ is an exchangeable sequence of mixtures of inhomogeneous Markov chains. Recall the definition of $\Pi_{-1} = \{\Pi_{-0}(t)\}_{t\geq 0}$ and $\Lambda = \{\Lambda_{\mathbf{j}}\}_{\mathbf{j}\in G}$ from proposition 6 and (5.22).

Lemma 19. *(i) The sequence $\{\boldsymbol{Y}_n\}$ is exchangeable.*
(ii) Given Π_{-0}, $\{p(\boldsymbol{i},t)_0\}_{\boldsymbol{i},t\geq 0}$ and Λ, $\boldsymbol{Y}_n$ is a time-inhomogeneous Markov chain, with initial distribution Λ, transition matrix Π_{-0} and termination probability $p_0 := \{p(\boldsymbol{i},t)_0\}_{\boldsymbol{i},t\geq 0}$, i.e.

$$\mathbb{P}(\bigcap_{t=0}^{k}\{Y_t^{(n)} = \boldsymbol{i}_t\}, \tau_n = k]|\Pi) = \lambda_{\boldsymbol{i}_0}\left(\prod_{t=1}^{k}\Pi_{(\boldsymbol{i}_t,t),(\boldsymbol{i}_{t+1},t)}\right)p(\boldsymbol{i}_m,m)_0 \tag{5.31}$$

Remark 17. *We may clarify the structure of the random transition probability of $\boldsymbol{Y}_n$ as stated in the last lemma. If we denote with $P_\Pi(\cdot) = \mathbb{P}(\cdot|\Pi_{-0},\Lambda,p_0)$, then for $\boldsymbol{i}$ and $\boldsymbol{j}$ in G the last lemma states that for $t \geq 0$*

$$\begin{aligned} &P_\Pi(Y_{t+1}^{(n)} = \boldsymbol{j}|Y_t^{(n)} = \boldsymbol{i}) \\ &\quad = P_\Pi(Y_{t+1,1}^{(n)} = j_1|Y_t^{(n)} = \boldsymbol{i}) \\ &\quad \times P_\Pi(Y_{t+1,2}^{(n)} = j_2|Y_t^{(n)} = \boldsymbol{i}, Y_{t+1,1}^{(n)} = j_1) \\ &\quad \times \cdots \\ &\quad \times P_\Pi(Y_{t+1,K}^{(n)} = j_K|Y_t^{(n)} = \boldsymbol{i}, Y_{t+1,1}^{(n)} = j_1, \cdots, Y_{t+1,K-1}^{(n)} = j_{K-1}), \end{aligned}$$

where $P_\Pi(Y_{t+1,1}^{(n)} = j_1|Y_t^{(n)} = \boldsymbol{i}) = p(\boldsymbol{i},t)_{j_1}$ and for $1 \leq K-1$

$$P_\Pi(Y_{t+1,k+1}^{(n)} = j_{k+1}|Y_t^{(n)} = \boldsymbol{i}, Y_{t+1,1}^{(n)} = j_1, \cdots, Y_{t+1,k}^{(n)} = j_k) = p((\boldsymbol{i},t)|j_{1:k})_{j_{k+1}}.$$

Meaning that the transition probability is factorized into K conditional dis-

tributions, which is the simplest way to express the probability of K events at the same time. But the disadvantage is clearly that the random probability is not invariant under reordering of the K components.

Similar to the previous two chapters we may use the sequence $\{\mathbf{Y}_n\}$ to predict operating characteristics of time-inhomogeneous Markov chains. We may observe M replica of the same underlying phenomena. Information about the phenomena is not obtained by observing the process over a long period, but instead we observe several replica. In the following two examples we look at two rather unusual applications of the sequence $\{\mathbf{Y}_n\}$.

Example 3 (A discrete-time survival model with covariates)**.** *Suppose two health indicators are monitored on a weekly base (or monthly or annually) for M patients. Each health indicator is measured on a binary scale $\{1, 2\}$. 2 indicates the presence and 1 the absence of a health problem. Suppose that the discrete survival status is recorded as well. A common assumption in non-parametric Bayesian regression is that the joined vector of regressor and covariates is exchangeable among patients. This clearly does not imply that the regressors, given the covariates, are exchangeable as well. We choose $G_1 = \{1\}$ and $G_2 = G_3 = \{1, 2\}$. Furthermore the event $\{\tau_n = t\}$ indicates a failure between the weak t and $t + 1$.*

Then in the terminology above, for $(i_1, i_2) \in G_2 \times G_3$, the element Λ_{1,i_1,i_2} denotes the random probability of the initial health conditions. The probability $p(1, i_1, i_2, t)_0$ denote the probability for a death between week t and $t+1$ given the health condition (i_1, i_2) in week t. Furthermore for (i_1, i_2) and (j_1, j_2) in $G_2 \times G_3$ the random element $\Pi_{-0}(t)_{(1,i_1,i_2),(1,j_1,j_2)}$ denote the change in health conditions from week t to $t + 1$.

Suppose we observe the lifetimes of M patients and we know the health condition of a new patient at weak t, say $(i_1, i_2) \in G_2 \times G_3$. The predicted probability of surviving from weak t to weak $t + 1$ for the new patient is now

given as

$$\begin{aligned} P(\tau_{M+1} > t | Y_t^{(M+1)} = (i_1, i_2), \{ \boldsymbol{Y}_n \}_{n \leq M}) \\ &= \mathbb{E}[p(1, i_1, i_2, t)_1 | \{ \boldsymbol{Y}_n \}_{n \leq M})] \\ &= \frac{m_{(i_1, i_2, t)}(1) + t_{v(M)}(i_1, i_2, t)(1, t+1)}{m_{(\boldsymbol{i}, t)} + t_{v(M)}(i_1, i_2, t)}. \end{aligned}$$

Similarly the probability of a transition from health state $(i_1, i_2) \in G_2 \times G_3$ *to state* $(j_1, j_2) \in G_2 \times G_3$*, under survival, is given by*

$$\begin{aligned} &P(\tau_{M+1} > t, Y_t^{(M+1)} = (j_1, j_2) | Y_t^{(M+1)} = (i_1, i_2), \{ \boldsymbol{Y}_n \}_{n \leq M}) \\ &= \frac{m_{(i_1, i_2, t)}(1) + t_{v(M)}(i_1, i_2, t)(1, t+1)}{m_{(\boldsymbol{i}, t)} + t_{v(M)}(i_1, i_2, t)} \\ &\times \frac{m_{(i_1, i_2, t)}(1) + t_{v(M)}(i_1, i_2, t)(1, j_1, t+1)}{m_{(\boldsymbol{i}, t)} + t_{v(M)}(i_1, i_2, t)(1, t+1)} \\ &\times \frac{m_{(i_1, i_2, t)}(1) + t_{v(M)}(i_1, i_2, t)(1, j_1, j_2, t+1)}{m_{(\boldsymbol{i}, t)} + t_{v(M)}(i_1, i_2, t)(1, j_1, t+1)}. \end{aligned}$$

Predictive probabilities of other quantities, can be found using the generic estimator derived in the next section.

The incorporation of some covariates into the model, as shown in the last example, should be considers with caution. If covariates are constant over time like the sex of patients we basically have two independent sub-processes in the model. Hence estimation will be inefficient, since no transition from one to the other sex is possible. Also, the more covariates we include into the model, say $K-1$, the less transition counts can be observed among fixed states in $G_2 \times G_3 \times \cdots \times G_K$ and hence the less efficient and reliable is the estimator of the survival function given the covariates.

Example 4. *Suppose a system consists of* K *interacting components, where the state of each component changes on a daily basis dependent on daily demand. The demand may be different dependent on the day of the week* $1, \cdots, 5$*. Suppose that component* $1 \leq k \leq K$ *of the system runs in* $\{1, \cdots, N_{k+1}\}$ *different states. We take* $G_1 = \{1\}$ *and* $G_k = \{1, \cdots, N_k\}$ *for* $2 \leq k \leq K+1$*. Now, let* $\boldsymbol{Y}_n = (Y_0^n, \cdots, Y_4^n)$ *represent the working history of the system in*

week n. Since each block terminates deterministically every Friday we set $m_{(1,\boldsymbol{j},t)}(0) = 0$ *for* $t = 0, \cdots 4$ *and* $m_{(1,\boldsymbol{j},4)}(1) = 0$ *for all* $\boldsymbol{j}$ *in* G*. This system does not satisfy assumption 3(i) but it is not difficult to see that the resulting urn process is recurrent.*

In the following section we derive some general predictive estimator for summary measures of inhomogeneous Markov chains based on the sequence $\{\mathbf{Y}_n\}$.

5.3.1 Prediction for multivariate Markov chains

In this section we will use the sequence $\{\mathbf{Y}_n\}$ and the predictive distribution of $\mathbf{Y}_{M+1}$ given $\{\mathbf{Y}_n\}_{n\leq M}$ for prediction of summary measure of multivariate time-inhomogeneous Markov chains.

In many situations we may observe several replicas of a stochastic phenomenon which evolves over time. In system reliability this might be the history of a system, which contains K interacting components, between successive shut-downs. In event history analysis we may observed lifetimes of persons, where persons with the same initial conditions are exchangeable. Or in clinical trials we may monitor several different health indicators of a patient simultaneously until a random or predefined endpoint.

We may model the phenomena by a trajectory $\mathbf{Y}_m$, where for each $1 \leq k \leq K$ we code G_k to represent the set of possible states taken at the k-th dimension. For example in reliability G_k may represent the distinct working states of the K-th components of the system or in health monitoring G_k may represent the different states of the k-th health status.

We start with the predicted transition probabilities.

Lemma 20. *Let* $\{\boldsymbol{Y}_n\}_{n\leq M}$ *be a sample from the process defined in (5.30). Then, given* $\{\boldsymbol{Y}_n\}_{n\leq M}$*,*
(i) the expected transition probability from $\boldsymbol{j} \in G$ *to* $\boldsymbol{i} \in G$ *at time* t *is given by*

$$\mathbb{E}[P_{\Pi}(Y_{t+1}^{(M+1)} = \boldsymbol{i} | Y_t^{(M+1)} = \boldsymbol{j}) | \{\boldsymbol{Y}_n\}_{n\leq M}] = \frac{m^*_{(\boldsymbol{j},t)}(i_1)}{m^*_{(\boldsymbol{j},t)}} \prod_{k=1}^{K-1} \frac{m^*_{(\boldsymbol{j},t|i_{1:k})}(i_{k+1})}{m^*_{(\boldsymbol{j},t|i_{1:k})}};$$

(ii) the expected transition probability at time t, at the d-component from $\boldsymbol{j} \in G$ to $v \in G_d$ is given by

$$\mathbb{E}[P_\Pi(Y_{t+1,d}^{(M+1)} = v | Y_t^{(M+1)} = \boldsymbol{j}) | \{\boldsymbol{Y}_n\}_{n\le M}] = \sum_{i_k \in G_k, k \le d-1} \frac{m^*_{(\boldsymbol{j},t)}(i_1)}{m^*_{(\boldsymbol{j},t)}} \left[\prod_{k=1}^{d-2} \frac{m^*_{(\boldsymbol{j},t|i_{1:k})}(i_{k+1})}{m^*_{(\boldsymbol{j},t|i_{1:k})}}\right] \frac{m^*_{(\boldsymbol{j},t|i_{1:(d-1)})}(v)}{m^*_{(\boldsymbol{j},t|i_{1:(d-1)})}};$$

(iii) and the expected transition probability from $\boldsymbol{j} \in G$ to $v \in G_d$ and $w \in G_{d'}$ $(1 \le d < d' \le K)$ at time t is given by

$$\begin{aligned}\mathbb{E}[P_\Pi(Y_{t+1,d}^{(M+1)} = v, Y_{t+1,d'}^{(M+1)} = w | Y_t^{(M+1)} = \boldsymbol{j}) | \{\boldsymbol{Y}_n\}_{n\le M}] = \sum_{\substack{i_k \in G_k;\\ k \notin \{d,d'\},\\ k \le d'-1}} \Bigg[& \frac{m^*_{(\boldsymbol{j},t)}(i_1)}{m^*_{(\boldsymbol{j},t)}} \prod_{k=1}^{d-2} \frac{m^*_{(\boldsymbol{j},t|i_{1:k})}(i_{k+1})}{m^*_{(\boldsymbol{j},t|i_{1:k})}} \prod_{k=d+1}^{d'-2} \frac{m^*_{(\boldsymbol{j},t|i_{1:(d-1)},v,i_{(v+1):k})}(i_{k+1})}{m^*_{(\boldsymbol{j},t|i_{1:(d-1)},v,i_{(v+1):k})}} \\ & \frac{m^*_{(\boldsymbol{j},t|i_{1:(d-1)})}(v)}{m^*_{(\boldsymbol{j},t|i_{1:(d-1)})}} \frac{m^*_{(\boldsymbol{j},t|i_{1:(d-1)},v)}(i_{d+1})}{m^*_{(\boldsymbol{j},t|i_{1:(d-1)},v)}} \frac{m^*_{(\boldsymbol{j},t|i_{1:(d-1)},v,i_{(d+1):(d'-1)})}(w)}{m^*_{(\boldsymbol{j},t|i_{1:(d-1)},v,i_{(d+1):(d'-1)})}} \Bigg].\end{aligned}$$

In the next lemma we predict the probability of staying in a certain state at time t for the complete vector of states and a single component of the vector. To keep notations simple, for $t \ge 0$ let $\hat{\Pi}_{-0}^{(M)}(t) = \mathbb{E}[\Pi_{-0}(t) | \mathbf{Y}_j, j \le M]$ and $\hat{p}_0^{(M)}(t) = \mathbb{E}[(p(\mathbf{i},t)_0)_{\mathbf{i}\in G} | \mathbf{Y}_j, j \le M]'$ with elements

$$\hat{\Pi}_{-0}^{(M)}(t)_{\mathbf{i},\mathbf{j}} = \frac{m^*_{(\mathbf{i},t)}(j_1)}{m^*_{(\mathbf{i},t)}} \prod_{k=1}^{K-1} \frac{m^*_{(\mathbf{i},t|j_{1:k})}(j_1)}{m^*_{(\mathbf{i},t|j_{1:k})}} \qquad \text{for } \mathbf{i},\mathbf{j} \in G,$$

$$\hat{p}_0^{(M)}(t)_{\mathbf{i}} = \frac{m^*_{(\mathbf{i},t)}(0)}{m^*_{(\mathbf{i},t)}} \qquad \text{for } \mathbf{i} \in G.$$

and let $\hat{\Lambda}^{(M)} = \mathbb{E}[\Lambda | \mathbf{Y}_j, j \le M] = \left(\frac{\tilde{m}^*(i_1)}{\tilde{m}^*} \prod_{k=1}^{K-1} \frac{\tilde{m}^*_{(i_{1:k})}(i_{k+1})}{\tilde{m}^*_{(i_{1:k})}}; \mathbf{i} \in G\right)$.

Lemma 21. *Let $\{\boldsymbol{Y}_n\}_{n\le M}$ be a sample from the process defined in (5.30). Then, given $\{\boldsymbol{Y}_n\}_{n\le M}$,*

(i) the expected l-step transition probability into state $\boldsymbol{i} \in G$ *is given by*

$$\mathbb{E}[P_{\Pi}(Y_l^{(M+1)} = \boldsymbol{i})|\{\boldsymbol{Y}_n\}_{n\leq M}] = \left[\hat{\Lambda}^{(M)} \prod_{t=1}^{l-1} \hat{\Pi}_{-0}^{(M)}(t)\right]_{\boldsymbol{i}},$$

(ii) the expected l-step transition probability of the d-th component into $v \in G_d$ *given by*

$$\mathbb{E}[P_{\Pi}(Y_{l,d}^{(M+1)} = v)|\{\boldsymbol{Y}_n\}_{n\leq M}] = \sum_{\boldsymbol{i}\in G: i_d = v} \left[\hat{\Lambda}^{(M)} \prod_{t=1}^{l-1} \hat{\Pi}_{-0}^{(M)}(t)\right]_{\boldsymbol{i}}.$$

In certain cases we need to predict the survival time of a process and the expected occupation time in a given state, say $T(\mathbf{Y}_n)_{\mathbf{i}} = \sum_{t\geq 0} I(Y_t^{(n)} = \mathbf{i})$ and $T(\mathbf{Y}_n)_{d,i_d} = \sum_{t\geq 0} I(Y_{t,d}^{(n)} = i_d)$. The predictive estimator for both problems are given in the next lemma.

Lemma 22. *Let* $\{\boldsymbol{Y}_n\}_{n\leq M}$ *be a sample from the process defined in (5.30). Then, given* $\{\boldsymbol{Y}_n\}_{n\leq M}$,
(i) the expected survival function, survival time and failure rate is given by

$$\mathbb{E}[P_{\Pi}(\tau_m > l)|\{\boldsymbol{Y}_n\}_{n\leq M}] = \left[\hat{\Lambda}^{(M)} \prod_{t=1}^{l-1} \hat{\Pi}_{-0}^{(M)}(t)\right] 1_G$$

$$\mathbb{E}[\tau_m|\{\boldsymbol{Y}_n\}_{n\leq M}] = \sum_{l\geq 0} \left[\hat{\Lambda}^{(M)} \prod_{t=1}^{l-1} \hat{\Pi}_{-0}^{(M)}(t)\right] 1_G$$

$$\mathbb{E}[P_{\Pi}(\tau_M = l)|\{\boldsymbol{Y}_n\}_{n\leq M}] = \left[\hat{\Lambda}^{(M)} \prod_{t=1}^{l-1} \hat{\Pi}_{-0}^{(M)}(t)\right] \hat{p}_0^{(M)}(l)$$

where $I_G = (1, 1, \cdots, 1)'$ *of dimension* $|G| = \prod_{i=1}^{K} N_i$.
(ii) the expected mean occupation time in state $\boldsymbol{i} \in G$ *and sub-state* $v \in G_d$

is given by

$$\mathbb{E}[T(\boldsymbol{Y}_{M+1})_{\boldsymbol{i}}|\{\boldsymbol{Y}_n\}_{n\leq M}] = \sum_{l\geq 0}\left[\hat{\Lambda}^{(M)}\prod_{t=1}^{l-1}\hat{\Pi}_{-0}^{(M)}(t)\right]_{\boldsymbol{i}},$$

$$\mathbb{E}[T(\boldsymbol{Y}_{M+1})_{d,v}|\{\boldsymbol{Y}_n\}_{n\leq M}] = \sum_{\substack{l\geq 0,\\ \boldsymbol{i}\in G: i_d=v}}\left(\hat{\Lambda}^{(M)}\prod_{t=1}^{l-1}\hat{\Pi}_{-0}^{(M)}(t)\right)_{\boldsymbol{i}}.$$

Finally, we give an expression of the covariance between the d' and the d components of $\mathbf{Y}_{M+1}$ at time t given the observed data $\{\mathbf{Y}_n\}_{n\leq M}$. Please notice that for categorical data, the covariance might not be an appropriate measure of dependent. Nonetheless for some ordinary data structure it might make scene to consider covariance estimation.

Lemma 23. *Let $\{\boldsymbol{Y}_n\}_{n\leq M}$ be a sample from the process defined in (5.30). Then, given $\{\boldsymbol{Y}_n\}_{n\leq M}$,*
(i) the covariance between the d-th and the d′-th component at time $t+1$ given the value at time t is given by

$$\begin{aligned}
&Cov(Y_{t+1,d}^{(M+1)}, Y_{t+1,d'}^{(M+1))}|Y_t^{(M+1)}, \{\boldsymbol{Y}_n\}_{n\leq M}) \\
&= \sum_{\substack{i_k\in G_k;\\ k\leq d'}} i_d i_{d'}\left[\frac{m^*_{(Y_t,t)}(i_1)}{m^*_{(Y_t,t)}}\prod_{k=1}^{d'-1}\frac{m^*_{(Y_t,t|i_{1:k})}(i_{k+1})}{m^*_{(Y_t,t|i_{1:k})}}\right] \\
&\quad - \left[\sum_{i_k\in G_k, k\leq d} i_d\frac{m^*_{(\boldsymbol{j},t)}(i_1)}{m^*_{(\boldsymbol{j},t)}}\left[\prod_{k=1}^{d-1}\frac{m^*_{(Y_t,t|i_{1:k})}(i_{k+1})}{m^*_{(Y_t,t|i_{1:k})}}\right]\right] \\
&\quad \times\left[\sum_{i_k\in G_k, k\leq d'} i_d\frac{m^*_{(Y_t,t)}(i_1)}{m^*_{(\boldsymbol{j},t)}}\left[\prod_{k=1}^{d'-1}\frac{m^*_{(Y_t,t|i_{1:k})}(i_{k+1})}{m^*_{(Y_t,t|i_{1:k})}}\right]\right]
\end{aligned}$$

(ii) the covariance between the d^{th} and d'^{th} component process at time $t+1$

is given by

$$
\begin{aligned}
&Cov(Y_{t+1,d}^{(M+1)}, Y_{t+1,d'}^{(M+1))} | \{ \boldsymbol{Y}_n \}_{n \leq M}) \\
&\quad = \sum_{\boldsymbol{i} \in G} i_d i_{d'} \Big(\hat{\Lambda}^{(M)} \prod_{l=1}^{t-1} \hat{\Pi}_{-0}^{(M)}(l) \Big)_{\boldsymbol{i}} \\
&\quad - \left[\sum_{\boldsymbol{i} \in G} i_d \Big(\hat{\Lambda}^{(M)} \prod_{l=1}^{t-1} \hat{\Pi}_{-0}^{(M)}(l) \Big)_{\boldsymbol{i}} \right] \left[\sum_{\boldsymbol{i} \in G} i_{d'} \Big(\hat{\Lambda}^{(M)} \prod_{l=1}^{t-1} \hat{\Pi}_{-0}^{(M)}(l) \Big)_{\boldsymbol{i}} \right]
\end{aligned}
$$

From the expressions above it is unfortunately not obvious how $Y_{t+1,d}^{(M+1)}$ and $Y_{t+1,d'}^{(M+1))}$ are dependent on each other. Furthermore for large M the elements in transition sub-matrices $\hat{\Pi}_{-0}^{(M)}$ are dominated by the empirical transition counts. Hence for large M the covariance will largely agree with the empirical covariance between components.

5.4 Conclusion

In the present paper we used a sequence of hierarchical Pólya urns to construct an exchangeable sequence of random elements $\{\mathbf{Y}_n\}$, where each element $\mathbf{Y}_n$ is a mixture of inhomogeneous Markov chains on a finite multivariate state space. The sequence $\{\mathbf{Y}_n\}$ and the predictive distribution of $\mathbf{Y}_{M+1}$ given $\{\mathbf{Y}_n\}_{n \leq M}$ was used to predict summary measure for inhomogeneous Markov chains. In this way the process may serve as a black box model for prediction in discrete event history models or in system reliability. In such models information are usually collected through observing the same underlying phenomena for M observation, where one may assumes that observations and covariates are jointly exchangeable.

Future extensions will mainly focus on modelling continuous time multistate processes though a process with reinforcement. But implementation of such a problem becomes substantially more cumbersome. Such problems may be modelled with mixture of Semi-Markov chains and a focus towards the matrix of successor states and holding time for every state of the state-space [51].

5.5 Appendix: Proofs

Proof of Proposition 3. By definition, if $(\mathbf{i}^j, s_j)_{0\le j\le n}$ is inadmissible then by (5.5) and (5.6), $\bigcap_{j=0}^{n}\{X_j = (\mathbf{i}^j, s_j)\}$ is clearly a null event. On the other hand if $(\mathbf{i}^j, s_j)_{0\le j\le n}$ is admissible

$$
\begin{aligned}
&P[X_0 = (\mathbf{i}^0, s_0), X_1 = (\mathbf{i}^1, s_1), \cdots, X_n = (\mathbf{i}_n, s_n)]\\
&= \prod_{u=0}^{n-1}\left[\frac{\tilde m(i_1^u) + t^*_{u-1}(i_1^u)}{\tilde m + t^*_{u-1}}\prod_{k=1}^{K-1}\frac{\tilde m_{(i^u_{1:k})}(i^u_{k+1}) + t^*_{u-1}(i^u_{1:k+1})}{\tilde m_{(i^u_{1:k})} + t^*_{u-1}(i^u_{1:k})}\right]^{I\{s_u=0\}}\\
&\times\prod_{j=1}^{n-1}\Bigg(\left[m_{(\mathbf{i}^{(j)},s_j)}(i_1^{j+1}) + t_j(\mathbf{i}^{(j)}, s_j)(i_1^{j+1}, s_j+1)\right]^{I_{\{s_{j+1}=s_j+1\}}}\\
&\qquad\times\frac{\Big(m_{(\mathbf{i}^{(j)},s_j)}(0) + t_j(\mathbf{i}^{(j)}, s_j)(\cdot,0)\Big)^{I_{\{s_{j+1}=0\}}}}{m_{(\mathbf{i}^{(j)},s_j)} + t_j(\mathbf{i}^{(j)}, s_j)}\Bigg)\\
&\quad\times\left[\prod_{k=1}^{K-1}\frac{m_{(\mathbf{i}^{(j)},s_j|i^{j+1}_{1:k})}(i^{j+1}_{k+1}) + t_j(\mathbf{i}^{(j)}, s_j)(i^{j+1}_{1:k+1}, s_j+1)}{m_{(\mathbf{i}^{(j)},s_j|i^{j+1}_{1:k})} + t_j(\mathbf{i}^{(j)}, s_j)(i^{j+1}_{1:k}, s_j+1)}\right]^{I\{s_{j+1}=s_j+1\}}
\end{aligned}
\tag{5.32}
$$

Among $\{(\mathbf{i}_j, s_j); 0\le j\le n\}$ are $t_n^*\ge 1$ elements which belong to the subset $G\times\{0\}$. We re-label this elements as $(\mathbf{v}_u, 0)_{1\le u\le t_n^*}$ and rewrite the first product in (5.32) as

$$
\begin{aligned}
&\frac{\tilde m(v_1^1))}{\tilde m}\prod_{k=1}^{K-1}\frac{\tilde m_{(v^1_{1:k})}(v^1_{k+1})}{\tilde m_{(i^1_{1:k})}}\\
&\times\frac{\tilde m(v_1^2) + I\{v_1^1 = v_1^2\}}{\tilde m+1}\prod_{k=1}^{K-1}\frac{\tilde m_{(v^2_{1:k})}(v^2_{k+1}) + I\{v^1_{1:k+1} = v^2_{1:k+1}\}}{\tilde m_{(i^2_{1:k})} + I\{v^1_{1:k} = v^2_{1:k}\}}\\
&\times\cdots\\
&\times\frac{\tilde m\big(v_1^{t_n^*}\big) + \sum_{j=1}^{t_n^*-1} I\{v_1^j = v_1^{t_n^*}\}}{\tilde m + t_n^* - 1}\prod_{k=1}^{K-1}\frac{\tilde m_{(v^{t_n^*}_{1:k})}\big(v^{t_n^*}_{k+1}\big) + \sum_{j=1}^{t_n^*-1} I\{v^j_{1:k+1} = v^{t_n^*}_{1:k+1}\}}{\tilde m_{(i^{t_n^*}_{1:k})} + \sum_{j=1}^{t_n^*-1} I\{v^j_{1:k} = v^{t_n^*}_{1:k}\}}\\
&= \prod_{\mathbf{v}\in G}\left(\frac{\tilde m(v_1)^{[t_n^*(v_1)]}}{\tilde m^{[t_n^*]}}\prod_{k=1}^{K-1}\frac{\tilde m_{(v_{1:k})}(v_{k+1})^{[t_n^*(v_{1:k+1})]}}{\tilde m^{[t_n^*(v_{1:k})]}_{(v_{1:k})}}\right).
\end{aligned}
\tag{5.33}
$$

Furthermore let $(\mathbf{i}, s)$ be an arbitrary state in S among $(\mathbf{i}_j, s_j)_{0\leq j\leq n}$. Let $(\mathbf{i}, s)$ have successor states $(\mathbf{v}^j, l_j)_{1\leq j\leq t_n(\mathbf{i},s)}$. By definition $l_j \in \{0, s+1\}$ for $j = 1, \cdots, t_n(\mathbf{i}, s)$. The product component associated to successor of $(\mathbf{i}, s)$ in the second product of (5.32) can be written as the product of $t_n(\mathbf{i}, s)$ terms each standing for one successor. Successors with $l_j = 0$ are only involved in this product by the draw of a zero colors from the first level urn $U^{(1)}(\mathbf{i}, s)$ (the remaining part was already accounted for in (5.33). The first terns is

$$\frac{m_{(\mathbf{i},s)}(v_1^1)}{m_{(\mathbf{i},s)}}\left(\prod_{k=1}^{K-1}\frac{m_{(\mathbf{i},s|v_{1:k}^1)}(v_{k+1}^1)}{m_{(\mathbf{i},s|v_{1:k}^1)}}\right)^{I_{\{l_1=s+1\}}} \tag{5.34}$$

and the second term is

$$\frac{[m_{(\mathbf{i},s)}(0)+I\{l_1=0\}]^{I_{\{l_2=0\}}}[m_{(\mathbf{i},s)}(v_1^2)+I\{v_1^1=v_1^2,l_1=s+1\}]^{I_{\{l_2=s+1\}}}}{m_{(\mathbf{i},s)}+1}$$
$$\left(\prod_{k=1}^{K-1}\frac{m_{(\mathbf{i},s|v_{1:k}^2)}(v_{k+1}^2)+I\{v_{1:k+1}^1=v_{1:k+1}^2,l_1=s+1\}}{m_{(\mathbf{i},s|v_{1:k}^1)}+I\{v_{1:k}^1=v_{1:k}^2,l_1=s+1\}}\right)^{I_{\{l_2=s+1\}}}. \tag{5.35}$$

And the $t_n(\mathbf{i}, s)$-th term is given by

$$\left[m_{(\mathbf{i},s)}(0)+\sum_{j=1}^{t_n(\mathbf{i},s)-1}I\{l_j=0\}\right]^{I_{\{l_{t_n(\mathbf{i},s)}=0\}}}$$
$$\times\frac{[m_{(\mathbf{i},s)}(v_1^{t_n(\mathbf{i},s)})+\sum_{j=1}^{t_n(\mathbf{i},s)-1}I\{v_1^j=v_1^{t_n(\mathbf{i},s)},l_j=s+1\}]^{I_{\{l_{t_n(\mathbf{i},s)}=s+1\}}}}{m_{(\mathbf{i},s)}+t_n(\mathbf{i},s)-1}$$
$$\times\left(\prod_{k=1}^{K-1}\frac{m_{(\mathbf{i},s|v_{1:k}^{t_n(\mathbf{i},s)})}(v_{k+1}^{t_n(\mathbf{i},s)})+\sum_{j=1}^{t_n(\mathbf{i},s)-1}I\{v_{1:k+1}^j=v_{1:k+1}^{t_n(\mathbf{i},s)},l_j=s+1\}}{m_{(\mathbf{i},s|v_{1:k}^{t_n(\mathbf{i},s)})}+\sum_{j=1}^{t_n(\mathbf{i},s)-1}I\{v_{1:k}^j=v_{1:k}^{t_n(\mathbf{i},s)},l_j=s+1\}}\right)^{I_{\{l_{t_n(\mathbf{i},s)}=s+1\}}}. \tag{5.36}$$

Combining all products, we can write the likelihood associated to successors of $(\mathbf{i}, s)$ as

$$= \prod_{j \in G} \frac{m_{(\mathbf{i},s)}(0)^{[t_n(\mathbf{i},s)(\cdot,0)]} m_{(\mathbf{i},s)}(j_1)^{[t_n(\mathbf{i},s)(j_1,s+1)]}}{m_{(\mathbf{i},s)}^{[t_n(\mathbf{i},s)]}} \prod_{k=1}^{K-1} \frac{m_{(\mathbf{i},s|j_{1:k})}(j_{k+1})^{[t_n(\mathbf{i},s)(j_{1:k+1})]}}{m_{(\mathbf{i},s|i_{1:k})}^{[t_n(\mathbf{i},s)(j_{1:k+1})]}} \tag{5.37}$$

Now take the product of (5.37) over all states in S and combine the product with (5.33) yields the claim. □

Proof of Corollary 5. It is straight forward to show that the system of finite dimensional distributions defined in (5.14) satisfies Kolmogorov's consistency conditions.

Alternative one may check that the sequence of transition probabilities defined in (5.5) and (5.6) satisfies the requirements of the Ionescu Tulcea extension theorem. Which yields the existence of a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process $\{X_n\}_{n\geq 0}$ having (5.5) and (5.6) as transition probabilities. But by the previous proposition this process has finite dimensional distribution as specifies in (5.14). □

Proof of Proposition 4. Let $\mu = \big((\mathbf{i}^j, s_j); 0 \leq j \leq n\big)$ and $\sigma = \big((\mathbf{v}^j, w_j); 0 \leq j \leq n\big)$ be two equivalent sequences of elements in S. Denote for all states in S with $t_n(\mathbf{i}, s)(\mathbf{i}^*, s^*)_\mu$ the number of transitions from $(\mathbf{i}, s)$ to $(\mathbf{i}^*, s^*)$ in μ and define $t_n(\mathbf{i}, s)(\mathbf{i}^*, s^*)_\sigma$ similar for σ. By equivalence between μ and σ we have

$$(\mathbf{i}^0, s_0) = (\mathbf{v}^0, w_0), \text{ and } t_n(\mathbf{i}, s)(\mathbf{i}^*, s^*)_\mu = t_n(\mathbf{i}, s)(\mathbf{i}^*, s^*)_\sigma \ \ \forall (\mathbf{i}, s), (\mathbf{i}^*, s^*) \in S.$$

Suppose that μ is inadmissible, than either $s_0 \neq 0$ or there exists $(\mathbf{i}^j, s_j)$ among $(\mathbf{i}^j)_{1\leq j\leq n}$ such that $s_j \notin \{0, s_{j-1}\}$ and $1 \leq t_n(\mathbf{i}^{j-1}, s_{j-1})(\mathbf{i}^j, s_j)_\mu = t_n(\mathbf{i}^{j-1}, s_j)(\mathbf{i}^j, s_j)_\sigma$. Which implies that σ is inadmissible too and therefore

$$P[\cap_{t=0}^n \{X_t = (\mathbf{i}^t, s_t)\}] = 0 = P[\cap_{t=0}^n \{X_t = (\mathbf{v}^t, w_t)\}].$$

On the other hand if μ is admissible and hence σ too, then by equivalence

of σ an μ for any $(\mathbf{i}, s)$ and $(\mathbf{j}, v)$ in S also

$$\begin{aligned} t_\mu(\mathbf{i}, s) &= t_\sigma(\mathbf{i}, s), \\ t_\mu(\mathbf{i}, s)(j_{1:k}, v) &= t_\sigma(\mathbf{i}, s)(j_{1:k}, v), \\ t_\mu(\mathbf{i}, s)(\cdot, 0) &= t_\sigma(\mathbf{i}, s)(\cdot, 0) \end{aligned}$$

and furthermore

$$\begin{aligned} t^*_\mu(\mathbf{i}) &= I\{\mathbf{i}^0 = \mathbf{i}\} + \sum_{(\mathbf{j},s)\in S} t_\mu(\mathbf{j}, s)(\mathbf{i}, 0) = t^*_\sigma(\mathbf{i}), \\ t^*_\mu(i_{1:k}) &= t^*_\sigma(i_{1:k}), \\ t^*_\mu &= t^*_\sigma. \end{aligned}$$

Since the finite dimensional law of the process $\{X_n\}$ depends only on the six transition counts above which are all equal for μ and σ both sequences have the same probability. □

In order to prove Proposition 5, we need the following two lemmas.

Lemma 24. *Suppose assumption 3 holds, then* $P(\cap_n\{\zeta_n < \infty\}) = 1$.

Proof. We proof the lemma by induction on n. Without loss of generality we may assume that $t_\epsilon = 1$. For ζ_1 we have

$$\begin{aligned} &P(\zeta_1 > t) = P(\cap_{n=0}^t\{X_n \in G \times \{n\}\}) \\ &= \sum_{\substack{\mathbf{i}^n\in G,\\ 0\le n\le t}} \frac{\tilde m(i_1^{(0)})}{\tilde m} \prod_{k=1}^{K-1} \frac{\tilde m_{(i^0_{1:k})}(i^{(0)}_{k+1})}{\tilde m_{(i^{(0)}_{1:k})}} \prod_{n=0}^{t-1} \frac{m_{(\mathbf{i}^{(n)},n)}(i_1^{(n+1)})}{m_{(\mathbf{i}^{(n)},n)}} \prod_{k=1}^{K-1} \frac{m_{(\mathbf{i}^{(n)},n|i^{(n+1)}_{1:k})}(i^{(n+1)}_{k+1})}{m_{(\mathbf{i}^{(n)},n|i^{(n+1)}_{1:k})}} \\ &\le (1-\epsilon)^t \sum_{\substack{\mathbf{i}^n\in G,\\ n=0,\cdots,t-1\\ i_1^t\in G_1}} \frac{\tilde m(i_1^{(0)})}{\tilde m} \prod_{k=1}^{K-1} \frac{\tilde m_{(i^0_{1:k})}(i^{(0)}_{k+1})}{\tilde m_{(i^{(0)}_{1:n})}} \prod_{n=0}^{t-2}\prod_{k=1}^{K-1} \frac{m_{(\mathbf{i}^{(n)},n|i^{(n+1)}_{1:k})}(i^{(n+1)}_{k+1})}{m_{(\mathbf{i}^{(n)},n|i^{(n+1)}_{1:k})}} \\ &= (1-\epsilon)^t \times (N_1 t) \sum_{\mathbf{i}^0\in G} \left(\frac{\tilde m(i_1^{(0)})}{\tilde m} \prod_{k=1}^{K-1} \frac{\tilde m_{(i^0_{1:k})}(i^{(0)}_{k+1})}{\tilde m_{(i^{(0)}_{1:n})}} \right) \\ &= \exp\{-\not{\beta}\} \times N_{\ 1}t = o(t) \quad \text{as } t \to +\infty \end{aligned}$$

where $\beta = -\log(1-\epsilon) \in [-\infty, 0)$. Hence $P(\zeta_1 < \infty) = 1 - \lim_{t\to\infty} P(\zeta_1 > t) = 0$.

For the induction step, assume that $P(\cap_{j=1}^{n}\{\zeta_j < \infty\}) = 1$ for $n > 1$. Let $T = 1 + \max_{1\le j\le n}\{\zeta_j - \zeta_{j-1}\}$ which is finite with probability one. For $s \ge T$ all $t_{\zeta_n}[(\mathbf{i}, s); (\mathbf{j}, v)]$ are zero for all states $(\mathbf{i}, s)$ and $(\mathbf{j}, v)$ in S.

Now, let $P_v^*(\cdot) = P(\cdot|\cap_{j=1}^{n}\{\zeta_j < \infty\}, T, \zeta_n = v)$. Then for $t > T$ we have that

$$
\begin{aligned}
&P_v^*(\zeta_{n+1} > t)\\
&= P(\cap_{l=0}^{t}\{X_{\zeta_n+l} \in G \times \{l\}\}|\{\zeta_j < \infty\}, \zeta_n = v, T)\\
&\le P^*(\cap_{l=T}^{t}\{X_{v+l} \in G \times \{l\}\}|\{\zeta_j < \infty\}, \zeta_n = v, T)\\
&= \sum_{\substack{\mathbf{i}^{T+n}\in G,\\ 0\le n\le t-T}} \prod_{n=0}^{t-T-1}\left(\frac{m_{(\mathbf{i}^{(T+n)},T+n)}(i_1^{(T+n+1)})}{m_{(\mathbf{i}^{(T+n)},T+n)}} \prod_{k=1}^{K-1}\frac{m_{(\mathbf{i}^{(T+n)},T+n|i_{1:k}^{(T+n+1)})}(i_{k+1}^{(T+n+1)})}{m_{(\mathbf{i}^{(T+n)},T+n|i_{1:k}^{(T+n+1)})}}\right)\\
&\le (1-\epsilon)^{(t-T)}N_1(t-T) = o(t) \ \text{ as } t \to +\infty
\end{aligned}
$$

Since by assumption $P(\zeta_n < \infty) = 1$, for every $\delta > 0$ we can find a $k \in \mathbb{N}$ such that $P(\zeta_n \le k) > 1 - \delta$. Then

$$
\begin{aligned}
&P(\zeta_{n+1} > t|\cap_{j=1}^{n}\{\zeta_j < \infty\}, T)\\
&\quad = E[P_{\zeta_n}^*(\zeta_{n+1} > t)I(\zeta_n > k)] + E[P_{\zeta_n}^*(\zeta_{n+1} > t)I(\zeta_n \le k)]\\
&\quad \le P(\zeta_n > k) + \sum_{i=0}^{k} E[P_i^*(\zeta_{n+1} > t)I(\zeta_n = i)]\\
&\quad \le \delta + \sum_{i=0}^{k} N_1 t \times \exp\{-\beta(t-T)\} = \delta + o(t) \text{ as } t \to \infty.
\end{aligned}
$$

Hence $\lim_{t\to\infty} P^*(\zeta_{n+1} > t) \le \delta$ a.s. and since δ was arbitrary $\lim_{t\to\infty} P^*(\zeta_{n+1} > t) = 0$ a.s.. The induction step follows now from the dominated convergence theorem by

$$
\begin{aligned}
P(\zeta_{n+1} < \infty) &= E[P(\zeta_{n+1} < \infty|\cap_{j=1}^{n}\{\zeta_j < \infty\}, T)]\\
&= E[1 - \lim_{t\to\infty} P^*(\zeta_{n+1} > t)] = 1.
\end{aligned}
$$

□

Lemma 25. *If assumption 3 hold, then* $P(Z_n = \boldsymbol{i}\ i.o.) = 1$ *for all* $\boldsymbol{i}$ *in* G.

Proof. Notice first that for any $n > 1$ and for any sequence $(\mathbf{i}^0, \cdots, \mathbf{i}^n)$ of elements in G a similar derivation as in (5.33) yields that

$$P(Z_0 = \mathbf{i}^0, Z_1 = \mathbf{i}^1, \cdots, Z_n = \mathbf{i}^n) = \prod_{\mathbf{i}\in G} \frac{\tilde{m}(i_1)^{[t_n^*(i_1)]}}{\tilde{m}^{[n]}} \prod_{k=1}^{K-1} \frac{\tilde{m}_{(i_{1:k})}(i_{k+1})^{[t_n^*(i_{1:(k+1)})]}}{\tilde{m}_{(i_{1:k})}^{[t_n^*(i_{1:k})]}}$$

where for $\mathbf{i} \in G$ $t_n^*(\mathbf{i}) = \sum_1^n \delta_{Z_k}(\mathbf{i})$ and $t_n^*(i_{1:j}) = \sum_{i_{(j+1):K}\in \times_{j+1}^K} t_n^*(\mathbf{i})$ for $1 \le j \le K$. Therefore clearly $\{Z_n\}_n$ is exchangeable. Furthermore, from the previous lemma we know that $P(\cap_{n\le 0}\{\zeta_n < \infty\}) = 1$, hence $\{Z_n\}$ is an infinite exchangeable sequence. By de Finetti's theorem and using the same argumentation along the lines of Proposition 6 one can show that

$$\begin{aligned} Z_n|\Lambda &\sim^{iid} \Lambda \\ P(Z_n = \mathbf{i}|\Lambda) &= \lambda_{i_1} \times \lambda(i_1)_{i_2} \times \lambda(i_{1:2})_{i_3} \times \cdots \times \lambda(i_{1:(K-1)})_{i_K} \end{aligned}$$

with $\lambda \sim Dir(\tilde{m}(1), \cdots, \tilde{m}(N_1))$ and for $2 \le k \le K-1$

$$\lambda(i_{1:k}) \sim Dir\big(\tilde{m}_{(i_{1:k})}(1), \cdots, \tilde{m}_{(i_{1:k})}(N_{k+1})\big).$$

Where the random distributions $\{\lambda, \lambda(i_{1:k})\}_{i_{1:k}\in\times_{j=1}^k G_j, k=1,\cdots,K-1}$ are mutually independent. Fix an arbitrary $\mathbf{i} = (i_1, \cdots, i_k) \in G$, by assumption

$$\tilde{m}(i_1) \prod_{k=1}^{K-1} \tilde{m}_{(i_{1:k})(i_{k+1})} > 0.$$

Since $\lambda_{i_1} \sim Beta\big(\tilde{m}(i_1), \sum_{j\in G_1\setminus\{i_1\}} \tilde{m}(j)\big)$ it follows that $\lambda_{i_1} > 0$ a.s.. Similarly $\lambda(i_{1:k})_{i_{k+1}} \sim Beta\big(\tilde{m}_{(i_{1:k})}(i_{k+1}), \sum_{j\in G_{k+1}\setminus\{i_{k+1}\}} \tilde{m}_{(i_{1:k})}(j)\big)$ for $1 \le k \le K-1$ and therefore $\lambda(i_{1:k})_{i_{k+1}} > 0$ a.s.. Therefore

$$P(Z_n = \mathbf{i}|\Lambda) = \lambda_{i_1}\ \lambda(i_1)_{i_2}\ \cdots\ \lambda(i_{1:(K-1)}) > 0\ a.s..$$

Since, given Λ, $\{Z_n\}$ are independent and identical distributed such that

$P(Z_n = \mathbf{i}|\Lambda) > 0$ $a.s$, we may use Borel-Cantelli lemma conditional on the σ-filed of Λ to deduce that $P(Z_n = \mathbf{i}\ i.o.|\Lambda) = 1$ on a set of probability one. An application of Fubini theorem yields that

$$P(Z_n = \mathbf{i}\ i.o.) = E[P(Z_n = \mathbf{i}\ i.o.|\Lambda)] = 1.$$

Since $\mathbf{i} \in G$ was arbitrary the result follows. □

Now, using the previous two lemmas we can proof Proposition 5.

Proof of Proposition 5. We need to show (5.16), which by conditioning on the first state holds if we can show that $P[X_n = (\mathbf{i}, 0)\ i.o.] = 1$ for all $\mathbf{i} \in G$.

Hence, fix an arbitrary $\mathbf{i} \in G$. By Lemma 24 and assumption 3

$$P(\cap_{n=1}^{\infty}\{\zeta_n < \infty\}) = P(\{X_n = (j_n, 0)\ \exists j_n \in G\}\ i.o.) = 1.$$

Since $|G| = \prod_{k=1}^{K} N_k < \infty$ there must by a subsequence and some states $\mathbf{j} \in G$ such that $P(\cap_{k=1}^{\infty}\{\zeta_{n_k} < \infty\}) = P(X_k = (\mathbf{j}, 0)\ i.o.) = 1$. But by Lemma 25 also $P(Z_n = \mathbf{i}\ i.o.) = P(\cap_k\{Z_{n_k} = \mathbf{i}\}) = 1$. Therefore

$$P(X_n = (\mathbf{i}, 0)\ i.o.) = P\Big(\cap_{n=1}^{\infty}\{\zeta_n < \infty\} \bigcap \{Z_n = \mathbf{i}\ i.o.\}\Big) = 1.$$

Since $\mathbf{i}$ was arbitrary the result follows. □

Proof of Corollary 6. Proposition 4 showed that $\{X_n\}$ is partial exchangeable and assumption 3 imply the recurrence of $\{X_n\}$. An application of de Finetti's theorem for recurrent, partial exchangeable processes (see [44], Theorem 7) yields the claim. □

The proof of proposition 6 will follow directly from the following two lemmas.

Lemma 26. *Proposition 6 holds for every* $(\boldsymbol{i}, s) \in \mathcal{H}$ *such that* $P(X_n = (\boldsymbol{i}, s)\ i.o.) = 1$.

Proof. The proof is rather lengthy and will be given in two steps. The main idea is the following. Consider that the space of stochastic matrices on $S \times$

S with the topology of coordinate-wise convergence. Denote for $n \in \mathbb{N}$ with $\hat{\Pi}_n = \{\hat{\Pi}_{n,(\mathbf{i},s)(\mathbf{j},l)};\ (\mathbf{i},s),(\mathbf{j},v) \in S\}$ the empirical transition matrix with element, for $(\mathbf{i},s)$ and $(\mathbf{j},v)$ in S,

$$\hat{\Pi}_{n,(\mathbf{i},s)(\mathbf{j},l)} = \begin{cases} \dfrac{t_n(\mathbf{i},s)(\mathbf{j},l)}{t_n(\mathbf{i},s)} & \text{if } t_n(\mathbf{i},s) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.38}$$

Diaconis and Freedman [44] showed that for recurrent partial exchangeable processes $\hat{\Pi}_n \to \Pi_\infty$ in the topology of coordinate-wise convergence, i.e. coordinates-wise $\Pi_{(\mathbf{i},s)(\mathbf{j},v)} \stackrel{d}{=} \lim_{n\to\infty} \frac{t_n(\mathbf{i},s)(\mathbf{j},l)}{t_n(\mathbf{i},s)}$ for $(\mathbf{i},s)$ and $(\mathbf{j},l)$ in S whenever the denominator is positive. Step one consists of finding the limit of the scaled transition counts and in step 2 we use this limits to express the coordinate-wise limit of the empirical transition probabilities.

Step 1. Let $(\mathbf{i},s) \in \mathcal{H}$. By assumption $(\mathbf{i},s)$ is visited infinitely often by $\{X_n\}$ with probability one. Now, let $\xi^n_{(\mathbf{i},s)}$ denotes the n^{th} hitting time to $(\mathbf{i},s)$, that is $\xi^n_{(\mathbf{i},s)} = \inf\{k > \xi^{n-1}_{(\mathbf{i},s)} : X_k = (\mathbf{i},s)\}$ and let $\{V(\mathbf{i},s)_n\}$ be the sequence of successor states of $(\mathbf{i},s)$, where $V(\mathbf{i},s)_n = X_{\xi^n_{(\mathbf{i},s)}+1} \in G \times \{0, s+1\}$. Notice that from the form of the transition probability (5.6)

$$P(V(\mathbf{i},s)_{n+1}|\mathcal{F}_{\xi^n_{(\mathbf{i},s)}}) = P(V(\mathbf{i},s)_{n+1}|\{V(\mathbf{i},s)_v\}_{v\le n-1}, \{Z_k\}_{k\le t^*_{\xi^n_{(\mathbf{i},s)}}}, \xi^n_{(\mathbf{i},s)}).$$

Meaning that the probability of successor states of $(\mathbf{i},s)$ depend on the history of the process only through the previous successor states and visits to states in $G \times \{0\}$ and the time of the n^{th} hit to $(\mathbf{i},s)$. This become also clear from step (B) if look at the sampling scheme 5.1 to sample $V(\mathbf{i},s)_{n+1} = (\mathbf{j},l) \in G \times \{0, s+1\}$.

Let $(\mathbf{j},l) \in R(\mathbf{i},s) \subset G \times \{0, s+1\}$. Consider the first case, where $l = s+1$, i.e. $m_{(\mathbf{i},s)}(j_1) \prod_{k=1}^{K-1} m_{(\mathbf{i},s|j_{1:k})}(j_{k+1}) > 0$. Denote with $\{\theta(\mathbf{i},s)_n\}$ the sequence of labels drawn from Pólya urn $U^{(1)}(\mathbf{i},s)$ and for $1 \le k \le K-1$ denote with $\{\theta(\mathbf{i},s|j_{1:k})_n\}$ be sequence of labels drawn from $U^{(k)}(\mathbf{i},s|j_{1:k})$. Hence from

the sampling scheme 5.1

$$\begin{aligned}&\big\{V(\mathbf{i},s)_{n+1}=(\mathbf{j},s+1)\big\}\\&=\big\{\theta(\mathbf{i},s)_{n+1}=j_1\big\}\bigcap\bigcap_{k=1}^{K-1}\big\{\theta(\mathbf{i},s|j_{1:k})_{t_{\xi^n_{(\mathbf{i},s)}}(\mathbf{i},s)(j_{1:k},s+1)+1}=j_{k+1}\big\}.\end{aligned}$$

We precede recursively as follows. Since $P(X_n=(\mathbf{i},s)\ i.o)=1$ we sample infinity often from urn $U^{(1)}(\mathbf{i},s)$ and for $I(\mathbf{i},s)_{j,n}=\#\{k\leq n:\theta(\mathbf{i},s)_k=j\}$, by [21], as $n\to+\infty$,

$$\begin{aligned}n^{-1}\Big(I(\mathbf{i},s)_{0,n},\cdots,I(\mathbf{i},s)_{N_1,n}\Big)&\overset{a.s.}{\to}p(\mathbf{i},s)\sim Dir(m_{(\mathbf{i},s)}(0),\cdots,m_{(\mathbf{i},s)}(N_1)),\\\{\theta(\mathbf{i},s)_n\}&\overset{iid}{\sim}\text{Multinomial}(1,p(\mathbf{i},s))\end{aligned}$$

where $p(\mathbf{i},s)_{j_1}>0\ a.s$ since $m_{(\mathbf{i},s)}(j_1)>0$.
Now, apply Borel-Cantelli lemma to $\{\theta(\mathbf{i},s)_n=j_1\}_n$ to deduce that

$$P(\theta(\mathbf{i},s)_n=j_1\ i.o.|p(\mathbf{i},s))\overset{a.s.}{=}1 \tag{5.39}$$

By Fubini theorem this implies that $P(\theta(\mathbf{i},s)_n=j_1\ i.o.)=1$ and hence we sample infinitely often from $U^{(2)}(\mathbf{i},s|j_1)$. Since we sample infinitely often from $U^{(2)}(\mathbf{i},s|j_1)$, we may use again [21], to deduce that, for $I(\mathbf{i},s|j_1)_{j,n}=\#\{k\leq n:\theta(\mathbf{i},s|j_1)_k=j\}$, as $n\to+\infty$

$$\begin{aligned}n^{-1}\Big(I(\mathbf{i},s|j_1)_{1,n},\cdots,I(\mathbf{i},s|j_1)_{N_2,n}\Big)&\overset{a.s.}{=}p(\mathbf{i},s|j_1)\\p(\mathbf{i},s|j_1)&\sim Dir(m_{(\mathbf{i},s|j_1)}(1),\cdots,m_{(\mathbf{i},s|j_1)}(N_2)),\\\{\theta(\mathbf{i},s|j_1)_n\}&\overset{iid}{\sim}\text{Multinomial}(1,p(\mathbf{i},s|j_1))\end{aligned}$$

where again $p(\mathbf{i},s|j_1)_{j_2}>0\ a.s$ since $m_{(\mathbf{i},s|j_1)}(j_2)>0$. Again by Borel-Cantelli lemma and Fubini theorem we sample infinitely often from $U^{(2)}(\mathbf{i},s|j_1)$. Proceed further this way and use that $m_{(\mathbf{i},s)}(j_1)\prod_{k=1}^{K-1}m_{(\mathbf{i},s|j_{1:k})}(j_{k+1})>0$ we deduce that for $2\leq k\leq K-1$ we sample infinity often from $U^{(k+1)}((\mathbf{i},s)|j_{1:k})$

and

$$p(\mathbf{i},s|j_{1:k}) \overset{a.s.}{=} n^{-1}(I(\mathbf{i},s|j_{1:k})_{1,n},\cdots,I(\mathbf{i},s|j_{1:(K-1)})_{N_{k+1},n}),$$
$$p(\mathbf{i},s|j_{1:k}) \sim Dir(m_{(\mathbf{i},s|j_{1:k})}(1),\cdots,m_{(\mathbf{i},s|j_{1:k})}(N_{k+1})),$$
$$\{\theta(\mathbf{i},s|j_{1:k})_n\} \overset{iid}{=} Multinomial(1,p(\mathbf{i},s|j_{1:k})).$$

Similarly, for the case $l=0$, by definition of $R^2(\mathbf{i},s)$ and from assumption 3 $m_{(\mathbf{i},s)}(0)\tilde{m}(j_1)\prod_{k=1}^{K-1}\tilde{m}_{j_{1:k})}(j_{k+1})>0$ and

$$\begin{aligned}&\left\{V(\mathbf{i},s)_{n+1}=(\mathbf{j},0)\right\}\\&=\left\{\theta(\mathbf{i},s)_{n+1}=0\right\}\bigcap\left\{\alpha(\mathbf{i},s)_{t^*_{\xi^n_{(\mathbf{i},s)}}+1}=j_1\right\}\bigcap\bigcap_{k=1}^{K-1}\left\{\alpha(j_{1:k})_{t^*_{\xi^n_{(\mathbf{i},s)}}(j_{1:k})+1}=j_{k+1}\right\}\end{aligned}$$

where $\{\alpha\}$ and $\{\alpha(j_{1:k})_n\}$ denotes the sequence of labels draws from urn $\tilde{U}^{(*1)}$ and $\tilde{U}^{(*k)}(j_{1:k})$ for $1\leq k\leq K-1$. Using exactly the same argumentation as before we have

$$n^{-1}(J_{1,n},\cdots,J_{N_K,n}) \overset{a.s.}{\to} \lambda \sim Dir(\tilde{m}(1),\cdots,\tilde{m}(N_1)),$$
$$n^{-1}(J(j_{1:k})_{1,n},\cdots,J(j_{1:k})_{N_{k+1},n}) \overset{a.s.}{\to} \lambda(j_{1:k}) \sim Dir(\tilde{m}_{(j_{1:k})}(1),\cdots,\tilde{m}_{(j_{1:k})}(N_{k+1}))$$

for $1\leq k\leq K-1$.

Step 2. We can now use step one to find the coordinate-wise limit of the empirical transition probabilities and hence the law of the transition row $\Pi_{\mathbf{i},s}=\{\Pi_{(\mathbf{i},s),(\mathbf{j},l)}\}_{(\mathbf{j},l)\in S}$.

For $(\mathbf{j},l)\notin R_{(\mathbf{i},s)}$, $t_n(\mathbf{i},s)(\mathbf{j},l)=0$ for all n, and therefore

$$\Pi_{(\mathbf{i},s)(\mathbf{j},l)} \overset{d}{=} \lim_n \frac{t_n(\mathbf{i},s)(\mathbf{j},l)}{t_n(\mathbf{i},s)}=0$$

For $(\mathbf{j},l)\in R_{(\mathbf{i},s)}, l=s+1$ and by step one we have that

$$\Pi_{(\mathbf{i},s)(\mathbf{j},l)} \overset{d}{=} \lim_n \frac{t_n(\mathbf{i},s)(\mathbf{j},s+1)}{t_n(\mathbf{i},s)}$$

$$\begin{aligned}
&\overset{a.s}{=} \lim_n \frac{t_n(\mathbf{i},s)(j_1,s+1)}{t_n(\mathbf{i},s)} \prod_{k=1}^{K-1} \lim_n \frac{t_n(\mathbf{i},s)(j_{1:(k+1)},s+1)}{t_n(\mathbf{i},s)(j_{1:k},s+1)} \\
&\overset{a.s}{=} \lim_l \frac{I(\mathbf{i},s)_{l,j_1}}{l} \prod_{k=1}^{K-1} \lim_l \frac{I(\mathbf{i},s|j_{1:k})_{k,j_{k+1}}}{k} \\
&\overset{a.s}{=} p(\mathbf{i},s)_{j_1} \prod_{k=1}^{K-1} p(\mathbf{i},s|_{j_{1:k}})_{j_{k+1}}
\end{aligned}$$

and for $(\mathbf{j},l) \in R_{(\mathbf{i},s)}, l=0$ by step one we have that

$$\begin{aligned}
\Pi_{(\mathbf{i},s)(\mathbf{j},l)} &\overset{d}{=} \lim_n \frac{t_n(\mathbf{i},s)(\mathbf{j},l)}{t_n(\mathbf{i},s)} \overset{a.s}{=} \lim_n \left(\frac{t_n(\mathbf{i},0)(\cdot,0)}{t_n(\mathbf{i},s)} \frac{t_n^*(j_1)}{t_n^*} \prod_{k=1}^{K-1} \frac{t^*(j_{1:(k+1)})}{t^*(j_{1:k})} \right) \\
&\overset{a.s}{=} \lim_n \frac{t_n(\mathbf{i},0)(\cdot,0)}{t_n(\mathbf{i},s)} \lim_n \frac{t_n^*(j_1)}{t_n^*} \prod_{k=1}^{K-1} \lim_n \frac{t^*(j_{1:(k+1)})}{t^*(j_{1:k})} \\
&\overset{a.s}{=} \lim_l \frac{I(\mathbf{i},s)_{l,0}}{l} \lim_l \frac{J_{l,j_1}}{l} \prod_{k=1}^{K-1} \lim_l \frac{J(j_{1:k})_{l,j_{k+1}}}{l} \ a.s. \\
&\overset{a.s}{=} p(\mathbf{i},s)_0 \lambda_{j_1} \prod_{k=1}^{K-1} \lambda(j_{1:k})_{j_{k+1}}.
\end{aligned}$$

Finally if $(\mathbf{j},s+1) \in S \setminus R(\mathbf{i},s)$ then $m_{(\mathbf{i},s)}(j_1) \prod_{k=1}^{K-1} m_{(\mathbf{i},s|j_{1:k})}(j_{k+1}) = 0$. Let k' be the first index for which $m_{(\mathbf{i},s|j_{1:k'})}(j_{k'+1}) = 0$. Then $t_n(\mathbf{i},s)(\mathbf{j},s+1) = 0$ for all $n \geq 0$ and

$$t_n(\mathbf{i},s)(j_{1:k},s+1) > 0 \qquad \text{for } k \leq k' \text{ and } n \text{ large} \tag{5.40}$$

$$t_n(\mathbf{i},s)(j_{1:k},s+1) = 0 \qquad \text{for } k = k'+1,\cdots,K \text{ and } \forall n \geq 0 \quad . \tag{5.41}$$

Hence

$$\begin{aligned}
0 &\overset{d}{=} \lim_n \frac{t_n(\mathbf{i},s)(\mathbf{j},s+1)}{t_n(\mathbf{i},s)} \\
&\overset{a.s.}{=} \lim_n \frac{t_n(\mathbf{i},s)(j_1,s+1)}{t_n(\mathbf{i},s)} \prod_{k=2}^{k'} \lim_n \frac{t_n(\mathbf{i},s)(j_{1:k},s+1)}{t_n(\mathbf{i},s)(j_{1:(k-1)},s+1)} \times \frac{t_n(\mathbf{i},s)(j_{1:(k'+1)},s+1)}{t_n(\mathbf{i},s)(j_{1:k'},s+1)}
\end{aligned}$$

$$\stackrel{a.s.}{=} \lim_l \frac{I(\mathbf{i},s)_{l,j_1}}{l} \prod_{k=1}^{k'} \lim_l \frac{I(\mathbf{i},s|j_{1:k})_{l,j_{k+1}}}{l} \times 0$$

$$\stackrel{a.s.}{=} p(\mathbf{i},s)_{j_1} \prod_{k=1}^{k'-1} p(\mathbf{i},s|j_{1:k})_{j_{k+1}} \times p(\mathbf{i},s|j_{1:k'})_{j_{k'+1}} \times \prod_{v=k'}^{K-1} p(\mathbf{i},s|(j_{1:k'},j_{(k'+1):v}))_{j_{v+1}}.$$

Where the last equality hold since $p(\mathbf{i},s|j_{1:k'})_{j_{k'+1}} = 0$ a.s. and the random conditional distributions $p(\mathbf{i},s|(j_{1:k'},j_{(k'+1):v}))_{j_{v+1}}$ can be choose arbitrary on the set

$$\Big\{X_n = (\mathbf{i},s), X_{n+1} \in \{i_{1:(k'+1)}\} \times_{v=k'+2}^{K} G_v \times \{s+1\}\Big\}. \tag{5.42}$$

□

Lemma 27. *If the initial urn compositions satisfy assumption 3, then for all states* $(\boldsymbol{i},s)$ *in* $\mathcal{H}$ $P(X_n = (\boldsymbol{i},s)\ i.o.) = 1$.

Proof. It suffices to show that $P(X_n = (\mathbf{i},s)\ i.o.) = 1$ for $(\mathbf{i},s) \in \mathcal{H}_m$ and all $m \geq 0$. For $m = 0$, $\mathcal{H}_0 = G \times \{0\}$ and the claim is true by recurrence of $\{X_n\}$. Now, assume that the $P(X_n = (\mathbf{i},s)\ i.o.) = 1$ for $(\mathbf{i},s)$ in $\mathcal{H}_k$ for all $1 \leq k \leq m$. Let $(\mathbf{i},s) \in \mathcal{H}_{m+1} = \bigcup_{(\mathbf{j},v)\in\mathcal{H}_m} R_{(\mathbf{j},v)}$. If $s = 0$, then $(\mathbf{i},s) \in G \times \{0\} = \mathcal{H}_0$ and hence $P(X_n = (\mathbf{i},s)\ i.o.) = 1$.

If $s > 0$, there exists a $(\mathbf{j},v) \in \mathcal{H}_m$ such that $(\mathbf{i},s) \in R_{(\mathbf{j},v)}$ and $s = l+1$. By the induction assumption $P(X_n = (\mathbf{v},l)\ i.o.) = 1$ we may use the previous lemma and that $m_{(\mathbf{j},l)}(i_1)\prod_{k=1}^{K-1} m_{(\mathbf{j},v|i_k)}(i_{k+1}) > 0$ to write

$$0 < P[X_{m+1} = (\mathbf{i},v+1)|X_m = (\mathbf{j},v)] = \mathbb{E}[p(\mathbf{j},v)_{i_1} \prod_{k=1}^{K} p(\mathbf{j},v|i_{1:k})_{i_{k+1}}].$$

where $p(\mathbf{j},v)_{i_1} \sim Beta(m_{(\mathbf{j},v)}(i_1), m_{(\mathbf{j},v)} - m_{(\mathbf{j},v)}(j_1))$ and for $1 \leq k \leq K-1$

$$p(\mathbf{j},v|i_{1:k})_{i_{k+1}} \sim Beta(m_{(\mathbf{j},v)|i_{1:k})}(i_{k+1}), m_{(\mathbf{j},v|i_{1:k})} - m_{(\mathbf{j},v|i_{1:k}}(i_{k+1})).$$

Since $m_{(\mathbf{j},l)}(i_1)\prod_{k=1}^{K-1} m_{(\mathbf{j},v|i_k)}(i_{k+1}) > 0$, there exists a measurable set N such

that $P(N) = 0$ and $\Pi_{(\mathbf{j},v),(\mathbf{i},v+1)}(\omega) > 0$ for $\omega \in N^c$. Furthermore

$$P(X_n = (\mathbf{j}, l)\ i.o.) = \mathbb{E}[P(X_n = (\mathbf{j}, l)\ i.o.|\Pi)] = 1.$$

hence there exists a measurable set M such that $P(M) = 0$ and for $\omega \in M^c$ $P(X_n = (\mathbf{j}, l)\ i.o.|\Pi(\omega)) > 0$. But, given Π, $\{X_n\}$ is Markov therefore $P(X_n = (\mathbf{j}, l)\ i.o.|\Pi(\omega)) = 1$ for every $\omega \in M^c$. Now $P(X_n = (\mathbf{j}, v)\ i.o.|\Pi(\omega)) = 1$ and $\Pi_{(\mathbf{j},v),(\mathbf{i},v+1)}(\omega) > 0$ and, since the class of recurrence states of $\Pi(\omega)$ is closes, $P(X_n = (\mathbf{i}, v+1)\ i.o.|\Pi(\omega)) = 1$. Hence

$$\begin{aligned} P(X_n = (\mathbf{i}, s)\ i.o.) &= E[P(X_n = (\mathbf{i}, v+1)\ i.o.|\Pi)] \\ &\geq E[P(X_n = (\mathbf{i}, v+1)\ i.o.|\Pi)I((N \cup M)^c)] \\ &= P((N \cup M)^c) = 1 \end{aligned}$$

□

Proof of lemma 17. The case $v_n = n$ follows by construction of the process. The case $v_n = \zeta_n$ follows from the former case, since given $\{X_k\}_{k\leq\zeta_n} \cap \{\zeta_n = t\}$, $\{X_{\zeta_n+k}\}_{k\geq 0}$ is a RHUP. Take the union over all $t \in \mathbb{N}_0$ and notice that $P(\zeta_n < \infty) = 1$ the result follows. □

Proof of Corollary 7. Consider the case where $v_n = n$ first. By Lemma 17, given a sample $\{X_k = (\mathbf{i}_k, t_k)\}_{k=0,\cdots,n}$, $\{X_{n+k}\}_{k\geq 0}$ is again a RHUP with initial state $(\mathbf{i}_n, t_n)$. It is not difficult to show that assumption 3 also imply that $\{X_{n+k}\}_{k\geq 0}$ is recurrence. Hence, for $Z_k := X_{n+k}$, $\{Z_n\}_n$ is a recurrent partial exchangeable process and the result follows now by Corollary 6, Proposition 6 and Lemma 17.

The case $v_n = \zeta_n$ follows from the first case and Corollary 17 (ii). □

Proof of Lemma 18. Fix an $n > 1$. Consider the first N elements of $\{\mathbf{B}_j\}_{j\geq 1}$ and let σ be an arbitrary permutation of $\{1, 2, \cdots n\}$. We want to show that

$$(\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_n) \stackrel{d}{=} (\mathbf{B}_{\sigma(1)}, \mathbf{B}_{\sigma(2)}, \cdots, \mathbf{B}_{\sigma(n)}). \tag{5.43}$$

We will express the finite dimensional law of both sides of (5.43) through the

finite dimensional law of the RHUP. For $(\mathbf{i}, s)$ and $(\mathbf{j}, v)$ in S let $t_{B_1,\cdots,B_n}(\mathbf{i}, s)(\mathbf{j}, v)$ denote the number of transitions from $(\mathbf{i}, s)$ to $(\mathbf{j}, v)$ in the string $(\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_n)$ and define $t_{B_{\sigma(1)},\cdots,B_{\sigma(n)}}(\mathbf{i}, s)(\mathbf{j}, v)$ similar for the permutation $(\mathbf{B}_{\sigma(1)}, \mathbf{B}_{\sigma(2)}, \cdots, \mathbf{B}_{\sigma(n)})$. Notice that the number of transitions from $(\mathbf{i}, s)$ to $(\mathbf{j}, v)$ in $\mathbf{B}_v$ is one if $v = s+1$ and $s \leq \zeta_v - \zeta_{v-1} - 2$ and also $(X_{\zeta_{v-1}+s}, X_{\zeta_{v-1}+s+1}) = ((\mathbf{i}, s), (\mathbf{j}, s+1))$ and zero otherwise. Furthermore let $l_v = \zeta_n - \zeta_{n-1} - 1$ denote the length of the vector $\mathbf{B}_v$. Hence

$$\begin{aligned}
&t_{B_1,\cdots,B_n}(\mathbf{i}, s)(\mathbf{j}, s+1) \\
&= \sum_{v=1}^{n} I(X_{\zeta_{v-1}+s} = (\mathbf{i}, s), X_{\zeta_{v-1}+s+1} = (\mathbf{j}, s+1), s \leq l_v - 1) \\
&= \sum_{v=1}^{n} I(X_{\zeta_{\sigma(v)-1}+s} = (\mathbf{i}, s), X_{\zeta_{\sigma(v)-1}+s+1} = (\mathbf{j}, s+1), s \leq l_{\sigma(v)} - 1) \\
&= t_{B_{\sigma(1)},\cdots,B_{\sigma(n)}}(\mathbf{i}, s)(\mathbf{j}, s+1)
\end{aligned}$$

and

$$\begin{aligned}
t_{B_1,\cdots,B_n}(\mathbf{i}, s)(\cdot, 0) &= \sum_{v=1}^{n} I(X_{\zeta_{v-1}+s} = (\mathbf{i}, s), l_v = s) \\
&= \sum_{v=1}^{n} I(X_{\zeta_{\sigma(v)-1}+s} = (\mathbf{i}, s), l_{\sigma(v)} = s) \\
&= t_{B_{\sigma(1)},\cdots,B_{\sigma(n)}}(\mathbf{i}, s)(\cdot, 0).
\end{aligned}$$

This implies that $t_{B_1,\cdots,B_n}(\mathbf{i}, s)(j_{1:k}, s+1) = t_{B_{\sigma(1)},\cdots,B_{\sigma(n)}}(\mathbf{i}, s)(j_{1:k}, s+1)$ for $1 \leq k \leq K-1$ and that $t_{B_1,\cdots,B_n}(\mathbf{i}, s) = t_{B_{\sigma(1)},\cdots,B_{\sigma(n)}}(\mathbf{i}, s)$. Furthermore, for $\mathbf{i} \in G$ clearly

$$t^*_{B_1,\cdots,B_n}(\mathbf{i}) = \sum_{v=1}^{n} I\{X_{\zeta_{j-1}} = \mathbf{i}\} = \sum_{v=1}^{n} I\{X_{\zeta_{\sigma(j)-1}} = \mathbf{i}\} = t^*_{B_{\sigma(1)},\cdots,B_{\sigma(n)}}(\mathbf{i})$$

and therefore also $t^*_{B_1,\cdots,B_n}(i_{1:k}) = t^*_{B_{\sigma(1)},\cdots,B_{\sigma(n)}}(i_{1:k})$ for $1 \le k \le K-1$. Now

$$
\begin{aligned}
&P\big[\cap_{v=1}^n\{\mathbf{B}_v = \mathbf{b}_v\}\big] \\
&= P\big[\cap_{v=1}^n\{(X_{\zeta_{v-1}},\cdots,X_{\zeta_v-1}) = \mathbf{b}_v\}\big] \\
&= \prod_{\substack{\mathbf{i}\in G,\\ s\le\tau_n}}\Bigg[\left(\frac{\tilde m(i_1)^{[t^*_{B_1,\cdots,B_n}(i_1)]}}{\tilde m^{t^*_{B_1,\cdots,B_n}}}\prod_{k=1}^{K-1}\frac{\tilde m_{(i_{1:k})}(i_{k+1})^{[t^*_{B_1,\cdots,B_n}(i_{1:k+1})]}}{\tilde m_{(i_{1:k})}^{[t^*_{B_1,\cdots,B_n}(i_{1:k})]}}\right) \\
&\qquad\qquad\times\left(\frac{m_{(\mathbf{i},s)}(0)^{[t_{B_1,\cdots,B_n}(\mathbf{i},s)(\cdot,0)]}}{m_{(\mathbf{i},s)}^{[t_{B_1,\cdots,B_n}(\mathbf{i},s)]}}\right) \\
&\qquad\qquad\times\prod_{\mathbf{j}\in G}\Bigg(m_{(\mathbf{i},s)}(j_1)^{[t_{B_1,\cdots,B_n}(\mathbf{i},s)(j_1,s+1)]} \\
&\qquad\qquad\qquad\times\prod_{k=1}^{K-1}\frac{m_{(\mathbf{i},s|j_{1:k})}(j_{k+1})^{[t_{B_1,\cdots,B_n}(\mathbf{i},s)(j_{1:k+1},s+1)]}}{m_{(j_{1:k})}^{[t_{B_1,\cdots,B_n}(\mathbf{i},s)(j_{1:k},s+1)]]}}\Bigg)\Bigg] \\
&= P\big[\cap_{v=1}^n\{\mathbf{B}_v = \mathbf{b}_{\sigma(v)}\}\big]
\end{aligned}
$$

Where the last equality holds by equality of all counting processes of $\{B_{\sigma(j)}\}_{1\le j\le n}$ and $\{B_j\}_{1\le j\le n}$. □

Proof of Lemma 5.31. The first part is obvious. For the second part we simple rewrite the left hand side of (5.31) as

$$
\begin{aligned}
&\overset{a.s}{=} P\Big[\{X_{\zeta_{n-1}-1}\in S\}\cap\cap_{t=0}^k\big\{X_{\zeta_{n-1}+t} = (\mathbf{i}_t,t)\big\}\cap\big\{X_{\zeta_{n-1}+k+1}\in G\times\{0\}\big\}\Big|\Pi\Big] \\
&\overset{a.s}{=} \sum_{\mathbf{w}\in G,(\mathbf{j},v)\in S} P\Big[X_{\zeta_{n-1}-1} = (\mathbf{j},v)|\Pi\Big]\Pi_{(\mathbf{j},v),(\mathbf{i}_0,0)}\left[\prod_{t=0}^{k-1}\Pi_{(\mathbf{i}_t,t),(\mathbf{i}_{t+1},t+1)}\right]\Pi_{(\mathbf{i}_k,k),(\mathbf{w},0)} \\
&\overset{a.s}{=} \sum_{\mathbf{w}\in G,(\mathbf{j},v)\in S}\Big(P\Big[X_{\zeta_{n-1}-1} = (\mathbf{j},v)|\Pi\Big]p(\mathbf{j},v)_0\Big)\Lambda_{\mathbf{i}_0}\left[\prod_{t=0}^{k-1}\Pi_{(\mathbf{i}_t,t),(\mathbf{i}_{t+1},t+1)}\right]p(\mathbf{i}_k,k)_0\Lambda_{\mathbf{w}} \\
&\overset{a.s}{=} \Lambda_{\mathbf{i}_0}\left[\prod_{t=0}^{k-1}\Pi_{(\mathbf{i}_t,t),(\mathbf{i}_{t+1},t+1)}\right]p(\mathbf{i}_k,k)_0, \qquad (5.44)
\end{aligned}
$$

where the last equality holds by

$$\begin{aligned}
&\sum_{(\mathbf{j},v)\in S\in G} P\Big[X_{\zeta_{n-1}-1}=(\mathbf{j},v)|\Pi\Big]p(\mathbf{j},v)_0\\
&\overset{a.s}{=} \sum_{(\mathbf{j},v)\in S,\mathbf{i}\in G} P\Big[X_{\zeta_{n-1}-1}=(\mathbf{j},v)|\Pi\Big]p(\mathbf{j},v)_0\Lambda_{\mathbf{i}}\\
&\overset{a.s}{=} \sum_{(\mathbf{j},v)\in S,\mathbf{i}\in G} P\Big[X_{\zeta_{n-1}-1}=(\mathbf{j},v)|\Pi\Big]P(X_{\zeta_{n-1}}\in G\times\{0\}|X_{\zeta_{n-1}-1}=(\mathbf{j},v),\Pi)\\
&\overset{a.s}{=} P(X_{\zeta_{n-1}-1}\in S, X_{\zeta_{n-1}}\in G\times\{0\}|\Pi)\overset{a.s}{=}1
\end{aligned}$$

□

Proof of Lemma 20. The computations are straight forward. We demonstrate only (iii),

$$\begin{aligned}
&P[Y_{t+1}^{(M+1,d)}=v, Y_{t+1}^{(M+1,d')}=w|Y_t^{(M+1)}=\mathbf{j},\{\mathbf{Y}_n\}_{n\le M}]\\
&=\sum_{\mathbf{i}\in G:i_d=w,i_{d'}=w} P[X_{\zeta_M+t+1}=(\mathbf{i},t+1)|X_{\zeta_M+t}=(\mathbf{j},t),\{X_l\}_{l=0,\cdots,\zeta_M}]
\end{aligned}$$

$$\begin{aligned}
=&\mathbb{E}\Big[\sum_{\mathbf{i}\in G:i_d=w,i_{d'}=w} P[X_{\zeta_M+t+1}=(\mathbf{i},t+1)|X_{\zeta_n+t}=(\mathbf{j},t),\Pi^{(\zeta_M)}]\Big]\\
=&\sum_{\substack{i_v\in G_v;\\ v\notin\{d,d'\},\\ 1\le v\le d'-1}} \mathbb{E}[p^{(\zeta_M)}(\mathbf{i},t)_{i_1}]\Big[\prod_{k=1}^{d-2}\mathbb{E}[p^{(\zeta_M)}(\mathbf{i},t|i_{1:k})_{i_{k+1}}]\Big]\\
&\times\mathbb{E}[p^{(\zeta_M)}(\mathbf{i},t|i_{1:(d-1)})_v]\mathbb{E}[p^{(\zeta_M)}(\mathbf{i},t|i_{1:k},v)_{i_{d+1}}]\\
&\times\Big[\prod_{k=d}^{h-2}\mathbb{E}[p^{(\zeta_M)}(\mathbf{i},t|i_{1:k},v,i_{(d+1):k})_{i_{k+1}}]\Big]\mathbb{E}[p^{(\zeta_M)}(\mathbf{i},t|i_{1:k},v,i_{(d+1):(h-1)})_w]
\end{aligned}$$

where for the second equality we use lemma 17 and a Corollary 7. Solving the expectations we obtain (iii). □

Proof of Lemma 21. We compute only (ii)

$$\begin{aligned}
&P[Y_l^{(M+1,d)} = v|\{\mathbf{Y}_n\}_{n\leq M}] \\
&= \sum_{\mathbf{i}\in G:i_d=v} P[\cap_{j=0}^{t-1}\{X_{\zeta_M+j}\in G\times\{j\}\cap\{X_{\zeta_M+t}=(\mathbf{i},t)|\{X_j\}_{j\leq\zeta_M}] \\
&= \mathbb{E}\Big[\sum_{\mathbf{i}\in G:i_d=v} P[\cap_{j=0}^{t-1}\{X_{\zeta_M+j}\in G\times\{j\}\cap\{X_{\zeta_M+t}=(\mathbf{i},t)\}|\Pi^{(\zeta_M)}]\Big] \\
&= \sum_{\mathbf{i}\in G:i_d=v}\Big[\hat{\Lambda}^{(M)}\prod_{t=1}^{l-1}\hat{\Pi}_{-0}^{(M)}(t)\Big]_{\mathbf{i}}.
\end{aligned}$$

The derivation of (i) is similar. □

Proof of lemma 22. We compute the failure rate first

$$\begin{aligned}
&P[\tau_{M+1}=l|\{\mathbf{Y}_n\}_{n\leq M}] \\
&= P[\{\tau_{M+1}=l\}\cap\cap_{t=1}^{l}\{Y_t^{(M+1)}\in G\}|\{\mathbf{Y}_n\}_{n\leq M}] \\
&= \mathbb{E}\Big[P[\cap_{t=1}^{l}\{X_{\zeta_M+t}\in G\times\{t\}\}\cap\{X_{\zeta_m+l+1}\in G\times\{0\}\}|\Pi^{(\zeta_M)}]\Big] \\
&= \mathbb{E}\Big[\sum_{\mathbf{i},\mathbf{j}\in G}\Big(\Lambda^{(\zeta_M)}\prod_{t=0}^{l-1}\Pi_{-0}^{(\zeta_M)}(t)\Big)_{\mathbf{i}}\Big(p^{(\zeta_M)}(\mathbf{i},l)_0\Lambda_{\mathbf{j}}^{(\zeta_M)}\Big)\Big] \\
&= \Big[\hat{\Lambda}^{(M)}\prod_{t=1}^{l-1}\hat{\Pi}_{-0}^{(M)}(t)\Big]\hat{p}_0^{(M)}(l).
\end{aligned}$$

The computation of the survival function is similar. Furthermore, notice that

$$E[\tau_M||\{\mathbf{Y}_n\}_{n\leq M}] = \sum_{l\geq 0}P(\tau_m>l)|\{\mathbf{Y}_n\}_{n\leq M}) \quad = \sum_{l\geq 0}\Big[\hat{\Lambda}^{(M)}\prod_{t=1}^{l-1}\hat{\Pi}_{-0}^{(M)}(t)\Big]1_G$$

For (ii), observe that for $v \in G_d$

$$\begin{aligned}\mathbb{E}[T(\mathbf{Y}_{M+1})_{d,v}|\{\mathbf{Y}_n\}_{n\leq M}] &= \mathbb{E}[\sum_{l\geq 0} I(Y_{l,d}^{(M+1)} = v)|\{\mathbf{Y}_n\}_{n\leq M}] \\ &= \sum_{l\geq 0} P(Y_{l,d}^{(M+1)} = v|\{\mathbf{Y}_n\}_{n\leq M}) \\ &= \sum_{\substack{l\geq 0,\\ \mathbf{i}\in G: i_d = v}} \Big[\hat{\Lambda}^{(M)} \prod_{t=1}^{l-1} \hat{\Pi}_{-0}^{(M)}(t)\Big]_{\mathbf{i}}.\end{aligned}$$

the predicted average occupation time in $\mathbf{i}$ can be calculates similar. □

Proof of Lemma 23. The expectation of the d^{th} component at time $(t+1)$, given $\{\mathbf{Y}_n\}_{n\leq M}$ is given by

$$\begin{aligned}E[Y_{t+1,d}^{(M+1)}|\{\mathbf{Y}_n\}_{n\leq M}] &= \sum_{i_d\in G_d} i_d P(Y_{t+1,d}^{(M+1)} = i_d|\{\mathbf{Y}_n\}_{n\leq M}) \\ &= \sum_{\mathbf{i}\in G} i_d \Big[\hat{\Lambda}^{(M)} \prod_{t=1}^{l-1} \hat{\Pi}_{-0}^{(M)}(t)\Big]_{\mathbf{i}}\end{aligned}$$

and the expectation of the product of the d^{th} and the d'^{th} components at time $(t+1)$ is given by

$$\begin{aligned}E[Y_{t+1,d}^{(M+1)} Y_{t+1,d'}^{(M+1)}|\{\mathbf{Y}_n\}_{n\leq M}] &= \sum_{i\in G_d, j\in G_{d'}} ij P[Y_{t+1,d}^{(M+1)} = i, Y_{t+1,d'}^{(M+1)} = j|\{\mathbf{Y}_n\}_{n\leq M}] \\ &= \sum_{\mathbf{i}\in G} i_d i_{d'} \Big(\hat{\Lambda}_M \prod_{l=0}^{t-1} \hat{\Pi}_{sub,M}(l)\Big)_{\mathbf{i}}\end{aligned}$$

Hence (ii) follows by definition of the covariance. (i) can be computed similarly. □

Part II

Inference for some one-counter Queues

Chapter 6

Some Preliminaries On Queuing Processes

6.1 Introduction

In the present chapter we shortly summarize the general structure of queuing processes without going into details. A queuing model is a stochastic description of a service system. The generic term service can be very general like the digestion of food or a thunder storm which enters a certain area. In an abstract way this means that items enter into a system to inquire a service and each item leaves the system after termination of its service. The queuing process $X = \{X(t), t \geq 0\}$ records the number of items in the system, where $X(t)$ denoted the number if items waiting for a service and in service at time $t \geq 0$. Kendall [88] introduced a standard classification of a queue according to its main features. A queue is summarized as $F^B_{IA}/F_S/c/C_C/C_P/D$:

(i) Here F_{IA} denoted the cdf of the inter-arrival times IA_n $n \geq 1$ (the time between two successive arrivals) where the arrival time of the n-th bath of items is given by $A_n = \sum_{i=1}^{n} IA_i$. A standard assumption is that $IA_n \overset{iid}{\sim} F_{IA}$.

(ii) B is random variable with support $\mathbb{N}$, the positive integers. Where $B_n \overset{\mathcal{L}}{=} B$ are iid copies of B with bath size cdf F_B. If $A_n = t$ then

$B_n = k$ indicates that at time t a bath of k items enters the queuing system. It is assumed that the $B_n = k$ items enter are ordered within the queue, starting with the the first fortunate item in the n-th bath up to the most unfortunate (k-th) item.

(iii) F_S denotes the service time cdf. The n-th item which enters the system has a service-time of length $S_n \overset{iid}{\sim} F_S$. Here n indicates the index of a single item not the bath-index.

(iv) $c \in \mathbb{N} \cup \{+\infty\}$ denotes the number of counter (service stations) of the system.

(v) $C_C \in \mathbb{N} \cup \{+\infty\}$ denote the total waiting space of the system (the storing space for items which are put on hold).

(vi) Whereas $C_P \in \mathbb{N} \cup \{+\infty\}$ denotes the total number of the items which can inquire a service (the risk population which can enter the system).

(vi) D denotes the service policy of the system. In most real live systems this is a $FIFO$, first in first out policy. But there may be other policies such as the $LIFO$, last in first out policy or a $RANDOM$, choose an item randomly for a service policy.

Most systems are by default of the form $C_F = C_p = +\infty$, $D = FIFO$ and $B \sim \delta_1(\cdot)$. In this case the system is abbreviated as $F_{IA}/F_S/c$. For a system with exponential distributed inter-arrival times (or service times) the cdf F_{IA} (or F_S) is denoted by M, where M stands for the memoryless property of the exponential distribution.

The simplest queuing model is the $M/M/1$ queue. This is a one counter model with exponential distributed inter-arrival times and service times of rate $\mu > 0$ and $\lambda > 0$. The process is a simple birth-and-death process with birth rate μ and death rate λ. The model has closed form expression for the transition probabilities and the invariant distribution, but is usually fare too simplistic to be of much use in reality. For queuing models with non-exponential inter-arrival or service times the finite state transition probabilities or other finite-state characteristics do usually have no closed

form expression. This is because these models are neither Markov nor semi-Markov processes. A way to get around this problem is to analyze queuing processes by their imbedded Semi-/Markov chains or by supplementary variable techniques [89, 121, 14, 109]

6.2 Relevant literature

Queuing models have a broad range of applications in operation research, electronic engineering, physics, biology and many further fields. Therefore scientific analysis of queuing models has a long history, starting in the early 1930's and 1940's and is still an active field of research. The majority of these studies are purely probabilistic with little attention to statistical analysis. Armero and Bayarri [9] gave a complete listed a reference on Bayesian analysis of queuing models until 1996 which contained seven article, three of them by Armero itself. In comparison to thousands of papers on Bayesian analysis of regression models this number is negligible small.

Nonetheless, in recent years there has been a growing interest in statistical analysis of queuing systems. In principle a queuing process supply an endless amount of data and any consistent estimation procedure will do just fine. But unlike most statistical estimation problems a queuing process produces continuously data. Since queuing process are generally not Semi-/Markov, different experimental designs and statistical ways of recording or summarizing the evolution of a queuing process will procure different likelihoods. A Bayesian analysis reflects the experimental design trough Bayes theorem. Hence the same queuing process with two different experimental designs may lead to different posterior even for identical prior. In most statistical problems the object is estimation of hypothesis testing. In queuing models the statistical problem is rather prediction. Prediction does not mean predicting the next observation, but predicting certain summary measures of the systems.

Since the mid-1990s also Bayesian analyses of Queuing models have become more popular. The early articles focus mostly on simple Markovian queuing [7, 10, 8, 9, 11, 12]. Such models involve only two parameters the

birth and death rate μ and λ usually re-parameterized in terms of the traffic intensity ($\rho = \mu/\lambda, \lambda$) with restricted Gamma prior on $\{(x, y) \in [0, +\infty) : x < y\}$ for μ and λ. Extension to bulk queues are obtains by [11] assuming the bath-size to be $B_n \overset{iid}{\sim} Geom(p)$. Inference was implemented based on MCMC and numerical-inversion methods. Hence even for the simple $M^{Geom}/M/1$ queues , which requires the bath-sizes, inter-arrival-times and service-times to have the memoryless property, a closed-form Bayesian analysis is intractable.

There are only a few Bayesian studies for Non-Markovian queuing models. Most of them focus on the $M/G/1$ model or the dual model $G/M/1$, where G denotes a generic inter-arrival time cdf (or service time cdf). Wiper [155] conducted a Bayesian analysis for inter-arrival times with Erlangen distribution and memoryless service times, whereas [27] models the service times with an inverse-Gaussian distribution. A semi-parametric Bayesian approach based in a finite mixture model of Exponential, Gamma and Erlangen distributions for the service times was studied in [77, 156, 15]. Semi-parametric analysis for the $G/G/1$ models was considered in [16]. They model the inter-arrival times and the service times with a mixture of convolutions of exponential distributions (called Coxian distributions). Recently Sutton and Jordan [146] proposed a simulation based approach to Bayesian inference for queuing networks. The only non-parametric Bayesian analysis of queuing models we are aware of is Conti [32, 33] who analysis the discrete-time $Geo/G/1$ queue, where G the discrete service time distribution on $\mathbb{N}$ was modeled with the classical Dirichlet random probability on $\mathbb{N}$.

6.3 Outline of Part II

In the following two chapters we analyze the $M/G/1$ queue and the $M^X/G/1$ bulk queue in a Bayesian non-parametric framework. In the first chapter we study the $M/G/1$ queue using as statistical design the imbedded Semi-Markov chain of the $M/G/1$ queue and some tools of the theory of de Finetti's theorem for Semi-/Markov chains developed in [51, 44]. In particular we define a continuous-time jump process predictively to approximate the

Semi-Markov process of the $M/G/1$ queue. To be of some use for statistical estimation, we will need to disregard some aspects the Semi-Markov process of the $M/G/1$ queue and focus on a simple statistical learning procedure. The advantage of our procedure is that our statistical estimation procedures do not require time consuming MCMC simulations. Since for queuing processes data-sets ranges easily between 5,000 to 500,000 observations the later approach can requires several hours or even several days of computing time, whereas our approach will take a few minutes of computing time regardless of the sample-size.

In the second chapter we analyze the $M^X/G/1$ queue using non-parametric prior processes for the batch-size distribution and a semi-parametric model for the service time distribution. The implementation is similar to parametric Bayesian models. We use a combination of several simulation tool, like a Bootstrap scheme, MCMC and numerical inversion methods for Laplace transforms. The second approach is computational much more demanding, bit it is also much more flexible. In principle any steady-state performance measure can be estimated easily. This is not possible with the somewhat inflexible de Finetti's theory of Semi-Markov processes.

Chapter 7

Approximate Prediction in the $M/G/1$ model via mixture of Semi-Markov chain.

7.1 Introduction

In the following we consider a non-parametric framework to the analysis of the $M/G/1$ model. The design we consider is the embedded Semi-Markov process where the process is summarized through the departure times of the queue and the number of arrivals during successive departures. The statistical analysis is based on a process constructed in a predictive way with the aim to approximate the embedded Semi-Markov process of the $M/G/1$ queue. Using de Finetti's theorem we show that this is equivalent to a Bayesian non-parametric analysis of the $M/G/1$ using discrete- and continuous time Neutral-to-the-right processes.

The outline of the paper is as follows. In section 7.2 we recall the well-known embedded Semi-Markov process of the $M/G/1$ queue. Section 7.3 introduces a process which we will use to analysis the $M/G/1$ queue and establishes some basic properties of the constructed process. In section 7.4 we will use the process for statistical analysis the $M/G/1$ process. We also provide a numerical illustration through a simple simulation experiments

for some simple $M/G/1$ queues. The final section 7.5 is devoted to some concluding remakes and some critical aspects of the proposed analysis.

7.2 The M/G/1 Queue and its imbedded Semi-Markov Process

Consider a one counter queue where customers arrive according to a homogeneous Poisson process with rate λ, with independent and identical distributed service times with service time distribution G on $[0, \infty)$. We denote with $X(t)$ the number of items in the queuing system at time $t \geq 0$.

In the special case where the service time law is an exponential distribution $G = Exp(\mu)$ with mean $1/\mu$, the process $X = (X(t), t \geq 0)$ can be expressed as a Semi-Markov process through the jump chain and holding times. Denote the jump chain with $Y = \{Y_n\}_{n\geq 0}$ and the holding times with $T = \{T_n\}_{n\geq 1}$. Let $Z = \{Z_n\}_{n\geq 1}$ represents the sequence of time points where an item enters or departures from the system, i.e. $Z_n = \sum_{l=1}^{n} T_l$. The jump chain Y indicates the sequence of states immediately after the time points T. Then (Y, T) is semi-Markov and the transition kernel is given by

$$\begin{aligned} Q_i(\{j\}, [0,t]) &= P(Y_{n+1} = j, T_{n+1} \leq t | (Y_j, T_j)_{j=0}^{n}) \\ &= \begin{cases} 1 - e^{-\lambda t} & \text{if } Y_n = 0, j = 1 \\ \dfrac{\mu}{\mu+\lambda}(1 - e^{-(\lambda+\mu)t}) & \text{if } j = Y_n - 1 \geq 0 \\ \dfrac{\lambda}{\mu+\lambda}(1 - e^{-(\lambda+\mu)t}) & \text{if } j = Y_n + 1 \geq 2 \\ 0 & \text{otherwise .} \end{cases} \end{aligned} \tag{7.1}$$

Furthermore one can recover X at time $t \geq 0$ by $X(t) = Y_n$ if $Z_n \leq t < Z_{n+1}$. If the service time distribution G is not exponential, then (Y, T) is unfortunately neither Markov nor Semi-Markov. But one can introduce a second process $Y = (Y(t), t \geq 0)$ that equals to X at a certain sequence of random points in time. Define the process $Y = \{Y(t), t \geq 0\}$ as follows. Denote with S_n the n-th departure epoch, i.e. the random point in time

immediately after the n-th item terminates his service and leaves the system. Let

$$Y(t) = Y_n \text{ if } S_n \leq t < S_{n+1} \tag{7.2}$$

where now $Y_n = X(S_n)$ and $S_n = \sum_{j=1}^{n} \triangle S_n$ such that the random time $\triangle S_n$ denotes the n-th intra-departure time. The process is well defined if $G(0) = 0$, $G(\infty) = 1$ and $0 < \lambda < \infty$. Both X and Y are identical at departure epochs but differ between departure epochs. By definition of Y, the process satisfies the relation

$$Y_n = \min(Y_{n-1} - 1, 0) + A_n \tag{7.3}$$

where A_n denotes the number of arriving items during the n-th service time (see for example [138], p.89 or [132], p.71).

Since the arrival process of a $M/G/1$ model is Markov, the number of arriving costumers between successive service times are iid. If G would be known, say up to unknown parameters, then $a_K := P(A_1 = k) = E[(\lambda S)^k e^{-\lambda S}]/j!$ for $k \geq 0$ where $S \sim G$. Hence the embedded process $\{Y_n\}$ is Markov with transition matrix

$$\Pi = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ 0 & 0 & 0 & a_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}. \tag{7.4}$$

In practice, if the model G is unknown or difficult to determine, we do not know Π. Still it is possible to determine, at least empirically, whether the queue is ergodic or transient. Suppose for the moment that G is known and hence Π is known as well. Then the embedded jump chain is positive recurrent if there exists a probability distribution $\pi = (\pi_i)$ on the non-negative integers such that $\pi\Pi = \pi$. This (see for example [121]) translates into the

requirement that

$$E[A_1] \begin{cases} > 1 & Y \text{ is transient} \\ = 1 & Y \text{ is null recurrent and} \\ < 1 & Y \text{ is positive recurrent.} \end{cases}$$

The last requirement can, at least empirically, be check more easily than finding π if the model G is unknown.

7.3 Approximate analysis of $M/G/1$ Queues with mixtures of Semi-Markov processes

The heuristic and somewhat naive idea of this section is to replace the embedded Semi-Markov model of the $M/G/1$ queue

$$\begin{pmatrix} S_n \\ Y_n \end{pmatrix} = \begin{pmatrix} S_{n-1} + \Delta S_n \\ \min(Y_{n-1} - 1, 0) + A_n \end{pmatrix} \tag{7.5}$$

where $\Delta S = (\Delta S_n)_{n\geq 0}$ and $A = (A_n)_{n\geq 0}$ are two independent iid sequences by a mixture of Semi-Markov processes. But both ΔS and A are taken to be exchangeable. For the exchangeable case the transition probability of the vector (S_{n+1}, X_{n+1}) will depends on the whole history $\mathcal{F}_n = \sigma(\{(X_j, S_j)\}_{0\leq j\leq n})$. Hence it is possible to learn sequentially from the past about the future behavior of the embedded process of the $M/G/1$ queue. The next subsection describes the detailed construction of the process whereas section 7.3.2 presents some properties of the constructed process.

7.3.1 Construction of the process

We proceed as follows. For $i = 1, 2$, let $\alpha^{(i)}$ be a measure on the non-negative real line equipped with the Borel σ-field $\mathbb{B}([0, \infty))$ and $\beta^{(i)} : [0, \infty) \to (0, \infty)$ be a measurable step function. For easy of exposure we decompose the measure $\alpha^{(i)}$ into a discrete part $\alpha_d^{(i)}$ and a continuous part $\alpha_c^{(i)}$. The set

$\mathcal{J}^{(i)} = \{t_j\}_{j\geq 1} = \{t \geq 0 : \alpha^{(i)}\{t\} > 0\}$ will denote the collection of points of discontinuity of $\alpha^{(i)}$, i.e. the domain of $\alpha_d^{(i)}$. Assume that $\mathcal{J}^{(i)}$ is countable, so that

$$\alpha^{(i)}[0,t] = \alpha_c^{(i)}(0,t] + \sum_{j:t_j\in\mathcal{J}^{(i)},t_i\leq t} \alpha_d^{(i)}(\{t_i\}).$$

For $i = 1, 2$, we will assume that $0 \notin \mathcal{J}^{(i)}$ and that $\alpha^{(i)}$ and $\beta^{(i)}$ satisfy

$$\prod_{t_j\in\mathcal{J}^{(i)}} \left(1 - \frac{\alpha_d^{(i)}\{t_j\}}{\alpha_d^{(i)}\{t_j\} + \beta^{(i)}(t_j)}\right) \exp\left\{-\int_{[0,\infty)} \frac{\alpha_c^{(i)}(ds)}{\beta^{(i)}(s)}\right\} = 0. \tag{7.6}$$

Furthermore let $s = \{s_n\}_{n\geq 0}$ and $m = \{m_n\}_{n\geq 0}$ denote two non-negative functions with domain $\mathbb{N}_0$, where $\mathbb{N}_0$ consists of the set of non-negative integers. Assume that for all $k \geq 0$ s and m satisfy $s_k < m_k$ and $\sup_j m_j < \infty$ and also $\inf_j s_j > 0$. Furthermore also assume that

$$\mu(m,s) := \sum_{k\geq 0} \prod_{j=0}^{k} \frac{m_j - s_j}{m_j} \in (0,1). \tag{7.7}$$

We now define a sequence of increments for a renewal process and a biased random walk. We start with the renewal process first. For $i = 1, 2$ and $n = 1$ let $\Delta S_1^{(i)}$ be a random variables on $[0,\infty)$ with complementary cdf (ccdf) given by

$$\begin{aligned}&\mathbb{P}\Big(\Delta S_1^{(i)} > t\Big)\\ &\quad= \prod_{j:t_j\in\mathcal{J}^{(i)},t_j\leq t} \left(1 - \frac{\alpha_d^{(i)}\{t_j\}}{\beta^{(i)}(t_j) + \alpha_d^{(i)}\{t_j\}}\right) \exp\left\{-\int_{(0,t]} \frac{\alpha_c^{(i)}(ds)}{\beta^{(i)}(s)}\right\}.\end{aligned} \tag{7.8}$$

Furthermore for $n \geq 1$ and given the σ-filed $\mathcal{S}_n^{(i)} = \sigma(\{\Delta S_j^{(i)}\}_{1\leq j\leq n})$ define the random variable $\Delta S_{n+1}^{(i)}$ having ccdf

$$\begin{aligned}
&\mathbb{P}\Big(\Delta S_{n+1}^{(i)} > t|\mathcal{S}_n^{(i)}\Big) \\
&\quad = \prod_{j:t_j\in\mathcal{J}_n^{(i)},t_j\leq t}\left(1-\frac{\alpha_{n,d}^{(i)}\{t_j\}}{\beta^{(i)}(t_j)+Y_n^{(i)}(t)+\alpha_d^{(i)}\{t_j\}}\right) \\
&\quad \times \exp\left\{-\int_{(0,t]}\frac{\alpha_c^{(i)}(ds)}{Y_n^{(i)}(s)+\beta^{(i)}(s)}\right\},
\end{aligned} \tag{7.9}$$

where $Y_n^{(i)}(t) = \sum_{1\leq j\leq n} I(\Delta S_j^{(i)} \geq t)$. Furthermore the discrete part of $\alpha^{(i)}$ is updated by the counting process $N_n^{(i)}(t) = \sum_{1\leq j\leq n} I(\Delta S_j^{(i)} \leq t)$ via

$$\alpha_{n,d}^{(i)}\{t_j\} = \alpha_d^{(i)}\{t_j\} + dN_n^{(i)}(t_j), \tag{7.10}$$

where $\alpha_d^{(i)}\{t_j\} = 0$ if $t_j \notin \mathcal{J}^{(i)}$ and the set of discontinuity points is updated to $\mathcal{J}_n^{(i)} = \mathcal{J}^{(i)} \cup \{\Delta S_j^{(i)}\}_{1\leq j\leq n}$. Note that assumption (7.6) and $0 \notin \mathcal{J}^{(i)}$ assures that all predictive cdf's are proper and without positive mass at the origin, i.e $\mathbb{P}(\Delta S_n \in \{0, +\infty\} \;\; \exists\, n \geq 1) = 0$.

Parallel to $\Delta S = \{(\Delta S_n^{(1)}), (\Delta S_n^{(2)})\}_{n\geq 1}$, we define a second sequence $A = \{A_n\}_{n\geq 1}$ as follows. For $n = 1$, let A_1 be a discrete random variable with values in $\mathbb{N}_0$ and probability mass function

$$\mathbb{P}(A_1 = k) = \frac{s_k}{m_k}\prod_{j=0}^{k-1}\left(1-\frac{s_j}{m_j}\right) \quad \text{for } k \in \mathbb{N}_0. \tag{7.11}$$

For $n \geq 1$ we define A_{n+1}, given $\mathcal{A}_n = \sigma(\{A_j\}_{1\leq j\leq n})$, by

$$\mathbb{P}(A_{n+1} = k|\mathcal{A}_n) = \frac{s_k + v_n(k)}{m_k + w_n(k)}\prod_{j=0}^{k-1}\left(1-\frac{s_j + v_n(j)}{m_j + w_n(j)}\right) \tag{7.12}$$

where we keep track of the counting process $v_n(k) = \sum_{i=1}^n I(A_i = k)$ and the risk process $w_n(k) = \sum_{i=1}^n I(A_i \geq k)$ for $k \in \mathbb{N}_0$. We assume ΔS and A to be independent and use the filtration $(\mathcal{F}_n)_{n\geq 0}$ where $\mathcal{F}_n = \sigma(\mathcal{S}_n^{(1)}, \mathcal{S}_n^{(2)}, \mathcal{A}_n)$

for $n \geq 1$ and $\mathcal{S}_0^{(1)}, \mathcal{S}_0^{(2)}, \mathcal{A}_0$ denote the corresponding trivial sigma-algebras. Finally, similar to (7.5), define the $(\mathcal{F}_n)_n$ measurable process $(X, S) = (\{X_n, S_n\}; n \geq 0)$ as follows. For $n \geq 1$, let

$$X_{n+1} = \max(X_n - 1, 0) + A_{n+1} \tag{7.13}$$

$$S_{n+1} = \begin{cases} S_n + \Delta S_{h(n)}^{(1)} & \text{if } X_n = 0 \\ S_n + \Delta S_{n+1-h(n)}^{(2)} & \text{if } X_n \geq 1 \end{cases} \tag{7.14}$$

where $h(n) = \sum_{t=0}^{n} I(X_t = 0)$ denotes the number of hits to 0 in the first n transitions and we set $S_0 = 0$ and $X_0 = a_0 \in \mathbb{N}_0$. Hence the sequence S constitutes a renewal sequence with non-iid renewal increments and X constitutes a random walk with bias of one (if X is greater than zero) and non-iid drifts A. The interpretation of the process (X, S) is similar to the embedded Semi-Markov process of an $M/G/1$ queue. Even those they are not the same things. We will stress this fact further in the later part. The distinction between the holding times for the non-idle and idle case is not necessary, but has advantages for statistical applications and prediction purposes later on. The definition (7.5) models the departure-epochs marginally without conditioning on the size of the queue. But (7.5) could have been written in the same conditional form as $\Delta S = E + \Delta\tilde{S}$, if the queue is idle, where $E \sim Exp(\lambda)$ and the service time $\Delta\tilde{S} \sim G$ or as $\Delta S = \Delta\tilde{S}$ of the queue is non-idle. In the following we shortly deriving a representation of the process (X, S) as a mixture of Semi-Markov processes and use of this property for predictive estimation in the $M/G/1$ model.

7.3.2 Some basic properties of the constructed process

We consider first the process X separately. The transition probability of X_{n+1} given $\mathcal{F}_n$ can be computed by noticing that

$$\begin{aligned}
&\mathbb{P}(X_{n+1}=j|\mathcal{F}_n)\\
&=\mathbb{P}(A_{n+1}=j-X_n+I(X_n\geq 1)|\mathcal{F}_n)\\
&=\begin{cases}
\left(\frac{s_j+v_n(j)}{m_j+w_n(j)}\right)\prod_{k=0}^{j-1}\left(1-\frac{s_k+v_n(k)}{m_k+w_n(k)}\right) & \text{if } j\geq X_n=0\\
\left(\frac{s_{j-x_n+1}+v_n(j-x_n+1)}{m_{j-x_n+1}+w_{j-x_n+1}(n)}\right)\prod_{k=0}^{j-x_n}\left(1-\frac{s_k+v_n(k)}{m_k+w_n(k)}\right) & \text{if } j\geq X_n-1\geq 0\\
0 & \text{otherwise}
\end{cases}
\end{aligned}$$

where $\prod_{j=0}^{-1}:=1$ and we can rewrite the counting- and risk-process as

$$v_n(k)=\sum_{1\leq i\leq n}I(X_i=X_{i-1}+k-1+\delta_{X_{i-1}}(0))=\sum_{1\leq i\leq n}I(A_i=k) \quad (7.15)$$

$$w_n(k)=\sum_{1\leq i\leq n}I(X_i\geq X_{i-1}+k-1+\delta_{X_{i-1}}(0))=\sum_{1\leq i\leq n}I(A_i\geq k). \quad (7.16)$$

Since the process X has a restricted set of possible transition, i.e. from $i\geq 1$ to $i-1,i,i+1,\cdots$ and from $i=0$ to $\mathbb{N}_0$, we will introduce admissible sequence of states to rule out some null events. To be specific we will call a sequence $(j_k)_{k=0}^n$ of non-negative integers admissible if for every $1\leq k\leq n$ whenever $j_{k-1}>0$, then the successor state satisfies $j_k\geq j_{k-1}-1$. We will need to determine the finite dimensional law of the discrete process X. This establishes at the same time an invariance property of the process. We recall that two integer valued strings of length $n+1$, say $\tau=(\tau_0,\cdots,\tau_n)$ and $\zeta=(\zeta_0,\cdots,\zeta_n)$, are called equivalent if $\tau_0=\zeta_0$ and for every pair of non-negative integers s,v the number of transitions from s to v among successors in τ equals to the same number of transitions in ζ, i.e. $t_\tau(s,v):=\sum_{k=0}^{n-1}I(\tau_k=s,\tau_{k+1}=v)=\sum_{k=0}^{n-1}I(\zeta_k=s,\zeta_{k+1}=v)=:t_\zeta(s,v)$.

Lemma 28. *(i) For every $n \geq 1$ and every sequence of states $(j_k)_{k=0}^n$*

$$\mathbb{P}(\cap_{k=0}^n \{X_k = j_k\}) = \begin{cases} \prod_{l \geq 0} \frac{s_l^{[v_n(l)]}(m_l - s_l)^{[h_n(l)]}}{m_l^{[w_n(l)]}} & \text{if } (j_k)_{k=1}^n \text{ is admissible} \\ 0 & \text{otherwise} \end{cases}$$

where $a^{[n]} = a(a+1)\cdots(a+n-1)$ and $h_l(n) = w_l(n) - v_l(n)$.
(ii) The process X is partial exchangeable according [44], i.e. for every $n > 0$ and every couple of equivalent sequences of length $n+1$, say τ and ζ, of elements in $\mathbb{N}_0$

$$\mathbb{P}(\cap_{k=0}^n \{X_k = \tau_k\}) = \mathbb{P}(\cap_{k=0}^n \{X_k = \zeta_k\}). \tag{7.17}$$

The proof of this and all remaining result are given in the appendix.

Partial exchangeable processes have an interesting property which will be of use for us. If in addition X is also recurrence, meaning $\mathbb{P}(X_n = X_0 \text{ } i.o.) = 1$, then by Diaconis and Freedman's representation theorem the process X can be represented as a mixture of Markov chains. To obtain a representation for (X, S) as a mixture of Semi-Markov chains, we need the following lemma which states the recurrence of the discrete process X and already yields a representation of X as a mixture of Markov chains.

Lemma 29. *For every fixed initial value $x_0 \in \mathbb{N}_0$, $\mathbb{P}(X_n = x_0 \text{ } i.o.) = 1$.*

Having established the fact that X is a recurrent, partial exchangeable process, we now want to derive a joined mixture representation of (X, S). For this purpose let us define first, for every state $i \in \mathbb{N}_0$, the sequence of hitting times to i by setting $\tau_i(1) = \inf\{m \geq 0 : X_m = i\}$ and for $n > 1$

$$\tau_i(n+1) = \inf\{m > \tau_i(n) : X_m = i\}. \tag{7.18}$$

Hence $\tau_i(n)$ denote the n-th time ar which X visits the state i, where $\tau_i(n+1) = +\infty$ if $X_k \neq i$ for all $k > \tau_i(n)$. Furthermore, following [51], we

introduce the sequence of successor states of i by

$$V_i(n) = \begin{cases} X_{\tau_i(n)+1} & \text{if } \tau_i(n) < \infty \\ \infty & \text{otherwise ;} \end{cases} \tag{7.19}$$

together with the sequence of holding times to state i

$$T_i(n) = \begin{cases} S_{\tau_i(n)+1} - S_{\tau_i(n)} & \text{if } \tau_{i(n)} < \infty \\ +\infty & \text{otherwise.} \end{cases} \tag{7.20}$$

The following lemma gives now a simple way to obtain a mixture representation of (X, S) using row-wise partial exchangeability as an invariance property.

Lemma 30. *Define the matrix (V, T) of successor states and holding times as*

$$(V,T) = \begin{pmatrix} (V_0(1),T_0(1)) & (V_0(2),T_0(2)) & \cdots \\ (V_1(1),T_1(1)) & (V_1(2),T_1(2)) & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix} = \Big\{ \big(V_i(n), T_i(n)\big) \Big\}_{i\geq 0, n\geq 1}.$$

Then (V, T) is row-wise partial exchangeable, i.e. for all $N \geq 0$ and $n > 1$

$$\begin{aligned} &\mathbb{P}\Big(\bigcap_{0\leq i\leq N} \bigcap_{1\leq k\leq n} \big\{ V_i(k) = j_{i,k}, T_i(k) \leq t_{i,k} \big\} \Big) \\ =&\mathbb{P}\Big(\bigcap_{0\leq i\leq N} \bigcap_{1\leq k\leq n} \big\{ V_i(k) = j_{i,\sigma_i(k)}, T_i(k) \leq t_{i,\sigma_i(k)} \big\} \Big) \end{aligned} \tag{7.21}$$

for any permutation σ_i of $\{1, \cdots, n\}$ for $1 \leq i \leq N$.

Now we can proceed identical to [51, 118] and apply de Finetti's theorem for row-wise partial exchangeable arrays to obtain a representation of the probability law of (X, S) as a mixture of Semi-Markov transition kernels. Recall that a continuous time jump process is Semi-Markov if the Jump chain and holding time are jointly Markov on $\mathcal{S} = \mathbb{N}_0 \times [0, \infty)$. Also recall that a Semi-Markov kernel $\mathbf{W} = (W_i, i \in \mathbb{N}_0)$ is a sequence of probability

measures on $\mathcal{S}$. Denote with $\mathbb{W} = M(\mathcal{S})^{\infty}$ the space of all sequences of probability measures on $\mathcal{S}$. We can make $\mathbb{W}$ into a topological space via the product topology, say $\mathcal{W}$, generated though products of open sets of $\mathcal{M}(\mathcal{S})$. $\mathcal{M}(\mathcal{S})$ denotes the topology of weak-convergence on the space of all probability measure on $\mathcal{S}$.

Corollary 8. *(i) There exists an $\mathbb{W}$-valued random element $\boldsymbol{W}$ such that for any $n \geq 1$, every admissible sequence $(i_k)_{0\leq k\leq n}$ of states and all positive real numbers $(t_k)_{1\leq k\leq n}$*

$$\mathbb{P}\Big(\bigcap_{0\leq k\leq n}\{X_k = i_k, S_k - S_{k-1} \leq t_k\}\big|X_0 = i_0\Big) = \mathbb{E}\Big[\prod_{1\leq k\leq n} W_{i_{k-1}}(\{j_k\},[0,t_k])\Big].$$

(ii) The random Semi-Markov kernel $\boldsymbol{W} = \{W_i(\cdot,\cdot)\}_{i\geq 0}$ has the form

$$W_i(\{j\},[0,t]) = \begin{cases} \Pi_{0,j}F^{(1)}(t) & \text{if } i = 0, t \geq 0 \\ \Pi_{i,j}F^{(2)}(t) & \text{if } i \geq 1, t \geq 0; \end{cases} \tag{7.22}$$

where $\Pi_{i,j} = Q_{j-i+1-\delta_0(i)}$ if $j \geq i - 1 + \delta_0(i) \geq 0$ and zero otherwise; and

$$Q_j = \theta_j \prod_{k<j}(1-\theta_j)\ k \geq 0,\ \text{for } \theta_j \overset{ind.}{\sim} Beta(s_j, m_j - s_j). \tag{7.23}$$

Furthermore, for $i = 1, 2$, the process $F^{(i)}$ is a Beta-Stacy process on $\mathbb{R}_+$ with parameters $(\beta^{(i)}, \alpha^{(i)})$, i.e.

$$1 - F^{(i)}(t) = \exp\{-Z_c^{(i)}(t)\} \prod_{j:t_j\in\mathcal{J}^{(i)}, t_j\leq t}(1-\zeta_j^{(i)}) \quad \text{for } t \in [0,\infty]. \tag{7.24}$$

Where $\zeta_j^{(i)} \overset{ind.}{\sim} Beta(\alpha_d^{(i)}\{t_j\}, \beta^{(i)}(t_j))$ random variables and $Z_c^{(i)}$ is an independent increment process on $\mathbb{R}_+$ with Levy measure

$$v(ds,dt) = \frac{\exp\{-s\beta^{(i)}(t)\}}{1-\exp\{-s\}} ds\alpha_c^{(i)}(dt). \tag{7.25}$$

Remark 18. *Given $\boldsymbol{W}$, our constructed process (X, S) behaves similar to the embedded Semi-Markov process of the $M/G/1$ queue, meaning that it*

mimics the basic relations (7.5) conditional on the distinction between non-/idle states as described above. But the constructed process purposely lacks other characteristics of such an embedded process.

Most importantly, in the embedded Semi-Markov process of a $M/G/1$ model, the distribution function $F^{(1)}$ is the convolution of the service time cdf G with an exponential cdf of mean $1/\lambda$. Which we do not consider here. Secondly, in the embedded Semi-Markov process of a $M/G/1$, the discrete distribution $Q = \{Q_j, j \geq 0\}$ is the marginal of a compound Poisson distribution, compounded over the rate $\lambda \times S$ where $S \sim G$. Which we do not model either.

Disregarding this two points and modeling $Q, F^{(1)}$ and $F^{(2)}$ basically independent has advantages for statistical applications, discussed in the next section. These simplification in our process, where we do not model random-convolutions and random-compounding, retain tractable calculations. Given the model as stated in the previous corollary and it marginal specification via the predictive definition, we sequentially learn about our targets (i) Q simply from X and (ii) about $F^{(2)}$ simply from (S_n) with the additional information that $X_n > 0$.

7.4 Statistical Application

In the present section we discuss the application of the process constructed in the last section to statistical estimation problems of some characteristics of the $M/G/1$ model. Suppose data for an $M/G/1$ process are observed until the N-th departure epoch S_N and summarized into the embedded Semi-Markov process $(X, S)|_N := \{(X_n, S_n); 0 \leq n \leq N\}$. Based on our prior information we want to predict some characteristics of the system without knowing the exact sampling model of service times or the arrival rate. As new observations arrive we want to update our inferential scheme rather fast.

Number of items at departure epochs and Service time

Predicting the number of items at the next departure epoch can be done directly using the predictive rule of the discrete process X. In particulate, the transition probability from $i \geq 0$ items to $j \geq i - 1 + \delta_0(i)$ items, given the past $\mathcal{F}_N$, is given by

$$\begin{aligned}\mathbb{E}[\Pi_{i,j}|\mathcal{F}_N] &= \mathbb{P}(X_{N+1} = j|X_N = i, \mathcal{F}_n) \\ &= \begin{cases} \dfrac{s_{j-i+1-\delta_0(i)} + v_N(j - i + 1 - \delta_0(i))}{m_{j-i+1-\delta_0(i)} + w_N(j - i + 1 - \delta_0(i))} \\ \times \prod_{l=0}^{j-i-\delta_0(i)} \left(1 - \dfrac{s_l + v_N(l)}{m_l + w_N(l)}\right) & \text{if } j \geq i - 1 + \delta_0(i) \geq 0 \\ 0 & \text{otherwise.} \end{cases}\end{aligned} \tag{7.26}$$

Similar the tail probability of more than $j \geq i-1+\delta_0(i)$ items in the system at the next departure epoch is given by

$$\mathbb{P}(X_{N+1} > j|X_N = i, \mathcal{F}_n) = \prod_{l=0}^{j-i+1-\delta_0(i)} \left(1 - \frac{s_l + v_N(l)}{m_l + w_N(l)}\right). \tag{7.27}$$

The expected number of items at the next departure epoch, given the information $\mathcal{F}_n$, is given by

$$\begin{aligned}\mathbb{E}[X_{N+1}|\mathcal{F}_N] &= X_N + 1 - \delta_0(X_N) + \mathbb{E}[A_{N+1}|\mathcal{F}_N] \\ &= X_N + 1 - \delta_0(X_N) + \sum_{j\geq 0} \prod_{l=0}^{j} \left(1 - \frac{s_l + v_N(l)}{m_l + w_N(l)}\right).\end{aligned} \tag{7.28}$$

The distribution function of the service-times given the current information can be estimated by

$$\hat{F}_S(t) = \mathbb{E}[F^{(2)}(t)|\mathcal{F}_N] = 1 - \prod_{j:t\geq t_j\in\mathcal{J}^{(2)}} \left(1 - \frac{\alpha^{(2)}_{d,N}\{t_j\}}{\beta^{(2)}(t_j) + Y^{(2)}_N(t_j) + \alpha^{(2)}_d\{t_j\}}\right) \times \exp\left\{-\int_{(0,t]} \frac{\alpha^{(2)}_c(dx)}{\beta^{(2)}_N(x) + Y^{(2)}_N(x)}\right\}. \tag{7.29}$$

The evaluation of (7.29) for varying t is quite tedious in practice. In the actual numerical implementation, see the next section, we prefer to sample repeatedly $F^{(2)} \sim \mathbb{P}(dF^{(2)}|\mathcal{F}_N)$ from the posterior. In this way we can provide Highest-Posterior-density (HPD) credibility sets for the service-time cdf F_S and the mean service time $\mathbb{E}[\Delta S^{(2)}|F_S]$. We developed a simple and fast algorithm to sample from the posterior $\mathbb{P}(dF^{(2)}|\mathcal{F}_N)$. The detailed simulation steps of the algorithm are described in the appendix.

Traffic intensity and arrival rate

Given that we know the "true" service time model of the $M/G/1$ process the traffic intensity and the arrival rate of the process are given by

$$\rho = \mathbb{E}[A_n|Q] \quad \text{and} \quad \lambda = \frac{\mathbb{E}[A_n|Q]}{\mathbb{E}[S|F_S]}. \tag{7.30}$$

The expected traffic intensity, given the current information $\mathcal{F}_N$ is given by

$$\hat{\rho}_N = E[A_N|\mathcal{F}_N] = \sum_{j\geq 0}\prod_{l=0}^{j}\left(1 - \frac{s_l + v_N(l)}{m_l + w_N(l)}\right). \tag{7.31}$$

Since the traffic intensity is a crucial summary measure of a queuing system we may want to express our uncertainty about ρ as well. Quantification of the uncertainty connected with the estimator can be obtained by repeatedly simulate from the posterior: for $c = 1, \cdots, C$

(i) sample $\theta^{(c)}_j \sim \text{Beta}(s_j + v_N(j), m_j + w_N(j) - v_N(j))$ for $j \geq 0$ and

(ii) set $\rho^{(c)} := \sum_{j\geq 0}\prod_{l=1}^{j}(1-\theta_l)$.

In this way we approximate the posterior law of ρ by $\mathcal{L}(\rho|\mathcal{F}_N)$ and construct a HPD-credibility interval.

The expected arrival rate $E[\lambda|\mathcal{F}_N]$ of the process, given the current information, will in general be different from the plugging-estimator $E[A_{N+1}|\mathcal{F}_N]/\mathbb{E}[S|\mathcal{F}_N]$. But, since we already have draws $\rho^{(c)} = \mathbb{E}[A_{N+1}|Q^{(c)}]$ and $\mathbb{E}[S|F_S^{(c)}]$ from the previous calculations at our disposal, we can approximate the posterior probability and point-/interval estimator without any further computational expense by

$$\mathcal{L}(\lambda|\mathcal{F}_N) \approx C^{-1} \sum_{1\leq c\leq C} \delta_{\rho^{(c)}/\mathbb{E}[S|F_S^{(c)}]}(\cdot). \tag{7.32}$$

Invariant distribution and lengths of busy periods in equilibrium

Finally, having estimated the traffic intensity $\rho = \mathbb{E}[A_1|Q] = \lambda\mathbb{E}[\Delta S^{(2)}|F_S]$, we may formally decide whether the system is stable (meaning ergodic) and reaches a stationary level as $S_N \to \infty$. It $\mathbb{P}(\rho \geq 1|\mathcal{F}_N) \approx C^{-1}\sum_{1\leq c\leq C} I(\rho^{(c)} \geq 1) < \gamma$ for some predefined small value γ, then we may accept that $\rho < 1$ [1] and want to predict the invariant distribution of the $M/G/1$ system.

Note that for Poisson arrivals, departing items "see" the same items left behind as arriving items "see" at arrival. And Poisson arrivals "see" time averages [109], p.257 and p.273. Hence the invariant distribution of the $M/G/1$ queue is the same as the invariant distribution of the embedded Markov chain (X_n) of the $M/G/1$ queue (see [109], p.273).

The invariant distribution ψ which solves $\psi\Pi = \psi$, for

$$\Pi = \begin{pmatrix} Q_0 & Q_1 & Q_2 & \cdots \\ Q_0 & Q_1 & Q_2 & \cdots \\ 0 & Q_0 & Q_1 & \cdots \\ 0 & 0 & Q_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{7.33}$$

[1] More formal we may fix a weighted $0-1$-loss function first, with weighs a_0, a_1 for false rejection of $\rho < 1$ and false acceptance of $\rho < 1$ and set $\gamma = a_0/(a_0 + a_1)$. In this case we would accept or reject based on a Bayes rule with weighted $0-1$-loss function.

with $Q_k = \theta_k \prod_{l<k}(1-\theta_l)$, can be computed recursively. After some simple algebraic manipulations we can express the invariant distribution as

$$\psi_0 = 1 - \rho = 1 - \sum_{j\geq 0} S(j) \text{ and} \tag{7.34}$$

$$\psi_k = \frac{\psi_0 S_{k-1} + \sum_{i=1}^{k-1} \psi_i S(k-i)}{Q_0} \text{ for } k \geq 1, \tag{7.35}$$

where $S(k) = \sum_{j\geq k+1} Q_j = \prod_{j=1}^{k}(1-\theta_j)$.

Hence we may obtain realizations from the law of the random invariant distributions ψ which corresponds the law of the random transitions matrix Π simply by using the already simulated realizations $Q^{(c)}$: for $c = 1, \cdots, C$

(i) sample $\theta_j^{(c)} \sim \text{Beta}(s_j + v_N(j), m_j + w_N(j) - v_N(j))$ for $j \geq 0$ and

(ii) set $\rho^{(c)} := \sum_{j\geq 0} \prod_{l=1}^{j}(1-\theta_l)$

(iii) if $\rho^{(c)} < 1$, set $Q_0^{(c)} = \theta_0^{(c)}$, compute $S^{(c)}$ as above and compute $\psi^{(c)}$ recursively according to (7.34) and (7.35).

Finally, if the "true" service time law and arrival rate would be known, then the mean length of the busy period in equilibrium is given by

$$\mu(Q, F_S) := \mathbb{E}[\text{non-idle Period}|F_S, Q] = \frac{\mathbb{E}[S|F_S]}{1 - \mathbb{E}[A|Q]}. \tag{7.36}$$

Without any further computational expense, we can estimate this quantity, given our current information $\mathcal{F}_N$, by using the already simulated random quantities

$$\mathcal{L}(\mu(Q, F_S)|\mathcal{F}_N) \approx \frac{1}{C^*} \sum_{c:\rho^{(c)}<1} \delta_{\mathbb{E}[S|F_S^{(c)}]/(1-\rho^{(c)})}(\cdot) \tag{7.37}$$

where C^* denotes the number is simulation with $\rho^{(c)} < 1$.

Table 7.1: Model characteristics

Model	parameters	traffic intensity ρ	transition row Q
$M/Exp(\mu)/1$	$(\lambda, \mu) = (.3, .4)$	$\lambda/\mu = .75$	Geom$(\mu/(\mu+\lambda))$
$M/U(t_1, t_2)/1$	$(\lambda, t_1, t_2) = (.3, 1, 4)$	$.5\lambda(t_1 + t_2) = .75$	Poisson-Uni(λ, t_1, t_2)
$M/Gam(a, b)/1$	$(\lambda, a, b) = (.3, 5, 2)$	$\lambda a/b = .75$	NegBi$(a, \lambda/(\lambda+b))$

7.4.1 Numerical illustration

In this subsection we shortly illustrate the statistical estimation methods proposed in the last section applied to three simple $M/G/1$ models. In particular consider the three models summarized in table 7.1.

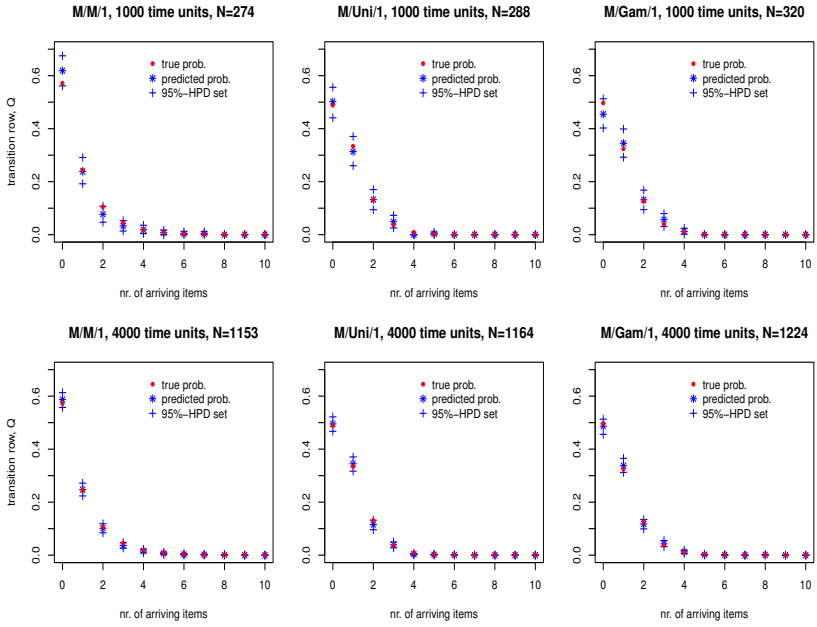
The first model is the classical $M/M/1$ queue with exponential service time with mean 2.5. The remaining two $M/G/1$ models have a service time uniformly distributed between one and five and a Gamma distributed service time with shape parameter five and rate two. For all models, the mean service time is identical equal to 2.5 and we take the same arrival rate of $\lambda = .3$. Therefore all three models are ergodic with traffic intensity $\rho = .75$.

We simulated one $M/G/1$ trajectory for all tree models for a process time S_N the first departure epoch after $t = 1000, 4000$ units. For each of the six cases (tree models for two lengths of observation) we estimate (i) the transition row, (ii) the service time complementary-cdf, (iii) the mean service time, (iv) the traffic intensity and (v) the invariant distribution together with (vi) the mean length of the busy period. For all tree models we center the service-time cdf on a Weibull model by choosing $\alpha(0, t] = \alpha_c(0, t] = vt^a$ and $\not{\beta}t) = 1$, i.e, $\mathbb{P}(\Delta S_1^{(2)} > t) = \exp\{-vt^a\}$. The parameters (a, v) where determined by an empirical Bayes procedure. Furthermore we center the initial distribution of (A_n) on a Geometric distribution with mean .9 by setting $s_j \equiv .1$ and $m_j \equiv .19$.

Transient-State

Figure 7.1 shows the point estimator and the 95-percent point-wise credibility intervals (Highest-posterior density) for all three models. In all figures the first row represents the smaller sample size and the second row the estimates

Figure 7.1: Predictive Estimator of Q, the number of new arrivals between two successive departure times for $T = 1000$ time units at the first row and for $T = 4000$ time units at the second row.



based on the bigger sample size. The red dots indicate always the true values. The estimator seems quite all-right especially for the bigger sample size with a substantial reduction of variation around the true values. The same holds for the estimated complementary service-time cdf shown in figure 7.2. Even those the model was initially always centred on a Weibull distribution the data point very quickly towards the right model. The point estimator are practically identical to the true values with a higher variation around the true service time ccdf for the $M/M/1$ model for the small sample size. For the bigger sample size almost all simulated trajectories of the complementary service-time cdf are basically "identical" to the true complementary service-

Figure 7.2: Predicted complementary-cdf of the service-times, for $T = 1000$ time units at the first row and for $T = 4000$ time units at the second row.

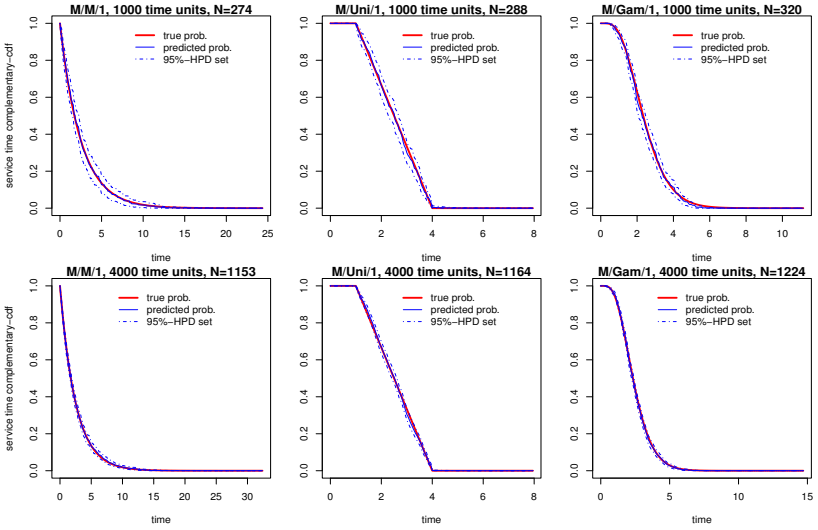
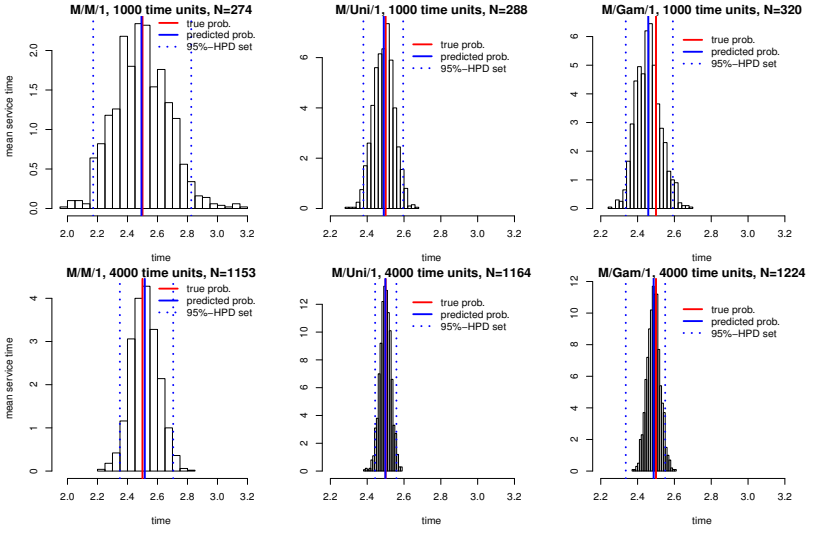
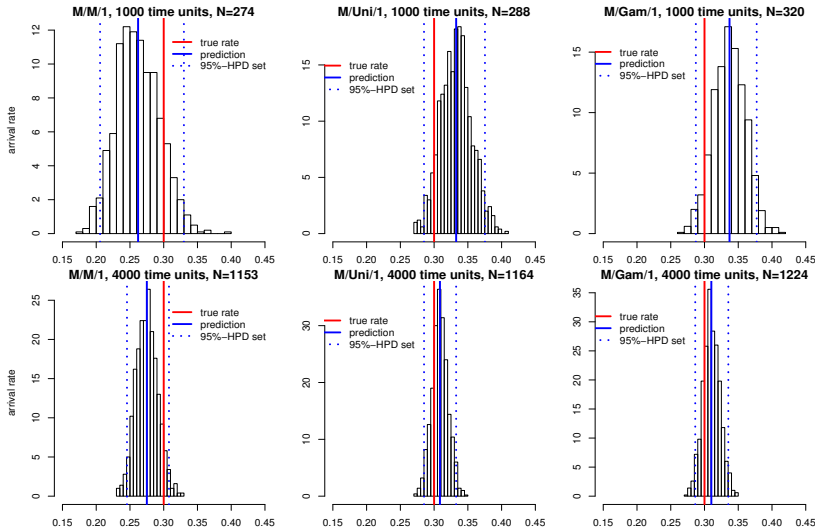


Figure 7.3: Predicted mean service-time of the $M/G/1$ queue, for $T = 1000$ time units at the first row and for $T = 4000$ time units at the second row.



time cdf. The same picture can be found for the mean service time. Even for the small sample size the point estimator are basically identical to the true value. Again the widest credibility set can be found for the $M/M/1$ model and a small sample size. But an increase in sample size reduces the length of the intervals by more than one third. A somewhat different picture can be

Figure 7.4: Predicted arrival rate of the $M/G/1$ queue, for $T = 1000$ time units at the first row and for $T = 4000$ time units at the second row.



found for the estimation of the arrival rates, which are indirectly estimated trough the relation (7.30) by (7.32). Since we divide at every iteration the simulated traffic intensities by the simulated service-time means, we require a higher sample size to reduce variation. Except for the $M/M/1$ queue, we generally overestimate the arrival rate. But this bias is reduced when we increase the sample size. Again the precision of the estimator of the arrival rate is lower in the $M/M/1$ model than in the other models.

Steady-state

We now turn to the estimation of the traffic intensity and the steady-state performance measure.

Figure 7.5 shows the histogram for the simulated posterior distribution of the traffic intensity. Recall that the prior mean of the intensity was set to .9. We observe the same patters as before. The point estimators are closer to the true estimator for the non-Markovian queues. As the sample size increases the posterior mean comes again close to the true traffic intensity of .75 except for the $M/M/1$ model.

In all six cases the posterior draws concentrate with more that 95-percent on the interval $(0, 1)$. For the larger sample size all draws lie within $(0, 1)$ and the maxima are substantially bounded away from one. This explains the histogram of the MC-draws from the approximate posterior of the mean busy period at steady state. The histogram for the lower sample size shows a long tail to the right, which can be explained by the former observation since $(1 - \rho^{(c)})^{-1}$ becomes quite large for draws of the traffic intensity close to one. Since for the larger sample size all MC-draws of the traffic intensity are bounded away from one the histogram becomes much more symmetric and close to the true mean busy period.

We close the simulation study by looking at the invariant distribution of the number of items in steady-state. Incidentally, we recover the invariant distribution for the M/Uni/1 queue for a lower sample size almost perfectly. But there is generally a substantial amount of uncertainty around the point estimator. The increase in sample size improves the quality of estimation, both variance and precision, substantially. This should be expected, since the queue will be closer to its steady-state equilibrium at the end of the observation for the larger sample size compared to the lower sample size.

7.5 Conclusions

In this present paper we used the theory of de Finetti's theorem for Semi/-Markov chains and non-parametric prior processes to implement a non-

Figure 7.5: Predicted traffic intensity of the $M/G/1$ queue, for $T = 1000$ time units at the first row and for $T = 4000$ time units at the second row.

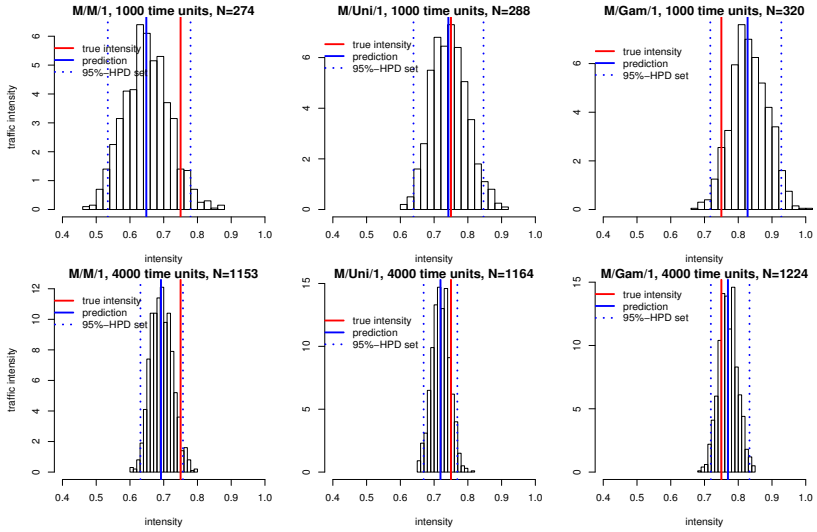


Figure 7.6: Predicted mean busy period at steady state, for $T = 1000$ time units at the first row and for $T = 4000$ time units at the second row.

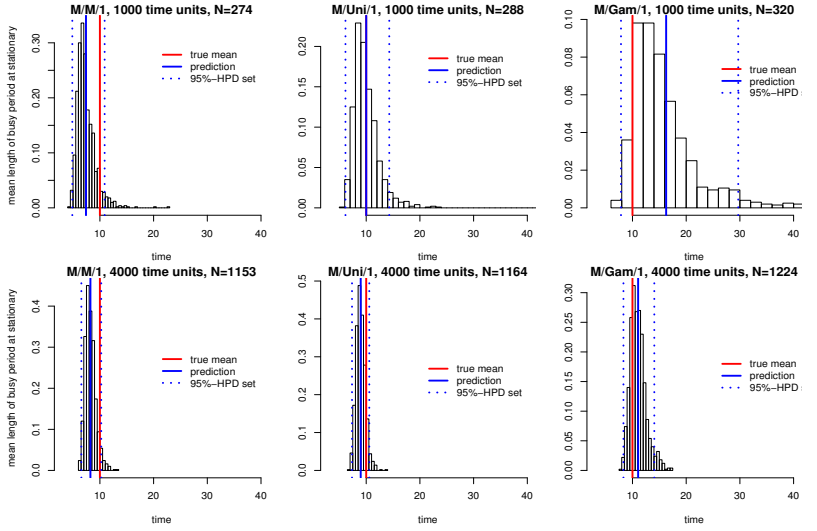
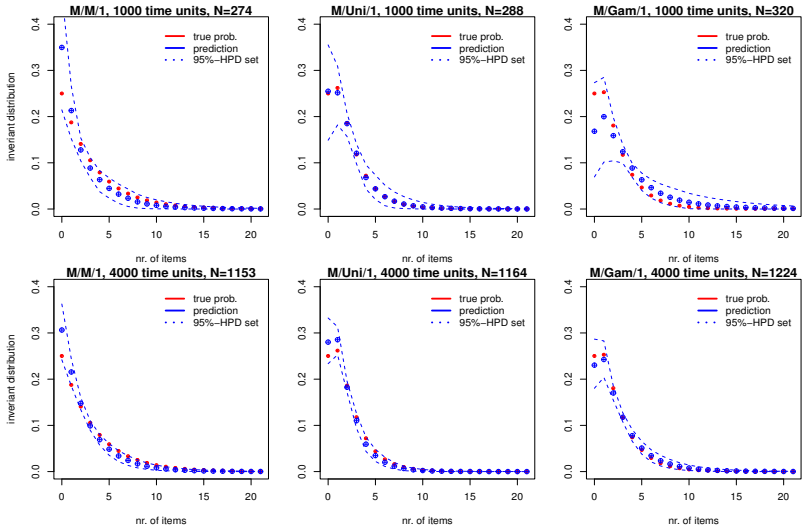


Figure 7.7: Predicted invariant distribution of the $M/G/1$ queue, for $T = 1000$ time units at the first row and for $T = 4000$ time units at the second row.



parametric analysis for the $M/G/1$ queue. We derived estimator for some main characteristics of the $M/G/1$ process for transitive- and steady-state characteristics. The analysis can be implemented with simple MC-simulations and avoids therefore time-consuming MCMC schemes. The computational costs are only a few minutes and is basically indifferent to the sample size. This is clearly superior to other Bayesian semi-parametric approaches based on Mixture models where the computational time increases substantially with the increase of sample size. On the down side the non-parametric nature of our approach requires a certain amount of data to pin-point the estimator and, more important, the credibility sets to the "true" model. This may not be much of a problem in applications since a queuing system produces very quickly a huge amount of data. A more substantial, steady-state characteristics, which are often only characterizes through generating functions, like

the cdf of the waiting in the queue before service, cannot be handled with our approach easily.

Future research will be directed toward the $M^X/G/1$ bulk queue with limited population size and to queuing networks. Both problems can be analyzed using a similar approach as the one suggested in the current paper. This work is presently ongoing.

7.6 Appendix

7.6.1 Simulating trajectories from a Beta-Stacy process

In this subsection we summarize the computational steps to simulate a trajectory of a random cumulative distribution function (cdf) from a Beta-Stacy Process (BSP) with parameters α and β For detailed information on the BSP see [148]. α will be a positive Borel measure decomposed into an absolutely continuous part α_c and a discrete part α_d with discontinuity set $\mathcal{J}$. βwill be a positive piecewise continuous function with domain $[0, +\infty)$. A random cdf F on the non-negative real line is called a BSP, denoted by $F \sim \text{BSP}(\alpha, \beta)$, if $S = 1 - F = \exp\{-Z_c - Z_d\}$ where Z_c and Z_d are two non-decreasing independent increment processes on $\mathbb{R}_+$. The process $Z_c = \sum_j Z_{c,j}\delta_{T_j}$ is an independent increment process (without deterministic or Gaussian component) on $[0, \infty)$ with Levy measure

$$\nu(ds, dt) = k(s|t)ds\alpha_c(dt) = \frac{\exp\{-s\beta(t)\}ds}{1 - \exp\{-s\}}\alpha_c(dt). \tag{7.38}$$

This means that Z_c is a pure jump process with random jumps of sized $Z_{c,j} \geq 0$ at random locations $T_j > 0$ [57, 158, 157, 149, 99].

Whereas the later process $Z_d = \sum_{t_j \in \mathcal{I}} Z_{d,j}\delta_{t_j}$ is a jump process with random jumps at fixed locations $\mathcal{J} = \{t_j\}$, such that the jump sizes are mutually independent and $Z_{d,j}$ has density

$$f_{J_i}(s) = \frac{\Gamma(\alpha_d\{t_j\} + \beta(t_j))}{\Gamma(\alpha_d\{t_j\})\Gamma(\beta(t_j))}(1 - e^{-s})^{\alpha_d\{t_j\}-1}e^{-s\beta(t_j)}. \tag{7.39}$$

Which is equivalent to $W_j := 1 - e^{-Z_{d,j}} \sim \text{Beta}(\alpha_d\{t_j\}, \beta(t_j))$. Therefore we can represent $S = 1 - F$ as

$$S(t) = \prod_{t_j \leq t}(1 - W_j)\exp\{-Z_c(t)\} \quad \text{for } t \geq 0. \tag{7.40}$$

The posterior distribution

Now suppose that $X_1, \cdots, X_n$ is an iid sample from F and $F \sim BSP(\beta\alpha)$ then $F_n := [F|(X_i)_1^n]$ is again a BSP such that $\log(1-F_n) = Z_{n,d} + Z_{n,c}$. The continuous part $Z_{n,c}$ has Levy density

$$v_n(ds, dt) = e^{-sY(t)}k(s|t)ds\alpha_c(dt) = \frac{\exp\{-s(\beta t) + Y_n(t))\}}{1-\exp\{-s\}}ds\alpha_c(dt)$$

where $Y_n(t) = \sum_{1\le i\le n} I(X_i \ge t)$. Whereas the discrete part $Z_{n,d} = \sum_{j:t_j\in\mathcal{J}_n} Z_{d,j}\delta_{t_j}$ has random jumps of size $Z_{d,j}$ at deterministic locations $t_j \in \mathcal{J}_n = \mathcal{J} \cup \{X_i\}_{i\le n}$ with jump-size density

$$f_{t_j}(s) \propto (1-e^{-s})^{dN_n(t_j)+\alpha_d\{t_j\}-1}e^{-s(\beta t_j)+Y_n(t_j)-dN_n(t_j))} \tag{7.41}$$

where $N_n(t) = \sum_{1\le i\le n} I(X_i \le t)$.

Simulation via Transformation

For simulation purposes we note the following neat fact. Let $Z_c = \sum_j Z_{c,j}\delta_{T_j}$ be as described before. Making a change of variable $H = \sum_j (1-e^{-Z_{c,j}})\delta_{T_j}$ the new process H has Levy measure $v_H(ds, dt)$ given by

$$v_H(ds, dt) = \frac{k(-\log(1-s)|t)}{1-s}\alpha_c(dt) = \beta(t)s^{-1}(1-s)^{\beta(t)-1}\frac{\alpha_c(dt)}{\beta(t)}. \tag{7.42}$$

This is the Levy measure of a Beta process [74] with parameters β and α_c/β say $H \sim \mathrm{BP}(\beta, \alpha_c/\beta)$. Hence we can use the same trick as used in [99] for simulating the homogeneous process. Suppose we want to simulate trajectories of the posterior cdf F_n. Instead of simulating F_n directly (meaning simulate $Z_{n,d}$ and $Z_{n,c}$) we can simulate a Beta process $H = \sum_j H_j\delta_{T_j} \sim \mathrm{BP}(\beta_n, \alpha_c/\beta_n)$ where $\beta_n = \beta + Y_n$ and set $Z_c = \sum_j Z_{c,j}\delta_{T_j}$ where $Z_{c,j} := -\log(1-H_j)$. We sample the actual Beta process by the ϵ-truncations method of [99]. All computational steps are summarized in figure 7.8. The algorithm is stated for generic $\beta\alpha_c$. To simulate from the posterior we only need to replace β by $\beta + Y_n$ and $\mathcal{J}$ by $\mathcal{J}_n$.

Figure 7.8: Simulation of a Beta-Stacy Process $F = (F(t), 0 \leq t \leq T)$ by the transformation method from a Beta process

1. Fix $\epsilon > 0$ small, $T > 0$ large and sample $M \sim \text{Poisson}(\epsilon^{-1}\alpha_c(0,T])$.
2. Sample $T_j \overset{iid}{\sim} f(t) = \dfrac{\alpha_c(dt)}{\alpha_c(0,T]} I_{(0,T]}(t)$ for $i = 1, \cdots, M$.
3. Given $\{T_i\}_{i=1}^M$, sample $H_j \sim \text{Beta}(\epsilon, \beta T_{(j)})$ for $j = 1, \cdots, M$.
4. Set $Z_{c,\epsilon} = \sum_{j=1}^M -\log(1-H_j)\delta_{T_{(j)}}$.
5. Sample $W_j \sim \text{Beta}(\alpha_d\{t_j\}, \beta t_j)$ for $t_j \in \mathcal{J}$.
6. Set $F_\epsilon(t) = 1 - \left(\prod_{t_j<t}(1-W_j)\right)\left(\prod_{k=1}^M (1-H_j)^{I(T_j \leq t)}\right)$ for $t \in [0,T]$.

Remark 19. *Using theorem 2 in [100], when* $0 < \inf_{[0,T]} \beta(t) \leq \sup_{[0,T]} \beta(t) < \infty$, *then as* $\epsilon \to 0$, $H_\epsilon := \sum_{j=1}^M H_j \delta_{T_{(j)}} \to BP(\beta, \alpha_c/\beta)$ *in distribution on* $D[0,T]$ *equipped with the Skorohod topology. Now an application of the continuous mapping theorem shows that* $Z_{c,\epsilon} \to Z_c$ *in distribution on* $D[0,T]$ *as* $\epsilon \to 0$.

7.6.2 Proofs

Proof of lemma 28. (i) Let $s_n(j) = s_j + v_n(j)$ and $m_n(j) = m_j + w_n(j)$. Then using the transition law (7.15)

$$\begin{aligned} P(\cap_{k=0}^n \{X_k = j_k\}) &= \prod_{k=1}^n P(A_k = j_k - j_{k-1} - I(j_{k-1} > 0) | \mathcal{F}_k) \\ &= \prod_{k=1}^n \left[\frac{s_{k-1}(j_k)}{m_{k-1}(j_k)} \prod_{i=0}^{j_k-1} \left(1 - \frac{s_{k-1}(i)}{m_{k-1}(i)}\right)\right]^{I(j_{k-1}=0)} \\ &\times \prod_{k=1}^n \left[\frac{s_{k-1}(j_k - j_{k-1} + 1)}{m_{k-1}(j_k - j_{k-1} + 1)} \prod_{i=0}^{j_k - j_{k-1}} \left(1 - \frac{s_{k-1}(i)}{m_{k-1}(i)}\right)\right]^{I(j_{k-1}>0)}. \end{aligned}$$

Simplifying the last expression we obtain

$$= \frac{\prod_{k=1}^{n} s_{k-1}(j_k - (j_{k-1}-1)_+) \prod_{k=1}^{n} \prod_{i=0}^{j_k-(j_{k-1}-1)_+-1} [m_i(k-1) - s_{k-1}(i)]}{\prod_{k=1}^{n} \prod_{i=0}^{j_k-(j_{k-1}-1)_+-1} m_{k-1}(i)}$$
$$= \prod_{j\geq 0} \frac{s_j^{[\tilde{v}_j(n)]}(m_j - s_j)^{[\tilde{h}_j(n)]}}{m_j^{[\tilde{w}_j(n)]}}.$$

(ii) Let τ and η be defined as in lemma 28 (ii), we need to show (7.17). Consider first the case where one sequence, say τ, is inadmissible. Therefore, there exists a $1 \leq k \leq n$ and an element $\tau_{k-1} > 0$ such that $\tau_k < \tau_{k-1} - 1$. But, since $1 \leq t_\tau(\tau_{k-1}, \tau_k) = t_\zeta(\tau_{k-1}, \tau_k)$, there exists some k' such that $\zeta_{k'} < \zeta_{k'-1} - 1$ for $\zeta_{k'-1} = \tau_{k-1} > 0$ and $\zeta_{k'} = \tau_k$. Therefore ζ is inadmissible too and $P(\cap_{k=1}\{X_k = \tau_l\}) = 0 = P(\cap_{k=1}\{X_k = \zeta_l\})$.

Now, consider the case where at least one sequence is admissible, which therefore implies that both τ and ζ are admissible. From part (i) the finite dimensional law of X depends only on the 'counting' processes $v_n(i), h_n(i), w_n(i)$ for $i \in \mathbb{N}_0$ and $w_n(i) = \sum_{k=i}^{\infty} v_n(k)$, $h_n(i) = w_n(i) - v_n(i)$. Therefore, it suffices to show that $v_\tau(i) = v_\zeta(i)$ for $i \geq 0$, where

$$v_\tau(i) = \sum_{k=1}^{n} I(\tau_k = \tau_{k-1} + i - 1 + \delta_{\tau_{k-1}}(0))$$

and $v_\zeta(i)$ is defined similar for ζ. Using the fact that τ and ζ are equivalent we obtain for $i \geq 0$

$$\begin{aligned} v_\tau(i) &= \sum_{k=1}^{n} I(\tau_k = i, \tau_{k-1} = 0) + \sum_{k=1}^{n} \sum_{l=1}^{\infty} I(\tau_k = l + i - 1, \tau_{k-1} = l) \\ &= t_\tau(0, i) + \sum_{l=1}^{\infty} t_\tau(l, l + i - 1) \\ &= t_\zeta(0, i) + \sum_{v=1}^{\infty} t_\zeta(l, l + i - 1) = v_\zeta(i) \end{aligned}$$

This completes the proof. □

To proof lemma 29 we need the following lemma.

Lemma 31. *For each fix state $x \in \mathbb{N}_0$ define*

$$S_n^{(x)} = \prod_{i=0}^{x} \frac{m_i - s_i + h_i(n)}{m_i + w_i(n)} \quad \text{for } n \geq 0. \tag{7.43}$$

(i) Then, as $n \to +\infty$, we have $S_n^{(x)} \xrightarrow{a.s.} S_\infty^{(x)}$.
(ii) Furthermore the random variable $S_\infty^{(x)} > 0$ with probability one.

Proof. First, we show that for a fixed $x \in \mathbb{N}_0$ the sequence $(S_n^{(x)})$ is a martingale with respect to the filtration $(\mathcal{F}_n)$. We may use the fact that $A = (A_k)_{k\geq 1}$ are exchangeable [150, 117]. Hence, given the tail sigma-field of A, say $\mathcal{A}_\infty$ all $(A_k)_{k\geq 1}$ are independent and identical distributed. Now, fix $n \geq 1$, then

$$\begin{aligned}
\mathbb{E}[S_{n+1}^{(x)}|\mathcal{F}_n] &= \mathbb{E}[\mathbb{P}[A_{n+2} > x)|\mathcal{F}_{n+1}]|\mathcal{F}_n] \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[I(A_{n+2} > x)|\mathcal{A}_\infty]|\mathcal{F}_{n+1}]|\mathcal{F}_n] \\
&= \mathbb{E}[\mathbb{E}[I(A_{n+2} > x)|\mathcal{A}_\infty]|\mathcal{F}_n] \quad \text{(by tower the property)} \\
&= \mathbb{E}[\mathbb{E}[I(A_{n+1} > x)|\mathcal{A}_\infty]|\mathcal{F}_n] \quad \text{(by conditional iid)} \\
&= \mathbb{E}[I(A_{n+1} > x)|\mathcal{F}_n] = S_n^{(x)}.
\end{aligned}$$

Since $S_n^{(x)} \in [0, 1]$, an application of the Martingale-convergence theorem shows that $S_n^{(x)}$ convergence a.s. to some random variable, say $S_\infty^{(x)}$. This gives (i).

(ii) Furthermore, applying Proposition 5.2 in [59] also gives the convergence in distribution of $S_n^{(x)}, x \in \mathbb{N}_0$ to a random complementary cdf $S^{(\cdot)} = (S^{(x)}, x \in \mathbb{N}_0)$. From [150, 117] the limit equals $S^{(x)} = \prod_{k=1}^{x}(1 - \theta_k)$ where θ_k are independent $Beta(s_k, m_k - s_k)$ random variables. Therefore

$$P(S^{(x)} > 0) = P\left(\prod_{k=1}^{x}(1 - \theta_k) > 0\right) = \prod_{k=0}^{x} P(\theta_k < 1) = 1$$

since $m_k - s_k > 0$ for $k = 0, \cdots, x$. Since limits are unique a.s., we conclude that $S_\infty^{(x)} > 0$ a.s.. □

Proof of lemma 29. For any positive integer n define the predictive comple-

mentary cdf $S(\cdot|\mathcal{F}_n)$ of A_{n+1} given $\mathcal{F}_n$ as above by

$$S(k|\mathcal{F}_n) = \begin{cases} \prod_{i=0}^{k} \frac{m_i - s_i + h_i(n)}{m_i + w_i(n)} & \text{if } k \geq 0 \\ 1 & \text{if } k < 0. \end{cases} \tag{7.44}$$

Since $S(\cdot|\mathcal{F}_n)$ is a complementary cdf it is clearly non-increasing. Now fix $x_0 \in \mathbb{N}_0$ arbitrary. We can bound the predictive law of X from below by

$$\begin{aligned} &P(X_{n+1} = x_0|\mathcal{F}_n, X_0 = x_0) \\ &= \frac{s_{x_0 - X_n + 1 - \delta_0(X_n)} + v_n(x_0 - X_n + 1 - \delta_0(X_n))}{m_{x_0 - X_n + I(X_n \geq 1)} + w_n(x_0 - X_n + 1 - \delta_0(X_n))} S(x_0 - X_n + 1 - \delta_{X_n}(0)|\mathcal{F}_n) \\ &\geq \frac{\inf_i s_i}{\sup_i m_i + n} S(x_0 + 1|\mathcal{F}_n) \ a.s. \end{aligned} \tag{7.45}$$

where the last inequality follows from the fact that $X_n \geq 0$ a.s. and $S(\cdot|\mathcal{F}_n)$ is non-increasing. Hence we have that

$$\sum_{n=1}^{\infty} P(X_n = x_0|\mathcal{F}_n) \geq \sum_{n=1}^{\infty} \frac{\inf_i s_i}{\sup_i m_i + n} S(x_0 + 1|\mathcal{F}_n) \ a.s. \tag{7.46}$$

Suppose that the sum on the right hand side of (7.46) divergence with probability one. Then we can use Levy's extension of the Borel-Cantelli Lemma [154] to deduce that

$$P(X_n = x_0 \ \text{ for infinitely many n } |X_0 = x_0) = 1. \tag{7.47}$$

Choose $j > x_0$. For every finite n $S(x_n + 1|\mathcal{F}_n) \geq S(j|\mathcal{F}_n) > 0$ $a.s.$. Since the first term in the sum on the right hand side of (7.46) is of order $O(n^{-1})$ it suffices to show that

$$S(j|\mathcal{F}_n) \overset{a.s.}{\to} S_j \quad \text{and} \quad S_j > 0 \ \ a.s.. \tag{7.48}$$

But both facts follow from the previous lemma. □

Remark 20. *Note that if we assume that the process starts initially empty, i.e. $x_0 = 0$, then (7.45) is trivially of order $O(1/n)$ and the process is re-*

current without using any argumentation involving the convergence of the predictive distribution.
One could have shown the general case also by fixing $x_0 = 1$, showing only that $P(\liminf_n S(2|\mathcal{F}_n) > 0) = 1$. The case $x_0 = \{0, 1\}$ would be sufficient since for all positive integers $x_0 \geq 1$ we can use the relation

$$\begin{aligned} P_{x_0}(X_n = x_0 \ i.o.) &= P\bigg(x_0 + n + \sum_{i=1}^{n} A_i - n(0) = x_0 \ i.o.\bigg) \\ &= P\bigg((x_0 + j) + n + \sum_{i=1}^{n} A_i - n(0) = x_0 + j \ i.o.\bigg) \\ &= P_{x_0+j}(X_n = x_0 + j \ i.o.) \end{aligned}$$

where $n(0) = \sum_{0 \leq j \leq n-1} \delta_{X_j}(0)$.

Proof of lemma 30. Lemma 1 (ii) and Lemma 2 state that X is recurrent and partial exchangeable. By Theorem 2 in [60] this is equivalent to the row-wise exchangeability of V, i.e.

$$\begin{pmatrix} V_0(1) & V_0(2) & \cdots \\ \cdots & \cdots & \cdots \\ V_k(1) & V_k(2) & \cdots \end{pmatrix} \stackrel{\mathcal{L}}{=} \begin{pmatrix} V_0(\sigma_0(1)) & V_0(\sigma_0(2)) & \cdots \\ \cdots & \cdots & \cdots \\ V_k(\sigma_k(1)) & V_k(\sigma_k(2)) & \cdots \end{pmatrix} \tag{7.49}$$

for any finite permutation σ_k of $\mathbb{N}$ every $k \geq 0$. Furthermore from [150, 118] $\Delta S^{(i)} = (\Delta S_n^{(i)})_{n \geq 1}$ are exchangeable as well for $i = 1, 2$.

Note that (V, T) are tied together only via $(\tau_i(n); n \geq 1, i \in \mathbb{N}_0)$. We may assume that $j_{0,k} \geq 0$ and $j_{i,k} \geq i - 1$ for all $1 \leq k \leq N$ and $1 \leq i \leq n$, since

otherwise both sides of (7.21) are equal to 0. Then

$$\begin{aligned}
&\mathbb{P}\Big(\bigcap_{0\le i\le N}\bigcap_{1\le k\le n}\{V_i(k)=j_{i,k},T_i(k)\le t_{i,k}\}\Big)\\
&=\mathbb{E}\Big[\mathbb{P}\Big(\bigcap_{0\le i\le N}\bigcap_{1\le k\le n}\{T_i(k)\le t_{i,k}\}\Big|\bigcap_{0\le i\le N}\bigcap_{1\le k\le n}\{V_i(k)=j_{i,k},\tau_i(n)\}\Big)\\
&\qquad\qquad\Big|\bigcap_{0\le i\le N}\bigcap_{1\le k\le n}\{V_i(k)=j_{i,k}\}\Big]\\
&\times\mathbb{P}\Big(\bigcap_{0\le i\le N}\bigcap_{1\le k\le n}\{V_i(k)=j_{i,k}\}\Big)
\end{aligned}$$

By definition of T and using the independent of $\{(\Delta S_n^{(1)}),(\Delta S_n^{(2)})\}$ and X we can rewrite the conditional probability within the conditional expectation as

$$\begin{aligned}
&\mathbb{P}\Big(\bigcap_{0\le i\le N}\bigcap_{1\le k\le n}\{T_i(k)\le t_{i,k}\}\Big|\bigcap_{0\le i\le N}\bigcap_{1\le k\le n}\{V_i(k)=j_{i,k},\tau_i(n)\}\Big)\\
&=\mathbb{P}\Big(\bigcap_{1\le k\le n}\{\Delta S_k^{(1)}\le t_{0,k}\}\bigcap\bigcap_{1\le i\le N}\bigcap_{1\le k\le n}\{\Delta S^{(2)}_{\tau_i(k)+1-h(\tau_i(n))}\le t_{i,k}\}\\
&\qquad\qquad\Big|\bigcap_{0\le i\le N}\bigcap_{1\le k\le n}\{\tau_i(n)\}\Big)\\
&=\mathbb{P}\Big(\bigcap_{1\le k\le n}\{\Delta S_k^{(1)}\le t_{0,k}\}\bigcap\bigcap_{1\le i\le N}\bigcap_{1\le k\le n}\{\Delta S^{(2)}_{(i-1)n+k}\le t_{i,k}\}\\
&\qquad\qquad\Big|\bigcap_{0\le i\le N}\bigcap_{1\le k\le n}\{\tau_i(n)\}\Big)\\
&=\mathbb{P}\Big(\bigcap_{1\le k\le n}\{\Delta S_k^{(1)}\le t_{0,\sigma_0(k)}\}\bigcap\bigcap_{1\le i\le N}\bigcap_{1\le k\le n}\{\Delta S^{(2)}_{(i-1)n+k}\le t_{i,\sigma_i(k)}\}\\
&\qquad\qquad\Big|\bigcap_{0\le i\le N}\bigcap_{1\le k\le n}\{\tau_i(n)\}\Big)
\end{aligned}$$

where the second and third equality follow be exchangeability. Applying the

same steps backward the last equation equals

$$= \mathbb{P}\Big(\bigcap_{1\le k\le n} \{\Delta S_k^{(1)} \le t_{0,\sigma_0(k)}\} \bigcap \bigcap_{1\le i\le N} \bigcap_{1\le k\le n} \{\Delta S^{(2)}_{\tau_i(k)+1-h(\tau_i(k))} \le t_{i,\sigma_i(k)}\} \Big| \bigcap_{0\le i\le N} \bigcap_{1\le k\le n} \{\tau_i(n)\}\Big)$$

$$= \mathbb{P}\Big(\bigcap_{0\le i\le N} \bigcap_{1\le k\le n} \{T_i(k) \le t_{i,\sigma_i(k)}\} \Big| \bigcap_{0\le i\le N} \bigcap_{1\le k\le n} \{V_i(k) = j_{i,\sigma_i(k)}, \tau_i(n)\}\Big).$$

For the last equality we used the fact that V is row-wise partial exchangeable. Now stick the last expression back into the expectation and use again that V is row-wise partial exchangeable we obtain the result. Note that the result could have been shown equivalently by stopping-time results from [84] and its generalization in [5]. □

Proof of Corollary 1. (i) Since (V, T) is row-wise partial exchangeable de Finetti's theorem for row-wise partial exchangeable arrays (see [102, 38, 51, 118]) gives the existence of a $\mathcal{W}$-valued random element $W = (W_i)_{i\in\mathbb{N}_0}$ such that for every state $i \ge 0$ the element W_i belongs to $\mathbb{M}(\mathcal{S})$. And given W the sequence $(V_i(n), T_i(n))_n \overset{iid}{\sim} W_i$ for any $i \in \mathbb{N}_0$. And since $\{V_i(n), T_i(n)|\tau_i(n) = k\} = \{X_{k+1}, S_{k+1} - S_k|X_k = i, \tau_i(n) = k\}$ this gives (i).

Furthermore for any $i \in \mathbb{N}_0$ and for $j \ge i - 1 + \delta_0(i)$, $t \ge 0$ the empirical transition kernel converges weakly

$$\lim n^{-1} \sum_{1\le k\le n} \delta_{(V_i(k),T_i(k))}(j, [0, t]) \overset{\mathcal{L}}{=} W_i(\{j\}, [0, t]) \overset{\mathcal{L}}{=} \Pi_{i,j} F_{i,j}(t). \tag{7.50}$$

Also, for $T_{i,j}(n) = S_{\tau_{i,j}(n)+1} - S_{\tau_{i,j}(n)}$ if $\tau_{i,j}(n) = \inf\{k > \tau_{i,j}(n-1) : X_k = i, X_{k+1} = j\} < \infty$ and $T_{i,j}(n) = +\infty$ otherwise, we have that

$$\Pi_{i,j} \overset{\mathcal{L}}{=} \lim n^{-1} \sum_{1\le k\le n} \delta_{V_i(n)}(j) \quad \text{and } F_{i,j}(t) \overset{\mathcal{L}}{=} \lim n^{-1} \sum_{1\le k\le n} \delta_{T_{i,j}(n)}([0, t]).$$

Form [59, 150, 117]

$$\lim_n n^{-1} \sum_{1\leq k\leq n} I(A_k = j) \stackrel{\mathcal{L}}{=} \lim_n P(A_{n+1} = j|\mathcal{F}_n) \stackrel{\mathcal{L}}{=} Q_j \ a.s. \tag{7.51}$$

where $Q = (Q_j, j \geq 0)$ is the random discrete probability stated in the corollary. Since $A_n \geq 0$ and

$$\{V_i(n) = j\} = \{A_{\tau_i(n)+1} = j - i + I(i > 0)\} \tag{7.52}$$

$$\stackrel{d}{=} \{A_n = j - i + I(i > 0)\} \tag{7.53}$$

where the second equality follows from a stopping-time result of [84, 5], (7.51) gives $\Pi_{i,j} = Q_{j-i+\delta_i(0)}$ if $j \geq i-1+\delta_0(i)$ and zero otherwise. This establishes the distributional form of Π.

Similarly, since $\Delta S^{(2)}$ is exchangeable and independent of X, by the same stopping-time property for exchangeable random variables [5][pp.47], for $i > 0$

$$F_{i,j} \stackrel{\mathcal{L}}{=} \lim_n n^{-1}|\{k \leq n : \Delta S^{(2)}_{\tau_{i,j}(k)+1-h(\tau_{i,j}(k))} \leq \cdot\}| \tag{7.54}$$

$$\stackrel{\mathcal{L}}{=} \lim_n n^{-1}|\{k \leq n : \Delta S^{(2)}_k \leq \cdot\}| \quad (\stackrel{\mathcal{L}}{=} F^{(2)}) \tag{7.55}$$

$$\stackrel{\mathcal{L}}{=} \lim_n n^{-1}|\{k \leq n : \Delta S^{(2)}_{\tau_i(k)+1-h(\tau_i(k))} \leq \cdot\}| \stackrel{\mathcal{L}}{=} F_i. \tag{7.56}$$

Provided that $\tau_i(n), \tau_{i,j}(n) < \infty$ a.s. for every n, this shows that the limit, if it exists, does not dependent on $i > 0$ or $j \geq 0$ and is independent of Π as well.

The fact that the limit, say $F^{(2)}$, exists and is in fact equal the one stated above, corresponds to the fact that the system of predictive distributions of $\Delta S^{(2)}$ uniquely determines the finite dimensional joined law of $\Delta S^{(2)}$. Hence two sequences with the same system of predictive distributions will have the same system of finite dimensional laws. But the sequence of predictive distributions of $\Delta S^{(2)}$ correspond to the system of predictive distributions of an exchangeable sequence of random variable with range space $\mathbb{R}_+$ and Beta-Stacy directing random probability with parameters $(\alpha^{(2)}, \beta^{(2)})$ [150, 118].

Hence, by de Finetti's theorem the limit $F^{(2)}$ is equal in distribution to a Beta-Stacy process with parameters $(\alpha^{(2)}, \beta^{(2)})$. The claim $F_{0,j} = F^{(1)}$ for all $j \geq 0$ where $F^{(1)}$ is Beta-Stacy process with parameters $(\alpha^{(1)}, \beta^{(1)})$ can be shown in a similar way.

To show that $\tau_i(n), \tau_{i,j}(n) < \infty$ a.s. for every n, we use

$$\mathbb{P}\Big(\bigcap_{n\geq 1}\{\tau_{i,j}(n) < \infty\}\Big) = \mathbb{P}(X_n = i \ i.o.[n], V_i(k) = j \ i.o.[k]) \tag{7.57}$$

and $\mathbb{P}(X_n = i \ i.o.[n]) = \mathbb{P}(\cap_{n\geq 1}\{\tau_i(n) < \infty\})$. Now, given $\mathcal{F}_\infty = \cap_{n\geq 0}\mathcal{F}_n$

$$\begin{aligned} P(X_{n+k} = i, \exists k \geq 0 | X_n = x_0, \mathcal{F}_\infty) &= \sum_{k\geq 0}(\Pi^k)_{x_0,i} \\ &\geq \begin{cases} Q_{i-x_0+1-\delta_0(x_0)} & \text{if } i > x_0 \\ (Q_0)^{x_0-1} & \text{if } i < x_0 \end{cases} \end{aligned} \tag{7.58}$$

where $\Pi^k = \Pi^{k-1}\Pi$. Since $Q_t = \theta_t \prod_{l<t}(1-\theta_l)$ and $0 < s_t < m_t < \infty$ for all t, both quantities in the last expression are positive (non-zero) a.s.. Similarly $P(X_{n+k} = x_0, \exists k \geq 0 | X_n = i, \mathcal{F}_\infty) > 0$ a.s.. Hence, given $\mathcal{F}_\infty$, both x_0 and i commute. Since x_0 is recurrence, given $\mathcal{F}_\infty$, state i is recurrent as well, i.e. $P(\cap_{n\geq 1}\{\tau_i(n) < \infty\}|\mathcal{F}_\infty) = 1$ a.s.. Therefore, by Fubini theorem, $P(\cap_{n\geq 1}\{\tau_i(n) < \infty\}) = 1$.

By a similarly argument

$$P(V_i(k) = j|\mathcal{F}_\infty) = P(A_{\tau_i(k)} = j - i + 1 - \delta_0(i)|\mathcal{F}_\infty) = Q_{j-i+1-\delta_0(i)} > 0 \ a.s..$$

Now, use Borel-Cantelli lemma first and then Fubini theorem, to show that $P(V_i(n) = j \ i.o.[n]) = 1$. □

Chapter 8

Semiparametric Bayesian Inference for the $M^X/G/1$ Bulk Queue

8.1 Introduction

Bayesian analysis of bulk queues have be studied previously in [11]. This analysis focuses on parametric inference for the $M^X/M/1$ queue, where the inter-arrival and service-times are exponential distributed and the batch-sizes are geometrically distributed. The model consist only of tree parameters, hence statistical estimation can be implemented easily. Unfortunately, service times and bath-sizes do almost never satisfy the memoryless property. Hence the $M^X/M/1$ queue may not be sufficient to analysis every one-counter bulk queue one encounters in applications.

In the following we consider a semi-parametric analysis of the $M^X/G/1$ bulk queue. The bath-size distribution will be modelled non-parametrically using a discrete-time Beta-Stacy process. The service-time distribution will be assumes to be absolutely continuous and follows a Poisson-Dirichlet process mixture model with gamma kernel. The arrival rate is assumes to follow the standard conjugated prior. We focus on the prediction of performance measure in transient- and steady-state.

The outline of the paper is as follows. In section 8.2 we shortly introduce the main quantities of the $M^X/G/1$ process. We will carefully choose a prior distribution for all unknown quantities of the model and derive the corresponding posterior distribution.

8.2 Inference for the $M^X/G/1$ bulk queue

Consider the following data generating process. At each arrival time $0 < A_1 < A_2 < \cdots$ a batch of $B_1, B_2, \ldots$ items enters a system. By time t a total of $N_B(t) = \sum_{i=1}^{N(t)} B_i$ items entered the system, where $N(t) = \sup\{n : A_n \leq t\}$. Each of the single items, in an arbitrary batch, has a service time S_i independently of the arrival process. We assume a first-in-fist-out (FIFO) service strategy. If we denote with $IA_n = A_n - A_{n-1}$ the n-th intra-arrival time, then the $M^X/G/1$ process assumes the following probability model

$$[IA_n|\lambda] \overset{iid}{\sim} \text{Exp}(\lambda), \tag{8.1}$$

$$[B_n|P_B] \overset{iid}{\sim} P_B \quad \text{and} \tag{8.2}$$

$$[S_n|P_S] \overset{iid}{\sim} P_S \tag{8.3}$$

for $\lambda \geq 0$ and some probability measures P_B and P_S on $\mathbb{N}$ and $[0, +\infty)$, i.e.

$$\mathbb{P}[B_n = k|P_B] = P_B(k) \qquad k \geq 1 \tag{8.4}$$

$$\mathbb{P}[S_n \in (t, t+h]|P_S] = P_S(t, t+h] \qquad t, h \geq 0. \tag{8.5}$$

Also assume that given the triple (λ, P_B, P_S) the sequences (IA_n), (B_n) and (S_n) are independent of each other.

In statistical estimation one frequently considers (for computational convenience) an exponential service time distribution (the $M^X/M/1$ queue) or an Erlangen distributional (the $M^X/E_r/1$ queue) or more generally mixture of Erlangen distribution with a random but finite number of mixture components [10, 77, 139, 11, 156, 15]. A typical model for the bath-size distribution is a geometric law [11]. Apart for computational convenience an exponential service time law and a geometric bath-size distribution may not be appro-

priate for most applications involving a $M^X/G/1$ model. Therefore one may prefers to work with an unspecified batch-size and service time probability. This is most convenient if the observed empirical distribution functions of the batch-sizes and the service-times do not match any standard parametric family of distribution functions.

In the following, we will denote with Θ the parameter space of the $M^X/G/1$ process, i.e.

$$\Theta = \{(\lambda, P_B, P_S) : \lambda \geq 0, P_B \in M_P(\mathbb{N}), P_S \in M_P(\mathbb{R}^+)\} \tag{8.6}$$

where $M_P(\mathcal{X})$ denotes the space of all probability measures on a Polish space $\mathcal{X}$ endowed with the topology of weak convergence [70].

8.2.1 Prior specification

In a Bayesian analysis of the $M^X/G/1$ batch queue one needs to specify a prior probability on the parameter space Θ. The prior information are revised trough the evidence provided by the data, which are sampled from the data generating process (8.1), (8.2) and (8.3). By definition of Θ, this involves specifying prior probabilities on spaces of probability measures. Or differently stated, we treat probability measures as random functions (i.e. as stochastic processes). In the following we specify the following prior on Θ:

1. The arrival rate is distributed according to a Gamma distribution with shape parameter u_0 and rate v_0, $\lambda \sim Gamma(u_0, v_0)$.

2. The batch size distribution $P_B = (P_B(k), k \geq 1)$ is assumed to follow a discrete beta-Stacy process (BSP) with parameters (α, β), that is

$$P_B(k) = V_k \prod_{j=1}^{k-1}(1 - V_j) \quad \text{for } k \geq 1, \tag{8.7}$$

where V_k are independent Beta$(\alpha(k), \beta(k))$ distributed random variables and $\alpha, \beta : \mathbb{N} \to [0, \infty)$ are two functions such that $\prod_{k=1}^{\infty} \frac{\beta(j)}{\beta(j)+\alpha(j)} = 0$ (see [150] for details).

3. The random service time distribution $P_S = \{P_S(A), A \in \mathbb{B}(\mathbb{R}_+)\}$ is a kernel mixture with random mixing distribution G, that is

$$\begin{aligned} P_S(A) &= \int_A \int_{(0,\infty)^2} f_{\text{Gam}}(t|\nu, \nu/\mu) G(d\nu, d\mu) dt \\ &= \sum_{k=1}^{\infty} w_k \int_A f_{\text{Gam}}(t|\nu_k^*, \nu_k^*/\mu_k^*) dt \end{aligned} \tag{8.8}$$

where $f_{\text{Gam}}(t|\nu, \nu/\mu_2)$ denotes the density of a Gamma distribution with shape parameter ν and mean μ. Furthermore the random mixing probability

$$G(\cdot) = \sum_k w_k \delta_{(\nu_k^*, \mu_k^*)}(\cdot) \tag{8.9}$$

is a two parameter Poisson-Dirichlet Process (PDP) on $M_P([0,\infty)^2)$ with parameter $(a, b, G_0(\cdot|\eta))$. That is, for $k \geq 1$

$$w_k = V_k \prod_{j<k}(1 - V_j) \quad \text{with } V_k \sim^{ind} \text{Beta}(1-a, b+ka) \tag{8.10}$$

$$(\nu_k^*, \mu_k^*) \overset{iid}{\sim} G_0(d(\nu_k^*, \mu_k^*)|\eta). \tag{8.11}$$

Where (w_k) and (θ_k) are independent and G_0 is a probability on $[0,\infty)^2$ (see [127, 128] for details on the PDP).

4. The random elements λ, P_B and F_S are independent.

The gamma prior for λ was chosen for convenience. It is the natural conjugated prior for the arrival rate of a homogeneous Poisson process.

There are many potential non-parametric prior for P_B. The simplest alternative to the beta-Stacy process would be a Dirichlet process prior on $M_P(\mathbb{N})$[54, 55], i.e. for all $k > 1$ $(P_B(1), \dots P_B(k), 1 - \sum_1^k P_B(k))$ has a Dirichlet distribution with parameters $(\gamma(1), \dots, \gamma(k), \gamma - \sum_1^k \gamma(j))$ for some finite measure $\gamma(\cdot)$ with total mass $\gamma = \sum_j \gamma(j)$. For example, [32, 33] used this prior for the service time law of the $Geo/G/1$ queue. Our primary motivation comes from the fact that the Dirichlet prior is a special case of

the beta-Stacy prior [150]. But whereas the beta-Stacy prior is conjugated to right-censoring the Dirichlet prior is not [58, 145]. In fact under censoring a Dirichlet prior would yields a beta-Stacy posterior [150]. Furthermore the beta-Stacy process can is basically identical the the random weights of a generic stick breaking prior probability defined in [78], hence it shares many theoretical properties with this class of prior.

The prior for the random service time distribution was chosen by theoretical reasoning and for computational simplicity. The service time distribution is most reasonably modelled as an absolutely continuous distribution. Hence any 'Neutral-to-the-Right' (NTR) random distribution (see [46] for a definition, and [54, 55, 58, 150] for examples) would be inappropriate. NTR random distributions are a.s. discrete at fixed and/or random locations (see [46], corollary 3.2). Alternatively we could have choose a Pólya-tree random distribution [55, 97, 98] which is in theory continuous for some special choices of hyper-parameters. But in practice one would need to terminate an actual infinite partition tree at a fixed level, which makes the random distribution discrete. Furthermore, our actual object is the probability law of performance measures of the queuing system. Hence, we will need to work with the Laplace-Stieltjes transform of the service-time distribution which for a Pólya-tree random measure is not easy to handle. As a last point we mention that we could have chosen also the famous Dirichlet-Process mixture model (DPM) ([56, 103, 52, 53]. But the DPM has a tendency to create a few dominating cluster [1], which can be resolved by using the PDP-mixture model. Furthermore the DPM is a special case of the PDP-mixture model with parameter (a, b, G_0) where $a = 0$. Instead of the usual normal kernel, we used the gamma kernel to avoid a log-transformation of the data. A log-transformation would force us to work in the later part with the Laplace transform of the log-normal distribution which does not have a neat closed form expression. Furthermore the space of continuous distributions is dense in the space if gamma mixtures [14]. Hence we can approximate the service-

[1] In the present context a cluster, say j, represents a mixture components θ_j and the corresponding cluster specific service time law Gamma$(\cdot|\theta_j)$ which may or may not have a physical interpretation as the random service time of the j-th kind of service.

time distribution arbitrary close with a mixture of gamma distributions.

8.2.2 Posterior computation

Suppose that we observe data generated from a $M^X/G/1$ batch queue summarized as $\{(IA_n, B_n)\}_{n=1}^{N_1}$ and $\{S_n\}_{n=1}^{N_2}$. As before IA_n denotes the waiting time between the $(n-1)$-th and n-th arrival and B_n denotes the bath size at the n-th arrival time. Again S_n denotes the n-th service time.

For the prior on the parameter space Θ as specified in the previous subsection, we can write the likelihood as

$$L(\Theta|\mathbf{IA}, \mathbf{B}, \mathbf{S}) \propto \left(\lambda^{N_1} \exp\{-\lambda A_{N_1}\}\right)\left(\prod_{k=1}^{\infty} V_k^{n(k)}(1-V_k)^{m(k)}\right) L(S) \quad (8.12)$$

where, for $i \geq 1$, $n(i) = \sum_1^{N_1} I(B_n = i)$ denotes the number of batches of size i among the first N_1 batches and $m(i) = \sum_1^{N_1} I(B_n > i)$ denote the number of batches of size greater than i. The remaining term $L(S) = \prod_1^{N_2} P_S(dS_i)$ denotes the likelihood of the service-time distribution, where P_S is given in (8.8). Combining the prior and likelihood, the posterior is given by

$$\begin{aligned}
&\mathbb{P}(d\Theta|\mathbf{IA}, \mathbf{B}, \mathbf{S}) \\
&\propto Ga(d\lambda|u_0 + N_1, v_0 + A_{N_1}) \times \left(\prod_{k=1}^{\infty} Beta(dV_k|\alpha(k) + n_k, \not{\beta}k) + m_{\ k})\right) \\
&\quad \times L(S)\Pi(dF_S). \qquad (8.13)
\end{aligned}$$

Clearly by construction, given the data, the arrival rate λ, the batch-size distribution P_B and the service-time distribution P_S are still independent. Furthermore the posterior distribution of λ and the batch-size distribution P_B are given in closed form as a Gamma distribution and a beta-Stacy process, each with updated parameter

$$\begin{aligned}
(u_0, v_0) &\to (u_0 + N_1, v_0 + A_{N_1}) \quad \text{and} \\
(\alpha(\cdot), \not{\beta}\cdot)) &\to (\alpha(\cdot) + n(\cdot), \not{\beta}\cdot) + m(\cdot)).
\end{aligned}$$

The distribution of the service times is difficult to represent in closed form, but we can use Monte-Carlo (MC) methods developed for the DPM and the PDP-mixtrure model in [120, 78, 67, 151, 122, 86] to sample from the posterior law $\mathcal{L}(P_s|\mathbf{S})$ and $\mathcal{L}(\phi(P_s)|\mathbf{S})$ for functionals ϕ of the service time law P_S (see section 8.3.2 below for details).

Recall that our targets are performance measure of the queuing system. Most of these performance measure are characterized directly or indirectly through the corresponding Laplace-Stieltjes transforms (LTS) or probability generating functions (PGF). These transforms are functions of the means and the transform of the batch-size and service-times distribution. Since the posterior of the service-time distribution does not have a closed form, we need to implement the inference by MC-simulations. Hence we need to simulate approximately the mean and the transforms from the posterior of the bath-size and service time distribution. This is described in the next section.

8.3 Approximations to simulate from the posterior of performance measures.

In the next two subsection we shortly discuss how to sample functions of the random batch distribution P_B and the random service time distribution P_S. We require this draws to obtain realizations of the transform of the probability law of a performance measure in steady-state. In the last subsection we summaries how to invert the transform to obtain a realization of the actual probability distribution.

8.3.1 Bootstrap approximation of functional of P_B.

The posterior of the batch-size distribution $P_B(\cdot)$ is known on closed form a circect computation of the random mean $\mu_B = \mathbb{E}[B|P_B]$ and the random probability generating function (PGF), say P_B^*, of P_B is intractable. Both expressions involve a sum over an infinite number of elements. To overcome this problem we could use a simple approximation and truncate the posterior,

i.e fix $\epsilon > 0$ small and choose $H > 1$ large such that

$$\mathbb{E}[P(B_1 > H|P_B)|\mathbf{B}] = \prod_{j=1}^{H} \frac{\not{\beta}(j) + m(j)}{\alpha(j) + \not{\beta}(j) + n(j) + m(j)} \leq \epsilon. \tag{8.14}$$

Now, we could repeatedly sample $P_{B,H}(\cdot) = \sum_{j=1}^{H+1} w_j \delta_j(\cdot)$ where $w_j = \theta_j \prod_{l<j}(1-\theta_l)$ for $1 \leq j \leq H+1$ and $\theta_j \sim \text{Beta}(\alpha(j) + n(j), \not{\beta}(j) + m(j))$ for $1 \leq j \leq H$ and $\theta_{H+1} := 1$. A similar strategy was proposed in [67].

We prefer a different random truncation, which is related to computational strategies used in [115, 78, 116, 79, 33] for the Dirichlet-Multinomial process, the Dirichlet process and the Gamma process. The main idea is to approximate the posterior parameter of the beta-Stacy process by an empirical measure based on an iid sample from the posterior predictive distribution. The difference to the approach in [78, 79] is that we approximate the posterior parameter after model-fitting rather than to approximate the prior parameter before model fitting, whereas [115, 116, 33] applied a similar bootstrap approximation to the Dirichlet posterior.

To be more specific, given the observed batches $\mathbf{B} = (B_j)_1^N$ (for simplicity we set $N_1 = N$), the predictive distribution is given by

$$P_{B,N}(k) = \mathbb{E}[P_B(j)|\mathbf{B}] = \frac{\alpha_N(k)}{v_N(k)} \prod_{l=1}^{k-1} \frac{\beta_N(l)}{v_n(l)} \qquad \text{for } k \geq 1, \tag{8.15}$$

where $\alpha_N(k) = \alpha(k) + n(k)$, $\beta_N(k) = \beta_k + m(k)$ and $v_N(k) = \alpha_N(k) + \beta_N(k)$. Now fix M large and consider the following procedure.

1. Sample $B_1^*, \ldots, B_M^*$ iid from $P_{B,N}$ and denote with $M(k) = \sum_1^M I(B_j^* = k)$ and $\bar{M}(k) = \sum_1^M I(B_j^* > k)$ the counting and risk process.

2. For $k \geq 1$ set

$$\alpha_M^*(k) = \begin{cases} v_N(k)\dfrac{M(k)}{\bar{M}(k-1)} & \text{if } \bar{M}(k-1) > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{8.16}$$

$$\beta_M^*(k) = \begin{cases} v_N(k)\dfrac{\bar{M}(k)}{\bar{M}(k-1)} & \text{if } \bar{M}(k-1) > 0 \\ 0 & \text{otherwise .} \end{cases} \tag{8.17}$$

3. Simulate $Q_{B,M} = (Q_{B,M}(j), j \geq 1) \sim \text{BSP}(\alpha_M^*, \beta_M^*)$. Where $Q_{B,M}(j) = \theta_{M,j}^* \prod_{l<j}(1-\theta_{M,l}^*)$ and $\theta_{M,j}^* \sim \text{Beta}(\alpha_M^*(j), \beta_M^*(j))$.

4. Compute the mean and PGF of $Q_{B,M}$, i.e.

$$\mu_{B,M} = 1 + \sum_{k\geq 1}\prod_{1}^{k}(1-\theta_l^*) \quad \text{and} \quad Q_{M,B}^*(s) = \sum_k s^k Q_{B,M}(k) \quad s \in [0,1].$$

5. Repeat step (1) to (4) C times to obtain a sample of approximate draws from the posterior $\mathcal{L}(P_B|\mathbf{B})$, $\mathcal{L}(\mu_B|\mathbf{B})$ and $\mathcal{L}(P_B^*|\mathbf{B})$.

The main point of the algorithm is that for any iteration c after sampling $(B_i^{*(c)})_1^M$, the random probability equals $Q_{B,M}^{(c)}(k) = 0$ a.s. for $k > \sup((B_i^{*(c)})_1^M)$. This makes computation tractable. But since we do not truncated in advance at a fixed H as in (8.14), the Bootstrap will explore once in a while some value $Q_{M,B}^{(c)}(j) > 0$ a.s. for $j > H$. Furthermore notice that since $Q_M^{*(c)}(\cdot)$ is a finite polynomial a.s., the map $s \to Q_M^{*(c)}(s)$ is finite for any compact interval $s \in [0,a]$ for any $a \in [0,\infty)$ fixed.

We want to justify the proposed method in the sense that, if M becomes large, then a single bootstrap draw become an exact draw from the true posterior. The following three lemmas state this fact for $Q_{M,B}$, $\mu_{B,M}$ and $Q_{B,M}^*$. The proofs of the lemmas are given in the appendix.

Lemma 32. *Let $Q_{B,M}(\cdot)$ denote the random bootstrap probability measure, for a sample of M replicas at a fixed bootstrap iteration. Then as $M \to +\infty$, given $\boldsymbol{B}$, $Q_{B,M}(\cdot) \to P_B(\cdot)$ weakly in $M_P(\mathbb{N})$.*

Lemma 33. *Denote with $Q_{B,M}$ the random bootstrap probability measure defined above and let $\mu_{M,B}$ denote the random mean corresponding to $Q_{M,B}$. Then as $M \to +\infty$, given $\boldsymbol{B}$, $\mu_{B,M} \xrightarrow{d} \mu_B$ in distribution.*

Lemma 34. *Let $Q_{B,M}$ be the random bootstrap probability measure defined above and denote with $Q^*_{M,B}(\cdot)$ the probability generating function corresponding to $Q_{M,B}$. Then as $M \to +\infty$, given $\boldsymbol{B}$, $Q^*_{B,M}(\cdot) \to P^*_B(\cdot)$ weakly in $C([0,1])$ with respect to the uniform metric.*

8.3.2 MCMC approximation of functions of P_S.

In this subsections we summarize the computational steps we need to sample (approximately) from the posterior $\mathcal{L}(\phi(P_S)|\mathbf{S})$ for functions of the random service-time distribution P_S defined in (8.8). Such a function is actually a function of the posterior law of the random mixing probability G with respect to the data $\mathbf{S}$, i.e.$\mathcal{L}(\phi(P_S)|\mathbf{S}) = \mathcal{L}(\phi^*(G)|\mathbf{S})$. We need to simulate the Laplace-Stieltjes transform and the mean from the posterior, i.e.

$$\mathbb{E}[S|F_S] \sim \mathcal{L}\Big(\int \mu G(d(\nu,\mu))|\mathbf{S}\Big) \qquad \text{and}$$
$$\mathbb{E}[e^{-(\cdot)S}|P_S] \sim \mathcal{L}\Big(\int \Big((1+\cdot\mu/\nu)^{-\nu} G(d(\nu,\mu))|\mathbf{S}\Big).$$

The computational steps of the algorithm are basic modifications of algorithms suggested in [120, 78, 67]. For simplicity we can write the sampling model of the service-times S_i in a hierarchical form as

$$\begin{aligned}
[S_i|\nu_i,\mu_i] &\overset{ind.}{\sim} \text{Gam}(\nu_i, \nu_i/\mu_i) && \text{for } i = 1,\cdots,N\\
[\chi_i|G] &\overset{iid}{\sim} G && \text{for } i = 1,\cdots,N\\
G &\sim \text{Poisson-Dirichlet-Process}(a,b,G_0(\cdot|\eta)).
\end{aligned}$$

where $\chi_i = (\nu_i, \mu_i)$. Furthermore we note that for any measurable and integrable functional ϕ with domain $M_P([0,\infty)^2)$ the following relation holds

$$\int e^{-u\phi(G)}\mathbb{P}(dG|\mathbf{S}) = \int\int e^{-u\phi(G)}\mathbb{P}(dG, d\chi|\mathbf{S}) = \int\int e^{-u\phi(G)}\mathbb{P}(dG|\chi)\mathbb{P}(d\chi|\mathbf{S})$$

where $\chi = (\chi_1, \ldots, \chi_N)$ and the last equality holds since G is independent of $\mathbf{S}$ given χ. Hence in theory we can sample from $\mathcal{L}(\phi(G)|\mathbf{S})$ in tree steps:

1. sample $\chi \sim \mathbb{P}(d\chi|\mathbf{S})$.
2. sample $\breve{G} := [G|\chi] \sim \mathbb{P}(dG|\chi)$
3. Set $\breve{\phi} := \phi(\breve{G})$.

This fact was noted in [78, 67]; and is our main strategy to sample functionals $\phi(G)$ approximately from $\mathcal{L}(\phi(G)|\mathbf{S})$. The description of the first step is rather long and is therefore given in the appendix. Whereas the second step is explained in the next subsection.

Sampling $\breve{G} \sim \mathcal{L}(G|\chi)$

To simulate $\phi(G)$ from $\mathcal{L}(\phi(G)|\mathbf{S})$ we need, in the second step, to sample the random probability $\breve{G} = [G|\chi]$ from the posterior law $\mathcal{L}(G|\chi)$. From [127] a draw from the posterior has the distributional form

$$\breve{G} \stackrel{\mathcal{L}}{=} \sum_{j=1}^{k_N} p_j \delta_{\chi_j^*}(\cdot) + p_{k_n+1}\tilde{G}(\cdot) \tag{8.18}$$

where $\chi = (\chi_i = \chi_{c_i}^*)_{1\leq i\leq N}$, $(\chi_j^*)_{1\leq i\leq K_N}$ are the unique values among χ which occur with frequencies $(n_i)_{1\leq i\leq K_N}$ and $c_i \in \{1, \cdots, k_N\}$. Furthermore $\tilde{G}$ is a Poisson-Dirichlet process with parameters $(a, b + ak_n, G_0)$ and

$$p := (p_1, \cdots, p_{k_n}, p_{k_n+1}) \sim \mathrm{Dir}(n_1 - a, \cdots, n_{k_n} - a, b + k_n a). \tag{8.19}$$

Where p, $(\chi_j^*)_1^{k_N}$ and $\tilde{G}$ are independent of each other. Since $\tilde{G}$ is infinite dimensional, it not possible to draw $\breve{G}$ exactly. There are several ways to obtain approximate draws from $\breve{G}$. One way would be to use distributional results for linear functional of the Poisson-Dirichlet process in combination with perfect simulation as described in [82, 41]. This approach has several limitations since the support of G_0 will be unbounded and perfect simulation can only be approximated [72]. Instead we use the following result, which follows directly from theorem 2 of [78].

Lemma 35. *Let $\breve{G}$ be defined as in (8.18) and define the truncated posterior random probability*

$$\breve{G}_T = \sum_{j=1}^{k_N} p_j \delta_{\chi_j^*}(\cdot) + p_{k_n+1} \sum_{j=1}^{T} \delta_{\tilde{\chi}_j} \tilde{w}_j. \tag{8.20}$$

where $\tilde{\chi}_j \sim^{iid} G_0(\cdot|\eta)$ and for $1 \leq j \leq T$

$$\tilde{w}_j = \tilde{V}_j \prod_{l<j} (1 - \tilde{V}_l) \quad with \quad \tilde{V}_T = 1, \tilde{V}_j \sim Beta(a, b + a(j + k_N)). \tag{8.21}$$

Furthermore denote with $\tilde{\boldsymbol{S}} = (\tilde{S}_i)_{i=1}^M$ a future exchangeable sample of size M. Then the L_1-distance between the predictive distributions of $\tilde{\boldsymbol{S}}$ given $\chi = (\chi)_1^N \sim P(\chi|\boldsymbol{S})$ under $\breve{G}$ and the truncated version $\breve{G}_T$ is bounded by

$$\begin{aligned} &\Big|\Big|\mathbb{P}(d\tilde{\boldsymbol{S}}|\chi) - \mathbb{P}(d\tilde{\boldsymbol{S}}|\chi)_T)\Big|\Big|_{L_1} \\ &\quad \leq \frac{4(b + k_N a)}{b + N}\Big(1 - \Big(1 - \prod_{j=1}^{T-1} \frac{b + (j + k_N)a}{b + (j + k_N - 1)a}\Big)^M\Big) \end{aligned} \tag{8.22}$$

where

$$\mathbb{P}(d\tilde{\boldsymbol{S}}|\chi)_T = \int \Big(\int_{[0,\infty)^2} \prod_{j=1}^{M} f(\tilde{S}_j|\tilde{\chi}) \breve{G}_T(d\tilde{\chi}) \Big) \mathbb{P}(d\breve{G}_T|\chi) \tag{8.23}$$

and $\mathbb{P}(d\tilde{\boldsymbol{S}}|\chi)$ is defined similarly with $\breve{G}_T$ replace by $\breve{G}$.

Since $a < 1$ a.s. the bound (8.23) shows that even for small T prediction with $\breve{G}_T$ instead of $\breve{G}$ yield quite accurate results for cases in which the number of required mixture components k_N is substantially smaller than the sample size N. And most certainly this hold in the context of a queuing process where N ranges easily in the area of several thousands of observations. This fact is also intuitively reasonable. Through sampling $[\chi|\mathrm{S}]$ in the first step, without truncating the number of mixture components in advance, we already obtain a substantial amount of information about $[G|\chi]$ from the first term in (8.18). This term is also shared by $\breve{G}_T$. Whereas the second term of

$\tilde{G}$ in (8.18) represents the part of information about $[G|\chi]$ not jet discovered by the data. This hidden part of information is quantified by the second term in the product (8.23) and can usually be represented quite accurate for T between 50 and 200 dependent in the required precision [78]. Hence, we monitor the number of clusters k_N from the MCMC first and choose a T which makes the predictive distribution close to the true predictive distribution. The mean and Laplace transforms are then computed as

$$\mu_{S,T} = \sum_{j=1}^{k_N} p_j \mu_j^* + p_{k_N+1} \sum_{j=1}^{T} w_j \tilde{\mu}_j \quad \text{and} \tag{8.24}$$

$$\hat{F}_{S,T}(s) = \sum_{j=1}^{k_N} p_j \Big(1 + s\mu_j^*/\nu_j^*\Big)^{-\nu_j^*} + p_{k_N+1} \sum_{j=1}^{T} w_j \Big(1 + s\tilde{\mu}_j/\tilde{\nu}_j\Big)^{-\tilde{\nu}_j} \tag{8.25}$$

8.3.3 Numerical Inversion of Transforms

In the implementation of inference for performance measure of the $M/G/1$ queue we need to invert probability generating functions and Laplace-transforms. We summarize the algorithms we choose to invert these transforms. For a general overview and comparison of inversion algorithms see [3].

Numerical Inversion of Probability generating functions

Denote in the following with $G_a(\cdot)$ the z-transform of a real sequence $a = (a_k)$

$$G_a(z) = \sum_k a_k z^k \quad z \in R_a := \Big\{ c \in \mathbb{C} : |\sum_k a_k c^k| < \infty \Big\}. \tag{8.26}$$

If $p = (p_k)_{k \geq 0}$ is a proper probability mass function then $(-1, 1] \in R_p$ and G_p is analytic for all $|z| < 1$. Furthermore, if $p = \{p_k\}$ has cdf $F = (F_k)$ and complementary cdf (ccdf) $S = 1 - F$, then the z-transforms of p,F and S are related by

$$G_F(z) = \frac{G_p(z)}{1 - z} \quad \text{and} \quad G_S(z) = \frac{1 - G_p(z)}{1 - z} \tag{8.27}$$

whenever the right hand side exist (which is the case for $|z| < 1$). The z-transform can be inverted via the Cauchy contour integral

$$a_k = \frac{1}{2\pi i}\int_{r-\infty i}^{r+\infty i} G_a(z)z^{-k-1}dz = \frac{1}{2\pi r^k}\int_0^{2\pi} \mathrm{Re}(G_a)(re^{iu})e^{-iku}du, \quad (8.28)$$

whenever we can obtain a closed form expression for the last integral. For cases we will encounter there does not exist a closed form expression for the inverse transform. Hence, we need to refer to some numerical inversion procedure. We use the "Lattice-Poisson algorithm" [4, 3]. The procedure can be applied directly to the PGF of the probability mass function p or to the z-transform of the cdf F or to the ccdf S using the relation (8.27).

For completeness we will describe the main steps of the inversion procedure for a generic real sequence a (having in mind $a = p, F, S = 1 - F$). The idea is to approximation (8.28) by a m-point trapezoidal integral with step size $h = 2\pi/k$. Using fact that a is real we can approximate a_k by

$$a_{k,m} = \frac{1}{mr^k}\sum_{j=1}^{m} \mathrm{Re}(G_a)(re^{2\pi ij/m})e^{-2\pi ikj/m}. \quad (8.29)$$

To use this approximation in practice it is important to control the approximation error $e_m = |a_k - a_{k,m}|$ uniformly in k. An alternative derivation of the algorithm via discrete Fourier transforms and a discrete Poisson summation formula for a derived in [4] shows that if $a_k \leq 1$ (which is the case if $a = p, F$ or S) then $e_m \leq r^m/(1 - r^m) \approx r^m$. To obtain a precision for $a_{k,m}$ of $10^{-\gamma}$ (meaning $|a_k - a_{k,m}| \leq 10^{-\gamma}$) one may choose $r = 10^{-\gamma/(2k)}$. For $m = 2k$ formula (8.29) is known at the Lattice-Poisson algorithm [3] which (since $e^{\pi ji} = (-1)^j$) can be written as

$$a_{k,m} = \frac{1}{2kr^k}\left(G_a(r) + (-1)^k G_a(-r) + 2\sum_{j=1}^{k-1}(-1)^j \mathrm{Re}(G_a)(re^{\pi ij/k})\right). \quad (8.30)$$

Numerical Inversion of Laplace Transforms

Denote in the following with Υ the Laplace transform of a non-negative piecewise continuous function $g : [0, \infty) \to [0, \infty)$, that is

$$\Upsilon(g)(s) = \int_{[0,\infty)} e^{-sx} g(x) dx \qquad \mathrm{Re}(s) \geq 0 \tag{8.31}$$

whenever the integral exist. We assume that there exist $s_0 = x_0 + ix_0 \in \mathbb{C}$ for which the integral (8.31) exists. This is true if $g \leq 1$, which is the case if g is a cdf or ccdf. Hence $\Upsilon(g)$ exists for any complex number with real part greater than x_0 (see [153], p.37). Furthermore we denote with $\sigma_c \in [-\infty, \infty)$ the abscissa of convergence of $\Upsilon(g)$. Meaning the extended real number such that, if $\sigma_c = -\infty$, then $\Upsilon(g)(s)$ exists for all $s \in \mathbb{C}$ and, if $\sigma_c \in \mathbb{R}$, then $\Upsilon(g)(s)$ exists for all s such that $\mathrm{Re}(s) > \sigma_c$, but diverges for all s with $\mathrm{Re}(s) < \sigma_c$ (see [153] for more details).

If g is an absolutely continuous density on $[0, \infty)$ with cdf F and ccdf $S = 1 - F$, then the Laplace transform Υ of all three functions are related by

$$\Upsilon(F)(s) = \int_0^\infty e^{-sx} F(x) dx = \frac{\Upsilon(g)(s)}{s} \tag{8.32}$$

$$\Upsilon(S)(s) = \int_0^\infty e^{-sx} S(x) dx = \frac{1 - \Upsilon(g)(s)}{s} \tag{8.33}$$

whenever the right hand side exists.

For a generic piecewise continuous function g which is exponentially restricted as $x \to +\infty$, the inverse of the Laplace transform is given by Bromwich's contour-integral [153, 3]

$$g(x) = \frac{1}{2\pi i} \int_{C_a} e^{sx} \Upsilon(g)(s) ds = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{(a+it)x} \Upsilon(g)(a + it) dt, \tag{8.34}$$

where the range of integration consists of the contour path $C_a = \{s \in \mathbb{C} : \mathrm{Re}(s) = a\}$ and $a > \sigma_c$. Again we will not be able to obtain a closed form expression of this integral and need a numerical approximation. We use the 'Euler-Inversion algorithm' [2, 3]. The main idea of the algorithm is

the following. Replace the integral in (8.34) by an infinite step trapezoidal integral of step size $h = \pi/(2t)$ at $a = A/(2t)$ we obtain

$$g_h(t) = \frac{e^{A/2}}{2t}\,\mathrm{Re}(\Upsilon)\Big(\frac{A}{2t}\Big) + \frac{e^{A/2}}{t}\sum_{j\geq 1}(-1)^j\,\mathrm{Re}(\Upsilon)\Big(\frac{A+2j\pi i}{2t}\Big). \tag{8.35}$$

[2] derived $g_h(t)$ through a Fourier series and a Poisson summation formula. The authors showed that, if $g(t) \leq 1$, the absolute error can be bounded uniformly in t by $e_h \leq \frac{e^{-A}}{1-e^{-A}} \approx e^{-A}$.

For an implementation one has to truncate the infinite sum (8.35) which will decrease the quality of approximation. But the higher order terms in the convergent sum (8.35) do not affect the approximation substantially (especially if the sign of $\mathrm{Re}(\Upsilon)((A+2j\pi i)/2t)$ remains constant as a function of j). As suggested in [2, 3] one may truncate the series and uses ordinary Euler summation to accelerate the convergence of the alternating sum (8.35). If we let

$$g_{h,k}(t) = \frac{e^{A/2}}{2t}\,\mathrm{Re}(\Upsilon)\Big(\frac{A}{2t}\Big) + \frac{e^{A/2}}{t}\sum_{j=1}^{k}(-1)^j\,\mathrm{Re}(\Upsilon)\Big(\frac{A+2j\pi i}{2t}\Big) \tag{8.36}$$

for $k \in n, \cdots, n+m$, then the "Euler-Inversion" algorithm of [2] is given by

$$g_{h,n,m}(t) = \sum_{k=1}^{m}\binom{m}{k}2^{-k} f_{h,n+k}(t). \tag{8.37}$$

The algorithm is quite simple, but our experience shows that even for small n, m, [2] suggested $m = 15, n = 11$, the method works extremely well. In most cases $A \approx 19$ yield $|g_{h,n,m}(t) - g(t)|$ in the order of 10^{-7} to 10^{-8}, which is more accuracy than we actually need.

Note that stochastic integration would require substantially more computational time to obtain this accuracy. Hence, since we need an inversion step at every MCMC-iteration, stochastic integration is out of question for our MC-procedure.

8.4 Estimating performance measure

In this section we discuss inference for performance measure of the $M^X/G/1$ process. Most of this performance measures are usually studied for the system in steady-state whenever the system is stable. This is partly due to the fact that the system should run cost-optimal in the long run, i.e. short idle periods and an average queuing-length close to one. Furthermore, the distribution of performance measures is usually only tractable for a system in steady-state.

The probability distribution of the performance measure is usually only characterized through the probability generating function (PGF) or the Laplace-Stieltjes transform (LST). In the sequel we denote the PGF for a positive integer valued random variable X, with probability mass function P_X, by $P_X^*(s) = \sum_k s^k P_X(k)$. For a non-negative random variable Y, with cdf F_Y, the LST is denoted by $\hat{F}_Y(s) = \int e^{-sy} F_Y(dy)$.

Although this transforms determine uniquely the probability law of the performance measure, they are usually note invertible in closed. Hence, we need to refer to a MC-sampling scheme and then invert the transform numerically. The generic algorithm will consists of the following steps:

1. Sample the random quantities $\left(\lambda, E[S|F_S], E[B|P_B], P_B^*(\cdot), \hat{F}_S(\cdot)\right)$ from the posterior $\Pi[\cdot|\mathbf{IA,B,S}, \rho < 1)$.

2. Compute the transform for the performance measure in steady-state say $P^*(\cdot)$ or $\hat{F}(\cdot)$.

3. Invert the transform numerically to obtain a realization of the random probability of the performance measures $(P(k), k \geq 1)$ (or $F(t), t \geq 0)$).

4. Repeat step (1) to (3) C times to obtain realizations of the probability of a performance measure in steady state, $P^{(c)}$ (or $F^{(c)}$), $c = 1, \cdots, C$. Compute the point estimator and a credibility interval for the probability measure P (or F) based on this sample.

We may use these sample also to estimate further summaries, like the median number of customers in the system in steady state.

8.4.1 Performance Measures in transient-state

In this section we discuss the estimation of basic performance measures in transient state. A usefully summary measure in transient state is the total number of arrivals until time $t > 0$ defined as usually by

$$N_c(t) = \sum_{j=1}^{N(t)} Y_j \quad \text{where} \quad N(t) = \sup\Big\{n : \sum_{i=1}^{n} IA_i := A_n \leq t\Big\}. \tag{8.38}$$

N_c represents an ordinary compound Poisson process. If λ and P_B would be known, then, by elementary properties of conditional expectations, the probability generating function of N_c at time t is given by

$$P^*_{N_c(t)}(s) = \exp\{\lambda t(P^*_B(u) - 1)\} \quad \text{where} \quad P^*_B(u) = \sum_{j\geq 1} u^j P_B(j) \tag{8.39}$$

and the mean number of arrivals by time $t > 0$ is given by

$$\mu_{N_c}(t) = \mathbb{E}[N_c(t)|\lambda, P_B] = \lambda \mathbb{E}[B_i|P_B]t. \tag{8.40}$$

The mean bath size $\mathbb{E}[B_{N_1+1}|\mathbf{B}] = \mathbb{E}[\mathbb{E}[B_i|P_B]|\mathbf{B}]$ involves an infinite sum and the power series $P^*_B(\cdot)$ has no closed form expectation. Hence, it is not possible to invert $P^*_{N_c(t)}$ in closed form or compute the predicted mean number of arrivals explicitly. We will refer to the following MC-estimation procedure: Fix M and for $c = 1, \cdots, C$ repeat

1. Draw $\lambda_c \sim \text{Gam}(u_0 + N_1, v_0 + A_{N_1})$ and sample $Q^{(c)}_{M,B}(\cdot)$ with the bootstrap scheme described in section 8.3.1.
2. Compute the mean $\mu^{(c)}_B$ and the PGF $Q^{(c)*}_{M,B}$ of $Q^{(c)}_{M,B}$.
3. Set $\mu_{N_c}(t) = \lambda_c \mu^{(c)}_B t$ and $P^{(c)*}_{N_c(t)}(\cdot) = \exp\Big\{\lambda_c t\big(Q^{(c)*}_{M,B}(\cdot) - 1\big)\Big\}$.
4. Invert $P^{(c)*}_{N_c(t)}(\cdot)$ numerically to obtain an approximate realization of $P^{(c)}_{N_c(t)} \sim \mathbb{P}[dP_{N_c(t)}(\cdot)|\mathbf{IA}, \mathbf{B}]$.

We then use $P^{(c)}_{N_c(t)}$ and $\mu_{N_c}(t)$ for $c = 1, \cdots, C$ to obtain summary measures for the arrival process at a fixed time-points t.

Furthermore, consider the number of completed services in $[t_1, t_1+t]$, given that the system is non-idle during the period $[t_1, t_1 + t]$. By homogeneity, this is equivalent (in distribution) to number of completed services in $[0, t]$, given that the queue is non-idle between $[0, t]$

$$N_S(t) = \sup\left\{ n : \sum_{i=1}^{n} S_i \leq t \right\} = \sum_{j\geq 1} I_{[0,t]}\left(\sum_{1\leq i\leq j} S_i \right). \tag{8.41}$$

If the service times distribution P_S is known, then the renewal function is given by

$$M_S(t) := E[N_S(t)|F_S] = \sum_{j\geq 1} F_S^{\#(j)}(t), \tag{8.42}$$

where $F_S^{\#(j)}(t)$ denotes the j-fold convolution of $F_S(\cdot) = P_S[0, \cdot]$ [138, 132]. By linearity of the LT, we can express the LT of $M_S(t)$ by

$$\Upsilon_{M_S}(s) = \frac{\hat{F}_S(s)}{1 - \hat{F}_S(s)}, \tag{8.43}$$

where $\hat{F}_S$ denotes the Laplace transform of F_S [138]. Since the posterior probability $\mathbb{P}(dF_S|\mathbf{S})$ is not available in closed form, we cannot take expectation in (8.43) and invert the transform. We need to refer again to MC-simulations, in particular we estimate the renewal function by using the following steps: For $c = 1, \cdots, C$ repeat

1. Simulate a service-time cdf $F_S^{(c)} = \int F(\cdot|\theta)\breve{G}_T^{(c)}(d\theta)$ with $\breve{G}_T$ as in (8.20).

2. Compute the corresponding LST $\hat{F}_{S,T}^{(c)}(\cdot)$ as in (8.25) and set $\Upsilon_{M_S}^{(c)}(\cdot) = \hat{F}_{S,T}^{(c)}(\cdot)/(1 - \hat{F}_{S,T}^{(c)}(\cdot))$.

3. Invert $\Upsilon_{M_S}^{(c)}(\cdot)$ numerically to obtain an approximate realization of $M_S^{(c)}(\cdot)$ from $\mathbb{P}(dM_S(\cdot)|\mathbf{S})$.

Again we use the sample $M_S^{(c)}(\cdot)$, $c = 1, \cdots, C$ to obtain summary measures of interest for $M_S(\cdot)$. For example the predicted renewal function (marginal

expected number if completed service) by time t is given by

$$\hat{M}_S(t) = \mathbb{E}[E[N_S(t)|F_S]|\mathbf{S}] \approx C^{-1} \sum_{c=1}^{C} M_S^{(c)}(t). \tag{8.44}$$

Traffic intensity of the System

The most important performance measure of the $M^X/G/1$ queue is the traffic intensity of the system which is given by

$$\rho = \frac{\mathbb{E}[B|P_B]\mathbb{E}[S|F_S]}{\mathbb{E}[IA|\lambda]} \tag{8.45}$$

see [109], p. 295. The traffic intensity ρ determines whether the queuing system reaches a probabilistic equilibrium at regenerative time-points when time goes to infinity. For the $M^X/G/1$ queue, the sequence of departure epochs creates a sequence of regenerative time points. An $M^X/G/1$ queuing system is stable ('ergodic') if $\rho < 1$, null recurrence if $\rho = 1$ and transient if $\rho > 1$. For a transient system the number of items on hold will grow without any limit and is therefore economically speaking cost-inefficient. Whereas for a null-recurrent system, the average time of return to the same queuing size just observed will be infinity. This is similarly undesirable.

The Bayesian point estimator for the traffic intensity with respect to a square loss function is given by

$$\begin{aligned}\mathbb{E}[\rho|\text{data}] &= \mathbb{E}[\lambda|\mathbf{IA}]\mathbb{E}[B_i|\mathbf{B}]\mathbb{E}[S_i|\mathbf{S}] \\ &= \frac{N_1 + u_0}{A_{N_1} + v_0}\left(1 + \sum_{k=1}^{\infty} \frac{\not{\beta}k) + m(k)}{\not{\beta}k) + \alpha(k) + n(k) + m(k)}\right) \sum_{k\geq 1} \mathbb{E}[w_k \mu_k^*|\mathbf{S}].\end{aligned}$$

The last two terms in the last expression involve an infinite sum which is intractable to compute in exact form. Hence, we use an MC-approximation and estimate ρ by

$$\mathbb{E}[\rho|(\mathbf{S}, \mathbf{B}, \mathbf{IA})] \approx \frac{N_1 + u_0}{A_{N_1} + v_0} \times C^{-1} \sum_{c=1}^{C} \mu_{B,M}^{(c)}\ \mu_S^{(c)}. \tag{8.46}$$

Where for a fixed iteration $\mu_{B,M}^{(c)}$ is computed as described in section 8.3.1 and $\mu_S^{(c)}$ is sampled as described in section 8.3.2. Alternatively and from a computational point inexpensive, one may estimate the complete posterior distribution of the traffic intensity by a Rao-Blackwellized estimator (after re-parameterization)

$$\mathcal{L}(d\rho|(\mathbf{S},\mathbf{B},\mathbf{IA})) \approx C^{-1}\sum_{k=1}^{C} f_{Gam}(d\rho|u_0+N_1,(v_0+A_{N_1})/(\mu_{B,M}^{(c)}\mu_S^{(c)})). \tag{8.47}$$

The advantage of approximating the posterior law $\mathcal{L}(\rho|\mathbf{S},\mathbf{B},\mathbf{IA})$ over the point estimator $\hat{\rho}$ is that we obtain information about the shape of the posterior and a possible multi-modality.

Testing Ergodicity of the Queuing System

Before conducting inference for performance measures in steady-state one needs to do a formal decision whether such a steady-state equilibrium even exists for the particular queue under consideration. If the system is unstable, no such performance measures would even exist. Hence, we need a decision rule $\delta(\mathbf{IA},\mathbf{S},\mathbf{B}) \in \{0,1\}$ based on the current state of information provided by the data $(\mathbf{IA},\mathbf{S},\mathbf{B})$. We denote with $\delta(\mathbf{IA},\mathbf{S},\mathbf{B})=1$ the acceptance and with $\delta(\mathbf{IA},\mathbf{S},\mathbf{B})=0$ the rejection of the null hypothesis

$$H_0 : \rho < 1 \quad \text{versus } H_1 : \rho \geq 1. \tag{8.48}$$

We may assume a weighted 0-1 loss function of the form

$$L(\rho,\delta(\mathbf{IA},\mathbf{S},\mathbf{B})) = \begin{cases} a_0 & \text{if } \rho<1, \delta(\mathbf{IA},\mathbf{S},\mathbf{B})=0 \\ a_1 & \text{if } \rho\geq 1, \delta(\mathbf{IA},\mathbf{S},\mathbf{B})=1 \\ 0 & \text{otherwise.} \end{cases} \tag{8.49}$$

A Bayesian decision rule for accepting or rejecting H_0 based on the current information can be expressed as

$$\delta(\mathbf{IA}, \mathbf{S}, \mathbf{B}) = \begin{cases} 1 & \text{if } P(\rho < 1|\mathbf{IA}, \mathbf{S}, \mathbf{B}) > \dfrac{a_1}{a_0 + a_1} \\ 0 & \text{otherwise.} \end{cases} \tag{8.50}$$

Where the Rao-Blackwellized estimator of the acceptance probability is given by

$$P(\rho < 1|\mathbf{IA}, \mathbf{S}, \mathbf{B}) \approx C^{-1} \sum_{k=1}^{C} F_{\text{Gam}}(1^-|u_0 + N_1, (v_0 + A_{N_1})/\mu_{B,M}^{(c)} \mu_S^{(c)}). \tag{8.51}$$

8.4.2 Performance measures in steady-state

Suppose the date provide sufficient evidence for the hypothesis $\rho < 1$ and based on the test outlined above we accept that the system is ergodic. Under this assumption we may assume that the system reaches a steady-state equilibrium at departure epochs. Inference for performance measure in steady-state needs to be implemented by restricting the posterior to the ergodic parameter space

$$\Theta_{ergodic} = \{(\lambda, P_B, P_S) \in \Theta : \lambda E[B|P_B] E[S|F_S] < 1\}. \tag{8.52}$$

This is important since even if $P(\rho < 1|\mathbf{IA}, \mathbf{B}, \mathbf{S}) =: p_\rho > \frac{a_1}{a_1+a_2}$, the probability p_ρ might still be fare away from one. In this case $(1-p_\rho) \times 100$ percent of each MC-simulation yields a traffic intensity ρ, which is greater or equal to one. For such draws (λ, P_B, P_S) no steady-state performance measure would exist. Therefore, if we want to do inference of performance measures in steady-state, we need to make sure that we sample for the right parameter

space. In practice we sample from $\Theta_{ergodic}$ by using the fact that

$$\begin{aligned} &P(d(\lambda, P_B, P_S)|\rho < 1, \mathbf{IA}, \mathbf{B}, \mathbf{S}) \\ &\qquad = \frac{P(d(\lambda, P_B, P_S)|\mathbf{IA}, \mathbf{B}, \mathbf{S})I(\rho < 1)}{P(\rho < 1|\mathbf{IA}, \mathbf{B}, \mathbf{S})} \\ &\qquad \propto P(dP_B|\mathbf{B})P(dP_S|\mathbf{S})P(d\lambda|\mathbf{IA})I(\lambda < (\mu_B\mu_S)^{-1}). \end{aligned}$$

Hence we may sample from the posterior of $\Theta_{ergodic}$ by first sampling (P_B, P_S) from $\mathcal{L}(P_B, P_S|\mathbf{B}, \mathbf{S})$ and then sample λ from a Gamma distribution restricted to the interval $(0, 1/(\mu_B\mu_S))$. We can draw random variables from a truncated Gamma distribution with some care with acceptance-rejection sampling [126] or by slice sampling within the MCMC as proposed in [35]. In practice we prefer the acceptance-reception algorithm of [126] since it produces independent draws. For the remaining part we will assume that the $M^X/G/1$ queue is ergodic and an equilibrium distribution at departure epochs exists with respect to $P(\cdot|\rho < 1, \mathbf{IA}, \mathbf{B}, \mathbf{S})$.

Number of items in the system at departure epochs

A typical steady-state performance measure of interest is the distribution of the number of items in the system at departure epochs in steady-state, say $\mathcal{L}(X_E)$. Assume first that the "true" parameter $\theta = (\lambda, P_B, F_S) \in \Theta_{ergodic}$ of the system is known. In this case, one can characterize the PGF of $\mathcal{L}(X_E)$ through the traffic intensity and the transform of P_B and F_S by Pollaczek-Khinchin formula for the $M^X/G/1$ system, see [109], p. 295

$$P^*_{X_E}(s) = h(s, \rho, \lambda, P^*_S, \hat{F}_S) = \frac{(1-\rho)(1-s)\hat{F}_S(\lambda[1 - P^*_B(s)])}{\hat{F}_S(\lambda[1 - P^*_B(s)]) - s} \tag{8.53}$$

where $\rho, \hat{F}_S$ and P^*_B are defined as before.

Except for some specific θ, it is usually not possible to invert the PGF (8.53) in closed form to obtain $\mathcal{L}(X_E)$ explicitly. The same holds for the non-parametric model $(\Theta_{ergodic}, P(d\theta|\rho < 1, \mathbf{IA}.\mathbf{B}, \mathbf{S}))$. Therefore we refer to simulation-based inference and estimate the probability measure of the

number of items in steady-state as follows: for $c = 1, \cdots, C$ repeat the following steps.

1. Sample a batch-size probability $Q_{B,M}^{(c)}$ with the bootstrap scheme as described in section 8.3.1 and
2. compute the mean $\mu_{B,M}^{(c)}$ and PGF $Q_{B,M}^{(c)*}$ corresponding to $Q_{B,M}^{(c)}$.
3. Simulate a service-time distribution $F_S^{(c)} = \int F(\cdot|\theta)\breve{G}_T(d\theta)$ where $\breve{G}_T$ is defined in (8.20) and
4. compute the mean $\mu_{S,T}^{(c)}$ and the LST $\hat{F}_{S,T}^{(c)}$ as in (8.24) and (8.25).
5. Draw $[\lambda^{(c)}|\mu_{B,M}^{(c)}, \mu_{S,T}^{(c)}] \sim \text{Gam}(u_0 + N_1, v_0 + A_{N_2})I(\lambda^{-1} < \mu_{B,M}^{(c)}\mu_{S,T}^{(c)})$.
6. Set $P_{X_E}^{(c)*}(s) = h(s, \rho^{(c)}, \lambda^{(c)}, P_{B,M}^{(c)*}, \hat{F}_{S,T}^{(c)})$ and invert $P_{X_E}^{(c)*}$ numerical as described in section 8.3.3 to obtain a realization $P_{X_E}^{(c)} := \mathcal{L}(X_E)^{(c)}$ from $\mathbb{P}(\mathcal{L}(X_E)|\rho < 1, \mathbf{IA}, \mathbf{B}, \mathbf{S})$.

Again, based on the MC-scheme we compute the predictive distribution and the median as usually by

$$\begin{aligned}\mathbb{P}(X_E = n|\mathbf{IA}, \mathbf{B}, \mathbf{S}) &= \mathbb{E}[P(X_E = n|\lambda, P_B, F_S)|\mathbf{IA}, \mathbf{B}, \mathbf{S}] \\ &\approx \begin{cases} C^{-1}\sum_1^C P_{X_E}^{(c)}(n) & \text{if } n \geq 1 \\ 1 - C^{-1}\sum_1^C \rho^{(c)} & \text{if } n = 0 \end{cases} \qquad (8.54)\end{aligned}$$

$$\text{median}(\mathcal{L}(X_E)||\mathbf{IA}, \mathbf{B}, \mathbf{S}) \approx C^{-1}\sum_1^C \text{median}(P_{X_E}^{(c)}) \qquad (8.55)$$

Note that we estimated the center of the distribution of X_E by the median instead of the mean. We choose the median since the posterior mean $\mathbb{E}[X_E|\rho < 1, \mathbf{IA}, \mathbf{B}, \mathbf{S}]$ does not exists. This is a consequence of the fact that the posterior density of ρ does not approach zero as ρ approaches one even for the restriction $(\lambda, P_B, F_S) \in \Theta_{ergodic}$ (see also [10, 155, 11, 15] for the same phenomena in the semi-/parametric case).

Waiting time for service and in the system

So fare we only considered the total number of items in the system at steady-state. A further quantity of interest is the waiting time from arrival to processing and the total processing time of an item in steady-state. The second quantity refers to the waiting time plus the actual service time.

This random variables can be studied for the first fortune item within an arriving batch, say W_1 for the waiting time in the queue and $W_1 + S$ for the total time in the system. The same quantities can also be analyzed for an average test-item in a batch, say $W_1 + W_2$ and $W_1 + W_2 + S$. The random time W_2 corresponds to the additional waiting time of the average test-item until all items, within the batch, in front of the test-item are processed. We could do the same analysis even for the last item within an arriving bath (which we do not consider here). The Laplace-Stieltjes transforms of the three random variables W_1, $W_1 + S$ and $W_1 + W_2 + S$ are given by

$$\hat{F}_{W_1}(s) = \frac{s(1-\rho)}{s - \lambda + \lambda P_B^*[\hat{F}_S(s)]}, \tag{8.56}$$

$$\hat{F}_{W_1+S}(s) = \hat{F}_{W_1}(s)\hat{F}_S(s) \qquad \text{and} \tag{8.57}$$

$$\hat{F}_{W_1+W_2+S}(s) = \frac{1 - P_B^*[\hat{F}_S(s)]}{\mu_B[1 - \hat{F}_S(s)]}\hat{F}_{W_1}(s)\hat{F}_S(s). \tag{8.58}$$

see [109], p. 297. We will consider only the estimation of the probability $\mathcal{L}(W_1)$. The estimation of $\mathcal{L}(W_1+S)$ and $\mathcal{L}(W_1+W_1+S)$ is straight forward. We treat the analysis of $\mathcal{L}(W_1)$ separately since $\mathcal{L}(W_1)$ is not absolutely continuous.

If the queuing system is empty at arrival, which happens with probability $1 - \rho$, then the waiting time before starting the service, for the first item in a bath, is 0. Hence the cdf of W_1 is given by

$$F_{W_1} = (1-\rho)I_{[0,\infty)}(\cdot) + \rho F_{\tilde{W}_1} = I_{[0,\infty)} - \rho(1 - F_{\tilde{W}_1}) \tag{8.59}$$

where $\tilde{W}_1 := [W_1|W_1 > 0]$ has absolutely continuous cdf $F_{\tilde{W}_1}$. Taking Laplace transform on both sides and using (8.32) we can express the LT of the ccdf

of $\tilde{W}_1$ as

$$\Upsilon(S_{\tilde{W}_1})(s) = \Upsilon(1 - F_{\tilde{W}_1})(s) = \frac{1 - \hat{F}_{W_1}(s)}{\rho s} \tag{8.60}$$

where $\hat{F}_{W_1}$ is given in (8.56).

Now, we can estimate the cdf F_{W_1} using the following algorithm: for $c = 1, \cdots, C$ repeat

1. Sample a batch-size probability $Q^{(c)}_{B,M}$ with the bootstrap scheme as described in section 8.3.1 and

2. compute the corresponding mean $\mu^{(c)}_{B,M}$ and PGF $Q^{(c)*}_{B,M}(\cdot)$ of $Q^{(c)}_{B,M}$.

3. Simulate a service-time distribution $F^{(c)}_S = \int F(\cdot|\theta)\breve{G}_T(d\theta)$ where $\breve{G}_T$ is defined in (8.20) and

4. compute the mean $\mu^{(c)}_{S,T}$ and the LST $\hat{F}^{(c)}_{S,T}(\cdot)$ as in (8.24) and (8.25).

5. Draw $[\lambda^{(c)}|\mu^{(c)}_{B,M}, \mu^{(c)}_{S,T}] \sim \text{Gam}(u_0 + N_1, v_0 + A_{N_2})I(\lambda^{-1} < \mu^{(c)}_{B,M}\mu^{(c)}_{S,T})$ and set $\rho^{(c)} = \lambda^{(c)}\mu^{(c)}_{B,M}\mu^{(c)}_{S,T}$.

6. Compute $\Upsilon(S^{(c)}_{\tilde{W}_1})$ according to (8.60) and invert the LT numerically to obtain a realization $S^{(c)}_{\tilde{W}_1}(\cdot) = P^{(c)}(\tilde{W}_1 > \cdot|\rho^{(c)} < 1, \lambda^{(c)}, Q^{(c)}_{B,M}, F^{(c)}_S)$ and take

$$F^{(c)}_{W_1}(\cdot) = 1 - \rho^{(c)}S^{(c)}_{\tilde{W}_1}(\cdot)$$

as an approximate realization form $\mathbb{P}(dF_{W_1}|\rho < 1, \mathbf{IA}, \mathbf{B}, \mathbf{S})$.

8.5 Numerical Illustration

The aim of this section is to numerical illustrate the inferential procedures described in the previous section. In particular we consider a $M^X/G/1$ model which a true arrival rate of $\lambda_0 = .1$ and a batch size and service time distri-

bution as follows

$$P_B(j) = 5/6Pois(j-1|1) + 1/6Pois(j-1|7) \quad \text{for } j \geq 1,$$
$$F_S(t) = .4F_{Exp}(t|1) + .4F_{Gam}(t|10, 10/3) + .15F_{Gam}(5, 5/6) + .05F_{Gam}(t|7, 7/10)$$

for $t \in [0, \infty)$. The true mean of the batch-sizes and mean service-time is $\mu_{B,0} = \mu_{S,0} = 3$. The corresponding traffic intensity of the system is $\rho_0 = .9$. We simulated two scenarios with a different sample size. In particular we draw $N_1 = 500$ and 1000 inter-arrival times and batch size variables together with $N_2 = 500$ and 1000 service times from the true model. Table 8.5 shows a summary of the simulated data together with the corresponding maximum likelihood estimator for the traffic intensity.

Table 8.1: Data Summary for the simulated $M^X/G/1$ queue

		sample means		
Sample size	A_{N_1}	Batch-size	Service-times	$\hat{\rho}_{MLE}$
500	5072.3	2.928	2.841	0.820
1000	10101.7	2.949	2.969	0.867

Since we know the true models we will choose relatively non-informative prior. For λ we choose $P(d\lambda) \propto \lambda^{-1} d\lambda$. This corresponds to the case where $u_0 = v_0 = 0$. Furthermore we center the beta-Stacy random distribution on a geometric distribution with mean 2, i.e.

$$\mathbb{E}[P_B(j)] = \frac{\alpha(j)}{\not{\beta}j) + \alpha(j)} \prod_{1 \leq l < j} \frac{\not{\beta}j)}{\not{\beta}j) + \alpha(j)} = p(1-p)^{j-1} \tag{8.61}$$

where $\alpha(j) = p_1, \not{\beta}j) = p_{\ 2}$ and $.5 = p_1/(p_1 + p_2)$. In particular we assumed a low precision by setting $p_1 = p_2 = .001$. Hence the prior is centered on a distribution with is rather different from the true batch-size distribution, but the data will dominate the prior guess quickly. For the Bootstrap approximation we used $M = 50000$ Bootstrap replicas drawn from the predictive bath-size distribution.

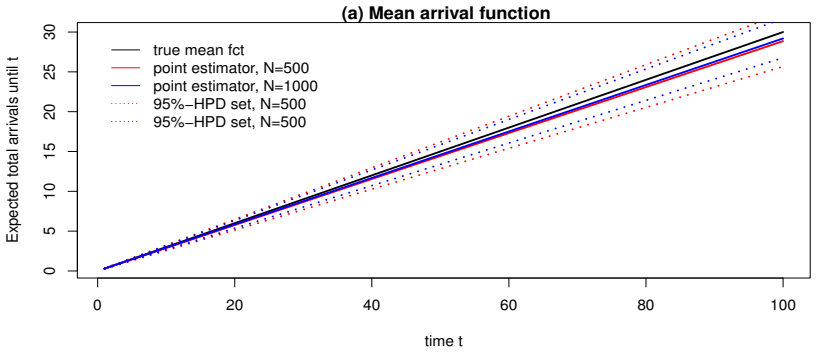
For the MCMC-simulation of service-time distribution we used a burn-in period of 2000 iterations. After the burn-in we ran 90 Markov chains (using

available parallel computing techniques). Each single chain has a total length of only 60 iterations giving a total of 5400 iterations. The autocorrelation plots showed very little correlation across iterations. Therefore we do not use any thinning.

Performance measure in transient state

We first show the estimator of performance measures in transient-state. The arrival process $N_C(t) = \sum_{j \leq N(t)} B_j$ will be treated first. The true mean-function of the arrival process is given by $E_0[N_c(t)] = .3t$. The mean function was estimated as described in section 8.4.1 and is shown in figure 8.5.

Figure 8.1: Predicted mean-function (solid red and blue lines), 95-percent point-wise credibility interval (dotted red and blue lines) of the compound arrival process and the true mean function (solid black line).



The black line shows the true mean function and the solid blue and red lines show the posterior mean for 500 and 1000 observations. By construction the accuracy of the estimator will decrease as time t increases. Even for a sample size of $N_1 = 500$ the point estimator and credibility intervals are quite acceptable.

The predicted probabilities of the total number of arrivals $N_c(t)$ at fixed values $t = 20, 40, 60$ and 80 are shown in figure 8.5. The probability of no

arrival until t can be estimates directly from the relation

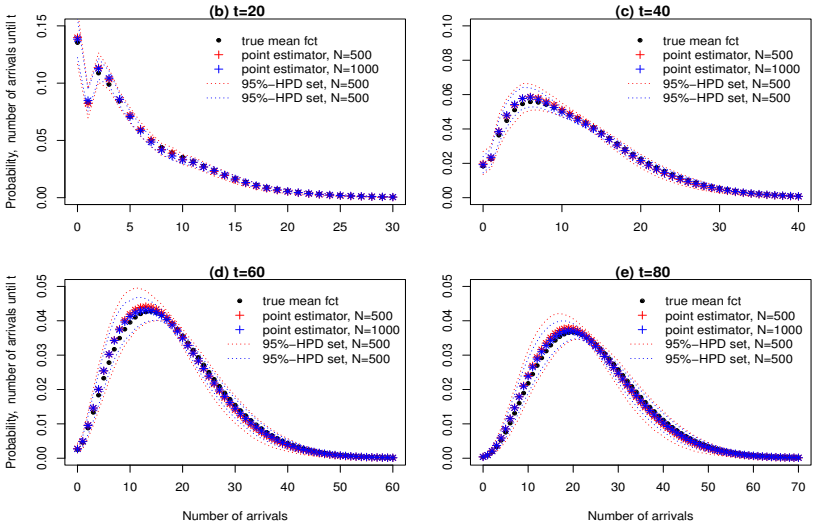
$$P(N_c(t) = 0|\lambda, P_B) = P(N(t) = 0|\lambda) = P(IA_1 > t|\lambda) = \exp\{-\lambda t\}. \quad (8.62)$$

Taking expectation with respect to the posterior, we get

$$\mathbb{E}[P(N_c(t) = 0|\lambda, P_B)|\mathbf{IA}, \mathbf{B}] = (1 + t\lambda_{MLE}/N)^{-N} \to \exp\{-\lambda_0 t\} \ a.s.$$

as $N \to +\infty$, which is a strongly consistent estimator for the true probability function $P_0(N_C(t) = 0)$. The remaining probabilities are estimated as described in the previous section.

Figure 8.2: Predicted probability of arrivals at $t = 20, 40, 60$ and 80 (red and blue crosses), 95-percent point-wise credibility interval (dotted red and blue lines) of the compound arrival process and the true probabilities (solid black points).



The true probability distribution can be recovers quite well. The increase in sample size helps reducing the length if the credibility intervals.

Traffic intensity and performance measures in steady-state

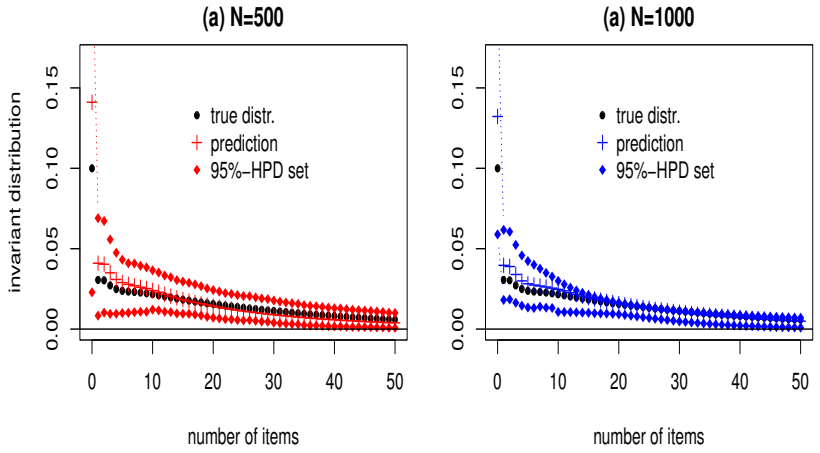
We now turn to the analysis of the stability of the system summarized by the posterior of the traffic intensity ρ. A summary of the posterior distribution is given in table 8.2. Even for $N = 500$ the posterior concentrates almost all its mass on the interval $[0, 1)$. For a sample size of 1000 observations the posterior become close to normal distributed around the maximum likelihood estimate. The posterior probability in favor of ergodicity of the system is .980 and .996 for 500 and 1000 observations. Any series couple of weights a_0 and a_1 for the weighted loss function would lead to acceptance of the ergodicity of the system.

Table 8.2: Posterior probability for the traffic intensity (true value $\rho_0 = .9$).

Sample	Posterior Quantile					MC-		$P(\rho < 1 \mid \text{data})$
Size	.025	.25	.5	.75	.975	Mean	MLE	
500	0.746	0.815	0.858	0.901	0.989	0.860	0.820	0.980
1000	0.786	0.837	0.867	0.898	0.959	0.868	0.867	0.996

The predicted distribution of the number of items in the system at departure epochs at steady-state is shown in figure 8.3. Note that the predicted probability of the event of no items in the queue is strongly overestimated. This may seem unreasonably, but hast to be expected. The true probability of no items in the system equals one minus the traffic intensity, which it therefore .1. But we choose relative non-informative prior. Therefore the predicted traffic intensity follows closely the MLE of the traffic intensity which is .86. Hence the predicted probability of no items in the system is given by .14. This basically means that the data did not provide precise information about the invariant distribution. And an increase in precision can only be obtained by better prior information, which we tried to avoid since we know the true model, or by an increase in sample size.

Figure 8.3: Predicted probability of the number of items in the system at steady-state at departure epochs (red and blue crosses) and the true probability (black dots).



8.6 Conclusions

In the present paper we proposed a semi-parametric analysis of the $M^X/G/1$ bulk queue. The bath-size and service-time distribution was modelled as a beta-Stacy process and a Poisson-Dirichlet process mixture model with Gamma kernel. Prediction for summary measure of the queuing model in transient- and steady-state are obtained by a combination of Bootstrap approximations and MCMC simulations. Our approach provides a flexible model for the unknown service-time and bath-size distribution without a restrictive assumption on a particular class of parametric models. Possible extensions will focus on alternative tools to approximate the posterior distribution of the service-time cdf in order to speed up the computational time. This may be obtained by Approximate Bayes Computations (ABS) or by variational methods. This will be necessary in order to study more complex data structures where data may be observed from several exchangeable queuing systems. For such problems hierarchical models of the same form as

above may be appropriate but this requires a shorter computational time.

8.7 Appendix

8.7.1 Sampling $\chi \sim \mathcal{L}(\chi|\mathbf{S})$

Note that, if G is a Poisson-Dirichlet process with parameters (a, b, G_0), then a draw of $\chi = (\chi_1, \cdots, \chi_N)$ from $\mathbb{P}(d\chi) = \mathbb{E}[\prod_{1\leq i\leq N} G(d\chi_i)]$ can be obtained through the generalized Pólya-urn scheme [127]

$$\begin{aligned} \chi_1 &\sim G_0(\cdot|\eta) \\ [\chi_i|\chi_l, l<i] &\sim \sum_{j=1}^{k_i} \frac{n_{j,i}-a}{b+i-1}\delta_{\chi^*_{j,i}} + \frac{b+k_i a}{b+i-1}G_0(\cdot|\eta) \quad \text{for} \quad 2\leq i\leq N \end{aligned}$$

where $(\chi^*_{j,i})_{j=1}^{k_i}$ are the unique values among $(\chi_l)_{l<i}$ which occurred with frequencies $n_{j,i} = |\{l < i : \chi_l = \chi^*_{j,i}\}|$ for $1 \leq j \leq k_i$. Since there are ties, one usually introduces cluster indicators, i.e. $c_l = j$ if $\chi_l = \chi^*_{i,j}$. Hence, a draw of χ within a Gibbs sampler can be implemented by sampling first the partition $[(c_i)_1^N|(\chi^*_j)_1^{K_N}, \mathbf{S}]$ and secondly sampling the unique values $[(\chi^*_j)_1^{K_N}|(c_i)_1^N, \mathbf{S}]$.

The full conditional for the cluster memberships is given by

$$\mathbb{P}[c_i = j|c_{-i}, \chi^*_{-i}, \mathbf{S}] \propto \begin{cases} (n_{j,-i}-a)f(S_i|\chi^*_{j,-i}) & \text{if } 1\leq j\leq k_{-i} \\ (b+ak_{-i})\int f(S_i|\chi)G_0(d\chi|\eta) & \text{if } j = k_{-i}+1 \end{cases}$$

where χ^*_{-i} denote the unique values among $(\chi)_{j\neq i}$ with frequencies $n_{j,-i} = |\{l \neq i : \chi_l = \chi^*_j\}|$ and $f(S_i|\chi^*_{j,-i})$ stands for the gamma density with shape parameter $\nu^*_{j,-i}$, mean $\mu^*_{j,-i}$ and $\chi^*_{j,i} = (\nu^*_{j,-i}, \mu^*_{j,-i})$. Unfortunately, for the gamma distribution the integral is intractable even for the standard conjugated prior G_0 [112, 130]. One could use numerical integration, but following [120] there are several reasons against this approach.

Instead we apply algorithm 8 in [120] to the Poisson-Dirichlet process to sample the clustering step. The algorithm introduces auxiliary variables which leave the full conditional distribution marginally unchanged. The step of partition and unique cluster representatives is summarized in the first two points of table 8.3.

It is not always possible to accumulate sufficient prior information to

choose appropriate parameters (a, b, η). Especially the parameter η of the base measure is crucial for a good mixing of the Markov chain. We essentially choose to open a new cluster in step $(i.b)$ if $f(S_i|\chi)$ is large as a function of $\chi \sim G_0(\cdot|\eta)$. Hence we may add two additional steps in the MCMC scheme and treat $\eta \sim P(d\eta)$ as a random variable itself. Similar we let $a \sim \text{Unif}(0,1)$ and $b|a$ be a truncated and shifted exponential distribution with support $(-a, b_{max})$ and rate λ. Therefore we add two more steps to our MCMC scheme in table 8.3.

The base measure $G_0(\cdot|\eta)$ and the full conditional $\mathbb{P}(d\chi^*|\mathbf{S}, c)$

Given the data $\mathbf{S}$ and a partition c the likelihood for the unique cluster representatives χ_j^* is given by

$$\prod_{i:c_i=j} f(S_i|\chi_j^*) \propto \frac{(\nu_j^*)^{n_j\nu_j^*} P(S,j)^{\nu_j^*}}{\Gamma(\nu_j^*)^{n_j}} \, (\mu_j^*)^{-n_j\nu_j^*} \exp\left\{-\frac{1}{\mu_j^*} T(S,j)\nu_j^*\right\}$$

where $P(S,j) := \prod_{i:c_i=j} S_i$ and $T(S,j) := \sum_{i:c_i=j} S_i$. The conjugated prior for μ_j^* is the inverse-gamma prior with parameters $(\vartheta_1, \vartheta_2)$ and the full conditional becomes

$$\mathbb{P}(d\mu_j^*|\nu_j^*, \mathbf{S}, c) = \text{InvGam}(d\mu_j^*|\vartheta_1 + \nu_j^* n_j, \vartheta_2 + \nu_j^* T(S,j)). \tag{8.67}$$

The full conditional for ν_j^* is given by

$$\begin{aligned}\mathbb{P}(d\nu_j^*|\mathbf{S}, c) &= \frac{P(d\chi_j^*|\mathbf{S}, c)}{P(d\mu_j^*|\nu_j^*, \mathbf{S}, c)} \\ &\propto \frac{\Gamma(\vartheta_1 + \nu_j^* n_j)}{\Gamma(\nu_j^*)^{n_j}} \frac{(\nu_j^*)^{n_j\nu_j^*} P(S,j)^{\nu_j^*}}{(\vartheta_2 + T(S,j)\nu_j^*)^{\vartheta_1 + n_j\nu_j^*}} P(d\nu_j^*).\end{aligned} \tag{8.68}$$

There is no standard conjugated prior for ν_j^*. We take a gamma prior $\text{Gam}(\psi_1, \psi_2)$ and the full-conditional (8.68) becomes

$$\mathbb{P}(d\nu_j^*|\mathbf{S}) \propto \frac{\Gamma(\vartheta_1 + \nu_j^* n_j)}{\Gamma(\nu_j^*)^{n_j}} \frac{(\nu_j^*)^{\psi_1 + n_j\nu_j^* - 1} e^{-\nu_j^*(\phi_2 - \log P(S,j))}}{(\vartheta_2 + T(S,j)\nu_j^*)^{\vartheta_1 + n_j\nu_j^*}}. \tag{8.69}$$

Table 8.3: **MCMC algorithm to sample the partition c, cluster representatives χ^* and hyper-parameter (a, b, η)**

- Fix an $m > 0$ over all iterations and sample at a fixed iteration t as follows:

(i) for i in $1, \cdots, N$

(a.1) if $c_i^{(t-1)} = j$ and $n_{j,-i}^{(t)} \geq 1$, draw $(\chi_{j,-i}^*)_{j=k_{-i}+1}^{k_{-i}+m} \overset{iid}{\sim} G_0(\cdot|\eta^{(t-1)})$.

(a.2) if $c_i^{(t-1)} = j$ and $n_{j,-i}^{(t)} = 0$, set $\chi_{k_{-i}+1,-i}^* = \chi_j^{*(t-1)}$ and draw $(\chi_{j,-i}^*)_{j=k_{-i}+2}^{k_{-i}+m} \overset{iid}{\sim} G_0(\cdot|\eta^{(t-1)})$.

(b) draw $c_i^{(t)} \in \{1, \cdots, k_{-i} + m\}$ from

$$\mathbb{P}(c_i^{(t)} = j | c_{-i}^{(t)}, \chi^{*(t-1)}, \mathbf{S}) = \begin{cases} h(n_{j,-i} - a) f(S_i|\chi_{j,-i}^{*(t-1)}) & \text{if } 1 \leq j \leq k_{-i} \\ h\dfrac{(b + k_{-i}a)}{m} f(S_i|\chi_{j,-i}^*) & \text{if } k_{-i} < j \leq k_{-i} + m \end{cases} \tag{8.63}$$

where $c_{-i}^{(t)} = \{c_l^{(t)}\}_{l<i} \cup \{c_l^{(t-1)}\}_{l>i}$ and remove all χ_j^*'s without cluster members.

(ii) Resample the unique values $(\chi_j^{*(t)})_1^{k_N}$, i.e of $j = 1, \ldots, k_N$ draw

$$\mathbb{P}[d\chi_j^{*(t)}|\mathbf{c}^{(t)}, \mathbf{S}] \propto G_0(d\chi_j^{*(t)}|\eta^{(t-1)}) \prod_{i:c_i^{(t)}=j} f(S_i|\chi_j^{*(t)}) \tag{8.64}$$

(iii) Given the unique values $\chi^* = (\chi_j^{*(t)})_1^{k_N}$ sample $\eta^{(t)}$ from the full conditional

$$\mathbb{P}(d\eta|\chi^{*(t)}) \propto P(d\eta) \prod_{j=1}^{k_N} G_0(\chi_j^{*(t)}|\eta) \tag{8.65}$$

(iv) Given the cluster-membership $n^{(t)} = (n_1^{(t)}, \cdots, n_{k_N^{(t)}}^{(t)})$ sample

$$\mathbb{P}(d(a,b)|n^{(t)}) \propto \mathbb{P}(d(a,b)) \frac{\prod_{l=0}^{k_N^{(t)}-1}(b + al)}{b^{[N]}} \prod_{j=1}^{k_N^{(t)}} (1-a)^{[n_j^{(t)}-1]} \tag{8.66}$$

The density isn't a standard density and not log-concave either. Hence we need to sample from (8.69) with a Metropolis-Hastings step within the Gibbs sampler. As suggested in [139] we propose a candidate ν_j^P from a Gamma density with shape r and mean $\nu_j^{*(t-1)}$ (hence $Var[\nu_j^P|\nu_j^{*(t-1)}] = E[\nu_j^P|\nu_j^{(t-1)}]^2 r^{-1}$).

Choice of the prior $P(d\eta)$ and the resulting $P(d\eta|\chi^*)$

For the base measure $G_0(d\chi|\eta) = Gam(d\nu|\psi_1,\psi_2)InvGam(d\mu|\vartheta_1,\vartheta_2)$ the hyper-parameter η equals $(\psi_1,\psi_2,\vartheta_1,\vartheta_2)$. The likelihood for η is given by

$$\prod_{j=1}^{k_N} G_0(d\chi_j^*|\eta) \propto \psi_2^{k_N\psi_1}\exp\{-\psi_2 T(\nu^*)\}\frac{P(\nu^*)^{\psi_1}}{\Gamma(\psi_1)^{k_N}} \tag{8.70}$$

$$\times\, \vartheta_2^{k_N\vartheta_1}\exp\{-\vartheta_2 T(\mu^*)\}\frac{P(\mu^*,2)^{\vartheta_1}}{\Gamma(\vartheta_1)^{K_N}} \tag{8.71}$$

where $T(\nu)=\sum_{j=1}^{k_N}\nu_j^*$, $T(\mu^*)=\sum_{j=1}^{k_N}1/\mu_j^*$ and $P(\nu^*)=\prod_{j=1}^{K_N}\nu_j^*$, $P(\mu^*)=\prod_{j=1}^{k_N}1/\mu_j^*$. The likelihood for η is similar to the likelihood of χ^*. The conjugated prior for (ψ_2,ϑ_2) are independent Gamma distributions with parameters $(\tau_{i,1},\tau_{i,2})$ for $i=1,2$ and the full-conditional for $(d\psi_2,d\vartheta_2)$ becomes

$$\mathbb{P}(d\psi_2,d\vartheta_2|\psi_1,\vartheta_1,\chi^*) = \mathrm{Gam}(d\psi_2|\tau_{1,1}+k_N\psi_1,\tau_{1,2}+T(\nu^*)) \tag{8.72}$$

$$\times\, \mathrm{Gam}(d\vartheta_2|\tau_{2,1}+k_N\vartheta_1,\tau_{1,2}+T(\mu^*)). \tag{8.73}$$

We use independent prior distributions for (ψ_1,ϑ_1) and the full conditional for (ψ_1,ϑ_1) is given by

$$\mathbb{P}(d\psi_1,d\vartheta_1)|\chi^*) = \frac{\mathbb{P}(d\psi_1,d\psi_2,d\vartheta_1,d\vartheta_2)|\chi^*)}{\mathbb{P}(d\psi_2,d\vartheta_2)|\psi_1,\vartheta_1,\chi^*)}$$

$$\propto \frac{\Gamma(\tau_{1,1}+k_N\psi_1)}{\Gamma(\psi_1)^{k_N}}\frac{P(\nu^*)^{\psi_1}}{(\tau_{1,2}+T(\nu^*))^{k_N\psi_1+\tau_{1,1}}}\mathbb{P}(d\psi_1) \tag{8.74}$$

$$\times \frac{\Gamma(\tau_{2,1}+k_N\vartheta_1)}{\Gamma(\vartheta_1)^{k_N}}\frac{P(\mu^*)^{\vartheta_1}}{(\tau_{2,2}+T(\mu^*))^{k_N\vartheta_1+\tau_{21}}}\mathbb{P}(d\vartheta_1). \tag{8.75}$$

ψ_1 and ϑ_1 are conditionally independent and if $P(d\psi_1)$ and $P(d\vartheta_1)$ are log-concave, then the full-conditionals will be log-concave as well (see [130] for a formal demonstration of this result). In particular we choose $\psi_1 \sim \text{Exp}(\lambda_1)$ and $\vartheta_1 \sim \text{Exp}(\lambda_2) + (1+\epsilon)$ for $\epsilon > 0$, where the support of ϑ_1 was restricted to $(1+\epsilon, \infty)$ in order we ensure that $E[\mu^*|\vartheta_1, \vartheta_2] < \infty$. Hence we can sample first (ψ_1, ϑ_1) independently by adaptive rejection sampling or by black-box sampling methods for log-concave densities as developed [39, 40] and then sample (ψ_2, ϑ_2) given ψ_1, ϑ_1 as independent Gamma random variables.

Sampling $P(d(a,b)|c)$ via latent variables

We introduce auxiliary random vectors $(\mathbf{V}, \mathbf{W}, \mathbf{U})$ which leave the full conditional of (a,b) marginal unchanged. The joined density is given by

$$P(d(a,b), \mathbf{V}, \mathbf{W}, \mathbf{U}|(n_j)_1^{k_N})$$
$$\propto P(da,db) \prod_{l=1}^{k_N-1} I(V_l < b + al) \prod_{l=1}^{N-1} I(W_l < (b+l)^{-1}) \prod_{j=1}^{k_N} \prod_{l=0}^{n_j-2} I(U_{j,l} < 1 - a + l).$$

Hence we can sample a, b and the augmented variables as

$$\begin{aligned}
[V_l|(a,b)] &\sim Unif(0, b+al) && 1 \le l \le K_N - 1 \\
[W_l|b] &\sim Unif(0, (b+l)^{-1}) && 1 \le l \le N-1 \\
[W_{j,l}|a] &\sim Unif(0, 1-a+l) && 0 \le l \le n_j - 2, 1 \le j \le K_N \\
[a|\mathbf{V}, \mathbf{W}, \mathbf{U}, b] &\sim Unif(a_{\min}, a_{\max}) \\
[b|\mathbf{V}, \mathbf{W}, \mathbf{U}, a] &\sim Exp(\lambda) I(b_{\min} < b < b^*_{\max})
\end{aligned}$$

where

$$b_{\min} = \max_{1 \le l \le K-1} (V_l - al, -a), \quad b^*_{\max} = \min_{\le l \le N-1} (1/W_l - l, b_{max}), \tag{8.76}$$

$$a_{\min} = \max_{1 \le l \le K-1} (0, (V_l - b)/l), \quad a_{\max} = 1 + \min_{1 \le j \le K} \min_{0 \le l \le n_j - 2} (l - U_{j,l}). \tag{8.77}$$

If N become to large it will be faster and more efficient to sample the full conditional directly by a Metropolis-Hastings step.

8.7.2 Proofs

Proof of Lemma 33. Note first that by the strong law of large numbers $M(j)/M \to P_{B,n}(j)$ a.s. and $\tilde{M}(j)/M \to 1 - F_{B,n}(j)$ a.s. as $M \to +\infty$. Therefore

$$\frac{M(j)}{\tilde{M}(j-1)} \to \frac{\alpha_N(j)}{\alpha_N(j)+\beta_N(j)} \ a.s. \ \text{and} \tag{8.78}$$

$$\frac{\tilde{M}(j)}{\tilde{M}(j-1)} \to \frac{\beta_N(j)}{\alpha_N(j)+\beta_N(j)} \ a.s. \text{ as } M \to +\infty. \tag{8.79}$$

Or equivalently $(\alpha^*_M(j), \beta^*_M(j)) \to (\alpha_N(j), \beta_N(j))$ a.s. as $M \to +\infty$.
Now for $M \geq 1$ let $Q_{B,M} = \{Q_{B,M}(j), j \geq 1\}$ be a discrete beta-Stacy process with parameter $\alpha^*_M(\cdot)$ and $\beta^*_M(\cdot)$. The $Q_{B,M}(j) = \theta^*_{M,j} \prod_{l<j}(1-\theta^*_{M,l})$ where $\{\theta^*_{M,j}$ are independent beta random variables with parameter $(\alpha^*_M(j), \beta^*_M(j))$. Fix $j \geq 1$, then

$$\begin{aligned}
\lim_{M\to+\infty} \mathbb{E}\big(\exp\{u\theta^*_{M,j}\}|\mathbf{B}\big) &= 1 + \lim_{M\to+\infty} \sum_{k\geq 1} \left(\prod_{l=1}^{k-1} \frac{\alpha^*_M(j)+l}{\alpha^*_M(j)+\beta^*_M(j)+l} \right) \frac{u^k}{k!} \\
&= 1 + \sum_{k\geq 1} \lim_{M\to+\infty} \left(\prod_{l=1}^{k-1} \frac{\alpha^*_M(j)+l}{\alpha^*_M(j)+\beta^*_M(j)+l} \right) \frac{u^k}{k!} \\
&= 1 + \sum_{k\geq 1} \left(\prod_{l=1}^{k-1} \frac{\alpha_N(j)+l}{\alpha_N(j)+\beta_N(j)+l} \right) \frac{u^k}{k!} \quad a.s. \\
&= \mathbb{E}\big(\exp\{u\theta_j\}\big|\mathbf{B}\big),
\end{aligned}$$

where $\theta_j \sim \text{Beta}(\alpha_N(j), \beta_N(j))$ and the interchange of limit and sum follows from the Dominated convergence theorem ($\{u^k/k!\}_k$ is proportional to a Poisson probability mass function with parameter u and the integrand is bounded by one for all k). Since $\{\theta^*_{M,j}\}_{j\geq 1}$ are independent, we have that for

every $k \geq 1$

$$\begin{aligned}\mathbb{E}\big(\exp\{\mathbf{u}'(\theta^*_{m,1},\ldots,\theta^*_{m,k})\}|\mathbf{B}\big) &= \prod_{j=1}^{k}\mathbb{E}\big(\exp\{u_j\theta^*_{m,j}\}|\mathbf{B}\big)\\ &\to \prod_{j=1}^{k}\mathbb{E}\big(\exp\{u_j\theta_j\}\big|\mathbf{B}\big)a.s.\\ &= \mathbb{E}\big(\exp\{\mathbf{u}'(\theta_1,\ldots,\theta^*_k)\}\big|\mathbf{B}\big).\end{aligned}$$

Since $[0,1]^\infty$ is separable and the class of finite dimensional sets is a convergence determining class ([20], p.19), given $\mathbf{B}$, the random element $\Theta^*_M = \{\theta^*_{M,j}\}_{j\geq 1}$ converges weakly to $\Theta = \{\theta_j; j \geq 1\}$ as a random element on $[0,1]^\infty$ as $M \to +\infty$.

Now define for every j the map $f_j : [0,1]^\infty \to [0,1]$ by $\to f_j(\mathbf{x}) = x_j\prod_{l<j}(1-x_l)$ which is easily seen to be continuous. For a fixed $j \geq 1$, by the continuous mapping theorem ([129], Theorem 12, p. 70), given $\mathbf{B}$,

$$Q_{B,M}(j) = f_j(\Theta^*_M) \xrightarrow{d} f_j(\Theta) = \theta_j\prod_{l=1}^{j-1}(1-\theta_l) =: P_B(j). \tag{8.80}$$

Similarly for $f : [0,1]^\infty \to [0,1]^\infty$ defined by $f(\mathbf{x}) = (f_j(\mathbf{x}), j \geq 1)$ an application of the continuous mapping theorem yields that, given $\mathbf{B}$,

$$\{Q_{B,M}(j), j \geq 1\} = f(\Theta^*_M) \Longrightarrow f(\Theta) = \{P_B(j), j \geq 1\} \tag{8.81}$$

weakly as a random element on $[0,1]^\infty$. This implies, by definition of convergence in distribution on $[0,1]^\infty$, that for every bounded and continuous function g we have that

$$E[g(Q^*_{B,M})|\mathbf{B}] \to E[g(P_B)|\mathbf{B}] \quad \text{as } M \to +\infty. \tag{8.82}$$

Furthermore, let $Q_{B,M}(\cdot) = \sum_{k\geq 1} Q_{B,m}(k)\delta_k(\cdot)$ be defined as a random element on $M_P(\mathbb{N})$ and let $f : \mathbb{N} \to [0,\infty)$ be a bounded, measurable function vanishing outside a compact set. Necessarily $f = \sum_{j=1}^{k} a_j\delta_{i_j}$ for some $k < \infty$,

positive integers $i_1, \dots i_k$ and positive real numbers $a_1, \dots, a_k$. Form (8.82)

$$\begin{aligned} E[\exp\{\int f dQ^*_{M,B}\}|\mathbf{B}] &= E[\exp\{\sum_{j=1}^{k} a_j Q^*_{B,M}(i_j)\}|\mathbf{B}] \\ &\to E[\exp\{\sum_{j=1}^{k} a_j P_B(i_j)\}|\mathbf{B}] \\ &= E[\exp\{\int f dP_B\}|\mathbf{B}]. \end{aligned}$$

Which shows that $Q^*_{B,M}(\cdot) \Longrightarrow P_B(\cdot)$ weakly in $M_P(\mathbb{N})$ as $M \to +\infty$. □

Proof of Lemma 34. Since for every j the sequence $\theta_{M,j}$ converges in distribution to θ_j as $M \to +\infty$ and $\{\theta_{M,j}\}_M$ is uniformly integrable we also have that $\mathbb{E}[\theta_{M,j}|\mathbf{B}] \to \mathbb{E}[\theta_j|\mathbf{B}]$. Hence for $S_M(j) := \prod_{l\le j}(1-\theta_{M,j})$ we have that

$$\mathbb{E}[Q_M(j)|\mathbf{B}] \to P_{B,N}(j) \;\text{ and }\; \mathbb{E}[S_M(j)|\mathbf{B}] \to S_{B,N}(j) = 1 - F_{B,N}(j) \quad (8.83)$$

as $M \to +\infty$. Fix $\epsilon > 0$ and choose $\delta \in (0, \epsilon)$. Since, for every fixed M, we have that $1 \ge \mathbb{E}[S_M(j)|\mathbf{B}] \downarrow 0$ as j increases and $\sum_{j>1} \mathbb{E}[S_M(j)|\mathbf{B}] < \infty$ we can find positive integers K and M_0 such that

$$\sum_{j>K} \mathbb{E}[S_M(j)|\mathbf{B}] < \delta^2/4 \;\; \forall M > M_0 \quad \text{and} \quad \sum_{j>K} S_{B,N}(j) < \delta^2/4 \qquad (8.84)$$

Let $t \ge 0$ be a point of continuity of $\mathbb{P}(\mu_B \le \cdot|\mathbf{B})$. Now, the following reasoning starts similar to [32], proof of lemma 1. Observe that for $M > M_0$

$$\begin{aligned} &\mathbb{P}\Big(\mu_{B,M} \le t|\mathbf{B}\Big) \\ &= \mathbb{P}\Big(\mu_{B,M} \le t, \sum_{j>K} S_M(j) > \epsilon|\mathbf{B}\Big) + \mathbb{P}\Big(\mu_{B,M} \le t, \sum_{j>K} S_M(j) \le \epsilon|\mathbf{B}\Big) \\ &\le \mathbb{P}\Big(\sum_{j>K} S_M(j) > \epsilon|\mathbf{B}\Big) + \mathbb{P}\Big(1 + \sum_{1\le j\le K} \prod_{l<j}(1-\theta_{l,M}) \le t + \epsilon|\mathbf{B}\Big) \\ &\le \epsilon^{-1} \sum_{j>K} \mathbb{E}[S_M(j)|\mathbf{B}] + \mathbb{P}\Big(1 + \sum_{1\le j\le K} \prod_{l<j}(1-\theta_{l,M}) \le t + \epsilon|\mathbf{B}\Big). \end{aligned}$$

Hence, using the fact that $\sum_{1\leq j\leq K}\prod_{l<j}(1-\theta_{l,M}) \to^{\mathcal{L}} \sum_{1\leq j\leq K}\prod_{l<j}(1-\theta_l)$ as $M\to+\infty$

$$\limsup_M \mathbb{P}\Big(\mu_{B,M}\leq t|\mathbf{B}\Big) \leq \epsilon/2 + \mathbb{P}\Big(1+\sum_{1\leq j\leq K}\prod_{l<j}(1-\theta_j)\leq t+\epsilon|\mathbf{B}\Big).$$

Let $\epsilon\to 0$ and $K\to+\infty$ and using the fact that t is a point of continuity of $\mathbb{P}(\mu_B\leq\cdot|\mathbf{B})$ we obtain

$$\limsup_M \mathbb{P}(\mu_{B,M}\leq t|\mathbf{B}) \leq \mathbb{P}(\mu_B\leq t|\mathbf{B}). \tag{8.85}$$

On the other hand

$$\begin{aligned}
&\mathbb{P}(\mu_B\leq t-\epsilon|\mathbf{B}) && (8.86)\\
&\leq \mathbb{P}\Big(\sum_{j>K}S_B(j)>\epsilon|\mathbf{B}\Big)+\mathbb{P}\Big(\mu_B\leq t-\epsilon,\sum_{j>K}S_B(j)\leq\epsilon|\mathbf{B}\Big) && (8.87)\\
&\leq \epsilon/2+\mathbb{P}\Big(1+\sum_{1\leq j\leq K}\prod_{l<j}(1-\theta_l)\leq t|\mathbf{B}\Big) && (8.88)
\end{aligned}$$

and therefore as $M\to+\infty$

$$\mathbb{P}(\mu_B\leq t-\epsilon|\mathbf{B}) \leq \epsilon/2+\liminf_M \mathbb{P}\Big(1+\sum_{1\leq j\leq K}\prod_{l<j}(1-\theta_{l,M})\leq t|\mathbf{B}\Big). \tag{8.89}$$

Again, let $\epsilon\to 0$ and $K\to+\infty$, with t being a point of continuity of $\mathbb{P}(\mu_B\leq\cdot|\mathbf{B})$, we obtain

$$\mathbb{P}(\mu_B\leq t|\mathbf{B}) \leq \liminf_M \mathbb{P}\Big(\mu_{M,B}\leq t|\mathbf{B}\Big). \tag{8.90}$$

□

Proof of Lemma 3. We show first that $(Q^*_M(s_1),Q^*_M(s_2),\cdots,Q^*_M(s_n))$ convergence in distribution to $(P^*_B(s_1),P^*_B(s_2),\cdots,P^*_B(s_n))$ for any n and any $0\leq s_1<s_2<s_n\leq 1$. Secondly we show that $\{Q^*_M(\cdot)\}_M$ is tight which by theorem 7.5 in [20] gives the result.

We show only the case $n=1$ for n arbitrary the computations are similar.

To show point-wise convergence in distribution we modify the proof of lemma 33. Fix ϵ, choose $\nu \in (0, \epsilon)$. Since $\mathbb{E}[Q_{B,N}(j)|\mathbf{B}] \to P_{B,N}(j)$ for every $j \geq 1$ and $\sum_{j\geq 1} \mathbb{E}[Q_{B,N}(j)|\mathbf{B}] = 1$ we can find positive integers M_0 and K such that

$$\sum_{j>K} P_{B,N}(j) < \nu^2/4 \quad \text{and} \quad \sum_{j>K} \mathbb{E}[Q_{B,N}(j)|\mathbf{B}] < \nu^2/4 \ \ \forall M > M_0. \tag{8.91}$$

Fix $s \in [0,1]$, and without loss of generality we may assume that $0 < t < 1$ (since $Q^*_M(1) = P^*_B(1) = 1$ and $Q^*_M(0) = P^*_B(0) = 0$ a.s.). Let t be a point of continuity of $\mathbb{P}(P^*_B(s) \leq \cdot|\mathbf{B})$, then

$$\begin{aligned}
&\mathbb{P}\Big(Q^*_M(s) \leq t|\mathbf{B}\Big) \\
&= \mathbb{P}\Big(Q^*_M(s) \leq t, \sum_{j\geq K} Q_M(j)s^j > \epsilon|\mathbf{B}\Big) + \mathbb{P}\Big(Q^*_M(s) \leq t, \sum_{j\geq K} Q_M(j)s^j \leq \epsilon|\mathbf{B}\Big) \\
&\leq \mathbb{P}\Big(\sum_{j\geq K} Q_M(j)s^j > \epsilon|\mathbf{B}\Big) + \mathbb{P}\Big(\sum_{1\leq j\leq K} Q_M(j)s^j \leq t+\epsilon|\mathbf{B}\Big) \\
&\leq \mathbb{P}\Big(\sum_{j\geq K} Q_M(j) > \epsilon|\mathbf{B}\Big) + \mathbb{P}\Big(\sum_{1\leq j\leq K} Q_M(j)s^j \leq t+\epsilon|\mathbf{B}\Big) \\
&\leq \epsilon^{-1}\sum_{j>K} \mathbb{E}[Q_M(j)|\mathbf{B}] + \mathbb{P}\Big(\sum_{1\leq j\leq K} Q_M(j)s^j \leq t+\epsilon|\mathbf{B}\Big).
\end{aligned}$$

Hence as $M \to +\infty$

$$\limsup_M \mathbb{P}\Big(Q^*_M(s) \leq t|\mathbf{B}\Big) \leq \epsilon/2 + \mathbb{P}\Big(\sum_{1\leq j\leq K} P_B(j)s^j \leq t+\epsilon|\mathbf{B}\Big). \tag{8.92}$$

Now, let $\epsilon \to 0$ such that $K \to +\infty$ and use the fact that t is a point of continuity of $\mathbb{P}(Q^*_M(s) \leq \cdot|\mathbf{B})$, we obtain

$$\limsup_M \mathbb{P}\Big(Q^*_M(s) \leq t|\mathbf{B}\Big) \leq \mathbb{P}\Big(P^*_B(s) \leq t|\mathbf{B}\Big). \tag{8.93}$$

Similar to the proof of lemma 33 the opposite inequality can be obtained

starting from

$$\begin{aligned}
&\mathbb{P}(P_B^*(s) \leq t-\epsilon|\mathbf{B}) \\
&\leq \mathbb{P}\Big(\sum_{j>K} P_B(j)s^j > \epsilon|\mathbf{B}\Big) + \mathbb{P}\Big(P_B^*(s) \leq t-\epsilon, \sum_{j>K} P_B(j)s^j \leq \epsilon|\mathbf{B}\Big) \\
&\leq \epsilon/2 + \mathbb{P}\Big(\sum_{1\leq j\leq K} P_B(j)s^j \leq t|\mathbf{B}\Big)
\end{aligned}$$

and therefore

$$\mathbb{P}(P_B^*(s) \leq t-\epsilon|\mathbf{B}) \leq \epsilon/2 + \liminf_M \mathbb{P}\Big(\sum_{1\leq j\leq K} Q_M(j)s^j \leq t|\mathbf{B}\Big).$$

Now, let $\epsilon \to 0$ such that $K \to +\infty$, we also obtain

$$\mathbb{P}(P_B^*(s) \leq t-\epsilon|\mathbf{B}) \leq \liminf_M \mathbb{P}(Q^*(s) \leq t|\mathbf{B}). \tag{8.94}$$

To show tightness we use theorem 7.4 [20] and its Corollary in [20]. We need to show that for any fixed $t \in [0,1)$ the following condition is satisfied

$$\begin{aligned}
&\forall\epsilon > 0\ \forall\delta > 0\ \exists\nu \in (0,1)\ \exists M_0\ : \\
&\nu^{-1}\mathbb{P}(\sup_{s\in[t,t+\nu]} |Q_M^*(s) - Q_M^*(t)| \geq \epsilon|\mathbf{B}) \leq \delta\ \ \forall M \geq M_0.
\end{aligned}$$

Now, using the fact that $(t-\nu)^j - t^j = t\nu\sum_{l=0}^{j-1}\binom{j}{l}t^{l-1}\nu^{j-1-l} < \nu(t+\nu)^{j-1}$ we can start similar as [32, 33]. We have

$$\begin{aligned}
\mathbb{P}(\sup_{s\in[t,t+\nu]} |Q_M^*(s) - Q_M^*(t)| \geq \epsilon|\mathbf{B}) &= \mathbb{P}\Big(\sum_{j\geq 1} Q_M(j)((t+\nu)^j - t^j) \geq \epsilon|\mathbf{B}\Big) \\
&\leq \mathbb{P}\Big(\sum_{j\geq 1} Q_M(j)(t+\nu)^{j-1} \geq \epsilon/\nu|\mathbf{B}\Big) \\
&\leq \mathbb{P}\Big(\sum_{j\geq 1} Q_M(j) \geq \epsilon/\nu|\mathbf{B}\Big) \\
&\leq \nu^2\epsilon^{-2}\mathbb{E}\Big(\Big[\sum_{j\geq 1} Q_M(j)\Big]^2|\mathbf{B}\Big)
\end{aligned}$$

Since $\sum_{j\geq 1} Q_M(j) = 1$ a.s. also $Y_M := \left(\sum_{j\geq 1} Q_M(j)\right)^2 = 1$ a.s. and hence $E[Y_M^2|\mathbf{B}] = 1$. Therefore

$$\nu^{-1}\mathbb{P}(\sup_{s\in[t,t+\nu]} |Q_M^*(s) - Q_M^*(t)| \geq \epsilon|\mathbf{B}) \leq \nu/\epsilon^2 \leq \delta \tag{8.95}$$

for $\nu < \min\{\delta\epsilon^2, 1-t\}$. □

Chapter 9

Conclusions

In the present thesis we tried to use the theory of stochastic processes with reinforcement and in particular Pólya urn processes to analyze multiple-state processes and queuing processes. We used only very simple tools, namely systems of finite-color Pólya urns. In the first part we started with univariate state-spaces and extended the analysis to the bivariate and multivariate case. We used the sequence of states between successes stopping times to model an exchangeable sequence of mixtures of time-inhomogeneous Markov chains. This is equivalent to regenerative cycles and the strong Markov property for ordinary Markov chains. But for ordinary Markov chains there is no learning about future cycles form past cycles. The use of reinforcement makes it possible to predict summary measure of future cycles by using the information from the past cycles summarized by the transition counts. All models considered in the first are discrete time processes.

In the second part we introduced a continuous time stochastic process with reinforcement which was used for approximate prediction for the $M/G/1$ queue. We developed a new technique to simulate trajectories from a Beta-Stacy random distribution function which simplified the implementation substantially. The estimation of summary measures of the queue can be done without the use of a MCMC sampler which is usually required for Bayesian analysis of $M/G/1$ queues. The quality of the estimator are quite accurate for a medium sample size. The sample size for queuing problems lies

usually within several thousand observations. In a second step we used Bayesian methods to analyze the $M^X/G/1$ bulk queue. A Beta-Stacy random probability was used to model the random bath-size probability and a Poisson-Dirichlet mixture model was used to approximate the service-time density. For the prediction of summary measures we developed a new Bootstrap method to sample functions of the random bath-size distribution which was motivated by Bootstrap techniques developed in [33, 78].

Future work could be directed towards the analysis of multi-state models in continuous time using reinforced stochastic processes. One very interesting approach in this direction was proposed in [118]. The authors modelled the usual semi-Markov transition kernel

$$P(T_n \leq t, X_n = j | X_{n-1} = i) = \Pi_{i,j} F_{i,j}(t), \tag{9.1}$$

where Π is some homogeneous transition matrix and $F_{i,j}$ is some cdf for each fixed state i, j, in the very interesting opposite direction

$$P(T_n \leq t, X_n = j | X_{n-1} = i) = \Pi_{i,j}(t) F_i(t) \tag{9.2}$$

where $\Pi(\cdot)$ is now a time dependent transition matrix, i.e. $\Pi(t)$ is a transition matrix for every fixed t. [118] modelled in an elegant way a sequence of mixture of semi-Markov processes. First, for every state i, a sequence of exchangeable holding times in state i is generated, $(T_{i,n})_{n\geq 1}$ say. Suppose that the holding time to state i at the n-th visit of state i is $T_{i,n}$. Then the successor state to state i is drawn by using a continuum of Pólya urns $U(t)_i$ $t \geq 0$ and some time transformation function $h_i()$. In particular the n-th successor state of state i will be the color of the ball drawn from the urn $U(\tilde{T}_{i,n})_i$ where $\tilde{T}_{i,n} = h_i(T_{n,i})$. There are several way of choosing the function $h_i(\cdot)$. A particular choice could be $h_i(t) = 1$, $h_i(t) = \sum_{n\geq 1} t_j I_{[t_j, t_{j+1})}(t)$ or $h_i(t) = t$. Choosing the identity function would probably be the natural choice. But this choice would have the disadvantage that if the actual observed holding times are continuous, then every Pólya urn would only be visited once with probability one. This makes prediction inefficient for the

case $h_i(t) = t$. This would actually be the case for every continuous function $h_i(\cdot)$. Clearly the choice $h_i(t) = 1$ or $h_i(t) = \sum_{n\geq 1} t_j I_{[t_j,t_{j+1})}(t)$ are legitimate choices. The random semi-Markov kernel (9.2) would reduce to $\Pi_{i,j}F_i(dt)$ for the first choice and $\Pi_{i,j}(t_j)F_i(dt)$, if $t \in [t_j, t_{j+1})$, for the second choice. I personally think that it could be interesting to study if there are other ways to construct mixture of semi-Markov chains using reinforced stochastic processes without the need of time-transformation functions like h_i.

For the analysis of queuing models, it is possible to extend the process formulation to more general queues by using a supplementary variable technique. These technique extends the original process to a higher dimensional process which can be analysis with a continuous time semi-Markov chains. One could introduce processes keeping track of the length of the current service times for each of the c counter of the $M/G/c$ queue at the arrival- and departure-times. The analysis of a $G/G/1$ queue with a non-parametric prior for the service-time and inter-arrival times could be of interest as well. But it will be necessary to focus in faster computational tools on order to simulate from the posterior.

Bibliography

[1] O. O. Aalen, O. Borgan, and H. K. Gjessing. *Survival and Event History Analysis: A Process Point of View*. Springer, 2008.

[2] J. Abate. Numerical inversion of Laplace transforms of probability distributions. *Journal on Computing*, 7:36–43, 1995.

[3] J. Abate and W. Whitt. The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems*, 10:5–88, 1992.

[4] J. Abate and W. Whitt. Numerical inversion of probability generating functions. *Operations Research Letters*, 12:245–251, 1992.

[5] D. Aldous. *Exchangeability and Related Topics*. Springer, 1985.

[6] G. Aletti, C. May, and P. Secchi. On the distribution of the limit proportion for a two-color, randomly reinforced urn with equal reinforcement distributions. *Advances in Applied Probability*, 39:690 – 707, 2007.

[7] C. Armero. *Bayesian Statistics 2*, chapter Bayesian Analysis of $M/M/1/\infty/FIFO$ queues, pages 613–617. North Holland, 1985.

[8] C. Armero and M. Bayarri. Prior assessments for prediction in queues. *The Statistician*, 43:139–153, 1994.

[9] C. Armero and M. Bayarri. *Bayesian Statistics 5*, chapter Bayesian Questions and Answers in Queues (with discussions), pages 3–23. Oxford University Press, 1996.

[10] C. Armero and M. J. Bayarri. Bayesian prediction in $M/M/1$ queues. *Queueing Systems*, 15:401–417, 1994.

[11] C. Armero and D. Conesa. Prediction in Markovian bulk arrival queues. *Queueing Systems 34*, 34:327 – 350, 2000.

[12] C. Armero and D. Conesa. Bayesian hierarchical models in manufacturing bulk service queues. *Journal of Statistical Planning and Inference*, 136:335 – 354, 2006.

[13] B. Arthure and Y. E. Y. Kaniovskii. A generalized urn problem and its applications. *Cybernetics*, 19:61–71, 1983.

[14] S. Asmussen. *Applied Probability and Queues*. Springer, 2003.

[15] M. Ausin, M. Wiper, and R. Lillo. Bayesian estimation for the $M/G/1$ queue using a phase-type approximation. *Journal of Statistical Planning and Inference*, 118:83–101, 2004.

[16] M. Ausin, M. Wiper, and R. Lillo. Bayesian prediction of the transient behaviour and busy period in short- and long-tailed $GI/G/1$ queueing systems. *Computational Statistics & Data Analysis*, 52:1615–1635, 2008.

[17] S. Bacallado. Bayesian analysis of variable-order, reversible Markov chains. *Annals of Statistics*, 39:838–864, 2011.

[18] T. Banerjee, M.-H. Chen, D. K. Dey, and S. Kim. Bayesian analysis of generalized odds-rate hazards models for survival data. *Lifetime Data Analysis*, 13:241 – 260, 2007.

[19] A. Beggs. On the convergence of reinforcement learning. *Journal of Economic Theory*, 122:1 – 36, 2005.

[20] P. Billingsley. *Convergence of Probability Measure*. Wiley, 1999.

[21] D. Blackwell and J. B. MacQueen. Ferguson distributions via Polya urn schemes. *The Annals of Statistics*, 1:353–355, 1973.

[22] L. Breiman. *Probability*. SIAM, 1992.

[23] H. Buehlmann. Austauschbare stochastische variablen und ihre grenzwertsaetze. *University California Publications in Statistics*, 3:31 – 36, 1960.

[24] P. Bulla and P. Muliere. Bayesian nonparametric estimation for reinforced Markov renewal processes. *Statistical Inference for Stochastic Process*, 10:283–303, 2007.

[25] P. Bulla, P. Muliere, and S. Walker. Bayesian nonparametric estimation of a bivariate survival function. *Statistica Sinica*, 17:427–444, 2007.

[26] P. Bulla, P. Muliere, and S. Walker. A Bayesian nonparametric estimator of a multivariate survival function. *Journal of Statistical Planning and Inference 139*, 139:3639–3648, 2009.

[27] R. Butler and A. Huzurbazar. Bayesian prediction for waiting times in stochastic models. *Canadian Journal of Statistics*, 28:311–325, 2000.

[28] L. Cai, M. Hayward, Y. Saito, J. Lubitz, A. Hagedorn, and E. Crimmins. Estimation of multi-state life table functions and their variability from complex survey data using the SPACE Program. *Demographic Research*, 22(6):129 – 158, 2010.

[29] C. Chiang. *The life table and its applications*. Robert E. Krieger Publishing Company, Malabar, Florida, 1984.

[30] C. L. Chiang. *Introduction to Stochastic Processes in Biostatistics*. Wiley, 1968.

[31] R. J. Connor and J. E. Mosimann. Concepts of independence for proportions with a generalization of the dirichlet distribution. *Journal of the American Statistical Association*, 64:94–206, 1969.

[32] P. L. Conti. Large sample Bayesian analysis for Geo/G/1 discrete-time queueing models. *Annals of Statistics*, 27:1785–1807, 1999.

[33] P. L. Conti. Bootstrap approximations for Bayesian analysis of $Geo/G/1$ discrete-time queueing models. *Journal of Statistical Planning and Inference*, 120:65–84, 2004.

[34] D. Coppersmith and P. Diaconis. Random walk with reinforcement. unpublished, 1986.

[35] P. Damien and S. G. Walker. Sampling truncated normal, beta, and gamma densities. *Journal of Computational and Graphical Statistics*, 10(2):206–215, 2001.

[36] B. Davis. Reinforced random walk. *Probability Theory and Related Fields*, 84:203–229, 1990.

[37] B. de Finetti. Funzione caratteristica di un fenomeno aleatorio. In *Memorie, Classe di Scienze Fisiche, Mathematice e Naturale,*, volume 6, pages 251 – 299. Atti della R. Academia Nazionale dei Lincei, 1931.

[38] B. de Finetti. *Studies in Inductive Logic and Probability*, chapter On the Condition of Partial Exchangeability, pages 193–205. University of California Press, 1980.

[39] L. Devroye. A simple algorithm for generating random variables with a log-concave density. *Computing*, 33:247–257, 1984.

[40] L. Devroye. A note on generating random variables with log-concave densities. Technical report, School of Computer Science, McGill University, 2012.

[41] L. Devroye and L. F. James. The double CFTP method. *ACM Trans. Model. Comput. Simul.*, 21(2):10:1–10:20, Feb. 2011.

[42] P. Diaconis. *Bayesian Statistics 3*, chapter Recent Progress on de Finetti's Notions of Exchangeability, pages 111–125. Oxford University Press, 1988.

[43] P. Diaconis and D. Freedman. Finite exchangeable sequences. *The Annals of Probability*, 8:745–/64, 1980.

[44] P. Diaconis and D. A. Freedman. De Finetti's theorem for Markov chains. *The Annals of Probability*, 8:115–130, 1980.

[45] P. Diaconis and S. W. Rolles. Bayesian analysis for reversible Markov chains. *The Annals of Statistics*, 34:1270–1292, 2006.

[46] K. Doksum. Tail-free and neutral random probabilities and their posterior distributions. *The Annals of Probability*, 2:183–201, 1974.

[47] L. E. Dubins and D. A. Freedman. Random distribution functions. *The Annals of Mathematical Statistics*, 34:548–551, 1963.

[48] L. E. Dubins and D. A. Freedman. Exchangeable processes need not be mixtures of independent, identically distributed random variables. *Probability Theory and Related Fields*, 48:115–132, 1979.

[49] M. Eckhoff and S. W. Rolles. Uniqueness of the mixing measure for a random walk in a random environment on the positive integers. *Electronic Communications in Probability*, 14:31–35, 2009.

[50] F. Eggenberger and G. Polya. Ueber die statistik verketteter vorgaenge. *Zeitschfrift fuer Angewandte Mathematik und Mechanik*, 3:279–289, 1923.

[51] I. Epifani, S. Fortini, and L. Ladellic. A characterization for mixtures of semi-Markov processes. *Statistics and Probability Letters*, 60:445–457, 2002.

[52] M. D. Escobar. *Estimating the Means of Several Normal Populations by Nonparametric Estimation of the Distribution of the Means*. PhD thesis, Yale University, Department of Statistics, 1988.

[53] M. D. Escobar. Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association*, 89:268–277, 1994.

[54] T. S. Ferguson. A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1:209–230., 1973.

[55] T. S. Ferguson. Prior distributions on the space of probability measures. *The Annals of Statistics*, 2:616–629, 1974.

[56] T. S. Ferguson. *Recent Advances in Statistics*, chapter Bayesian Density Estimation by Mixtures of Normal Distributions, pages 287–303. Academic Press, 1983.

[57] T. S. Ferguson and M. J. Klass. A representation of independent increment processes without Gaussian components. *Annals of Mathematical Statistics*, 43:1634 – 1643, 1972.

[58] T. S. Ferguson and E. G. Phadia. Bayesian nonparametric estimation based on censorded data. *The Annals of Stastistics*, 7:163–186, 1979.

[59] S. Fortin, L. Ladelli, and E. Regazzini. Exchangeability, predictive distributions and parametric models. *Sankhya*, 62:86 – 109, 2000.

[60] S. Fortini, L. Ladelli, G. Petris, and E. Regazzini. On mixtures of distributions of Markov chains. *Stochastic Processes and their Applications*, 100:147–165, 2002.

[61] S. Fortini and S. Petrone. Hierarchical reinforced urn processes. *Statistics and Probability Letters*, 82:1521 – 1529, 2012.

[62] D. Freedman. On the asymptotic behaviour of Bayes' estimates in the discrete cases. *The Annals of Mathematical Statistics*, 34:1386–1403, 1963.

[63] D. A. Freedman. Invariants under mixing which generalize de Finetti's theorem. *The Annals of Mathematical Statistics*, 33:916–923, 1962.

[64] D. A. Freedman. Poisson processes with random arrival rate. *The Annals of Mathematical Statistics*, 33:924–929, 1962.

[65] D. A. Freedman. Invariants under mixing which generalize de Finetti's theorem: Continuous time parameter. *The Annals of Mathematical Statistics*, 34:1194–1216, 1963.

[66] D. A. Freedman. De Finetti's theorem in continuous time. In *Statistics, probability and game theory: Papers in honor of David Blackwell*, 1996.

[67] A. E. Gelfand and A. Kottas. A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 11:289–305, 2002.

[68] A. Gelman. Objections to Bayesian statistics. *Bayesian Analysis*, 3:445 – 450, 2008.

[69] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, 2004.

[70] J. K. Ghosh and R. Ramamoorthi. *Bayesian Nonparametrics*. Springer, 2003.

[71] R. M. Gray. *Probability, Random Processes and Ergodic Properties*. Springer, 2009.

[72] A. Guglielmi, C. C. Holmes, and S. G. Walker. Perfect simulation involving functionals of a Dirichlet process. *Journal of Computational and Graphical Statistics*, 11:306–310, 2002.

[73] E. Hewitt and L. J. Savage. Symmetric measures on cartesian products. *Transactions of the American Mathematical Society*, 80:470 – 501, 1955.

[74] N. L. Hjort. Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 18:1259 – 1294, 1990.

[75] N. L. Hjort, C. Holmes, P. MÃ$\frac{1}{4}$ller, and S. G. Walker, editors. *Bayesian Nonparametrics*. Cambridge University Press, 2011.

[76] F. M. Hoppe. Polya-like urns and the Ewens' sampling formula. *Journal of Mathematical Biology*, 20:91–94, 1984.

[77] D. R. Insua, M. Wiper, and F. Ruggeri. Bayesian analysis of $M/E_r/1$ and $M/H_k/1$ queues. *Queueing Systems*, 30:289–308, 1998.

[78] H. Ishwaran and L. F. James. Gibbs sampling methods for stick-breaking prior. *Journal of the American Statistical Association*, 96:161–173, 2001.

[79] H. Ishwaran and L. F. James. Computational methods for multiplicative intensity models using weighted gamma process: Proportional hazards, marked point processes, and panel count data. *Journal of the American Statistical Association*, 99:175 – 190, 2004.

[80] C. Jackson. *Multi-state modelling with R: the msm package (available on http://cran.r-project.org)*, 2007.

[81] L. James. Poisson calculus for spatial neutral to the right processes. *The Annals of Statistics*, 34:416–440, 2006.

[82] L. F. James, A. Lijoi, and I. PrÃ$\frac{1}{4}$nster. Distributions of linear functionals of two parameter Poisson-Dirichlet random measures. *Annals of Applied Probability*, 18:521 – 551, 2008.

[83] N. Johnson and S. S. Kotz. *Urn models and Their Applications: An Approach to Modern Discrete Probability Theory*. Wiley, 1977.

[84] O. Kallenberg. Characterizations and embedding properties in exchangeability. *Probability theory and Related Fields*, 60:249–281, 1982.

[85] O. Kallenberg. *Probabilistic Symmetries and Invariance Principles*. Springer, 2005.

[86] M. Kalli, J. E. Griffin, and S. G. Walker. Slice sampling mixture models. *Statistics and Computing*, 21:93–105, 2011.

[87] M. S. Kean and S. W. Rolles. *Infinite dimensional stochastic analysis*, chapter Edge-Reinforced Random Walk on a finite Graph, pages 217–234. Koninklijke Nederlandse Akademie van Wetenschappen, 2000.

[88] D. G. Kendall. Some problems in the theory of queues. *Journal of the Royal Statistical Society. Series B*, 13:151–185, 1951.

[89] D. G. Kendall. Stochastic processes occurring in the theory of queues and their analysis by the method of imbedded Markov chains. *Annals of Mathematical Statistics*, 24:338–354, 1953.

[90] N. Keyfitz. *Applied Mathematical Demography*. Wiley, New York, 1977.

[91] N. Keyfitz. *Introduction to the mathematics of population*. second edition. Addison-Wesley, Reading, 1977.

[92] Y. Kim. Nonparametric Bayesian estimators for counting processes. *The Annals of Statistics*, 27:562–588, 1999.

[93] Y. Kim, L. James, and R. Weissbach. Bayesian analysis of multi-state event history data: Beta-Dirichlet process prior. *Biometrika*, 2011.

[94] Y. Kim and J. Lee. Bayesian analysis of proportional hazard models. *The Annals of Statistics*, 31:493–511, 2003.

[95] J. F. C. Kingman. Uses of exchangeability. *Annals of Probability*, 6:183–197, 1978.

[96] C. H. Kraft. A class of distribution function processes which have derivatives. *Journal of Applied Probability*, 1:385–388, 1964.

[97] M. Lavine. Some aspects of Pólya tree distributions for statistical modelling. *The Annals of Statistics*, 20:1222 – 1235, 1992.

[98] M. Lavine. More aspects of Pólya tree distributions for statistical modelling. *Annals of Statistics*, 22:1161 – 1176, 1994.

[99] J. Lee. Sampling methods of neutral to the right processes. *Journal of Computational and Graphical Statistics*, 16:656–671, 2007.

[100] J. Lee and Y. Kim. A new algorithm to generate beta processes. *Computational Statistics & Data Analysis*, 47:441 – 453, 2004.

[101] A. Lijoi and I. Pruenster. *Bayesian Nonparametric*, chapter Models beyond the Dirichlet process, pages 80–136. Cambridge University Press, 2010.

[102] G. Link. *Studies in Inductive Logic and Probability*, chapter Representation Theorems of the de Finetti Type for (Partial) Symmetric Probability Measures, pages 207–231. University of California Press, 1980.

[103] A. Y. Lo. On a class of Bayesian nonparametric estimates: I. density estimates. *Annals of Statistics*, 12:351–357, 1984.

[104] S. M. Lynch and J. S. Brown. A new approach to estimating life tables with covariates and constructing interval estimates of life table quantities. *Sociological Methodology*, 35:pp. 189–237, 2005.

[105] S. M. LYNCH and J. S. BROWN. Obtaining multi-state life table distributions for highly refined subpopulations from cross-sectional data: A Bayesian extension of Sullivan's method. *Demography*, 47(4):pp. 1053–1077, 2010.

[106] H. M. Mahmoud. *Polya Urn Models*. Chapman & Hall, 2009.

[107] R. D. Mauldin and S. C. Williams. Reinforced random walks and random distributions. *Proceedings of the American Mathematical Society*, 110:251–258, 1990.

[108] C. May and N. Flournoy. Asymptotics in response-adaptive designs generated by a two-color, randomly reinforced urn. *The Annals of Statistics*, 37:1058 – 1078, 2009.

[109] J. Medhi. *Stochastic Models in Queuing Theory*. Academic Press, 2003.

[110] F. Merkl and S. W. Rolles. Recurrence of edge-reinforced random walk on a two-dimensional graph. *The Annals of Probability*, 37:1679–1714, 2009.

[111] M. Mezzetti, P. Muliere, and P. Bulla. An application of reinforced urn processes to determining maximum tolerated dose. *Statistics and Probability Letters*, 77:740–747, 2007.

[112] R. Miller. Bayesian analysis of the two-parameter gamma distribution. *Technometrics*, 22:69–69, 1980.

[113] P. Muliere, A. M. Paganoni, and P. Secch. Randomly reinforced urns for clinical trials with continuous responses. In *SIS - Proceedings of the XLIII Scientific Meeting*, pages 403 – 414, Padova, 2006. Cleup.

[114] P. Muliere, A. M. Paganoni, and P. Secchi. A randomly reinforced urn. *Journal of Statistical Planning and Inference*, 136:1853–1874, 2006.

[115] P. Muliere and P. Secchi. Bayesian nonparametric predictive inference and bootstrap techniques. *The Annals of the Institute of Statistical Mathematics*, 48:663–673, 1996.

[116] P. Muliere and P. Secchi. Weak convergence of a Dirichlet-multinomial process. *Georgian Mathematical Journal*, 10:319–324, 2003.

[117] P. Muliere, P. Secchi, and S. G. Walker. Urn schemes and reinforced random walks. *Stochastic Processes and their Applications*, 88:59 – 78, 2000.

[118] P. Muliere, P. Secchi, and S. G. Walker. Reinforced random processes in continuous time. *Stochastic Processes and their Applications*, 104:117–130, 2003.

[119] P. Muliere, P. Secchi, and S. G. Walker. Partially exchangeable processes indexed by the vertices of a k-tree constructed via reinforcement. *Stochastic Processes and their Applications*, 115:662–677, 2006.

[120] R. M. Neal. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9:249–265, 2000.

[121] J. Norris. *Markov Chains*. Cambridge University Press, 1997.

[122] O. Papaspiliopoulos and G. O. Roberts. Retrospective MCMC for Dirichlet process hierarchical models. *Biometrika*, 95:169 – 186, 2008.

[123] R. Pemantle. Phase transition in reinforced random walk and RWRE on trees. *The Annals of Probability*, 16:1229–1241, 1988.

[124] R. Pemantle. Vertex-reinforced random walk. *Probability Theory and Related Fields*, 92:117–136, 1990.

[125] R. Pemantle. A survey of random processes with reinforcement. *Probability Surveys*, 4:1–49, 2007.

[126] A. Philippe. Simulation of right and left truncated gamma distributions by mixtures. *Statistics and Computing*, 7:173–181, 1997.

[127] J. Pitman. Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory; Papers in honour of David Blackwell*, pages 245–267, 1996.

[128] J. Pitman and M. Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Annals of Probability*, 25:855 – 900, 1997.

[129] D. Pollard. *Convergence of Stochastic Processes*. Springer, 1984.

[130] B. Pradhan and D. Kundu. Bayes estimation and prediction of the two-parameter gamma distribution. *Journal of Statistical Computation and Simulation*, 81:1187 – 1198, 2011.

[131] H. Putter, M. Fiocco, and R. B. Geskus. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med*, 26(11):2389–2430, May 2007.

[132] S. I. Resnick. *Advantures in Stochastic Processes*. Birkhaeuser, 1992.

[133] S. I. Resnick. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, 2007.

[134] A. Rogers. *Introduction to multiregional mathematical demography*. A Wiley-Inter-science publication. Wiley, New York, NY [u.a.], 1975.

[135] S. W. Rolles. How edge-reinforced random walk arises naturally. *Probability Theory and Related Fields*, 126:243–260, 2003.

[136] S. W. Rolles. On the recurrence of edge-reinforced random walk on $\mathbb{Z} \times G$. *Probability Theory and Related Fields*, 135:216–264, 2006.

[137] W. F. Rosenberger and F. Hu. Bootstrap methods for adaptive designs. *Statistics in Medicine*, 18:1757 – 1767, 1999.

[138] S. M. Ross. *Applied Probability Models With Optimization Applications*. Dover Publications, 1992.

[139] F. Ruggeri, M. Wiper, and D. R. Insua. Bayesian model selection for $M/G/1$ queues. In W. Racugno, editor, *Proceedings of the Workshop on Model Selection*. Pitagora, Bologna, 1999.

[140] R. Schoen. *Modeling multigroup populations*. Plenum Press, New York, 1988.

[141] R. Schoen. Practical uses of multi-state population models. *Annual Review of Sociology*, 14:pp. 341–361, 1988.

[142] J. Sethuraman. A constructive definition of Dirichlet priors. *Statistics Sinica*, 4:639–650, 1994.

[143] L. D. Sharples, C. H. Jackson, J. Parameshwar, J. Wallwork, and S. R. Large. Diagnostic accuracy of coronary angiography and risk factors for post-heart-transplant cardiac allograft vasculopathy. *Transplantation*, 76:679–682, 2003.

[144] M. Sung, R. Soyer, and N. Nhan. Bayesian analysis of non-homogeneous Markov chains: Application to mental health data. *Statistics in Medicine*, 26:3000 – 3017, 2007.

[145] V. Susarla and J. V. Ryzin. Nonparametric Bayesian estimation of survival curves from incomplete observations. *Journal of the American Statistical Association*, 71:897–902, 1976.

[146] C. Sutton and M. I. Jordan. Bayesian inference for queueing networks and modeling of internet services. *The Annals of Applied Statistics*, 5:254 – 282, 2011.

[147] L. Trippa, P. Bulla, and S. Petrone. Extended Bernstein prior via reinforced urn processes. *Annals of the Institute of Statistical Mathematics*, 2009.

[148] S. Walker. A nonparametric approach to a survival study with surrogate endpoints. *Biometrics*, 54:662–672, 1998.

[149] S. Walker and P. Damien. Representations of Lévy processes without Gaussian components. *Biometrika*, 87:477 – 483, 2000.

[150] S. Walker and P. Muliere. Beta-Stacy processes and a generalization of the Pólya-urn scheme. *Annals of Statistics*, 25:1762 – 1780, 1997.

[151] S. G. Walker. Sampling the Dirichlet mixture model with slices. *Communications in Statistics - Simulation and Computation*, 36:45 – 54, 2007.

[152] S. G. Walker, P. Damien, P. W. Laud, and A. F. M. Smith. Bayesian nonparametric inference for random distributions and related functions. *Journal of the Royal Statistical Society. Series B*, 61:485–527, 1999.

[153] D. W. Widder. *The Laplace Transform*. Princeton University Press, 1941.

[154] D. Williams. *Probability with Martingales*. Cambridge University Press, 1991.

[155] M. Wiper. Bayesian analysis of $E_r/M/1$ and $E_r/M/c$ queues. *Journal of Statistical Planning and Inference*, 69:65–79, 1998.

[156] M. Wiper, D. R. Insua, and F. Ruggeri. Mixture of gamma distributions with application. *Journal of Computational and Graphical Statistics*, 10:440–454, 2001.

[157] R. Wolpert and K. Ickstadt. *Practical Nonparametric and Semiparametric Statistics*, chapter Simulation of Levy Random Fields, pages 227–242. Springer, 1998.

[158] R. L. Wolpert and K. Ickstadt. Poisson/Gamme random field models for spatial statistics. *Biometrica*, 85:251–267, 1998.

[159] A. Zaman. Urn models for Markov exchangeability. *The Annals of Probability*, 11:223–229, 1984.