THESIS DECLARATION

The undersigned

Stefano, Rizzelli
PhD Registration Number: *1380575*

Thesis Title: *Asymptotic properties of nonparametric and semiparametric statistical methods for the extremal dependence*

PhD in Statistics
30^{th} Cycle
Advisor: *Simone A. Padoan*
Year of Discussion: *2019*

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 Societá 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 *Asymptotic properties of nonparametric and semiparametric statistical methods for the extremal dependence*;
   - by *Stefano Rizzelli*;
   - discussed at Università Commerciale Luigi Bocconi - Milano in *2019*;
   - 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 Universitá Commerciale Luigi Bocconi to reproduce the same for research and teaching purposes, quoting the source;

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) 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.

Date 22/07/2018

Rizzelli, Stefano

Contents

List of Figures

List of Tables

Acknowledgements

I thank Simone A. Padoan for his patient supervision. I am very grateful for his huge dedication during the last two years and for all the bright intuitions and knowledge he has shared with me. I thank Armelle Guillou for her guidance and excellent work. The findings presented in the final chapter of this manuscript are the result of a fruitful collaboration with her and Simone. I thank as well Enkelejd Hashorva, whose work and ideas sparked the investigation presented in Chapter 3. I finally thank Sonia Petrone for her continued support, going well beyond her role of PhD director.

Abstract

The study of the dependence structure associated to a set of random variables represents a crucial topic in probability and statistics. This task is particularly delicate when extremes are concerned. Extreme value theory (EVT) offers a specialized probabilistic description of extreme events. In particular, it provides a set of functional objects which allow to trace the degree and the structure of the extremal dependence. The so-called Pickands dependence function has received particular attention. Recently, several nonparametric estimators of the latter have been proposed, mostly adopting a frequentist approach. Bayesian nonparametric methods for inferring the extremal dependence are still at an early stage. To the best of our knowledge, no results concerning their asymptotic behavior are currently available. Herein, we establish almost-sure posterior consistency for prior distributions of the Pickands dependence function, that exploit polynomial and piecewise polynomial representations. A second main contribution of this work concerns modelling and inference of the extremal dependence beyond the theoretical framework described by classical EVT. Specifically, we consider aggregated data in the form of maxima computed on a random number of observations, deriving a new class of max-stable distributions. We find an explicit connection between the Pickands dependence function of such aggregated variables and that of the non-aggregated counterpart. We construct a class of frequentist semiparametric estimators for inferring the latter and establish their asymptotic properties. Finally, we deal with the notion of asymptotic independence, arising when the limiting distribution of linearly normalized maxima is the product of its marginals. We consider an alternative componentwise maxima approach and derive a new dependence function analogous to the Pickands. We design a frequentist semiparametric estimation procedure and investigate its asymptotic behavior. Furthermore, we propose a test of independence in the classical multivariate extreme value setting, which can be used to detect the occurrence of asymptotic independence.

Chapter 1

Introduction and probabilistic preliminaries

This work is concerned with the analysis of the dependence structure of a random vector, say $\mathbf{X}$, in the tails of its distribution. We refer to the latter as tail or extremal dependence. Such task can be accomplished by studying the dependence among the extremes of a random sample of replicates of $\mathbf{X}$. The notion of extremes that is predominantly used herein is the one of componentwise maxima, described in Section 1.2.1. Multivariate extreme value theory (MEVT) provides a class of limiting distributions for the latter, called multivariate extreme value distributions, having copulas of the form

$$C(\mathbf{u}) = \exp\left\{-\left(\sum_{j=1}^{d}\log u_j\right) A\left(\frac{\log u_1}{\sum_{j=1}^{d}\log u_j}, \dots, \frac{\log u_d}{\sum_{j=1}^{d}\log u_j}\right)\right\},$$

for $\mathbf{u} \in [0,1]^d$. The function A is known as Pickands dependence function - see [56] - and fully characterizes the dependence structure of a multivariate extreme value distribution. The estimation of such function will be one of the main objects of study in the following chapters.

Chapter 2 investigates the construction of strongly consistent Bayesian procedures for inferring A, within the classical multivariate extreme value paradigm. It constitutes a first step towards the understanding of the (frequentist) asymptotic properties of Bayesian nonparametric methods for the extremal dependence. As such, it can be seen as a contribution to the mathematical statistics literature for classical MEVT.

As we move forward, the work drifts away from this modelling framework, going beyond the relatively limited data structures it accomodates. Chapter 3 contributes to the extension of the latter for aggregated data in the form of maxima computed on a random number of observations. In this setting, a new class of max-stable distributions is derived. Their Pickands dependence function is linked to the Pickands dependence function of the max-stable distribution arising as a limit for the usual componentwise maxima. Exploiting this relation, we construct a class of frequentist semiparametric estimators for inferring the latter. The main asymptotic properties of such estimators are established. These findings are part of a joint work coauthored by E. Hashorva and S. A. Padoan.

Finally, Chapter 4 deals with the notion of asymptotic independence, arising when the limiting distribution of usual componentwise maxima factorizes into its marginals. We propose a test for detecting the occurrence of asymptotic independence. Then, to bypass the problems with modelling and estimation arising in this case, we consider an alternative componentwise maxima approach and derive a new dependence function analogous to the Pickands. We design a frequentist semiparametric estimation procedure and describe its asymptotic behaviour. These methodological developments are also presented in a paper cohautored by A. Guillou and S. A. Padoan, published by the Journal of Multivariate Analysis - see [38].

Overall, this work contributes to modelling and inference for the extremal dependence from both a Bayesian and a frequentist stance, adopting a nonparametric/semiparametric approach. A special attention is devoted to the derivation of the asymptotic properties of the proposed inferential methods. As treating extreme value models as the true data generating mechanism often represents a simplistic assumption, a more rigorous approach would require accounting for model misspecification and convergence bias. This would introduce additional technical difficulties, whose solution is here deferred to future work. On this ground, the present analysis aims at paving the way towards the answer to some challenging problems in asymptotic statistics for extremes.

1.1 Notation

Number sets We make use of the following symbol:

- $\mathbb{N}_+$ set of positive integers.

Function sets We denote by

- $C(B)$ the set of continuous function on $B \subset \mathbb{R}^d$;
- $C^+(B)$ the set of positive, continuous function on $B \subset \mathbb{R}^d$;
- $C_b^+(B)$ the set of positive, continuous and bounded functions on $B \subset \mathbb{R}^d$;
- $C^m([a,b])$ the set of functions that have m continuous derivatives on $[a,b]$, $-\infty < a \leq b < +\infty$;
- $\ell^\infty(\mathcal{X})$ the space of bounded real-valued functions on $\mathcal{X}$.

Probability The generic expressions $\mathbb{P}$ and $\mathbb{E}$ are used to denote: by $\mathbb{P}(B)$, the probability of some event B; by $\mathbb{E}X$, the expectation of a random vector X, without making explicit reference to the underlying probability measure. Moreover:

- for any probability measure P on the measurable space $(E, \mathcal{E})$, we denote by Pf the integral $\int_E f(x)P(dx)$, for any measurable function f;
- the above notation is used also for signed measures. In particular, denoting by $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$ the empirical measure of a random sample $(X_i)_{i=1}^n$ drawn from P, we use the abbreviations $\mathbb{P}_n f = n^{-1}\sum_{i=1}^n f(X_i)$ and

$$\sqrt{n}(\mathbb{P}_n - P)f = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n f(X_i) - \int_E f(x)P(dx)\right);$$

- for a sequence of random elements (X_n), we write $X_n = o_p(1)$ to denote both convergence in (outer) probability to 0 in $\mathbb{R}$, and convergence in outer probability to the 0 function in ℓ^∞;
- the symbol "$\rightsquigarrow$" denotes both weak convergence in the space ℓ^∞ and the usual convergence in distribution for random vectors; the symbols "$\xrightarrow{\text{p}}$" and "$\xrightarrow{\text{as}}$" denote usual convergence in probability and with (outer) probability one;
- for a distribution F on $\mathbb{R}^k$ and a k-dimensional random vector X, we occasionally use the short notation $X \overset{d}{\sim} F$ to claim that X is distributed according to F.

Operators and limits We make use of the following mathematical symbols:

- for any $x \in \mathbb{R}$, $\lfloor x \rfloor$ denotes the largest integer smaller than x;
- for a real valued function f on $B \subset \mathbb{R}$, $f^{\leftarrow}$ denotes its left inverse (provided it exists);
- for sequences of real numbers (x_n) and (y_n) we write $x_n \lesssim y_n$ if there exists $n' \in \mathbb{N}_+$ such that y_n dominates x_n up to a positive constant, for every $n \geq n'$;
- for (x_n) and (y_n) as above, we write $x_n \sim y_n$ if

$$\lim_{n\to\infty} \frac{x_n}{y_n} = 1;$$

- for a sequence (x_n), Δx_j denotes the forward difference $\Delta x_j = x_{j+1} - x_j$;
- for a generic set B of real valued functions, conv$B$ denotes its convex hall;
- for measures (μ_n), μ on $(E, \mathcal{E})$, we write $\mu_n \xrightarrow{\text{v}} \mu$ to denote vague convergence of μ_n to μ on a given subspace.

Norms We make use of following notation

- for a vector $\mathbf{x} \in \mathbb{R}^d$, we denote by $\|\mathbf{x}\|_1 = \sum_{j=1}^{d} |x_j|$. For a function f on a subset B of $\mathbb{R}^d$, we denote $\|f\|_1 = \int_B |f(\mathbf{x})| \mathrm{d}\mathbf{x}$;
- similarly, we denote $\|\mathbf{x}\|_\infty = \max_{1\leq j\leq d} |x_j|$ and $\|f\|_\infty = \sup_{\mathbf{x}\in B} |f(\mathbf{x})|$.

Additional norm symbols are introduced in the next chapters, where needed.

1.2 Multivariate extreme value theory

A wide range of applications study the occurrence of extreme events affecting a set of variables, which may be rare but carry the risk of great (e.g. financial, environmental) losses. Typically, a possibly infinite set of centered moments is not sufficient to characterize such large risks and their dependence structure – in case of heavy-tailed distributions moments do not even exist beyond a finite order. Also, the more sophisticated approaches to multivariate modelling which are suitable to describe the mean levels of real-world processes and analyze the bulk of the observations, are usually not ideal to hande extremal aspects.

Extreme value theory offers a specialized probabilistic description of extreme events and provides a mathematical basis for extrapolation beyond the observed levels into the tails of the distribution generating the data. In particular, it provides a set of coefficients and functional objects which allow to trace the degree and the structure of the extremal dependence among a set of variables. Loosely speaking, extremal or tail dependence is the amount of dependence characterizing a multivariate distribution in the upper (lower) regions of the nonnegative (nonpositive) orthant. In the following, we formalize this concept by embedding it within the limit theory of componentwise maxima.

1.2.1 Componentwise maxima

Let $\{\mathbf{X}_i, i \geq 1\}$ be a sequence of independent and identically distributed (i.i.d.) $\mathbb{R}^d$-valued random vectors with distribution F and define the vector of componentwise maxima

$$\mathbf{M}_n := \left(\bigvee_{i=1}^{n} X_{i,1}, \dots, \bigvee_{i=1}^{n} X_{i,d} \right), \quad n \in \mathbb{N}_+.$$

If there exist sequences of norming constants $\mathbf{a}_n > \mathbf{0}$, $\mathbf{b_n} \in \mathbb{R}^d$ and a distribution G with non-degenerate margins such that

$$\lim_{n\to\infty} \mathbb{P}(\mathbf{M}_n \leq \mathbf{a}_n\mathbf{x} + \mathbf{b}_n) = \lim_{n\to\infty} F^n(\mathbf{a}_n\mathbf{x} + \mathbf{b}_n) = G(\mathbf{x}) \tag{1.1}$$

for all continuity points $\mathbf{x}$ of G, we say that G is a multivariate extreme-value distribution (MEVD) and that F is in its max-domain of attraction, in symbols $F \in \mathcal{D}(G)$. In the following, the limiting distribution G is also referred to as the max-attractor of $F_{\mathbf{X}}$ or $\mathbf{X}$. More precisely, G takes the form

$$G(\mathbf{x}) = C(G_1(x_1), \dots, G_d(x_d)), \quad \mathbf{x} \in \mathbb{R}^d,$$

where the marginal distributions G_j, $1 \leq j \leq d$, are members of the family of generalized extreme-value distributions (GEV), i.e.

$$G_j(x) = \exp\left\{ - \left[1 + \xi_j \left(\frac{x - \mu_j}{\sigma_j} \right) \right]^{-1/\xi_j} \right\}, \tag{1.2}$$

defined on the set $\{x : 1 + \xi_j(x - \mu_j)/\sigma_j > 0\}$, where $\xi_j, \mu_j \in \mathbb{R}$, $\sigma_j \in (0, \infty)$ - see e.g. [12]. Furthermore, C is an extreme-value copula, i.e. it admits the following representation

$$C(\mathbf{u}) = \exp\{-L(-\log u_1, \cdots, -\log u_d)\}, \quad \mathbf{u} \in (0,1]^d, \tag{1.3}$$

where $L : [0,\infty)^d \mapsto [0,\infty)$ is the so-called stable dependence function. The function L is a homogenous function of order one, i.e. $L(a\mathbf{z}) = aL(\mathbf{z})$ for every $\mathbf{z} \in [0,\infty)$ and $a > 0$ - see e.g [2, Ch. 8.2.2]. From this property it follows that

$$L(\mathbf{z}) = (z_1 + \cdots + z_d)\, A(\mathbf{t}), \qquad \mathbf{z} \in [0,\infty)^d, \tag{1.4}$$

with $t_j = z_j/(z_1 + \cdots + z_d)$ for $j = 1, \ldots, d-1$, $t_1 = 1 - t_2 - \cdots - t_d$, where A is called Pickands dependence function – see [56] – and represents the restriction of L on the d-dimensional unit simplex,

$$\mathcal{S}_d := \left\{(v_1, \ldots, v_d) \in [0,1]^d : v_1 + \cdots + v_d = 1\right\}.$$

Properties 1.2.1 *The Pickands dependence function satisfies the following properties:*

(A1) *A is continuous and $A(\mathbf{e}_j) = 1$, where $\mathbf{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ is the j^{th} unit vector of $\mathbb{R}^d$;*

(A2) *$A(\mathbf{t})$ has lower and upper bounds $1/d \le \bigvee_{j=1}^d t_j \le A(\mathbf{t}) \le 1$;*

(A3) *A is convex, i.e. $\forall \mathbf{t}, \mathbf{w} \in \mathcal{S}_d, \alpha \in [0,1]$, $\alpha A(\mathbf{t}) + (1-\alpha)A(\mathbf{w}) \ge A(\alpha\mathbf{t} + (1-\alpha)\mathbf{w})$.*

The lower and upper bounds in (A3) represent the cases of complete dependence and independence, respectively. The extremal dependence structure represented by A is synthesized by

$$\theta \equiv \theta(G) := dA(1/d, \ldots, 1/d),$$

the so called extremal coefficient. It ranges from 1 to d, larger values corresponding to looser degrees of extremal dependence. The function A is linked to another important measure of extremal dependence, the so-called angular probability measure, denoted by H. In particular, it holds that

$$A(\mathbf{t}) = d \int_{\mathcal{S}_d} \bigvee_{j=1}^{d} t_j w_j H(\mathrm{d}\mathbf{w}), \quad \mathbf{t} \in \mathcal{S}_d,$$

where H satisfies the mean condition

$$\int_{\mathcal{S}_d} w_j H(\mathrm{d}\mathbf{w}) = 1/d, \quad 1 \leq j \leq d. \tag{1.5}$$

Both A and H are used to describe the dependence structure characterizing a vector of d extreme variables - as componentwise maxima - but they can also be used to describe the dependence in the tail of F and model the probability of large threshold exceedances. Indeed, letting $\mathbf{X} \overset{d}{\sim} F$, under the domain of attraction condition (1.1) it holds that

$$\begin{aligned}
&\lim_{t\to\infty} \mathbb{P}\left(\cup_{j=1}^{d} \{X_j > U_j(tz_j)\} \mid \cup_{j=1}^{d} \{X_j > U_j(t)\}\right) \\
&= \lim_{t\to\infty} \frac{\mathbb{P}\left(\cup_{j=1}^{d} \{X_j > U_j(tz_j)\}\right)}{\mathbb{P}\left(\cup_{j=1}^{d} \{X_j > U_j(t)\}\right)} \\
&= \frac{\int_{\mathcal{S}_d} \bigvee_{j=1}^{d} (w_j/z_j) H(\mathrm{d}\mathbf{w})}{\int_{\mathcal{S}_d} \bigvee_{j=1}^{d} w_j H(\mathrm{d}\mathbf{w})},
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}\left(\cap_{j=1}^{d} \{X_j > U_j(tz_j)\} \mid \cap_{j=1}^{d} \{X_j > U_j(t)\}\right) \\
&= \lim_{t\to\infty} \frac{\mathbb{P}\left(\cap_{j=1}^{d} \{X_j > U_j(tz_j)\}\right)}{\mathbb{P}\left(\cap_{j=1}^{d} \{X_j > U_j(t)\}\right)} \\
&= \frac{\int_{\mathcal{S}_d} \bigwedge_{j=1}^{d} (w_j/z_j) H(\mathrm{d}\mathbf{w})}{\int_{\mathcal{S}_d} \bigwedge_{j=1}^{d} w_j H(\mathrm{d}\mathbf{w})},
\end{aligned}$$

where

$$U_j(z) = F_j^{\leftarrow}(1 - 1/z) \tag{1.6}$$

for $z > 0$, $1 \leq j \leq d$. A formal justification of this fact is provided in [60, Proposition 5.15] and Proposition 1.2.2 in the next subsection.

1.2.2 Max-stability, regular variation

A distribution $G(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$, is said max-stable if for $i = 1, \dots, d$ and every $t > 0$ there exist functions $\alpha_i(t) > 0$, $\beta_i(t)$ such that

$$G^t(\mathbf{x}) = G(\alpha_1(t)x_1 + \beta_1(t), \dots, \alpha_d(t)x_d + \beta_d(t)).$$

According to Proposition 5.9 in [60], the class of multivariate extreme-value distributions arising from (1.1) is precisely the class of max-stable distribution functions with nondegenerate marginals. A characterization of such class of distributions can be given by establishing a connection with the theory of multivariate regular variation.

A measurable function $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is called regularly varying at ∞ with index α if

$$\lim_{t\to\infty} f(tx)/f(t) = x^\alpha,$$

for every $x > 0$. This notion can be extended to the multivariate domain as follows. Let $C \subset \mathbb{R}^d$ be a cone, then a measurable function f is (multivariate) regularly varying on C with limit function λ if, for all $\mathbf{x} \in C$, $\lambda(\mathbf{x}) > 0$ and

$$\lim_{t\to\infty} f(t\mathbf{x})/f(t\mathbf{1}) = \lambda(\mathbf{x}).$$

In fact, the function λ satisfying the above limit condition has to be homogeneous of some order, i.e. $\lambda(s\mathbf{x}) = s^\alpha\lambda(\mathbf{x})$ for every $s > 0$ and $\mathbf{x} \in C$. The following proposition collects some fundamental results presented in [60, Ch. 5], of which we make use throughout the next chapters.

Proposition 1.2.2 **(a)** *Suppose G is a multivariate distribution function with continuous marginals. Define for $j = 1, \dots, d$, $D_j(x) = G_j^{\leftarrow}(e^{-1/x})$, $x > 0$, and*

$$G_*(\mathbf{x}) = G(D_1(x_1), \dots, D_d(x_d)), \quad \mathbf{x} \geq \mathbf{0}.$$

Then G_* has marginal distributions $G_{*,j}(x) = \Phi_1(x)$, with Φ_1 denoting the unit Fréchet distribution, and G is a MEVD if and only if G_* is so.
(b) *Let F be a multivariate distribution with continuous marginals and define U_j, $j = 1, \dots, d$ as in equation (1.6). Let F_* be the distribution defined via*

$$F_*(\mathbf{x}) = F(U_1(x_1), \dots, U_d(x_d)), \quad \mathbf{x} \geq \mathbf{0}.$$

If (1.1) *holds, so that* $F \in \mathcal{D}(G)$*, then* $F_* \in \mathcal{D}(G_*)$*. Conversely, if* $F_* \in \mathcal{D}(G_*)$ *and* G_* *has nondegenerate margins, then* $F \in \mathcal{D}(G)$*.*
(c) $F_* \in \mathcal{D}(G_*)$ *if and only if* $1 - F_*$ *is multivariate regularly varying on the cone* $(0, \infty]^d$ *with limit function* $-\log G_*(\mathbf{x})/(-\log G_*(\mathbf{1}))$*.*
(d) *The two conditions in the previous point are equivalent to*

$$n\mathbb{P}(n^{-1}\mathbf{X}_* \in \cdot) \xrightarrow{\text{v}} \mu_* \quad \text{on} \quad [\mathbf{0}, \infty] \setminus \{\mathbf{0}\} \tag{1.7}$$

where $\mathbf{X}_* \overset{d}{\sim} F_*$ *and* μ_* *is a Radon measure on* $[\mathbf{0}, \infty] \setminus \{\mathbf{0}\}$*. In particular, it must be that* $-\log G_*(\mathbf{x}) = \mu_*([\mathbf{0}, \mathbf{x}]^c)$*, for all* $\mathbf{x} > \mathbf{0}$*.*
(e) *Finally,* (1.7) *holds true if and only if*

$$n\mathbb{P}\left\{(n^{-1}\|\mathbf{X}_*\|_1, \|\mathbf{X}_*\|_1^{-1}\mathbf{X}_*) \in (dr, \mathrm{d}\mathbf{w})\right\} \xrightarrow{\text{v}} r^{-2}dr \times H(\mathrm{d}\mathbf{w}) \quad \text{on} \quad (0, \infty) \times \mathcal{S}_d,$$

where H *denotes a probability measure on* $\mathcal{S}_d$ *satisfying* (1.5)*.*

A major implication of Proposition 1.2.2 is the following. When $F \in \mathcal{D}(G)$, it is possible to transform $\mathbf{X}$ into a random vector

$$\mathbf{X}_* = (U_1^{\leftarrow}(X_1), \dots, U_d^{\leftarrow}(X_d)),$$

whose marginal and joint distributions are in the max-domain of Φ_1 and G_*, respectively. In particular, G_* has the same extreme value copula of G and has the following representation

$$\begin{aligned} G_*(\mathbf{x}) &= \exp\left\{-\int_{\mathcal{S}_d} \bigvee_{j=1}^{d} \left(\frac{w_j}{x_j}\right) H(\mathrm{d}\mathbf{w})\right\} \\ &= \exp\left\{-\left(\sum_{j=1}^{d} \frac{1}{x_j}\right) A(\mathbf{t})\right\} \end{aligned}$$

for $t_j = x_j/(x_1 + \ldots + x_d)$, $j = 1, \ldots, d-1$, $t_d = 1 - t_1 - \ldots - t_{d-1}$ and $\mathbf{x} > \mathbf{0}$. Henceforth, studying the extremal behaviour of $\mathbf{X}_*$ in place of $\mathbf{X}$ allows to get rid from the task of modelling the extremal properties of the marginals, focusing only on the extremal dependence. This argument consitutes the basis of several derivations presented in this work.

Chapter 2

Consistent Bayesian methods for the extremal dependence

Within the univariate block maxima approach, frequentist estimation of the parameters in (1.2) has been extensively studied – see e.g. [12, 18] for a comprehensive account. An illustration of Bayesian inference on (ξ_j, μ_j, σ_j) using block maxima can be found in [2, Ch. 11.5] and [66, Ch. 13.1.6]. Concerning the dependence structure, parametric models have been widely discussed in the literature (e.g. [3]), but lack some flexibility. Recently, increasing attention has been devoted to the develpment of nonparametric estimation methods, both from a frequentist (e.g. [4, 6, 46]) and a Bayesian standpoint. Bayesian nonparametric estimation of A and H using a componentwise maxima approach have been discussed, among others, by [35, 39, 47]. In particular, these works exploit Bernstein polynomial representations of A and H, which are also one of the two main objects of this chapter. Focusing in the bivariate case, we investigate the asymptotic properties of Bayesian nonparametric inference on A for prior distributions that exploit polynomial and piecewise polynomial representations. To the best of our knowledge, this is the first work to establish the consistency of Bayesian nonparametric procedures for the extremal dependence.

The rest of the chapter is organized as follows. Section 2.1 introduces the modelling assumptions that will be adopted throughout the present asymptotic study. Section 2.2 provides technical results concerning the Kulback-Leibler support of a prior on the space of bivariate extreme value densities with standard Fréchet margins, induced by a prior on the space of Pickands dependence functions. In Section 2.3 we establish an almost-sure consistency result for Bayesian nonparametric inference on the extremal dependence using a Bernstein polynomial prior on A. In Section 2.4 we characterize a piecewise polynomial representation of the Pickands dependence function, using parabolic B-splines; a consistency result is provided which parallels that of Section 2.3. Section 2.5 ends this chapter, discussing directions for future research. All the proofs are deferred to Section 2.6.

2.1 Introduction

As in [36, 47], we only consider the subset of all valid angular measures whose elements are absolutely continuous on $(0,1)$. That is, we only consider angular probability measures H which satisfy

$$H([a,b]) = p_0\delta_0([a,b]) + \int_a^b h(w)dw + p_1\delta_1([a,b]), \quad 0 \le a < b \le 1, \tag{2.1}$$

where $p_0, p_1 \in [0,1/2]$, δ_x denotes the Dirac measure at $x \in \mathbb{R}$, and h is a nonnegative Lebesgue integrable function satisying $\int_0^1 h(w)dw = 1 - p_0 - p_1$ and, in light of (1.5),

$$\int_0^1 wh(w)dw = 1/2 - p_1, \quad \int_0^1 (1-w)h(w)dw = 1/2 - p_0. \tag{2.2}$$

The resulting class of angular probability measures is flexible enough to cover most of the dependence structures of practical interest. We denote by $\mathcal{H}$ and $\mathcal{A}$ the classes of the cumulative distribution functions and Pickands dependence function associated to a probability measure satisfying (2.1)-(2.2), respectively. In particular, since $d = 2$, $\mathcal{A}$ coincides with the class of functions $A : [0,1] \mapsto [1/2,1]$ satisfying Properties 1.2.1 and such that

$$A(t) = 1 + 2\int_0^t H(w)dw - t, \quad t \in [0,1], \tag{2.3}$$

for some $H \in \mathcal{H}$ – see [2, p. 269]. Therefore, each $A \in \mathcal{A}$ has first and second derivatives on $(0,1)$ which equal $A'(t) = -1 + 2H(t)$ and $A''(t) = 2h(t)$, respectively. From now on, we denote by A' the continuous extension of the first derivative on $[0,1]$, by imposing

$$A'(0) = 2p_0 - 1, \quad A'(1) = 1 - 2p_1. \tag{2.4}$$

Observe that, in light of (2.1), $A' \in C([0,1])$ and A can be seen as the restriction on $[0,1]$ of a function in $C^1([0,1])$. Consequently, we consider $\mathcal{A}$ as a subset of $C^1([0,1])$.

At this stage, the space $\mathcal{A}$ on which a prior has to be specified has been properly identified. In practice, the prior distributions considered herein are supported on a subset of $\mathcal{A}$ which is dense with respect to the uniform metric, having therefore full (uniform) support. The second ingredient for Bayesian inference on the extremal dependence is represented by a suitable likelihood for extremes. As it is common practice within a componentwise maxima approach, we assume that conditionally on $A \in \mathcal{A}$ the data are i.i.d according to $G_*(\cdot|A)$, a bivariate extreme-value distribution with common unit-Fréchet margins. Consequently the likelihood is derived from the densities

$$g_*(\mathbf{z}|A) = G_*(\mathbf{z}|A)\left(\frac{[A(t)-tA'(t)][A(t)+(1-t)A'(t)]}{(z_1z_2)^2} + \frac{A''(t)}{(z_1+z_2)^3}\right), \tag{2.5}$$

defined for $\mathbf{z} \in (0,\infty)^2$, with $t = z_1/(z_1+z_2)$. Extensions to more realistic settings are discussed in Section 2.5.

2.1.1 Notation

For $f \in C([a,b])$, we denote by

$$\omega(f;h) = \sup\{|f(x)-f(y)| : x, y \in [a,b], |x-y| \le h\}$$

its modulus of continuity. For any pair f, g of functions defined on $[0,1]$, we denote by $d_\infty(f,g) := \sup_{x\in[0,1]} |f(x)-g(x)|$ the uniform metric on $[0,1]$. If f and g are differentiable on $[0,1]$ and twice differentiable on $(0,1)$, with bounded first and second derivatives, we denote by $\|\cdot\|_{2,\infty}$ the norm

$$\|f\|_{2,\infty} := \max\left\{\sup_{x\in[0,1]}|f(x)|, \sup_{x\in[0,1]}|f'(x)|, \sup_{x\in(0,1)}|f''(x)|\right\}$$

and define the metric $\rho_\infty(f,g) := \|f-g\|_{2,\infty}$. For a function f which is differentiable m-times, we occasionally denote by $\mathcal{D}^m f$ the mth order derivative.

For two probability laws F, G which are absolutely continuous with respect to some measure ν, with densities f, g respectively, we denote by

$$K(F,G) := \int_E \log\left(\frac{f(x)}{g(x)}\right) f(x) \mathrm{d}\nu(x)$$

their Kullback-Leibler (K-L) divergence and by

$$d_H(f,g) := \left[\int_E (\sqrt{f(x)} - \sqrt{g(x)})\mathrm{d}\nu(x)\right]^{1/2}$$

the Hellinger distance between their densities.

2.2 KL support of a prior on $\mathcal{A}$

As discussed in [30, Ch. 6], a crucial condition for proving consistency of a Bayesian posterior distribution through extended Schwartz's theorem is that the true probaility density function is in the Kulback-Leibler support of the prior. In the present setting, this condition reduces to

$$\Pi_{\mathcal{A}}\{A : K(g_*(\cdot|A_0), g_*(\cdot|A)) < \epsilon\} > 0, \tag{2.6}$$

where $\Pi_{\mathcal{A}}$ is a prior on $\mathcal{A}$ and A_0 denotes the true Pickands dependence function, associated to the true data generating density $g_*(\cdot|A_0)$. When the above condition is satisfied, we say that $g_*(\cdot|A_0)$ possesses the K-L property relative to $\Pi_{\mathcal{A}}$. In the following, we provide sufficient conditions for a prior $\Pi_{\mathcal{A}}$ to comply with (2.6) for every A_0 in $\mathcal{A}_0 \subset \mathcal{A}$, a subset of sufficiently regular Pickands dependence functions. Precisely, we define $\mathcal{A}_0$ as follows.

Definition 2.2.1 *Let $\mathcal{A}_0$ be the set of $A \in \mathcal{A}$ associated to an angular density h such that:*

(i) $h \in C_b^+((0,1))$;

(ii) $h(0^+) = \lim_{x\downarrow 0} h(x)$ *and* $h(1^-) = \lim_{x\uparrow 1} h(x)$ *exist and are finite;*

(iii) *one of the following is satisfed:*

(iii.a) h *is bounded away from* 0;

(iii.b) $\inf_{x\in(0,1)} h(x) = 0$ *and there exists* $\delta^* > 0$ *such that every* $\delta \in (0, \delta^*)$ *admits values* $\epsilon_{\delta,1}, \epsilon_{\delta,2} \in (0,1)$ *satisfying* $h(x) \geq \delta$ *for* $x \in [\epsilon_{\delta,1}, 1-\epsilon_{\delta,2}]$ *and* $h(x) \leq \delta$ *for* $x \in [0, \epsilon_{\delta,1}) \cup (1-\epsilon_{\delta,2}, 1]$.

Recall that $A''(t) = 2h(t)$, $t \in (0,1)$, henceforth each $A \in \mathcal{A}_0$ can be considered twice continuously differentiable on $[0,1]$ and strictly convex. In particular, A'' is allowed to vanish only on the boundary of the unit interval, in this case being monotone in a right-interval of 0 and a left-interval of 1.

Before stating the main result of this section, we introduce two other subsets of $\mathcal{A}$.

Definition 2.2.2 *Let $\mathcal{A}^+$ be the set of $A \in \mathcal{A}$ satisfying $\|A\|_{2,\infty} < +\infty$, whose angular probability measure and density comply with* (ii)-(iii.a) *and*

(iv) $H(\{0\}) = p_0 > 0, \quad H(\{1\}) = p_1 > 0.$

In particular, observe that, in light of (2.4), each element of $A \in \mathcal{A}^+$ satisfies $A'(0) > -1$, $A'(1) < 1$. Differently from $\mathcal{A}_0$, A'' (and thus h) must be bounded away from 0 but need not be continuous.

Let $\mathcal{A}'$ be a generic $\|\cdot\|_{2,\infty}$-bounded subset of $\mathcal{A}$. The scene is finally set for stating the following.

Theorem 2.2.3 *Let $\mathcal{A}' = \cup_{k \geq k'} \mathcal{A}_k$, for some fixed $k' \in \mathbb{N}_+$. Assume that for every $A \in \mathcal{A}_0 \cap \mathcal{A}^+$ there exist $k'' \geq k'$ and a sequence $(A_k)_{k=k''}^{\infty}$ satisfying*

$$A_k \in \mathcal{A}_k, \ \forall k \geq k''; \qquad \rho_\infty(A, A_k) = o(1), \ k \to +\infty. \tag{2.7}$$

Let $\Pi_{\mathcal{A}}$ be a prior distribution assigning positive mass to every ρ_∞-ball centered at an element of $\mathcal{A}_k$, for all $k \geq k'$. Then $\Pi_{\mathcal{A}}$ satisfies (2.6) *for every $A_0 \in \mathcal{A}_0$.*

In the next two sections, specific examples of the sets $\mathcal{A}_k$ composing $\mathcal{A}'$ will be provided by linear combinations of Bernstein polynomials of degree k and parabolic B-splines with $k-1$ distinct knots. Yet, the reach of Theorem 2.2.3 seems to be broader, possibly accommodating other representations of the Pickands dependence function which provide ρ_∞-approximations of the elements of $\mathcal{A}_0 \cap \mathcal{A}^+$.

Remark 2.2.4 In Definition 2.2.1, point (iii.b), we could have allowed h to vanish also at a single point of the boundary set $\{0,1\}$. With few changes, the proof of Theorem 2.2.3 can be extended to accommodate this instance.

Several well established parametric models for the Pickands dependence function allow unbounded continuous spectral densities. For example, this is the case for the bivariate logistic model (e.g. [2, Ch. 9.2.2]) with dependence parameter larger than 1/2. Under the assumptions of Theorem 2.2.3, the K-L property is actually guaranteed also in such instances.

Corollary 2.2.5 *Let $A_0 \in \mathcal{A}$ be associated to an angular density $h_0 \in C^+((0,1))$ such that $\lim_{x\downarrow 0} h_0(x) = \lim_{x\uparrow 1} h_0(x) = +\infty$. Let $\mathcal{A}'$ and $\Pi_{\mathcal{A}}$ comply with the assumptions of Theorem 2.2.3. Then condition* (2.6) *is satisfied.*

2.3 Polynomial representation

In this section we consider a polynomial representation of the Pickands dependence function. That is, we adopt a representation in the form $\sum_{j=0}^{k} \beta_j b_{j,k}$, where $\beta_0, \ldots, \beta_k$ are linear coefficients and $\{b_{0,k}, \ldots, b_{k,k}\}$ constitutes a suitable basis for the vector space of polynomials of degree less than or equal to k, restricted on $[0,1]$. Polynomial modelling is very convenient in that it permits quick computer calculation, simple differentiation/integration and allows to approximate smooth functions to any degree of accuracy desired. Herein we make use of the Bernstein polynomial basis, due to its shape preserving properties. The use of Bernstein polynomials for Bayesian nonparametric modelling is illustrated in [1, 10, 31, 52, 53, 54], to name a few.

2.3.1 Bernstein polynomials

In this subsection, we list some properties of Bernstein polynomials of which we make use later on. Denote by $b_{j,k}(t) := \binom{k}{j} t^j (1-t)^{k-j}$, for $t \in [0,1]$, $1 \le j \le k$, $k \in \mathbb{N}_+$, the Bernstein polynomial basis of degree k.

Properties 2.3.1 *The basis $(b_{j,k})_{j=0}^{k}$ satisfies the following properties:*

(i) *the polynomials $b_{j,k}$, $j = 1, \ldots, k$, are nonnegative on* $[0,1]$ *and form a partiton of unity, i.e. $\sum_{j=0}^{k} b_{j,k}(t) = 1$;*

(ii) *derivatives of the k-th degree Bernstein polynomials are polynomials of degree $k-1$, i.e. $b'_{j,k}(t) = k(b_{j-1,k-1}(t) - b_{j,k-1}(t))$;*

(iii) $(k+1)b_{j,k}(t) = \mathrm{Be}(t|j+1, k-j+1)$, *where* $\mathrm{Be}(\cdot|a,b)$ *denotes the Beta density function with shape parameters* $a, b > 0$, *and*

$$b_{j,k}(0) = \delta_{j,0}, \quad b_{j,k}(1) = \delta_{j,1},$$

where $\delta_{j,i}$ *is the Kronecker delta function.*

Given a function f on $[0,1]$, we can define the polynomial

$$B_k(f;t) = \sum_{j=0}^{k} f\left(\frac{j}{k}\right) b_{j,k}(t)$$

for each $k \in \mathbb{N}_+$; B_k is called the Bernstein operator - see e.g. [55, Ch. 7].

Properties 2.3.2 *The Bernstein operator* B_k *satisfies the following properties:*

(iv) B_k *is linear, i.e.* $B_k(\lambda f + \mu g) = \lambda B_k(f) + \mu B_k(g)$ *for all functions* f *and* g *defined on* $[0,1]$ *and real* λ, μ.

(v) B_k *is monotone, i.e. whenever* $f \geq g$, $B_k(f) \geq B_k(g)$. *Moreover, if* f *is monotonically increasing (decreasing) on* $[0,1]$, *so is* $B_k(f)$.

(vi) *For any integer* $m \geq 0$, *the* m*-th derivative of* $B_{k+m}(f;t)$ *with respect to* t *is equal to*

$$\mathcal{D}^m B_{k+m}(f;t) = \frac{(m+k)!}{k!} \sum_{j=0}^{k} \Delta^m f\left(\frac{j}{m+k}\right) b_{j,k}(t),$$

where Δ *is the forward difference operator* $\Delta f(t) = f(t+\epsilon) - f(t)$ *applied with step size* $\epsilon = 1/(m+k)$.

(vii) *If* f *is convex on* $[0,1]$, *so is* $B_k(f)$; *moreover,* $B_k(f;t) \geq f(t)$, $0 \leq t \leq 1$.

(viii) *If* $f \in C^m([0,1])$, *for some integer* $m \geq 0$, *then* $\mathcal{D}^m B_k(f;t)$ *converges uniformly to* $\mathcal{D}^m f(t)$ *on* $[0,1]$ *as* $k \to +\infty$.

We now have all the technical tools for constructing consistent Bayesian procedures which exploit a Bernstein polynomial representation of A.

2.3.2 Bayesian inference

We denote

- by $\mathcal{A}_k$ the set of polynomials $\sum_{j=0}^{k} \beta_j b_{j,k}$ whose coefficients satisfy:

  (P1) $\beta_0 = \beta_k = 1 \geq \beta_j$, $j = 1, \ldots, k-1$;

  (P2) $\beta_1 = \frac{k-1+2p_0}{k}$ and $\beta_{k-1} = \frac{k-1+2p_1}{k}$, for some $p_0, p_1 \in [0, 1/2]$;

  (P3) $\beta_{j+2} - 2\beta_{j+1} + \beta_j \geq 0$, $j = 0, \ldots, k-2$;

- by $\mathcal{H}_{k-1}$ the set of polynomials $\sum_{j=0}^{k-1} \eta_j b_{j,k-1}$ whose coefficients satisfy:

  (P4) $0 \leq p_0 = \eta_0 \leq \eta_1 \leq \ldots \leq \eta_{k-1} = 1 - p_1 \leq 1$, for some $p_0, p_1 \in [0, 1/2]$;

  (P5) $\sum_{j=0}^{k-1} \eta_j = k/2$.

Then, by Propositions 3.1-3.3 in [47], $\cup_{k=1}^{\infty} \mathcal{A}_k$ and $\cup_{k=1}^{\infty} \mathcal{H}_k$ are dense subsets of the spaces $(\mathcal{A}, d_\infty)$ and $(\mathcal{H}, d_\infty)$, respectively. Moreover, for every $A_k \in \mathcal{A}_k$ it is possible to recover a polynomial $H_{k-1} \in \mathcal{H}_{k-1}$ and vice versa, by means of the following relation among coefficients:

$$\eta_j = \frac{k}{2}\left(\beta_{j+1} - \beta_j + \frac{1}{k}\right), \quad \beta_{j+1} = \frac{1}{k}\left(2\sum_{i=0}^{j} \eta_j + k - j - 1\right), \tag{2.8}$$

for $j = 0, \ldots, k-1$, with $\beta_0 = 1$ in the latter case. These facts have the following consequences. For each $k \in \mathbb{N}_+$, define $\mathcal{B}_k := \{(\beta_0, \ldots, \beta_k) : \text{(P1)-(P3) are satisfied}\}$. Let Π be a prior distribution that assigns positive mass to each integer k above some $k' \in \mathbb{N}_+$ and subset $B \subset \mathcal{B}_k$. Then, the prior distribution of the Pickands dependence function, $\Pi_{\mathcal{A}}$, induced by Π, has full support on $(\mathcal{A}, d_\infty)$. Similarly, define $\mathcal{E}_{k-1} := \{(\eta_0, \ldots, \eta_{k-1}) : \text{(P4)-(P5) are satisfied}\}$, for every $k \in \mathbb{N}_+$. Let $\widetilde{\Pi}$ be a prior assigning positive mass to each integer k above some $k' \in \mathbb{N}_+$ and subset $E \subset \mathcal{E}_k$. Then, the prior distribution of the angular distribution, $\widetilde{\Pi}_{\mathcal{H}}$, induced by $\widetilde{\Pi}$, has full support on $(\mathcal{H}, d_\infty)$. We refer to $\Pi_{\mathcal{A}}$ and $\Pi_{\mathcal{H}}$ as Bernstein polynomial priors. Moreover, any prior Π on $\cup_{k \geq k'}(\{k\} \times \mathcal{B}_k)$ induces a prior $\widetilde{\Pi}$ on $\cup_{k \geq k'-1}(\{k\} \times \mathcal{E}_k)$ and vice versa; thus, specifying a Bernstein polynomial prior on $\mathcal{A}$ is equivalent to doing so on $\mathcal{H}$.

We now provide an almost sure-consistency result for the posterior distribution on the space $(\mathcal{A}, \|\cdot\|_1)$, relying on the following assumption.

Assumption 2.3.3 *Assume that the following hypotheses are satisfied:*

(a) *the observables* $\{\mathbf{Z}_i, i \geq 1\}$ *are i.i.d. according to* $G_*(\cdot|A_0)$*, where* A_0 *denotes the true Pickands dependence function;*

(b) *either* $A_0 \in \mathcal{A}_0$*, or it is associated to an angular density* $h_0 \in C^+((0,1))$ *which satisfies* $\lim_{x\downarrow 0} h_0(x) = \lim_{x\uparrow 1} h_0(x) = +\infty$.

Denote by P_0 the probability measure associated to $G_*(\cdot|A_0)$ and by P_0^∞ the corresponding infinite product measure. Then, we can state the following result.

Theorem 2.3.4 *Let* $\Pi_{\mathcal{A}}$ *be a Bernstein polynomial prior on* $\mathcal{A}$*, induced by a prior* Π *on* $\cup_{k\geq k'}(\{k\} \times \mathcal{B}_k)$*,* $k' \in \mathbb{N}_+$*, which satisfies*

(c) $\Pi\{k\} > 0, \Pi(B) > 0, \quad \forall B \subset \mathcal{B}_k, \forall k \geq k'$,

(d) $\Pi\{k, k+1, k+2, \ldots\} \lesssim e^{-qk}$,

for some positive constant q*. Under Assumption 2.3.3, the posterior distribution* $\Pi_{\mathcal{A}}^{(n)}(\cdot) \equiv \Pi_{\mathcal{A}}(\cdot|\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ *satisfies*

$$\lim_{n\to\infty} \Pi_{\mathcal{A}}^{(n)}(\mathcal{U}_\varepsilon) = 0, \quad P_0^\infty - a.s.$$

for any $\varepsilon > 0$ *and set* $\mathcal{U}_\varepsilon := \{A \in \mathcal{A} : \|A - A_0\|_1 > \varepsilon\}$.

An example of prior distribution satisfying the requirement at point (c) can be found in Corollary 3.4 of [47]. We close the present discussion with a couple of remarks.

Remark 2.3.5 The assumption (c) of Theorem 2.3.4 guarantees that (2.6) is satisfied, whenever A_0 satisfies point (b) of Assumption 2.3.3. In this case, $\Pi_{\mathcal{A}}$ induces a prior $\Pi_{\mathcal{G}}$ on the space $\mathcal{G}$ of densities of the form (2.2) whose Kullback-Leibler support includes $g_*(\cdot|A_0)$. Theorem 6.50 in [30] then guarantees consistency of the Cesàro averages of the predictive densities

$$\hat{g}_{n+1}(\mathbf{z}) = \frac{\int_{\mathcal{G}} g(x) \prod_{i=1}^n g(\mathbf{Z}_i) d\Pi_{\mathcal{G}}(g)}{\int_{\mathcal{G}} \prod_{i=1}^n g(\mathbf{Z}_i) d\Pi_{\mathcal{G}}(g)},$$

also for strong metrics (e.g. d_H or $\|\cdot\|_1$). Yet, it remains an open problem whether the strong posterior consistency of $\Pi_{\mathcal{A}}$ is inherited by $\Pi_{\mathcal{G}}$ (e.g. endowing $\mathcal{G}$ with d_H or $\|\cdot\|_1$).

Remark 2.3.6 Equation (2.3) entails that $\|A - D\|_1 \le 2\|H_A - H_D\|_1$, for every $A, D \in \mathcal{A}$. Consequently, L^1 consistency does not directly extend to the space $\mathcal{H}$. Establishing an analogue of Theorem 2.3.4 with d_∞ in place of the L^1 distance would instead allow to inherit uniform consistency on the space $\mathcal{H}$, via Theorem 2.5.7 in [62] and the equality $A'(t) = -1 + 2H(t)$, $t \in (0, 1)$.

2.4 Piecewise polynomial representation

In this section we consider a piecewise polynomial (PP) representation of the Pickands dependence function. For $(\xi_j)_{j=1}^{l+1}$ a strictly increasing sequence of points on the real line, $q \in \mathbb{N}_+$ and a sequence $P_1, \dots, P_l$ of polynomials of order q, we define a PP function f of order q by the prescription

$$f(x) = \begin{cases} P_1(x), & x \le \xi_1 \\ P_j(x), & \xi_j \le x < \xi_{j+1} \\ P_l(x), & \xi_{l+1} \le x \end{cases} \tag{2.9}$$

and the interval $[\xi_1, \xi_{l+1}]$ is called the basic interval. While retaining the main properties of polynomial modelling, PP modelling using B-splines offers approximations of any smooth function with error bounds comparable to those of polynomials but of smaller complexity - see [16, pp. 26–27]. Examples of inference on the Pickands dependence function using B-splines can be found in [13, 34, 49], to name a few. Differently from these works, we do not rely on a smoothing procedure, but characterize a set of splines which are valid Pickands dependence functions. Such a characterization is then exploited to specify a prior on $\mathcal{A}$, whose posterior consistency is of interest. The next subsection introduces the preliminary concepts needed to fulfil this task.

2.4.1 B-splines

For $\mathbf{t} = (t_j)$ a nondecreasing sequence (which may be finite, infinite or biinfinite), define the jth B-spline of order 1 by $B_{j,\mathbf{t},1} = \mathbb{1}_{[t_j, t_{j+1})}$ and the jth spline of order $q > 1$ via the recurrence relation

$$B_{j,\mathbf{t},q} = \omega_{j,q} B_{j,\mathbf{t},q-1} + (1 - \omega_{j+1,q}) B_{j+1,\mathbf{t},q-1},$$

where

$$\omega_{j,q}(x) := \begin{cases} \frac{x-t_j}{t_{j+q-1}-t_j}, & t_{j+q-1} \neq t_j \\ 0, & \text{otherwise} \end{cases} .$$

The B-splines of order q represents a basis for the space of PP functions of order q with given numbers $\nu_2 - 1, \dots, \nu_l - 1$ of continuous derivatives at the internal breaks $\xi_2, \dots, \xi_l$, with $l \in \mathbb{N}_+$ controlling the number of polynomial pieces in (2.9). For simplicity, we restrict to the case $\nu_2 = \dots = \nu_l = q - 1$. Set $k = q + l - 1$ and let

$$\begin{aligned} t_1 = \dots = t_q = \xi_1; \\ t_{j+q} = \xi_{j+1},\ j = 1, \dots, k-q; \\ t_{k+1} = \dots = t_{k+q} = \xi_{l+1}. \end{aligned}$$

Then, any PP function f of order q with internal breaks $(\xi_j)_{j=2}^l$ which satisfies

$$\mathcal{D}^m f(\xi_j^-) = \mathcal{D}^m f(\xi_j^+), \quad m = 0, \dots, q-2;\ j = 2, \dots, l,$$

admits a unique representation in the form

$$f = \sum_{j=1}^k \alpha_j B_{j,\mathbf{t},q} \tag{2.10}$$

on $[t_q, t_{k+1}]$. This result is a consequence of Curry-Shoenberg theorem - see e.g. [16, Ch. 9]. The representation in (2.10) is known as the B-form of f. A converse result is also true: a linear combination of $(B_{j,\mathbf{t},q})_1^k$ originates a PP function of order q, with $q-2$ continuous derivatives at the internal breaks. We now review some properties of the B-spline basis.

Properties 2.4.1 *Let* $\mathbf{t} = (t_j)_{j=1}^{k+q}$ *be defined as above. Then the B-spline basis functions* $(B_{j,\mathbf{t},q})_{j=1}^k$ *satisfy the following properties:*

(i) $(B_{j,\mathbf{t},q})_{j=1}^k$ *provide a positive and local partition of unity, i.e.* $B_{j,\mathbf{t},q} > 0$ *on* (t_j, t_{j+q}), $B_{j,\mathbf{t},q} = 0$ *outside* $[t_j, t_{j+q}]$ *and* $\sum_{j=1}^{k+q} B_{j,\mathbf{t},q} = 1$ *on* $[t_q, t_{k+1}]$.

(ii) *For* $m = 1, 2$, *it holds that*

$$\mathcal{D}^m \left(\sum_{j=1}^k \alpha_j B_{j,\mathbf{t},q} \right) = \sum_{j=m+1}^k \alpha_j^{(m)} B_{j,\mathbf{t},q-m}$$

where

$$\alpha_j^{(1)} = (q-1)\frac{\alpha_j - \alpha_{j-1}}{t_{j+q-1} - t_j}, \quad \alpha_j^{(2)} = (q-2)\frac{\alpha_j^{(1)} - \alpha_{j-1}^{(1)}}{t_{j+q-2} - t_j}.$$

(iii) *For any* $x \in [t_q, t_{k+1}]$ *it holds that*

$$\int_{t_q}^{x} \left(\sum_{j=1}^{k} \alpha_j B_{j,\mathbf{t},q-1}(w) \right) dw = \sum_{j=1}^{k} \left(\sum_{r=1}^{j} \alpha_j \frac{t_{j+q-1} - t_j}{q-1} \right) B_{j,\mathbf{t},q}(x).$$

(iv) *If* $x \in [t_j, t_{j+1}]$, *for some* $j \in \{q, \dots, k\}$, *it holds that*

$$\bigwedge_{s=j+1-q}^{j} \alpha_s \leq \sum_{j=1}^{k} \alpha_j B_{j,\mathbf{t},q}(x) \leq \bigvee_{s=j+1-q}^{j} \alpha_s.$$

(v) *Let* $f \in C([a,b])$ *and* $t_q = \xi_1 = a$, $t_{k+1} = \xi_{l+1} = b$; *then, setting*

$$\tau_j := \begin{cases} (t_{j+(q-1)/2} + t_{j+(q+1)/2})/2, & q \text{ odd} \\ t_{j+q/2}, & q \text{ even} \end{cases}$$

for $j = 1, \dots, k$, *it holds that*

$$\sup_{x \in [a,b]} \left| \sum_{j=1}^{k} f(\tau_j) B_{j,\mathbf{t},q}(x) - f(x) \right| \leq \lfloor (q+1)/2 \rfloor \omega(f; |\mathbf{t}|),$$

with $|\mathbf{t}| = \max_{1 \leq j \leq k+q-1} \Delta t_j$.

(vi) *Let* $f \in C^{q-1}[a,b]$, *with* t_q, t_{k+1} *and* $(\tau_j)_1^k$ *being as above. Then, for* $m = 0, \dots, q-1$ *there exist constants* $c_{q,m}$ *such that*

$$\sup_{x \in [a,b]} \left| \mathcal{D}^m Q_k(f; x) - \mathcal{D}^m f(x) \right| \leq c_{q,m} M_{\mathbf{t}}^{(2m-q)_+} |\mathbf{t}|^{q-m-1} \omega(\mathcal{D}^{q-1} f; |\mathbf{t}|),$$

where $M_{\mathbf{t}} := \max\{\Delta t_r / \Delta t_s : |r-s| = 1, q \leq r, s \leq k\}$ *and*

$$Q_k(f; x) = \sum_{j=1}^{k} (\lambda_{j,q} f) B_{j,\mathbf{t},q},$$

with linear coefficients $(\lambda_{j,q} f)_{j=1}^k$ *defined via*

$$\lambda_{j,q} f := \sum_{r=1}^{q} \frac{(-\mathcal{D}^{q-r}) \psi_{j,q}(\tau_j)}{(q-1)!} \mathcal{D}^{r-1} f(\tau_j), \quad \psi_{j,q}(\tau_j) := \prod_{r=1}^{q-1} (t_{j+r} - \tau_j).$$

The approximation result at point (vi) is established in [17]. In the following, the sequence of quasi-interpolants $(Q_k(f))$ will play the pivotal role of providing ρ_∞-approximations of the elements of $\mathcal{A}_0 \cap \mathcal{A}^+$, needed for establishing posterior consistency via Theorem 2.2.3.

2.4.2 Bayesian inference

In this subsection, we develop a Bayesian procedure for modelling the extremal dependence, based on the PP representation of the Pickands dependence function, and establish its posterior consistency. As a first step, we construct a suitable support for such a prior, by means of linear combinations of parabolic B-spline basis functions, i.e.

$$A_{\mathbf{t},3} = \sum_{j=1}^{k} \alpha_j B_{j,\mathbf{t},3}. \tag{2.11}$$

The reason for choosing $q = 3$ is twofold. On one hand, it correspond to the minimum degree of smoothness required from an element of $\mathcal{A}$. On the other, splines in the form (2.11) are very well conditioned, making easier the identification of valid sets of coefficients. Henceforward, the knots $(t_j)_{j=1}^{k+3}$ are fixed as follows: $t_1 = t_2 = t_3 = 0$; $t_{j+3} = j/(k-2)$, for $j = 1, \dots, k-3$; $t_{k+1} = t_{k+2} = t_{k+3} = 1$, $k \geq 3$. This corresponds to choosing $k-1$ equally spaced breaks ξ_j on the unit interval. Since $A_{\mathbf{t},3}(1^-) = \alpha_k$, we continuously extend $A_{\mathbf{t},3}$ from $[0,1)$ to $[0,1]$ by imposing $A_{\mathbf{t},3}(1) = \alpha_k$. We now provide conditions on the coefficients $(\alpha_j)_{j=1}^k$ which are necessary and sufficient for $A_{\mathbf{t},3}$ to satisfy Properties 1.2.1.

Proposition 2.4.2 *The function $A_{\mathbf{t},3}$ is a valid Pickands dependence function if and only if $(\alpha_j)_{j=1}^k$ satisfies the following restrictions:*

(S1) $\alpha_1 = \alpha_k = 1 \geq \alpha_j$, $j = 2, \dots, k-1$;

(S2) $\alpha_2 = 1 + (p_0 - 1/2)/(k-2)$ *and* $\alpha_{k-1} = 1 + (p_1 - 1/2)/(k-2)$, *for some* $0 \leq p_0, p_1 \leq 1/2$;

(S3) $\alpha_3 - \alpha_2 \geq 2(\alpha_2 - \alpha_1)$, $2(\alpha_k - \alpha_{k-1}) \geq (\alpha_{k-1} - \alpha_{k-2})$ *and*

$$\alpha_j - \alpha_{j-1} \geq \alpha_{j-1} - \alpha_{j-2}, \quad j = 4, \dots, k-1.$$

When (S1)-(S3) are satisfied, $A_{\mathbf{t},3}$ is a valid Pickands dependence function and the function $H_{\mathbf{t},2}$ defined via

$$H_{\mathbf{t},2}(x) = \begin{cases} \frac{1+A'_{\mathbf{t},3}(x)}{2} = \sum_{j=2}^{k} \left[\frac{1}{2} + \frac{\alpha_j - \alpha_{j-1}}{t_{j+2}-t-j}\right] B_{j,\mathbf{t},2}(x), & x \in [0,1) \\ 1, & x = 1 \end{cases},$$

defines a valid angular distribution. In general, we can state the following.

Proposition 2.4.3 *Let $H_{\mathbf{t},2}$ be a function on $[0,1]$ defined via*

$$H_{\mathbf{t},2}(x) = \begin{cases} \sum_{j=2}^{k} \zeta_j B_{j,\mathbf{t},2}(x), & x \in [0,1) \\ 1, & x = 1 \end{cases}. \tag{2.12}$$

Then, $H_{\mathbf{t},2}$ is a valid angular distribution if and only if $(\zeta_j)_{j=2}^{k}$ satisfies the following restrictions:

(S4) $p_0 = \zeta_2 \leq \zeta_3 \leq \ldots \leq \zeta_{k-1} \leq \ \zeta_k = 1 - p_1$, *with* $0 \leq p_0, p_1 \leq 1/2$*;*

(S5) $\frac{1}{2}\zeta_2 + \sum_{l=3}^{k-1} \zeta_l + \frac{1}{2}\zeta_k = (k-2)/2.$

In particular, if $(\alpha_j)_{j=1}^{k}$ complies with (S1)-(S3)*, then setting*

$$\zeta_j = \frac{1}{2} + \frac{\alpha_j - \alpha_{j-1}}{t_{j+2} - t_j}, \quad j = 2, \ldots, k, \tag{2.13}$$

in (2.12)*, $H_{2,\mathbf{t}}$ is a valid angular distribution.*

In fact, we can establish a converse result. Starting with a valid angular distribution $H_{\mathbf{t},2}$ as in (2.12), it is possible to exploit point (iii) of Properties 2.4.1 and obtain a valid Pickands dependence function via

$$\begin{aligned} A_{\mathbf{t},3}(t) &= 1 + 2\int_0^t H_{\mathbf{t},2}(x)dx - t \\ &= 1 + 2\int_0^t \left(\sum_{j=2}^{k} (\zeta_j - 1/2) B_{j,\mathbf{t},2}(x)\right) dx \\ &= \sum_{j=1}^{k} \left(\sum_{r=1}^{j} (\zeta_r - 1/2)(t_{r+2} - t_r) + 1\right) B_{j,\mathbf{t},3}(t), \end{aligned}$$

where we impose $\zeta_1 = 1/2$. In other words, we can assert the following.

Proposition 2.4.4 *Let $(\zeta_j)_{j=2}^k$ satsisfy conditions* (S4)-(S5)*. Then, setting*

$$\alpha_1 = 1, \quad \alpha_j = \sum_{r=2}^{j} (\zeta_r - 1/2)(t_{r+2} - t_r) + 1, \; j = 2, \dots, k,$$

the coefficients $(\alpha_j)_{j=1}^k$ satisfy (S1)-(S3)*.*

For $k \geq 3$, denote by $\mathfrak{A}_k = \{(\alpha_j)_1^k : \text{ (S1)-(S3) are satisfied}\}$. From the arguments developed so far, we learn that a Pickands dependence function in the form (2.11) is associated to a vector of linear coefficients in $\mathfrak{A}_k$ and is paired with a valid angular distribution in the form (2.12). Since the latter is a piecewise linear function which is continuous on $(0, 1)$, we can conclude that the valid Pickands dependence functions in the form (2.11) constitute a subset of $\mathcal{A}$. In symbols, denote $\mathcal{A}_k = \{\sum_{j=1}^k \alpha_j B_{j,\mathbf{t},3} : (\alpha_j)_{j=1}^k \in \mathfrak{A}_k\}$, then $\cup_{k\geq 3}\mathcal{A}_k \subset \mathcal{A}$. Moreover, $\cup_{k\geq 3}\mathcal{A}_k$ is dense in $\mathcal{A}$ with respect to d_∞. Technically speaking, this is due to the fact that, for $q = 3$, the approximation at point (v) of Properties 2.4.1 becomes Shoenberg's variation diminishing approximation, which is shape preserving - see [16, Ch. 11]. As a consequence, any prior Π assigning positive mass to all the positive integers k above some $k' \geq 3$ and subsets of $\mathfrak{A}_k$ induces a prior $\Pi_{\mathcal{A}}$ with full support on $(\mathcal{A}, d_\infty)$. We refer to $\Pi_{\mathcal{A}}$ as parabolic spline prior. We now establish the L^1-posterior consistency of $\Pi_{\mathcal{A}}$.

Theorem 2.4.5 *Let $\Pi_{\mathcal{A}}$ be a parabolic spline prior on $\mathcal{A}$, induced by a prior Π on $\cup_{k\geq k'}(\{k\} \times \mathfrak{A}_k)$, $k' \geq 3$, which satisfies*

(c′) $\Pi\{k\} > 0, \Pi(B) > 0, \quad \forall B \subset \mathfrak{A}_k, \forall k \geq k'$,

(d′) $\Pi\{k, k+1, k+2, \dots\} \lesssim e^{-pk}$,

for some positive constant p. Under Assumption 2.3.3, the posterior distribution $\Pi_{\mathcal{A}}^{(n)}(\cdot) \equiv \Pi_{\mathcal{A}}(\cdot|\mathbf{Z}_1, \dots, \mathbf{Z}_n)$ satisfies

$$\lim_{n\to\infty} \Pi_{\mathcal{A}}^{(n)}(\mathcal{U}_\varepsilon) = 0, \quad P_0^\infty - a.s.$$

for any $\varepsilon > 0$ and set $\mathcal{U}_\varepsilon := \{A \in \mathcal{A} : \|A - A_0\|_1 > \epsilon\}$.

We now provide an example of prior distribution Π satisfying the assumptions of Theorem 2.4.5. We exploit the following result.

Proposition 2.4.6 *A vector of linear coefficients* $(\alpha_j)_{j=1}^k$ *satisfies* (S1)-(S3) *if and only if it complies with the following conditions*

(S6) $\alpha_1 = \alpha_k = 1$

(S7) $\alpha_2 \in [1 - \frac{1}{2(k-2)}, 1]$ *and* $\alpha_{k-1} \in [l_{k-1}, u_{k-1}]$*, with*

$$l_{k-1} = \max\left\{1 + (\alpha_2 - 1)[2(k-3)+1], 1 - \frac{1}{2(k-2)}\right\},$$
$$u_{k-1} = 1 + \frac{\alpha_2 - 1}{2(k-3)+1};$$

(S8) $\alpha_3 \in [l_3, u_3]$*, with*

$$l_3 = \max\left\{\frac{3\alpha_2 - \alpha_1}{2}, \alpha_{k-1}[2(k-3)-1] - 2(k-4)\right\},$$
$$u_3 = \min\left\{\frac{\alpha_{k-1} + \alpha_2(k-4)}{k-3}, 2(1-\alpha_{k-1}) + \alpha_2\right\},$$

and, for $j = 4, \dots, k-1$, $\alpha_j \in [l_j, u_j]$ *with*

$$l_j = \max\left\{2\alpha_{j-1} - \alpha_{j-2}, \alpha_{k-1}[2(k-j)-1] - 2(k-j-1)\right\},$$
$$u_j = \min\left\{\frac{1}{k-j}[\alpha_{k-1} + \alpha_{j-1}(k-j-1)], 2(1-\alpha_{k-1}) + \alpha_{j-1}\right\}.$$

A consequence of the above result is the following. Assume that a prior distribution on k has been specified, in such a way that assumption (d′) is satisfied. Then, the prescription

$$\alpha_i|k \sim \delta_1, \quad i = 1, k; \quad \alpha_2|k \sim \mathrm{Unif}(1 - 1/2(k-2), 1);$$
$$\alpha_{k-1}|k, \alpha_2 \sim \mathrm{Unif}(l_{k-1}, u_{k-1});$$
$$\alpha_j|k, \alpha_2, \dots, \alpha_{j-1}, \alpha_{k-1} \sim \mathrm{Unif}(l_j, u_j), \quad j = 3, \dots, k-2;$$

defines a prior which assigns positive mass to any subset of $\mathfrak{A}_k$.

Remark 2.4.7 The prior proposed above might be seen as an "objective" prior on the coefficient space $\cup_{k \geq k'}(\{k\} \times \mathfrak{A}_k)$ (and then on $\mathcal{A}$). It is mostly meant as an example of prior matching the assumptions of Theorem 2.4.5. Different prior distributions might be designed to account for prior beliefs. For example, conditionally on k, one might want to specify a prior on $\mathfrak{A}_k$ in such a way that $\Pi\{(\alpha_j)_{j=1}^k \in \mathfrak{A}_k : d(\sum_{j=1}^l \alpha_j B_{j,\mathbf{t},k}, \mathcal{A}^\Pi) < \epsilon | k\}$ is large, with $\mathcal{A}^\Pi$ a class of Pickands dependence functions on which the prior has to concentrate, d a semimetric and ϵ a positive constant. The specification of such priors goes beyond the scope of the present analysis, which is mostly focused on the asymptotic characterization.

Concluding, notice that the observations in Remarks 2.3.5-2.3.6 can be directly extended to the present setting. Yet, L^1-consistency on $\mathcal{H}$ is not of particular interest herein: the angular distributions in (2.12) are useful to characterize parabolic splines Pickands dependence functions, but are not of practical use for inference, due to their picewise linear structure. In order to satisfactorily carry out simultaneous inference about A and H, the present spline construction must be extended to higher polynomial orders (i.e. $q > 3$). Yet, in such cases, finding necessary and sufficient conditions on the coefficients $(\alpha_j)_{j=1}^k$ to define valid Pickands dependence functions might be more demanding.

2.5 Discussion

The results presented in this chapter provide a significant contribution to the development of the Bayesian nonparametric modelling of the extremal dependence. In particular, it makes a first step towards the understanding of the asymptotic behaviour of such statistical methods.

Theorems 2.3.4 and 2.4.5 establish posterior consistency of Bernstein polynomial and parabolic spline priors of the Pickands dependence function. A first refinement of such results consists in the derivation of posterior concentration rates. A simulation study aimed at inspecting the smaller-sample performances attainable with the two representations appears a natural complement of the present analysis. Another substantial extension consists in adapting the results of Theorems 2.3.4 and 2.4.5 to the case in which the observables are assumed i.i.d. according to $G(\cdot|A_0, \Theta_0)$, with $\Theta_0 = \{(\xi_{0,j}, \mu_{0,j}, \sigma_{0,j}) : j = 1, 2\}$ and $G(\cdot|A_0, \Theta_0)$ as in Section 1.2. In this case, the true marginal distributions are in the GEV class and the prior must be extended to the marginal parameters. Moreover, the likelihood is derived from the densities

$$g(\mathbf{y}|A,\Theta) = g_*(z(y;\Theta_1), z(y_2;\Theta_2)|A) \prod_{j=1,2} \frac{\partial}{\partial y_j} z(y_j;\Theta_j) \mathbb{1}_{\left\{1+\xi_j \frac{y_j-\mu_j}{\sigma_j}>0\right\}}$$

where $\Theta = (\Theta_1, \Theta_2)$, $\Theta_j = (\xi_j, \mu_j, \sigma_j)$ and

$$z(y_j;\Theta_j) = \left[1 + \xi_j \frac{y_j - \mu_j}{\sigma_j}\right]^{1/\xi_j}, \quad \forall y_j : 1 + \xi_j \frac{y_j - \mu_j}{\sigma_j} > 0.$$

for $j = 1, 2$. Observe that two different Θ and Θ' satisying $\|\Theta - \Theta'\| < \varepsilon$, for some norm $\|\cdot\|$ and small ε, might still correspond to distributions $G(\cdot|A,\Theta)$ and $G(\cdot|A',\Theta')$ with different supports. Consequently, the K-L divergence between the latter might be infinite. This could represent a potential problem for the application of extended Schwartz theorem, which requires the true distribution $G(\cdot|A_0,\Theta_0)$ to possess the K-L property relative to the prior.

The above discussion relies on the assumption that observables are drawn from a bivariate EVD. In applications, data are only approximately drawn from the latter and the use of a likelihood derived from the density of a bivariate EVD induces a problem of model misspecification. Nevertheless, we conjecture that when the observables are i.i.d. vectors of componentwise maxima $\mathbf{M}_{1,k_n}, \dots, \mathbf{M}_{n,k_n}$, the posterior distributions on $\mathcal{A}$ obtained via a misspecified likelihood can still be proved consistent in view of a contiguity argument - see [10, 11] for the analysis of analogous problems involving the spectral density of a time series. Deriving similar results for bivariate exceedances by the adoption of a bivariate generalized Pareto model (e.g. [63, 64]) would represent a further important goal. With the latter approach, a larger portion of the information contained in the data is exploited, typically resulting in an effiency gain. Our ultimate goal would be to carry the asymptotic analysis described above into the general multivariate setting.

2.6 Proofs

2.6.1 Proof of Theorem 2.2.3

This subsection is organized as follows. First, some notation is introduced. Then, in Lemmas 2.6.1 and 2.6.2, upper bounds for the K-L divergence between two densities in the form (2.5) are worked out. The result in Theorem 2.2.3 is finally established, as a direct consequence of those.

For every $A \in \mathcal{A}$, we will denote by $\tilde{g}_*(\cdot|A)$ the density

$$\tilde{g}_*(r,t|A) = \left[\frac{\phi_A(t) + r\psi_A(t)}{r^3t^2(1-t)^2}\right] \exp\left\{-\frac{A(t)}{rt(1-t)}\right\}, \quad (r,t) \in (0,\infty) \times (0,1),$$

where, for $t \in (0,1)$,

$$\phi_A(t) := [A(t) - tA'(t)][A(t) + (1-t)A'(t)], \tag{2.14}$$

$$\psi_A(t) := t^2(1-t)^2A''(t). \tag{2.15}$$

In particular, observe that conditionally on A

$$(R,W) := (Z_1 + Z_2, Z_1/(Z_1+Z_2)) \stackrel{d}{\sim} \tilde{g}_*(\cdot|A),$$

with $(Z_1, Z_2) \stackrel{d}{\sim} g_*(\cdot|A)$. Then, the first step for proving the main result consists in establishing the following one.

Lemma 2.6.1 *Let $A \in \mathcal{A}^+$, so that $d := \inf_{x\in(0,1)} h(x) > 0$. Then, there exists a positive constant m such that, for every $\varepsilon \in (0, d \wedge m/9)$ and $D \in \mathcal{A}$ satisfying $\rho_\infty(A,D) < \varepsilon$, it holds that*

$$K\left(g_*(\cdot|A), g_*(\cdot|D)\right) \leq 2\varepsilon + \log\left\{1 + \varepsilon\left(\frac{9}{m} + \frac{1}{d}\right)\right\}.$$

Proof. Observe that, by (2.3) and the mean constraint on H, for every $t \in (0,1)$

$$A(t) - tA'(t) > 0, \quad A(t) + (1-t)A'(t) > 0;$$

hence $\phi_A(t) > 0$. By condition (iv) in Definition 2.2.2, $\phi_A(0) = 2p_0 > 0$ and $\phi_A(1) = 2p_1 > 0$. Therefore, by continuity, there exists a positive constant m such that $\phi_A \geq m$. Furthermore, it can be shown that

$$m - \frac{9}{2}\varepsilon \leq \phi_D. \tag{2.16}$$

Indeed, since $\rho_\infty(A,D) < \varepsilon$, triangular inequality and few algebraic manipulations yield

$$\sup_{x\in[0,1]} |\phi_A(x) - \phi_D(x)| \leq \frac{9}{2}\varepsilon - \frac{5}{4}\varepsilon^2. \tag{2.17}$$

The lower bound in (2.16) now follows from the fact that $\phi_A \geq m$ and from (2.17).

Next, observe that by invariance of the K-L divergence with respect to continuous transformations

$$
\begin{aligned}
K\left(g_*(\cdot|A), g_*(\cdot|D)\right) &= K\left(\tilde{g}_*(\cdot|A), \tilde{g}_*(\cdot|D)\right) \\
&= \int_0^\infty \int_0^1 [D(t) - A(t)] \frac{\tilde{g}_*(r,t|A)}{rt(1-t)} dtdr \\
&\quad + \int_0^\infty \int_0^1 \log\left\{\frac{\phi_A(t) + r\psi_A(t)}{\phi_D(t) + r\psi_D(t)}\right\} \tilde{g}_*(r,t|A) dtdr \\
&\equiv T_1 + T_2.
\end{aligned} \tag{2.18}
$$

Since

$$
\int_0^\infty \int_0^1 \tilde{g}_*(r,t|A)/rt(1-t) dtdr = \mathbb{E}[Z_1^{-1} + Z_2^{-1}|A] = 2,
$$

it holds that $|T_1| \leq 2\varepsilon$. Furthermore, Jensen inequality, the fact that $d > 0$ and (2.16)-(2.17) yield

$$
\begin{aligned}
T_2 \leq \log\Bigg\{1 + \int_0^\infty \int_0^1 \Bigg[&\frac{\phi_D(t)}{\phi_D(t) + r\psi_D(t)} \frac{\phi_A(t) - \phi_D(t)}{\phi_D(t)} \\
&+ \frac{r\psi_D(t)}{\phi_D(t) + r\psi_D(t)} \frac{A''(t) - D''(t)}{D''(t)}\Bigg] \tilde{g}_*(r,t|A) dtdr\Bigg\} \\
\leq \log\left\{1 + \varepsilon(9m^{-1} + d^{-1})\right\}.&
\end{aligned}
$$

The result now follows. ∎

Denote by H_0 and h_0 the angular probability measure and density associated to A_0. Also, denote $p_{0,0} := H_0(\{0\})$, $p_{0,1} := H_0(\{1\})$ and $d_0 := \inf_{x\in(0,1)} h_0(x)$. Lemma 2.6.1 can now be used to establish the following result.

Lemma 2.6.2 *Let $A_0 \in \mathcal{A}_0$. Then, there exist $k'' \geq k'$ and a sequence $(\widetilde{A}_{0,k})_{k=k''}^\infty$ with $\widetilde{A}_{0,k} \in \mathcal{A}_k$, such that for any $\varepsilon > 0$, as $k \to \infty$, it holds that*

$$
K(g_*(\cdot|A_0), g_*(\cdot|\widetilde{A}_{0,k})) \leq 5\varepsilon + \log(1+\varepsilon). \tag{2.19}
$$

Proof. When $A_0 \in \mathcal{A}_0 \cap \mathcal{A}^+$, by hypothesis there exists a sequence $(A_{0,k})_{k=k''}^\infty$ satisfying (2.7) for $A = A_0$. Then, the result directly follows from Lemma 2.6.1

by choosing $\widetilde{A}_{0,k} = A_{0,k}$, since we can make the left-hand side of (2.19) arbitrarily small as $k \to \infty$.

Next, consider the case in which h_0 complies with (iii.b) in Definition 2.2.1 and group the possible configurations of $p_{0,0}$ and $p_{0,1}$ in the following way:

(I) $0 \le p_{0,1} < p_{0,0} < 1/2$;

(II) $p_{0,0} = p_{0,1} = 0$;

(III) $0 < p_{0,0} \le p_{0,1} < 1/2$ or $0 \le p_{0,0} < p_{0,1} < 1/2$.

Let $\delta \in (0, \delta^*)$ be any constant satisfying

$$\delta < \begin{cases} \max\{p_{0,0} - p_{0,1}, \gamma_{\varepsilon,1}\}, & \text{if (I) holds} \\ \max\{1/2, \gamma_{\varepsilon,1}\}, & \text{if (II) holds} \\ \max\{p_{0,1} - p_{0,0}, \gamma_{\varepsilon,0}\}, & \text{if (III) holds} \end{cases} \tag{2.20}$$

with

$$\gamma_{\varepsilon,i} := \frac{\varepsilon(1-2p_{0,i})}{1+2(1-2p_{0,i})(1+\varepsilon)}, \quad i = 0, 1.$$

Then, define

$$\begin{aligned} \widetilde{h}_0(w) &:= \frac{h_0(t) \vee \delta}{\int_0^1 [h_0(w) \vee \delta] dw}(1 - \widetilde{p}_0 - \widetilde{p}_1), \quad w \in (0,1), \\ \widetilde{H}_0(t) &:= \widetilde{p}_0 + \int_0^t \widetilde{h}_0(w) dw + \widetilde{p}_1 \mathbb{1}(t=1), \quad t \in [0,1], \\ \widetilde{A}_0(t) &:= 1 + 2\int_0^t \widetilde{H}_0(w) dw - t, \quad t \in [0,1], \end{aligned}$$

where

$$\begin{aligned} \widetilde{p}_0 &:= \begin{cases} \left[\frac{1}{2} - (1-\widetilde{p}_1)(1-c)\right]\frac{1}{c}, & \text{if (I)-(II) hold} \\ p_{0,0}(1-2\delta) + \delta, & \text{if (III) holds} \end{cases}, \\ \widetilde{p}_1 &:= \begin{cases} (1-2\delta)p_{0,1} + \delta, & \text{if (I)-(II) hold} \\ \left[\frac{1}{2} - (1-\widetilde{p}_0)c\right]\frac{1}{1-c}, & \text{if (III) holds} \end{cases}, \\ c &:= \frac{\int_0^1 w[h_0(w) \vee \delta] dw}{\int_0^1 [h_0(w) \vee \delta] dw}. \end{aligned}$$

From (2.20) it follows that $(\widetilde{p}_0, \widetilde{p}_1) \in (0, 1/2)^2$. Furthermore, it can be easily shown that $\widetilde{h}_0$ satisfies the analogue of (2.2) with $\widetilde{p}_0$ and $\widetilde{p}_1$ in place of p_0 and p_1, respectively. Therefore, $\widetilde{H}_0 \in \mathcal{H}$ and $\widetilde{A}_0 \in \mathcal{A}_0 \cap \mathcal{A}^+$.

Observe that the K-L divergence between $g_*(\cdot|A_0)$ and $g_*(\cdot|\widetilde{A}_0)$ satisfies

$$
\begin{aligned}
K(g_*(\cdot|A_0), g_*(\cdot|\widetilde{A}_0)) = & \int_0^\infty \int_0^1 \log \left\{ \frac{\phi_{A_0}(t) + r\psi_{A_0}(t)}{\phi_{\widetilde{A}_0}(t) + r\psi_{\widetilde{A}_0}(t)} \right\} \tilde{g}_*(r, t|A_0) dt dr \\
& + \int_0^\infty \int_0^1 [\widetilde{A}_0(t) - A_0(t)] \frac{\tilde{g}_*(r, t|A_0)}{rt(1-t)} dt dr \\
\equiv & \, T_1 + T_2.
\end{aligned}
$$

We now show that δ can be chosen in such a way that

$$
\frac{A_0''}{\widetilde{A}_0''} \leq 1 + \varepsilon, \quad \frac{\phi_{A_0}}{\phi_{\widetilde{A}_0}} \leq 1 + \varepsilon, \tag{2.21}
$$

whence it follows that $T_1 \leq \log(1 + \varepsilon)$. Few algebraic manipulations yield

$$
\begin{aligned}
\frac{A_0''(t)}{\widetilde{A}_0''(t)} &= \frac{h_0(t)}{h_0(t) \vee \delta} \frac{\int_0^1 [h_0(w) \vee \delta] dw}{1 - \widetilde{p}_0 - \widetilde{p}_1} \\
&\leq \frac{1/2 - p_{0,i} + \delta/2}{(1/2 - p_{0,i})(1 - 2\delta)}, \quad i = \begin{cases} 1, & \text{if (I)-(II) hold} \\ 0, & \text{if (III) holds} \end{cases}.
\end{aligned}
$$

Then, the first half of (2.21) follows from (2.20). Next, define S_1 and S_2 via

$$
\frac{\phi_{A_0}(t)}{\phi_{\widetilde{A}_0}(t)} = \left(\frac{A_0(t) - tA_0'(t)}{\widetilde{A}_0(t) - t\widetilde{A}_0'(t)} \right) \left(\frac{A_0(t) + (1-t)A_0'(t)}{\widetilde{A}_0(t) + (1-t)\widetilde{A}_0'(t)} \right) =: S_1(t) S_2(t).
$$

Letting ε' be a positive constant which satisfies

$$
\varepsilon' \leq \min \left(\tfrac{2\gamma_{\varepsilon,i}}{1 - \gamma_{\varepsilon,i}}, \sqrt{1 + \varepsilon} - 1 \right), \quad i = \begin{cases} 1, & \text{if (I)-(II) hold} \\ 0, & \text{if (III) holds} \end{cases}
$$

and choosing δ such that

$$
\frac{\varepsilon'}{2(1 + \varepsilon')} \leq \delta \leq \frac{\varepsilon'}{2}, \tag{2.22}
$$

we now prove that $S_j \leq (1 + \varepsilon')$, $j = 1, 2$. Then, the second half of (2.21) folows as a direct consequence. From (2.3) and (2.4), it follows that $S_1(0) = 1$, $S_1(1) = p_{0,1}/\widetilde{p}_1$ and

$$
S_1(t) = \frac{1 - 2\int_0^t \int_w^t h_0(s) ds dw}{1 - 2\int_0^t \int_w^t \widetilde{h}_0(s) ds dw}, \quad t \in (0, 1).
$$

On one hand, using the bounds in (2.22), it can be shown that $p_{0,1}/\widetilde{p}_1 < 1+\varepsilon'$. On the other hand, the inequality $S_1(t) \leq 1+\varepsilon'$ for $t \in (0,1)$ can be established by showing that

$$\int_0^t \int_w^t [(1+\varepsilon')\widetilde{h}_0(s) - h_0(s)] ds dw \leq \frac{\varepsilon'}{2}, \tag{2.23}$$

for every $t \in (0,1)$. Letting $\epsilon_{\delta,1}$ and $\epsilon_{\delta,2}$ be as in condition (iii.b) of Definition 2.2.1, observe that for every $s \in (\epsilon_{\delta,1}, 1-\epsilon_{\delta,2})$

$$\begin{aligned}(1+\varepsilon')\widetilde{h}_0(s) - h_0(s) &= h_0(s)\left[\frac{1-\widetilde{p}_0-\widetilde{p}_1}{\int_0^1 [h_0(w)\vee\delta]}(1+\varepsilon') - 1\right] \\ &\leq h_0(s)[(1-2\delta)(1+\varepsilon')-1]\end{aligned} \tag{2.24}$$

and that, by (2.22), the term on the right end side is negative. This fact, condition (iii.b) and few algebraic manipulations now yield (2.23). Next, observe that

$$S_2(t) = \frac{1/2 - \int_0^{1-t}\int_t^{1-v} h_0(s) ds dv}{1/2 - \int_0^{1-t}\int_t^{1-v} \widetilde{h}_0(s) ds dv}, \quad t \in (0,1),$$

$S_2(1) = p_{0,0}/\widetilde{p}_1$ and $S_2(1) = 0$. Then, the inequality $S_2 \leq (1+\varepsilon')$ follows as above.

Concerning T_2, by (2.3) we have that

$$\begin{aligned}T_2 &= \int_0^\infty \int_0^1 \left[2\int_0^t (\widetilde{H}_0(w) - H_0(w)) dw\right] \frac{\tilde{g}_*(r,t|A_0)}{rt(1-t)} dt dr \\ &\leq 2\int_0^\infty \int_0^1 \left[\int_0^t \int_0^w (\widetilde{h}_0(s) - h_0(s)) ds dw\right] \frac{\tilde{g}_*(r,t|A_0)}{rt(1-t)} dt dr \\ &\quad + 4\max\{\widetilde{p}_0 - p_{0,0}, 0\}.\end{aligned}$$

Furthermore, by (2.22)–(2.24) , the first summand on the right hand side can be bounded above by $2\varepsilon'$. The bounds in (2.20) and (2.22), together with few algebraic manipulations, yield the inequality: $\max\{\widetilde{p}_0 - p_{0,0}, 0\} \leq (1/2 - p_{0,0})\varepsilon$. As a consequence, $T_2 \leq 4\varepsilon$ and

$$K(g_*(\cdot|A_0), g_*(\cdot|\widetilde{A}_0)) \leq \log(1+\varepsilon) + 4\varepsilon. \tag{2.25}$$

Concluding, since $\widetilde{A}_0 \in \mathcal{A}_0 \cap \mathcal{A}^+$, there exists a sequence $(\widetilde{A}_{0,k})_{k=k''}^\infty$ satisfying (2.7) for $A = \widetilde{A}_0$. Moreover, it holds that

$$K(g_*(\cdot|A_0), g_*(\cdot|\widetilde{A}_{0,k})) = K(g_*(\cdot|A_0), g_*(\cdot|\widetilde{A}_0)) + K(g_*(\cdot|\widetilde{A}_0), g_*(\cdot|\widetilde{A}_{0,k})).$$

The statement now follows from (2.25) and Lemmma 2.6.1.

At this point, we only need to discuss the case in which condition (iii.a) of Definition 2.2.1 is satisfied, i.e. $d_0 > 0$, and $p_{0,0} = p_{0,1} = 0$. This setting can be dealt with letting $\delta \in (0, d_0\varepsilon(1+\varepsilon)^{-1} \wedge 1/4)$, $\widetilde{p}_0 = \widetilde{p}_1 = \delta/2$, $\widetilde{h}_0 = h_0 - \delta$ and defnining $\widetilde{H}_0$ and $\widetilde{A}_0$ as above. With this setup, the final result can be obtained by following the steps outlined for the previous case. Yet, derivations are much easier and are therefore omitted. The proof is now complete. ■

We can now complete the proof of Theorem 2.2.3. By Lemma 2.6.2, there exists a sequence $(\widetilde{A}_{0,k})_{k=k''}^{\infty} \in \times_{k=k''}^{\infty} \mathcal{A}_k$ such that the K-L divergence between $g_*(\cdot|A_0)$ and $g_*(\cdot|\widetilde{A}_{0,k})$ can be made arbitrarily small. In particular, $\widetilde{A}_{0,k}$ can be constructed in such a way that $\widetilde{A}_{0,k} \in \mathcal{A}^+, \forall k \geq k'''$, for some fixed $k''' \geq k''$. Then, in light of Lemma 2.6.1, the term

$$K\left(g_*(\cdot|A_0), g_*(\cdot|D)\right) = K\left(g_*(\cdot|A_0), g_*(\cdot|\widetilde{A}_{0,k})\right) + K\left(g_*(\cdot|\widetilde{A}_{0,k}), g_*(\cdot|D)\right)$$

can be made aribitrarily small by choosing D in a small ρ_∞-ball centered at $\widetilde{A}_{0,k}$, with k large. The statement now follows from the assumption that $\Pi_{\mathcal{A}}$ assigns positive probability to every ρ_∞-ball centered at an $A \in \mathcal{A}_k$, for $k \geq k'$. ■

2.6.2 Proof of Corollary 2.2.5

For the sake of brevity, we only sketch the main lines of the proof. As before, denote $p_{0,0} = H_0(\{0\})$ and $p_{0,1} = H_0(\{1\})$. Observe that for a pair of small constants $\boldsymbol{\epsilon} := (\epsilon_1, \epsilon_2) > \mathbf{0}$, there exist:

- positive bounded functions $f_{\epsilon_1}, g_{\epsilon_2}$ satisfying $f_{\epsilon_1}(t) \leq h_0(t)$, $\forall t \in (0, \epsilon_1)$, and $f_{\epsilon_2}(t) \leq h_0(t)$, $\forall t \in (1-\epsilon_2, 1)$,
- constants $\widetilde{p}_{0,\boldsymbol{\epsilon}} \in (p_{0,0}, 1/2]$, $\widetilde{p}_{1,\boldsymbol{\epsilon}} \in (p_{0,1}, 1/2]$,

such that the function

$$\widetilde{h}_{0,\boldsymbol{\epsilon}}(t) = \begin{cases} f_{\epsilon_1}(t), & t \in [0, \epsilon_1) \\ h_0(t), & t \in (\epsilon_1, 1-\epsilon_2) \\ g_{\epsilon_2}(t), & t \in [1-\epsilon_2, 1] \end{cases}$$

is continuous, is lower bounded by $d_0 = \inf_{t\in(0,1)} h_0(t) > 0$ and satisfies the analogue of (2.2) with p_0 and p_1 replaced by $\widetilde{p}_{2,\epsilon}$ and $\widetilde{p}_{1,\epsilon}$. In particular, for any small $\varepsilon > 0$, the quantities just introduced can be chosen in such a way that $d_\infty(H_0, \widetilde{H}_{0,\epsilon}) < \varepsilon$ and $d_\infty(A_0, \widetilde{A}_{0,\epsilon}) < \varepsilon$, where $\widetilde{H}_{0,\epsilon}$ and $\widetilde{A}_{0,\epsilon}$ are defined via

$$\begin{aligned}
\widetilde{H}_{0,\epsilon}(t) &= \widetilde{p}_{0,\epsilon} + \int_0^t \widetilde{h}_{0,\epsilon}(w)dw + \widetilde{p}_{1,\epsilon}\mathbb{1}(t=1),\\
\widetilde{A}_{0,\epsilon}(t) &= 1 + 2\int_0^t \widetilde{H}_{0,\epsilon}(w)dw - t.
\end{aligned}$$

Note that $\widetilde{A}_{0,\epsilon} \in \mathcal{A}_0 \cap \mathcal{A}^+$.

Next, observe that, by the invariance of the K-L divergence under continuous transformation and Jensen's inequality,

$$\begin{aligned}
K(g_*(\cdot|A_0), g_*(\cdot|\widetilde{A}_{0,\epsilon})) \leq \log&\left\{\int_0^\infty\int_0^1 \frac{\phi_{A_0}(t) + r\psi_{A_0}(t)}{\phi_{\widetilde{A}_{0,\epsilon}}(t) + r\psi_{\widetilde{A}_{0,\epsilon}}(t)}\tilde{g}_*(r,t|A_0)dtdr\right\}\\
&+ \int_0^\infty\int_0^1 [\widetilde{A}_{0,\epsilon}(t) - A_0(t)]\frac{\tilde{g}_*(r,t|A_0)}{rt(1-t)}dtdr\\
&\equiv T_1 + T_2,
\end{aligned}$$

where ϕ_{A_0} and $\phi_{\widetilde{A}_{0,\epsilon}}$ are defined as in (2.14), ψ_{A_0} and $\psi_{\widetilde{A}_{0,\epsilon}}$ as in (2.15). As for T_1, using similar arguments to those developed in the previous subsection, we obtain the inequality: $\phi_A/\phi_{\widetilde{A}_{0,\epsilon}} \leq 1$. Moreover, due to the integrability of h_0, there exists a constant $\beta \in (0,1)$ such that $\lim_{t\to 0} t^\beta h_0(t)$ and $\lim_{t\to 1}(1-t)^\beta h_0(t)$. As a consequence, for ϵ_1 and ϵ_2 sufficiently small

$$\begin{aligned}
\int_0^\infty\int_0^{\epsilon_1} \frac{h_0(t)}{\widetilde{h}_{0,\epsilon}(t)}\tilde{g}_*(r,t|A_0)dtdr &\leq \frac{1}{d_0}\int_0^\infty\int_0^{\epsilon_1}\frac{\tilde{g}_*(r,t|A_0)}{t^\beta}dtdr \leq \frac{\varepsilon}{2},\\
\int_0^\infty\int_{1-\epsilon_2}^1 \frac{h_0(t)}{\widetilde{h}_{0,\epsilon}(t)}\tilde{g}_*(r,t|A_0)dtdr &\leq \frac{1}{d_0}\int_0^\infty\int_{1-\epsilon_2}^1\frac{\tilde{g}_*(r,t|A_0)}{(1-t)^\beta}dtdr \leq \frac{\varepsilon}{2}.
\end{aligned}$$

These facts and few algebraic manipulations yield the inequality: $T_1 \leq \log(1+\varepsilon)$. Moreover, it can be easily seen that $T_2 \leq 2\varepsilon$.

In light of the above, we conclude that any arbitrarily small K-L neighborhood of $g_*(\cdot|A_0)$ contains a density of the form of $g_*(\cdot|\widetilde{A}_{0,\epsilon})$, with $\widetilde{A}_{0,\epsilon} \in \mathcal{A}_0 \cap \mathcal{A}^+$. The statement now follows from Theorem 2.2.3.

2.6.3 Proof of Theorem 2.3.4

The proof follows the lines of the extended Schwartz theorem – see e.g. [30, Theorem 6.17]. First, we establish the property in (2.6) when $\Pi_{\mathcal{A}}$ is the Bernstein polynomial prior. Then, we prove the existence of tests with exponentially bounded error probabilities on a suitable sequence of sieves.

In order to establish property (2.6) for Π_A the Berstein polynomial prior and $A_0 \in \mathcal{A}_0$, we prove that the assumptions of Theorem 2.2.3 are satisfied. Let $\mathcal{A}' = \cup_{k \geq k'} \mathcal{A}_k$, with $\mathcal{A}_k$ defined as at the beginning of Subsection 2.3.2. For any $A \in \mathcal{A}_0 \cap \mathcal{A}^+$, the existence of a sequence satisfying (2.7) is guaranteed by the fact that $B_k(A) \in \mathcal{A}_k$, for all $k \geq k'$, and by point (viii) of Properties 2.3.2. The fact that $\Pi_{\mathcal{A}}$ assigns positive mass to any ρ_∞-ball centered at an element $A_k \in \mathcal{A}_k$, for every $k \geq k'$, follows from assumption (c) and points (i)-(ii) of Properties 2.3.1. In fact, by the latter, for every $D_k \in \mathcal{A}_k$ it holds that for $m = 0, 1, 2$

$$d_\infty(\mathcal{D}^m A_k, \mathcal{D}^m D_k) \leq (m+2) \frac{(k+1)!}{(k-m)!} \max_{0 \leq j \leq k} |\beta_{A_k,j} - \beta_{D_k,j}|,$$

where $\beta_{A_k,j}$ and $\beta_{D_k,j}$ represent the linear coefficients of A_k and D_k, respectively. Notice that, by Corollary 2.2.5, property (2.6) is satisfied also in the case of A_0 with unbounded angular density.

As for testing, we can establish the following result.

Lemma 2.6.3 *For every $\varepsilon \in (0,2)$ and set $\mathcal{U}_\varepsilon = \{A \in \mathcal{A} : \|A - A_0\|_1 > \varepsilon\}$, there exist sequences of integers (ν_n) and tests (ϕ_n) such that, letting $\mathcal{A}_{1,n} := \cup_{j \leq \nu_n} (\mathcal{U}_\varepsilon \cap \mathcal{A}_j)$, it holds that*

$$P_0^n \phi_n \leq e^{-nc_1}, \quad \sup_{A \in \mathcal{A}_{n,1}} P_A^n (1 - \phi_n) \leq e^{-nc_2}, \tag{2.26}$$

where c_1, c_2 are positive constants.

Proof Observe that each set $\mathcal{U}_\varepsilon \cap \mathcal{A}_j$ can be covered by m_j balls

$$\mathbf{B}_{l,j} := \{A \in \mathcal{A}_j : \|A - A_l^{(j)}\|_1 < \varepsilon_1\}, \quad 1 \leq l \leq m_j,$$

with $\{A_1^{(j)}, \ldots, A_{m_j}^{(j)}\} \subset \mathcal{U}_\epsilon \cap \mathcal{A}_j$, $\varepsilon_1 = 2^{-1} e^{-1} \varepsilon / (e^{-1} + 2e^{-1/2})$ and

$$m_j \leq \left(\frac{6}{\varepsilon_1} \right)^j, \quad 1 \leq j \leq \nu_n. \tag{2.27}$$

Indeed, from (2.3), the first half of (2.8) and points (ii)-(iii) of Properties 2.3.1, it follows that for any $A, D \in \mathcal{A}_j$ with angular distributions H_A, H_D

$$\begin{aligned}\|A - D\|_1 &\leq 2\|H_A - H_D\|_1 \\ &\leq 2|\eta_{A,0} - \eta_{D,0}| + 2\sum_{i=0}^{j-2} |(\eta_{A,i+1} - \eta_{A,i}) - (\eta_{D,i+1} - \eta_{D,i})|,\end{aligned}$$

and the vectors $(\eta_{\bullet,0}, \eta_{\bullet,1} - \eta_{\bullet,0}, \ldots, \eta_{\bullet,j-1} - \eta_{\bullet,j-2})$ lie in $\{\mathbf{x} \in \mathbb{R}^j : \|\mathbf{x}\|_1 \leq 1\}$. Then, (2.27) follows by Proposition C.2 in [30]. Next, observe that by the Lipschitz continuity of order e of the map $x \mapsto -\log x$ on $(e^{-1}, e^{-1/2})$ we have that for any $A, D \in \mathcal{A}$ and $t \in (0, 1)$

$$\begin{aligned}|A(t) - D(t)| &\leq e|e^{-A(t)} - e^{-D(t)}| \\ &\leq e\|g_*(\cdot|A) - g_*(\cdot|D)\|_1 \\ &\leq 2ed_H(g_*(\cdot|A), g_*(\cdot|D)).\end{aligned}$$

Consequently, $\|A - D\|_1 \leq 2ed_H(g_*(\cdot|A), g_*(\cdot|D))$. These fact and some algebraic manipulations yield

$$\begin{aligned}&d_H\left(g_*(\cdot|A_0), \alpha g_*(\cdot|A) + (1-\alpha)g_*(\cdot|D)\right) \\ &\geq \frac{1}{2}\left(e^{-1}\|A_0 - D\|_1 - e^{-1/2}\|A - D\|_1\right) \\ &\geq \frac{1}{4e}\varepsilon,\end{aligned}$$

for any $\alpha \in [0, 1]$, $A, D \in \mathbf{B}_{l,j}$ and $1 \leq l \leq m_j$, $1 \leq j \leq \nu_n$. That is: defining

$$\mathbf{Q}_{l,j} = \{g_*(\cdot|A) : A \in \mathbf{B}_{l,j}\}$$

it holds that

$$d_H(g_*(\cdot|A_0), \mathrm{conv}\mathbf{Q}_{l,j}) \geq \frac{1}{4e}\varepsilon, \tag{2.28}$$

where $\mathrm{conv}\mathbf{Q}_{l,j}$ denotes the convex hall of $\mathbf{Q}_{l,j}$. As a consequence, by Corollary D.7 in [30] there exist tests $\psi_n^{(l,j)}$ to test $g_*(\cdot|A_0)$ versus the alternatives $\mathbf{Q}_{l,j}$ whose error probabilities satisfy

$$P_0^n\psi_n^{(l,j)} + \sup_{A \in \mathbf{B}_{l,j}} P_A^n\left(1 - \psi_n^{(l,j)}\right) \leq e^{-nc_1}, \tag{2.29}$$

with $c_1 = \varepsilon^2/32e^2$. Now, let $\phi_n := \max\{\psi_n^{(l,j)} : 1 \leq l \leq j,\, 1 \leq j \leq \nu_n\}$. Then, the first half of (2.26) directly follows from (2.29). Moreover, from (2.29) and (2.27) it follows that

$$\sup_{A \in \mathcal{A}_{n,1}} P_A^n(1-\phi_n) \leq e^{-c_1 n} \sum_{j=1}^{\nu_n} m_j \ \leq e^{\nu_n \log(6/\varepsilon_1) - c_1 n}.$$

Therefore, letting ξ and c_2 be positive constants which satisfy

$$\xi \leq c_1 \log^{-1}(6/\varepsilon_1), \quad c_2 = c_1 - \xi \log(6/\varepsilon_1), \tag{2.30}$$

and choosing $\nu_n = \lfloor n\xi \rfloor$, the second half of (2.26) follows. ∎

We can now proceed with the proof of the main result. Property (2.6) has been established. Henceforth, letting ϕ_n be as in Lemma 2.6.3, for every $c > 0$ and n large, the posterior distribution satisfies

$$\Pi_{\mathcal{A}}^{(n)}(\mathcal{U}_\varepsilon) \leq \phi_n + e^{cn} \int_{\mathcal{U}_\varepsilon} (1-\phi_n) R_n(A) \Pi_{\mathcal{A}}(dA), \quad P_0^\infty - a.s.$$

with $R_n(A) := \prod_{i=1}^n [g_*(\mathbf{Z}_i|A)/g_*(\mathbf{Z}_i|A_0)]$. On one hand, $\phi_n = o(1)$, P_0^∞-almost surely. On the other, using the partition $\mathcal{A}' \cap \mathcal{U}_\varepsilon = \mathcal{A}_{n,1} \cup \mathcal{A}_{n,2}$, with $\mathcal{A}_{n,2} = \cup_{j>\nu_n}(\mathcal{U}_\varepsilon \cap \mathcal{A}_j)$, the expectation of the second summand on the right hand side satisfies

$$\begin{aligned} P_0^n\left[e^{cn}\int_{\mathcal{U}_\varepsilon}(1-\phi_n)R_n(A)\Pi_{\mathcal{A}}(dA)\right] &\leq e^{cn}\int_{\mathcal{A}_{n,1}\cup\mathcal{A}_{n,2}} P_A^n(1-\phi_n)\Pi_{\mathcal{A}}(dA) \\ &\leq e^{cn}\int_{\mathcal{A}_{n,1}} P_A^n(1-\phi_n)\Pi_{\mathcal{A}}(dA) \\ &\quad + e^{cn}\Pi\{\nu_n, \nu_n+1, \ldots\}. \end{aligned}$$

Fixing $c \leq \min(c_2, q\xi)$, with ξ, c_2 as in (2.30), the final result follows from the second half of Lemma 2.6.3 and assumption (b), plus Markov inequality and Borel-Cantelli lemma. ∎

2.6.4 Proof of Proposition 2.4.2

First, observe that, since $A_{\mathbf{t},3}(0) = \alpha_1$ and $A_{\mathbf{t},3}(1) = \alpha_k$, $A_{\mathbf{t},3}$ satisfies the endopoint conditions $A_{\mathbf{t},3}(0) = A_{\mathbf{t},3}(1) = 1$ if and only if $\alpha_0 = \alpha_k = 1$. Next, observe that by points (i)-(ii) of Properties 2.4.1, for every $j = 3, \dots, k-1$

$$A''_{\mathbf{t},3}(t) = \frac{2}{t_{j+1} - t_j} \left[\frac{\alpha_j - \alpha_{j-1}}{t_{j+2} - t_j} - \frac{\alpha_{j-1} - \alpha_{j-2}}{t_{j+1} - t_{j-1}} \right], \quad x \in [t_j, t_{j+1}).$$

Therefore, a necessary and sufficient condition for the convexity of $A_{\mathbf{t},3}$ is the non-negativity of the terms in the above display, which reduces to condition (S3). By points (i)-(ii) of Properties 2.4.1

$$A'_{\mathbf{t},3}(0) = 2\frac{\alpha_2 - \alpha_1}{t_4 - t_2} B_{2,\mathbf{t},2}(0) = 2(\alpha_2 - \alpha_1)(k-2)$$

and

$$A'_{\mathbf{t},3}(1^-) = 2\frac{\alpha_k - \alpha_{k-1}}{t_{k+2} - t_k} B_{k,\mathbf{t},2}(1^-) = 2(\alpha_k - \alpha_{k-1})(k-2).$$

Thus, we can continuously extend $A'_{\mathbf{t},3}$ from $[0,1)$ to $[0,1]$ by imposing $A'_{\mathbf{t},3}(1) = 2(\alpha_k - \alpha_{k-1})(k-2)$. Under conditions (S1) and (S3), $A_{\mathbf{t},3}$ is convex, equals 1 on $\{0,1\}$ and is less than equal to 1 on $(0,1)$. In fact, by point (iv) of Properties 2.4.1,

$$A_{\mathbf{t},3}(x) \leq \max\{\alpha_{j-2}, \alpha_{j-1}, \alpha_j\} \leq 1, \quad x \in [t_j, t_{j+1}].$$

As a consequence, in order $A_{\mathbf{t},3}$ to satisfy the lower bound condition $\max\{t, 1-t\} \leq A_{\mathbf{t},3}(t)$, it must be that

$$-1 \leq A'_{\mathbf{t},3}(0) \leq 0, \quad 0 \leq A'_{\mathbf{t},3}(0) \leq 1. \tag{2.31}$$

In terms of coefficients α_2, α_{k-1}, the above inequalities read

$$1 - \frac{1}{2(k-2)} \leq \alpha_2, \alpha_{k-1} \leq 1,$$

which is equivalent to (S2). So far, it has been shown that (S1)-(S3) are sufficient conditions for $A_{\mathbf{t},3}$ to satisfy (A1)-(A3). To complete the proof of the converse implication, assume $A_{\mathbf{t},3}$ is a valid Pickands dependence function. Then, being convex and admitting bounds $\max\{t, 1-t\} \leq A_{\mathbf{t},3}(t) \leq 1$, $A_{\mathbf{t},3}$ must satisfy (2.31), which is (S2), together with (A3) and $\alpha_1 = \alpha_k = 1$. The fact that $\alpha_j \leq 1$, $j = 3, \dots, k-1$ is a necessary condition now follows from (S3) and the fact that by point (iv) of Properties 2.4.1

$$\min\{\alpha_{j-2}, \alpha_{j-1}, \alpha_j\} \leq A_{\mathbf{t},3}(t) \leq 1, \quad t \in [t_j, t_{j+1}], \quad j = 3, \dots, k.$$

The proof is now complete. ∎

2.6.5 Proof of Proposition 2.4.3

Observe that

$$H_{\mathbf{t},2}(0) = \zeta_2, \quad H_{\mathbf{t},2}(1^-) = \zeta_k;$$

hence, for $H_{\mathbf{t},2}$ to be a valid angular distribution, it must be that $\zeta_2 = p_0$, $\zeta_k = 1 - p_1$, $0 \leq p_0, p_1 \leq 1/2$. Furthermore, for any $x_1 \leq x_2 \in [0,1)$ we have that by point (ii) in Properties 2.4.1

$$\begin{aligned} H_{\mathbf{t},2}(x_2) - H_{\mathbf{t},2}(x_1) &= \int_{x_1}^{x_2} H'_{\mathbf{t},2}(x)dx \\ &= \int_{x_1}^{x_2} \left\{ \sum_{j=3}^{k} \left[\frac{\zeta_j - \zeta_{j-1}}{t_{j+1} - t_j} \right] B_{j,\mathbf{t},1}(x) \right\} dx \\ &= \sum_{j=3}^{k} \left[\frac{\zeta_j - \zeta_{j-1}}{t_{j+1} - t_j} \right] \int_{x_1}^{x_2} \mathbb{1}_{[t_j, t_{j+1})}(x)dx. \end{aligned}$$

Hence, $\zeta_j \geq \zeta_{j-1}$, $j = 3, \ldots, k$ is a sufficient condition for $H_{\mathbf{t},2}$ to be monotone nondecreasing. Yet, observing that in the case $t_j \leq x_1 \leq x_2 < t_{j+1}$ the above equality reduces to

$$H_{\mathbf{t},2}(x_2) - H_{\mathbf{t},2}(x_1) = \left[\frac{\zeta_j - \zeta_{j-1}}{t_{j+1} - t_j} \right] (x_2 - x_1),$$

it becomes clear that $\zeta_j \geq \zeta_{j-1}$ is also a necessary condition.

Next, observe that the mean constraint in (1.5) for $H_{\mathbf{t},2}$ are equivalent to

$$\begin{aligned} 1/2 &= p_1 + \int_0^1 x H'_{\mathbf{t},2}(x)dx \\ &= p_1 + \frac{1}{2} \sum_{j=3}^{k} (\zeta_j - \zeta_{j-1})(t_{j+1} + t_j) \end{aligned}$$

and

$$\begin{aligned} 1/2 &= p_0 + \int_0^1 (1-x) H'_{\mathbf{t},2}(x)dx \\ &= p_0 + \frac{1}{2} \sum_{j=3}^{k} (\zeta_j - \zeta_{j-1}) \left[1 - \frac{t_{j+1} + t_j}{2} \right], \end{aligned}$$

which both reduce to (S4) after few algebraic manipulations.

So far, (S4)-(S5) have been shown to be necessary and sufficient conditions for $H_{\mathbf{t},2}$ to be a valid angular distribution. We are thus left to prove that the presctiption in (2.13) yields coefficients $(\zeta_j)_{j=2}^k$ complying with (S4)-(S5). Start observing that by (S2)

$$\zeta_2 = \frac{1}{2} + (k-2)(\alpha_2 - 1) = p_0, \quad \zeta_k = \frac{1}{2} + (k-2)(1-\alpha_{k-1}) = 1 - p_1,$$

and that $\zeta_j \geq \zeta_{j-1}$ is equivalent to

$$\frac{\alpha_j - \alpha_{j-1}}{t_{j+2} - t_j} - \frac{\alpha_{j-1} - \alpha_{j-2}}{t_{j+1} - t_{j-1}} \geq 0,$$

which is satisfied in view of (S3). Hence $(\zeta_j)_{j=2}^k$ satisfies (S4). Next, observe that

$$\begin{aligned}\frac{1}{2}\zeta_2 + \sum_{l=2}^{k-1}\zeta_l + \frac{1}{2}\zeta_k &= \frac{1}{2}\left[\frac{1}{2} + (k-2)(\alpha_2 - 1)\right] + \sum_{l=3}^{k-1}\left[\frac{1}{2} + \frac{\alpha_j - \alpha_{j-1}}{2/(k-2)}\right] \\ &\quad + \frac{1}{2}\left[\frac{1}{2} + (k-2)(1-\alpha_{k-1})\right] \\ &= (n-2)/2.\end{aligned}$$

That is, $(\zeta_j)_{j=2}^k$ also satisfies (S5) and the proof is now complete. ∎

2.6.6 Proof of Proposition 2.4.4

Observe that, by (S5), the following holds true

$$\begin{aligned}\alpha_k &= \sum_{r=2}^{k}(\zeta_r - 1/2)(t_{r+2} - t_r) + 1 \\ &= \frac{1}{k-2}\zeta_2 + \frac{2}{k-2}\sum_{r=2}^{k-1}\zeta_r + \frac{1}{k-2}\zeta_k - \frac{1}{k-2} - \frac{k-3}{k-2} + 1 \\ &= \frac{2}{k-2}\left(\frac{1}{2}\zeta_2 + \sum_{r=2}^{k-1}\zeta_r + \frac{1}{2}\zeta_k\right) = 1.\end{aligned}$$

Next, we show that $\alpha_j \leq 1$, $j = 2, \dots, k-1$. For $j = 2$, the result is immediate, since $\alpha_2 = 1 + (\zeta_2 - 1/2)/(k-2)$ and $\zeta_2 \leq 1/2$. For $j = 3, \dots, k-1$, the condition $\alpha_j \leq 1$ is equivalent to

$$\frac{1}{2}\zeta_2 + \sum_{r=3}^{j}\zeta_r \leq \frac{j-2+1/2}{2}. \tag{2.32}$$

Assume that the above inequality does not hold true. Then, on one hand

$$\sum_{r=3}^{j} \zeta_r > \frac{j-2}{2} + \frac{1}{2}\left(\frac{1}{2} - \zeta_2\right) \Longrightarrow \frac{\sum_{r=3}^{j} \zeta_r}{j-2} > 1/2;$$

thus, $\max\{\zeta_3, \dots, \zeta_j\} > 1/2$. On the other,

$$\begin{aligned}
\frac{k-2}{2} &= \frac{1}{2}\zeta_k + \frac{1}{2}\zeta_2 + \sum_{j=3}^{k-1} \zeta_j \\
&> \frac{1}{4} + \frac{j-2+1/2}{2} + \sum_{r=j+1}^{k-1} \zeta_r \\
&\geq \frac{j-1}{2} + \max\{\zeta_3, \dots, \zeta_j\}(k-j-1) \\
&> \frac{k-2}{2},
\end{aligned}$$

yielding a contradiction. Thus condition (2.32) is satisfied. Property (S1) has now been established.

The first half of property (S2) has already been established, in light of the equality $\zeta_2 = p_0$. As for the second half, observe that

$$1 - \alpha_{k-1} = \alpha_k - \alpha_{k-1} = \frac{\zeta_k - 1/2}{k-2} = \frac{1/2 - p_1}{k-2}.$$

Finally, to establish (S3), notice that for $j = 3, \dots, k$

$$\frac{\alpha_j - \alpha_{j-1}}{t_{j+2} - t_j} \geq \frac{\alpha_{j-1} - \alpha_{j-2}}{t_{j+1} - t_{j-1}} \Longleftrightarrow \zeta_j \geq \zeta_{j-1},$$

and that the inequalities on the right hand side are satisfied by assumption. ∎

2.6.7 Proof of Theorem 2.4.5

The proof is constructed in the same way as that of Theorem 2.3.4, with few modifications. Herein, we only sketch the latter. In order to establish property (2.6) with $\Pi_{\mathcal{A}}$ the parabolic spline prior, we verify once more that $\Pi_{\mathcal{A}}$ meets the requirements of Theorem 2.2.3. Let $\mathcal{A}' = \cup_{k \geq k'} \mathcal{A}_k$, with $\mathcal{A}_k$ defined as in Subsection 2.4.2. By point (vi) of Properties 2.4.1, for any $A \in \mathcal{A}_0 \cap \mathcal{A}^+$, the sequence of quasi-interpolants

$$Q_k(A; t) = \sum_{j=0}^{k} \left[A\left(\frac{t_{j+1} + t_{j+2}}{2}\right) - \frac{(\Delta t_{j+1})^2}{8} A''\left(\frac{t_{j+1} + t_{j+2}}{2}\right) \right] B_{j,\mathbf{t},3}(t),$$

with $k \geq k''$ and fixed k'' large enough, satisfies the condition in (2.7). By points (i)-(ii) of Properties 2.4.1, for every pair $A_k, D_k \in \mathcal{A}_k$ and $m = 0, 1, 2$ it holds that

$$d_\infty(\mathcal{D}^m A_k, \mathcal{D}^m D_k) \leq (k-m)^{m+1} 2(m+1) \frac{3!}{(3-m-1)!} \max_{1 \leq j \leq k} |\alpha_{A_k,j} - \alpha_{D_k,j}|.$$

Consequently, the fact that $\Pi_{\mathcal{A}}$ assigns positive mass to any ρ_∞-ball centered at an element $A_k \in \mathcal{A}_k$, for any $k \geq k'$, follows from assumption (c′).

As for testing, an analogue of Lemma 2.6.3 can be established. This follows from the fact that, by (2.3) and points (i)-(ii) of Properties 2.4.1, every pair $A_k, D_k \in \mathcal{A}_k$, with angular distributions H_{A_k} and H_{D_k}, satisfy

$$\begin{aligned} \|A_k - D_k\|_1 &\leq 2\|H_{A_k} - H_{D_k}\|_1 \\ &\leq 2|\zeta_{A_k,2} - \zeta_{D_k,2}| + 2\sum_{j=3}^{k} |\Delta\zeta_{A_k,j-1} - \Delta\zeta_{D_k,j-1}|, \end{aligned}$$

the vectors $(\zeta_{\bullet,2}, \Delta\zeta_{\bullet,2}, \ldots, \Delta\zeta_{\bullet,k-1})$ be lying in $\{\mathbf{x} \in \mathbb{R}^{k-1} : \|\mathbf{x}\|_1 \leq 1\}$.

The final result now follows as in Subsection 2.6.3. ▮

2.6.8 Proof of Proposition 2.4.6

The result we are after is a consequence of the following.

Lemma 2.6.4 *The vector of linear coefficients* $(\zeta_j)_{j=2}^k$ *satisfies* (S4)-(S5) *if and only if it satisfies the following conditions*

(S9) $\zeta_2 \in [0, 1/2]$ *and* $\zeta_k \in [a_k, b_k]$, *with*

$$a_k = \max\left\{\frac{k-2-\zeta_2}{2(k-3)+1}, 0\right\}, \quad b_k = k-2-[2(k-3)+1]\zeta_2;$$

(S10) $\zeta_3 \in [a_3, b_3]$, *with*

$$\begin{aligned} a_3 &= \max\left\{\zeta_2, \frac{k-2}{2} - \frac{2(k-3)-1}{2}\zeta_k - \frac{1}{2}\zeta_2\right\}, \\ b_3 &= \min\left\{\zeta_k, \frac{k-2-\zeta_k-\zeta_2}{2(k-3)}\right\}, \end{aligned}$$

and, for $j = 4, \ldots, k-1$, $\zeta_j \in [a_j, b_j]$ *with*

$$a_j = \max\left\{\zeta_{j-1}, \frac{k-2}{2} - \frac{2(k-j)-1}{2}\zeta_k - \frac{1}{2}\zeta_2 - \sum_{l=3}^{j-1}\zeta_l\right\},$$
$$b_j = \min\left\{\zeta_k, \frac{k-2-\zeta_k\frac{1}{2}-\zeta_2\frac{1}{2}-\sum_{l=3}^{j-1}\zeta_l}{2(k-j)}\right\}.$$

Several of the arguments needed to prove Lemma 2.6.4 can be found in [47, pp. 3319-3320]. Therefore, an explicit proof is omitted.

We now proceed with the proof of Propositon 2.4.6. Assume (S1)-(S3) hold true. Then, by Proposition 2.4.3 it is possible to obtain a valid spectral distribution of the form in (2.12), by means of the prescription in (2.13). That is, the coefficients $(\zeta_j)_{j=2}^k$ in (2.13) satisfy (S4)-(S5). By Lemma 2.6.4, this entails that the coefficients $(\zeta_j)_{j=2}^k$ satisfy (S9)-(S10). By reexpressing (S9)-(S10) in terms of α_j's, conditions (S6)-(S8) follow. A converse reasoning, via Proposition 2.4.4, yields the "only if" part of the statement. ∎

Chapter 3

Extremes of aggregated data: modelling and inference

Risk analysis in the area of insurance and finance as well as the design of telecommunication networks are concerned with studying the probability that multiple extreme events take place simultaneously. As already discussed in Section 1.2, extreme value theory provides tools for modelling such a probability. Yet, when aggregated data are considered, few results are available, especially in high dimensions. As real-world processes are always more frequently based on complex systems, which generate extremely large amounts of data, keeping track of the entire information produced is often unfeasible and data aggregation becomes unavoidable. Other motivations for data aggregation are linked to privacy law enforcements and confidentiality agreements. Herein, we consider aggregated data in the form of maxima computed over a random number N of observations. In Section 3.2, we derive the asymptotic distribution of linearly normalized maxima of multivariate aggregated data, under appropriate conditions on the random number of observations. In Section 3.3, we analyze the extremal dependence structure associated to the latter and show that, when N is heavy-tailed, aggregation inflates the dependence among the usual multivariate maxima, studied in the classical theory of extremes. We connect the extremal dependence structures arising pre- and post-aggregation through random scaling and Pickands dependence functions. In Section 3.4, we exploit an inversion method to construct a class of semiparametric estimators for the extremal dependence of the unobservable (non aggragated) data, starting from estimators of the extremal dependence of the aggregated data. We establish the asymptotic properties of the estimators in such class and further explore their finite-sample performances by a simulation study. All the proofs are deferred to Section 3.6. Some supplementary material is provided in Section 3.7.

3.1 Introduction

Let $\mathbf{X} = (X_1, \ldots, X_d)$ denote a d-dimensional random vector with distribution $F_{\mathbf{X}}$ and margins F_{X_j}, $j = 1, \ldots, d$. Let $(\mathbf{X}_i)_1^\infty$ be i.i.d. copies of $\mathbf{X}$. Let N be a discrete random variable taking values in $\mathbb{N}_+$. Assume hereafter that N, with distribution F_N, is independent of the $\mathbf{X}_i$'s. In some applications the interest is in analyzing aggregated data such as the total or maximum amounts obtained on a random number of observations. The study of

$$\mathbf{S}_N = \left(\sum_{i=1}^N X_{i,1}, \ldots, \sum_{i=1}^N X_{i,d} \right) \tag{3.1}$$

and

$$\mathbf{M}_N = \left(\max_{1 \leq i \leq N} X_{i,1}, \ldots, \max_{1 \leq i \leq N} X_{i,d} \right), \tag{3.2}$$

may be of particular interest in insurance, finance and risk management, as well as in big-data problems such as the analysis of Internet traffic data. Since the number of measurements of Internet traffic is huge, the data processing is feasible only after suitable aggregation. For dimension $d = 1$ the tail behaviour of S_N has been extensively studied in the literature (e.g. [22]). In the multivariate case, there is no result concerning $\mathbf{S}_N$ and only few findings are available on the joint upper tail behaviour of $\mathbf{M}_N$ (e.g. [40, 25]).

The remainder of this chapter focuses on the extremal properties of aggregated data which are realizations of $\mathbf{M}_N$ in (3.2). Since N is assumed independent of the $\mathbf{X}_i$'s, knowledge of the extremal properties of $F_{\mathbf{X}}$ and F_N is sufficient to derive the main extremal properties of $F_{\mathbf{M}_N}$. Under the assumption that $\mathbb{E}N < \infty$, it is already known that the extremal properties of $F_{\mathbf{M}_N}$ are the same as those of $F_{\mathbf{X}}$ - see [40]. However, in the literature no results are available concerning the joint extremal behavior of $(\mathbf{M}_N, N)$. The case of $\mathbb{E}N = \infty$ is subtler and there is no heuristic argument which can point to how different the extremal properties of

$F_{\mathbf{M}_N}$ are with respect to those of $F_{\mathbf{X}}$, so that asymptotic theory in this case has to be established ex novo. Although $\mathbf{M}_N$ and N are dependent, it is possible to have (under some restriction) asymptotic independence, i.e., the extremal properties of $\mathbf{M}_N$ and N can be studied separately. In this context, it is of interest to study the joint extremal behavior of $(\mathbf{M}_N, N)$ and pinpoint the conditions that induce asymptotic independence. The inverse problem is also of interest from both a theoretical and practical point of view. Specifically, since in applications $\mathbf{M}_N$ and N are typically observable but $(\mathbf{X}_i)$ is not, it is of interest to recover the extremal properties of $F_{\mathbf{X}}$ out of those of $F_{\mathbf{M}_N}$. Both the joint extremal behaviour of $(\mathbf{M}_N, N)$ and the inverse problem just mentioned above shall be investigated herein by relying on multivariate extreme value theory.

3.2 Maximum-domain of attraction

The members of the GEV family of distributions are: the α-Fréchet (heavy-tailed distribution), Gumbel (light-tailed distribution) and Weibull (short-tailed distribution), in symbols, $\Phi_\alpha(x) = \exp(-x^{-\alpha})$, with $x > 0$ and $\alpha > 0$, $\Lambda(x) = \exp(-e^{-x})$ with $x \in \mathbb{R}$ and $\Psi_\alpha(x) = \exp(-(-x)^{-\alpha})$ with $x < 0$. In the following, we assume that $F_{\mathbf{X}} \in \mathcal{D}(G)$ and $F_N \in \mathcal{D}(H)$, where either $H \equiv \Phi_\alpha$ or $H \equiv \Lambda$ since N is discrete-valued (e.g. [61]). Under such assumptions, we establish a new limit result concerning the joint maximum-domain of attraction for a vector of componentwise maxima computed on a randomly sized block of iid random vectors, $\mathbf{M}_N$, and the random block size, N.

Theorem 3.2.1 *Assume that $F_{\mathbf{X}} \in \mathcal{D}(G)$ and $F_N \in \mathcal{D}(H)$ with $H \equiv \Phi_\alpha$ or $H \equiv \Lambda$. Then, there exist norming constants $\mathbf{c}_n > \mathbf{0}, c_n > 0$, $\mathbf{d}_n \in \mathbb{R}^d, d_n \in \mathbb{R}$ such that*

$$\lim_{n\to\infty} \mathbb{P}^n \left(\frac{\mathbf{M}_N - \mathbf{d}_n}{\mathbf{c}_n} \leq \mathbf{x}, \frac{N - d_n}{c_n} \leq y \right) = Q(\mathbf{x}, y),$$

for all continuity points $(\mathbf{x}, y)$ of Q, a $(d+1)$-variate max-stable distribution. Precisely, when

1. *$F_N \in \mathcal{D}(\Phi_\alpha)$, then*

$$-\ln Q(\mathbf{x}, y) = \begin{cases} y^{-\alpha} e^{-y\sigma(\mathbf{x},\alpha)} + \sigma^\alpha(\mathbf{x},\alpha)\,\gamma\{1-\alpha, y\,\sigma(\mathbf{x},\alpha)\}, & \alpha \in (0,1] \\ -\ln G(\mathbf{x}) + y^{-\alpha}, & \alpha > 1 \end{cases} \quad (3.3)$$

for $\mathbf{x} \in \mathbb{R}^d$ *and* $y > 0$*, where* $\sigma(\mathbf{x}, \alpha) = \{-\ln G(\mathbf{x})\}/\Gamma^{1/\alpha}(1-\alpha)$ *and* Γ*,* γ *denote the Euler Gamma and Lower Incomplete Gamma functions, with the convention* $\Gamma(0) := 1$*;*

2. $F_N \in \mathcal{D}(\Lambda)$*, then*

$$-\ln Q(\mathbf{x}, y) = -\ln G(\mathbf{x}) + e^{-y}, \quad \mathbf{x} \in \mathbb{R}^d,\ y \in \mathbb{R}. \tag{3.4}$$

In the reminder of this section, we first highlight some features of the limit distribution Q. Then, we provide some intuition behind them. The margins of Q are

$$\begin{cases} G_\alpha(\mathbf{x}) := \exp\{-(-\ln G(\boldsymbol{x}))^\alpha\}, & \alpha \in (0,1) \\ G(\mathbf{x}), & \alpha \geq 1 \end{cases} \tag{3.5}$$

and $\Phi_\alpha(y)$, when $F_N \in \mathcal{D}(\Phi_\alpha)$, $\alpha > 0$. They are equal to $G(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$, and $\Lambda(y)$, $y \in \mathbb{R}$, when $F_N \in \mathcal{D}(\Lambda)$. The distribution G_α, $\alpha \in (0,1)$, is a max-stable distribution with margins $G_{\alpha,j} = \exp\{-(-\ln G_j)^\alpha\}$, that are members of the GEV class, and extreme-value copula given by

$$C_{G_\alpha}(\mathbf{u}) = \exp\big(-L^\alpha\big((-\ln u_1)^{1/\alpha}, \dots, (-\ln u_d)^{1/\alpha}\big)\big), \quad \mathbf{u} \in (0,1]^d, \tag{3.6}$$

where L is the stable-tail dependence function of the max-stable distribution G. As for the overall dependence structure characterizing Q, it can be synthesized through the extremal coefficient as follows.

Corollary 3.2.2 *When the expression of* Q *is given in in* (3.3)*, the extremal coefficient is*

$$\theta(Q) = \begin{cases} \exp\left(-\frac{\theta(G)}{\Gamma^{1/\alpha}(1-\alpha)}\right) + \frac{(\theta(G))^\alpha}{\Gamma(1-\alpha)}\,\gamma\left(1-\alpha, \frac{\theta(G)}{\Gamma^{1/\alpha}(1-\alpha)}\right), & \alpha \in (0,1) \\ \exp(-\theta(G)) + \theta(G)(\mathrm{li}(\exp(-\theta(G))) + 1), & \alpha = 1 \\ \theta(G) + 1, & \alpha > 1 \end{cases}$$

where li *denotes the Logarithmic Integral function. When the expression of* Q *is given in* (3.4)*,* $\theta(Q) = \theta(G) + 1$*.*

Remark 3.2.3 Observe that when the expression of Q is given in in (3.3), i.e. when $F_N \in \mathcal{D}(\Phi_\alpha)$ and $\alpha \in (0,1)$, the extremal coefficient $\theta(Q)$ is increasing in

α. That is, the more F_N is heavy-tailed (i.e. the smaller is α), the larger is the amount of dependence associated to the random vector $(\mathbf{M}_N, N)$ at the extremes. The extremal coefficient of $\theta(G_\alpha)$ for the first d-components is given in the next section.

Let $(N_i)_{i=1}^{\infty}$ be a sequence of i.i.d. copies of N and define

$$S_n = N_1 + \cdots + N_n, \quad M_n = \max(N_1, \ldots, N_n).$$

An intuitive probabilistic interpretation of the results in Theorem 3.2.1 derives from the analysis of the joint limit behavior of (S_n, M_n), appropriately normalized (a.n.). Start observing that, for $\mathbf{x} \in \mathbb{R}^d$ and $y > 0$,

$$\mathbb{P}^n \left(\frac{\mathbf{M}_N - \mathbf{d}_n}{\mathbf{c}_n} \leq \mathbf{x}, \frac{N - d_n}{c_n} \leq y \right) = \mathbb{P} \left(\frac{\mathbf{M}_{S_n} - \mathbf{d}_n}{\mathbf{c}_n} \leq \mathbf{x}, \frac{M_n - d_n}{c_n} \leq y \right).$$

When $F_N \in \mathcal{D}(\Lambda)$ (light-tailed) or $F_N \in \mathcal{D}(\Phi_\alpha)$ (heavy-tailed), with $\alpha > 1$, then $\mu = \mathbb{E}(N) < \infty$. Since $n^{-1} S_n \xrightarrow{\text{as}} \mu$, then the asymptotic distributions of $\mathbf{M}_{S_n}$ and $\mathbf{M}_{\lfloor n\mu \rfloor}$ a.n. are approximately the same. Furthermore, $n^{-1/2} S_n$ and $c_n^{-1}(M_n - d_n)$ are asymptotically independent, see [74, Lemma 21.19]. Accordingly, the extremes of $\mathbf{M}_N$ and N are asymptotically independent. When $F_N \in \mathcal{D}(\Phi_\alpha)$ (heavy-tailed), with $0 < \alpha \leq 1$, then $\mathbb{E}N = \infty$ and S_n and M_n a.n. are asymptotically dependent, whence the extremes of $\mathbf{M}_N$ and N are asymptotically dependent too. Furthermore, if $\alpha \in (0,1)$ and $\tilde{c}_n := c_n \Gamma^{1/\alpha}(1-\alpha)$, then $S_n/\tilde{c}_n$ converges in distribution to S, a one-sided stable random variable with stability parameter α. That is, $L_S(r) = \mathbb{E}e^{-rS} = e^{-r^{-\alpha}}$, $r > 0$. In this case, the asymptotic distributions of $\mathbf{c}_n^{-1} \mathbf{M}_{S_n}$ and $\mathbf{c}_n^{-1} \mathbf{M}_{\lfloor \tilde{c}_n S \rfloor}$ coincide, producing a location-scale mixture of G, the max-stable distribution obtained in the case of deterministic block size. This point is formalized by the following proposition.

Proposition 3.2.4 *Assume $F_{\mathbf{X}} \in \mathcal{D}(G)$ and let $\mathbf{c}_n$, $\mathbf{d}_n$ and c_n denote the norming sequences introduced in Theorem 3.2.1 for the case $F_N \in \mathcal{D}(\Phi_\alpha)$, with $\alpha \in (0,1)$. Then, defining $\tilde{c}_n := c_n \Gamma^{1/\alpha}(1-\alpha)$, it holds that*

(i) $S_n/\tilde{c}_n \rightsquigarrow S$, as $n \to \infty$;

(ii) $\lim_{n\to\infty} \mathbb{P}(\mathbf{M}_{\lfloor \tilde{c}_n S \rfloor} \leq \mathbf{c}_n \mathbf{x} + \mathbf{d}_n) = G_\alpha(\mathbf{x})$, *for all the continuity points $\mathbf{x}$ of G_α.*

3.3 Characterization of the limit model G_α

Let S be as above and denote by $\mathbf{Z}$ a random vector distributed according to G_*, a max-stable distribution with common unit-Fréchet margins. That is,

$$G_*(\mathbf{z}) = \exp\{-L(1/x_1, \dots, 1/x_d)\}, \quad \mathbf{z} \in (0,\infty)^d,$$

where L is the stable tail dependence function of G_*. Assume that S and $\mathbf{Z}$ are independent and define the random vector $\mathbf{R} := (SZ_1, \dots, SZ_d)$. Since L is homogeneous of order 1, we have that for every $\mathbf{z} > \mathbf{0}$

$$\begin{aligned} \mathbb{P}(\mathbf{R} \le \mathbf{z}) &= \mathbb{E}G_*(\mathbf{z}/S) = \mathbb{E}G_*^S(\mathbf{z}) = L_S(-\ln G_*(\mathbf{z})) \\ &= \exp\{-[-\ln G_*(\mathbf{z})]^\alpha\}. \end{aligned} \tag{3.7}$$

In particular, each component R_j is distributed according to Φ_α. This simple random scaling construction can be used to link the dependence structure of G_α, the max-attractor derived for $F_{\mathbf{M}_N}$ when $F_N \in \mathcal{D}(\Phi_\alpha)$ and $\alpha \in (0,1)$, with that of G, the max-attractor of $F_{\mathbf{X}}$.

Proposition 3.3.1 *Let $(\mathbf{X}_i)_{i=1}^\infty$ be a sequence of i.i.d. copies of $\mathbf{X}$ and assume that $F_{\mathbf{X}} \in D(G)$. Let S be independent from $\mathbf{X}_i$, $i \ge 1$.*

(i) *Let*

$$\mathbf{M}_n = \left(\max_{1\le i\le n} X_{i,1}, \dots, \max_{1\le i\le n} X_{i,d} \right),$$

then there exist sequences of maps $(\boldsymbol{\lambda}_n)_{i=1}^\infty$ and $(\boldsymbol{\mu}_n)_{i=1}^\infty$, with $\boldsymbol{\lambda}_n : (0,\infty) \mapsto (0,\infty)^d$ and $\boldsymbol{\mu}_n : (0,\infty) \mapsto \mathbb{R}^d$, such that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\mathbf{M}_n - \boldsymbol{\mu}_n(S)}{\boldsymbol{\lambda}_n(S)} \le \mathbf{x} \right) = G_\alpha(\mathbf{x}),$$

for all the continuity points $\mathbf{x}$ of G_α.

(ii) *Let G_* and $\mathbf{X}_*$ be defined as in Proposition 1.2.2 and denote by $F_{R\mathbf{X}_*}$ the distribution of $R\mathbf{X}_*$. Then $F_{R\mathbf{X}_*} \in \mathcal{D}(G_\alpha^*)$ where*

$$\begin{aligned} G_\alpha^*(\mathbf{z}) &:= \exp\{-(-\ln G_*(\mathbf{z}))^\alpha\}, \\ C_{G_\alpha^*}(\mathbf{u}) &:= G_\alpha^*((G_{\alpha,1}^*)^\leftarrow(u_1), \dots, (G_{\alpha,1}^*)^\leftarrow(u_d)) = C_{G_\alpha}(\mathbf{u}), \end{aligned}$$

for $\mathbf{z} \in (0,\infty)^d$ and $\mathbf{u} \in [0,1]^d$.

The above results illustrate how different constructions yield a limiting distribution with the same copula as G_α. The latter is characterized by the Pickands dependence function A_α, whose explicit expression is given in the following proposition.

Proposition 3.3.2 *The Pickands dependence function of G_α is*

$$A_\alpha(\mathbf{t}) = \|\mathbf{t}\|_{1/\alpha} A^\alpha \left(\left(\mathbf{t}/\|\mathbf{t}\|_{1/\alpha} \right)^{1/\alpha} \right), \quad \mathbf{t} \in \mathcal{S}_d,\ \alpha \in (0,1), \tag{3.8}$$

where A is the Pickands dependence function of G and

$$\|\mathbf{t}\|_{1/\alpha} = \left((1 - \sum_{i=1}^{d-1} t_i)^{1/\alpha} + \sum_{i=1}^{d-1} t_i^{1/\alpha} \right)^\alpha, \quad \mathbf{t} \in \mathcal{S}_d,\ \alpha \in (0,1). \tag{3.9}$$

The following facts are direct consequences of Proposition 3.3.2.

Remark 3.3.3 Observe that the smaller the parameter α, the more A_α represents a stronger dependence level than A. This effect can be easily appreciated by analysing extremal coefficients. Since $\|1/d, \ldots, 1/d\|_{1/\alpha} = d^{\alpha-1}$, from the definition of extremal coefficient given in Section 1.2 we obtain the following equality

$$\theta(G_\alpha) = (\theta(G))^\alpha. \tag{3.10}$$

Clearly, the smaller is α, the smaller is $\theta(G_\alpha)$ than $\theta(G)$.

Remark 3.3.4 When $A \equiv 1$, then $A_\alpha = \|\cdot\|_{1/\alpha}$, i.e. A_α is the Pickands dependence function of the well-known Logistic or Gumbel model, see e.g. [12, p. 146]. In this case, the components of a random vector distributed according to G_α are dependent if $\alpha \in (0,1)$, they become nearly independent as $\alpha \to 1$ and perfectly dependent as $\alpha \to 0$.

Remark 3.3.5 By solving for A in equation (3.8), we obtain the inverse relation between A_α and A, i.e.,

$$A^\star(\mathbf{t}) := A\left(\left(\mathbf{t}/\|\mathbf{t}\|_{1/\alpha} \right)^{1/\alpha} \right) = \left(A_\alpha(\mathbf{t})/\|\mathbf{t}\|_{1/\alpha} \right)^{1/\alpha}, \quad \mathbf{t} \in \mathcal{S}_d, \tag{3.11}$$

and

$$A(\mathbf{t}) = A^\star(\mathbf{t}^\alpha / \|\mathbf{t}^\alpha\|_1), \quad \mathbf{t} \in \mathcal{S}_d. \tag{3.12}$$

3.4 Inverse method to estimate the Pickands dependence function

3.4.1 A semiparametric estimator

In this section we introduce a new semiparametric procedure for inferring the Pickands dependence function A, working with maxima of aggregated data and exploiting the inverse relation between A and A_α in (3.11). Observe that, since $\mathbf{t} \mapsto (\mathbf{t}/\|\mathbf{t}\|_{1/\alpha})^{1/\alpha}$ is a bijective map, $A^\star$ univocally identifies A and the problem of estimating the former is equivalent to that of estimating the latter. When the sequence of observables, say $(\boldsymbol{\eta}_i, \xi_i)_{i=1}^\infty$, consists of i.i.d. draws from the joint distribution in (3.3) with $\alpha \in (0,1)$, an estimator of $A^\star$ can be constructed by plugging estimators of α and A_α in (3.11). More precisely, assume that the sample $(\boldsymbol{\eta}_1, \xi_1), \ldots, (\boldsymbol{\eta}_n, \xi_n)$ is observable. Notice that $\xi_1, \ldots, \xi_n$ are i.i.d. according to Φ_α. For estimating α, we consider two well known estimators: the Generalized Probability Weighted Moment (GPWM) - see [37] - and the Maximum Likelihood (ML). In the first case the estimator is given by

$$\widehat{\alpha}_n^{\text{GPWM}} := \left(k - 2\frac{\widehat{\mu}_{1,k}}{\widehat{\mu}_{1,k-1}}\right)^{-1}, \tag{3.13}$$

for $k \in \mathbb{N}_+$, where

$$\widehat{\mu}_{a,b} = \int_0^1 H_n^{\leftarrow}(v) v^a (-\ln v)^b \mathrm{d}v, \quad a, b \in \mathbb{N}$$

and

$$H_n(y) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(\xi_i \leq y), \quad y > 0. \tag{3.14}$$

In the second case the estimator is defined as

$$\widehat{\alpha}_n^{\text{ML}} := \underset{\alpha' \in (0,\infty)}{\operatorname{argmax}} \sum_{i=1}^n \ln \dot{\Phi}_{\alpha'}(\xi_i), \tag{3.15}$$

where $\dot{\Phi}_{\alpha'}(x) = \partial/\partial x\, \Phi_{\alpha'}(x)$, $x > 0$ denotes the probability density function of the Fréchet distribution with shape parameter α'. Next, notice that $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n$ are i.i.d. according to G_α. For estimating A_α we consider three well established, rank-based estimators: Pickands (P), Capéraà-Fougère-Genest (CFG) and Madogram

(MD) - see [56, 46, 9]. In the first case the estimator is

$$\widehat{A}^{\mathrm{P}}_{\alpha,n}(\mathbf{t}) = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\vartheta}_i(\mathbf{t})\right)^{-1}, \tag{3.16}$$

where

$$\widehat{\vartheta}_i(\mathbf{t}) := \min_{1\leq j\leq d}\left\{-\frac{1}{t_j}\ln\left(\frac{n}{n+1}G_{n,j}(\eta_{i,j})\right)\right\}$$

and for every $x\in\mathbb{R}$ and $j\in\{1,\dots,d\}$

$$G_{n,j}(x) := \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(\eta_{i,j}\leq x). \tag{3.17}$$

In the second case the estimator is

$$\widehat{A}^{\mathrm{CFG}}_{\alpha,n}(\mathbf{t}) = \exp\left(-\frac{1}{n}\sum_{i=1}^{n}\ln\widehat{\vartheta}_i(\mathbf{t}) - \varsigma\right), \tag{3.18}$$

where ς is Euler's constant. Finally, in the third case the estimator is

$$\widehat{A}^{\mathrm{MD}}_{\alpha,n}(\mathbf{t}) := \frac{\widehat{\nu}_n(\mathbf{t}) + c(\mathbf{t})}{1-\widehat{\nu}_n(\mathbf{t}) - c(\mathbf{t})}, \tag{3.19}$$

$$\widehat{\nu}_n(\mathbf{t}) := \frac{1}{n}\sum_{i=1}^{n}\left(\max_{j=1,\dots,d} G^{1/t_j}_{n,j}(\eta_{i,j}) - \frac{1}{d}\sum_{j=1}^{d}G^{1/t_j}_{n,j}(\eta_{i,j})\right), \tag{3.20}$$

where $u^{1/0}=0$ for $0<u<1$ by convention and

$$c(\mathbf{t}) = \frac{t_1}{1+t_1} + \cdots + \frac{t_d}{1+t_d}.$$

Now, plugging in equation (3.11) the estimator "$\bullet$" for α and the estimator "$\circ$" for A_α, we obtain the following compositional estimator for $A^\star$

$$\widehat{A^\star}^{\circ,\bullet}_n(\mathbf{t}) := \left(\widehat{A}^{\circ}_{\alpha,n}(\mathbf{t})/\|\mathbf{t}\|_{1/\widehat{\alpha}^\bullet_n}\right)^{1/\widehat{\alpha}^\bullet_n}, \quad \mathbf{t}\in\mathcal{S}_d. \tag{3.21}$$

In the following, we establish the asymptotic properties of the compositional estimator in (3.21) defined for all the combinations of the GPWM and ML estimators for α with the P, CFG and MD estimators for A_α. We make use of two assumptions.

Condition 3.4.1 For $j \in \{1, \ldots, d\}$, let $\mathcal{U}_j = \{\mathbf{u} \in [0,1]^d : 0 < u_j < 1\}$. Consider the following conditions:

(i) for $j \in \{1, \ldots, d\}$, the first-order partial derivative

$$\dot{C}_{G_\alpha}(\mathbf{u}) := \partial/\partial u_j C_{G_\alpha}(\mathbf{u})$$

exists and is continuous in $\mathcal{U}_j$;

(ii) for $i, j \in \{1, \ldots, d\}$, the second-order partial derivative $\ddot{C}_{G_\alpha}(\mathbf{u}) := \partial/\partial u_i \dot{C}_{G_\alpha}(\mathbf{u})$ exists and is continuous in $\mathcal{U}_i \cap \mathcal{U}_j$ and

$$\sup_{\mathbf{u} \in \mathcal{U}_i \cap \mathcal{U}_j} \max(u_i, u_j) |\ddot{C}_{G_\alpha}(\mathbf{u})| < \infty.$$

We are now ready to state the following result.

Theorem 3.4.2 *Assume Condition 3.4.1(i) holds true . In addition, for estimators* $\widehat{A^\star}_n^{\mathrm{P},\bullet}$, $\widehat{A^\star}_n^{\mathrm{CFG},\bullet}$, *assume that Condition 3.4.1(ii) holds also true; for estimators* $\widehat{A^\star}_n^{\circ,\mathrm{GPWM}}$ *assume that* $\alpha > 1/(k-1)$. *Then,*

$$\sqrt{n}\left(\widehat{A^\star}_n^{\circ,\bullet} - A^\star\right) \rightsquigarrow \phi_{\circ,\bullet}(\mathbb{C}_Q), \quad n \to \infty, \tag{3.22}$$

in $\ell^\infty(\mathcal{S}_d)$, *where the map* $\phi_{\circ,\bullet}$ *is given in Section 3.6.5,* $\mathbb{C}_Q$ *is a zero-mean Gaussian process with covariance function given by*

$$\mathrm{Cov}(\mathbb{C}_Q(\mathbf{u}), \mathbb{C}_Q(\mathbf{v})) = C_Q(\min(\mathbf{u}, \mathbf{v})) - C_Q(\mathbf{u}) C_Q(\mathbf{v}), \tag{3.23}$$

for every $\mathbf{u}, \mathbf{v} \in [0,1]^{d+1}$, *where*

$$C_Q(\mathbf{u}, v) = Q(G^{\leftarrow}_{\alpha,1}(u_1), \ldots, G^{\leftarrow}_{\alpha,d}(u_d), \Phi^{\leftarrow}_\alpha(v))$$

and the minimum is taken componentwise. Moreover,

$$\|\widehat{A^\star}_n^{\circ,\bullet} - A^\star\|_\infty \xrightarrow{p} 0, \quad n \to \infty,$$
$$\|\widehat{A^\star}_n^{\mathrm{MD},\mathrm{GPWM}} - A^\star\|_\infty \xrightarrow{as} 0, \quad n \to \infty.$$

Remark 3.4.3 The results in Theorem 3.4.2 are still valid when the modified versions proposed in [32] and [38] of the estimators P, CFG and MD for A_α are considered in place of (3.16), (3.18) and (3.19), respectively - see Section 4.2.1 for the MD. In the mentioned works an adjusted version of the estimators is introduced to guarantee that $\widehat{A}^\circ_{\alpha,n}(\mathbf{e}_j) = 1$ for all $n = 1, 2 \ldots$ and $j = 1, \ldots, d$, where $\mathbf{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)$.

Remark 3.4.4 The asymptotic results provided in Theorem 3.4.2 concern specific estimators of α and A_α. In fact, they are derived using Propositions 3.6.4 and 3.6.5 in Section 3.6, which establish similar results for a wider class of generic estimators $\widehat{\alpha}_n$ and $\widehat{A}_{\alpha,n}$.

Remark 3.4.5 As for Conditions 3.4.1(i)-(ii), in light of the identity $L_\alpha(\mathbf{z}) = L(\mathbf{z}^{1/\alpha})^\alpha$, Proposition 1 in [33] guarantees that C_{G_α} has the required smoothness properties if L satisfies Assumption 2 thereof.

Remark 3.4.6 As discussed at the beginning of this subsection, inferential problems concerning A (e.g. testing) can be equivalently solved after rephrasing them in terms of $A^\star$. Still, a natural question is whether, defining an estimator $\widehat{A}_n$ of A via (3.12), i.e.

$$\widehat{A}_n(\mathbf{t}) = \widehat{A^\star}_n^{\circ,\bullet}\left(\mathbf{t}^{\widehat{\alpha}_n^\bullet}/\|\mathbf{t}^{\widehat{\alpha}_n^\bullet}\|_1\right), \quad \mathbf{t} \in \mathcal{S}_d,$$

its asymptotic properties can be directly desumed from Theorem 3.4.2. This is the case for uniform consistency. Weak converge is technically more involved. The following decomposition hols:

$$\begin{aligned}\sqrt{n}\left(\widehat{A}_n(\mathbf{t}) - A(\mathbf{t})\right) = &\sqrt{n}\left[\widehat{A^\star}_n^{\circ,\bullet}\left(\mathbf{t}^{\widehat{\alpha}_n^\bullet}/\|\mathbf{t}^{\widehat{\alpha}_n^\bullet}\|_1\right) - A\left(\mathbf{t}^{\widehat{\alpha}_n^\bullet}/\|\mathbf{t}^{\widehat{\alpha}_n^\bullet}\|_1\right)\right] \\ &+ \sqrt{n}\left[A\left(\mathbf{t}^{\widehat{\alpha}_n^\bullet}/\|\mathbf{t}^{\widehat{\alpha}_n^\bullet}\|_1\right) - A(\mathbf{t}^\alpha/\|\mathbf{t}^\alpha\|_1)\right].\end{aligned}$$

By exploiting the asymptotic tightness of $\sqrt{n}(\widehat{\alpha}_n^\bullet - \alpha)$ and an asymptotic equicontinuity argument – see e.g. [75, Theorem 1.5.7] – the first term might be shown to be asymptotically equivalent (in outer probability) to

$$\sqrt{n}\left[\widehat{A^\star}_n^{\circ,\bullet}(\mathbf{t}^\alpha/\|\mathbf{t}^\alpha\|_1) - A(\mathbf{t}^\alpha/\|\mathbf{t}^\alpha\|_1)\right],$$

which converges in distribution by Theorem 3.4.2. As for the second term, it would require a non-trivial application of the (functional) delta method and addtional smoothness assumptions on A might be needed. Once weak convergence is established for both terms, it can be concluded that the sum of the two processes is asymptotically tight, and its limit distribution might be deduced from consideration of the marginal distributions – see also [73, pp. 246-247].

3.4.2 Simulation study

We illustrate the finite sample performances of the estimator $\widehat{A^\star}_n^{\circ,\bullet}$ through a simulation study consisting of two experiments. Herein we consider the adjusted versions of the P, CFG and MD estimators mentioned in Remark 3.4.3.

First experiment: We sample from the limiting distribution G_α in (3.7). Precisely, we let $\mathbf{Z}$ be a two-dimensional random vector with a max-stable distribution G with common unit-Fréchet margins and a Symmetric Logistic dependence model with dependence parameter $\psi \in (0,1]$ (e.g., [68]). As above, we denote by S a positive α-stable random variable with $\alpha \in (0,1)$. With this setup, we ensure that $\mathbf{R} = S\mathbf{Z}$ is distributed according to G_α, with α-Fréchet margins - see Section 3.3. We implement the following routine:

1. we simulate n values from G_α and we set $\xi_i = \max(R_{i,1}, R_{i,2})$ and $\eta_{i,j} = R_{i,j}$, with $j = 1, 2$ and $i = 1, \dots, n$;
2. then, we estimate α by the GPWM estimator $\widehat{\alpha}_n^{\text{GPWM}}$ in equation (3.13), with $k = 5$, and the ML estimator $\widehat{\alpha}_n^{\text{ML}}$ in (3.15);
3. we estimate A_α by the P estimator $\widehat{A}_{\alpha,n}^{\text{P}}$ in (3.16), CFG estimator $\widehat{A}_{\alpha,n}^{\text{CFG}}$ in (3.18) and MD estimator $\widehat{A}_{\alpha,n}^{\text{MD}}$ in (3.19);
4. finally, we estimate $A^\star$ using the compositional estimator $\widehat{A^\star}_n^{\circ,\bullet}$ in equation (3.21).

Note that in this setting an estimator of α can be recovered using only a sample from G_α, i.e. draws from the $(d+1)$-th component of Q are unnecessary for estimation – in general, this might not possible and one should stick to the estimation procedure of Subsection 3.4.1. Yet, the asymptotic properties of the present version of estimator $\widehat{A^\star}_n^{\circ,\bullet}$ are similar to those reported in Theorem 3.4.2 (see Subsection 3.7). We repeat steps 1.-4. for different values of the dependence parameters α and ψ and different sample sizes. Specifically, we consider $\alpha = 0.5, 0.633, 0.767, 0.9$ and 15 equally spaced values in $[0.1, 1]$ for ψ; as for the sample sizes, we set $n = 50, 100$. For each configuration (α, ψ, n), we repeat the simulation and estimation steps 1000 times and we compute a Monte Carlo approximation of

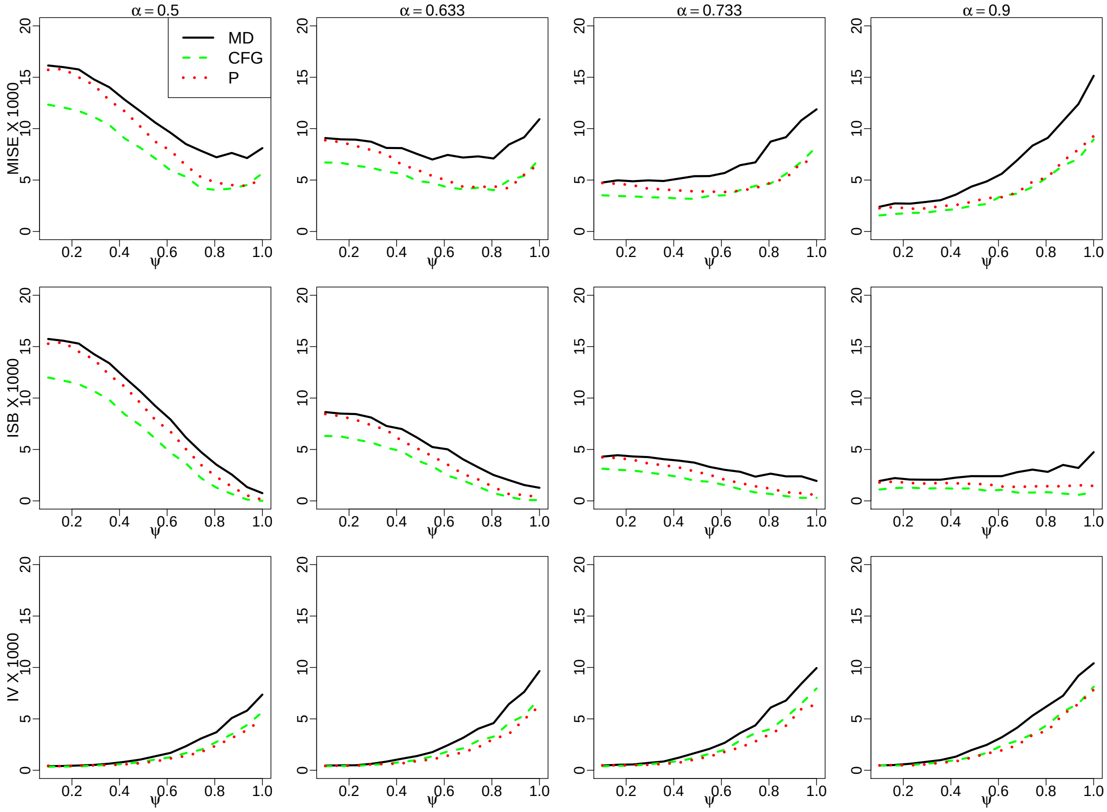

Figure 3.1: MISE, ISB and IV for 1000 samples of size 50 from the bivariate extreme-value copula in (3.6) with the logistic Pickands dependence model, for different values of the dependence parameters ψ and α. The function $A^\star$ is estimated by the compositional estimator $\widehat{A^\star}_n^{\mathrm{GPWM},\bullet}$ in formula (3.21).

the Mean Integrated Squared Error (MISE), i.e.,

$$\begin{aligned}\mathrm{MISE}(\widehat{A^\star}_n, A^\star) &= \mathbb{E}\left[\int_{\mathcal{S}_d} \left(\widehat{A^\star}_n(\mathbf{t}) - A^\star(\mathbf{t})\right)^2 \mathrm{d}\mathbf{t}\right] \\ &= \int_{\mathcal{S}_d} \left[\mathbb{E}\left(\widehat{A^\star}_n(\mathbf{t})\right) - A^\star(\mathbf{t})\right]^2 \mathrm{d}\mathbf{t} \\ &\quad + \int_{\mathcal{S}_d} \mathbb{E}\left[\widehat{A^\star}_n(\mathbf{t}) - \mathbb{E}\left(\widehat{A^\star}_n(\mathbf{t})\right)\right]^2 \mathrm{d}\mathbf{t},\end{aligned}$$

where the first and second summand on the right hand side are known as integrated squared bias (ISB) and integrated variance (IV) - see [29, Ch. 6.3].

The results for the sample size $n = 50$ are summarized in Figure 3.1. The MISE, ISB and IV ($\times 1000$) of the GPWM-based estimators are reported from the first to the third row. The solid black, dashed green and dotted red lines report the results obtained estimating A_α with P, CFG and MD estimators, respectively. The results for the different values of α are reported along the columns. For each fixed value of α we see that the IV is close to zero at the strongest dependence level ($\psi = 0.1$), then it increases as the dependence level decreases (i.e. ψ increases toward one). This is consistent with what has been observed in [76, Section 3.2], for a similar estimation problem. In contrast, for $\alpha = 0.5, 0.633$ the ISB takes the largest value at $\psi = 0.1$ and then it decreases as the dependence level decreases. This seems to be due to the fact that, when the dependence level is strong (ψ close to 0), the estimator $\widehat{A^\star}_n^{\circ,\bullet}$ tends to lie above the true function $A^\star$. In the cases of $\alpha = 0.733, 0.9$, the ISB displays less significant changes for different dependence levels and turns out to be generally below the values recorded for $\alpha = 0.5, 0.633$. This might be due to the fact that larger values of α inflates the initial dependence level, for any given ψ, inducing an upward bias for $\widehat{A^\star}_n^{\circ,\bullet}$. Overall, for $\alpha = 0.5$, the MISE takes the largest value at $\psi = 0.1$ and then it decreases as the dependence level decreases. In the case of $\alpha = 0.633$, the MISE does not change much all over the range of the dependence levels, since the ISB and the IV compensate each other. In the cases $\alpha = 0.767, 0.9$, the IV grows much more than the ISB decreases, implying that the MISE increases as the dependence level decreases. The smallest values of the ISB and the IV are obtained with the CFG-based and P-based estimator, respectively. Concluding, on the basis of the MISE, the best performances are obtained with the CFG-based estimator, although there is little difference with respect to the P-based estimator. In the supplementary material (Subsection 3.7.1), we show that the difference between the performances of the P-, CFG- and MD-based estimators are considerably mitigated already for sample size $n = 100$.

In Figure 3.2 the comparison between the estimation results obtained with the GPWM- and ML-based estimators is reported. Precisely, from the first to the third row the ratio between the MISE, ISB and IV computed estimating $A^\star$ via the GPWM- and ML-based estimators are displayed. On the basis of the ISB, in the cases $\alpha = 0.5, 0.633$, the GPWM- and ML-based estimators have almost the same performances. However, for $\alpha = 0.633$, the ML-based estimators outperform the GPWM-based estimators when ψ is close to 1, that is, at milder dependence levels. In the cases $\alpha = 0.767, 0.9$, the ML-based estimators clearly outperforms the

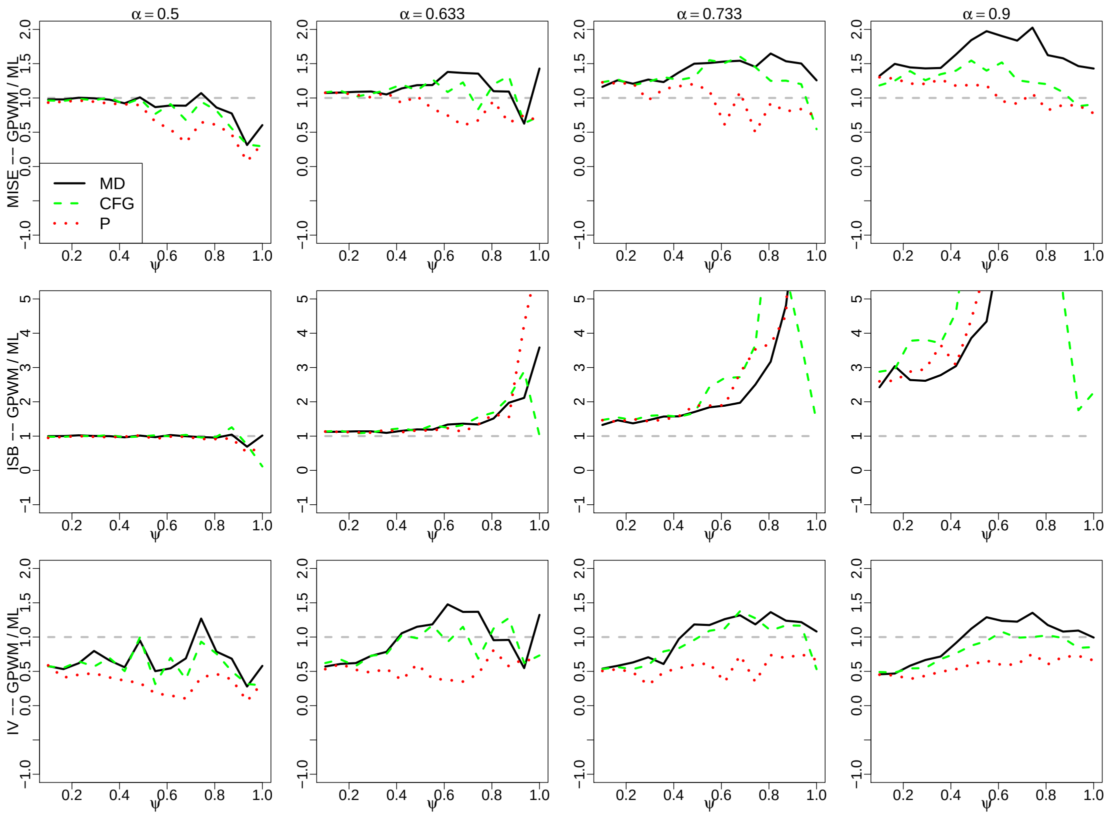

Figure 3.2: Ratio between MISE, ISB and IV computed estimating $A^\star$ by the estimator $\widehat{A^\star}_n^{\text{GPWM},\bullet}$ and $\widehat{A^\star}_n^{\text{ML},\bullet}$ in formula (3.21). The same setting of Figure 3.1 is considered.

GPWM-based estimators. On the basis of the IV, the GPWM-based estimators basically outperform the ML-based estimators, except for the GPWM-MD-based estimator. Indeed, for the cases $\alpha = 0.633, 0.767, 0.9$, the GPWM-MD-based estimator outperforms the ML-MD-based estimator only for strong dependence levels, the opposite being true for weak dependence levels. Overall, on the basis of the MISE, the GPWM-based estimators outperform the ML-based estimators for the case $\alpha = 0.5$, while they perform almost the same for the case $\alpha = 0.633$. On the contrary, for the cases $\alpha = 0.767, 0.9$, the ML-based estimators outperform the GPWM-based estimators, except for the GPWM-P-based and ML-P-based estimators, which have almost the same performances.

Second experiment: we study the performances of the compositional estimator $\widehat{A^\star}_n^{\circ,\bullet}$ in (3.21) when the observations are only approximately drawn from the limit distribution Q. This is a more realistic scenario. Specifically, we set $N = \lceil N' \rceil$, where N' denotes a standard Pareto distribution with shape parameter $\alpha \in (0,1)$. We denote by $\mathbf{X}$ a two-dimensional random vector distributed according to a standard bivariate Student-t distribution, with given values of the correlation ρ and the degrees of freedom υ. We recall that a Student-t distribution is in the domain of attraction of a multivariate extreme-value distribution, G, with an extreme-value copula, C_G, known as the extremal-t copula (e.g., [51]). In the bivariate case, the extremal coefficient of the extremal-t copula is $\theta(G) = 2T_{\upsilon+1}[\{(\upsilon+1)(1-\rho)/(1+\rho)\}^{1/2}]$, where $T_{\upsilon+1}$ is a univariate standard Student-t distribution with $\upsilon + 1$ degrees of freedom. We implement the following routine

1. first we sample N;
2. then we sample a number of i.i.d. replicates of $\mathbf{X}$ which equals the observed value of N;
3. computing maxima componentwise on the block of observations obtained at point 2., we now obtain a realization of $\mathbf{M}_N$ in (3.2);
4. we repeat steps 1.–3. $n' = 500$ times, generating n' independent observations from the pair $(N, \mathbf{M}_N)$. With those, we obtain a realization of the random variable $\xi = \max(N_1, \dots, N_{n'})$ and vector $\boldsymbol{\eta} = \max(\mathbf{M}_{N_1}, \dots, \mathbf{M}_{N_{n'}})$, maxima being meant componentwise;
5. we repeat steps 1.–4. n times generating a data sample approximately drawn from the distribution Q, in the first line of (3.3);
6. we estimate α by the GPWM and ML estimators in (4.5.5) and (3.15), using the realization of the sequence $\xi_1, \dots, \xi_n$;
7. we estimate A_α by the P, CFG and MD estimators in (3.16)-(3.19), using the realization of the sequence $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n$;
8. finally, we estimate $A^\star$ by the compositional estimator $\widehat{A^\star}_n^{\circ,\bullet}$.

The above routine is iterated 1000 times and an approximation of the MISE is computed. The experiment is repeated for different values of the model parameters and different sample sizes. In particular, as for the parameters of the Student-t

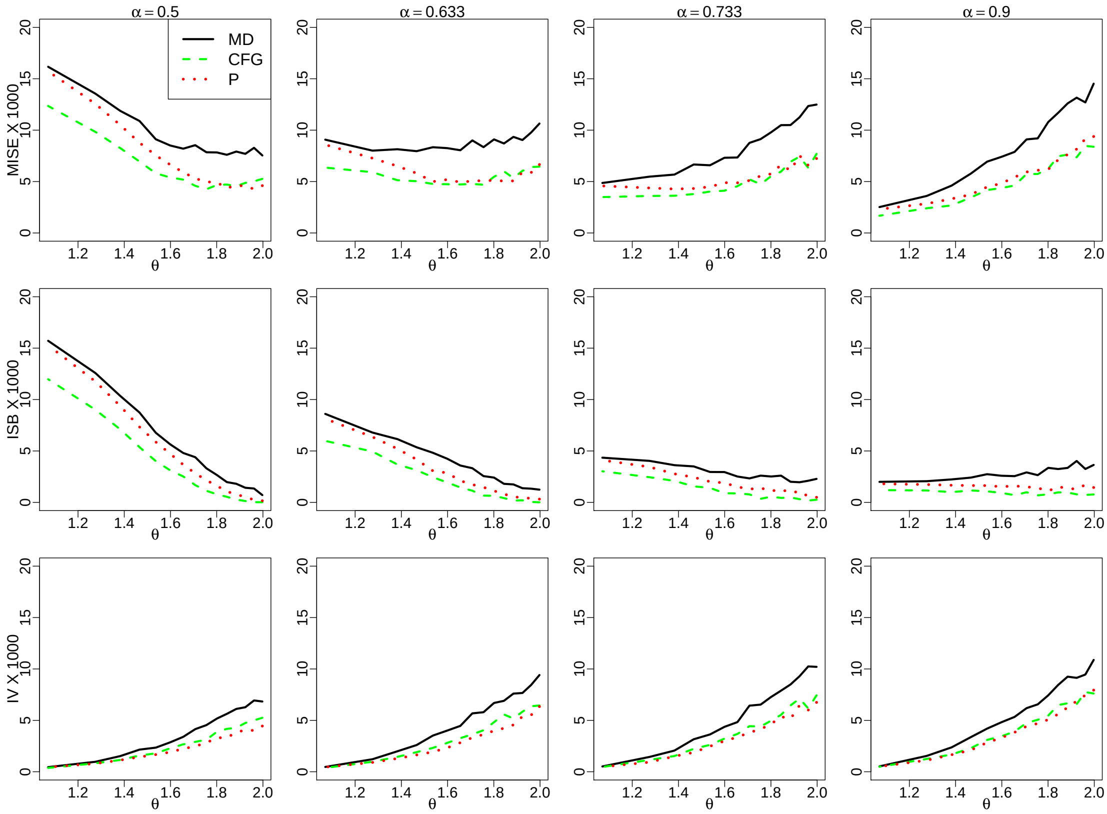

Figure 3.3: MISE, ISB and IV for 1000 samples of size 50 drawn from a distribution in the max-domain of Q, obtained on the basis of the standard Pareto distribution for N and the bivariate Student-t distribution for $\mathbf{X}$, for different values of the parameters α and (ρ, υ). The parameter $\theta \equiv \theta(G)$ is the extremal coefficient related to the corresponding extreme-value copula, known as extremal-t copula.

distribution, we consider degrees of freedom $\upsilon = 1$ and 15 equally spaced values of ρ in $[-0.99, 0.99]$. With these parameters values the extremal coefficient $\theta(G)$ (related to the extremal-t copula) takes values in $[1, 2]$, the lower and upper bounds representing the cases of complete dependence and independence. Furthermore, the considered values of α are the same as those of the previous experiment and the considered sample sizes are $n = 50, 100$.

Figure 3.3 displays the results obtained with the GPWM-based estimators for the sample size $n = 50$. Although in this experiment we consider synthetic data from a more complicated model than that of the previous experiment, and notwithstanding that the data are only approximately drawn from the distribution

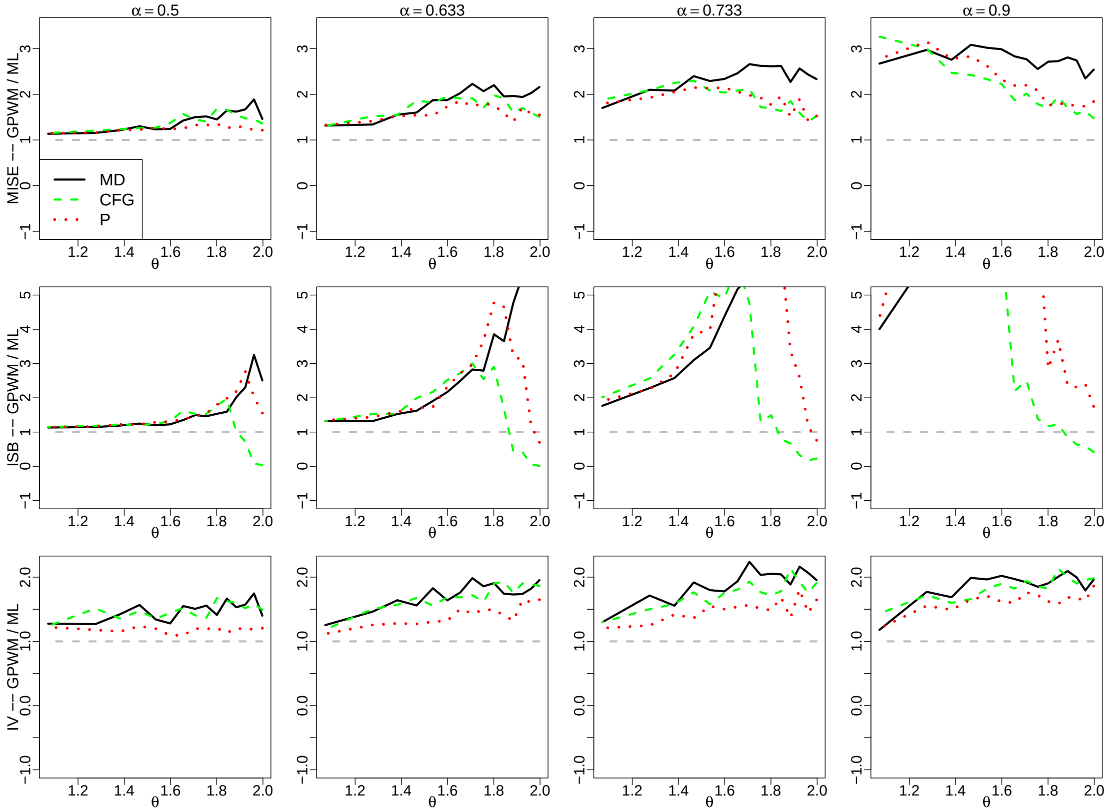

Figure 3.4: Ratio between MISE, ISB and IV computed estimating $A^\star$ by the estimator $\widehat{A^\star}_n^{\mathrm{GPWM},\bullet}$ and $\widehat{A^\star}_n^{\mathrm{ML},\bullet}$ in formula (3.21). The same setting of Figure 3.3 is considered.

Q, the results summarised by the ISB, the IV and the MISE are very similar in both experiments. In particular, the compositional estimators we propose display a small ISB, despite the uniform consistency guarantee provided by Theorem 3.4.2 for random samples drawn from Q does not directly extend to the present setting. Similar conclusions are obtained with sample size $n = 100$, the results for this case being reported in the supplementary material (see Subsection 3.7.1).

Figure 3.4 displays the comparison between the estimation performances obtained with the GPWM- and ML-based estimators. Concerning the configurations with $\alpha > 0.5$, the ML-based estimators considerably outperform the GPWM-based estimators in terms of MISE. Specifically, the IV is smaller for the ML-based estimators, which provide better performances for weaker dependence structures (i.e. $\theta(G)$ approaches 2). The ML-based estimators are much less biased than then

GPWM-based estimators and the difference is much more pronounced for larger values of α and milder dependence levels (i.e. $\theta > 1.5$). Yet, for the P- and CFG-based estimators, such difference is mitigated as θ approaches 2.

3.5 Discusssion

The analysis of aggregated data is becoming a very important topic in statistics. In insurance, finance, risk management and big-data problems, examples of aggregated data structures of particular interest are represented by total and maximum amounts, computed on a random number of observations. These can be described through the random vectors $\mathbf{S}_N$ and $\mathbf{M}_N$ in (3.1)-(3.2). Beyond the univariate case, in which several results on the extremal behaviour of S_N are available, a multivariate extreme-value theory that characterizes the extremal behaviour of $\mathbf{S}_N$ and $\mathbf{M}_N$ is not yet available. This chapter makes a first step in establishing such a theory. Although observations of the form of $\mathbf{S}_N$ are involved in a wider range of applications, we have studied the extremal behaviour of $\mathbf{M}_N$ as a starting point, since it is simpler to analyze. Nevertheless, simpler does not imply trivial, and there are still several open problems. For instance, an exact algorithm to draw samples from the limit distribution Q (first line of 3.3) is not available yet. The modelling of the joint upper tail associated to the distribution function of $\mathbf{M}_N$ would benefit from the derivation of new nonparametric estimators, defined on the basis of threshold exceedances (for at least one component). This would make the theory useful to all applications in which few block-maxima are available, but a larger number of threshold exceedances is at hand.

Concluding, an important step forward in the construction of a multivariate extreme-value theory of aggregated data consists in extending our results (probabilistic and inferential) to the case of the random vector $\mathbf{S}_N$. Once again, this would broaden the practical applicability of our theoretical framework.

3.6 Proofs

3.6.1 Proof of Theorem 3.2.1

Let $U_j(t) := F_{X_j}^{\leftarrow}(1 - 1/t)$, $t > 1$, $D_j(x) = G_j^{\leftarrow}(e^{-1/x})$, $x \in \text{supp}(G_j)$, for $j = 1, \ldots, d$, and

$$\begin{aligned} F_*(\cdot) &= F_{\mathbf{X}}(U_1(\cdot), \ldots, U_d(\cdot)), \\ G_*(\mathbf{y}) &= G(D_1(y_1), \ldots, D_d(y_d)), \quad \mathbf{y} \in (0, \infty)^d, \end{aligned}$$

Since $F_{\mathbf{X}} \in \mathcal{D}(G)$, then

$$\lim_{n\to\infty} \frac{1 - F_*(n\mathbf{y})}{1 - F_*(n\mathbf{1})} = \frac{-\ln G_*(\mathbf{y})}{\theta(G)}, \quad \mathbf{y} \in (0, \infty)^d, \tag{3.24}$$

see Proposition 1.2.2 and [60, Ch. 5].

The proof is organized in three parts: the derivation of the norming constants, a preliminary result and the main body of the proof.

Norming constants

We first derive c_n and d_n. When $F_N \in \mathcal{D}(\Phi_\alpha)$, $\alpha \in (0, 1]$, then $\bar{F}_N(y) = 1 - F_N(y)$, $y > 0$, satisfies

$$\bar{F}_N(y) \sim \mathcal{L}(y)\, y^{-\alpha}, \quad y \to \infty. \tag{3.25}$$

Set $z_n = F_N^{\leftarrow}(1 - [\theta(G)]^\alpha / [n\Gamma(1 - \alpha)])$, where $\Gamma(0) := 1$ by convention, then

$$\mathcal{L}(z_n) z_n^{-\alpha} \sim [\theta(G)]^\alpha / [n\Gamma(1 - \alpha)], \quad n \to \infty$$

and

$$\Gamma(1 - \alpha)[\bar{F}_*(m_n)/\theta(G)]^\alpha \mathcal{L}\left(1/\bar{F}_*(m_n)\right) \sim n^{-1}, \quad n \to \infty. \tag{3.26}$$

The sequence m_n in the above display is defined as $m_n = F_*^{\leftarrow}(1 - 1/z_n)$, with $F_*(y) = F_*(y\mathbf{1})$ for $y > 0$, so that z_n satisfies $z_n \sim 1/\bar{F}_*(m_n)$ as $n \to \infty$. Hence, in this case we set $d_n = 0$ and

$$c_n = \theta(G)/[\bar{F}_*(m_n)\Gamma^{1/\alpha}(1 - \alpha)].$$

Instead, when $F_N \in \mathcal{D}(\Phi_\alpha)$, $\alpha > 1$, or $F_N \in \mathcal{D}(\Lambda)$, since $\mathbb{E}N \in (0, \infty)$ we denote $m_n = n\mathbb{E}N$ and define c_n and d_n in the standard way, as described for instance in [60, pp. 48-54].

We now derive $\mathbf{c}_n$ and $\mathbf{d}_n$. When $F_N \in \mathcal{D}(\Phi_\alpha)$, $\alpha > 0$,

(i) if $F_{X_j} \in \mathcal{D}(\Phi_{\beta_j})$, then we set $d_{n,j} = 0$ and

$$c_{n,j} = \begin{cases} U_j(m_n), & \alpha \in (0,1] \\ U_j(n)[\mathbb{E}N]^{1/\beta_j}, & \alpha > 1 \end{cases}$$

(ii) if $F_{X_j} \in \mathcal{D}(\Lambda)$, then we set

$$\begin{aligned} c_{n,j} &= \begin{cases} \omega_j(d_{n,j}), & \alpha \in (0,1] \\ \omega_j(\Upsilon_j^{\leftarrow}(1 - 1/\delta_j n)), & \alpha > 1 \end{cases} \\ d_{n,j} &= \begin{cases} \Upsilon_j^{\leftarrow}(1 - 1/\delta_j m_n), & \alpha \in (0,1] \\ c_{n,j} \ln \mathbb{E}N + \Upsilon_j^{\leftarrow}(1 - 1/\delta_j n), & \alpha > 1, \end{cases} \end{aligned}$$

where Υ_j is the Von Mises function associated to $\bar{F}_{X_j}$, ω_j its auxiliary function (e.g., [60, pp. 40-43]) and $\delta_j = \lim_{x\to\infty} \bar{F}_{X_j}(x)/(1 - \Upsilon_j(x))$;

(iii) if $F_{X_j} \in \mathcal{D}(\Psi_{\beta_j})$, then we set $d_{n,j} = x_{0,j}$, where $x_{0,j} = \sup\{x : F_{X_j}(x) < 1\}$, and

$$c_{n,j} = \begin{cases} [\tilde{F}_{X_j}^{\leftarrow}(1 - 1/m_n)]^{-1}, & \alpha \in (0,1] \\ \{\tilde{F}_{X_j}^{\leftarrow}(1 - 1/n)[\mathbb{E}N]^{1/\beta_j}\}^{-1}, & \alpha > 1, \end{cases}$$

where $\tilde{F}_{X_j}(x) = F_{X_j}(x_{0,j} - 1/x)$.

When $F_N \in \mathcal{D}(\Lambda)$, then we set $\mathbf{c}_n$ and $\mathbf{d}_n$ equal to the case of $F_N \in \mathcal{D}(\Phi_\alpha)$, with $\alpha > 1$.

With these norming constants, we obtain the following approximations

$$U_j^{\leftarrow}(c_{n,j}x_j + d_{n,j}) \sim m_n\, D_j^{\leftarrow}(x_j) = \begin{cases} m_n\, x_j^{\beta_j}, & F_{X_j} \in \mathcal{D}(\Phi_{\beta_j}) \\ m_n\, e^{x_j}, & F_{X_j} \in \mathcal{D}(\Lambda) \\ m_n\, (-x_j)^{-\beta_j}, & F_{X_j} \in \mathcal{D}(\Psi_{\beta_j}) \end{cases} \tag{3.27}$$

as n goes to infinity.

Preliminary result

Lemma 3.6.1 *Let* $p_n(\mathbf{x}) = F_{\mathbf{X}}(\mathbf{c}_n\mathbf{x} + \mathbf{d}_n)$. *If* $F_N \in \mathcal{D}(\Phi_\alpha), \alpha \in (0,1]$, *then for all* $\mathbf{x} \in \mathbb{R}^d$ *we have*

$$\lim_{n\to\infty} n[1 - \mathbb{E}p_n^N(\mathbf{x})] = [-\ln G(\mathbf{x})]^\alpha.$$

If $F_N \in \mathcal{D}(\Lambda)$ *or* $F_N \in \mathcal{D}(\Phi_\alpha), \alpha > 1$, *then we have*

$$\lim_{n\to\infty} n[1 - \mathbb{E}p_n^N(\mathbf{x})] = -\ln G(\mathbf{x}).$$

Proof. When $F_N \in \mathcal{D}(\Phi_\alpha)$, $\alpha \in (0,1]$, then by Corollary 8.17 in [5] we have that

$$1 - L_N(s) \sim \Gamma(1-\alpha)\mathcal{L}(1/s)s^\alpha, \quad s \to 0^+,$$

with L_N the Laplace transform of N. Therefore, as $n \to \infty$

$$\begin{aligned} 1 - \mathbb{E}p_n^N(\mathbf{x}) &= 1 - L_N(-\ln p_n(\mathbf{x})) \\ &\sim \Gamma(1-\alpha)\mathcal{L}(-1/\ln p_n(\mathbf{x}))[-\ln p_n(\mathbf{x})]^\alpha \\ &\sim \Gamma(1-\alpha)\mathcal{L}(1/(1-p_n(\mathbf{x})))[1-p_n(\mathbf{x})]^\alpha. \end{aligned}$$

By (3.24) and (3.27) we obtain

$$\begin{aligned} 1 - p_n(\mathbf{x}) &\sim 1 - F_*(m_n\, D_1^{\leftarrow}(x_1), \dots, m_n\, D_d^{\leftarrow}(x_d)) \\ &\sim \bar{F}_*(m_n)[-\ln G(\mathbf{x})]/\theta(G) \end{aligned} \tag{3.28}$$

as $n \to \infty$. Using this last approximation and (3.26), we conclude that $n[1 - \mathbb{E}p_n^N(\mathbf{x})]$ is asymptotically equivalent to

$$\begin{aligned} &\Gamma(1-\alpha)n\{\bar{F}_*(m_n)[-\ln G(\mathbf{x})]/\theta(G)\}^\alpha \mathcal{L}(1/\{\bar{F}_*(m_n)[-\ln G(\mathbf{x})]/\theta(G)\}) \\ &\sim \Gamma(1-\alpha)n\mathcal{L}\left(1/\bar{F}_*(m_n)\right)[\bar{F}_*(m_n)/\theta(G)]^\alpha[-\ln G(\mathbf{x})]^\alpha \\ &\sim [-\ln G(\mathbf{x})]^\alpha. \end{aligned}$$

as $n \to \infty$.

When $F_N \in \mathcal{D}(\Phi_\alpha)$, $\alpha > 1$, or $F_N \in \mathcal{D}(\Lambda)$, it holds that $\mathbb{E}N < \infty$. Consequently, denoting by L_N the Laplace transform of N, we have that

$$1 - L_N(s) \sim s\mathbb{E}N, \quad s \to 0^+.$$

Therefore, the following asymptotic equivalence holds true

$$\begin{aligned} 1 - \mathbb{E}p_n^N(\mathbf{x}) &= 1 - L_N(-\ln p_n(\mathbf{x})) \\ &\sim -\ln p_n(\mathbf{x})\mathbb{E}N \\ &\sim [1 - p_n(\mathbf{x})]\mathbb{E}N, \end{aligned}$$

as $n \to \infty$. We recall that in these cases $m_n = n\mathbb{E}N \in (0, \infty)$, thus by (3.24) and (3.27) we have that

$$\begin{aligned} 1 - p_n(\mathbf{x}) &\sim 1 - F_*(n\mathbb{E}N\, D_1^{\leftarrow}(x_1), \dots, n\mathbb{E}N\, D_d^{\leftarrow}(x_d)) \\ &\sim \bar{F}_*(n\mathbb{E}N)[-\ln G(\mathbf{x})]/\theta(G) \\ &\sim -\ln G(\mathbf{x})/(n\mathbb{E}N), \end{aligned} \tag{3.29}$$

as $n \to \infty$. Finally, we have that

$$\lim_{n\to\infty} n[1 - \mathbb{E}p_n^N(\mathbf{x})] = -\ln G(\mathbf{x}).$$

The statement now follows. ∎

Main body of the proof

As $n \to \infty$, the term $-n \ln \mathbb{P}(\mathbf{M}_N \leq \mathbf{c}_n\mathbf{x} + \mathbf{d}_n, N \leq c_n y + d_n)$ is asymptotically equivalent to

$$\begin{aligned} & n[1 - \mathbb{P}(\mathbf{M}_N \leq \mathbf{c}_n\mathbf{x} + \mathbf{d}_n, N \leq c_n y + d_n)] \\ &= n\left\{1 - \mathbb{E}\left[p_n^N(\mathbf{x})\mathbb{1}(N \leq u_n(y))\right]\right\} \\ &= n[1 - \mathbb{E}p_n^N(\mathbf{x})] + n\mathbb{E}\left[p_n^N(\mathbf{x})\mathbb{1}(N > u_n(y))\right] \\ &\equiv T_{1,n} + T_{2,n}, \end{aligned}$$

where $u_n(y) = c_n y + d_n$. The limiting behavior of $T_{1,n}$ has been established in Lemma 3.6.1. As for $T_{2,n}$, start observing that, for every $v \in [0, 1]$,

$$\begin{aligned} & \mathbb{P}\left\{p_n^N(\mathbf{x})\mathbb{1}(N > u_n(y)) \leq v\right\} \\ &= F_N(u_n(y)) + \mathbb{P}(N \geq \ln v/\ln p_n(\mathbf{x}), N > u_n(y)) \\ &= \begin{cases} F_N(u_n(y)), & v = 0 \\ F_N(u_n(y)) + \bar{F}_N(\ln v/\ln p_n(\mathbf{x})), & 0 < v < p_n^{u_n(y)}(\mathbf{x}) \\ 1, & v \geq p_n^{u_n(y)}(\mathbf{x}) \end{cases}. \end{aligned}$$

Therefore,

$$\begin{aligned} T_{2,n} &= n\int_0^{p_n^{u_n(y)}(\mathbf{x})} \mathbb{P}\left\{p_n^N(\mathbf{x})\mathbb{1}(N > u_n(y)) > v\right\} \mathrm{d}v \\ &= np_n^{u_n(y)}(\mathbf{x})\bar{F}_N(u_n(y)) - n\int_0^{p_n^{u_n(y)}(\mathbf{x})} \bar{F}_N\left(\frac{\ln v}{\ln p_n(\mathbf{x})}\right) \mathrm{d}v \\ &\equiv nI_{n,1} + nI_{n,2}. \end{aligned}$$

When $F_N \in \mathcal{D}(\Phi_\alpha)$ with $\alpha \in (0,1]$, from (3.25), (3.26), (3.28) and the definition of c_n it follows that, as $n \to \infty$

$$\begin{aligned} nI_{n,1} &\sim n\mathcal{L}(c_n y)(c_n y)^{-\alpha} \exp\{-y\, c_n[1-p_n(\mathbf{x})]\} \\ &\sim n\Gamma(1-\alpha)\mathcal{L}\left(\frac{\theta(G)y}{\bar{F}_*(m_n)\Gamma^{1/\alpha}(1-\alpha)}\right)\left(\frac{\theta(G)}{\bar{F}_*(m_n)}\right)^{-\alpha} \\ &\quad \times y^{-\alpha}\exp\left\{-y[-\ln G(\mathbf{x})]/\Gamma^{1/\alpha}(1-\alpha)\right\} \\ &\sim y^{-\alpha}\exp\left\{-y[-\ln G(\mathbf{x})]/\Gamma^{1/\alpha}(1-\alpha)\right\} =: y^{-\alpha}\pi(\mathbf{x},y). \end{aligned}$$

Furthermore, by uniform convergence (e.g. [60, Proposition 0.5]) we also obtain

$$\begin{aligned} nI_{n,2} &\sim -n\int_0^{\pi(\mathbf{x},y)} \bar{F}_N\left(\frac{\ln v}{\ln p_n(\mathbf{x})}\right)\mathrm{d}v \\ &\sim -n\bar{F}_N\left(\frac{1}{-\ln p_n(\mathbf{x})}\right)\int_0^{\pi(\mathbf{x},y)}(-\ln v)^{-\alpha}\mathrm{d}v \\ &\sim -\frac{(-\ln G(\mathbf{x}))^\alpha}{\Gamma(1-\alpha)}\int_{-\ln(\pi(\mathbf{x},y))}^{\infty} t^{-\alpha}\exp(-t)\mathrm{d}t \\ &= \frac{(-\ln G(\mathbf{x}))^\alpha}{\Gamma(1-\alpha)}\gamma\left(1-\alpha, \frac{y\,(-\ln G(\mathbf{x}))}{\Gamma^{1/\alpha}(1-\alpha)}\right) - (-\ln G(\mathbf{x}))^\alpha, \end{aligned}$$

as $n \to \infty$ and therefore the first part of (3.3) follows.

When $F_N \in \mathcal{D}(\Phi_\alpha)$ with $\alpha > 1$, by (3.29) we have that as $n \to \infty$ $p_n^{c_n y}(\mathbf{x}) \to 1$ and, in turn, $n\, I_{n,1} \to y^{-\alpha}$. Recall that in this case $u_n(y) = c_n y$, with $c_n/n = o(1)$. Furthermore, by Karamata's theorem (e.g. [60, p. 17]) and (3.29) we also obtain

$$\begin{aligned} n|I_{n,2}| &= -n\ln p_n(\mathbf{x})\int_{u_n(y)}^{\infty}\bar{F}_N(t)\exp(-t(-\ln p_n(\mathbf{x})))\mathrm{d}t \\ &\leq -n\ln p_n(\mathbf{x})\int_{u_n(y)}^{\infty}\bar{F}_N(t)\mathrm{d}t \\ &\sim -n\ln p_n(\mathbf{x})u_n(y)\bar{F}_N(u_n(y)), \quad n\to\infty \\ &\sim [-\ln G(\mathbf{x})/\mathbb{E}N]u_n(y)\bar{F}_N(u_n(y)), \quad n\to\infty. \end{aligned} \tag{3.30}$$

By point (v) in [60, p. 23],

$$u_n(y)\bar{F}_N(u_n(y)) = c_n y\bar{F}_N(c_n y) \sim y^{1-\alpha}c_n/n \to 0, \quad n\to\infty$$

then we have that $\lim_{n\to\infty} n|I_{n,2}| = 0$. The second part of (3.3) follows.

When $F_N \in \mathcal{D}(\Lambda)$, using Propositions 1.9, 0.10 and 0.16 plus Lemma 1.2 in [60] it can be shown that $u_n(y)/n \to 0$ as $n \to \infty$ - see Subsection 3.7.3 for details. Consequently, by (3.29) it holds that

$$\begin{aligned} nI_{n,1} &\sim \exp\{-y\} \exp\{u_n(y)[1 - p_n(\mathbf{x})]\} \\ &\sim \exp\{-y\} \exp\{u_n(y)[-\ln G(\mathbf{x})/\mathbb{E}N]/n\}, \quad n \to \infty \\ &\sim \exp\{-y\}, \end{aligned} \tag{3.31}$$

as $n \to \infty$. Furthermore, by (3.30)

$$\begin{aligned} n|I_{n,2}| &\leq -n \ln p_n(\mathbf{x}) \int_{u_n(y)}^{\infty} \bar{F}_N(t)\mathrm{d}t \\ &\sim [-\ln G(\mathbf{x})/\mathbb{E}N] \int_{u_n(y)}^{\infty} \bar{F}_N(t)\mathrm{d}t, \quad n \to \infty. \end{aligned} \tag{3.32}$$

Since $F_N \in \mathcal{D}(\Lambda)$, the term on the right hand side can be shown to converge to zero as $n \to +\infty$ (see Subsection 3.7.3), consequently $n|I_{n,2}| \to 0$. Equation (3.4) now follows and the proof is complete. ∎

3.6.2 Proof of Proposition 3.2.4

Let $\mathbf{c}_n$, $\mathbf{d}_n$, c_n be the norming sequences defined in Subsection 3.6.1 for the case $F_N \in \mathcal{D}(\Phi_\alpha)$, $\alpha \in (0,1)$. In particular $\mathbf{d}_n = \mathbf{0}$ and $n F_N(c_n) \sim 1$, whence it follows that as $n \to \infty$

$$\tilde{c}_n \sim F_N^{\leftarrow}\left(1 - \frac{1}{n\Gamma(1-\alpha)}\right).$$

The result at point (i) now follows from Theorem 5.4.2 in [72].

Denoting $p_n(\mathbf{x}) = F_{\mathbf{X}}(\mathbf{c}_n\mathbf{x} + \mathbf{d}_n)$, in the previous subsection it has been established that as $n \to \infty$

$$p_n^{c_n}(\mathbf{x}) \sim \exp\left\{-\frac{[-\ln G(\mathbf{x})]}{\Gamma^{1/\alpha}(1-\alpha)}\right\}.$$

Consequently, by dominated convergence

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left(\mathbf{M}_{\lfloor \tilde{c}_n S\rfloor} \le \mathbf{c}_n\mathbf{x}+\mathbf{d}_n\right) &= \lim_{n\to\infty}\int_0^\infty p_n^{\lfloor \tilde{c}_n s\rfloor}(\mathbf{x})\mathrm{d}F_S(s)\\
&= \lim_{n\to\infty}\int_0^\infty \left[p_n^{c_n}(\mathbf{x})\right]^{\lfloor \tilde{c}_n s\rfloor/c_n}\mathrm{d}F_S(s)\\
&= \int_0^\infty \left(e^{-\frac{[-\ln G(\mathbf{x})]}{\Gamma^{1/\alpha}(1-\alpha)}}\right)^{s\Gamma^{1/\alpha}(1-\alpha)}\mathrm{d}F_S(s)\\
&= \mathbb{E}G^S(\mathbf{x}) = L_S\left(-\ln G(\mathbf{x})\right) = G_\alpha(\mathbf{x}).
\end{aligned}$$

The result at point (ii) is now established. ∎

3.6.3 Proof of Proposition 3.3.1

As for the result at point (i), observe that by the max-stability of G, for every $s>0$ and $j=1,\dots,d$, there exist functions $\alpha_j(s)>0$, $\beta_j(s)$ such that $G^s(\mathbf{x}) = G(\alpha_1(s)x_1+\beta_1(s),\dots,\alpha_d(s)x_d+\beta_d(s))$. Therefore, by the dominated convergence theorem we can conclude

$$\begin{aligned}
&G_\alpha(\mathbf{x}) = \mathbb{E}G^S(\mathbf{x}) = \int_0^\infty G^s(\mathbf{x})\mathrm{d}F_S(s)\\
&= \lim_{n\to\infty}\int_0^\infty F_{\mathbf{X}}^n\left(\lambda_{n,1}(s)x_1+\mu_{n,1}(s),\dots,\lambda_{n,d}(s)x_d+\mu_{n,d}(s)\right)\mathrm{d}F_S(s)\\
&= \lim_{n\to\infty}\mathbb{P}\left(\frac{\mathbf{M}_n-\boldsymbol{\mu}_n(S)}{\boldsymbol{\lambda}_n(S)}\le \mathbf{x}\right),
\end{aligned}$$

where the maps $\boldsymbol{\lambda}_n$ and $\boldsymbol{\mu}_n$ are defined via

$$\begin{aligned}
\boldsymbol{\lambda}_n &: (0,\infty)\mapsto(0,\infty)^d : s\mapsto \left(\alpha_j(s)a_{n,j}\right)_{j=1}^d,\\
\boldsymbol{\mu}_n &: (0,\infty)\mapsto\mathbb{R}^d : s\mapsto \left(\beta_j(s)a_{n,j}+b_{n,j}\right)_{j=1}^d.
\end{aligned}$$

The result at point (ii) is a direct consequence. Indeed, by $F_{\mathbf{X}}\in\mathcal{D}(G)$ and Proposition 1.2.2, it holds that

$$\lim_{n\to\infty} F_*^n(n\,y_1,\dots,n\,y_d) = G_*(y_1,\dots,y_d).$$

Then the fact that $F_{S\mathbf{X}_*}\in\mathcal{D}(G_\alpha^*)$ follows from the above arguments, replacing G_α, G and $F_{\mathbf{X}}$ with G_α^*, G_* and F_*, repsectively, and setting $\boldsymbol{\lambda}_n(s)=s^{-1}n\mathbf{1}$ and

$\boldsymbol{\mu}_n(s) = \mathbf{0}$, for every $s > 0$ and $n \in \mathbb{N}_+$. Next, observe that for every $j = 1, \ldots, d$ and $u \in [0,1]$, the j-th marginal distribution of G^*_α satisfies

$$(G^*_{\alpha,j})^{\leftarrow}(u) = G^{\leftarrow}_{*,j}\left(e^{-(-\ln u)^{1/\alpha}}\right).$$

As a consequence, for every $\mathbf{u} \in [0,1]^d$ it holds that

$$\begin{aligned}
&C_{G^*_\alpha}(\mathbf{u}) = G^*_\alpha\left((G^*_{\alpha,1})^{\leftarrow}(u_1), \ldots, (G^*_{\alpha,1})^{\leftarrow}(u_d)\right) \\
&= \exp\left\{-\left[-\ln G_*\left(G^{\leftarrow}_{*,1}\left(e^{-(-\ln u_1)^{1/\alpha}}\right), \ldots, G^{\leftarrow}_{*,d}\left(e^{-(-\ln u_d)^{1/\alpha}}\right)\right)\right]^\alpha\right\} \\
&= \exp\left\{-\left[-\ln C_{G_*}\left(e^{-(-\ln u_1)^{1/\alpha}}, \ldots, e^{-(-\ln u_d)^{1/\alpha}}\right)\right]^\alpha\right\}.
\end{aligned}$$

Using the fact that $C_{G_*} = C_G$ and exploiting (1.3), we can now conclude that $C_{G^*_\alpha}(\mathbf{u})$ equals the expression in (3.6). The first half of the statement at point (ii) now follows. The second half of the statement follows from similar arguments. ∎

3.6.4 Proof of Proposition 3.3.2

Since G_α is max-stable, then its copula satisfies

$$C_{G_\alpha}(\mathbf{u}) = \exp\{-L_\alpha(-\ln u_1, \ldots, -\ln u_d)\}$$

for $\mathbf{u} \in [0,1]^d$, where L_α is the stable tail dependence function associated to G_α. By (3.6) it must be that for every $\mathbf{u} \in [0,1]^d$

$$L_\alpha(-\ln u_1, \ldots, -\ln u_d) = L^\alpha((-\ln u_1)^{1/\alpha}, \ldots, (-\ln u_d)^{1/\alpha}),$$

which is equivalent to the condition $L_\alpha(\mathbf{z}) = L^\alpha(\mathbf{z}^\alpha)$, for every $\mathbf{z} \in [0,\infty)^d$. Now, choosing $\mathbf{z} \in \mathcal{S}_d$ and exploiting the homogeneity of stable tail dependence functions we obtain

$$\begin{aligned}
A_\alpha(\mathbf{z}) = L_\alpha(\mathbf{z}) &= \left[\|\mathbf{z}^{1/\alpha}\|_1 L_\alpha\left(\mathbf{z}^{1/\alpha}/\|\mathbf{z}^{1/\alpha}\|_1\right)\right]^\alpha \\
&= \left[\|\mathbf{z}\|_{1/\alpha}^{1/\alpha} A_\alpha\left((\mathbf{z}/\|\mathbf{z}\|_{1/\alpha})^{1/\alpha}\right)\right]^\alpha \\
&= \|\mathbf{z}\|_{1/\alpha} A^\alpha\left((\mathbf{z}/\|\mathbf{z}\|_{1/\alpha})^{1/\alpha}\right),
\end{aligned}$$

which is the result. ∎

3.6.5 Proof of Theorem 3.4.2

This subsection is organized in three parts: notation, some preliminary results and the main body of the proof of Theorem 3.4.2. The derivations presented herein will rely on some auxiliary results, collected in Subsection 3.6.6.

Notation

Recall that $(\boldsymbol{\eta}_1, \xi_1), \dots, (\boldsymbol{\eta}_n, \xi_n)$, $n = 1, 2, \dots$, are iid random vectors with distribution Q in (3.3) with $\alpha \in (0,1)$. For $i = 1, 2, \dots$, and $j \in \{1, \dots, d\}$, let

$$V_i := \Phi_\alpha(\xi_i), \quad U_{i,j} := G_{\alpha,j}(\boldsymbol{\eta}_{i,j}), \quad \widehat{U}_{i,j} := G_{n,j}(\eta_{i,j}), \tag{3.33}$$

where Φ_α denotes the one-parameter Fréchet distribution, $G_{\alpha,j}$ denotes the j-th margin of the distribution G_α in the first line of (3.5) and $G_{n,j}$ is as in (3.17), respectively. Set $\mathbf{U}_i = (U_{i,1}, \dots, U_{i,d})$ and $\widehat{\mathbf{U}}_i = (\widehat{U}_{i,1}, \dots, \widehat{U}_{i,d})$. In the following we shall sometimes drop the index i and refer to a single observation.

For every $\mathbf{u} \in [0,1]^d$ and $v \in [0,1]$, define the copula functions

$$C_{Q,n}(\mathbf{u}, v) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(\mathbf{U}_i \leq \mathbf{u}, V_i \leq v) \tag{3.34}$$

and

$$C_{G_\alpha,n}(\mathbf{u}) = C_{Q,n}(\mathbf{u}, 1), \quad W_n(v) = C_{Q,n}(\mathbf{1}, v), \tag{3.35}$$

where $\mathbf{1} = (1, \dots, 1)$. Then, define the the copula processes

$$\mathbb{C}_{Q,n}(\mathbf{u}, v) := \sqrt{n}(C_{Q,n}(\mathbf{u}, v) - C_Q(\mathbf{u}, v)) \tag{3.36}$$

and

$$\mathbb{C}_{G_\alpha,n}(\mathbf{u}) = \mathbb{C}_{Q,n}(\mathbf{u}, 1), \quad \mathbb{W}_n(v) = \mathbb{C}_{Q,n}(\mathbf{1}, v). \tag{3.37}$$

Let $\mathbb{C}_{G_\alpha}(\mathbf{u}) := \mathbb{C}_Q(\mathbf{u}, 1)$, for $\mathbf{u} \in [0,1]^d$, and observe that $\mathbb{C}_{G_\alpha}$ has covariance function as in (3.23), with C_Q replaced by C_{G_α}. Next, for every $\mathbf{u} \in [0,1]^d$, define the empirical copula function and process

$$\widehat{C}_{G_\alpha,n}(\mathbf{u}) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(\widehat{\mathbf{U}}_i \leq \mathbf{u}), \quad \widehat{\mathbb{C}}_{G_\alpha,n} = \sqrt{n} \left(\widehat{C}_{G_\alpha,n} - C_{G_\alpha} \right). \tag{3.38}$$

According to [65, Proposition 3.1], under Condition 3.4.1(i) it holds that

$$\widehat{\mathbb{C}}_{G_\alpha,n}(\mathbf{u}) = \mathbb{C}_{G_\alpha,n} - \sum_{j=1}^{d} \dot{C}_{G_\alpha,j}(\mathbf{u})\mathbb{C}_{G_\alpha,n}(1,\dots,1,u_j,1,\dots,1) + R_n, \tag{3.39}$$

where $R_n = o_p(1)$, uniformly on $[0,1]^d$. The representation in (3.39) is originally due to Stute - see [67, 71]. If Condition 3.4.1(ii) is also satisfied, [65, Proposition 4.2] guarantees the following almost sure rate for the remainder term:

$$R_n = O(n^{-1/4}(\ln n)^{1/2}(\ln\ln n)^{1/4}), \quad n \to \infty. \tag{3.40}$$

Equations (3.39)–(3.40) play a pivotal role in the derivations presented later on.

For every $f \in \ell([0,1]^p)$, let $g_\epsilon : \ell([0,1]^p) \mapsto \ell([0,1]^p)$ be the weighting map defined by

$$(g_\epsilon(f))(\mathbf{z}) = \begin{cases} w_\epsilon^{-1}(\mathbf{z})f(\mathbf{z}), & \mathbf{z} \in (0,1]^p \setminus \{1,\dots,1\} \\ 0, & \text{otherwise} \end{cases} \tag{3.41}$$

where, for a fixed $\epsilon \in [0, 1/2)$, w_ϵ is the weighting function

$$w_\epsilon : [0,1]^p \mapsto [0,1] : \mathbf{z} \mapsto \min_{1\le j\le p} z_j^\epsilon \left(1 - \min_{1\le j\le p} z_j\right)^\epsilon .$$

In the following, we shall consider $p = d$ and $p = d + 1$ and, with a little abuse of notation, use the symbols g_ϵ and w_ϵ in both cases, the dimension of the underlying unit hypercube being clear from the context. Then, as $n \to \infty$ it holds that

$$g_\epsilon(\mathbb{C}_{Q,n}) \rightsquigarrow g_\epsilon(\mathbb{C}_Q), \quad g_\epsilon(\mathbb{C}_{G_\alpha,n}) \rightsquigarrow g_\epsilon(\mathbb{C}_{G_\alpha}). \tag{3.42}$$

Specifically, when $\epsilon = 0$ it holds that $w_\epsilon \equiv 1$, and the above convergence result follows from standard empirical process arguments. When $\epsilon \in (0, 1/2)$, weak convergence of the weighted copula process $g_\epsilon(\mathbb{C}_{\bullet,n})$ to the corresponding wighted $C_\bullet$-Brownian bridge $g_\epsilon(C_\bullet)$ is established in [28, 33]. As a direct consequence, for every composition $\phi \circ g_\epsilon$, with ϕ a continuous map, it will be possible to claim $\phi \circ g_\epsilon(\mathbb{C}_{\bullet,n}) \rightsquigarrow \phi \circ g_\epsilon(C_\bullet)$ as $n \to \infty$.

For $f \in \ell([0,1]^{d+1})$, let $S_{1,\dots,d} : \ell([0,1]^{d+1}) \mapsto \ell([0,1]^d)$ and $S_{d+1} : \ell([0,1]^{d+1}) \mapsto \ell([0,1])$ be the selection maps defined by

$$(S_{1,\dots,d}(f))(\mathbf{u}) := f(u_1,\dots,u_d,1), \quad (S_{d+1}(f)) := f(1,\dots,1,v). \tag{3.43}$$

For every $\alpha \in (0,1)$, $\epsilon \in [0,1/2)$ and $\mathbf{u} \in (0,1]^d \setminus \{1,\ldots,1\}$, let $\omega_{\epsilon,\mathbf{u}} : [0,1]^d \mapsto \mathbb{R}$ be the weighted function defined by

$$\omega_{\epsilon,\mathbf{u}}(\mathbf{v}) := \frac{\mathbb{1}(\mathbf{v} \leq \mathbf{u}) - C_{G_\alpha}(\mathbf{u})}{(S_{1,\ldots,d}(w_\epsilon))(\mathbf{u})}. \tag{3.44}$$

For every $\mathbf{v} \in [0,1]^d$ and $\mathbf{u} \in [0,1]^d$, set

$$\omega'_{\epsilon,\mathbf{u}}(\mathbf{v}) = \begin{cases} \omega_{\epsilon,\mathbf{u}}(\mathbf{v}), & \mathbf{u} \in (0,1]^d \setminus \{1,\ldots,1\} \\ 0, & \text{otherwise} \end{cases}. \tag{3.45}$$

Preliminary results

The results presented herein rely on the following conditions.

Condition 3.6.2 Let $\widehat{A}_{\alpha,n}$ be an estimator of A_α admitting the following representation for some $\epsilon \in [0,1/2)$

$$\sqrt{n}(\ln \widehat{A}_{\alpha,n} - \ln A_\alpha) = \phi_{g_\epsilon}(\mathbb{C}_{G_\alpha,n}) + o_p(1),$$

where $\phi : \ell^\infty([0,1]^d) \mapsto \ell^\infty(\mathcal{S}_d)$ is a continuous linear map, $\phi_{g_\epsilon} = \phi \circ g_\epsilon$, g_ϵ is as in (3.41) and $\mathbb{C}_{G_\alpha,n}$ is as in (3.36).

Condition 3.6.3 Let $\widehat{\alpha}_n$ be an estimator of α satisfying one of the following properties:

(i) there exists a continuous linear map $\tau : \ell^\infty([0,1]) \mapsto \mathbb{R}$ such that

$$\sqrt{n}(\widehat{\alpha}_n - \alpha) = \tau(\mathbb{W}_n) + o_p(1),$$

where $\mathbb{W}_n$ is as in (3.36);

(ii) there exists a measurable function $\zeta : (0,+\infty) \mapsto \mathbb{R}$ with $\Phi_\alpha \zeta = 0$ and $\Phi_\alpha \zeta^2 < \infty$, such that

$$\sqrt{n}(\widehat{\alpha}_n - \alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta(\xi_i) + o_p(1).$$

The following propositions establish two weak convergence results for the compositional estimator $\widehat{A^\star}_n$ in (3.21), obtained by combining an estimator of A_α that satisfies Condition 3.6.2 with an estimator of α that satisfies Condition 3.6.3(i) and 3.6.3(ii), respectively.

Proposition 3.6.4 *Let $\widehat{A^\star}_n$ be the estimator in* (3.21) *obtained by the composition of two estimators $\widehat{A}_{\alpha,n}$ and $\widehat{\alpha}_n$ satisfying Condition 3.6.2 and Condition 3.6.3(i), respectively. Then, in $\ell^\infty([0,1]^d)$*

$$\sqrt{n}(\widehat{A^\star}_n - A^\star) \rightsquigarrow A_\alpha \left\{\phi'_{g_\epsilon}(\mathbb{C}_Q) + K_\alpha \tau'_{g_\epsilon}(\mathbb{C}_Q)\right\}, \quad n \to \infty. \tag{3.46}$$

Specifically, $\phi'_{g_\epsilon} = \alpha^{-1}\phi \circ S_{1,\dots,d} \circ g_\epsilon$, $\tau'_{g_\epsilon} = \tau \circ S_{d+1} \circ (w_\epsilon g_\epsilon)$, where $S_{1,\dots,d}$ and S_{d+1} are as in (3.43)*, w_ϵ is as in* (3.41)*, and for any $\mathbf{t} \in \mathcal{S}_d$*

$$K_\alpha(\mathbf{t}) = \alpha^{-2}\left\{\|\mathbf{t}\|_{1/\alpha}^{-1/\alpha} \sum_{1\le j\le d: t_j>0} t_j^{1/\alpha} \ln t_j - \ln A_\alpha(\mathbf{t})\right\}. \tag{3.47}$$

Proof. For simplicity we focus on $\ln A^\star$ and $\ln \widehat{A^\star}_n$. Then, we obtain

$$\begin{aligned}\sqrt{n}(\ln \widehat{A^\star}_n - \ln A^\star) &= \sqrt{n}\left(\frac{1}{\widehat{\alpha}_n}\ln \widehat{A}_{\alpha,n} - \frac{1}{\alpha}\log A_\alpha\right) \\ &\quad - \sqrt{n}\left(\ln\|\cdot\|_{1/\widehat{\alpha}_n}^{1/\widehat{\alpha}_n} - \ln\|\cdot\|_{1/\alpha}^{1/\alpha}\right) \\ &=: T_{1,n} + T_{2,n}.\end{aligned}$$

By Condition 3.6.3(i), the functional version of Slutsky's lemma (e.g., [75, page 32]) and an application of ordinary delta method (e.g., [74, Theorem 3.1]) it holds that

$$T_{1,n} = \frac{1}{\alpha}\sqrt{n}(\ln \widehat{A}_{\alpha,n} - \ln A_\alpha) - \alpha^{-2}\ln A_\alpha \sqrt{n}(\widehat{\alpha}_n - \alpha) + o_p(1).$$

From Lemma 3.6.6 and an application of functional delta method ([74, Ch. 20]) it follows that

$$T_{2,n} = (K_\alpha + \alpha^{-2}\log A_\alpha)\sqrt{n}(\widehat{\alpha}_n - \alpha) + o_p(1)$$

where K_α is as in (3.47). Consequently, exploiting Condition 3.6.2 and 3.6.3(i), we bring $T_{1,n} + T_{2,n}$ into the form

$$\begin{aligned}T_{1,n} + T_{2,n} &= \frac{1}{\alpha}\phi_{g_\epsilon}(\mathbb{C}_{G_\alpha,n}) + K_\alpha\sqrt{n}(\widehat{\alpha}_n - \alpha) + o_p(1) \\ &= \frac{1}{\alpha}\phi_{g_\epsilon}(\mathbb{C}_{G_\alpha,n}) + K_\alpha \tau(\mathbb{W}_n) + o_p(1) \\ &= \phi'_{g_\epsilon}(\mathbb{C}_{Q,n}) + K_\alpha \tau'_{g_\epsilon}(\mathbb{C}_{Q,n}) + o_p(1).\end{aligned} \tag{3.48}$$

Applying the continuous mapping theorem and the functional delta method in the last line of (3.48), we finally obtain the result in (3.46). ∎

For the next result, we shall refine Condition 3.6.2 by assuming that the map ϕ admits the following representation for every $f \in \ell^\infty([0,1]^d)$ and $\mathbf{t} \in \mathcal{S}_d$

$$(\phi(f))(\mathbf{t}) = \sum_{i=0}^{m} \int_a^b f(B_{i,1}(z;t_1), \ldots, B_{i,d}(z;t_d)) K_i(z;\mathbf{t}) \mathrm{d}z, \qquad (3.49)$$

where $-\infty \leq a < b \leq \infty$, $m \in \mathbb{N}_+$, for $i = 0, \ldots, m$, $j = 1, \ldots, d$ and $\mathbf{t} \in \mathcal{S}_d$ the functions $z \mapsto B_{i,j}(z;t_j)$ are bijective and continuous, and the functions K_i satisfy

$$\sup_{\mathbf{t} \in \mathcal{S}_d} \max_{0 \leq i \leq m} |K_i(z;\mathbf{t})| \leq K(z), \quad z \in (a,b),$$

for some integrable function K.

Proposition 3.6.5 *Let $\widehat{A^\star}_n$ be the estimator in* (3.21) *obtained by the composition of two estimators $\widehat{A}_{\alpha,n}$ and $\widehat{\alpha}_n$ satisfying Condition 3.6.2 and Condition 3.6.3(ii), respectively, with ϕ admitting the representation in* (3.49) *and $\varphi = \zeta \circ \Phi_\alpha^\leftarrow$ satisfying*

$$-\infty < \mathbb{E}\{\omega'_{\epsilon,\mathbf{u}}(\mathbf{U})\varphi(V)\} < \infty \qquad (3.50)$$

with $\omega'_{\epsilon,\mathbf{u}}$ as in (3.45)*. Then, in $\ell^\infty(\mathcal{S}_d)$ as $n \to \infty$*

$$\sqrt{n}\left(\widehat{A^\star}_n - A^\star\right) \rightsquigarrow A_\alpha \phi''_{g'_{\epsilon,\varphi}}(\mathbb{C}_Q).$$

Specifically, $\phi''_{g'_{\epsilon,\varphi}} = \phi'' \circ g'_{\epsilon,\varphi}$, where $\phi'' : \ell^\infty([0,1]^d) \mapsto \ell^\infty(\mathcal{S}_d) : f \mapsto \alpha^{-1}\phi(f) + K_{m+1} f(1, \ldots, 1)$,

$$K_{m+1}(\mathbf{t}) = K_\alpha(\mathbf{t}) - \frac{1}{\alpha} \sum_{i=0}^{m} \int_a^b K_i(z;\mathbf{t}) \mathrm{d}z, \quad \mathbf{t} \in \mathcal{S}_d$$

and $g'_{\epsilon,\varphi}(\mathbb{C}_Q)$ is a zero-mean Gaussian process with covariance function defined in (3.51).

Proof. For any $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \subset [0,1]^d$, the random vectors

$$(\omega'_{\epsilon,\mathbf{u}_1}(\mathbf{U}_i), \ldots, \omega'_{\epsilon,\mathbf{u}_k}(\mathbf{U}_i), \varphi(V_i)),$$

$i = 1, \ldots, n$, are iid with zero-mean and covariance matrix $\Sigma_{\mathbf{u}}$ with finite entries, by arguments in [28, 33], Condition 3.6.3(ii) and (3.50). For $\mathbf{u} \in [0,1]^d$, let

$$g'_\epsilon(\mathbb{C}_{Q,n})(\mathbf{u}) = n^{-1/2} \sum_{i=1}^{n} \omega'_{\epsilon,\mathbf{u}}(\mathbf{U}_i), \quad \bar{\varphi}_n(\mathbf{u}) = \delta(\mathbf{u}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(V_i),$$

where $\delta(\mathbf{u}) = 1$. Note that $g'_\epsilon(\mathbb{C}_{Q,n})(\mathbf{u}) = g_\epsilon(\mathbb{C}_{Q,n})(\mathbf{u})$. Both $g'_\epsilon(\mathbb{C}_{Q,n})$ and $\bar{\varphi}_n$ are asymptotically tight [75, Definition 1.3.7] and by the central limit theorem we have that

$$(g'_\epsilon(\mathbb{C}_{Q,n})(\mathbf{u}_1), \ldots, g'_\epsilon(\mathbb{C}_{Q,n})(\mathbf{u}_k), \bar{\varphi}_n(\mathbf{u}_1), \ldots, \bar{\varphi}_n(\mathbf{u}_k)) \rightsquigarrow N(\mathbf{0}, \Sigma_{\mathbf{u}})$$

as $n \to \infty$. Arguing as in [75, p. 42 point 3], we can now claim that the class of functions $\mathcal{Y}_{\epsilon,\varphi} := \{(\mathbf{a}, b) \mapsto \omega'_{\epsilon,\mathbf{u}}(\mathbf{a}) + \delta(\mathbf{u})\varphi(b) : \mathbf{u} \in [0,1]^d\}$ is C_Q-Donsker [75, pp. 80-82]. Indeed, introducing the map

$$g'_{\epsilon,\varphi} : M \mapsto \{Mf : f \in \mathcal{Y}_{\epsilon,\varphi}\}$$

defined on the space of signed measure M on $[0,1]^{d+1}$, we have that $g'_{\epsilon,\varphi}(\mathbb{C}_{Q,n})(\omega'_{\epsilon,\mathbf{u}} + \delta(\mathbf{u})\varphi) = g'_\epsilon(\mathbb{C}_{Q,n})(\mathbf{u}) + \bar{\varphi}_n(\mathbf{u})$, $\forall\, \mathbf{u} \in [0,1]^d$. Then, as $n \to \infty$, $g'_{\epsilon,\varphi}(\mathbb{C}_{Q,n}) \rightsquigarrow g'_{\epsilon,\varphi}(\mathbb{C}_Q)$ in $\ell^\infty(\mathcal{Y}_{\epsilon,\varphi})$, where $g'_{\epsilon,\varphi}(\mathbb{C}_Q)$ is a zero-mean Gaussian process with covariance function

$$\begin{aligned}
&\operatorname{Cov}\left\{g'_{\epsilon,\varphi}(\mathbb{C}_Q)(\omega'_{\epsilon,\mathbf{u}} + \delta(\mathbf{u})\varphi), g'_{\epsilon,\varphi}(\mathbb{C}_Q)(\omega'_{\epsilon,\mathbf{v}} + \delta(\mathbf{v})\varphi)\right\} \\
&\quad = \begin{cases} \mathrm{E}(\{\omega_{\epsilon,\mathbf{u}}(\mathbf{U}) + \varphi(V)\}\{\omega_{\epsilon,\mathbf{v}}(\mathbf{U}) + \varphi(V)\}), & \mathbf{u}, \mathbf{v} \in \mathcal{V} \\ \mathrm{E}(\{\omega_{\epsilon,\mathbf{u}}(\mathbf{U}) + \varphi(V)\}\varphi(V)), & \mathbf{u} \in \mathcal{V}, \mathbf{v} \in \mathcal{V}^c \\ \mathrm{E}(\varphi^2(V)), & \mathbf{u}, \mathbf{v} \in \mathcal{V}^c \end{cases}
\end{aligned} \tag{3.51}$$

$\mathcal{V} = (0,1]^d \backslash \{1, \ldots, 1\}$. Since each element of $\mathcal{Y}_{\epsilon,\varphi}$ corresponds to a unique $\mathbf{u} \in [0,1]^d$, we can consider the processes $g'_{\epsilon,\varphi}(\mathbb{C}_{Q,n})$, $g'_{\epsilon,\varphi}(\mathbb{C}_Q)$ as indexed on the latter set. By (3.50), Condition 3.6.3(ii) and the first line of (3.48) it follows that

$$\sqrt{n}(\widehat{A^\star}_n - A^\star) = \phi'' \circ g'_{\epsilon,\varphi}(\mathbb{C}_{Q,n}) + o_p(1).$$

The final result follows from sequentially applying the continuous mapping theorem and the functional delta method. ∎

Main body of the proof

We start analyzing the case in which α is estimated with the ML estimator in (3.15) and A_α with the MD estimator in (3.19).

We recall that the estimator in (3.15) is the unique solution of the log-likelihood equation: $n^{-1}\sum_{i=1}^n \dot{\ell}_{\alpha'}(\xi_i) = 0$, where for all $x > 0$

$$\dot{\ell}_{\alpha'}(x) = \partial/\partial\alpha' \ln \dot{\Phi}_{\alpha'}(x) = 1/\alpha' + \ln x(x^{-\alpha'} - 1)$$

denotes the score function of the one-parameter Fréchet family of distributions $\{\Phi_{\alpha'} : \alpha' \in (0, +\infty)\}$. Noting that ξ^{-1} is a Weibull random variable with unit scale parameter and shape parameter equal to α, we can resort to Theorems 5.41-5.42 in [74] and arguments thereof to claim that $\widehat{\alpha}_n^{\text{ML}} \xrightarrow{\text{p}} \alpha$ as $n \to \infty$ and

$$\sqrt{n}\,(\widehat{\alpha}_n^{\text{ML}} - \alpha) = \frac{1}{i_\alpha}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \dot{\ell}_\alpha(\xi_i) + o_p(1), \tag{3.52}$$

where $i_\alpha = \alpha^{-2}\{(1-\varsigma)^2 + \pi^2/6\}$ is the Fisher information at α and ς is the Euler's constant. Under Condition 3.4.1(i), $\widehat{A}_{\alpha,n}^{\text{MD}}$ satisfies Condition 3.6.2 by Lemma 3.6.7, with $\phi = \phi_{\text{MD}}$ where ϕ_{MD} is in (3.55). Furthermore, by (3.52), $\widehat{\alpha}_n^{\text{ML}}$ satisfies Condition 3.6.3(ii), with $\zeta = \zeta_{\text{ML}} = i_\alpha^{-1}\dot{\ell}_\alpha$. For $v \in (0,1)$, define

$$\varphi_{\text{ML}}(v) := \zeta_{\text{ML}} \circ \Phi_\alpha^{\leftarrow}(v) = i_\alpha^{-1}\alpha^{-1}\{1 + (1+\ln v)\ln(-\ln v)\} \tag{3.53}$$

then (3.50) is satisfied with $\varphi = \varphi_{\text{ML}}$, by Lemma 3.6.9. Therefore, from Proposition 3.6.5 it follows that in $\ell^\infty([0,1]^d)$,

$$\sqrt{n}\left(\widehat{A^\star}_n^{\text{MD,ML}} - A^\star\right) \rightsquigarrow \phi_{\text{MD,ML}}(\mathbb{C}_Q), \quad n \to \infty.$$

Specifically, $\phi_{\text{MD,ML}} = A_\alpha \phi'' \circ g'_{0,\varphi_{\text{ML}}}$, with $(\phi''(f))(\mathbf{t}) = \alpha^{-1}(\phi_{\text{MD}}(f))(\mathbf{t}) + K_{d+1}^{MD}(\mathbf{t})f(1,\ldots,1)$ for every $f \in \ell^\infty([0,1]^d)$ and $\mathbf{t} \in \mathcal{S}_d$, where ϕ_{MD} is given in (3.55), K_{d+1}^{MD} is defined via

$$K_{d+1}^{MD}(\mathbf{t}) = K_\alpha(\mathbf{t}) - \frac{\{1 + A_\alpha(\mathbf{t})\}^2}{\alpha A_\alpha(\mathbf{t})}\left(\sum_{j=1}^{d}\int_0^1 \dot{C}_{G_\alpha,j}(v^{t_1},\ldots,v^{t_d})dv - 1\right)$$

and K_α is as in (3.47). In particular, observe that the covariance of the zero-mean Gaussian process $g'_{0,\varphi_{\text{ML}}}(\mathbb{C}_Q)$ reduces to

$$\begin{aligned}&\text{Cov}\{g'_{0,\varphi_{\text{ML}}}(\mathbb{C}_Q)(\mathbf{u}), g'_{0,\varphi_{\text{ML}}}(\mathbb{C}_Q)(\mathbf{v})\}\\ &\quad = C_{G_\alpha}(\min(\mathbf{u},\mathbf{v})) - C_{G_\alpha}(\mathbf{u})C_{G_\alpha}(\mathbf{v}) + T_\alpha(\mathbf{u}) + T_\alpha(\mathbf{v}) + 1,\end{aligned}$$

for every $\mathbf{u},\mathbf{v} \in [0,1]^d$, with

$$T_\alpha(\cdot) = \frac{1}{i_\alpha \alpha}\left(C_{G_\alpha}(\mathbf{u}) - \int_0^1 \frac{\partial}{\partial v}C_Q(\mathbf{u},v)(1+\ln v)\ln(-\ln v)\mathrm{d}v\right).$$

Finally, from the above weak convergence result, it follows that that $\|\widehat{A^\star}_n^{\text{MD,ML}} - A^\star\|_\infty \xrightarrow{\text{p}} 0$ as $n \to \infty$, by the functional version of Slutsky's lemma.

Next, we study the case in which α is estimated with the GPWM estimator in (3.13) and A_α with the MD estimator in (3.19). Herein, we additionally assume that $\alpha > 1/(k-1)$. By Lemma 3.6.10, the estimator $\widehat{\alpha}_n^{\text{GPWM}}$ satisfies Condition 3.6.3(i) with $\tau = \tau_{\text{GPWM}}$ given in (3.59). Then, from Proposition 3.6.4 it follows that in $\ell^\infty([0,1]^d)$

$$\sqrt{n}\left(\widehat{A^\star}_n^{\text{MD,GPWM}} - A^\star\right) \rightsquigarrow \phi_{\text{MD,GPWM}}(\mathbb{C}_Q), \quad n \to \infty,$$

where $\phi_{\text{MD,GPWM}}(f) = A_\alpha\{\phi'_{g_0}(f) + K_\alpha \tau'_{g_0}(f)\}$, with $\phi = \phi_{\text{MD}}$ given in (3.55), K_α in (3.47) and τ'_{g_0} defined via $\tau = \tau_{\text{GPWM}}$ in (3.59). Furthermore, by Lemma 3.6.10 and the fact that $n^{-1/2}\|\mathbb{W}_n\|_\infty \xrightarrow{\text{as}} 0$, we have that $\widehat{\alpha}_n^{\text{GPWM}} \xrightarrow{\text{as}} \alpha$ as $n \to \infty$. Therefore, from Lemma 3.6.8 it follows that

$$\left\|\widehat{A^\star}_n^{\text{MD,GPWM}} - A^\star\right\|_\infty \xrightarrow{\text{as}} 0, \quad n \to \infty.$$

Henceforward, it is assumed that both Conditions 3.4.1(i)-(ii) are satisfied. We first consider the case in which α is estimated with the ML estimator in (3.15) and A_α with the P and CFG estimator in (3.16) and (3.18), respectively. Both (3.39) and (3.40) hold true, then using arguments in [33, pp. 3082–3083] plus functional delta method (only for the case of the P estimator), we have that $\widehat{A}_{\alpha,n}^{\text{P}}$ and $\widehat{A}_{\alpha,n}^{\text{CFG}}$ satisfy Condition 3.6.2 with $\phi = \phi_{\text{P}}$ and $\phi = \phi_{\text{CFG}}$. Specifically, for any fixed $\epsilon \in (0, 1/2)$

$$\sqrt{n}(\ln \widehat{A}_{\alpha,n}^{\circ} - \ln A_\alpha) = \phi_{\circ,g_\epsilon}(\mathbb{C}_{G_\alpha,n}) + o_p(1),$$

where $\phi_{\circ,g_\epsilon} = \phi_\circ \circ g_\epsilon$ and $\phi_\circ : \ell^\infty([0,1]^d) \to \ell^\infty(\mathcal{S}_d)$ is defined via

$$\begin{aligned}
&(\phi_\circ(f))(\mathbf{t}) \\
&= \int_0^\infty f(e^{-vt_1}, \ldots, e^{-vt_d}) B_\epsilon(e^{-v\max(\mathbf{t})}) h_\circ(\mathbf{t}; v)\mathrm{d}v \\
&- \sum_{j=1}^d \int_0^\infty \dot{C}_{G_\alpha,j}(v^{t_1}, \ldots, v^{t_d}) f(1, \ldots, 1, v^{t_j}, 1, \ldots, 1) B_\epsilon(e^{-vt_j}) h_\circ(\mathbf{t}; v)\mathrm{d}v,
\end{aligned} \tag{3.54}$$

with $B_\epsilon(v) = v^\epsilon(1-v)^\epsilon$ for $v \in (0,1)$, $h_{\text{P}}(\mathbf{t}; v) = -A_\alpha^{-1}(\mathbf{t})$ and $h_{\text{CFG}}(\mathbf{t}; v) = 1/v$ for $v > 0$, $\mathbf{t} \in \mathcal{S}_d$. The map $\phi_\circ$ admits the representation in (3.49). Moreover, by Lemma 3.6.9, $\varphi = \varphi_{\text{ML}}$ satisfies the moment condition in (3.50) for any $\epsilon \in (0, 1/2)$. Then, by Proposition 3.6.5 we have that in $\ell^\infty([0,1]^d)$,

$$\sqrt{n}\left(\widehat{A^\star}_n^{\circ,\text{ML}} - A^\star\right) \rightsquigarrow \phi_{\circ,\text{ML}}(\mathbb{C}_Q), \quad n \to \infty.$$

Precisely, $\phi_{\circ,\mathrm{ML}} = A_\alpha \phi''_\circ \circ g'_{\epsilon,\varphi_{\mathrm{ML}}}$, where, for every $f \in \ell^\infty([0,1])$ and $\mathbf{t} \in \mathcal{S}_d$, $(\phi''_\circ(f))(\mathbf{t}) = \alpha^{-1}(\phi_\circ(f))(\mathbf{t}) + K^\circ_{d+1}(\mathbf{t})f(1,\ldots,1)$, with $\phi_\circ$ as in (3.54) and

$$K^\circ_{d+1}(\mathbf{t}) = K_\alpha(\mathbf{t}) - \frac{1}{\alpha}\int_0^\infty B_\epsilon(e^{-v\max(\mathbf{t})})h_\circ(\mathbf{t};v)\mathrm{d}v + \frac{1}{\alpha}\sum_{j=1}^d \int_0^\infty \dot{C}_{G_\alpha,j}(v^{t_1},\ldots,v^{t_d})B_\epsilon(e^{-vt_j})h_\circ(\mathbf{t};v)\mathrm{d}v.$$

From this result and the functional version of Slutsky's lemma, it follows that $\|\widehat{A^\star}_n^{\mathrm{P,ML}} - A^\star\|_\infty \xrightarrow{\mathrm{p}} 0$ and $\|\widehat{A^\star}_n^{\mathrm{CFG,ML}} - A^\star\|_\infty \xrightarrow{\mathrm{p}} 0$ as $n \to \infty$.

Concluding, we study the case in which α is estimated with the GPWM estimator in (3.13) and A_α with the P and CFG estimators in (3.16) and (3.18). Assuming in addition to the previous case that $\alpha > 1/(k-1)$, from Proposition 3.6.4 it follows that in $\ell^\infty([0,1]^d)$

$$\sqrt{n}\left(\widehat{A^\star}_n^{\circ,\mathrm{GPWM}} - A^\star\right) \rightsquigarrow \phi_{\circ,\mathrm{GPWM}}(\mathbb{C}_Q), \quad n \to \infty,$$

where for every $f \in \ell^\infty([0,1]^{d+1})$, $\mathbf{t} \in \mathcal{S}_d$ and any fixed $\epsilon \in (0,1/2)$ we have $\phi_{\circ,\mathrm{GPWM}}(f) = A_\alpha\{\phi'_{g_\epsilon}(f) + K_\alpha \tau'_{g_\epsilon}(f)\}$, with ϕ'_{g_ϵ} now obtained via $\phi = \phi_\circ$ in (3.54), K_α in (3.47) and $\tau = \tau_{\mathrm{GPWM}}$ in (3.59). Ultimately, from this result and the functional version of Slutsky's lemma, it follows that that $\|\widehat{A^\star}_n^{\mathrm{P,GPWM}} - A^\star\|_\infty \xrightarrow{\mathrm{p}} 0$ and $\|\widehat{A^\star}_n^{\mathrm{CFG,GPWM}} - A^\star\|_\infty \xrightarrow{\mathrm{p}} 0$ as $n \to \infty$.

3.6.6 Auxiliary results

Hadamard differentiability

Lemma 3.6.6 *For $\|\mathbf{t}\|_{1/\alpha}$ in (3.9), the map*

$$g : (0,\infty) \mapsto \ell^\infty(\mathcal{S}_d) : h \mapsto \left(\ln\|\mathbf{t}\|_{1/h}^{1/h}\right)_{\mathbf{t}\in\mathcal{S}_d}$$

is Hadamard differentiable at α with derivative

$$\{(\dot{g}_\alpha(h))(\mathbf{t})\}_{\mathbf{t}\in\mathcal{S}_d} = \left\{-h\alpha^{-2}\|\mathbf{t}\|_{1/\alpha}^{-1/\alpha}\sum_{1\le j\le d: t_j>0} t_j^{1/\alpha}\ln t_j\right\}_{\mathbf{t}\in\mathcal{S}_d}, \quad 0 < h < \infty.$$

Proof. Consider any $\mathbf{t} \in \mathcal{S}_d$ and sequences $\{r_n\}$, $\{h_n\}$, such that $r_n \downarrow 0$, $h_n \to h > 0$, and $r_h h_n + \alpha > 0$ for every n. Fix a small $\epsilon > 0$ such that $\alpha + \epsilon < 1$. For

n sufficiently large it holds that $0 < r_n h_n < \epsilon$ and a Taylor expansion at α with Lagrange reminder yields

$$\sup_{\mathbf{t}\in\mathcal{S}_d}\left|\frac{1}{r_n}\left[\ln\left(\|\mathbf{t}\|_{1/(\alpha+r_nh_n)}^{1/(\alpha+r_nh_n)}\right)-\ln\left(\|\mathbf{t}\|_{1/\alpha}^{1/\alpha}\right)\right]-\frac{\sum_{j\in J}t_j^{1/\alpha}\ln t_j}{\alpha^2\|\mathbf{t}\|_{1/\alpha}^{1/\alpha}}(-h)\right|$$
$$=\left\|\sum_{i=1}^{3}T_{n,i}\right\|_\infty$$

where, for every $\mathbf{t}\in\mathcal{S}_d$, $J\equiv J(\mathbf{t})=\{j\in(1,\dots,d):t_j>0\}$ and

$$T_{n,1}(\mathbf{t})=\frac{\sum_{j\in J}t_j^{1/\alpha}\ln t_j}{\alpha^2\|\mathbf{t}\|_{1/\alpha}^{1/\alpha}}(h-h_n),\quad T_{n,2}(\mathbf{t})=-\frac{\left(\sum_{j\in J}t_j^{1/\alpha_*}\ln t_j\right)^2}{2\alpha_*^4\|\mathbf{t}\|_{1/\alpha_*}^{2/\alpha_*}}r_nh_n^2,$$

$$T_{n,3}(\mathbf{t})=\frac{1}{2\|\mathbf{t}\|_{1/\alpha_*}^{1/\alpha_*}}\left[2\alpha_*^{-3}\sum_{j\in J}t_j^{1/\alpha_*}\ln t_j+\alpha_*^{-4}\sum_{j\in J}t_j^{1/\alpha_*}\ln^2 t_j\right]r_nh_n^2$$

with $\alpha_*\in(\alpha,\alpha+r_nh_n)\subset(\alpha,\alpha+\epsilon)$. Observe that for any $a,b>0$, the map $w:(0,1)\mapsto\mathbb{R}:t\mapsto t^a\ln^b t$ is contiuous and vanishes on the boundary of $(0,1)$. Then, there exists a positive constant $\kappa_{a,b}$ such that $\|w\|_\infty<\kappa_{a,b}$. Consequently, it holds that

$$\|T_{n,1}\|_\infty\le\alpha^{-2}d^{1/\alpha}\kappa_{1/\alpha,1}|h-h_n|\to 0,$$
$$\|T_{n,2}\|_\infty\le\alpha^{-4}d^{2(1/\alpha-1)}\left(d\kappa^2_{1/(\alpha+\epsilon),1}+d(d-1)\kappa^2_{1/(\alpha+\epsilon),1}\right)r_nh_n^2\to 0,$$
$$\|T_{n,3}\|_\infty\le d^{1/\alpha-1}\left(2\alpha^{-3}d\kappa_{1/(\alpha+\epsilon),1}+\alpha^{-4}d\kappa_{1/(\alpha+\epsilon),2}\right)r_nh_n^2\to 0,$$

as $n\to\infty$. The result now follows by triangular inequality. ∎

Madogram estimator of A_α

Lemma 3.6.7 *Under Condition 3.4.1(i) we have*

$$\sqrt{n}\left(\ln\widehat{A}^{\mathrm{MD}}_{\alpha,n}-\ln A_\alpha\right)=\phi_{\mathrm{MD}}(\mathbb{C}_{G_\alpha,n})+o_p(1),$$

where $\phi_{\mathrm{MD}}:\ell^\infty([0,1]^d)\mapsto\ell^\infty(\mathcal{S}_d)$ *is defined, for every* $\mathbf{t}\in\mathcal{S}_d$, *by*

$$(\phi_{\mathrm{MD}}(f))(\mathbf{t})=\frac{\{1+A_\alpha(\mathbf{t})\}^2}{A_\alpha(\mathbf{t})}\left(-\int_0^1 f(v^{t_1},\dots,v^{t_d})\mathrm{d}v\right.$$
$$\left.+\sum_{j=1}^{d}\int_0^1\dot{C}_{\alpha,j}(v^{t_1},\dots,v^{t_d})f(1,\dots,1,v^{t_j},1,\dots,1)\mathrm{d}v\right).\tag{3.55}$$

Proof. Let

$$\nu_{G_\alpha}(\mathbf{t}) = \mathrm{E}\left(\max_{1\le j\le d} U_j^{1/t_j} - \frac{1}{d}\sum_{j=1}^d U_j^{1/t_j}\right)$$

be the madogram function [46], where $u^{1/0} = 0$ for $0 < u < 1$ by convention, $\mathbf{U} = (G_{\alpha,1}(\eta_1), \dots, G_{\alpha,d}(\eta_d))$ and $\boldsymbol{\eta}$ is distributed according to G_α. For $\mathbf{u} \in [0,1]^d$, define

$$\widetilde{\mathbb{C}}_{G_\alpha,n}(\mathbf{u}) = \mathbb{C}_{G_\alpha,n}(\mathbf{u}) - \sum_{j=1}^d \dot{C}_{G_\alpha,j}(\mathbf{u})\mathbb{C}_{G_\alpha,n}(1,\dots,1,u_j,1,\dots,1),$$

where $\mathbb{C}_{G_\alpha,n}$ is as in (3.37) and $\dot{C}_{G_\alpha,j}(\mathbf{u}) := \partial/\partial u_j C_{G_\alpha}(\mathbf{u})$.

Then, by [46] we have that

$$\begin{aligned}\sqrt{n}(\widehat{\nu}_n - \nu_{G_\alpha}) = \Bigg(\frac{1}{d}\sum_{j=1}^d \int_0^1 \widehat{\mathbb{C}}_{G_\alpha,n}(1,\dots,1,v^{t_j}1,\dots,1)\mathrm{d}x \\ - \int_0^1 \widehat{\mathbb{C}}_{G_\alpha,n}(v^{t_1},\dots,v^{t_d})\mathrm{d}v\Bigg)_{\mathbf{t}\in\mathcal{S}_d},\end{aligned}$$

where $\widehat{\nu}_n$ and $\widehat{\mathbb{C}}_{G_\alpha,n}$ are as in (3.20) and (3.38), respectively. Under Condition 3.6.3(i), Stute's representation in (3.39) holds true and we have that $\|\widehat{\mathbb{C}}_{G_\alpha,n} - \widetilde{\mathbb{C}}_{G_\alpha,n}\|_\infty = o_p(1)$. Consequently,

$$\begin{aligned}\sqrt{n}(\widehat{\nu}_n - \nu_{G_\alpha}) &= \Bigg(\frac{1}{d}\sum_{j=1}^d \int_0^1 \widetilde{\mathbb{C}}_{G_\alpha,n}(1,\dots,1,x^{t_j}1,\dots,1)\mathrm{d}x \\ &\quad - \int_0^1 \widetilde{\mathbb{C}}_{G_\alpha,n}(x^{t_1},\dots,x^{t_d})\mathrm{d}x\Bigg)_{\mathbf{t}\in\mathcal{S}_d} + o_p(1) \\ &= \left(-\int_0^1 \widetilde{\mathbb{C}}_{G_\alpha,n}(x^{t_1},\dots,x^{t_d})\mathrm{d}x\right)_{\mathbf{t}\in\mathcal{S}_d} + o_p(1),\end{aligned}$$

where the second line follows from the fact that $\widetilde{\mathbb{C}}_{G_\alpha,n}(1,\dots,1,v^{t_j},1,\dots,1) = 0$ for $j = 1,\dots,d$, $v \in (0,1)$ and $\mathbf{t} \in \mathcal{S}_d$. Applying the functional delta method and resorting to [46, Proposition 2.2], we obtain

$$\begin{aligned}\sqrt{n}(\widehat{A}^{\mathrm{MD}}_{\alpha,n} - A_\alpha) &= (1+A_\alpha)^2\left(-\int_0^1 \widetilde{\mathbb{C}}_{G_\alpha,n}(x^{t_1},\dots,x^{t_d})\mathrm{d}x\right)_{\mathbf{t}\in\mathcal{S}_d} + o_p(1) \\ &= A_\alpha(\phi_{\mathrm{MD}}(\mathbb{C}_{G_\alpha,n})) + o_p(1).\end{aligned} \tag{3.56}$$

A further application of functional delta method now yields the result. ∎

Lemma 3.6.8 *Let $\widehat{\alpha}_n$ be an estimator of α satisfying $\widehat{\alpha}_n \xrightarrow{as} \alpha$ as $n \to \infty$ and $\widehat{A}^{\mathrm{MD}}_{\alpha,n}$ be the estimator of A_α given in* (3.19). *Let*

$$\widehat{A^{\star}}_n(\mathbf{t}) = \left(\widehat{A}^{\mathrm{MD}}_{\alpha,n}(\mathbf{t})/\|\mathbf{t}\|_{1/\widehat{\alpha}_n}\right)^{1/\widehat{\alpha}_n}, \quad \mathbf{t} \in \mathcal{S}_d.$$

Then,

$$\left\|\widehat{A^{\star}}_n - A^{\star}\right\|_\infty \xrightarrow{as} 0, \quad n \to \infty. \tag{3.57}$$

Proof. For simplicity we show

$$\| \ln \widehat{A^{\star}}_n - \ln A^{\star}\|_\infty \xrightarrow{\text{as}} 0, \quad n \to \infty. \tag{3.58}$$

The final result will then follows by a Lipschitz continuity argument.

Note that

$$\begin{aligned}
&\| \ln \widehat{A^{\star}}_n - \ln A^{\star}\|_\infty \\
= \; & \sup_{\mathbf{t}\in\mathcal{S}_d} \left| \widehat{\alpha}_n^{-1} \ln \widehat{A}^{\mathrm{MD}}_{\alpha,n}(\mathbf{t}) - \ln \|\mathbf{t}\|^{1/\widehat{\alpha}_n}_{1/\widehat{\alpha}_n} - \alpha^{-1}\ln A_\alpha(\mathbf{t}) + \ln\|\mathbf{t}\|^{1/\alpha}_{1/\alpha} \right| \\
\leq \; & \frac{1}{\widehat{\alpha}_n} \| \ln \widehat{A}^{\mathrm{MD}}_{\alpha,n} - \ln A_\alpha\|_\infty + \|\ln A_\alpha\|_\infty \left| \frac{1}{\widehat{\alpha}_n} - \frac{1}{\alpha}\right| \\
& + \sup_{\mathbf{t}\in\mathcal{S}_d} \left| \ln\|\mathbf{t}\|^{1/\widehat{\alpha}_n}_{1/\widehat{\alpha}_n} - \ln\|\mathbf{t}\|^{1/\alpha}_{1/\alpha}\right| \\
\equiv \; & I_{n,1} + I_{n,2} + I_{n,3}.
\end{aligned}$$

By Theorem 2.4 in [46] we have $\|\widehat{A}^{\mathrm{MD}}_{\alpha,n} - A_\alpha\|_\infty \xrightarrow{\text{as}} 0$. For n sufficiently large and any $\varepsilon > 0$ we have that, almost surely, $d^{-1} - \varepsilon \leq \widehat{A}^{\mathrm{MD}}_{\alpha,n} \leq 1 + \varepsilon$. From the fact that, by assumption, $\widehat{\alpha}_n \xrightarrow{\text{as}} \alpha$ and from the Lipschitz continuity of the map $x \mapsto \ln x$ on $[d^{-1} - \varepsilon, 1 + \varepsilon]$, it follows that

$$I_{n,1} \leq \widehat{\alpha}_n^{-1} K \|\widehat{A}^{\mathrm{MD}}_{\alpha,n} - A_\alpha\|_\infty \xrightarrow{\text{as}} 0, \quad n \to \infty,$$

where K is a Lipschitz constant. Furthermore, as $n \to \infty$, we have

$$I_{n,2} \leq -\ln(1/d)|\widehat{\alpha}_n^{-1} - \alpha^{-1}| \xrightarrow{\text{as}} 0.$$

Now, for any $v \in (0,1)$ and a small $\varepsilon > 0$ the map $x \mapsto v^{1/x}$ is Lipschitz continuous on $[\alpha - \varepsilon, \alpha + \varepsilon]$ and it satisfies $|v^{1/x} - v^{1/y}| \leq K(v)|x - y|$ for any $x, y \in [\alpha - \varepsilon, \alpha + \varepsilon]$, where $K(v) = v^{1/\alpha + \varepsilon}|\ln v|(\alpha - \varepsilon)^{-2}$. The function $K(v)$ is continuous and vanishes on the boundary of $(0,1)$, therefore there exist a positive constant K such that $\sup_{v\in(0,1)} K(v) = K$. In light of $\widehat{\alpha}_n \xrightarrow{\text{as}} \alpha$, for n sufficiently large and any $\varepsilon > 0$ we have $\alpha - \varepsilon \leq \widehat{\alpha}_n \leq \alpha + \varepsilon$, almost surely. Thus by the above arguments we have that

$$\left| \|\mathbf{t}\|_{1/\widehat{\alpha}_n}^{1/\widehat{\alpha}_n}(\mathbf{t}) - \|\mathbf{t}\|_{1/\alpha}^{1/\alpha} \right| \leq \sum_{j=1}^{d} \left| t_j^{1/\widehat{\alpha}_n} - t_j^{1/\alpha} \right| \leq dK|\widehat{\alpha}_n - \alpha| \xrightarrow{\text{as}} 0, \quad n \to \infty.$$

Consequently

$$\sup_{\mathbf{t}\in\mathcal{S}_d} \left| \|\mathbf{t}\|_{1/\widehat{\alpha}_n}^{1/\widehat{\alpha}_n}(\mathbf{t}) - \|\mathbf{t}\|_{1/\alpha}^{1/\alpha} \right| \xrightarrow{\text{as}} 0, \quad n \to \infty.$$

Finally, for n sufficiently large and an $\varepsilon > 0$ we have that almost surely $d^{1-1/\alpha} - \varepsilon \leq \|\mathbf{t}\|_{1/\widehat{\alpha}_n}^{1/\widehat{\alpha}_n} \leq 1 + \varepsilon$. Since the map $x \mapsto \ln x$ is Lipschitz continuous on $[d^{1-1/\alpha} - \varepsilon, 1 + \varepsilon]$, it holds that

$$I_{3,n} \leq K' \sup_{\mathbf{t}\in\mathcal{S}_d} \left| \|\mathbf{t}\|_{1/\widehat{\alpha}_n}^{1/\widehat{\alpha}_n} - \|\mathbf{t}\|_{1/\alpha}^{1/\alpha} \right| \xrightarrow{\text{as}} 0, \quad n \to \infty.$$

The result in (3.58) is now established. As a consequence, observing that

$$\|\widehat{A^\star}_n - A^\star\|_\infty = \|\exp(\ln \widehat{A^\star}_n) - \exp(\ln A^\star)\|_\infty,$$

a Lipschitz continuity argument similar to those used above yields

$$\|\widehat{A^\star}_n - A^\star\|_\infty \leq K''\|\ln \widehat{A^\star}_n - \ln A^\star\|_\infty \xrightarrow{\text{as}} 0, \quad n \to \infty.$$

■

Moment condition for Proposition 3.6.5 using $\widehat{\alpha}_n^{ML}$

Lemma 3.6.9 *Let* $(\mathbf{U}, V)$ *be defined as in* (3.33). *Let* $\omega'_{\epsilon,\mathbf{u}}$ *and and* φ_{ML} *be the functions defined in* (3.45) *and* (3.53), *respectively. Then, for every* $\epsilon \in [0, 1/2)$ *and* $\mathbf{u} \in [0,1]^d$, *we have* $-\infty < E\{\omega'_{\epsilon,\mathbf{u}}(\mathbf{U})\varphi_{\text{ML}}(V)\} < \infty$.

Proof. When $\epsilon = 0$, the result is immediate. Hence, in the reminder we only consider the case $\epsilon \in (0, 1/2)$. Let

$$M_\epsilon(\mathbf{x}) = d \max_{1\leq j\leq d}\{\max(x_j^{-\epsilon}, (1-x_j)^{-\epsilon})\}, \quad \epsilon \in (0, 1/2), \quad \mathbf{x} \in (0,1)^d.$$

By arguments in [28, p. 3019] and [33, p. 3082] and triangular inequality we obtain

$$\begin{aligned}
&\mathbb{E}|\omega'_{\epsilon,\mathbf{u}}(\mathbf{U})\varphi_{\mathrm{ML}}(V)| \\
= &\int_{(0,1)^{d+1}} |\omega'_{\epsilon,\mathbf{u}}(\mathbf{x})\varphi_{\mathrm{ML}}(v)|\mathrm{d}C_Q(\mathbf{x},v) \\
\leq &\int_{(0,1)^{d+1}} M_\epsilon(\mathbf{x})|\varphi_{\mathrm{ML}}(v)|\mathrm{d}C_Q(\mathbf{x},v) \\
\leq &\ (i_\alpha\alpha)^{-1}\int_{(0,1)^{d+1}} M_\epsilon(\mathbf{x})\{1-(1+\ln v)\ln(-\ln v)\}\mathrm{d}C_Q(\mathbf{x},v),
\end{aligned}$$

where $\omega'_{\epsilon,\mathbf{u}}$, i_α and φ_{ML} are as in (3.45), (3.52) and (3.53), respectively. On one hand, since $\mathbb{P}(M_\epsilon(\mathbf{U})>x)\leq 2d\min\{(x/d)^{-1/\epsilon},1\}$, it holds that

$$(i_\alpha\alpha)^{-1}\int_{(0,1)^{d+1}} M_\epsilon(\mathbf{x})\mathrm{d}C_Q(\mathrm{d}\mathbf{x},\mathrm{d}dv) = \frac{\mathbb{E}M_\epsilon(\mathbf{U})}{i_\alpha\alpha} < \infty.$$

On the other,

$$\begin{aligned}
&(i_\alpha\alpha)^{-1}\int_{(0,1)^{d+1}} M_\epsilon(\mathbf{x})(-1-\ln v)\ln(-\ln v)\mathrm{d}C_Q(\mathbf{x},v) \\
\leq &\ (i_\alpha\alpha)^{-1}\int_{(0,1)^{d+1}} \max\left\{M_\epsilon^2(\mathbf{x}),(-1-\ln v)^2\ln^2(-\ln v)\right\}\mathrm{d}C_Q(\mathbf{x},v) \\
\leq &\ (i_\alpha\alpha)^{-1}\mathbb{E}M_\epsilon^2(\mathbf{U}) + (i_\alpha\alpha)^{-1}\int_0^1 (-1-\ln v)^2\ln^2(-\ln v)\mathrm{d}v := A+B.
\end{aligned}$$

Since $\mathbb{P}(M_\epsilon^2(\mathbf{U})>x)\leq 2d\min\{(x/d^2)^{-1/2\epsilon},1\}$, then $A<\infty$. Furthermore, $B\approx 2.82(i_\alpha\alpha)^{-1}$. The integrability of $\omega'_{\epsilon,\mathbf{u}}\varphi_{\mathrm{ML}}$ with respect to C_Q now follows for every $\mathbf{u}\in[0,1]^d$. ∎

GPWM estimator of α

Lemma 3.6.10 *Assume that $\alpha>1/(k-1)$. Then, almost surely as $n\to\infty$*

$$\sqrt{n}(\widehat{\alpha}_n^{\mathrm{GPWM}}-\alpha)=\tau_{\mathrm{GPWM}}(\mathbb{W}_n)+o(1),$$

where $\tau_{\mathrm{GPWM}}:\ell^\infty([0,1])\to\mathbb{R}$ is defined as

$$\tau_{\mathrm{GPWM}}(f) = -2\int_0^1 f(v)\frac{v(-\ln v)^k\{\mu_{1,k-1}-\mu_{1,k}/(-\ln v)\}}{\dot{\Phi}_\alpha(\Phi_\alpha^{\leftarrow}(v))(k\mu_{1,k-1}-2\mu_{1,k})^2}\mathrm{d}v, \qquad (3.59)$$

with

$$\mu_{a,b} = \int_0^1 \Phi_\alpha^{\leftarrow}(v) v^a (-\ln v)^b \mathrm{d}v, \quad a,b \in \mathbb{N}_+$$

and $\dot{\Phi}_\alpha(v) = \partial/\partial v\, \Phi_\alpha(v)$, $v \in (0,1)$.

Proof. First, observe that

$$\begin{aligned}\sqrt{n}(\widehat{\alpha}_n^{\mathrm{GPWM}} - \alpha) &= \sqrt{n}\left[\left(k - 2\frac{\widehat{\mu}_{1,k}}{\widehat{\mu}_{1,k-1}}\right)^{-1} - \left(k - 2\frac{\mu_{1,k}}{\mu_{1,k-1}}\right)^{-1}\right] \\ &= \frac{\int_0^1 \mathbb{Q}_n(v)\tau_k(v)\mathrm{d}v}{n^{-1/2}\int_0^1 \mathbb{Q}_n(v)\rho_k(v)\mathrm{d}v + (k\mu_{1,k-1} - 2\mu_{1,k})^2} \\ &\equiv \frac{T_{n,1}}{T_{n,2}},\end{aligned}$$

where $\mathbb{Q}_n = \sqrt{n}(H_n^{\leftarrow} - \Phi_\alpha^{\leftarrow})$ and for every $v \in (0,1)$

$$\begin{aligned}\tau_k(v) &= 2v(-\ln v)^k\{\mu_{1,k-1} - \mu_{1,k}(-\ln v)^{-1}\}, \\ \rho_k(v) &= v(-\ln v)^{k-1}\left[k^2\mu_{1,k-1} - 2k\mu_{1,k} - 2k\mu_{1,k-1}(-\ln v) + 4\mu_{1,k}(-\ln v)\right].\end{aligned}$$

The term $T_{n,1}$ can be re-expressed as follows

$$\int_0^1 \mathbb{Q}_n(v)\dot{\Phi}_\alpha(\Phi_\alpha^{\leftarrow}(v))\tau_{k,\alpha}(v)\mathrm{d}v$$

where $\tau_{k,\alpha} = \tau_k/\dot{\Phi}_\alpha(\Phi_\alpha^{\leftarrow})$ has the following explicit expression

$$\tau_{k,\alpha}(v) = 2\alpha^{-1}\{\mu_{1,k-1}(-\ln v)^{k-1-1/\alpha} - \mu_{1,k}(-\ln v)^{k-2-1/\alpha}\}$$

and is integrable on the unit interval, since $\alpha > 1/(k-1) > 1/k$. Then, consecutive application of Theorem 3 in [14] and Bahadur-Kiefer theorem (see e.g., [21]) allows to conclude that almost surely

$$T_{n,1} = -\int_0^1 \mathbb{W}_n(v)\tau_{k,\alpha}(v)\mathrm{d}v + o(1),$$

where $\mathbb{W}_n$ is as in (3.37). Since $\alpha > 1/(k-1) > 1/k$, also $\rho_k(v)/\dot{\Phi}_\alpha(\Phi_\alpha^{\leftarrow}(v))$ is integrable on $(0,1)$ and a similar argument allows to conclude that almost surely

$$T_{n,2} = (k\mu_{1,k-1} - 2\mu_{1,k})^2 + o(1).$$

The result now follows. ∎

3.7 Supplementary material

3.7.1 Complements to Section 3.4.2: figures

First experiment: Figure 3.5 displays the simulation results of the first experiment of Section 3.4 for the sample size $n = 100$. The setting is the same as the one described thereof for the sample size $n = 50$. From the first to the third row, the MISE, ISB and IV computed estimating $A^\star$ by the GPWM-based estimators are reported. We can observe that the performances are quite similar to those illustrated in Figure 4.1. However, the effects of a larger sample size can be synthesized as follows:

- the upper bounds of the considered measures of performance are considerably reduced (as expected);
- the performance gap among the three GPWM-based estimators is substancially narrowed.

From the fourth to the sixth row, the ratio between the MISE's, ISB's and IV's computed estimating $A^\star$ by the GPWM- and ML-based estimators are reported. Overall, the trends observed for $n = 50$ are confirmed.

Second experiment: Figure 3.6 displays the simulation results of the second experiment of Subsection 3.4.2 for the sample size $n = 100$. The setting is the same as the one considered thereof for the sample size $n = 50$. The first three rows report the MISE, ISB and IV, computed estimating the Pickands dependence function by the GPWM-based estimators, while the last three rows report the ratio between the MISE, ISB and IV, computed estimating the Pickands dependence function by the GPWM- and ML-based estimators. Once again, as a consequence of a larger sample size, the gap among the performances of the GPWM-P, GPWM-CFG and GPWM-MD estimators is considerably narrowed. Differently from the first experiment, all the ML-based estimators now display less variability than the GPWM-based estimators. When $\alpha > 0.5$, the ML-based estimators are also less biased than the GPWM-based estimators, especially for weak dependence levels (θ approaching 2).

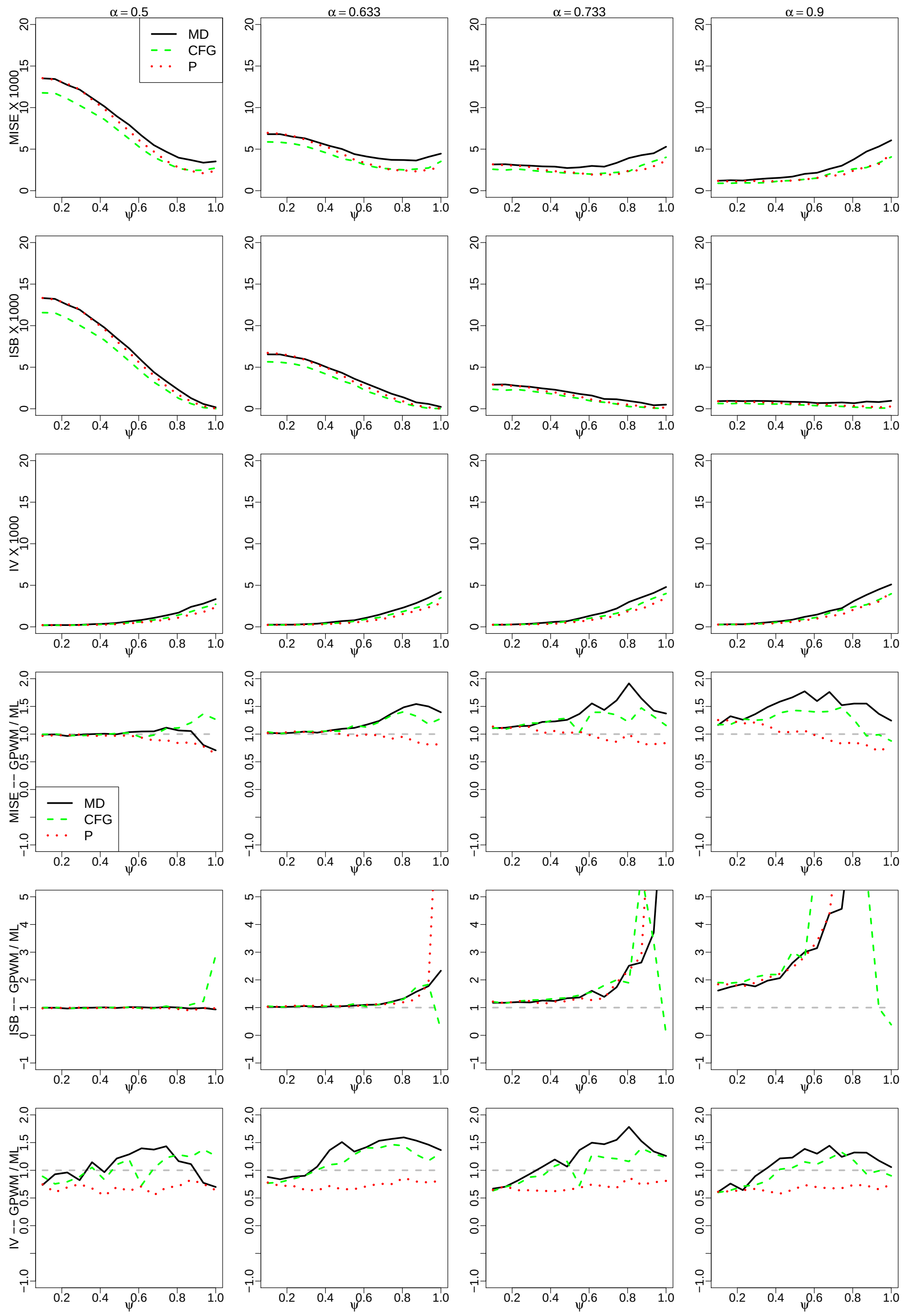

Figure 3.5: MISE, ISB, IV and ration between MISE, ISB and IV computed estimating $A^\star$ by $\widehat{A^\star}_n^{\text{GPWM},\bullet}$ and $\widehat{A^\star}_n^{\text{ML},\bullet}$ for 1000 samples of size 100. The setting of the first simulation experiment is considered.

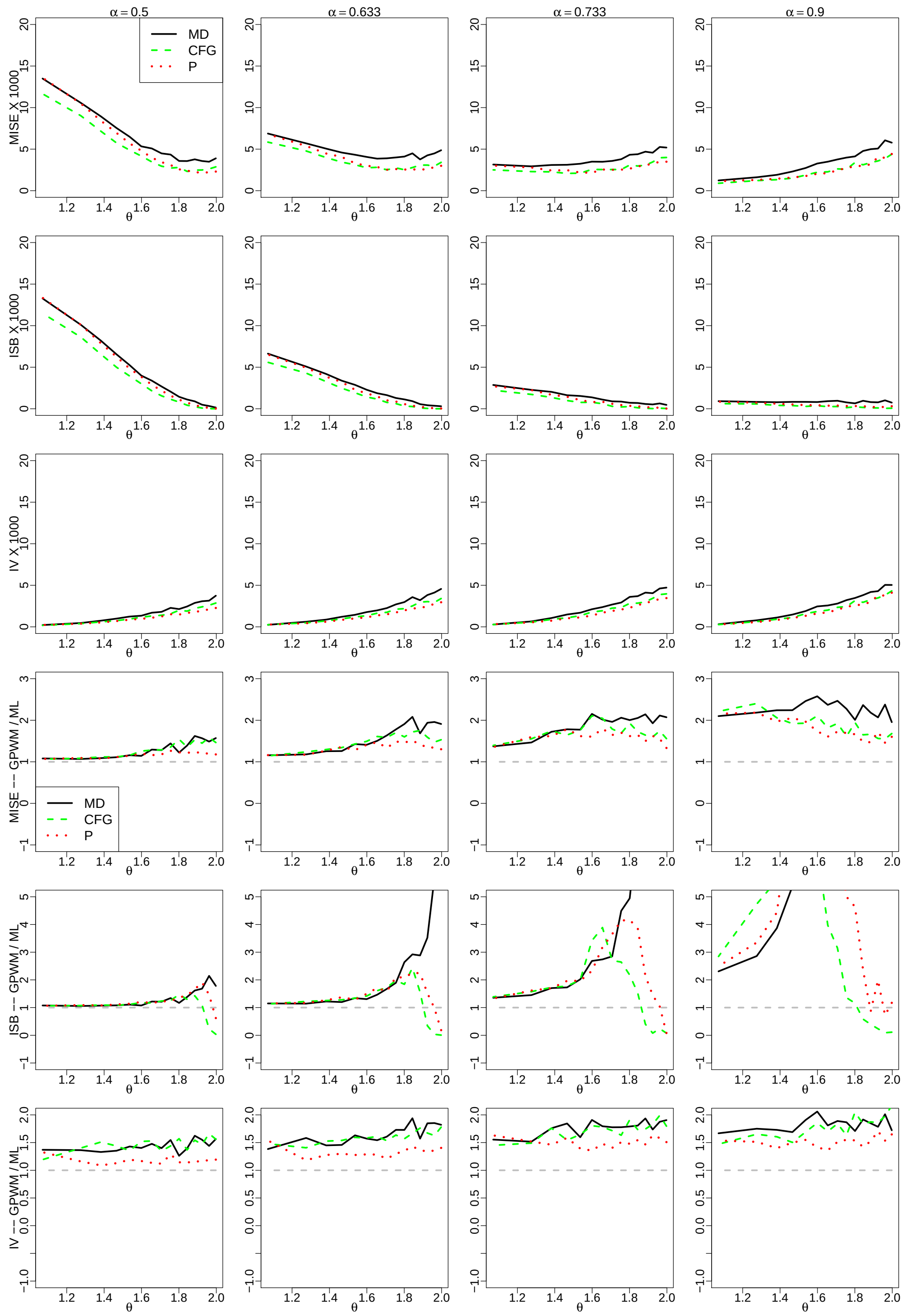

Figure 3.6: MISE, ISB, IV and ration between MISE, ISB and IV computed estimating $A^\star$ by $\widehat{A^\star}_n^{\mathrm{GPWM},\bullet}$ and $\widehat{A^\star}_n^{\mathrm{ML},\bullet}$ for 1000 samples of size 100. The setting of the second simulation experiment is considered.

3.7.2 Complements to Section 3.4.2: asymptotics

Consider the setup of the first simulation experiment in Section 3.4.2. According to the notation thereof, the random variables $\xi_i = \max\{R_{i,1}, R_{i,2}\}$, $i = 1, \ldots, n$, are iid according to $\Phi_{\alpha,\theta(G_\alpha)}(y) = \exp\{-\theta(G_\alpha)y^{-\alpha}\}$, $y > 0$, which differs from the case considered in Lemma 3.6.10 only by a scaling constant. Analogously, it can be shown that

$$\sqrt{n}(\widehat{\alpha}_n^{\text{GPWM}} - \alpha) = \tau_{\text{GPWM}}(\widetilde{\mathbb{W}}_n) + o_p(1),$$

where now $\widetilde{\mathbb{W}}_n(u) = \mathbb{C}_{G_\alpha,n}(u\mathbf{1})$, with $\mathbb{C}_{G_\alpha,n}$ as in (3.37), and τ_{GPWM} is defined via

$$\tau_{\text{GPWM}}(f) = -2\int_0^1 f(v)\frac{v(-\ln v)^k\{\mu_{1,k-1} - \mu_{1,k}/(-\ln v)\}}{\dot{\Phi}_{\alpha,\theta(G_\alpha)}(\Phi^{\leftarrow}_{\alpha,\theta(G_\alpha)}(v))(k\mu_{1,k-1} - 2\mu_{1,k})^2}\mathrm{d}v$$

$$\mu_{a,b} = \int_0^1 \Phi^{\leftarrow}_{\alpha,\theta(G_\alpha)}(v)v^a(-\ln v)^b\mathrm{d}v,$$

for $f \in \ell^\infty(0,1)$ and $a, b \in \mathbb{N}_+$. As for the ML estimator $\widehat{\alpha}_n^{\text{ML}}$, it now represents the unique zero of the function

$$\alpha' \mapsto \frac{1}{\alpha'} + \frac{\frac{1}{n}\sum_{i=1}^n \xi_i^{-\alpha'}\ln\xi_i}{\frac{1}{n}\sum_{i=1}^n \xi_i^{-\alpha'}} - \frac{1}{n}\sum_{i=1}^n \ln\xi_i$$

and the standardized sequence $\sqrt{n}(\widehat{\alpha}_n^{\text{ML}} - \alpha)$ admits a representation in terms of score function and Fisher information analogous to the one in (3.52) – see e.g. [8].

Consquently, in light of the above, the asymptotic behaviour of $\widehat{A^\star}_n^{\circ,\bullet}$ can be now derived by adapting to the present case the machinery of Propositions 3.6.4-3.6.5. Observe that, in this setting, weak convergence will be entirely driven by the empirical copula process $\mathbb{C}_{G_\alpha,n}$ rather than $\mathbb{C}_{Q,n}$.

3.7.3 Complements to the proof of Theorem 3.2.1

Let $F_N \in \mathcal{D}(\Lambda)$. We show that $u_n(y)/n \to 0$ and $n|I_{n,2}| \to 0$, as $n \to \infty$. We recall that in this case (e.g., [60], Proposition 1.9), $d_n = (1/\bar{F}_N)^{\leftarrow}(n)$ and $c_n = t_N(d_n)$, with

$$t_N(y) = \int_y^{+\infty} \bar{F}_N(s)\mathrm{d}s/\bar{F}_N(y), \quad y \in \mathbb{R}.$$

Observe that by Propositions 0.10 and 0.16 in [60], $(1/\bar{F}_N)^{\leftarrow}$ is Π-varying and there exists a further continuous, strictly increasing Π-varying function D such that $D > (1/\bar{F}_N)^{\leftarrow}$. Since D is monotone and Π-varying, D is slowly varying - see [60, p. 35]. Consequently, as $n \to \infty$

$$\frac{d_n}{n} < \frac{D(n)}{n} \to 0, \quad n \to \infty.$$

Moreover, observe that $\bar{F}_N = q_N[1 - \Upsilon_N]$, with q_N satisfying $\lim_{y\to+\infty} q_N(y) = \kappa_N > 0$ and Υ_N a Von Mises function. Therefore, for a small $\epsilon > 0$ and n sufficiently large, $q_N(d_n) \in (\kappa_N - \epsilon, \kappa_N + \epsilon)$ and

$$\begin{aligned}\frac{c_n}{n} &\le \frac{1}{n}\frac{\kappa_N + \epsilon}{\kappa_N - \epsilon}\int_{d_n}^{+\infty}[1 - \Upsilon_N(s)]\mathrm{d}s/[1 - \Upsilon_N(d_n)] \\ &\sim \frac{d_n}{n}\frac{\kappa_N + \epsilon}{\kappa_N - \epsilon}\frac{\omega_N(d_n)}{d_n} \to 0, \quad n \to \infty,\end{aligned}$$

where ω_N denotes the auxiliary function of Υ_N, which satisfies $\omega_N(y) \sim t_N(y)$ as $y \to \infty$ (e.g, [60], p. 49), and the conclusion follows from Lemma 1.2 in [60]. This entails that $u_n(y)/n \to 0$ as $n \to \infty$, for every $y \in \mathbb{R}$. The asymptotic equivalence $nI_{n,1} \sim \exp(-y)$ in (3.31) is a direct consequence.

Next, we provide some additional detail concerning (3.32). We have that

$$\begin{aligned}n|I_{2,n}| &= n[-\ln p_n(\mathbf{x})]\int_{u_n(y)}^{\infty}\bar{F}_N(t)e^{-t(-\ln p_n(\mathbf{x}))}\mathrm{d}t \\ &\le n[-\ln p_n(\mathbf{x})]\int_{u_n(y)}^{\infty}\bar{F}_N(t)\mathrm{d}t \\ &\le n[-\ln p_n(\mathbf{x})]\bar{F}_N(u_n(y))\frac{\xi_N + \epsilon}{\xi_N - \epsilon}\frac{\int_{u_n(y)}^{\infty}[1 - \Upsilon_N(t)]\mathrm{d}t}{1 - \Upsilon_N(u_n(y))} \\ &\sim [1 - p_n(\mathbf{x})]e^{-y}\frac{\xi_N + \epsilon}{\xi_N - \epsilon}\omega_N(u_n(y)), \quad n \to \infty \\ &\sim \frac{-\ln G(\mathbf{x})}{\mathrm{E}\,(N)}\frac{\xi_N + \epsilon}{\xi_N - \epsilon}\frac{u_n(y)}{n}\frac{\omega_N(u_n(y))}{u_n(y)}e^{-y} \to 0, \quad n \to \infty.\end{aligned}$$

Precisely, the fourth line follows from arguments in [60, p. 49], the fifth line from equation (3.29) and the sixth line from [60, Lemma 1.2].

Chapter 4

Inference for asymptotically independent data

An important topic in multivariate extreme-value theory is the development of probabilistic models and statistical methods to describe and measure the strength of dependence among extreme observations. The theory is well established for random vectors whose extremal dependence structure is compatible with that of asymptotically dependent models. This is the case when the distribution of sample maxima converges to a multivariate extreme-value distribution which does not factor into the product of its marginals. In practice, asymptotically dependent models are often not viable. In several applications, the data generating mechanism is associated to a limiting independence model for maxima. Despite this, observed data may still exhibit some residual dependence at the extremes, since convergence to asymptotically independent models can be very slow, while data is by nature finite. To account for such residual dependence, Ledford and Tawn [44, 45] developed a theory for the rate of decay of the dependence under asymptotic independence. Later theoretical developments are due to Resnick and co-authors (e.g. [48, 59]), under the name of "hidden regular variation". Since then, the literature on asymptotic independence has experienced a substantial growth. Yet, the state-of-art statistical procedures for handling asymptotically independent observations are far from being mature, especially in higher dimensions. This chapter contributes to the methodological development of such a context. In Section 4.2, we propose a statistical test based on the classical Pickands dependence function to diagnose the occurrence of asymptotic independence. In Section 4.3 we adopt the alternative componentwise maxima approach outlined in [58] and introduce a new dependence function, analoguous to the Pickands, which allows us to measure the residual (tail) dependence under asymptotic independence. We propose a semiparametric estimator, establish its asymptotic properties and illustrate its finite sample performance via a simulation study. In Section 4.4, we provide a discussion on the methodological assumptions and possible extensions of our work. All the proofs are deferred to Section 4.5.

4.1 Introduction

Let $\mathbf{X}$ be a multivariate random vector of dimension d, with distribution function F and marginals $F_1, \dots, F_d$. Recall from Section 1.2 that F is in the max-domain of attraction of a multivariate extreme-value distribution G if there exist sequences of constants $\mathbf{a}_n > \mathbf{0}$ and $\mathbf{b}_n \in \mathbb{R}^d$ such that

$$\lim_{n\to\infty} F^n(\mathbf{a}_n\mathbf{x} + \mathbf{b}_n) = G(\mathbf{x}).$$

for all continuity points $\mathbf{x} \in \mathbb{R}^d$ of G. Under this condition, a particular case arises when G is equal to the product of its marginal distributions. In this setting, we say that $\mathbf{X}$ satisfies the property of asymptotic independence (or tail independence) which is equivalent to saying that the pairwise conditional distributions satisfy

$$\lim_{u\to 1} \mathbb{P}\{F_j(X_j) > u | F_i(X_i) > u\} = 0$$

for all $1 \leq i \neq j \leq d$. If the above limits are positive, then the elements of $\mathbf{X}$ are called asymptotically dependent and G is not merely the product of its marginals. The classical theory expects asymptotic dependence and independence as the only two possible scenarios, conceiving the componentwise maxima as becoming independent in the second case. Many efforts have been made to characterize a residual tail dependence in the data (if there is any) by offering new dependence coefficients or probabilistic and statistical models under asymptotic independence; see [44], Section 8.4 in [12] or [48, 57, 58, 59, 77, 78], to list a few. In the bivariate case, several tests for checking asymptotic independence or tail independence have been proposed; see, e.g., [20, 41, 44], Chapter 6.5 in [23] and the references therein. However, extensions to dimensions $d > 2$ are still in their infancy. Recent proposals are based on the kth largest order statistics of the sample. Although these approaches are simple to implement, the performance of the resulting tests strongly depends on the choice of k; see, e.g., [42]. In Section 4.2, we propose an alternative approach to test asymptotic independence for an arbitrary dimension $d \geq 2$, based on the componentwise maxima. We illustrate the performances of our proposal up to dimension $d = 4$. Next, in Section 4.3, our focus will move on modelling and inferring residual depdendence within asymptotic independence.

4.2 A test for asymptotic independence

From equations (1.3)-(1.4) in Section 1.2, it is clear that an extreme-value copula is completely characterized by its Pickands dependence function. Consequently, a natural approach for testing independence at the extremes is through the Pickands dependence function.

4.2.1 A slightly modified version of the Pickands dependence estimator proposed by [46]

The estimator presented in this subsection is based on the concept of madogram, a notion borrowed from geostatistics in order to capture spatial structures. Starting from independent and identically distributed (iid) copies $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ of $\mathbf{Y} \overset{d}{\sim} G$, our estimator is defined as

$$\widehat{A}_n(\mathbf{t}) = \frac{\widehat{\nu}_n(\mathbf{t}) + c(\mathbf{t})}{1 - \widehat{\nu}_n(\mathbf{t}) - c(\mathbf{t})} \tag{4.1}$$

with

$$\begin{aligned} \widehat{\nu}_n(\mathbf{t}) &= \frac{1}{n} \sum_{i=1}^{n} \left[\bigvee_{j=1}^{d} \left\{ G_{n,j}^{(1)}(Y_{i,j}) \right\}^{1/t_j} - \frac{1}{d} \sum_{j=1}^{d} \left\{ G_{n,j}^{(1)}(Y_{i,j}) \right\}^{1/t_j} \right], \qquad (4.2) \\ c(\mathbf{t}) &= \frac{1}{d} \sum_{j=1}^{d} \frac{t_j}{1 + t_j}, \end{aligned}$$

where, for all $j \in \{1, \ldots, d\}$ and $a > 0$,

$$G_{n,j}^{(a)}(Y_{i,j}) = G_{n,j}(X_{i,j}) \left\{ \frac{1 + a}{a} \frac{1}{n} \sum_{k=1}^{n} G_{n,j}^{1/a}(Y_{k,j}) \right\}^{-a},$$

and

$$G_{n,j}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(Y_{i,j} \le x)}.$$

By convention, here $u^{1/0} = 0$ for $u \in (0,1)$. The estimator (4.2) is a slightly modified version of that proposed in [46], with $G_{n,j}^{(1)}$ in place of $G_{n,j}$. This ensures that the modified Pickands estimator $\widehat{A}_n$ now satisfies $\widehat{A}_n(\mathbf{e}_j) = 1$ for all $j \in \{1,\dots,d\}$, where $\mathbf{e}_j = (0,\dots,0,1,0,\dots,0)$ is the jth canonical unit vector; see Subsection 4.5.1. Reacall that this is a necessary condition that a function needs to satisfy in order to be a valid Pickands dependence function; see Properties 1.2.1. Although, as established in 4.5.1, our modified estimator shares the same asymptotic properties of the estimator proposed in [46], our modification greatly improves the performances of the latter in finite samples.

4.2.2 Construction of our statistical test

Using our estimator of A, we construct a statistical test to check asymptotic independence in dimensions $d \ge 2$. To this end, we consider the following system of hypotheses

$$\begin{cases} \mathcal{H}_0: & \forall \mathbf{t} \in \mathcal{S}_d \;\; A(\mathbf{t}) = 1, \\ \mathcal{H}_1: & \exists \mathbf{t} \in \mathcal{S}_d \;\; A(\mathbf{t}) < 1. \end{cases}$$

Note that $\mathcal{H}_0$ means that all the components of $\mathbf{X}$ are asymptotically independent, whereas under $\mathcal{H}_1$ some elements of $\mathbf{X}$ are asymptotically dependent.

Assuming that the extreme-value copula C has continuous partial derivatives over sets of the form $\{\mathbf{u} \in [0,1]^d : 0 < u_i < 1\}$ for all $i \in \{1,\dots,d\}$, by Theorem 2.4 in [46] and arguments in Subsection 4.5.1, under $\mathcal{H}_0$ we have that, as $n \to \infty$,

$$\sqrt{n}\,(\widehat{A}_n(\mathbf{t}) - 1)_{\mathbf{t}\in\mathcal{S}_d} \rightsquigarrow -4\left\{\int_0^1 \mathbb{A}(v^{t_1},\dots,v^{t_d})\mathrm{d}v\right\}_{\mathbf{t}\in\mathcal{S}_d}, \tag{4.3}$$

where $\mathbb{A}$ is a centered Gaussian process on $[0,1]^d$ with continuous sample paths and covariance function equal to

$$\mathrm{cov}\{\mathbb{A}(\mathbf{u}),\mathbb{A}(\mathbf{v})\} = \prod_{j=1}^{d} u_j \wedge v_j - \sum_{j=1}^{d}\left(u_j \wedge v_j \prod_{i\neq j} u_i v_i\right) + (d-1)\prod_{j=1}^{d} u_j v_j.$$

As a consequence, by the Continuous Mapping Theorem (see, e.g., Chapter 2.1 in [74]), it follows that, as $n \to \infty$,

$$\widehat{S}_n = \sup_{\mathbf{t} \in \mathcal{S}_d} \sqrt{n} \, |\widehat{A}_n(\mathbf{t}) - 1| \rightsquigarrow S = \sup_{\mathbf{t} \in \mathcal{S}_d} 4 \left| \int_0^1 \mathbb{A}(v^{t_1}, \ldots, v^{t_d}) \, \mathrm{d}v \right| .$$

This convergence result is the cornerstone of our test. Denoting by $Q_S(\alpha)$ with $\alpha \in (0,1)$, the $(1-\alpha)$-quantile function for the distribution of the random variable S, $\mathcal{H}_0$ can be rejected at approximate $100 \times \alpha\%$ significance level whenever $\widehat{s}_n$, the observed value of $\widehat{S}_n$, exceeds $Q_S(\alpha)$. Unfortunately, there is no closed form for the function $Q_S(\alpha)$. However, an approximation can still be computed with a Monte Carlo simulation as follows.

Note that for any $u, v \in [0,1]$ and $\mathbf{t}, \mathbf{w} \in \mathcal{S}_d$, the covariance function of the Gaussian process $\mathbb{A}$ in (4.3), evaluated at the indexes $u^{\mathbf{t}}, v^{\mathbf{w}} \in [0,1]^d$, is equal to

$$\begin{aligned} &\operatorname{cov}\{\mathbb{A}(v^{t_1}, \ldots, v^{t_d}), \mathbb{A}(u^{w_1}, \ldots, u^{w_d})\} = \\ &\prod_{j=1}^{d} (v^{t_j} \wedge u^{w_j}) - \sum_{j=1}^{d} (v^{t_j} \wedge u^{w_j}) v^{1-t_j} u^{1-w_j} + (d-1)uv. \end{aligned} \tag{4.4}$$

Thus, for any fixed $\alpha \in (0,1)$, an approximation of the quantile $Q_S(\alpha)$ can be obtained by adhering to the following four steps:

1. Divide the unit interval $(0,1)$ and the simplex $\mathcal{S}_d$ in p and m equally spaced points, where p and m are positive integers. Let $v_1, \ldots, v_m$ and $\mathbf{t}_1, \ldots, \mathbf{t}_p$ be the two sequences of points partitioning $(0,1)$ and $\mathcal{S}_d$, respectively. The sequences $v_1, \ldots, v_m$ and $\mathbf{t}_1, \ldots, \mathbf{t}_p$ form a finite sequence of positions

$$(v_r^{t_{k,1}}, \ldots, v_r^{t_{k,d}}) \in [0,1]^d,$$

   with $r \in \{1, \ldots, m\}$ and $k \in \{1, \ldots, p\}$, on which the process $\mathbb{A}$ is simulated.

2. For each $i \in \{1, \ldots, n^*\}$, sample a realization

$$x_i(v_1^{t_{1,1}}, \ldots, v_1^{t_{1,d}}), \ldots, x_i(v_m^{t_{p,1}}, \ldots, v_m^{t_{p,d}})$$

   of a zero-mean Gaussian process at sites $(v_r^{t_{k,1}}, \ldots, v_r^{t_{k,d}})$, for $r \in \{1, \ldots, m\}$ and $k \in \{1, \ldots, p\}$, with an $mp \times mp$ variance-covariance matrix defined through the covariance function in (4.4).

3. Simulate samples that approximately follow the distribution of the random variable S, the integral and the sup in the expression of S being approximated by the sum and the max for sufficiently large values of m and p. This leads to n^* realizations given, for all $i \in \{1, \ldots, n^*\}$, by

$$\widetilde{s}_i = \max_{k \in \{1,\ldots,p\}} \frac{4}{m} \left| \sum_{r=1}^{m} x_i(v_r^{t_k,1}, \ldots, v_r^{t_k,d}) \right|.$$

4. An approximation of the quantile $Q_S(\alpha)$, denoted by $\widetilde{Q}_S(\alpha)$, can then be obtained by computing the sample quantile of the realizations $\widetilde{s}_1, \ldots, \widetilde{s}_{n^*}$ for sufficiently large n^*.

4.2.3 Numerical results

We illustrate the performance of our statistical test through a simulation study. Precisely, we estimate some values of the significance level α and the power $1 - \beta$ of the test by computing the empirical proportion of simulated samples under the null hypothesis and the alternative hypothesis that rejected the null hypothesis, respectively. For simplicity we focus on the significance levels $\alpha = 0.05$ and 0.01. The study consists of five experiments.

First experiment: As a first step, we compute the approximate quantile $\widetilde{Q}_S(\alpha)$, for a given α, following the above algorithm. The quality of the approximation relies on the values of the indexes m, p and n^*. Clearly, the larger their values are, the more accurate the approximation is. We set $n^* = 500{,}000$. We consider increasing values of m and p and for each combination we compute $\widetilde{Q}_S$. We stop the search when the value of $\widetilde{Q}_S(\alpha)$ does not increase anymore, up to the second decimal. The computation of $\widetilde{Q}_S$ requires a considerable computational effort; therefore, we derive its values only for vectors $\mathbf{Y}$ of dimension $d \in \{2, 3, 4\}$. In a second step, we compute the rejection rates. We focus on the multivariate logistic extreme-value model introduced in [69], with dependence parameter $\psi \in (0, 1]$, $\psi = 1$ corresponding to independent components of $\mathbf{Y}$, whereas complete dependence can be reached as $\psi \to 0$.

We consider 20 equally spaced values of ψ in $(0, 1]$. For each of them, we simulate n independent observations from a logistic extreme-value distribution with unit Fréchet margins. Then we estimate the Pickands dependence function by (4.1) and we compute $\widehat{s}_n$. We repeat this task 5000 times and we compute the proportion of times that $\widehat{s}_n > \widetilde{Q}_S(\alpha)$. This experiment is repeated for different values of the samples size n and different dimension d of $\mathbf{Y}$. The middle part of Table 4.1 reports the estimated values of the significance levels α in the case of $\psi = 1$.

We see that accurate estimates of α are already obtained with sample size $n = 50$. Figure 4.1 displays the estimated power of the test. In the first and second rows the results obtained with $\alpha = 0.05$ and $\alpha = 0.01$ are reported, respectively. The panels from left to right illustrate the results for dimensions 2, 3 and 4. Once again, the test shows a good performance already with sample size $n = 50$. In the case $d = 2$ we see that the power of the test reaches 1 at mild dependence levels, i.e., $\psi = 0.5$. This figure also outlines that the power of the test improves as the dimension of $\mathbf{Y}$ increases and that, as expected, for any fixed dimension $d \in \{2, 3, 4\}$, it also improves as the sample size increases.

Second experiment: We repeat the second step of the first experiment using a different approximate quantile. Precisely, we simulate n values from d independent univariate Fréchet distributions, then we estimate the Pickands dependence function by (4.1) and we compute $\widehat{s}_n$. We repeat this task 5000 times and we compute the empirical quantile, for a given α, denoted by $Q_{\widehat{S}_n}(\alpha)$. The right-hand side of Table 4.1 reports their values for different values of n and d. We see that the empirical quantiles rapidly approach the asymptotic (approximate) quantiles, as the sample size increases. With sample size $n = 50$ already, the two types of quantiles are very close. The third and fourth lines of Figure 4.1 display the comparison between the estimates of $1 - \beta$ obtained with the two types of quantiles, but also the estimates of α since ($\psi = 1$ corresponds to independent components and thus the proportion of rejections reported in the figures represents an approximation of α in that case). Since the inferential results are the same for both $\alpha = 0.01$ and $\alpha = 0.05$, only the latter are reported. The performances obtained with the empirical quantiles are very close to those obtained with the asymptotic quantiles, already with sample size $n = 50$.

Third experiment: We repeat the second experiment using the Genest–Rémillard (GR) omnibus rank-based test of independence [27] and our proposed test with

Table 4.1: Estimated significance levels α. From left to right: the dimension of $\mathbf{Y}$, the true significance level, the approximate asymptotic $(1-\alpha)$-quantile, the empirical proportion of simulated samples under $\mathcal{H}_0$ that rejected the null hypothesis and the empirical $(1-\alpha)$-quantile. Here $\psi = 1$.

			$\hat{\alpha}$				$Q_{\widehat{S}_n}(\alpha)$			
			n				n			
d	α	$\widetilde{Q}_S(\alpha)$	25	50	100	200	25	50	100	200
2	0.05	0.960	0.0380	0.0460	0.0512	0.0524	0.9190	0.9393	0.9512	0.9541
	0.01	1.204	0.0060	0.0082	0.0102	0.0102	1.1359	1.1739	1.1926	1.1992
3	0.05	1.300	0.0364	0.0452	0.0508	0.0574	1.2540	1.2755	1.3036	1.3210
	0.01	1.540	0.0056	0.0068	0.0084	0.0092	1.4126	1.4904	1.5295	1.5601
4	0.05	1.480	0.0398	0.0454	0.0548	0.0576	1.5312	1.5508	1.5883	1.5745
	0.01	1.740	0.0064	0.0082	0.0096	0.0126	1.7715	1.7135	1.7849	1.7867

the Capéraà–Fougères–Genest (CFG) estimator [9, 79] of the Pickands dependence function in place of (4.1). The GR test was performed using the R package `Copula` [70, 43]. Figure 4.2 shows the estimated powers obtained with the GR test and our test (with both the CFG and the Madogram-based (4.1) estimators). For brevity, we show the results for $\alpha = 0.05$ and sample sizes $n \in \{25, 50\}$. We see that our test always outperforms the GR test, with the best results provided by the CFG estimator in dimension $d = 2$, whereas, in higher dimensions, similar results can be reached with either the Madogram-based estimator or the CFG.

Fourth experiment: We repeat the third experiment by sampling from three alternative distributions. In the first case, we draw samples from a three-dimensional random vector with a pair that follows the logistic extreme-value distribution and where the last variable is independent from the other two. In the second case, we consider a four-dimensional random vector with two pairs that follow the logistic extreme-value distribution and where the components of one pair are independent from each component of the second pair. In the last case, we consider a four-dimensional random vector, where one pair has independent components and each of these are independent from the components of the other pair, which in turn follows the logistic extreme-value distribution. The results are collected in the third ($n = 25$) and fourth row ($n = 50$) of Figure 4.2. In these cases we see that our test loses power and provides inferential results very similar to those provided by the GR test. However, the latter outperforms our test in the case of the largest number of independent variables.

Fifth experiment: We consider the multivariate inverted symmetric logistic model (see, e.g., [45, 78]), with dependence parameter $\psi \in (0, 1]$, $\psi = 1$ corresponding to exact independence of the components of $\mathbf{Y}$, whereas asymptotic dependence is reached as $\psi \to 0$. This time, we consider 10 equally spaced values of ψ in $(0, 1]$. For each of them, we simulate 366 values (for similarity with annual maxima) from an inverted logistic distribution with exponential margins. Then, we compute the normalized componentwise maxima and we repeat this procedure in order to obtain n normalized maxima from which we estimate the Pickands dependence function using (4.1) and we compute $\widehat{s}_n$. We repeat this task 5000 times and we compute the proportion of times that $\widehat{s}_n > Q_{S_n}(0.05)$. This procedure has been repeated for different values of d and n and the results are summarized in

Table 4.2. With $d = 2$, the rejection rates are close to 0.05 whenever ψ is larger than 0.5. Otherwise, the rejection rate is greater than 0.05 and it reaches 1 when ψ approaches 0. In these cases, it can be observed that the normalized maxima show quite a strong dependence, which indeed resembles that of an asymptotically dependent model more than an asymptotically independent one. The strength of the dependence is reduced whenever the normalized maxima are computed on sequences larger than 366, resulting in improvements in the performances of our test. The test performance deteriorates as the dimension of $\mathbf{Y}$ increases.

In conclusion, this study highlights the good performance of our statistical test for detecting the exact independence of sample maxima. However, our test is also useful to diagnose the occurrence of asymptotic independence for multivariate data, as long as the residual dependence is mild. By contrast, in the cases of strong residual dependence, our test fails to recognize the data as asymptotically independent. Clearly, these cases are the most naturally difficult to detect and more specific tools are required.

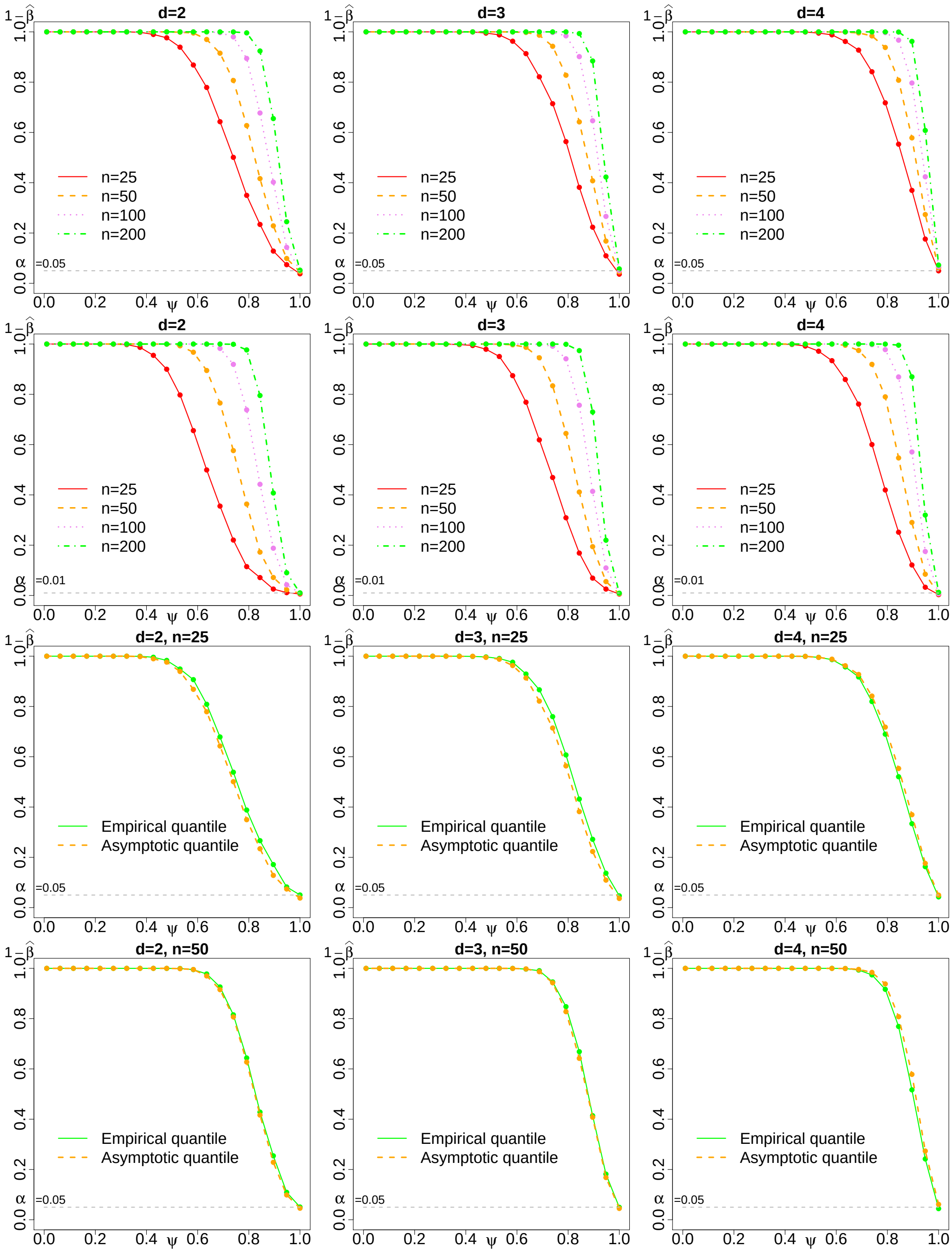

Figure 4.1: Estimated power functions. Points report the empirical proportion of simulated samples under $\mathcal{H}_1$ that rejected $\mathcal{H}_0$ as a function of ψ. Samples are simulated from a symmetric logistic model with parameter ψ. From the left to the right, the dimension is 2, 3 and 4, respectively. Comparison of the estimates of $1-\beta$ obtained with four sample sizes when $\alpha = 0.05$ (first row) and $\alpha = 0.01$ (second row), and obtained with the empirical and asymptotic quantiles when $\alpha = 0.05$, $n = 25$ (third row) and $n = 50$ (fourth row).

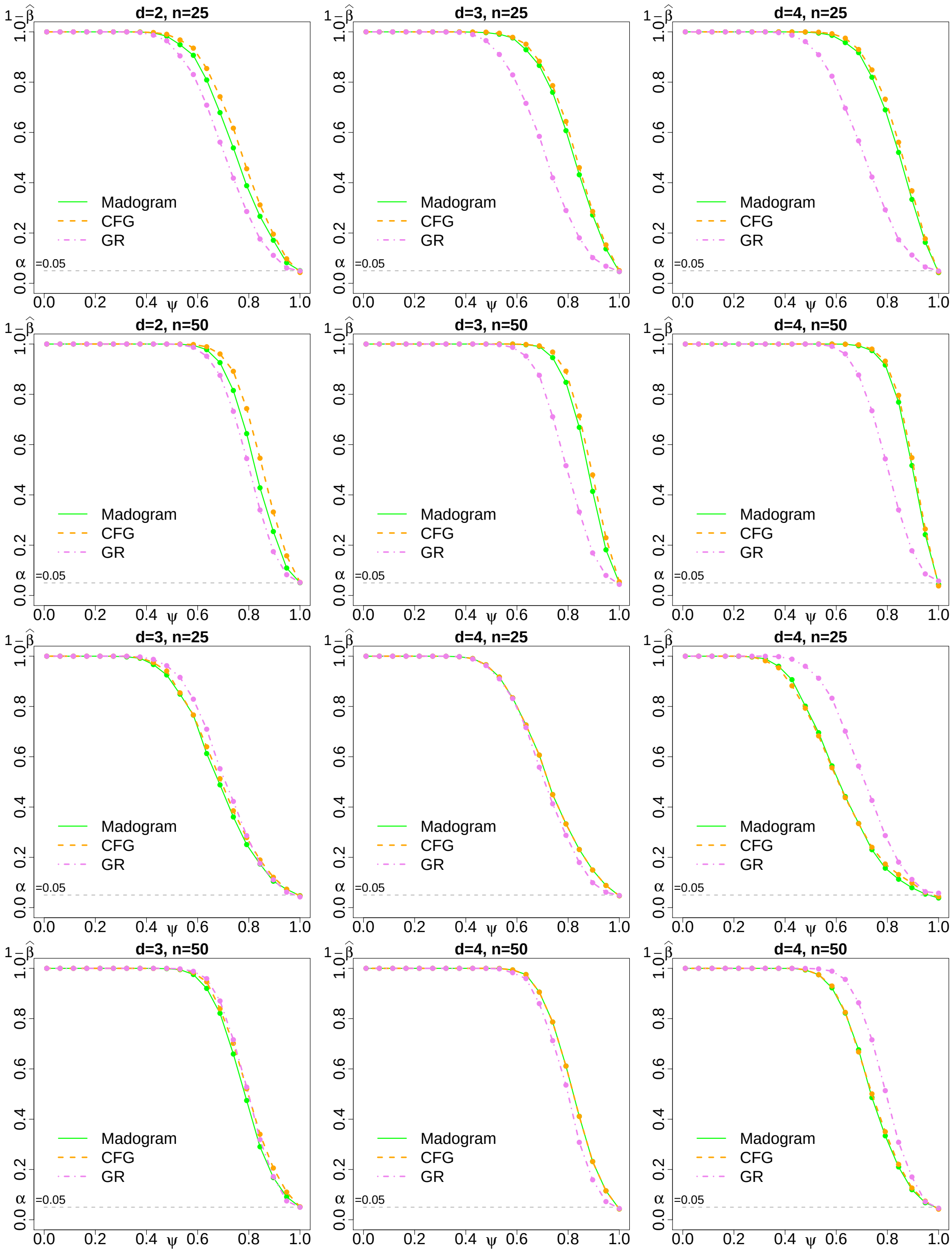

Figure 4.2: Estimated power functions. Points report the empirical proportion of simulated samples under $\mathcal{H}_1$ that rejected $\mathcal{H}_0$ as a function of ψ. Samples are simulated from a symmetric logistic model with parameter ψ. Comparison between GR test and our test (with both the CFG and the Madogram-based (4.1) estimators) when $\alpha = 0.05$, $n = 25$ (first row) and $n = 50$ (second row), from the left to the right, the dimension is 2, 3 and 4, respectively. The third and fourth rows are constructed similarly as the two first ones, but where samples are replications of random vectors with dependent and independent components.

Table 4.2: Estimated significance levels α. From left to right: the dimension of $\mathbf{Y}$, the sample size and the empirical proportion of simulated samples under $\mathcal{H}_0$ that rejected the null hypothesis for different values of ψ.

		ψ									
d	n	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
2	25	0.0522	0.0468	0.0578	0.0580	0.0584	0.0842	0.1294	0.2730	0.6372	0.9920
	50	0.0554	0.0524	0.0542	0.0562	0.0590	0.0980	0.1772	0.4538	0.8932	1.0000
	100	0.0476	0.0470	0.0506	0.0561	0.0594	0.1208	0.2938	0.7128	0.9962	1.0000
	200	0.0486	0.0570	0.0568	0.0560	0.0601	0.1856	0.4934	0.9482	1.0000	1.0000
3	25	0.0470	0.0508	0.0522	0.0582	0.0808	0.1076	0.2190	0.4234	0.8604	0.9990
	50	0.0542	0.0538	0.0528	0.0589	0.8320	0.1604	0.3606	0.7424	0.9950	1.0000
	100	0.0536	0.0468	0.0550	0.0594	0.0922	0.2084	0.5274	0.9486	1.0000	1.0000
	200	0.0540	0.0500	0.0506	0.0652	0.1242	0.3050	0.8096	0.9996	1.0000	1.0000
4	25	0.0488	0.0418	0.0448	0.0504	0.0692	0.1336	0.2676	0.6332	0.9582	1.0000
	50	0.0452	0.0438	0.0536	0.0574	0.1018	0.1854	0.4736	0.8852	0.9996	1.0000
	100	0.0496	0.0484	0.0468	0.0598	0.1130	0.2746	0.7052	0.9920	1.0000	1.0000
	200	0.0486	0.0448	0.0560	0.0770	0.1596	0.4768	0.9440	1.0000	1.0000	1.0000

4.3 Asymptotic independence for componentwise maxima

Being able to test asymptotic independence versus asymptotic dependence is obviously important. However, since asymptotic independence often arises in applications, it is also crucial to develop some general models that accommodate both situations. In this section, we consider the framework of [57]; see also [45]. Letting $\mathbf{X}$ be a d-dimensional random vector with common unit Fréchet margins, i.e., $\mathbb{P}(X_j \leq x) = e^{-1/x}$ for every $x > 0$ and $j \in \{1, \ldots, d\}$, we assume that the joint survival function of $\mathbf{X}$ is multivariate regularly varying with index $-1/\eta$, where $\eta \in (0, 1]$, i.e., $\mathbb{P}(\mathbf{X} > \mathbf{x}) = \tau(\mathbf{x})(x_1 \cdots x_d)^{-1/d\eta}$ with τ a slowly varying function satisfying

$$\lim_{r \to \infty} \frac{\tau(r\, x_1, \ldots, r\, x_d)}{\tau(r, \ldots, r)} = g(\mathbf{x})$$

for all $\mathbf{x} \in (0, \infty)^d$. The function g is homogeneous of order 0: $g(a\, x_1, \ldots, a\, x_d) = g(x_1, \ldots, x_d)$ for any $a > 0$. This assumption implies that the conditional joint survival function admits the following limit representation for every $\mathbf{x} \geq \mathbf{1}$, the vector of 1s,

$$\begin{aligned} \lim_{u \to \infty} \mathbb{P}(\mathbf{X} > u\mathbf{x} | \mathbf{X} > u\mathbf{1}) &= \lim_{u \to \infty} \frac{\mathbb{P}(\mathbf{X} > u\mathbf{x})}{\mathbb{P}(\mathbf{X} > u\mathbf{1})} \\ &= \eta \int_{\mathcal{S}_d} \bigwedge_{j=1}^{d} \left(\frac{w_j}{x_j} \right)^{1/\eta} \mathrm{d}H_\eta(\mathbf{w}), \end{aligned} \tag{4.5}$$

where H_η is a non-negative measure satisfying the condition

$$\eta \int_{\mathcal{S}_d} \bigwedge_{j=1}^{d} w_j^{1/\eta} \mathrm{d}H_\eta(\mathbf{w}) = 1. \tag{4.6}$$

This measure H_η is a particular case of the hidden angular measure introduced by [59] (see also [48]) when $\eta < 1$ and it is a rescaled version of the classical angular measure when $\eta = 1$ and asymptotic dependence holds true; see [57] for details. Following [58], we assume that H_η is a finite measure. We recall that η is the so-called coefficient of tail dependence, which measures the level of dependence within the asymptotic independence framework. Specifically, $\eta < 1$ corresponds to the case of asymptotic independence. More precisely, when the coefficient η falls in the following sets: $(1/d, 1)$, $\{1/d\}$ or $(0, 1/d)$, then we say that among the variables there is a positive association, independence or negative association, respectively, within asymptotic independence; see, e.g., [44]. The case $\eta = 1$ corresponds either to asymptotic dependence or independence; see, e.g., [78], Section 4 and Appendix A.

4.3.1 A $\eta-$Pickands dependence function

Consider now iid copies $\mathbf{X}_1, \ldots, \mathbf{X}_n$ of $\mathbf{X}$ and for a small $\varepsilon > 0$ define $\mathbf{M}_{n,\varepsilon} = (M_{n,1,\varepsilon}, \ldots, M_{n,d,\varepsilon})$ as the vector of componentwise maxima given by

$$M_{n,j,\varepsilon} = \bigvee_{i \in I_n(\varepsilon)} X_{i,j}, \quad j \in \{1, \ldots, d\},$$

with $I_n(\varepsilon) = \{1 \leq i \leq n : \mathbf{X}_i > \mathbf{1}\varepsilon\}$. Let b_n be a sequence of normalizing constants defined by the equation $n\mathbb{P}(\mathbf{X} > b_n) = 1$. Differently from the classical theory, the limiting distribution for the normalized vector of componentwise maxima $\mathbf{M}_{n,\varepsilon}$ is now obtained via

$$G_\eta(\mathbf{y}) = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \mathbb{P}(\mathbf{M}_{n,b_n\varepsilon} \leq b_n \mathbf{y}). \tag{4.7}$$

for each $\mathbf{y} \in (0, \infty]^d$; see [58] for details. When a limiting distribution exists with nondegenerate margins, we call G_η a multivariate η-extreme-value distribution. Specifically, G_η has Fréchet univariate margins $G_{\eta,j}(y) = \exp(-\sigma_{\eta,j} y^{-1/\eta})$, for all $y > 0$ and $j \in \{1, \ldots, d\}$, and satisfy:

$$G_\eta(\mathbf{y}) = C_\eta\{G_{\eta,1}(y_1), \ldots, G_{\eta,d}(y_d)\}, \tag{4.8}$$

for all $\mathbf{y} \in (0, \infty]^d$, where C_η is an η-extreme-value copula, i.e., for all $\mathbf{u} \in (0, 1]^d$,

$$C_\eta(\mathbf{u}) = \exp\left[-V_\eta\left\{\left(\frac{\sigma_{\eta,1}}{-\ln u_1}\right)^\eta, \ldots, \left(\frac{\sigma_{\eta,d}}{-\ln u_d}\right)^\eta\right\}\right],$$

with $V_\eta : (0, \infty]^d \to [0, \infty)$ a homogeneous function of order $-1/\eta$ given by

$$V_\eta(\mathbf{y}) = \eta \int_{\mathcal{S}_d} \bigvee_{1 \leq j \leq d} \left(\frac{w_j}{y_j}\right)^{1/\eta} \mathrm{d}H_\eta(\mathbf{w}),$$

for all $\mathbf{y} \in (0, \infty]^d$, with

$$\sigma_{\eta,j} = V_\eta(\infty, \dots, \infty, 1, \infty, \dots, \infty) = \eta \int_{\mathcal{S}_d} w_j^{1/\eta} \mathrm{d}H_\eta(\mathbf{w}). \tag{4.9}$$

Define $L_\eta(\mathbf{z}) := V_\eta\{(\boldsymbol{\sigma}_\eta/\mathbf{z})^\eta\}$, for all $\mathbf{z} \in [0, \infty)^d$. This function is referred to as the η-stable tail dependence function. Using the homogeneity property, it can be reexpressed, for all $\mathbf{z} \in [0, \infty)^d$, as

$$L_\eta(\mathbf{z}) = (z_1 + \cdots + z_d) A_\eta(\mathbf{t}),$$

where $t_j = z_j/(z_1 + \cdots + z_d)$ for $j \in \{2, \dots, d\}$ and $t_1 = 1 - t_2 - \cdots - t_d$. Henceforward, we refer to the function A_η as the η-Pickands dependence function.

Proposition 4.3.1 *The η-Pickands dependence function A_η satisfies the following properties:*

1. *For all $\eta \in (0, 1]$, $A_\eta(\mathbf{e}_j) = 1$ for all $j \in \{1, \dots, d\}$.*
2. *Under asymptotic dependence, we have $A_1(\mathbf{t}) = A(\mathbf{t})$, for all $\mathbf{t} \in \mathcal{S}_d$.*
3. *For every $\eta \in (0, 1]$ and $\mathbf{t} \in \mathcal{S}_d$, $1/d \leq \max(t_1, \dots, t_d) \leq A_\eta(\mathbf{t}) \leq 1$.*
4. *$A_\eta(\mathbf{t})$ is convex, i.e., $A_\eta\{a\mathbf{t}_1 + (1-a)\mathbf{t}_2\} \leq aA_\eta(\mathbf{t}_1) + (1-a)A_\eta(\mathbf{t}_2)$, for all $a \in [0, 1]$ and $\mathbf{t}_1, \mathbf{t}_2 \in \mathcal{S}_d$.*

Let $\mathbf{Z} \overset{d}{\sim} G_\eta$. Similarly to the classical literature, a $\eta-$madogram function can be defined as the expected distance between the maximum and the mean of the variables $G_{\eta,1}^{1/\eta t_1}(Z_1), \dots, G_{\eta,d}^{1/\eta t_d}(Z_d)$, i.e.,

$$\nu_\eta(\mathbf{t}) = \mathbb{E}\left[\bigvee_{j=1}^{d} \left\{G_{\eta,j}^{1/\eta t_j}(Z_j)\right\} - \frac{1}{d}\sum_{j=1}^{d} G_{\eta,j}^{1/\eta t_j}(Z_j)\right].$$

This function can be linked to the $\eta-$Pickands dependence function as follows.

Proposition 4.3.2 *Any random vector $\mathbf{Z}$ with a η-extreme-value distribution admits a η-Pickands dependence function A_η satisfying*

$$A_\eta(\mathbf{t}) \;=\; \frac{1}{\eta}\frac{\nu_\eta(\mathbf{t}) + c_\eta(\mathbf{t})}{1 - \nu_\eta(\mathbf{t}) - c_\eta(\mathbf{t})} \tag{4.10}$$

for all $\mathbf{t} \in \mathcal{S}_d$, where

$$c_\eta(\mathbf{t}) = \frac{1}{d}\sum_{j=1}^{d} \frac{t_j}{t_j + 1/\eta}. \tag{4.11}$$

This $\eta-$Pickands dependence function can be used to represent the level of dependence among the elements of $\mathbf{Z}$. In the next section, we propose an estimator of this function and derive its main asymptotic properties.

4.3.2 An estimator of the $\eta-$Pickands dependence function

Let $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ be iid copies of $\mathbf{Z}$ with distribution G_η and define, for all $\mathbf{y} \in (0, \infty]^d$,

$$H_n(\mathbf{y}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(\mathbf{Z}_i \leq \mathbf{y})},$$

and its associated empirical process given, for all $\mathbf{y} \in (0, \infty]^d$, by $\mathbb{H}_n(\mathbf{y}) = \sqrt{n}\,\{H_n(\mathbf{y}) - G_\eta(\mathbf{y})\}$. In order to estimate the $\eta-$Pickands dependence function we first assume that we have at our disposal an estimator $\widehat{\eta}_n$ of η satisfying the condition:

Condition 4.3.3 Let $\widehat{\eta}_n$ be an estimator of η satisfying:

(i) $\widehat{\eta}_n \to \eta$ a.s. as $n \to \infty$.

(ii) One of the following holds true:

(a) $\sqrt{n}\,(\widehat{\eta}_n - \eta) = n^{-1/2} \sum_{i=1}^n \rho(\mathbf{Z}_i) + o_p(1)$, where $\rho : (0, \infty]^d \mapsto \mathbb{R}$ is a measurable function such that $\mathrm{E}\rho(\mathbf{Z}) = 0$ and $\mathrm{E}\rho^2(\mathbf{Z}) < \infty$.

(b) $\sqrt{n}\,(\widehat{\eta}_n - \eta) = \chi(\mathbb{H}_n) + o_p(1)$, where $\chi : \ell^\infty((0, \infty]^d) \mapsto \mathbb{R}$ is a bounded linear functional.

In the spirit of (4.1) in Section 4.2, we propose the following estimator for A_η:

$$\widehat{A}_{\widehat{\eta}_n,n}(\mathbf{t}) = \frac{1}{\widehat{\eta}_n} \frac{\widehat{\nu}_{\widehat{\eta}_n,n}(\mathbf{t}) + \widehat{c}_{\widehat{\eta}_n,n}(\mathbf{t})}{1 - \widehat{\nu}_{\widehat{\eta}_n,n}(\mathbf{t}) - \widehat{c}_{\widehat{\eta}_n,n}(\mathbf{t})} \tag{4.12}$$

where

$$\widehat{\nu}_{\widehat{\eta}_n,n}(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^{n} \left[\bigvee_{j=1}^{d} \left\{ H_{n,j}^{(\widehat{\eta}_n)}(Z_{i,j}) \right\}^{1/\widehat{\eta}_n t_j} - \frac{1}{d} \sum_{j=1}^{d} \left\{ H_{n,j}^{(\widehat{\eta}_n)}\,(Z_{i,j}) \right\}^{1/\widehat{\eta}_n t_j} \right],$$

$$\widehat{c}_{\widehat{\eta}_n,n}(\mathbf{t}) = \frac{1}{n\,d} \sum_{i=1}^{n} \sum_{j=1}^{d} \left\{ H_{n,j}^{(\widehat{\eta}_n)}\,(Z_{i,j}) \right\}^{1/\widehat{\eta}_n t_j}, \tag{4.13}$$

where, for all $j \in \{1, \ldots, d\}$ and $a > 0$,

$$H_{n,j}^{(a)}(Z_{i,j}) = H_{n,j}(Z_{i,j}) \left\{ \frac{1+a}{a} \frac{1}{n} \sum_{k=1}^{n} H_{n,j}^{1/a}(Z_{k,j}) \right\}^{-a},$$

and

$$H_{n,j}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(Z_{i,j} \le x)}.$$

Note that (4.13) stems from the fact that c_η defined in Proposition 4.3.2 can be reexpressed as

$$c_\eta(\mathbf{t}) = \mathrm{E}\left[\frac{1}{d} \sum_{j=1}^{d} \{G_{\eta,j}(Z_j)\}^{1/\eta t_j} \right]$$

and thus in (4.13) we use the empirical counterpart. Another option would have been to replace η by an estimator in (4.11).

In the following, we will also make use of a smoothness condition on copulas.

Condition 4.3.4 The distribution G_η has an η-extreme-value copula C_η, i.e.

$$C_\eta(\mathbf{u}) = G_\eta(G_{\eta,1}^{\leftarrow}(u_1), \ldots, G_{\eta,d}^{\leftarrow}(u_d)),$$

for $\mathbf{u} \in [0,1]^d$, such that, for all $j \in \{1, \ldots, d\}$, the partial derivative $\dot{C}_{\eta,j}(\mathbf{u}) = \partial C_\eta / \partial u_j(\mathbf{u})$ exists and is continuous on $\{\mathbf{u} \in [0,1]^d : 0 < u_j < 1\}$.

We can now state our main result on the convergence of a rescaled version of $\widehat{A}_{\widehat{\eta}_n,n}$.

Theorem 4.3.5 *Let $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be iid copies of $\mathbf{Z}$ with distribution G_η satisfying Condition 4.3.4. Let $\widehat{A}_{\widehat{\eta}_n,n}$ be the estimator for A_η given in* (4.12). *Under Condition 4.3.3 (i), we have that, as $n \to \infty$,*

$$\|\widehat{A}_{\widehat{\eta}_n,n} - A_\eta\|_\infty \to 0 \ a.s. \tag{4.14}$$

Under Conditions 4.3.3 (ii) and 4.3.4, it holds that as $n \to \infty$

$$\begin{aligned} &\sqrt{n}\,\{\widehat{A}_{\widehat{\eta}_n,n}(t) - A_\eta(t)\}_{\mathbf{t} \in \mathcal{S}_d} \\ &\qquad \rightsquigarrow \left[-\frac{\{1+\eta A_\eta(\mathbf{t})\}^2}{\eta} \int_0^1 \mathbb{A}_\eta(v^{\eta t_1}, \ldots, v^{\eta t_d})\,\mathrm{d}v \right]_{\mathbf{t} \in \mathcal{S}_d}, \end{aligned} \tag{4.15}$$

in $\ell^\infty(\mathcal{S}_d)$, *where* $\mathbb{A}_\eta$ *is a stochastic process defined, for all* $\mathbf{u} \in [0,1]^d$, *as*

$$\mathbb{A}_\eta(\mathbf{u}) = \mathbb{B}_\eta(\mathbf{u}) - \sum_{j=1}^{d} \dot{C}_{\eta,j}(\mathbf{u})\mathbb{B}_\eta(1,\ldots,1,u_j,1,\ldots,1),$$

with $\mathbb{B}_\eta$ *a* C_η*-Brownian bridge, i.e., a zero-mean Gaussian process on* $[0,1]^d$ *with continuous sample paths and covariance function given, for all* $\mathbf{u}$, $\mathbf{v} \in [0,1]^d$, *by*

$$\operatorname{cov}\{\mathbb{B}_\eta(\mathbf{u}), \mathbb{B}_\eta(\mathbf{v})\} = C_\eta(\mathbf{u} \wedge \mathbf{v}) - C_\eta(\mathbf{u})C_\eta(\mathbf{v}).$$

4.3.3 Examples of estimators satisfying Condition 4.3.3

Our $\eta-$Pickands dependence function requires an estimator of η which satisfies Condition 4.3.3. Below, two examples of such estimators are proposed.

Example 1 Let $Z^* = \max(Z_1, \ldots, Z_d)$, where $\mathbf{Z} \overset{d}{\sim} G_\eta$, with G_η as in (4.8). Then, for any $y > 0$, the distribution of Z^* is $G_\eta(y) = G_\eta(y, \ldots, y)$. This distribution is a member of the two-parameter Fréchet family of distributions. Let $\widehat{\eta}_n$ be the Maximum Likelihood (ML) estimator. By Propositions 3.1 and 3.3 in [7], it follows that the ML estimator satisfies Conditions 4.3.3(i) and 4.3.3(ii)(a).

Example 2 Let $\widehat{\eta}_n$ be the Generalized Probability Weighted Moment (GPWM) estimator of η introduced by [37]. The following theorem shows that the GPWM estimator admits a stochastic representation implying that Condition 4.3.3 (ii)(b) is satisfied. The almost-sure consistency of $\widehat{\eta}_n$ is a direct consequence.

Theorem 4.3.6 *Let* $\widehat{\eta}_n$ *be the GPWM estimator. For* $a, b \in \mathbb{N}_+$ *and* $Q_\eta(u) = G_\eta^{\leftarrow}(u)$, *introduce the parameter*

$$\mu_{a,b} := \int_0^1 Q_\eta(u) u^a (-\ln u)^b \mathrm{d}u$$

and for $u \in (0,1)$ *the two functions*

$$\gamma(u) = \mu_{1,2} u(-\ln u) - \mu_{1,1} u(-\ln u)^2, \quad \varphi(u) = \frac{1}{\eta\, V_\eta^\eta(1,\ldots,1)}\, u\,(-\ln u)^{1+\eta}.$$

Then,

$$\sqrt{n}\,(\widehat{\eta}_n - \eta) = -\frac{2}{\mu_{1,1}^2} \int_0^1 \mathbb{H}_n\{Q_\eta(u), \ldots, Q_\eta(u)\}\, \frac{\gamma(u)}{\varphi(u)} \mathrm{d}u + \mathrm{o}(1) \;\; a.s. \qquad (4.16)$$

Consequently, as $n \to \infty$,

$$\widehat{\eta}_n \to \eta \ a.s., \quad \sqrt{n}(\widehat{\eta}_n - \eta) \rightsquigarrow -\frac{2}{\mu_{1,1}^2} \int_0^1 \mathbb{H}\{Q_\eta(u), \dots, Q_\eta(u)\} \frac{\gamma(u)}{\varphi(u)} \, \mathrm{d}u,$$

where $\mathbb{H}$ *is a tight centered Gaussian process on* $(0, \infty]^d$, *with covariance function given, for all* $\mathbf{z}, \mathbf{y} \in (0, \infty]^d$, *by*

$$\operatorname{cov}\{\mathbb{H}(\mathbf{z}), \mathbb{H}(\mathbf{y})\} = G_\eta(\mathbf{z} \wedge \mathbf{y}) - G_\eta(\mathbf{z}) G_\eta(\mathbf{y}).$$

4.3.4 Simulation

The performances of the $\widehat{A}_{\widehat{\eta}_n, n}$ are illustrated in a simulation study with two different experiments.

First experiment: We consider the bivariate η-asymmetric logistic dependence model introduced by [58]. Such a dependence structure is characterized by the following features. The hidden spectral measure H_η has density given, for all $w \in (0, 1)$, by

$$h_\eta(w) = \frac{\eta - \psi}{\psi \eta^2 N_\varrho} \left\{ (\rho w)^{-1/\psi} + \left(\frac{1 - w}{\varrho} \right)^{-1/\psi} \right\}^{\psi/\eta - 2} \{w(1 - w)\}^{-(1 + 1/\psi)}, \quad (4.17)$$

where $N_\varrho = \varrho^{-1/\eta} + \varrho^{1/\eta} - (\varrho^{-1/\eta} + \varrho^{1/\eta})^{\psi/\eta}$ and $\psi \in (0, 1]$, $\varrho > 0$, $\eta \in (0, 1]$ are dependence parameters. This satisfies condition (4.6), i.e., in the case $d = 2$,

$$\eta^{-1} = \int_0^{1/2} w^{1/\eta} dH_\eta(w) + \int_{1/2}^1 (1 - w)^{1/\eta} dH_\eta(w).$$

The associated limiting distribution in (4.7) takes the form

$$G_\eta(y_1, y_2) = \begin{cases} \exp\left[-N_\varrho^{-1} \left\{ (\varrho y_1)^{-1/\psi} + \left(\frac{y_2}{\varrho} \right)^{-1/\psi} \right\}^{\psi/\eta} \right] & \text{for } \psi < \eta \\ 0, & \text{for } \psi \geq \eta, \end{cases} \quad (4.18)$$

where the degenerate case arises when H_η is infinite. In the sequel we focus on the case $\psi < \eta$ and for simplicity we consider $\varrho = 1$. Distribution (4.18) is the attractor for the distribution of normalized componentwise maxima obtained from a random vector whose survival function is

$$\begin{aligned}\mathbb{P}(X_1 > x_1, X_2 > x_2) \\ = \frac{\lambda u^{1/\eta}}{N_\varrho}\left[(\varrho x_1)^{-1/\eta} + \left(\frac{x_2}{\varrho}\right)^{-1/\eta} - \left\{(\varrho x_1)^{-1/\psi} + \left(\frac{x_2}{\varrho}\right)^{-1/\psi}\right\}^{\psi/\eta}\right]\end{aligned} \tag{4.19}$$

where $(x_1, x_2) \in [u, \infty) \times [u, \infty)$, with u being a high threshold and λ the joint threshold exceedance probability (see [57] for details). The survival function (4.19) satisfies (4.5) and constitutes an asymptotically independent joint probability model for any $\eta \in (0, 1)$, where the strength of the dependence, within asymptotic independence, increases for decreasing values of the parameter ψ. We call (4.18) and (4.19) the η-asymmetric logistic distribution and survival function, respectively.

We simulate n values from the η-asymmetric logistic distribution and we estimate the η-Pickands dependence function with $\widehat{A}_{\widehat{\eta}_n,n}$. We repeat these steps 1000 times and we compute a Monte Carlo approximation of the Mean Integrated Squared Error (MISE), i.e.,

$$\text{MISE}(\widehat{A}_{\widehat{\eta}_n,n}, A_\eta) = \text{E}\left[\int_{\mathcal{S}_d}\left\{\widehat{A}_{\widehat{\eta}_n,n}(\mathbf{t}) - A_\eta(\mathbf{t})\right\}^2 \mathrm{d}\mathbf{t}\right].$$

This study is done for different values of the sample size n and different values of the dependence parameter ψ. The results are summarized in Table 4.3. For each value of ψ, between the second and the fifth column the mean of the estimates for η obtained with the GPWM (first row) and ML (second row) estimator are reported, for increasing sample size. In parentheses is the standard deviation. Between the sixth and ninth columns the approximated MISE is reported. Accurate estimates are obtained with all the dependence levels. GPWM and ML estimators provide similar results, although those of the former seem slightly better. According to the MISE, the better performances are obtained with stronger dependence levels. For every dependence level the estimation accuracy increases as the sample size increases (as expected).

Second experiment: We illustrate the performances of the estimator $\widehat{A}_{\widehat{\eta}_n,n}$ under a more realistic scenario. We simulate $n \times 366$ independent observations from a distribution whose survival function is given in (4.19). To do this we use the

Table 4.3: Estimates (standard deviation) of η and MISE for the η-Pickands dependence function, based on a bivariate η-asymmetric logistic dependence model with $\eta = 0.7$. The first line corresponds to the GPWM method, whereas the second line is the ML method.

	$\widehat{\eta}_n$				MISE$(\widehat{A}_{\widehat{\eta}_n,n}, A_\eta)$			
	n				n			
ψ	25	50	100	200	25	50	100	200
0.1	0.661(0.115)	0.678(0.084)	0.690(0.062)	0.695(0.043)	0.0111	0.0037	0.0013	0.0005
	0.800(0.201)	0.763(0.128)	0.741(0.085)	0.728(0.055)	0.0110	0.0036	0.0013	0.0005
0.2	0.667(0.116)	0.679(0.084)	0.688(0.062)	0.692(0.044)	0.0480	0.0195	0.0088	0.0041
	0.807(0.204)	0.761(0.128)	0.740(0.088)	0.724(0.057)	0.0457	0.0187	0.0086	0.0040
0.3	0.665(0.116)	0.680(0.087)	0.692(0.064)	0.696(0.046)	0.1176	0.0542	0.0262	0.0133
	0.811(0.211)	0.768(0.130)	0.745(0.087)	0.730(0.059)	0.1133	0.0527	0.0256	0.0131
0.4	0.673(0.114)	0.687(0.088)	0.694(0.062)	0.697(0.045)	0.2177	0.1021	0.0523	0.0260
	0.810(0.204)	0.770(0.129)	0.744(0.084)	0.729(0.057)	0.2118	0.1000	0.0514	0.0257
0.5	0.670(0.113)	0.684(0.085)	0.692(0.062)	0.695(0.044)	0.3602	0.1795	0.0952	0.0481
	0.805(0.201)	0.766(0.129)	0.742(0.085)	0.728(0.057)	0.3531	0.1765	0.0940	0.0476
0.6	0.670(0.115)	0.685(0.085)	0.691(0.062)	0.696(0.046)	0.4566	0.2252	0.1192	0.0556
	0.822(0.206)	0.778(0.127)	0.751(0.089)	0.734(0.063)	0.4758	0.2275	0.1156	0.0585

Table 4.4: Estimates (standard deviation) of η and MISE for the η-Pickands dependence function, based on componentwise maxima with approximate bivariate η-asymmetric logistic model with $\eta = 0.7$. The first line corresponds to the GPWM method, whereas the second line is the ML method.

	$\widehat{\eta}_n$				$\mathrm{MISE}(\widehat{A}_{\widehat{\eta}_n,n}, A_\eta)$			
	n				n			
ψ	25	50	100	200	25	50	100	200
0.1	0.668(0.115)	0.684(0.089)	0.692(0.061)	0.695(0.044)	0.0108	0.0034	0.0013	0.0005
	0.800(0.204)	0.764(0.128)	0.742(0.086)	0.730(0.060)	0.0106	0.0033	0.0013	0.0005
0.2	0.664(0.116)	0.681(0.086)	0.687(0.061)	0.693(0.045)	0.0456	0.0187	0.0088	0.0040
	0.810(0.213)	0.765(0.133)	0.743(0.091)	0.728(0.064)	0.0442	0.0183	0.0079	0.0039
0.3	0.670(0.120)	0.686(0.089)	0.696(0.063)	0.698(0.045)	0.1088	0.0563	0.0257	0.0119
	0.804(0.194)	0.766(0.119)	0.744(0.078)	0.732(0.055)	0.1080	0.0546	0.0255	0.0117
0.4	0.684(0.119)	0.699(0.091)	0.707(0.063)	0.711(0.045)	0.2265	0.1146	0.0593	0.0279
	0.829(0.209)	0.783(0.125)	0.759(0.082)	0.745(0.054)	0.2207	0.1129	0.0584	0.0276
0.5	0.711(0.119)	0.727(0.088)	0.734(0.063)	0.738(0.044)	0.3820	0.1970	0.1113	0.0657
	0.846(0.197)	0.806(0.122)	0.786(0.080)	0.773(0.049)	0.3819	0.1973	0.1112	0.0661
0.6	0.751(0.125)	0.766(0.094)	0.775(0.070)	0.781(0.050)	0.7105	0.4298	0.3018	0.2333
	0.898(0.201)	0.849(0.121)	0.831(0.082)	0.820(0.053)	0.7293	0.4387	0.3096	0.2381

algorithm described in Theorem 1.1 and Appendix B of [57]. The simulation procedure in [57] relies on the condition

$$\mathbb{P}(\mathbf{X} > \mathbf{x}) = \lambda\eta u^{1/\eta} \int_{(0,1)} \min\{w/x_1, (1-w)/x_2\}^{1/\eta} \, \mathrm{d}H_\eta(w),$$

for $\mathbf{x} > u\mathbf{1}$, where u and λ are as in the first experiment. This condition implies that, for every $x > u$,

$$1 - \mathbb{P}(X_1 > x, X_2 \leq u) = e^{-1/x} + \lambda\eta \int_{(0,1)} \min\left(\frac{w}{x/u}, 1-w\right)^{1/\eta} \mathrm{d}H_\eta(w),$$
$$1 - \mathbb{P}(X_1 \leq u, X_2 > x) = e^{-1/x} + \lambda\eta \int_{(0,1)} \min\left(w, \frac{1-w}{x/u}\right)^{1/\eta} \mathrm{d}H_\eta(w).$$

The values of u and λ must be selected in such a way that both functions of y above are monotonically increasing. When the density of H_η is given by (4.17) with $\varrho = 1$ and $\eta = 0.7$, the monotonicity conditions are satisfied for every $\psi < \eta$ by setting $u = 10$ and $\lambda = 1 - e^{-0.1} - 0.02$.

With simulated data we compute $\widehat{b}_{366}$, that is the empirical $(1-1/366)$-quantile of the minimum between pairs of all observations. For each block of 366 observations we compute the componentwise maxima using $\varepsilon = Q(0.07)/\widehat{b}_{366}$, where $Q(0.07)$ is the ninety-third percentile of a unit Fréchet distribution, i.e., by retaining only the pairs that are both greater than $\varepsilon\widehat{b}_{366}$. We standardize the maxima by dividing them by $\widehat{b}_{366}$. With the n normalized maxima we estimate the η-Pickands dependence function by $\widehat{A}_{\widehat{\eta}_n,n}$. We repeat these steps 1000 times and we compute an approximation of the MISE. Table 4.4 collects the results.

We see that the estimation results are similar to those obtained in Table 4.3. We mention that in each block of 366 observations the componentwise maxima are computed, after the truncation, on average on approximately 17 pairs. Although maxima are obtained with a small number of observations, the estimation results suggest that they are enough to obtain accurate estimates. Estimates are less accurate for $\psi = 0.6$ for the following reason. The simulation method for maxima produces observations that are approximately drawn from the non-degenerate distribution G_η in (4.18), provided that $\psi < \eta$, since G_η is a degenerate distribution for $\psi \geq \eta$. Furthermore, for this example it can be empirically verified that G_η provides a very accurate approximation for the distribution of the simulated maxima when $\psi < 0.6$. Instead, whenever ψ is close to η (a case that resembles the degenerate case), e.g., $0.6 \leq \psi < 0.7 = \eta$, the quality of the approximation deteriorates. In this case the mismatch between G_η and the distribution of simulated maxima is no longer negligible, thus affecting the estimation results.

Finally, note that the asymptotic properties of our estimator established in Theorem 4.3.5 are no longer valid in this experiment, although our estimator still performs well. Indeed, they should be re-established under the assumption that the data belong to the domain of attraction of G_η. In that case, additional technical difficulties arise, whose solution would make the program of this work far too ample. We therefore defer such refinements to future work.

4.4 Discussion

The framework for modeling the dependence within asymptotic independence, based on componentwise maxima, relies on the assumption that H_η is a finite measure (see Section 4.3) Before applying our estimation method, it is desirable to check somehow whether such an assumption is reasonable for the available data. To this end, we propose a diagnostic tool. We motivate it on the basis of the following discussion. For simplicity we focus on the bivariate case although our proposal is easily extendable to higher dimensions. Let $\mathbf{X}$ be a two-dimensional random vector defined as in Section 4.3. Define, for $j \in \{1,2\}$, $s > 0$, and $x > 1$,

$$\begin{aligned}\overline{\overline{F}}_j(x) &= \lim_{s\to\infty} \mathbb{P}(X_j > sx \,|\mathbf{X} > s\mathbf{1}) \\ &= \eta \left\{ x^{-1/\eta} \int_0^{x/(1+x)} w^{1/\eta} \mathrm{d}H_\eta(w) + \int_{x/(1+x)}^1 (1-w)^{1/\eta} \mathrm{d}H_\eta(w) \right\} \\ &\leq 2\eta M x^{-1/\eta},\end{aligned}$$

where $M = H_\eta\{(0,1)\} < \infty$, and

$$\overline{\overline{F}}_{\min}(x) = \lim_{s\to\infty} \mathbb{P}\{\min(X_1, X_2) > sx \,|\mathbf{X} > s\mathbf{1}\}.$$

Then, it follows that

$$\begin{aligned}1 &\leq 2\eta M \min\left\{ x^{-1/\eta}/\overline{\overline{F}}_1(x), x^{-1/\eta}/\overline{\overline{F}}_2(x) \right\} \\ &= 2\eta M \min\left\{ \overline{\overline{F}}_{\min}(x)/\overline{\overline{F}}_1(x), \overline{\overline{F}}_{\min}(x)/\overline{\overline{F}}_2(x) \right\}.\end{aligned}$$

Consequently marginal survival functions $\overline{\overline{F}}_j$, $j \in \{1,2\}$, heavier than $\overline{\overline{F}}_{\min}$, i.e., $\overline{\overline{F}}_{\min}(y)/\overline{\overline{F}}_j(y) \to 0$ as $y \to \infty$, $j \in \{1,2\}$, provide empirical evidence against the hypothesis that H_η is finite.

On the contrary, evidence in favor of a finite H_η is provided by the conditions $\overline{\overline{F}}_{\min}(x)/\overline{\overline{F}}_j(x) \to c_j$ as $y \to \infty$, where c_j, $j \in \{1,2\}$, are positive constants. On this basis, we suggest plotting the curves

$$x \mapsto \widehat{r}_j(x) = \frac{\widehat{\overline{\overline{F}}}_{\min}(x)}{\widehat{\overline{\overline{F}}}_j(x)}, \quad x \in [1, m^*/s],$$

where, for $j \in \{1,2\}$ and $x > 1$,

$$\widehat{\overline{\overline{F}}}_j(x) = \frac{1}{n_s}\sum_{i=1}^{n_s} \mathbb{1}_{(x_{i,j}>sx)}, \quad \widehat{\overline{\overline{F}}}_{\min}(x) = \frac{1}{n_s}\sum_{i=1}^{n_s} \mathbb{1}_{(m_i>sx)},$$

$x_{i,j}$ are observations larger than a positive threshold s, $m_i = \min(x_{i,1}, x_{i,2})$, n_s is the number of m_i and m^* is the (n_s-1)th order statistic of the sample $m_1, \ldots, m_{n_s}$. When H_η is finite, for $j \in \{1,2\}$, $\widehat{r}_j(x)$ approaches a positive constant as $y \to m^*/s$, whereas in the infinite mass case, it decreases toward zero.

We illustrate the diagnostic tool with some examples. We draw samples of 500×366 values from six different models that satisfy equations (4.5) and (4.6). We consider three models with an η-asymmetric logistic (η-AL) survival function and parameters $\psi = 0.1$, $\psi = 0.4$ and $\psi = 0.6$, respectively, while $u = 10$, $\lambda = 1 - e^{-0.1} - 0.02$, $\varrho = 1$, $\eta = 0.7$ are the same for all the three cases.

In these examples H_η is finite and both $\overline{\overline{F}}_{\min}$ and $\overline{\overline{F}}_j$, $j \in \{1,2\}$, behave approximately as $x^{-1/\eta}$ for large values of x. Figure 4.3 displays in the first line the plots of $\widehat{r}_j$, obtained using $s = Q(0.07)$, where $Q(0.07)$ is the 93rd percentile of the unit Fréchet distribution. As expected, for large values of x, $\widehat{r}_j(x)$ stays away from zero and it approaches 1 when $\psi = 0.1, 0.4$ and a smaller constant when $\psi = 0.6$ (value close to $\psi = 0.7$ with which H_η is infinite). We also consider an η-asymmetric logistic model with $u = 8$, $\lambda = 1 - e^{-1/8} - 0.04$, $\varrho = 1$, $\eta = 0.7$ and $\psi = 0.8$, a bivariate standard Gaussian distribution with $\rho = 0.5$ and a bivariate inverted symmetric logistic (SL) model with $\psi = 0.8$ (see Section 4.2.3). In the latter two models the marginal distributions of the data are transformed into unit Fréchet. In these three cases H_η is infinite. Furthermore, for large x, $\overline{\overline{F}}_{\min}$ behaves approximately as $x^{-1/\eta}$ with $\eta = 0.7$, $\eta = (1+\rho)/2 = 0.75$ and $\eta = 2^{-\psi} \simeq 0.57$,

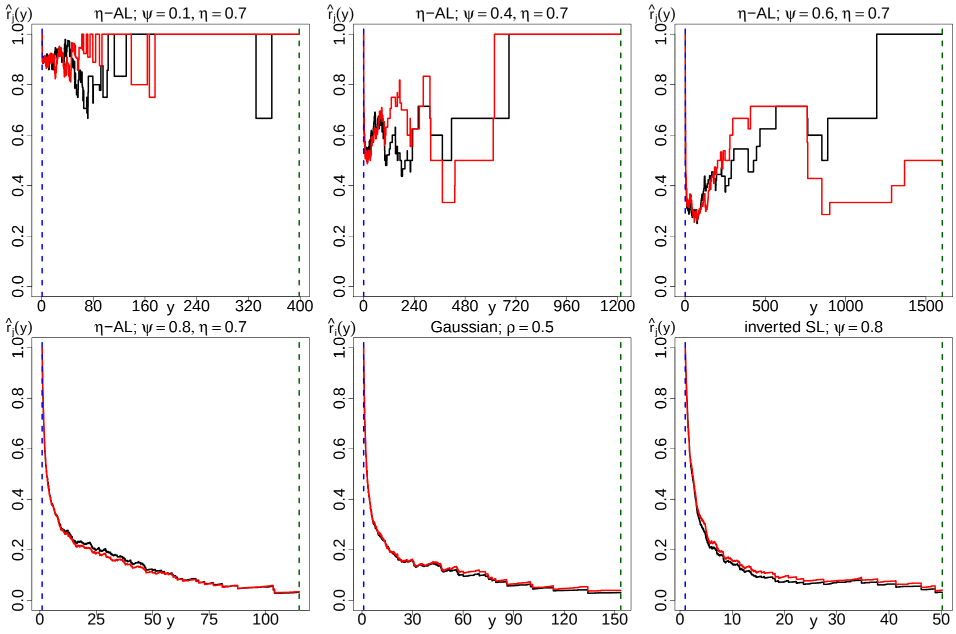

Figure 4.3: Diagnostic plots to check the finiteness of H_η. The left-hand vertical dotted line crosses the abscissas at 1, while the right-hand one at the value m^*/s. The red line is the case $j = 1$ and the black line the case $j = 2$.

respectively. Moreover, $\overline{\overline{F}}_1$, $\overline{\overline{F}}_2$ behave approximately as $x^{-1/k}$ with $k = \psi = 0.8$, $k = 1 + \rho = 1.5$ and $k = 2^{1-\psi} \simeq 1.15$, respectively. For these three examples the diagnostic plots are displayed in the second line of Figure 4.3. As expected, $\widehat{r}_j(x)$ goes to zero for large values of x.

The procedure for inferring A_η discussed in Section 4.3.2, when possible, provides useful means to extrapolate the probability of joint high thresholds exceedances as we describe next. For simplicity we focus on the bivariate case. By (4.5), (4.9) and the definition of the η-Pickands dependence function we have that the approximation

$$\begin{aligned}
&\mathbb{P}(X_1 > sx_1, X_2 > sx_2 \,|\mathbf{X} > s\mathbf{1}) \\
&\quad\approx \quad \eta \int_0^1 \min\left(\frac{w}{x_1}, \frac{1-w}{x_2}\right)^{1/\eta} \mathrm{d}H_\eta(w) \\
&\quad= \quad -\{\ln G_{\eta,1}(x_1) + \ln G_{\eta,2}(x_2)\} \\
&\qquad\quad \times \left[1 - A_\eta\left\{\frac{\ln G_{\eta,1}(x_1)}{\ln G_{\eta,1}(x_1) + \ln G_{\eta,2}(x_2)}, \frac{\ln G_{\eta,2}(x_2)}{\ln G_{\eta,1}(x_1) + \ln G_{\eta,2}(x_2)}\right\}\right]
\end{aligned}$$

holds for a large threshold s and $x_1, x_2 > 1$. Set $s = \widehat{b}_n$, where $\widehat{b}_n$ is the empirical $(1-1/n)$-quantile of the sequence $\min(x_{i,1}, x_{i,2})$ with $i \in \{1, \dots, n\}$, with $(x_{i,1}, x_{i,2})$ that are independent realizations of $\mathbf{X}$, see (4.7). Then, the above probability can be approximated by

$$\begin{aligned} &- \{\ln H_{n,1}(x_1) + \ln H_{n,2}(x_2)\} \\ &\times \left[1 - \widehat{A}_{\widehat{\eta}_n,n} \left\{ \frac{\ln H_{n,1}(x_1)}{\ln H_{n,1}(x_1) + \ln H_{n,2}(x_2)}, \frac{\ln H_{n,2}(x_2)}{\ln H_{n,1}(x_1) + \ln H_{n,2}(x_2)} \right\} \right], \end{aligned} \tag{4.20}$$

where $H_{n,1}$, $H_{n,2}$ are the empirical distribution functions; see Section 4.3.2.

We illustrate the extrapolation of the probability of high thresholds exceedances with two examples. We simulate 500×366 independent realizations from two distributions with an η-asymmetric logistic survival function and parameters $\psi = 0.1$, $\psi = 0.4$, respectively, while $u = 10$, $\lambda = 1 - e^{-0.1} - 0.02$, $\varrho = 1$, $\eta = 0.7$ are the same for both the cases. Then, we obtain the sample of maxima, using $\widehat{b}_{366}$, $\varepsilon = Q(0.07)/\widehat{b}_{366}$ (see Section 4.3.1 and the second experiment of Section 4.3.4 for details) and we estimate the Pickands dependence function by $\widehat{A}_{\widehat{\eta}_{500},500}$, where $\widehat{\eta}_{500}$ is the GPWM estimator of η. For $x_1, x_2 \in [1, 10]$, we extrapolate the probability of joint high thresholds exceedances by exploiting (4.20).

Figure 4.4 displays the estimated probabilities for the two models. The left and right panels report the results for the cases $\psi = 0.1$ and $\psi = 0.4$, respectively. To go further with this idea, a topic of interest would be to establish the asymptotic properties of the estimator defined in (4.20). This is outside the scope of the present work but it will lead to further investigations.

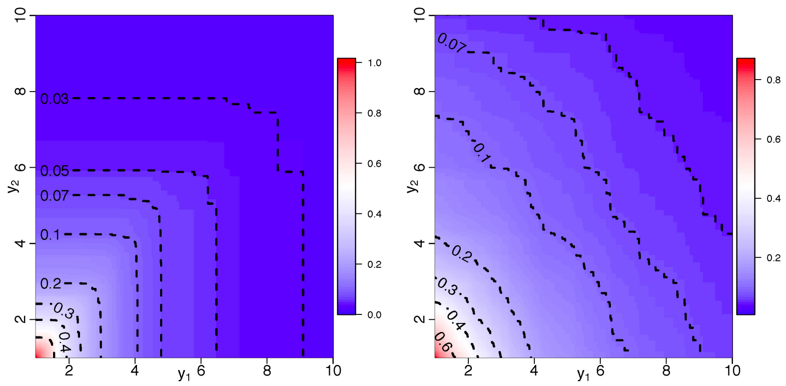

Figure 4.4: Estimated probabilities of joint high thresholds exceedances with $\psi = 0.1$ (left panel) and $\psi = 0.4$ (right panel).

4.5 Proofs

4.5.1 Some properties of $\widehat{A}_n$

Note that, for all $j \in \{1, \ldots, d\}$,

$$\begin{aligned}\widehat{\nu}_n(\mathbf{e}_j) &= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{G_{n,j}(Y_{i,j})}{2\,n^{-1}\sum_{k=1}^{n} G_{n,j}(Y_{k,j})} - \frac{1}{d}\frac{G_{n,j}(Y_{i,j})}{2\,n^{-1}\sum_{k=1}^{n} G_{n,j}(Y_{k,j})}\right\}\\ &= \frac{1}{2} - \frac{1}{2d}.\end{aligned}$$

Therefore, $\widehat{A}_n(\mathbf{e}_j) = 1$ for all $j \in \{1, \ldots, d\}$.

The distribution function of the iid random variables $Y_{1,j}, \ldots, Y_{n,j}$ with $j \in \{1, \ldots, d\}$ being continuous, almost surely there are no ties and thus

$$G_{n,j}^{(1)}(Y_{i,j}) = G_{n,j}(Y_{i,j})\left\{\frac{2}{n}\sum_{k=1}^{n} G_{n,j}(Y_{k,j})\right\}^{-1} = \frac{n}{n+1}\,G_{n,j}(Y_{i,j}).$$

Then, with simple adjustments of the proof of Theorem 2.4 in [46], the weak convergence of $\widehat{A}_n$ and its almost sure consistency follow. ∎

4.5.2 Proof of Proposition 4.3.1

The definition of V_η and L_η entails that, for all $t \in \mathcal{S}_d$,

$$A_\eta(\mathbf{t}) = \eta \int_{\mathcal{S}_d} \max\left(\frac{t_1 w_1^{1/\eta}}{\sigma_{\eta,1}}, \ldots, \frac{t_d w_d^{1/\eta}}{\sigma_{\eta,d}}\right) \mathrm{d}H_\eta(\mathbf{w}).$$

Then, Property 1 follows by the definition of $\sigma_{\eta,j}$ given in (4.9). When $\eta = 1$, according to Section 4.3, we have

$$\lim_{n\to\infty} \frac{\mathbb{P}(\mathbf{X} > n\mathbf{x})}{\mathbb{P}(\mathbf{X} > n\mathbf{1})} = \int_{\mathcal{S}_d} \bigwedge_{j=1}^{d} \left(\frac{w_j}{x_j}\right) \mathrm{d}H_1(\mathbf{w}).$$

Now, when asymptotic dependence holds, this limit can also be rephrased with the classical theory (see, e.g., Chapter 6 in [19]), where

$$\lim_{n\to\infty} \frac{\mathbb{P}(\mathbf{X} > n\mathbf{x})}{\mathbb{P}(\mathbf{X} > n\mathbf{1})} = \frac{d \int_{\mathcal{S}_d} \bigwedge_{j=1}^{d} (w_j/x_j) \mathrm{d}H(\mathbf{w})}{R(1, \ldots, 1)},$$

with H as in Section 1.2 and R defined on pages 218 and 225 in [19]. Thus Property 2 follows from the relation $d^{-1}R(1,\ldots,1)\mathrm{d}H_1(\mathbf{w}) = \mathrm{d}H(\mathbf{w})$, valid for all $\mathbf{w} \in \mathcal{S}_d$, and $\sigma_{1,j} = 1/R(1,\ldots,1)$ for all $j \in \{1,\ldots,d\}$.

For every $\mathbf{t} \in \mathcal{S}_d$ we have

$$\eta \int_{\mathcal{S}_d} \max\left(\frac{t_1 w_1^{1/\eta}}{\sigma_{\eta,1}}, \ldots, \frac{t_d w_d^{1/\eta}}{\sigma_{\eta,d}}\right) \mathrm{d}H_\eta(\mathbf{w}) \le \eta \int_{\mathcal{S}_d} \sum_{j=1}^{d} \left(\frac{t_j w_j^{1/\eta}}{\sigma_{\eta,j}}\right) \mathrm{d}H_\eta(\mathbf{w})$$
$$= 1,$$

from which the upper bound in Property 3 follows. To derive the lower bound, it is sufficient to remark that for every $\mathbf{t} \in \mathcal{S}_d$, we have

$$\begin{aligned}
&\eta \int_{\mathcal{S}_d} \max\left(\frac{t_1 w_1^{1/\eta}}{\sigma_{\eta,1}}, \ldots, \frac{t_d w_d^{1/\eta}}{\sigma_{\eta,d}}\right) \mathrm{d}H_\eta(\mathbf{w}) \\
&\quad \ge \bigvee_{1\le i<j\le d} \left\{\eta \int_{\mathcal{S}_d} \max\left(\frac{t_i w_i^{1/\eta}}{\sigma_{\eta,i}}, \frac{t_j w_j^{1/\eta}}{\sigma_{\eta,j}}\right) \mathrm{d}H_\eta(\mathbf{w})\right\} \\
&\quad = \bigvee_{1\le i<j\le d} \left\{t_i + t_j - \eta \int_{\mathcal{S}_d} \min\left(\frac{t_i w_i^{1/\eta}}{\sigma_{\eta,i}}, \frac{t_j w_j^{1/\eta}}{\sigma_{\eta,j}}\right) \mathrm{d}H_\eta(\mathbf{w})\right\} \\
&\quad \ge \bigvee_{1\le i<j\le d} \{t_i + t_j - \min(t_i, t_j)\} = \bigvee_{1\le j\le d} t_j.
\end{aligned}$$

Finally, the convexity in Property 4 can been shown similarly to the convexity of A. ∎

4.5.3 Proof of Proposition 4.3.2

For all $\eta \in (0,1]$ and $\mathbf{t} \in \mathcal{S}_d$, set, for all $\mathbf{u} \in [0,1]^d$,

$$\nu_\eta(\mathbf{u};\mathbf{t}) = \bigvee_{j=1}^{d} u_j^{1/\eta t_j} - \frac{1}{d}\sum_{j=1}^{d} u_j^{1/\eta t_j}.$$

By convention $u^{1/\eta t} = 0$ when $t = 0$ and $u \in [0,1]$. By Lemma A.1 in [46] we have

$$\begin{aligned}
\nu_\eta(\mathbf{t}) &= \int_{[0,1]^d} \nu_\eta(\mathbf{u};\mathbf{t})\mathrm{d}C_\eta(\mathbf{u}) &(4.21)\\
&= \frac{1}{d}\sum_{j=1}^{d}\int_0^1 C_\eta(1,\ldots,1,v^{\eta t_j},1,\ldots,1)\mathrm{d}v - \int_0^1 C_\eta(v^{\eta t_1},\ldots,v^{\eta t_d})\mathrm{d}v\\
&= \frac{1}{d}\sum_{j=1}^{d}\int_0^1 v^{\eta t_j}\mathrm{d}v - \int_0^1 v^{\eta A_\eta(\mathbf{t})}\mathrm{d}v\\
&= \frac{1}{d}\sum_{j=1}^{d}\frac{1}{1+\eta\, t_j} - \frac{1}{1+\eta\, A_\eta(\mathbf{t})}.
\end{aligned}$$

Result (4.10) follows by solving the above equality for A_η. ∎

4.5.4 Proof of Theorem 4.3.5

We start with some notation. Let $\widehat{\mathbb{C}}_n = \sqrt{n}\,(\widehat{C}_n - C_\eta)$, where $\widehat{C}_n$ is the empirical copula defined, for all $\mathbf{u} \in [0,1]^d$, by

$$\widehat{C}_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(\widehat{\mathbf{U}}_i \le \mathbf{u})},$$

with $\widehat{\mathbf{U}}_i = (H_{n,1}(Z_{i,1}),\ldots,H_{n,d}(Z_{i,d}))$. Define now, for all $\mathbf{t} \in \mathcal{S}_d$,

$$\begin{aligned}
M(\cdot,\mathbf{t}) &= 1 - \int_0^1 C_\eta(v^{\cdot\, t_1},\ldots,v^{\cdot\, t_d})\mathrm{d}v, &(4.22)\\
\widehat{M}_n(\cdot,\mathbf{t}) &= 1 - \int_0^1 \widehat{C}_n(v^{\cdot\, t_1},\ldots,v^{\cdot\, t_d})\mathrm{d}v. &(4.23)
\end{aligned}$$

We will prove Theorem 4.3.5 with $H_{n,j}^{(\widehat{\eta}_n)}$ in $\widehat{\nu}_{\widehat{\eta}_n,n}$ and $\widehat{c}_{\widehat{\eta}_n,n}$ replaced by $H_{n,j}$. Indeed, this slight modification has no impact on the convergences (4.14) and (4.15) since

$$H_{n,j}^{(\widehat{\eta}_n)}(Z_{i,j}) = H_{n,j}(Z_{i,j}) \left\{ 1 + \frac{1+\widehat{\eta}_n}{\widehat{\eta}_n} O\left(\frac{1}{n}\right) \right\}^{-\widehat{\eta}_n} \equiv H_{n,j}(Z_{i,j})\, e_n^{-1},$$

and the terms in (4.22) and (4.23) can be slightly changed by replacing in the integrals $v^{\cdot t_j}$ by $v^{\cdot t_j} e_n$ for each $j \in \{1, \dots, d\}$, without any impact. In view of this remark, we pursue the proof of Theorem 4.3.5 with $M(\cdot, \mathbf{t})$ and $\widehat{M}_n(\cdot, \mathbf{t})$ defined in (4.22) and (4.23) without taking care of the adjustment with e_n.

We start to prove (4.15). To this end, note that from (4.21) we have

$$M(\eta, \mathbf{t}) = \frac{\eta A_\eta(\mathbf{t})}{1 + \eta A_\eta(\mathbf{t})}$$

and thus the following decomposition holds

$$\begin{aligned}
&\sqrt{n}\,\{\widehat{A}_{\widehat{\eta}_n,n}(\mathbf{t}) - A_\eta(\mathbf{t})\} \\
&= \sqrt{n} \left\{ \frac{1}{\widehat{\eta}_n} \frac{\widehat{M}_n(\widehat{\eta}_n, \mathbf{t})}{1 - \widehat{M}_n(\widehat{\eta}_n, \mathbf{t})} - \frac{1}{\eta} \frac{M(\eta, \mathbf{t})}{1 - M(\eta, \mathbf{t})} \right\} \\
&= \frac{\sqrt{n}}{\widehat{\eta}_n} \left\{ \frac{\widehat{M}_n(\widehat{\eta}_n, \mathbf{t})}{1 - \widehat{M}_n(\widehat{\eta}_n, \mathbf{t})} - \frac{M(\eta, \mathbf{t})}{1 - M(\eta, \mathbf{t})} \right\} + \frac{M(\eta, \mathbf{t})}{1 - M(\eta, \mathbf{t})} \sqrt{n} \left(\frac{1}{\widehat{\eta}_n} - \frac{1}{\eta} \right) \\
&\equiv L_n(\mathbf{t}) + R_n(\mathbf{t}),
\end{aligned}$$

for all $\mathbf{t} \in \mathcal{S}_d$. We derive a tractable expression for L_n by means of the following three results.

Lemma 4.5.1 *We have the following decomposition*

$$\begin{aligned}
\sqrt{n}\,\{\widehat{M}_n(\widehat{\eta}_n, \mathbf{t}) - M(\eta, \mathbf{t})\} &= \sqrt{n}\,\{\widehat{M}_n(\eta, \mathbf{t}) - M(\eta, \mathbf{t})\} \\
&\quad + \sqrt{n}\,\{M(\widehat{\eta}_n, \mathbf{t}) - M(\eta, \mathbf{t})\} + o_p(1).
\end{aligned}$$

Proof. The proof uses arguments from [73]. Since

$$\begin{aligned}
&\sqrt{n}\,\{\widehat{M}_n(\widehat{\eta}_n, \mathbf{t}) - M(\eta, \mathbf{t})\} \\
&= \left[\sqrt{n}\,\{\widehat{M}_n(\widehat{\eta}_n, \mathbf{t}) - M(\widehat{\eta}_n, \mathbf{t})\} - \sqrt{n}\,\{\widehat{M}_n(\eta, \mathbf{t}) - M(\eta, \mathbf{t})\} \right] \\
&\quad + \sqrt{n}\{\widehat{M}_n(\eta, \mathbf{t}) - M(\eta, \mathbf{t})\} + \sqrt{n}\{M(\widehat{\eta}_n, \mathbf{t}) - M(\eta, \mathbf{t})\},
\end{aligned}$$

it remains to show that

$$\|\sqrt{n}\,\{\widehat{M}_n(\widehat{\eta}_n,\mathbf{t})-M(\widehat{\eta}_n,\mathbf{t})\}-\sqrt{n}\,\{\widehat{M}_n(\eta,\mathbf{t})-M(\eta,\mathbf{t})\}\|_\infty=o_p(1). \tag{4.24}$$

By Condition 4.3.3(ii) we have that $\sqrt{n}\,(\widehat{\eta}_n-\eta)$ is asymptotically tight. Thus, for every $\varepsilon>0$, there exists a compact set $K\equiv K_\varepsilon\subseteq\mathbb{R}$ such that

$$\liminf_{n\to\infty}\mathbb{P}\{\sqrt{n}\,(\widehat{\eta}_n-\eta)\in K\}>1-\varepsilon.$$

Furthermore, by the compactness of K, there exist $\delta>0$, $p=p(\delta)\in\mathbb{N}$ and $\{h_1,\ldots,h_p\}\subseteq K$ such that $K\subseteq\cup_{1\leq s\leq p}(h_s-\delta,h_s+\delta)$. Therefore,

$$\begin{aligned}\{\sqrt{n}(\widehat{\eta}_n-\eta)\in K\}&\subseteq\left\{\sqrt{n}(\widehat{\eta}_n-\eta)\subseteq\bigcup_{s=1}^{p}(h_s-\delta,h_s+\delta)\right\}\\&=\bigcup_{s=1}^{p}\left\{\widehat{\eta}_n\in\left(\eta+n^{-1/2}(h_s-\delta),\eta+n^{-1/2}(h_s+\delta)\right)\right\}.\end{aligned}$$

Consequently, it follows that, with probability at least $1-\varepsilon$,

$$\begin{aligned}&\|\sqrt{n}\,\{\widehat{M}_n(\widehat{\eta}_n,\mathbf{t})-M(\widehat{\eta}_n,\mathbf{t})\}-\sqrt{n}\,\{\widehat{M}_n(\eta,\mathbf{t})-M(\eta,\mathbf{t})\}\|_\infty\\&\leq\sup_{\mathbf{t}\in\mathcal{S}_d}\max_{s\in\{1,\ldots,p\}}\sup_{|h-h_s|<\delta}\\&\qquad|\sqrt{n}\,\{\widehat{M}_n(\eta_{n,h},\mathbf{t})-M(\eta_{n,h},\mathbf{t})\}-\sqrt{n}\,\{\widehat{M}_n(\eta,\mathbf{t})-M(\eta,\mathbf{t})\}|\\&\leq\sup_{\mathbf{t}\in\mathcal{S}_d}\max_{s\in\{1,\ldots,p\}}|\sqrt{n}\,\{\widehat{M}_n(\eta_{n,h_s},\mathbf{t})-M(\eta_{n,h_s},\mathbf{t})\}-\sqrt{n}\,\{\widehat{M}_n(\eta,\mathbf{t})-M(\eta,\mathbf{t})\}|\\&+\sup_{\mathbf{t}\in\mathcal{S}_d}\max_{s\in\{1,\ldots,p\}}\sup_{|h-h_s|<\delta}\\&\qquad|\sqrt{n}\,\{\widehat{M}_n(\eta_{n,h_s},\mathbf{t})-M(\eta_{n,h_s},\mathbf{t})\}-\sqrt{n}\,\{\widehat{M}_n(\eta_{n,h},\mathbf{t})-M(\eta_{n,h},\mathbf{t})\}|\\&\equiv I_{n,1}+I_{n,2},\end{aligned}$$

where $\eta_{n,\bullet}=\eta+n^{-1/2}\bullet$. Thus to show (4.24) it is sufficient to prove that both $I_{n,1}$ and $I_{n,2}$ tends to 0 in probability, as $n\to\infty$. Using (4.22) and (4.23) we obtain

$$\begin{aligned}I_{n,1}&=\sup_{\mathbf{t}\in\mathcal{S}_d}\max_{s\in\{1,\ldots,p\}}\left|\int_0^1\left\{\widehat{\mathbb{C}}_n(v^{\eta_{n,h_s}t_1},\ldots,v^{\eta_{n,h_s}t_d})\widehat{\mathbb{C}}_n(v^{\eta t_1},\ldots,v^{\eta t_d})\right\}\mathrm{d}v\right|\\&\leq\sup_{\mathbf{t}\in\mathcal{S}_d}\max_{s\in\{1,\ldots,p\}}\sup_{v\in(0,1)}\left|\widehat{\mathbb{C}}_n(v^{\eta_{n,h_s}t_1},\ldots,v^{\eta_{n,h_s}t_d})-\widehat{\mathbb{C}}_n(v^{\eta t_1},\ldots,v^{\eta t_d})\right|\end{aligned}$$

and

$$I_{n,2} = \sup_{\mathbf{t}\in\mathcal{S}_d} \max_{s\in\{1,\dots,p\}} \sup_{|h-h_s|<\delta} \left| \int_0^1 \left\{ \widehat{\mathbb{C}}_n(v^{\eta_{n,h_s}t_1}, \dots, v^{\eta_{n,h_s}t_d}) \widehat{\mathbb{C}}_n(v^{\eta_{n,h}t_1}, \dots, v^{\eta_{n,h}t_d}) \right\} \mathrm{d}v \right|$$
$$\leq \sup_{\mathbf{t}\in\mathcal{S}_d} \max_{s\in\{1,\dots,p\}} \sup_{|h-h_s|<\delta} \sup_{v\in(0,1)} \left| \widehat{\mathbb{C}}_n(v^{\eta_{n,h_s}t_1}, \dots, v^{\eta_{n,h_s}t_d}) - \widehat{\mathbb{C}}_n(v^{\eta_{n,h}t_1}, \dots, v^{\eta_{n,h}t_d}) \right|.$$

Now, for every $v \in (0,1)$ and small $\epsilon > 0$, the map $\varphi : (0,1) \to \ell^\infty([\eta-\epsilon, \eta+\epsilon]) : v \mapsto \varphi(v)$, defined by $(\varphi(v))(x) = v^x$, induces continuously differentiable functions on $[\eta-\epsilon, \eta+\epsilon]$ for every $v \in (0,1)$. The first derivative of such functions is $\{\dot\varphi(v)\}(x) = v^x \ln v$, whose absolute value is bounded above by $\xi_v = v^{\eta-\epsilon}|\ln v|$. Therefore, $\{\varphi(v)\}(x)$ is a Lipschitz function and for all $x, y \in [\eta-\epsilon, \eta+\epsilon]$, it satisfies the condition

$$|\{\varphi(v)\}(x) - \{\varphi(v)\}(y)| \leq \xi_v |x-y|.$$

Furthermore, there exists a positive constant ξ such that $\sup_{v\in(0,1)} \xi_v < \xi$, and thus for n sufficiently large ensuring that $\eta_{n,h}, \eta_{n,h_s} \in [\eta-\epsilon, \eta+\epsilon]$, we have

$$|v^{\eta_{n,h_s}t_j} - v^{\eta t_j}| \leq \xi\, |\eta - \eta_{n,h_s}| = \xi\, n^{-1/2}|h_s| \to 0$$
$$|v^{\eta_{n,h_s}t_j} - v^{\eta_{n,h}t_j}| \leq \xi\, |\eta_{n,h_s} - \eta_{n,h}| = \xi\, n^{-1/2}|h_s - h| \leq \xi\, \delta n^{-1/2} \to 0,$$

as $n \to \infty$, for every $\mathbf{t} \in \mathcal{S}_d$, indexes $s \in \{1,\dots,p\}$, $j \in \{1,\dots,d\}$ and for every $|h-h_s| < \delta$. These results imply that, as $n \to \infty$,

$$\sup_{\mathbf{t}\in\mathcal{S}_d} \max_{s\in\{1,\dots,p\}} \sup_{v\in(0,1)} \max_{j\in\{1,\dots,d\}} |v^{\eta_{n,h_s}t_j} - v^{\eta t_j}| \to 0, \tag{4.25}$$

and

$$\sup_{\mathbf{t}\in\mathcal{S}_d} \max_{s\in\{1,\dots,p\}} \sup_{|h-h_s|<\delta} \max_{j\in\{1,\dots,d\}} \sup_{v\in(0,1)} |v^{\eta_{n,h_s}t_j} - v^{\eta_{n,h}t_j}| \to 0. \tag{4.26}$$

Since the first partial derivative of C_η exists and is continuous on $\{\mathbf{u} \in [0,1]^d : 0 < u_j < 1\}$ for all $j \in \{1,\dots,d\}$, $\widehat{\mathbb{C}}_n \rightsquigarrow \mathbb{A}_\eta$ in $\ell^\infty([0,1]^d)$ as $n \to \infty$; see, e.g., [26, 24, 65]. Therefore the sequence $\widehat{\mathbb{C}}_n$ is asymptotically uniformly equicontinuous in probability; see Theorem 1.5.7 in [75]. Combining this result with (4.25) and (4.26) we can conclude that $I_{n,1}$ and $I_{n,2}$ converge to 0 in probability, as $n \to \infty$. Therefore (4.24) is established and the statement of Lemma 4.5.1 now follows. ∎

Lemma 4.5.2 *We have*

$$\sqrt{n}\,\{M(\widehat{\eta}_n,\mathbf{t})-M(\eta,\mathbf{t})\}=\frac{A_\eta(\mathbf{t})}{\{\eta A_\eta(\mathbf{t})+1\}^2}\,\sqrt{n}\,(\widehat{\eta}_n-\eta)+o_p(1).$$

Proof. Let $\varphi:((0,\infty),|\cdot|)\to(\ell^\infty(\mathcal{S}_d),\|\cdot\|_\infty):a\mapsto M(a,\cdot)$ be the map defined by

$$M(a,\cdot)=\frac{a\,A_\eta(\cdot)}{1+a\,A_\eta(\cdot)}.$$

Its Hadamard derivative at $\eta\in(0,1]$ is

$$h\mapsto\{\dot{\varphi}_\eta(h)\}=\frac{h\,A_\eta}{(\eta A_\eta+1)^2}.$$

Indeed, for every $\epsilon_n\downarrow 0$ and $h_n\to h\in(0,\infty)$, as $n\to\infty$, such that $\eta+\epsilon_n h_n\in(0,\infty)$, we have

$$\begin{aligned}
&\lim_{n\to\infty}\sup_{\mathbf{t}\in\mathcal{S}_d}\left|\frac{\{\varphi(\eta+\epsilon_n h_n)\}(\mathbf{t})-\{\varphi(\eta)\}(\mathbf{t})}{\epsilon_n}-\{\dot{\varphi}_\eta(h)\}(\mathbf{t})\right|\\
&=\lim_{n\to\infty}\sup_{\mathbf{t}\in\mathcal{S}_d}\left|\frac{1}{\epsilon_n}\left\{\frac{(\eta+\epsilon_n h_n)A_\eta(\mathbf{t})}{(\eta+\epsilon_n h_n)A_\eta(\mathbf{t})+1}-\frac{\eta A_\eta(\mathbf{t})}{\eta A_\eta(\mathbf{t})+1}\right\}-\frac{hA_\eta(\mathbf{t})}{\{\eta A_\eta(\mathbf{t})+1\}^2}\right|\\
&=\lim_{n\to\infty}\sup_{\mathbf{t}\in\mathcal{S}_d}\left|\frac{A_\eta(\mathbf{t})}{\eta A_\eta(\mathbf{t})+1}\right|\left|\frac{h_n}{(\eta+\epsilon_n h_n)A_\eta(\mathbf{t})+1}-\frac{h}{\eta A_\eta(\mathbf{t})+1}\right|\\
&\le\lim_{n\to\infty}d^2\,\frac{|h_n-h|+|hh_n|\epsilon_n}{(d+\eta)(d+\eta+\epsilon_n h_n)}=0.
\end{aligned}$$

The statement of Lemma 4.5.2 now follows from [74, Theorem 20.8] and Condition 4.3.3(ii). ∎

Lemma 4.5.3 *We have*

$$\sqrt{n}\left\{\frac{\widehat{M}_n(\widehat{\eta}_n,\mathbf{t})}{1-\widehat{M}_n(\widehat{\eta}_n,\mathbf{t})}-\frac{M(\eta,\mathbf{t})}{1-M(\eta,\mathbf{t})}\right\}=\{1+\eta A_\eta(\mathbf{t})\}^2\,\sqrt{n}\{\widehat{M}_n(\widehat{\eta}_n,\mathbf{t})-M(\eta,\mathbf{t})\}$$
$$+\,o_p(1).$$

Proof. The proof of this lemma is based on an application of the functional delta method after proving the Hadamard differentiability of the map $\varphi:\ell^\infty(\mathcal{S}_d)\mapsto\ell^\infty(\mathcal{S}_d):f\mapsto f/(1-f)$, with f in $\ell^\infty(\mathcal{S}_d)$, and the existence of the weak limit of $\sqrt{n}\,\{M(\widehat{\eta}_n,\cdot)-M(\eta,\cdot)\}$ in $\ell^\infty(\mathcal{S}_d)$.

First, we show that the Hadamard derivative of φ at $M = M(\eta, \cdot)$ is

$$h \mapsto \{\dot\varphi_M(h)\} = \frac{h}{(1-M)^2},$$

with h in $\ell^\infty(\mathcal{S}_d)$. Indeed, for every sequence $\epsilon_n \downarrow 0$ and $h_n \to h$ as $n \to \infty$, such that $M + \epsilon_n h_n$ in $\ell^\infty(\mathcal{S}_d)$, we have

$$\begin{aligned}
&\lim_{n\to\infty} \sup_{\mathbf{t}\in\mathcal{S}_d} \left| \frac{\{\varphi(M+\epsilon_n h_n)\}(\mathbf{t}) - \{\varphi(M)\}(\mathbf{t})}{\epsilon_n} - \{\dot\varphi_M(h)\}(\mathbf{t}) \right| \\
&= \lim_{n\to\infty} \sup_{\mathbf{t}\in\mathcal{S}_d} \left| \frac{1}{\epsilon_n} \left\{ \frac{M(\eta,\mathbf{t}) + \epsilon_n h_n(\mathbf{t})}{1 - M(\eta,\mathbf{t}) - \epsilon_n h_n(\mathbf{t})} - \frac{M(\eta,\mathbf{t})}{1-M(\eta,\mathbf{t})} \right\} - \frac{h(\mathbf{t})}{\{1-M(\eta,\mathbf{t})\}^2} \right| \\
&= \lim_{n\to\infty} \sup_{\mathbf{t}\in\mathcal{S}_d} \{1+\eta A_\eta(\mathbf{t})\}^2 \left| \frac{h_n(\mathbf{t}) - h(\mathbf{t}) + h(\mathbf{t})\epsilon_n h_n(\mathbf{t})\{1+\eta A_\eta(\mathbf{t})\}}{1 - \epsilon_n h_n(\mathbf{t})\{1+\eta A_\eta(\mathbf{t})\}} \right| \\
&\leq \lim_{n\to\infty} (1+\eta)^2 \, \frac{\|h_n - h\|_\infty + \epsilon_n \|h_n\, h\|_\infty (1+\eta)}{1 - \epsilon_n \|h_n\|_\infty (1+\eta)} \ = 0.
\end{aligned}$$

Then, combining Lemmas 4.5.1, 4.5.2 with Proposition 3.1 in [65], under Condition 4.3.3 (ii)(b) we have that

$$\sqrt{n}\,\{\widehat{M}_n(\widehat\eta_n,\cdot) - M(\eta,\cdot)\} = T_{n,1}(\cdot) + T_{n,2}(\cdot) + o_p(1),$$

where, for all $\mathbf{t} \in \mathcal{S}_d$,

$$\begin{aligned}
T_{n,1}(\mathbf{t}) = -\int_0^1 \Bigg\{ &\mathbb{C}_n(v^{t_1\eta}, \ldots, v^{t_d\eta}) \\
&- \sum_{j=1}^d \dot C_{\eta,j}(v^{t_1\eta}, \ldots, v^{t_d\eta}) \mathbb{C}_n(1,\ldots,1,v^{t_j\eta},1,\ldots,1) \Bigg\} \mathrm{d}v
\end{aligned}$$

and

$$T_{n,2} = \frac{A_\eta}{(1+\eta\, A_\eta)^2}\, \chi(\mathbb{H}_n).$$

For any $\mathbf{u} \in [0,1]^d$, $\mathbb{C}_n(\mathbf{u}) = \mathbb{H}_n\{G^{\leftarrow}_{\eta,1}(u_1), \ldots, G^{\leftarrow}_{\eta,d}(u_d)\}$, so both terms can be expressed as continuous transformations of the empirical process $\mathbb{H}_n$. Therefore, the weak convergence of $T_{n,1} + T_{n,2}$ follows from the continuous mapping theorem. A similar reasoning can be obtained if Condition 4.3.3(ii)(b) is replaced by Condition 4.3.3(ii)(a). In that case, we have the following decomposition

$$\sqrt{n}\,\{\widehat{M}_n(\widehat\eta_n,\cdot) - M(\eta,\cdot)\} \equiv T_{n,1} + \widetilde{T}_{n,2} + o_p(1),$$

where, for all $\mathbf{t} \in \mathcal{S}_d$,

$$T_{n,1}(\mathbf{t}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{W_{i,\mathbf{t}} - \mathrm{E}(W_{i,\mathbf{t}})\}, \quad \widetilde{T}_{n,2}(\mathbf{t}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{W}_{i,\mathbf{t}},$$

and

$$\begin{aligned} W_{i,\mathbf{t}} &= \bigvee_{j=1}^{d} G_{\eta}^{1/\eta t_j}(Z_{i,j}) + \sum_{j=1}^{d} \int_0^1 \dot{C}_{\eta,j}(v^{t_1\eta}, \ldots, v^{t_d\eta}) \mathbb{1}_{\{v > G_{\eta,j}^{1/\eta t_j}(Z_{i,j})\}} \mathrm{d}v, \\ \widetilde{W}_{i,\mathbf{t}} &= \frac{A_\eta(\mathbf{t})}{\{1+\eta A_\eta(\mathbf{t})\}^2} \rho(\mathbf{Z}_i). \end{aligned}$$

Note that the new expression for $T_{n,1}$ is obtained by applying Fubini's theorem. The pair $(T_{n,1}, \widetilde{T}_{n,2})$ is asymptotically tight and so, to show that its weak limit exists, it remains to establish the convergence of all its finite-dimensional distributions. This can be done by applying the central limit theorem since, for all $k \in \{1, 2, \ldots\}$, the iid random vectors

$$\left(W_{i,\mathbf{t}_1}, \ldots, W_{i,\mathbf{t}_k}, \widetilde{W}_{i,\mathbf{t}_1}, \ldots, \widetilde{W}_{i,\mathbf{t}_k}\right)$$

with $i \in \{1, \ldots, n\}$ have finite second order moments under the assumptions of Theorem 4.3.5; see also [50, Theorem 2.2.7]. This completes the proof of Lemma 4.5.3. ∎

We come back now to the proof of Theorem 4.3.5. Combining the three previous lemmas with the definition of $M(\eta, \mathbf{t})$, we have

$$\begin{aligned} L_n + R_n &= \frac{\{1+\eta A_\eta(\mathbf{t})\}^2}{\widehat{\eta}_n} \sqrt{n} \{\widehat{M}_n(\eta, \mathbf{t}) - M(\eta, \mathbf{t})\} \\ &\quad + \frac{A_\eta(\mathbf{t})}{\widehat{\eta}_n} \sqrt{n} (\widehat{\eta}_n - \eta) + \eta A_\eta(\mathbf{t}) \sqrt{n} \left(\frac{1}{\widehat{\eta}_n} - \frac{1}{\eta}\right) + o_p(1) \\ &= \frac{\{1+\eta A_\eta(\mathbf{t})\}^2}{\widehat{\eta}_n} \sqrt{n} \{\widehat{M}_n(\eta, \mathbf{t}) - M(\eta, \mathbf{t})\} + o_p(1) \\ &= -\frac{\{1+\eta A_\eta(\mathbf{t})\}^2}{\eta} \int_0^1 \widehat{\mathbb{C}}_n(v^{\eta t_1}, \ldots, v^{\eta t_d}) \mathrm{d}v + o_p(1). \end{aligned}$$

As in the proof of Lemma 4.5.1, using again the convergence $\widehat{\mathbb{C}}_n \rightsquigarrow \mathbb{A}_\eta$ in $\ell^\infty([0,1]^d)$ as $n \to \infty$, (4.15) follows from the continuous mapping theorem and Slutsky's lemma.

It now remains to prove (4.14). Note that

$$
\begin{aligned}
&\|\widehat{A}_{\widehat{\eta}_n,n} - A_\eta\|_\infty \\
&= \sup_{\mathbf{t}\in\mathcal{S}_d}\left|\frac{1}{\widehat{\eta}_n}\frac{\widehat{M}_n(\widehat{\eta}_n,\mathbf{t})}{1-\widehat{M}_n(\widehat{\eta}_n,\mathbf{t})} - \frac{1}{\eta}\frac{M(\eta,\mathbf{t})}{1-M(\eta,\mathbf{t})}\right| \\
&= \sup_{\mathbf{t}\in\mathcal{S}_d}\left|\frac{1}{\widehat{\eta}_n\eta\{1-\widehat{M}_n(\widehat{\eta}_n,\mathbf{t})\}\{1-M(\eta,\mathbf{t})\}}\right| \\
&\quad\times\sup_{\mathbf{t}\in\mathcal{S}_d}\left|\eta\{1-M(\eta,\mathbf{t})\}\widehat{M}_n(\widehat{\eta}_n,\mathbf{t}) - \widehat{\eta}_n\{1-\widehat{M}_n(\widehat{\eta}_n,\mathbf{t})\}M(\eta,\mathbf{t})\right| \\
&\equiv T_{n,1}\times T_{n,2}.
\end{aligned}
$$

Since $\widehat{\eta}_n \to \eta$ a.s., for a small $\varepsilon > 0$ and large n, we have almost-surely that

$$
T_{n,1} \le \frac{1+1/\eta}{\widehat{\eta}_n\int_0^1 \widehat{C}_n(v^{1+\varepsilon},\dots,v^{1+\varepsilon})\mathrm{d}v} \longrightarrow \frac{1+1/\eta}{\eta\int_0^1 C_\eta(v^{1+\varepsilon},\dots,v^{1+\varepsilon})\mathrm{d}v} < \infty.
$$

Now, using the Lipschitz continuity of order $k > 0$ of C_η, we have

$$
\begin{aligned}
T_{n,2} &\le \|\eta\{1-M(\eta,\mathbf{t})\} - \widehat{\eta}_n\{1-\widehat{M}(\widehat{\eta}_n,\mathbf{t})\}\|_\infty \\
&\quad + \|\{1-\widehat{M}_n(\widehat{\eta}_n,\mathbf{t})\}\{1-M(\eta,\mathbf{t})\}\|_\infty|\widehat{\eta}_n-\eta| \\
&\le |\widehat{\eta}_n-\eta|\|1-M(\eta,\mathbf{t})\|_\infty + \widehat{\eta}_n\|M(\eta,\mathbf{t})-\widehat{M}_n(\widehat{\eta}_n,\mathbf{t})\|_\infty + |\widehat{\eta}_n-\eta| \\
&\le 2|\widehat{\eta}_n-\eta| + \widehat{\eta}_n\|M(\widehat{\eta}_n,\mathbf{t})-\widehat{M}_n(\widehat{\eta}_n,\mathbf{t})\|_\infty + \widehat{\eta}_n\|M(\eta,\mathbf{t})-M(\widehat{\eta}_n,\mathbf{t})\|_\infty \\
&\le 2|\widehat{\eta}_n-\eta| + \widehat{\eta}_n\|\widehat{C}_n - C_\eta\|_\infty \\
&\quad + \widehat{\eta}_n k\int_0^1 \|v^{\widehat{\eta}_n t_1} - v^{\eta t_1},\dots,v^{\widehat{\eta}_n t_d} - v^{\eta t_d}\|_\infty \mathrm{d}v.
\end{aligned}
$$

Under our assumptions, each term on the right-hand side of this inequality converges to 0 a.s. – for the last term, similar arguments to those developed for Lemma 4.5.1 can be exploited. Thus (4.14) is established and the proof of Theorem 4.3.5 is now complete. ∎

4.5.5 Proof of Theorem 4.3.6

According to [37], η can be rewritten as $\eta = 2(1-\mu_{1,2}/\mu_{1,1})$. A natural estimator can thus be obtained by replacing $Q_\eta(u)$ by the empirical version $G_n^{\leftarrow}(u)$ where $G_n(u) = G_n(u,\dots,u)$. This entails

$$
\widehat{\eta}_n = 2(1-\widehat{\mu}_{1,2}/\widehat{\mu}_{1,1}),
$$

where

$$\widehat{\mu}_{a,b} = \int_0^1 Q_n(u)u^a(-\ln u)^b \mathrm{d}u.$$

Consequently, we can decompose the left-hand side of (4.16) as

$$\begin{aligned}\sqrt{n}\,(\widehat{\eta}_n - \eta) &= 2\sqrt{n}\left(\frac{\mu_{1,2}}{\mu_{1,1}} - \frac{\widehat{\mu}_{1,2}}{\widehat{\mu}_{1,1}}\right)\\ &= 2\frac{\int_0^1 \mathbb{Q}_n(u)\gamma(u)\mathrm{d}u}{n^{-1/2}\mu_{1,1}\int_0^1 \mathbb{Q}_n(u)u(-\ln u)\mathrm{d}u + \mu_{1,1}^2} \equiv 2\frac{N_n}{D_n}\end{aligned}$$

with $\mathbb{Q}_n(u) = \sqrt{n}\,\{Q_n(u) - Q_\eta(u)\}$. We start to study the numerator N_n. To this aim, we define the empirical and quantile processes, for all $u \in (0,1)$, by

$$\widetilde{\mathbb{H}}_n(u) = \sqrt{n}\,\{\widetilde{G}_n(u) - u\}, \quad \widetilde{\mathbb{Q}}_n(u) = \sqrt{n}\,\{\widetilde{Q}_n(u) - u\},$$

where for iid copies $U_1, \ldots, U_n$ of $U = G_\eta\{\max(Z_1, \ldots, Z_d)\}$, we denote, for all $u \in (0,1)$,

$$\widetilde{G}_n(u) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(U_i \le u)},$$

and as before $\widetilde{Q}_n = \widetilde{G}_n^{\leftarrow}$. Let $\dot{G}_\eta(y)$ and $\ddot{G}_\eta(y)$ be the first and second derivatives of $G_\eta(y)$ with respect to $y > 0$. The function defined in Theorem 4.3.6 is then equal, for all $u \in (0,1)$, to $\varphi(u) = \dot{G}_\eta\{Q_\eta(u)\}$. We can easily check that G_η satisfies the conditions of Theorem 3 in [15], whence

$$\sup_{u\in(0,1)} |\varphi(u)\,\mathbb{Q}_n(u) - \widetilde{\mathbb{Q}}_n(u)| = \mathrm{o}(1) \text{ a.s.} \tag{4.27}$$

and by Bahadur–Kiefer theorem (see, e.g., [21]) we have

$$\sup_{u\in(0,1)} |\widetilde{\mathbb{Q}}_n(u) + \widetilde{\mathbb{H}}_n(u)| = \mathrm{o}(1) \text{ a.s.}. \tag{4.28}$$

As by direct computations $\int_0^1 |\gamma(u)/\varphi(u)|\,\mathrm{d}u < \infty$, (4.27) and (4.28) entail

$$N_n = -\int_0^1 \widetilde{\mathbb{H}}_n(u)\frac{\gamma(u)}{\varphi(u)}\mathrm{d}u + \mathrm{o}(1) \text{ a.s.}.$$

A similar reasoning implies that almost surely

$$D_n = -n^{-1/2}\mu_{1,1}\int_0^1 \widetilde{\mathbb{H}}_n(u)\frac{u(-\ln u)}{\varphi(u)}\mathrm{d}u + \mu_{1,1}^2 + o(1) = \mu_{1,1}^2 + o(1).$$

We can now conclude that

$$\sqrt{n}\,(\widehat{\eta}_n - \eta) = -\frac{2}{\mu_{1,1}^2}\int_0^1 \mathbb{H}_n\{Q_\eta(u),\ldots,Q_\eta(u)\}\frac{\gamma(u)}{\varphi(u)}\mathrm{d}u + \mathrm{o}(1) \text{ a.s.},$$

where we use the fact that $\widetilde{\mathbb{H}}_n(u) = \mathbb{H}_n\{Q_\eta(u),\ldots,Q_\eta(u)\}$. Thus (4.16) is established. The other statements of the theorem are direct consequences. ∎

Bibliography

[1] A. F. Barrientos, A. Jara, and F. A. Quintana. Fully nonparametric regression for bounded data using dependent bernstein polynomials. *Journal of the American Statistical Association*, 112(518):806–825, 2017.

[2] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons Ltd.,Chichester, 2004.

[3] B. Beranger and S. A. Padoan. Extreme dependence models. In D. Dey and J. Yan, editors, *Extreme Value Modeling and Risk Analysis: Methods and Applications*. Chapman and Hall/CRC, 2015.

[4] B. Berghaus, A. Bücher, and H. Dette. Minimum distance estimators of the Pickands dependence function and related tests of multivariate extreme-value dependence. *Journal de la Société Française de Statistique*, 154(1):116–137, 2013.

[5] N. Bingham, C. Goldie, and J. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1989.

[6] A. Bücher, H. Dette, and S. Volgushev. New estimators of the Pickands dependence function and a test for extreme-value dependence. *The Annals of Statistics*, 39(4):1963–2006, 2011.

[7] A. Bücher and J. Segers. On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes*, 20:839–872, 2017.

[8] A. Bücher and J. Segers. Maximum likelihood estimation for the Fréchet

distribution based on block maxima extracted from a time series. *Bernoulli*, 24(2):1427–1462, 2018.

[9] P. Capéraà, A.-L. Fougères, and C. Genest. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84:567–577, 1997.

[10] N. Choudhuri, S. Ghosal, and A. Roy. Bayesian estimation of the spectral density of a time series. *JASA*, 99(468):1050–1059, 2004.

[11] N. Choudhuri, S. Ghosal, and A. Roy. Contiguity of the whittle measure for a gaussian time series. *Biometrika*, 91(1):211–218, 2004.

[12] S. G. Coles. *An Introduction to Statistical Modelling of Extreme Values*. Springer, London, 2001.

[13] E. Cormier, C. Genest, and J. G. Něsleová. Using b-splines for nonparametric inference on bivariate extreme-value copulas. *Extremes*, 17:633–659, 2014.

[14] M. Csörgő and P. Révész. Strong approximations of the quantile process. *The Annals of Statistics*, 6(4):882–894, 1978.

[15] M. Csörgő and P. Révész. Strong approximations of the quantile process. *Ann. Statist.*, 6:882–894, 1978.

[16] C. de Boor. *A practical guide to splines*, volume 27 of *Applied Mathematical Sciences*. Springer-Verlag, 2001.

[17] C. de Boor and G. J. Fix. Spline approximation by quasi-interpolants. *Journal of Approximation Theory*, 8:19–45.

[18] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer, 2006.

[19] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer Science & Business Media, 2006.

[20] G. Draisma, H. Drees, A. Ferreira, and L. de Haan. Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli*, 10:251–280, 2004.

[21] J. Einmahl. A short and elementary proof of the main bahadur–kiefer theorem. *Ann. Probab.*, (24):526–531, 1996.

[22] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. For insurance and finance.

[23] M. Falk, J. Hüsler, and R.-D. Reiß. *Laws of Small Numbers: Extremes and Rare Events (third ed.)*. Birkhäuser, Boston, 2010.

[24] J.-D. Fermanian, D. Radulović, and M. Wegkamp. Weak convergence of empirical copula processes. *Bernoulli*, 10:847–860, 2004.

[25] A. Freitas, J. Hüsler, and M. G. Temido. Limit laws for maxima of a stationary random sequence with random sample size. *Test*, 21(1):116–131, 2012.

[26] C. Genest, J. Nešlehová, and B. Rémillard. Asymptotic behavior of the empirical multilinear copula process under broad conditions. *J. Multivariate Anal*, 159:82–110, 2017.

[27] C. Genest and B. Rémillard. Test of independence and randomness based on the empirical copula process. *TEST*, 13:335–369, 2004.

[28] C. Genest and J. Segers. Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, 37(5B):2990–3022, 2009.

[29] J. E. Gentle. *Computational statistics*. Springer Science & Business Media, 2009.

[30] S. Ghosal and A. van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, 2017.

[31] S. Ghoshal. Convergence rates for density estimation with bernstein polynomials. *The Annals of Statistics*, 29(5):1264–280, 2001.

[32] G. Gudendorf and J. Segers. Nonparametric estimation of an extreme-value copula in arbitrary dimensions. *Journal of Multivariate Analysis*, 102(1):37–47, 2011.

[33] G. Gudendorf and J. Segers. Nonparametric estimation of multivariate extreme-value copulas. *Journal of Statistical Planning and Inference*, 142(12):3073–3085, 2012.

[34] S. Guillotte and F. Perron. A bayesian estimator for the dependence function of a bivariate extreme-value distribution. *Canadian Journal of Statistics*, 36(3):383–396, 2008.

[35] S. Guillotte and F. Perron. Polynomial pickands functions. *Bernoulli*, 22(1):213–241, 2016.

[36] S. Guillotte, F. Perron, and J. Segers. Non-parametric bayesian inference on bivariate extremes. *J. R. Statist. Soc. B*, 73(3):377–406, 2011.

[37] A. Guillou, P. Naveau, and A. Schorgen. Madogram and asymptotic independence among maxima. *REVSTAT–Statistical Journal*, 12(2):119–134, 2014.

[38] A. Guillou, S. A. Padoan, and S. Rizzelli. Inference for asymptotically independent sample of extremes. *Journal of Multivariate Analysis*, 167(114–135), 2018.

[39] T. E. Hanson, M. de Carvalho, and Y. Chen. Bernstein polynomial angular densities of multivariate extreme value distributions. *Statistics and Probability Letters*, 128:60–66, 2017.

[40] E. Hashorva, G. Ratomovirija, M. Tamraz, and Y. Bai. Some mathematical aspects of price optimisation. *Scandinavian Actuarial Journal*, (5), 2018.

[41] J. Hüsler and D. Li. Testing asymptotic independence in bivariate extremes. *J. Statist. Plann. Inference*, 139:990–998, 2009.

[42] A. Kiriliouk, J. Segers, and M. Warchoł. Nonparametric estimation of extremal dependence. *In: Extreme Value Modeling and Risk Analysis: Methods and Applications*, pages 353–376, 2016.

[43] I. Kojadinovic and J. Yan. Modeling multivariate distributions with continuous margins using the copula R package. *J. Stat. Softw.*, 34:1–20, 2010.

[44] A. Ledford and J. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83:169–187, 1996.

[45] A. Ledford and J. Tawn. Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B*, 59:475–499, 1997.

[46] G. Marcon, S. Padoan, P. Naveau, P. Muliere, and J. Segers. Multivariate nonparametric estimation of the pickands dependence function using bernstein polynomials. *Journal of Statistical Planning and Inference*, 183:1–17, 2017.

[47] G. Marcon, S. A. Padoan, and I. Antoniano-Villalobos. Bayesian inference for the extremal dependence. *Electronic Journal of Statistics*, 10(2):3310–3337, 2016.

[48] K. Maulik and S. Resnick. Characterizations and examples of hidden regular variation. *Extremes*, 7:31–67, 2002.

[49] L. Mhalla, V. Chavez-Demoulin, and P. Naveau. Non-linear models for extremal dependence. *Journal of Multivariate Analysis*, 159:49–66, 2017.

[50] R. Nelsen. An introduction to copulas, 2nd ed. 2006.

[51] A. K. Nikoloulopoulos, H. Joe, and H. Li. Extreme value properties of multivariate t copulas. *Extremes*, 12(2):129–148, 2009.

[52] S. Petrone. Bayesian density estimation using bernstein polynomials. *Canadian Journal of Statistics*, 27(1):105–126, 1999.

[53] S. Petrone. Random bernstein polynomials. *Scandinavian Journal of Statistics*, 26:373–393, 1999.

[54] S. Petrone and L. Wasserman. Consistency of bernstein polynomial posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(1):79–100, 2002.

[55] G. M. Phillips. *Interpolation and approximation by polynomials.* CMS Books in Mathematics. Springer, 2003.

[56] J. Pickands, III. Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, volume 49, pages 859–878, 894–902, 1981. With a discussion.

[57] A. Ramos and A. Ledford. A new class of models for bivariate joint tails. *J. R. Stat. Soc. Ser. B*, 71:219–241, 2009.

[58] A. Ramos and A. Ledford. An alternative point process framework for modeling multivariate extreme values. *Comm. Statist. Theory Methods*, 40:2205–2224, 2011.

[59] S. Resnick. Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5:303–336, 2002.

[60] S. I. Resnick. *Extreme Values, Regular Variation, and Point Processes*, volume 4. Springer Science & Business Media, 2007.

[61] C. Y. Robert and J. Segers. Tails of random sums of a heavy-tailed number of light-tailed terms. *Insurance: Mathematics and Economics*, 43(1):85–92, 2008.

[62] R. T. Rockafellar. *Convex analysis*. Princeton university press, 2015.

[63] H. Rootzén, J. Segers, and J. L. Wadsworth. Multivariate peaks over thresholds models. *Extremes*, 21(1):115–145, 2018.

[64] H. Rootzén and N. Tajvidi. Multivariate generalized pareto distributions. *Bernoulli*, 12(5):917–930, 2006.

[65] J. Segers. Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli*, 18(3):764–782, 2012.

[66] A. Stephenson. Bayesian inference for extreme value modelling. In D. Dey and J. Yan, editors, *Extreme Value Modeling and Risk Analysis: Methods and Applications*. Chapman and Hall/CRC, 2015.

[67] W. Stute. The oscillation behaviour of empirical processes: the multivariate case. *The Annals of Probability*, 12(2), 1984.

[68] J. A. Tawn. Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3):397–415, 1988.

[69] J. A. Tawn. Modelling multivariate extreme value distributions. *Biometrika*, 77(2):245–253, 1990.

[70] R Core Team. R: A language and environment for statistical computing. 2014.

[71] H. Tsukahara. Semiparametric estimation in copula models. *Canad. J. Statist.*, 33(3):357–375, 2005.

[72] V. V. Uchaikin and V. M. Zolotarev. *Chance and Stability Stable Distributions and Their Applications.* VSP International Science Publishers, Utrecht, 1999.

[73] A. van der Vaart and J. Wellner. Empirical processes indexed by estimated functions. *In: IMS Lecture Notes Monogr. Ser. 55*, pages 234–252, 2007.

[74] A. W. Van der Vaart. Asymptotic statistics (cambridge series in statistical and probabilistic mathematics). 2000.

[75] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer Verlag, New York, 1996.

[76] S. Vettori, R. Huser, and M. G. Genton. A comparison of dependence function estimators in multivariate extremes. *Statistics and Computing.*, DOI:10.1007/s11222-017-9745-7, 2018.

[77] J. Wadsworth and J. Tawn. A new representation for multivariate tail probabilities. *Bernoulli*, 19:2689–2714, 2013.

[78] J. Wadsworth, J. Tawn, A. Davison, and D. Elton. Modelling across extremal dependence classes. *J. R. Stat. Soc. Ser. B*, 79:149–175, 2017.

[79] D. Zhang, M. Wells, and L. Peng. Nonparametric estimation of the dependence function for a multivariate extreme value distribution. *J. Multivariate Anal.*, 99:577–588, 2008.