INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION
PAPER

1045

NONPARAMETRIC BAYESIAN INFERENCE ON BIVARIATE EXTREMES

GUILLOTTE, S., PERRON, F. and J. SEGERS

Nonparametric Bayesian Inference on Bivariate Extremes

Simon Guillotte[*], François Perron[†], and Johan Segers[‡]

October 26, 2010

Abstract

The tail of a bivariate distribution function in the domain of attraction of a bivariate extreme-value distribution may be approximated by the one of its extreme-value attractor. The extreme-value attractor has margins that belong to a three-parameter family and a dependence structure which is characterised by a probability measure on the unit interval with mean equal to one half, called spectral measure. Inference is done in a Bayesian framework using a censored-likelihood approach. A prior distribution is constructed on an infinite-dimensional model for this measure, the model being at the same time dense and computationally manageable. A trans-dimensional Markov chain Monte Carlo algorithm is developed and convergence to the posterior distribution is established. In simulations, the Bayes estimator for the spectral measure is shown to compare favorably with frequentist nonparametric estimators. An application to a data-set of Danish fire insurance claims is provided.

Keywords. Bayes, Bivariate Extreme-Value Distribution, Extreme Conditional Quantiles, MCMC, Metropolis-within-Gibbs, ϕ-irreducibility, Prediction, Rare Event Probabilities, Reversible Jumps, Spectral Measure.

1 Introduction

In areas such as engineering or financial risk management, decisions have to be made which depend on the extreme outcomes of two or more variables. Particular examples of interesting questions may be: how high should a dike be in order to withstand exceptionally high levels of a river at several sites simultaneously? How much capital to set aside in order to have sufficient reserve in times of financial crises, affecting the values of multiple financial securities at once?

[*]University of Prince Edward Island

[†]Université de Montréal

[‡]Corresponding author. Institut de statistique, biostatistique et sciences actuarielles (ISBA), Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. E-mail: Johan.Segers@uclouvain.be.

These kind of problems require inference on the distribution of bivariate (or more generally multivariate) extremes. In particular, we are interested in the bivariate density of a random pair (X_1, X_2) on a quadrant $[u_1, \infty) \times [u_2, \infty)$, where u_1 and u_2 are large thresholds, that is $\mathrm{P}(X_1 > u_1)$ and $\mathrm{P}(X_2 > u_2)$ are positive but small. This density can be used not only for the computation of rare-event probabilities over joint tail regions, but also of extreme conditional quantiles of one variable given an extreme outcome in the other variable. For instance, it may be of interest to find a level x_2 such that for a given probability $p > 0$ (small) and a given value of x_1 (large) we have $\mathrm{P}(X_2 > x_2 \mid X_1 = x_1) = p$. Up to the best of our knowledge, the latter type of problem has not yet been properly addressed in the extreme-value literature.

More specifically, let (x_{i1}, x_{i2}), $i = 1, \ldots, n$, be an observed sample from an unknown bivariate distribution F in the max-domain of attraction of a bivariate extreme-value distribution G. We want to infer on the (bivariate) tail of F using the sample. Except for the case of asymptotic independence, we know that the tail of F is well approximated by the one of G. The margins of the latter distribution are characterised by three parameters each, for a total of six parameters. We shall refer to these as the tail parameters. In addition, the dependence structure of G is characterised by a spectral measure, which can be any probability measure on the unit interval $[0, 1]$ with mean $1/2$. Thus, an approximation formula for F in a bivariate tail region is obtained by the specification of six marginal tail parameters and a spectral measure. It follows that with the above sample at hand, approximate inference on the tail of F may be done via inference on these parameters. Inference on the spectral measure may be done within parametric families, see for instance Boldi and Davison (2007); Coles and Tawn (1991, 1994); de Haan et al. (2008); Einmahl et al. (2008); Joe et al. (1992); Ledford and Tawn (1996); Smith (1994). Alternatively, one may prefer to proceed nonparametrically, see for instance de Haan and de Ronde (1998); de Haan and Sinha (1999); Einmahl et al. (2001, 2006); Einmahl and Segers (2009); Schmidt and Stadtmüller (2006). Surveys of these methods can be found for instance in the monographs Coles (2001); Beirlant et al. (2004); de Haan and Ferreira (2006); Kotz and Nadarajah (2000).

Data about extreme events being scarce by nature, the statistical uncertainty in extreme-value analysis is quite substantial. The question is how to deal with this uncertainty for practical purposes. A typical such purpose is prediction, the task being to compute a high return level, that is a level which is exceeded once, on average, during a long future time interval. The uncertainty due to the random nature of future outcomes then has to be combined with the statistical uncertainty on the parameter estimates. In a frequentist setting, it may be unclear how to do so: should one use a high quantile's point estimate or rather the upper bound of a certain confidence interval? As argued for instance in Coles and Tawn (1996) and Coles and Tawn (2005), the Bayesian approach via the predictive density (Aitchison and Dunsmore, 1975) seems more coherent. However, extreme-value dependence structures are essentially infinite-dimensional. Therefore, nonparametric Bayesian methodology for multivariate extremes should be further developed. Our paper aims to take a step in that direction.

Essentially, we extend the censored-likelihood method developed in Ledford and Tawn (1996) to the case of arbitrary (i.e. infinite-dimensional) spectral measures in a Bayesian setup. To this end, we select a prior on the six marginal tail parameters and on the set of spectral measures. Via the censored likelihood, the joint posterior of the tail parameters and the spectral measure is computed and converted into a posterior distribution for the tail quantities of interest. The actual inference, based on the posterior, is performed by means of a trans-dimensional Markov chain Monte Carlo algorithm; see for instance Guillotte and Perron (2008) for similar work in the context of annual maxima, that is when data can be modelled directly by a bivariate extreme-value distribution. Our methodology enables the evaluation of the predictive density in a bivariate tail region, which can be used, for instance, for prediction of high future levels of one variable given such outcomes of the other one.

The prior selection for the spectral measure is the more delicate part, and this is the main contribution of the paper. We need to put a prior on the set of all cumulative distribution functions on $[0, 1]$ with mean $1/2$. The actual prior will be concentrated on a countable union of finite-dimensional families of smooth spectral measures which is dense in the set of all spectral measures. For maximal flexibility, it is important that the spectral measures in our model allow for atoms at 0 and 1 and at the same time are absolutely continuous on $(0, 1)$.

The outline of the paper is as follows. Since the extreme-value model provides an approximation of the bivariate tail of F only, special care has to be taken on how to define the likelihood of the parameters given the data (Section 2). The construction of the subspace of spectral measures is done in Section 3. The selection of a prior for the spectral measures is explained in Section 4, which then is used as a starting point for Bayesian inference in Section 5. The MCMC algorithm employed for numerical computations along with a proof of its convergence are given in Section 6. In Section 7, the Bayes estimator for the spectral measure is compared via simulations with two nonparametric frequentist estimators. Furthermore, the methodology is applied to a data-set of Danish fire insurance claims McNeil (1997). Finally, a discussion in Section 8 concludes the paper.

2 Modelling bivariate tails

The domain-of-attraction condition on a bivariate cumulative distribution function F (Subsection 2.1) yields good approximations for F on quadrants of the form $[u_1, \infty) \times [u_2, \infty)$, where u_1 and u_2 are high thresholds (Subsection 2.2). Here, a high threshold means that $F_j(u_j)$ is less than but close to 1, where F_j is the marginal distribution, $j \in \{1, 2\}$. From this approximation for F, approximations for interesting tail quantities can be derived. Since F is only well approximated on a subset of its support, care has to be taken when writing down the likelihood (Subsection 2.3).

2.1 The domain-of-attraction condition

We assume that there exist sequences of constants $a_{nj} > 0$ and b_{nj}, for $j \in \{1, 2\}$, and a bivariate cumulative distribution function G with non-degenerate margins such that

$$\lim_{n\to\infty} F^n(a_{n1}x_1 + b_{n1}, a_{n2}x_2 + b_{n2}) = G(x_1, x_2) \tag{2.1}$$

for all continuity points (x_1, x_2) of G. Here, G is called an *extreme-value* cumulative distribution function and is necessarily of the following form:

- Its marginal cumulative distribution functions G_1 and G_2 are those of a univariate extreme-value distribution:

$$-\log G_j(x_j) = \left(1 + \xi_j \frac{x_j - \mu_j}{\sigma_j}\right)^{-1/\xi_j}, \qquad j \in \{1, 2\},$$

  for x_j such that $\sigma_j + \xi_j(x_j - \mu_j) > 0$, with shape parameter ξ_j (the extreme-value index), location parameter μ_j, and positive scale parameter σ_j;

- Its dependence structure is given by

$$-\log G(x, y) = \ell\big(-\log G_1(x), -\log G_2(y)\big), \tag{2.2}$$

  for all $(x, y) \in \mathbb{R}^2$ such that $G_1(x) > 0$ and $G_2(y) > 0$. The *stable tail dependence function* ℓ (sometimes called *tail copula*) admits the representation

$$\ell(s, t) = 2\int_{[0,1]} \max\big(ws, (1 - w)t\big)\, dH(w), \qquad (s, t) \in [0, \infty)^2,$$

  the *spectral measure* H being a probability measure on $[0, 1]$ with mean equal to $1/2$. Note that the name "spectral measure" is quite often reserved for the measure $2H$.

See for instance the monograph of Coles (2001) for an elementary introduction to extreme-value theory, in particular Section 3.1 for the univariate theory and Section 8.2 for the bivariate theory. The function ℓ is associated to the function V of Theorem 8.1 in Coles (2001) via $\ell(s, t) = V(1/s, 1/t)$.

Given a pair of large thresholds, u_1 and u_2, it will be convenient to rewrite the marginal distributions using a different parametrisation which incorporates the thresholds in their expressions. More precisely, writing $\eta_j = (\xi_j, \zeta_j, \sigma_j)$, $j \in \{1, 2\}$, the marginal cumulative distribution functions can be written as

$$-\log G_j(x_j \mid \eta_j) = \zeta_j \left(1 + \xi_j \frac{x_j - u_j}{\sigma_j}\right)^{-1/\xi_j}, \qquad j \in \{1, 2\}, \tag{2.3}$$

for x_j such that $\sigma_j + \xi_j(x_j - u_j) > 0$, where ξ_j is again the extreme-value index, $\sigma_j > 0$ is a scale parameter, and $0 < \zeta_j = -\log G_j(u_j \mid \eta_j)$. As u_j is large, we have $\zeta_j \approx 1 - G_j(u_j \mid \eta_j)$, the marginal probability of exceeding the threshold u_j.

Therefore, a bivariate extreme-value cumulative distribution function G is parameterised by its marginal parameter vectors η_1 and η_2 and its spectral measure H. From now on, we shall make this explicit. Such a cumulative distribution function $(x_1, x_2) \mapsto G(x_1, x_2 \mid H, \eta_1, \eta_2)$ is absolutely continuous provided that the restriction of the spectral measure H to the interior $(0, 1)$ of the unit interval is absolutely continuous. Still, the spectral measure is allowed to have atoms at 0 and 1, so as to include, for instance, the case of independence where H is equal to the Bernoulli$(1/2)$ measure.

In statistical practice, the spectral measure is often modelled parametrically. In this article, however, we model H nonparametrically for maximum flexibility.

2.2 The tail approximation

In Ledford and Tawn (1996) and Nadarajah et al. (1998), the domain-of-attraction condition (2.1) is exploited to construct a good approximation for the tail of F. This construction enables us to stipulate that for large thresholds u_1 and u_2, and for $(x_1, x_2) \in [u_1, \infty) \times [u_2, \infty)$, the form of $F(x_1, x_2)$ is that of a bivariate extreme-value cumulative distribution function. The justification comes from equation (2.1) and the fact that extreme-value distributions are max-stable: for $t > 0$, the function G^t is also a bivariate extreme-value cumulative distribution function and it differs from G by location and scale only.

Thus, as in Ledford and Tawn (1996), we postulate that $F(x_1, x_2) = F(x_1, x_2 \mid H, \eta_1, \eta_2)$ for $(x_1, x_2) \in [u_1, \infty) \times [u_2, \infty)$, where $F(x_1, x_2 \mid H, \eta_1, \eta_2)$ has dependence structure given by (2.2), for some spectral measure H, and marginal distributions given by (2.3), for some parameter vectors η_1 and η_2. Here, $F(x_1, x_2 \mid H, \eta_1, \eta_2)$ for $(x_1, x_2) \in [u_1, \infty) \times [u_2, \infty)$ is called the *tail approximation* to F.

Furthermore, we assume that H is absolutely continuous on $(0, 1)$, so that $F(x_1, x_2 \mid H, \eta_1, \eta_2)$ is absolutely continuous on the region $(x_1, x_2) \in [u_1, \infty) \times [u_2, \infty)$. Its density

$$f(x_1, x_2 \mid H, \eta_1, \eta_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F(x_1, x_2 \mid H, \eta_1, \eta_2), \qquad x_1 \geqslant u_1,\ x_2 \geqslant u_2,$$

is a (rather complicated) function of the six marginal tail parameters given by η_1 and η_2 together with the spectral measure H. It is derived in Appendix A.1.

2.3 The censored likelihood

Let (X_1, X_2) be a random vector with (unknown) cumulutative distribution function F, and let (x_1, x_2) be a realisation of this vector. Recall that the tail approximation to F is defined only on the tail region $[u_1, \infty) \times [u_2, \infty)$. Therefore, by using the tail approximation alone, it is not necessarily possible to specify the likelihood contribution of the datum (x_1, x_2), since it may or may not belong to the tail region. Let us introduce a pseudo-likelihood instead. We follow Ledford and Tawn (1996) by adopting a censoring approach: if $x_j \leqslant u_j$, then we pretend that x_j is censored by u_j.

Let $(x_1^*, x_2^*) = (x_1 \vee u_1, x_2 \vee u_2)$, let $d = (\mathbb{1}_{[u_1,\infty)}(x_1), \mathbb{1}_{[u_2,\infty)}(x_2))$, and define

$$f^*(x_1^*, x_2^* \mid H, \eta_1, \eta_2) = \begin{cases} F(u_1, u_2 \mid H, \eta_1, \eta_2) & \text{if } d = (0,0), \\ \dfrac{\partial}{\partial x_1} F(x_1, u_2 \mid H, \eta_1, \eta_2) & \text{if } d = (1,0), \\ \dfrac{\partial}{\partial x_2} F(u_1, x_2 \mid H, \eta_1, \eta_2) & \text{if } d = (0,1), \\ \dfrac{\partial^2}{\partial x_1 \partial x_2} F(x_1, x_2 \mid H, \eta_1, \eta_2) & \text{if } d = (1,1). \end{cases}$$

The exact expression for f^* is given in Appendix A.1. Note also that if $H(\{0\}) > 0$ and $H(\{1\}) > 0$, that is if the spectral measure has atoms at 0 and 1, then f^* is positive on the set $\{(x_1^*, x_2^*) : \sigma_j + \xi_j(x_j^* - u_j) > 0, j = 1, 2\}$.

Finally, let $x = \{(x_{i1}, x_{i2}) : i = 1, \ldots, n\}$ be an observed sample from F, with corresponding censored sample $x^* = \{(x_{i1}^*, x_{i2}^*) : i = 1, \ldots, n\}$. The likelihood is defined as

$$L(H, \eta_1, \eta_2 \mid x^*) = \prod_{i=1}^{n} f^*(x_{i1}^*, x_{i2}^* \mid H, \eta_1, \eta_2), \tag{2.4}$$

which depends implicitly on the thresholds u_1 and u_2.

3 Modelling the spectral measure

Let $\mathscr{H}^{\mathrm{c}}$ be the set of spectral measures whose only atoms, if any, are at 0 and at 1. For a spectral measure H, the same symbol will denote its cumulative distribution function, that is, $H(w) = H([0, w])$ for $w \in [0, 1]$. In this section we will first construct a class $\mathscr{H}^{\mathrm{d}}$ of discrete spectral measures (Subsection 3.1), and, by smoothing these, a class $\mathscr{H}^{\mathrm{s}}$ of smooth spectral measures (Subsection 3.2). Both classes are countable unions of finite-dimensional models and they are dense in $\mathscr{H}^{\mathrm{c}}$ with respect to the topology of uniform convergence of cumulative distribution functions.

3.1 The approximation step

For integer $m \geqslant 1$, let $\mathscr{H}_m^{\mathrm{d}}$ be the set of discrete spectral measures whose restriction to $(0, 1)$ is uniformly distributed over m distinct points. According to the following proposition, the cumulative distribution function of a spectral measure in $\mathscr{H}^{\mathrm{c}}$ can be approximated arbitrarily closely, and in a uniform way, by the one of a spectral measure in $\mathscr{H}^{\mathrm{d}} = \bigcup_{m \geqslant 1} \mathscr{H}_m^{\mathrm{d}}$. For convenience, we exclude the Bernoulli(1/2) spectral measure, which is in $\mathscr{H}^{\mathrm{c}}$ and can be thought of as the only member of $\mathscr{H}_0^{\mathrm{d}}$.

Proposition 3.1. *For every $H \in \mathscr{H}^{\mathrm{c}}$ with $H((0,1)) > 0$ and every integer $m \geqslant 1$ there exists $H^* \in \mathscr{H}_m^{\mathrm{d}}$ such that $H(\{a\}) = H^*(\{a\}) =: h_a$ for $a \in \{0, 1\}$ and such that*

$$\sup_{w \in [0,1]} |H(w) - H^*(w)| \leqslant (1 - h_0 - h_1)/m. \tag{3.1}$$

Proof. Let G be the cumulative distribution function of the probability measure supported on $(0,1)$ defined by $G(w) = H((0,w])/H((0,1))$ for $0 < w < 1$. By assumption, G is continuous, and

$$H(w) = h_0 + (1 - h_0 - h_1)\, G(w) + h_1\, \delta_1(w), \qquad w \in [0,1], \tag{3.2}$$

with δ_1 the cumulative distribution function of the Dirac measure at 1. Let $G^{-1}(u) = \inf\{w \in [0,1] : G(w) \geqslant u\}$, for $u \in [0,1]$, be the generalised inverse of G. Since G is continuous, it follows that G^{-1} is strictly increasing. Therefore, by setting $q_0 = 0$ and

$$q_i = G^{-1}(i/m), \qquad\qquad y_i = m \int_{(i-1)/m}^{i/m} G^{-1}(u)\, du,$$

for $i \in \{1, \ldots, m\}$, we get that $q_{i-1} < y_i < q_i$ for all $i \in \{1, \ldots, m\}$.

Let G^* be the cumulative distribution function of the discrete uniform distribution on the set $\{y_1, \ldots, y_m\}$. By construction, $G(q_i) = G^*(q_i) = i/m$ for all $i \in \{0, \ldots, m\}$ (recall that G is continuous) and

$$\int_0^1 w\, dG^*(w) = \frac{1}{m} \sum_{i=1}^{m} y_i = \int_0^1 G^{-1}(u)\, du = \int_0^1 w\, dG(w).$$

In addition, for $w \in (0, q_m]$ we can find $i \in \{1, \ldots, m\}$ such that $q_{i-1} < w \leqslant q_i$, whence

$$\sup_{w \in [0,1]} |G(w) - G^*(w)| \leqslant 1/m. \tag{3.3}$$

Let H^* be the discrete measure with masses h_0 and h_1 at the points 0 and 1 respectively and with masses $(1-h_0-h_1)/m$ on each of the points $y_1, \ldots, y_m$. The cumulative distribution function of H^* is given by

$$H^*(w) = h_0 + (1 - h_0 - h_1)\, G^*(w) + h_1\, \delta_1(w), \qquad w \in [0,1]. \tag{3.4}$$

Clearly, $H^*(1) = 1$ and

$$\begin{aligned} \int_0^1 w\, dH^*(w) &= (1 - h_0 - h_1) \int_0^1 w\, dG^*(w) + h_1 \\ &= (1 - h_0 - h_1) \int_0^1 w\, dG(w) + h_1 = \int_0^1 w\, dH(w) = \frac{1}{2}, \end{aligned}$$

that is, H^* is a spectral measure; more precisely, $H^* \in \mathscr{H}^{\mathrm{d}}$. Finally, by (3.2), (3.3) and (3.4), we arrive at (3.1). □

3.2 The smoothing step

Let $H^* \in \mathscr{H}^{\mathrm{d}}_m$ for some $m \geqslant 1$ with masses $0 \leqslant h_0 < 1/2$ and $0 \leqslant h_1 < 1/2$ at 0 and 1 respectively, and with mass $(1 - h_0 - h_1)/m$ on each of the points $y_1 < \cdots < y_m$ in $(0,1)$.

Write $\vartheta = \vartheta(H^*) = (h_0, y_1, \ldots, y_m, h_1)$, and note that the relation between H^* and ϑ is one-to-one. We will associate to ϑ a spectral measure H_ϑ in $\mathscr{H}^{\mathrm{c}}$ whose cumulative distribution function is close to the one of H^* itself.

Let S be a map that associates to a vector a of abscissas $a_1 < \cdots < a_k$ and a vector b of ordinates $b_1 \leqslant \cdots \leqslant b_k$ a piecewise cubic polynomial function $\varphi = S(a, b)$ such that $\varphi(a_i) = b_i$ for all $i \in \{1, \ldots, k\}$ and such that φ is nondecreasing and continuously differentiable on $[a_1, a_k]$. Such a construction is proposed in Fritsch and Butland (1984), see also Appendix A.2.

Given $H^* \in \mathscr{H}_m^{\mathrm{d}}$ as above, define functions φ_- and φ_+ on $[0, 1]$ by $\varphi_\pm = S(a; b_\pm)$ where

$$\begin{aligned} a &= (\quad 0, \quad y_1, \quad y_2, \quad \ldots, \quad y_m, \quad 1), \\ b_- &= \big(H^*(0), \quad H^*(0), \quad H^*(y_1), \quad \ldots, \quad H^*(y_{m-1}), \quad H^*(y_m)\big), \\ b_+ &= \big(H^*(0), \quad H^*(y_1), \quad H^*(y_2), \quad \ldots, \quad H^*(y_m), \quad H^*(y_m)\big). \end{aligned}$$

Define cumulative distribution functions H_- and H_+ of probability measures on $[0, 1]$ by $H_\pm(w) = \varphi_\pm(w)$ if $0 \leqslant w < 1$ and $H_\pm(1) = 1$. Clearly, $H_-(w) \leqslant H^*(w) \leqslant H_+(w)$ for all $w \in [0, 1]$. Put

$$H_\vartheta = \alpha\, H_- + (1 - \alpha)\, H_+, \tag{3.5}$$

where $\alpha = \alpha(\vartheta) \in [0, 1]$ is determined by the mean constraint

$$1/2 = \int_0^1 w\, dH_\vartheta(w) = \alpha \int_0^1 w\, dH_-(w) + (1 - \alpha) \int_0^1 w\, dH_+(w).$$

By construction, the masses of H_-, H_+ and H_ϑ at 0 and 1 are again h_0 and h_1 respectively, and

$$\sup_{w \in [0,1]} |H_\vartheta(w) - H^*(w)| \leqslant (1 - h_0 - h_1)/m, \tag{3.6}$$

the same bound holding true with H_ϑ replaced by either H_- or H_+.

Let $\mathscr{H}^{\mathrm{s}} = \bigcup_{m \geqslant 1} \mathscr{H}_m^{\mathrm{s}}$, where $\mathscr{H}_m^{\mathrm{s}} = \{H_\vartheta : \vartheta = \vartheta(H^*),\, H^* \in \mathscr{H}_m^{\mathrm{d}}\}$. As a corollary to Proposition 3.1, we obtain the following result, illustrated in Figure 1.

Proposition 3.2. *For every $H \in \mathscr{H}^{\mathrm{c}}$ such that $H((0, 1)) > 0$ and every integer $m \geqslant 1$ there exists $H_\vartheta \in \mathscr{H}_m^{\mathrm{s}}$ such that $H(\{a\}) = H_\vartheta(\{a\}) =: h_a$ for $a \in \{0, 1\}$ and such that*

$$\sup_{w \in [0,1]} |H_\vartheta(w) - H(w)| \leqslant 2\,(1 - h_0 - h_1)/m.$$

Proof. Let $H^* \in \mathscr{H}_m^{\mathrm{d}}$ be the discrete approximant from Proposition 3.1. Let $\vartheta = \vartheta(H^*)$. The measure H_ϑ in (3.5) enjoys all the required properties; in particular, see (3.6). □

Note that the Bernoulli(1/2) spectral measure can still be approximated arbitrarily closely by members in $\mathscr{H}^{\mathrm{d}}$ or $\mathscr{H}^{\mathrm{s}}$: take h_0 and h_1 close to 1/2.

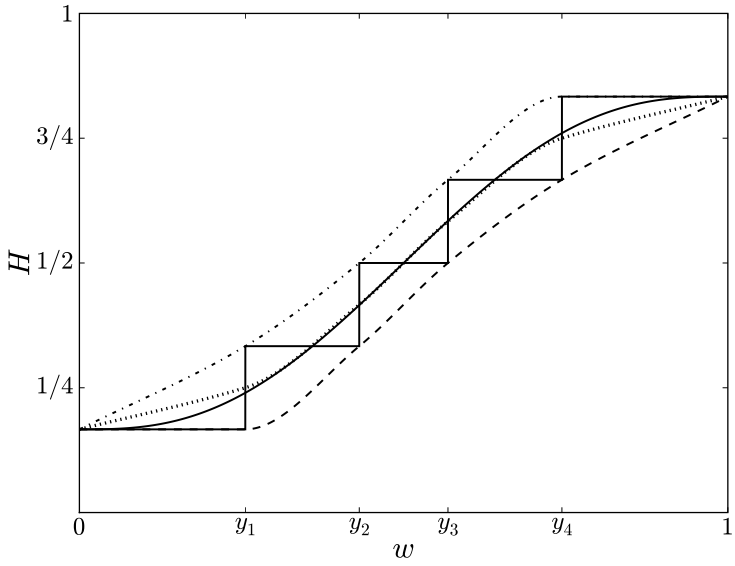


Figure 1: The illustration shows the distribution functions from Proposition 3.2. The step function is H^*, the dashed line is H_- and the dash-dotted line is H_+. The thick line is the true spectral measure H and the dotted line is the smooth approximation H_ϑ.

4 Prior distribution for the spectral measure

We will now construct a prior on the approximating set of spectral measures $\mathscr{H}^{\mathrm{s}} = \bigcup_{m\geqslant 1} \mathscr{H}^{\mathrm{s}}_m$. The spectral measures in $\mathscr{H}^{\mathrm{s}}_m$ are represented by vectors $\vartheta = (h_0, y_1, \ldots, y_m, h_1) \in \mathbb{R}^{m+2}$, with $0 \leqslant h_0 < 1/2$ and $0 \leqslant h_1 < 1/2$ and with $0 < y_1 < \cdots < y_m < 1$ subject to the constraint

$$1/2 = (1 - h_0 - h_1)\bar{y} + h_1. \tag{4.1}$$

The sigma-field on $\mathscr{H}^{\mathrm{s}}$ is the smallest one such such that the map $\vartheta \mapsto H_\vartheta$ measurable, that is, a set $\mathscr{G} \subset \mathscr{H}^{\mathrm{s}}$ is measurable if the set of $\vartheta \in \bigcup_{m\geqslant 1} \mathbb{R}^{m+2}$ verifying (4.1) and such that $H_\vartheta \in \mathscr{G}$ is a Borel set.

It will be convenient to work with a different parametrisation. Let

$$\Theta = \bigcup_{m\geqslant 1} \big(\{m\} \times \Theta_m\big), \tag{4.2}$$

with Θ_m defined as follows:

$$\begin{aligned}
\Theta_1 &= \big\{(h_0, h_1) : 0 < h_0 < 1/2 \text{ and } 0 < h_1 < 1/2\big\}, \\
\Theta_m &= \big\{(h_0, y_1, \ldots, y_{m-1}, h_1) : 0 < h_0 < 1/2,\ 0 < h_1 < 1/2, \\
&\qquad\qquad \text{and } 0 < y_i < 1 \text{ for } i \in \{1, \ldots, m\}\big\}, \quad m \geqslant 2,
\end{aligned}$$

the point y_m being a function of h_0, h_1 and $y_1, \ldots, y_{m-1}$ via the mean constraint (4.1):

$$y_m = m\,\frac{1/2 - h_1}{1 - h_0 - h_1} - \sum_{i=1}^{m-1} y_i. \tag{4.3}$$

Similarly, if $m = 1$, the constraint (4.1) implies $y_1 = (1/2 - h_1)/(1 - h_0 - h_1)$.

Elements of Θ_m with $m \geqslant 2$ for which there are ties among the points $y_1, \ldots, y_m$ do not correspond to spectral measures H_ϑ in $\mathscr{H}^{\mathrm{s}}$. Let $N \subset \Theta$ be the set of such parameter vectors. For $(m, \theta) \in \Theta \setminus N$, put $H_{m,\theta} = H_\vartheta$ where $\vartheta = \vartheta(m, \theta) = (h_0, y_{(1)}, \ldots, y_{(m)}, h_1)$ with $y_{(1)} < \cdots < y_{(m)}$, the ordered values of $y_1, \ldots, y_m$, and with y_m determined by (4.3). For definiteness, put $H_{m,\theta} = \text{Bernoulli}(1/2)$ for $(m, \theta) \in N$. Since the map $(m, \theta) \mapsto \vartheta$ is measurable and since N is measurable, the map

$$\Theta \to \mathscr{H}^{\mathrm{s}} : (m, \theta) \mapsto H_{m,\theta} \tag{4.4}$$

is also measurable.

Each spectral measure H_ϑ in $\mathscr{H}^{\mathrm{s}}$ for which $h_0 > 0$ and $h_1 > 0$ can be represented by a vector $(m, \theta) \in \Theta$ in this way. Because we do not insist on the points y_i of the vector θ to be ordered, there are actually $m!$ different parameter vectors (m, θ) that yield the same spectral measure H_ϑ. The situation is similar to the case of mixture distributions, where the order of the mixture components can be left non-identifiable, see Marin et al. (2005). The exclusion of zero masses $h_0 = 0$ or $h_1 = 0$ at 0 and 1 has the technical advantage that the parameter space is open and does not affect the flexibility of the model, since it is always possible to take h_0 and h_1 arbitrarily small.

We now construct a probability measure on the parameter space Θ in (4.2). With respect to this probability measure, the set N is a null set. Moreover, the map $(m, \theta) \mapsto H_{m,\theta}$ induces a probability measure on $\mathscr{H}^{\mathrm{s}}$ which we select as the prior for H.

First, we draw an integer $m \geqslant 1$ from a 0–truncated Poisson(λ) distribution. Next, inside each model m, the prior for $\theta \in \Theta_m$ is defined with a uniform prior on the atoms h_0 and h_1 in $(0, 1/2)^2$, and, if $m \geqslant 2$, conditionally on $(h_0, h_1) \in (0, 1/2)^2$, a uniform prior for $(y_1, \ldots, y_{m-1})$ on the set

$$\Theta_{m,h_0,h_1} = \big\{(y_1, \ldots, y_{m-1}) : (h_0, y_1, \ldots, y_{m-1}, h_1) \in \Theta_m\big\} \subset \mathbb{R}^{m-1}. \tag{4.5}$$

Specifically, the prior on $(m, \theta) \in \Theta$ is specified by $\pi(m, \theta) = \pi(\theta \mid m)\,\pi(m)$ where $\theta \in \Theta_m$ and

$$\pi(m) \propto \lambda^m/m!, \qquad \text{for } m \geqslant 1,$$

$$\pi(\theta \mid m) = \begin{cases} \pi(h_0, h_1) & \text{if } m = 1, \\ \pi(y_1, \ldots, y_{m-1} \mid m, h_0, h_1)\,\pi(h_0, h_1) & \text{if } m \geqslant 2, \end{cases}$$

where

$$\pi(h_0, h_1) \propto \mathbb{1}_{(0,1/2)^2}(h_0, h_1),$$
$$\pi(y_1, \ldots, y_{m-1} \mid m, h_0, h_1) \propto \mathbb{1}_{\Theta_{m,h_0,h_1}}(y_1, \ldots, y_{m-1})$$

The set $N \subset \Theta$ is indeed a null set and the mapping (4.4) is well defined everywhere on Θ. The prior on (m, θ) induces a prior on the spectral measure H via

$$\int_{\mathscr{H}^{\mathrm{s}}} g(H)\, d\pi(H) = \int_{\Theta} g(H_{m,\theta})\, d\pi(m, \theta) \tag{4.6}$$

for nonnegative measurable functions g on $\mathscr{H}^{\mathrm{s}}$.

Because of its trans-dimensional nature, implementation of the MCMC algorithm in Section 6 requires the exact knowledge of $\pi(y_1, \ldots, y_{m-1} \mid m, h_0, h_1)$ and thus of the normalising constants $\lambda_{m-1}(\Theta_{m,h_0,h_1})$ for integer $m \geqslant 2$ and $(h_0, h_1) \in (0, 1/2)^2$; here λ_{m-1} denotes the $(m-1)$–dimensional Lebesgue measure. These normalising constants are given by the following result, illustrated in Figure 2, and proved in Appendix A.3.

Lemma 4.1. *For integer $m \geqslant 2$ and $(h_0, h_1) \in (0, 1/2)^2$, we have*

$$\lambda_{m-1}(\Theta_{m,h_0,h_1}) = \frac{1}{(m-1)!} \left(\frac{m(1/2 - h_1)}{1 - h_0 - h_1} \right)^{m-1} \mathrm{P}\left(Y_{(m)} < \frac{1 - h_0 - h_1}{m(1/2 - h_1)} \right), \tag{4.7}$$

where $Y_{(m)}$ is the maximum of $Y = (Y_1, \ldots, Y_m)$, and Y is distributed according to the uniform distribution on the $(m-1)$-dimensional unit simplex, that is, the Dirichlet$(1, \ldots, 1)$ distribution. Furthermore,

$$\mathrm{P}\left(Y_{(m)} < \frac{1 - h_0 - h_1}{m(1/2 - h_1)} \right) = 1 - \sum_{k=1}^{K} (-1)^{k-1} \binom{m}{k} \left(1 - k \frac{1 - h_0 - h_1}{m(1/2 - h_1)} \right)^{m-1}, \tag{4.8}$$

where K is the greatest integer less than or equal to $m(1/2 - h_1)/(1 - h_0 - h_1)$.

Formally, putting $\Theta_{1,h_0,h_1} = \{0 < y_1 < 1 : (1 - h_0 - h_1)y_1 + h_1 = 1/2\} = \{(1/2 - h_1)/(1 - h_0 - h_1)\}$, formula (4.7) is also true for $m = 1$, by defining λ_0 as the counting measure.

5 Bayesian inference

The model approximating the tail of the (unknown) bivariate cumulative distribution F is specified via a spectral measure H in $\mathscr{H}^{\mathrm{s}}$ and marginal parameters $(\eta_1, \eta_2) \in \Xi^2$, $\eta_j = (\xi_j, \zeta_j, \sigma_j)$, $j \in \{1, 2\}$, where

$$\Xi = (-\infty, \infty) \times (0, \infty) \times (0, \infty).$$

The parameter space for $\mathscr{H}^{\mathrm{s}}$ is given by $\Theta = \bigcup_{m \geqslant 1} (\{m\} \times \Theta_m)$ in (4.2). Thus, the parameters defining a spectral measure consist of a model index $m \geqslant 1$ and, given this model index, a

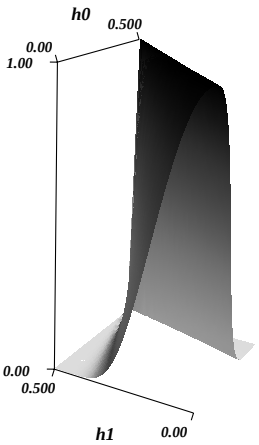


Figure 2: Illustration of $\lambda_{m-1}(\Theta_{m,h_0,h_1})$, for $m = 10$, $(h_0, h_1) \in [0, 1/2)^2$. The amplitude has been rescaled by a factor of 0.735.

parameter vector $\theta = (h_0, y_1, \ldots, y_{m-1}, h_1) \in \Theta_m$. Therefore, the complete parameter space is

$$\Omega = \Theta \times \Xi^2. \tag{5.1}$$

We assume a priori that (m, θ) and η_1, η_2 are independent. The prior for (m, θ), and hence for H, has been constructed in Section 4. Concerning the marginal parameters, we do not claim any originality but essentially follow the methodology proposed in Beirlant et al. (2004), in which a careful literature review has been made on the subject. We consider independent priors for both margins given by

$$\pi(\eta_j) \propto \exp(-\xi_j^2/2)\ \exp(-4\zeta_j)\,\sigma_j \exp(-\sigma_j/2)\,\mathbb{1}_{(0,\infty)}(\zeta_j)\,\mathbb{1}_{(0,\infty)}(\sigma_j), \tag{5.2}$$

that is, ξ_j, ζ_j and σ_j follow independent normal, exponential and gamma distributions respectively, for $j \in \{1, 2\}$. While arbitrary, such a choice guarantees that the prior for $(m, \theta, \eta_1, \eta_2)$ is proper, and so this is also true for the posterior. Alternatively, expert knowledge on certain marginal return levels (quantiles) may be incorporated in the prior as in Coles and Tawn (1996). The priors considered in the latter refererence are proper as well.

We develop Bayesian inference in the same spirit as that of model selection, see for instance Robert (2007). Again, let $x = \{(x_{11}, x_{12}), \ldots, (x_{n1}, x_{n2})\}$ be a sample from F. Let x^* be the corresponding censored sample. The joint posterior density for the parameter $(m, \theta, \eta_1, \eta_2)$ is given by

$$\pi(m, \theta, \eta_1, \eta_2 \mid x^*) \propto L(H_{m,\theta}, \eta_1, \eta_2 \mid x^*)\,\pi(\theta \mid m)\,\pi(m)\,\pi(\eta_1)\,\pi(\eta_2), \tag{5.3}$$

where L is the likelihood given by equation (2.4). In Section 6, we give an MCMC algorithm which is used for numerical computations in every inference procedure that we propose throughout the rest of the paper.

In particular, for integer $m \geqslant 1$, let

$$\pi(m \mid x^*) = \int_{\Theta_m \times \Xi^2} \pi(m, \theta, \eta_1, \eta_2 \mid x^*)\, d\theta\, d\eta_1\, d\eta_2$$

be the posterior probability of selecting model m. We define the Bayes estimator for the spectral measure H as the mixture

$$\hat{H}(w) = \sum_{m \geqslant 1} \pi(m \mid x^*)\, \hat{H}^{(m)}(w), \qquad w \in [0,1], \tag{5.4}$$

where

$$\hat{H}^{(m)}(w) = \frac{1}{\pi(m \mid x^*)} \int_{\Theta_m \times \Xi^2} H_{m,\theta}(w)\, \pi(m, \theta, \eta_1, \eta_2 \mid x^*)\, d\theta\, d\eta_1\, d\eta_2$$

is the $\mathscr{L}^2$-Bayes estimator of $H(w)$ inside model m. The estimator (5.4) is evaluated numerically via the sample mean of a trans-dimensional Markov chain constructed in the next section.

Further, tail-related quantities of F are derived from the joint predictive (censored) density

$$f^*(x_1^*, x_2^* \mid x^*) = \sum_{m \geqslant 1} \int_{\Theta_m \times \Xi^2} f^*(x_1^*, x_2^* \mid H_{m,\theta}, \eta_1, \eta_2) \pi(m, \theta, \eta_1, \eta_2 \mid x^*)\, d\theta\, d\eta_1\, d\eta_2, \tag{5.5}$$

where (x_1^*, x_2^*) can be thought of as a future (censored) observation, and from the conditional predictive density

$$f^*_{2|1}(x_2^* \mid x_1^*, x^*) = \frac{f^*(x_1^*, x_2^* \mid x^*)}{f_1^*(x_1^* \mid x^*)}. \tag{5.6}$$

See Section 7.2 for an illustration.

6 Markov chain Monte Carlo sampling scheme

We now construct an MCMC algorithm that generates an irreducible Markov chain having the posterior as its stationary distribution. The algorithm is described in Subsection 6.1. Irreducibility and convergence are treated in Subsection 6.3, based upon a general result on irreducibility provided in Subsection 6.2.

Recall that the complete parameter space for the vector $\omega = (m, \theta, \eta_1, \eta_2)$ is given by $\Omega = \Theta \times \Xi^2$ in (5.1). Here (η_1, η_2) determines both margins, and (m, θ) determines the spectral measure.

6.1 The algorithm

We consider a *block-at-a-time* algorithm, see Chib and Greenberg (1995) and Gamerman and Lopes (2006), constructed using two types of moves. At each iteration, a randomly chosen move is proposed. The first move (M_1) proposes a new candidate ω' for which only (η_1, η_2) is updated, and the second move (M_2) updates only (m, θ). The algorithm is thus a mixture of

two subalgorithms given by M_1 and M_2. While M_1 is a standard *full dimensional* algorithm on Ξ^2, M_2 is more involved since the proposed (m', θ') can either belong to the same space $\{m\} \times \Theta_m$ as the current (m, θ) ($M_{2.1}$), or to $\{m+1\} \times \Theta_{m+1}$ ($M_{2.2}$) or $\{m-1\} \times \Theta_{m-1}$ ($M_{2.3}$). In Roberts and Rosenthal (2006), M_2 is called a *trans-dimensional, Metropolis-within-Gibbs* algorithm.

The acceptance probabilities and target densities for both moves are given as follows:

1. The acceptance probability for M_1 is given by the usual Metropolis–Hastings acceptance probability, see for instance Tierney (1994), the target distribution here being the conditional posterior distribution of (η_1, η_2) given (m, θ).

2. The acceptance probability used in M_2 is the one proposed by Green (1995). The target distribution in M_2 is the conditional posterior distribution of (m, θ) given (η_1, η_2).

It then follows that the (joint) posterior distribution of $\omega = (m, \theta, \eta_1, \eta_2)$ is the stationary distribution of the generated Markov chain, see Chib and Greenberg (1995) and Gamerman and Lopes (2006).

The two moves are detailed below. Assume that the current state of the parameter is $\omega = (m, \theta, \eta_1, \eta_2)$ and that at this point, the posterior density $\pi(\omega \mid x^*)$ given in (5.3) is positive.

M_1 – **Propose new margins.**

1.1 Draw η_1' and η_2' independently and according to the prior in (5.2).

1.2 Let

$$\begin{aligned} A &= \frac{L(H_{m,\theta}, \eta_1', \eta_2' \mid x^*)\, \cancel{\pi_1(\eta_1')\, \pi_2(\eta_2')}\, \cancel{\pi_1(\eta_1)\, \pi_2(\eta_2)}}{L(H_{m,\theta}, \eta_1, \eta_2 \mid x^*)\, \cancel{\pi_1(\eta_1)\, \pi_2(\eta_2)}\, \cancel{\pi_1(\eta_1')\, \pi_2(\eta_2')}} \\ &= \frac{L(H_{m,\theta}, \eta_1', \eta_2' \mid x^*)}{L(H_{m,\theta}, \eta_1, \eta_2 \mid x^*)}, \end{aligned}$$

where L is the likelihood given by equation (2.4), and accept (η_1', η_2') with probability $A \wedge 1$.

M_2 – **Propose a new spectral measure.** We update (m, θ), where $\theta \in \Theta_m$ is given by

$$\theta = (\theta_0, \theta_1, \ldots, \theta_{m-1}, \theta_m) = \begin{cases} (h_0, h_1) & \text{if } m = 1, \\ (h_0, y_1, \ldots, y_{m-1}, h_1) & \text{if } m \geqslant 2. \end{cases}$$

If $m = 1$, select a submove among $M_{2.1}$ and $M_{2.2}$ with equal probability; if $m \geqslant 2$, select a submove among $M_{2.1}$, $M_{2.2}$ and $M_{2.3}$ with equal probability. Let $\mathscr{I}_m = \{0, 1, \ldots, m-1, m\}$ be the set of indices corresponding to the components of θ.

$M_{2.1}$ – **Propose a candidate in Θ_m.** This is a *random-scan Metropolis-within-Gibbs move*. Select at random a component $i \in \mathscr{I}_m$ of θ to be updated, the other

components remaining fixed. Draw θ_i' uniformly on the interval $I_{\theta_{-i}}$, associated to the selected component, with endpoints on the boundary of Θ_m, and let

$$q_i(\theta_i' \mid \theta_{-i}) = \frac{1}{\lambda_1\left(I_{\theta_{-i}}\right)} \mathbb{1}_{I_{\theta_{-i}}}(\theta_i'),$$

be the resulting (instrumental) density; λ_1 denotes Lebesgue measure. Let $\theta' \in \Theta_m$ be the proposed candidate, and let

$$\begin{aligned} A &= \frac{L(H_{m,\theta'}, \eta_1, \eta_2 \mid x^*)\, \pi(m, \theta')\, \cancel{q_i(\theta_i \mid \theta'_{-i})}}{L(H_{m,\theta}, \eta_1, \eta_2 \mid x^*)\, \pi(m, \theta)\, \cancel{q_i(\theta_i' \mid \theta_{-i})}} \\ &= \frac{L(H_{m,\theta'}, \eta_1, \eta_2 \mid x^*)\, \pi(m, \theta')}{L(H_{m,\theta}, \eta_1, \eta_2 \mid x^*)\, \pi(m, \theta)}. \end{aligned}$$

Accept θ' to be the new candidate with probability $A \wedge 1$.

$\mathrm{M}_{2.2}$ – **Propose a candidate in Θ_{m+1}.** Insert $\bar{y} = (1/2 - h_1)/(1 - h_0 - h_1)$ in θ as the component with index m, and relabel $\theta = (\theta_0, \theta_1, \ldots, \theta_m, \theta_{m+1})$, leaving θ_i fixed for $i < m$, and with $\theta_m = \bar{y}$ and $\theta_{m+1} = h_1$. Note that $\theta \in \Theta_{m+1}$. Draw θ_m' uniformly on the interval $I_{\theta_{-m}}$, associated to the component θ_m, with endpoints on the boundary of Θ_{m+1}, and let

$$q_m(\theta_m' \mid \theta_{-m}) = \frac{1}{\lambda_1\left(I_{\theta_{-m}}\right)} \mathbb{1}_{I_{\theta_{-m}}}(\theta_m'),$$

be the resulting (instrumental) density. Let $\theta' \in \Theta_{m+1}$ be the proposed candidate, and let

$$A = \frac{L(H_{m+1,\theta'}, \eta_1, \eta_2 \mid x^*)\, \pi(m+1, \theta')}{L(H_{m,\theta}, \eta_1, \eta_2 \mid x^*)\, \pi(m, \theta)\, q_m(\theta_m' \mid \theta_{-m})}.$$

Accept the proposed candidate with probability $A \wedge 1$.

$\mathrm{M}_{2.3}$ – **Propose a candidate in Θ_{m-1}.** Remove the component with index $m-1$ from θ, relabel $\theta' = (\theta_0, \theta_1, \ldots, \theta_{m-2}, \theta_{m-1})$, leaving θ_i fixed for $i < m-1$, and with $\theta_{m-1} = h_1$. Accept θ' with probability $(1/A) \wedge 1$, where A is as in $\mathrm{M}_{2.2}$ above with m replaced by $m-1$.

By way of example, we use move M_2 in the above algorithm to draw random spectral measures $H = H_{m,\theta}$ from the prior (substitute L by 1 in the two definitions of A above). Figure 3 below shows a shaded region $\mathscr{S} \subset [0,1) \times [0,1]$ such that for every vertical section $\mathscr{S}(w)$, $0 \leqslant w < 1$, we have $\mathrm{P}[H(w) \in \mathscr{S}(w)] = 0.95$ a priori, together with the prior pointwise mean $\mathrm{E}[H(w)]$. Note that the priors on $H(0) = h_0$ and $\lim_{w \uparrow 1} H(w) = 1 - h_1$ are uniform on $(0, 1/2)$ and $(1/2, 1)$, respectively.

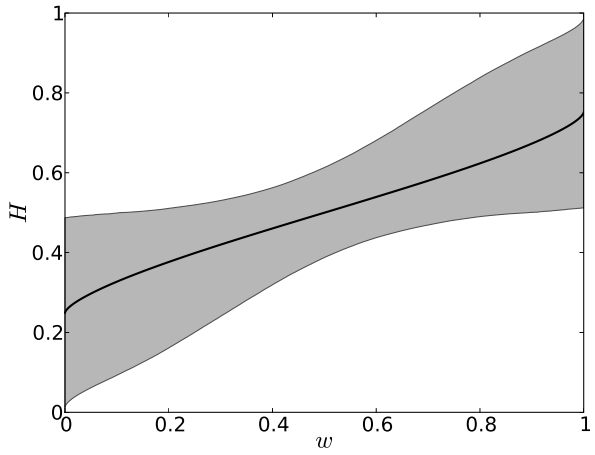


Figure 3: A shaded region $\mathscr{S} \subset [0,1) \times [0,1]$ such that $\mathrm{P}[H(w) \in \mathscr{S}(w)] = 0.95$ a priori for every vertical section $\mathscr{S}(w) = \{v \in [0,1] : (w,v) \in \mathscr{S}\}$, $0 \leqslant w < 1$, together with the prior pointwise mean $\mathrm{E}[H(w)]$ (full line). The hyperparameter for the model index m is set to $\lambda = 10$, and we use 500 000 iterations of the algorithm for generating the Markov chain.

6.2 Locally uniform irreducibility

In order to show irreducibility of the Markov chain generated by the algorithm in Subsection 6.1, we make use of a general result, Proposition 6.1 below, providing sufficient conditions for the irreducibility of a Markov chain which makes transitions according to a mechanism as in move $\mathrm{M}_{2.1}$. A crucial feature is that the number $k \geqslant 1$ of moves needed for the chain to be able to reach the destination set is the same for every starting point in an open cube. We do not claim originality of this result, and, in fact, similar conditions have been used by Roberts and Smith (1994) and Hobert et al. (1997) in the context of the standard *Gibbs* sampler. Since we have not found this particular result anywhere in the literature, we provide a detailed proof.

First we need to introduce some notation and terminology. For $x \in \mathbb{R}^d$, put $\|x\|_\infty = \max(|x_1|, \ldots, |x_d|)$. A set $C \subset \mathbb{R}^d$ is called an open *cube* if and only if there exist $x \in \mathbb{R}^d$ and $r > 0$ such that $C = \{y \in \mathbb{R}^d : \|x - y\|_\infty < r\} =: C(x,r)$. We call x the *center* and r the *radius* of the cube. A non-empty set $E \subset \mathbb{R}^d$ is said to be *path-connected* if for every $x, y \in E$ there exists a continuous map $\varphi : [0,1] \to E$, called a *path*, such that $\varphi(0) = x$ and $\varphi(1) = y$. For $E \subset \mathbb{R}^d$ and $x \in E$, write $E_i(x) = \{y_i \in \mathbb{R} \colon (x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_d) \in E\}$, that is, $E_i(x)$ is the set of coordinates y_i such that replacing the ith coordinate x_i of the vector x by y_i produces a new vector that still belongs to E. For a Borel set $E \subset \mathbb{R}^d$, let λ_E denote d-dimensional Lebesgue measure restricted to E.

Proposition 6.1 (Locally uniform irreducibility)**.** *Let $\varnothing \neq E \subset \mathbb{R}^d$ be open and let $(Y_t)_{t\in 0}^{\infty}$ be a homogeneous Markov chain on E such that the transition between $Y_0 = x$ and Y_1 is*

determined by the following mechanism:

1. *an index $i \in \{1, \ldots, d\}$ is drawn with probability $p_i(x) > 0$;*
2. *a new coordinate $y_i \in E_i(x)$ is drawn from a positive density $q_i(\cdot \mid x)$ on $E_i(x)$;*
3. *the move from x to $x' = (x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_d)$ is accepted with probability $0 < \alpha_i(y_i \mid x) < 1$, in which case $Y_1 = x'$, or else $Y_1 = x$.*

If E is path-connected, then for every cube $C_0 \subset E$ and every Borel set $B \subset E$ such that $\lambda_E(B) > 0$, there exists an integer $k \geqslant 1$ such that

$$\mathrm{P}[Y_k \in B \mid Y_0 = y] > 0, \qquad \textit{for all } y \in C_0.$$

In particular, the chain $(Y_t)_{t=0}^{\infty}$ is λ_E-irreducible.

The proof of the proposition, given in detail in Appendix A.4, rests upon two ideas: first, any two points in an open and path-connected set E can be connected by a finite number of overlapping cubes; second, within a d-dimensional cube, every destination set can be reached in at most d moves. These ideas are illustrated in Figure 4.

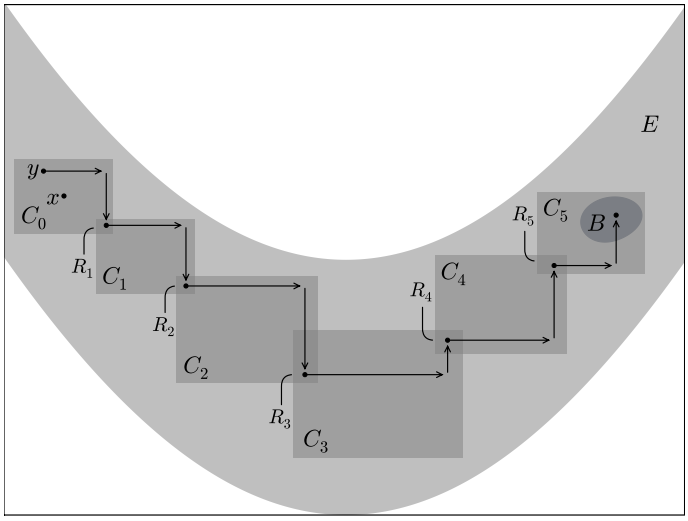


Figure 4: Proposition 6.1: moving from $y \in C(x, \varepsilon) = C_0$ via $R_i = C_{i-1} \cap C_i$, $i = 1, \ldots, 5$, to $B \subset C = C_5$, when $E \subset \mathbb{R}^2$.

6.3 Convergence

We show ϕ-irreducibility of the Markov chain generated by the algorithm in Subsection 6.1 (Theorem 6.2), where ϕ is to be specified. Let $\pi(\omega \mid x^*)$, for all $\omega \in \Omega = \Theta \times \Xi^2$, be the posterior density given the censored sample x^*, and denote

$$\Omega_{x^*} = \{\omega \in \Omega : \pi(\omega \mid x^*) > 0\},$$
$$(\Xi^2)_{x^*} = \{(\eta_1, \eta_2) \in \Xi^2 : \pi(\eta_1, \eta_2 \mid x^*) > 0\}.$$

Note that from the definition of the likelihood (2.4) and the prior (4.6) on (m, θ), we have $\pi(m, \theta \mid x^*) > 0$ for all $(m, \theta) \in \Theta$ (recall we excluded $h_0 = 0$ and $h_1 = 0$), and also $\Omega_{x^*} = \Theta \times (\Xi^2)_{x^*}$.

Here ϕ-irreducibility means that from any starting point $\omega \in \Omega_{x^*}$, and for every set $B \subseteq \Omega_{x^*}$ with $\phi(B) > 0$, the chain can reach B in a finite number of steps with positive probability, for some measure ϕ on Ω_{x^*}.

To select ϕ, let λ_d denote d-dimensional Lebesgue measure, and consider the product measure $\lambda_\oplus \times \lambda_6$, where $\lambda_\oplus$ is the direct sum measure, that is, for all $A = \bigcup_{d \geqslant 2} \{d\} \times A_d$, where $A_d \subseteq \mathbb{R}^d$, $d \geqslant 2$, $\lambda_\oplus(A) = \sum_{d \geqslant 2} \lambda_d(A_d)$, see Fremlin (2003). We take $\phi = \lambda^*$, the restriction of the above product measure to Ω_{x^*}. In fact λ^* is the reference measure for the posterior distribution on Ω_{x^*}. Since its density is positive on Ω_{x^*}, the posterior distribution is in fact equivalent to λ^*, but for notational convenience, we have decided to work with $\phi = \lambda^*$.

Our main theorem for assessing convergence is the following. Its proof is given below.

Theorem 6.2. *The MCMC algorithm in Subsection 6.1 generates a λ^*-irreducible, aperiodic Markov chain on Ω_{x^*}.*

In view of the classical result found for instance in Tierney (1994), Roberts and Rosenthal (2006) or Meyn and Tweedie (2009), Theorem 6.2 implies that there is a set of posterior probability one for which the chain converges to the posterior if the initial state belongs to this set. If $\pi_{m,\theta}$ denotes the marginal posterior of (m, θ), then Theorem 6.2 also implies ergodicity on a set of $\pi_{m,\theta}$-probability one, see Tierney (1996), that is convergence of the sample means of the generated spectral measures to the Bayes estimator (5.4) for $\pi_{m,\theta}$-almost every starting point $(m, \theta) \in \Theta$.

There may be a set of posterior probability zero inside Ω_{x^*}, and therefore of λ^*-measure zero, such that convergence to the posterior does not occur if the initial state is chosen from this set. If one wants to have convergence for every starting point, then a stronger property called Harris recurrence needs to be proved. Harris recurrence is equivalent to what is known as the Ergodic Theorem for irreducible Markov chains, see for example Robert and Casella (2004) or Meyn and Tweedie (2009). Although sufficient conditions for Harris recurrence are provided by Roberts and Rosenthal (2006), they seem to be difficult to verify in our setup. However, the set of measure zero on which convergence does not necessarily occur can be avoided by chosing the initial state using a continuous distribution on Ω_{x^*}.

Proof of Theorem 6.2. Suppose that the following claim has been established:

Claim 6.3. For $(\eta_1, \eta_2) \in (\Xi^2)_{x^*}$, the homogeneous Markov chain $(Z_t)_{t=0}^\infty$ on Θ generated by move M_2 of the MCMC algorithm generates a $\lambda_\oplus$-irreducible, aperiodic Markov chain on Θ.

Then the proof of the theorem can be concluded as follows. Let $\omega = (m, \theta; \eta) \in \Omega_{x^*} = \Theta \times (\Xi^2)_{x^*}$, where $\eta = (\eta_1, \eta_2)$, and let B be a λ^*-measurable set in Ω_{x^*} with $\lambda^*(B) > 0$. For every $(m', \theta') \in \Theta$, consider the section $B_2(m', \theta') = \{\eta' \in (\Xi^2)_{x^*} : (m', \theta'; \eta') \in B\}$, and

let $B_1 = \{(m', \theta') \in \Theta : \lambda_6(B_2(m', \theta')) > 0\}$. Here B_1 is a $\lambda_\oplus$-measurable set, and Fubini's theorem implies that $\lambda_\oplus(B_1) > 0$. We will move from ω into B in two stages:

$$\begin{array}{lll} \omega = (m, \theta; \eta) \xrightarrow{\text{(1) } k \text{ times move } \mathrm{M}_2} & (m', \theta'; \eta) & \in \quad B_1 \times \{\eta\} \\ & \downarrow \text{ (2) move } \mathrm{M}_1 & \\ & (m', \theta'; \eta') & \in \quad \{(m', \theta')\} \times B_2(m', \theta') \subset B \end{array}$$

(1) By Claim 6.3, there exists a positive integer k such that with positive probability, the chain generated by the MCMC algorithm can reach the set $B_1 \times \{\eta\}$ starting from the point $(m, \theta; \eta)$ after k times move M_2.

(2) From a point $(m', \theta'; \eta) \in B_1 \times \{\eta\}$, the chain can reach the set $\{(m', \theta')\} \times B_2(m', \theta') \subset B$ in a single application of move M_1.

By the Chapman-Kolmogorov equations, it then follows that starting from the point ω, the chain can reach the set B in $k + 1$ iterations. Hence, the Markov chain generated by the MCMC algorithm is λ^*-irreducible. In addition, from the proof of aperiodicity in Claim 6.3, it also follows that with positive probability, the chain may not move at all. Therefore, it is aperiodic.

It then remains to prove Claim 6.3. This is done below, first for irreducibility and then for aperiodicity.

Proof of Claim 6.3: irreducibility. Let $z = (m, \theta) \in \Theta = \bigcup_{m \geqslant 1} \big(\{m\} \times \Theta_m\big)$ and let $B \subset \Theta$ be a Borel set such that $\lambda_\oplus(B) > 0$. We have to show that there exists an integer $k = k(z, B)$ such that $\mathrm{P}[Z_k \in B \mid Z_0 = z] > 0$.

Without loss of generality, we may assume that B is of the form $\{d\} \times B_d$ where B_d is a Borel set in Θ_d with $\lambda_{d+1}(B_d) > 0$. We have to show that there exists a positive integer k such that $\mathrm{P}[Z_k \in \{d\} \times B_d \mid Z_0 = (m, \theta)] > 0$. There are three cases to consider: $d = m$, $d > m$ and $d < m$.

Case $d = m$: Suppose first that $d = m$, that is, no change of dimension is needed. By Lemma A.3 in Appendix A.5, the set Θ_m is path-connected. In view of Proposition 6.1, successive calls to move $\mathrm{M}_{2.1}$ will leave the model index m unchanged and will displace the initial vector θ to a point in the set B_m. Formally, let $(Y_t)_{t=0}^{\infty}$ be the homogeneous Markov chain on $E = \Theta_m$ with transition mechanism given by move $\mathrm{M}_{2.1}$. By Proposition 6.1, there exists an integer $k \geqslant 1$ such that $\mathrm{P}[Y_k \in B_m \mid Y_0 = \theta] > 0$. It follows that $\mathrm{P}[Z_k \in \{m\} \times B_m \mid Z_0 = (m, \theta)] \geqslant (1/3)^k \, \mathrm{P}[Y_k \in B_m \mid Y_0 = \theta] > 0$, since at each step, there is a chance of $1/3$ or $1/2$ that move $\mathrm{M}_{2.1}$ is selected.

Case $d > m$: For integer $j \geqslant 1$, write $c_j = (1/4; 1/2, \ldots, 1/2; 1/4) \in \Theta_j$, the center of Θ_j. Since Θ_j is open, there exists $r_j > 0$ such that $C(c_j, r_j) \subset \Theta_j$. Let $r = \min(r_m, \ldots, r_d)$. Recall that move $\mathrm{M}_{2.2}$ of the algorithm permits to increment the model index by one. We

will show that with positive probability, it is possible to move in the following way:

$$\begin{array}{ccc} & \{d\} \times C(c_d, r) & \xrightarrow{k_2 \text{ times move } \mathrm{M}_{2.1}} \{d\} \times B_d \\ & \uparrow \text{ move } \mathrm{M}_{2.2} & \\ & \cdots & \\ & \uparrow \text{ move } \mathrm{M}_{2.2} & \\ & \{m+1\} \times C(c_{m+1}, r) & \\ & \uparrow \text{ move } \mathrm{M}_{2.2} & \\ (m, \theta) & \xrightarrow{k_1 \text{ times move } \mathrm{M}_{2.1}} \{m\} \times C(c_m, r) & \end{array}$$

First, as in the case $d = m$, the existence of the required numbers k_1 and k_2 of moves is guaranteed by Proposition 6.1. A crucial fact here is that the number k_2 does not depend on the particular transit point in $\{d\} \times C(c_d, r)$.

Second, for each $j \in \{m, \ldots, d-1\}$ and each $\theta^{(j)} \in C(c_j, r)$, it is possible to move from $(j, \theta^{(j)})$ upwards to $\{j+1\} \times C(c_{j+1}, r)$ with a single call to move $\mathrm{M}_{2.2}$: an additional new coordinate in the interval $(1/2 - r, 1/2 + r)$ is proposed and the enlarged resulting vector is accepted.

Finally, by the Chapman–Kolmogorov equations, there is a positive probability that the chain will move from (m, θ) along the path above into $\{d\} \times B_d$ in $k_1 + (d - m) + k_2$ steps.

Case $d < m$: The proof is similar as in the case $d > m$, this time with $m - d$ downward moves ($\mathrm{M}_{2.3}$) from $\{m\} \times C(c_m, r)$ down to $\{d\} \times C(c_d, r)$, with $r = \min(r_d, \ldots, r_m)$. At each such step, the final coordinate of the vector $(y_1, \ldots, y_{j-1})$ is suppressed and the resulting vector is accepted.

This completes the proof of irreducibility in Claim 6.3. □

Proof of Claim 6.3: aperiodicity. It is sufficient to show that events of the form $\{Z_{t+1} = Z_t\}$ occur with positive probability. Consider the vector $\theta = (1/4; 1/4, 1/4, 1/4, 7/8; 1/4)$ in Θ_5. Applying move $\mathrm{M}_{2.3}$ to $(5, \theta)$, attempting to move into the set $\{4\} \times \Theta_4$, involves the suppression of the coordinate $7/8$. However, the resulting point $(4, \theta')$, where $\theta' = (1/4; 1/4, 1/4, 1/4; 1/4)$, does not belong to Θ_4, as $y_4 = 5/4 > 1$. As a consequence, the proposed vector must be rejected.

By continuity, there exists a small cube $C = C(\theta, \varepsilon) \subset \Theta_5$ so that for every point in $\{5\} \times C$, the move $\mathrm{M}_{2.3}$ yields a proposal that violates the constraint $0 < y_4 < 1$ and therefore must be rejected with probability one. Finally, by irreducibility, the chain has positive probability of reaching the set $\{5\} \times C$, and this completes the proof of aperiodicity. □

Claim 6.3 being proven, the proof of Theorem 6.2 is complete too. □

7 Examples

In order to provide some validation for our methodology, we first apply it to artificial data (Subsection 7.1) and compare our estimator with two frequentist estimators. Next, in Sub-

section 7.2, we analyse a dataset of Danish fire insurance claims, a univariate version of which has been previously studied in McNeil (1997) and Resnick (1997).

7.1 Simulation experiment

We compare the performance of our estimator with two frequentist estimators: the empirical spectral measure in Einmahl et al. (2001) and the maximum empirical likelihood estimator (MELE) in Einmahl and Segers (2009).

Let $x = \{(x_{i1}, x_{i2}), i = 1, \ldots, n\}$ be an observed sample from a distribution in the domain of attraction of a bivariate extreme-value distribution. While our estimator is based on the associated censored sample x^*, the two frequentist estimators are based on the following rank-transformed sample. Let $z_{ij} = n/(n + 1 - r_{ij})$, with r_{ij} the rank of x_{ij} among $x_{1j}, \ldots, x_{nj}$, for $i = 1, \ldots, n$ and $j = 1, 2$. For $i = 1, \ldots, n$, put $s_i = z_{i1} + z_{i2}$ and $w_i = z_{i1}/s_i$; think of (s_i, w_i) as the (pseudo-)polar coordinates of (z_{i1}, z_{i2}) with respect to the sum-norm. Now for a fixed $k > 0$, let $I_{n,k} \subset \{1, \ldots, n\}$ be the set of $i \in \{1, \ldots, n\}$ such that $s_i > n/k$. Here k plays a similar role to our threshold (u_1, u_2) and should be thought of as being large but of smaller order than n; in fact, asymptotically, $k = k_n \to \infty$ but $k_n/n \to 0$. Write $N = |I_{n,k}|$. The empirical spectral measure is simply

$$\hat{H}_{\mathrm{emp}}(w) = \frac{1}{N} \sum_{i \in I_{n,k}} \mathbb{1}_{[0,w_i]}(w), \qquad w \in [0, 1].$$

Note that its mean is not necessarily equal to $1/2$, so $\hat{H}_{\mathrm{emp}}$ is not a genuine spectral measure. The MELE is equal to

$$\hat{H}_{\mathrm{MELE}}(w) = \frac{1}{N} \sum_{i \in I_{n,k}} p_i \, \mathbb{1}_{[0,w_i]}(w), \qquad w \in [0, 1],$$

where the weights p_i are given by

$$p_i = \frac{1}{N} \frac{1}{1 + \mu(w_i - 1/2)}, \qquad i = 1, \ldots, N,$$

in terms of a Lagrange multiplier μ determined by the moment constraint

$$1/2 = \int_0^1 w \, d\hat{H}_{\mathrm{MELE}}(w) = \sum_{i \in I_{n,k}}^{N} p_i \, w_i.$$

We consider three models in the domain of attraction of some bivariate extreme-value distribution, two of them with atoms at 0 and 1, and one without such atoms; see Appendix A.6. To asses the performance of the estimators, we do the same type of simulation experiment as in Einmahl and Segers (2009). For each model, we draw 1 000 samples, each of size $n = 1\,000$. We consider ten different thresholds $(u_{1\alpha}, u_{2\alpha})$, with α ranging over an equidistant grid in $[0.72, 0.92]$. For each α and for $j \in \{1, 2\}$, the threshold $u_{j\alpha}$ is equal to the empirical quantile

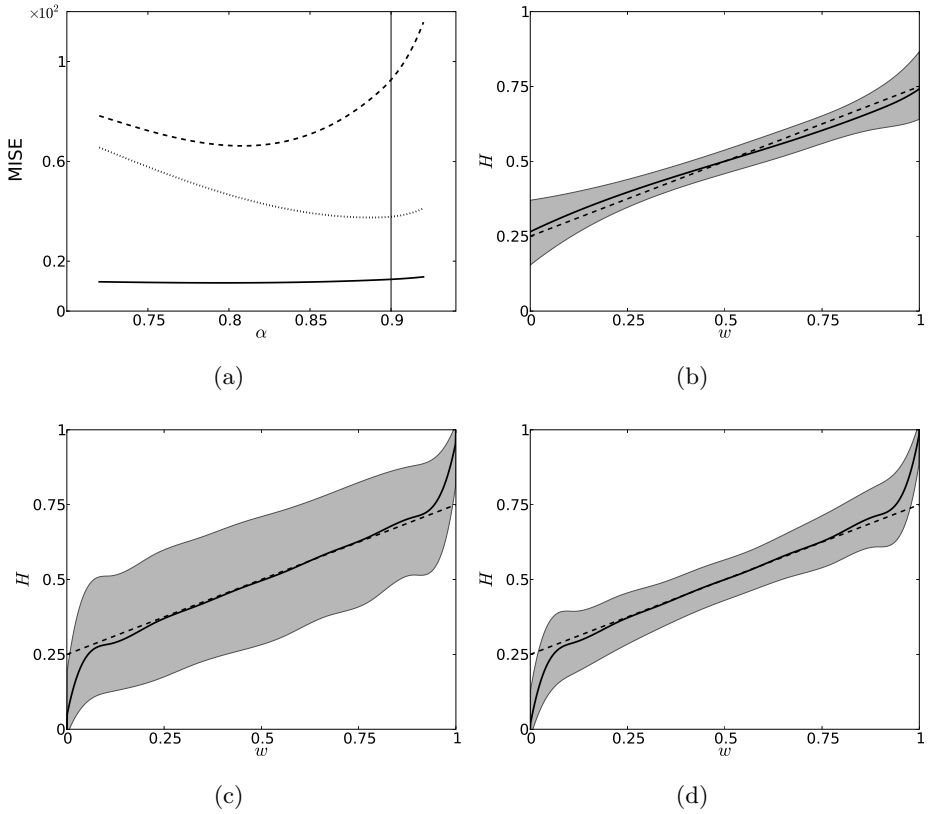


Figure 5: *Mixture model.* Figure (a) shows the MISE of each estimator as a function of the threshold (determined by α) for 1 000 samples, each of size $n = 1\,000$. The thick line is the MISE of the Bayes estimator, the dotted line is the MISE of the MELE, and the dashed line is the MISE of the empirical estimator. The vertical line indicates the threshold for which the results of the simulation are plotted in (b), (c), and (d). We summarise the results by providing 95% pointwise confidence bands (shaded regions) and the medians (thick lines) for each estimator: Bayes (b), empirical (c), and MELE (d). In the three latter plots, the dashed line is the true spectral measure.

of probability α for margin j. For the two frequentist estimators, we compute corresponding values k_α in such a way that for each α and each sample, there are the same number of observations in the tail region determined by k_α as in the tail region $[u_{1\alpha}, \infty) \times [u_{2\alpha}, \infty)$. For each value of the threshold, the performance of the estimators is measured via the mean integrated square error

$$\mathrm{MISE}(\hat{H}) = \mathrm{E}\left[\int_0^1 \{\hat{H}(w) - H(w)\}^2 \, dw\right],$$

estimated by the sample mean of the integrated squared errors over the 1 000 samples. In Figures 5–7, the results are plotted as a function of the threshold.

As the results indicate, there does not seem to be a grand winner, although there is a grand loser, the empirical spectral measure. When comparing the results for the Bayes and

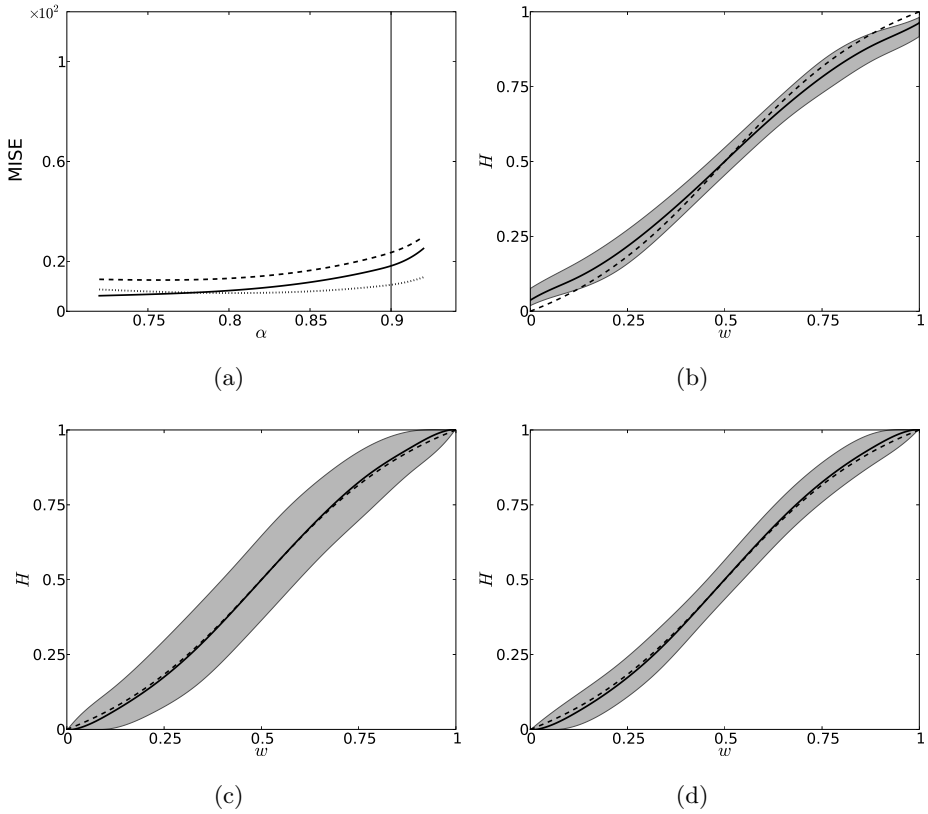


Figure 6: *Logistic model.* Figure (a) shows the MISE of each estimator as a function of the threshold (determined by α) for 1 000 samples, each of size $n = 1\,000$. The thick line is the MISE of the Bayes estimator, the dotted line is the MISE of the MELE, and the dashed line is the MISE of the empirical estimator. The vertical line indicates the threshold for which the results of the simulation are plotted in (b), (c), and (d). We summarise the results by providing 95% pointwise confidence bands (shaded regions) and the medians (thick lines) for each estimator: Bayes (b), empirical (c), and MELE (d). In the three latter plots, the dashed line is the true spectral measure.

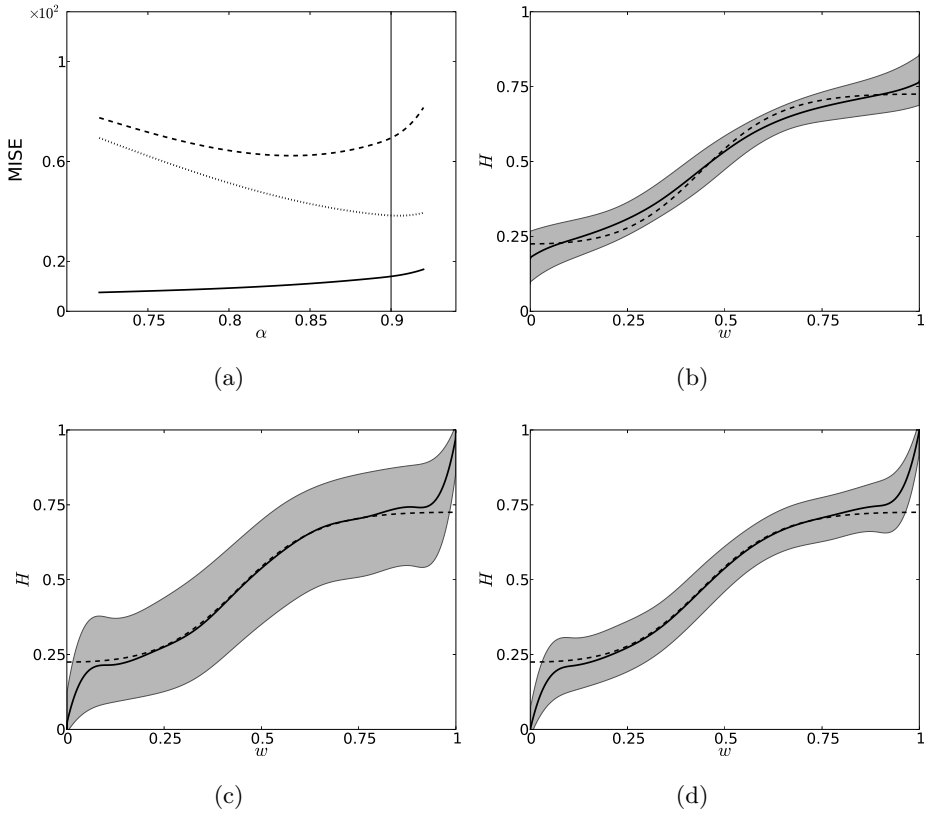


Figure 7: *Asymmetric logistic model.* Figure (a) shows the MISE of each estimator as a function of the threshold (determined by α) for 1 000 samples, each of size $n = 1\,000$. The thick line is the MISE of the Bayes estimator, the dotted line is the MISE of the MELE, and the dashed line is the MISE of the empirical estimator. The vertical line indicates the threshold for which the results of the simulation are plotted in (b), (c), and (d). We summarise the results by providing 95% pointwise confidence bands (shaded regions) and the medians (thick lines) for each estimator: Bayes (b), empirical (c), and MELE (d). In the three latter plots, the dashed line is the true spectral measure.

the MELE estimators, we see that both have their strenghts and their weaknesses. For instance, the MELE seems to do very well in the interior of the interval $(0, 1)$, but has problems at the end points, since the MELE estimator has no atoms at 0 and 1. In fact, much of its MISE accounts for this bias. One clear advantage of our estimator over the other ones is that it allows for atoms at 0 and 1, similar to many parametric models (Beirlant et al., 2004, chapter 9). Note also that the two frequentist estimators are not smooth; here we have smoothed them out using cubic splines. The Bayes estimator, on the other hand, is smooth by construction.

7.2 Danish fire insurance data

The data set comprises 2 167 industrial fire losses and was collected from the Copenhagen Reinsurance Company over the period 1980 to 1990. We are indebted to Alexander McNeil (Heriot-Watt University) for making these data available through his personal homepage[1]. The company's figure for compensatory damage is divided in three categories: *damage to building* (X_1), *damage to furniture and personal property* (X_2) and *loss of profits due to the incident* (X_3), see Figure 8.

In McNeil (1997) and Resnick (1997), the data-set is analysed as if it was univariate, by combining the three categories into a single loss figure, $X_1 + X_2 + X_3$. However, the three types of loss compensations involve different portfolios, and in view of this, it is in the insurance company's interest to know the dependence among these types of compensations. This is where our methodology becomes useful. Essentially, we are interested in the rare-event probabilities and the extreme conditional quantiles described in the introduction. Here, these tail quantities are derived respectively from the joint predictive density in (5.5) and the conditional predictive density in (5.6).

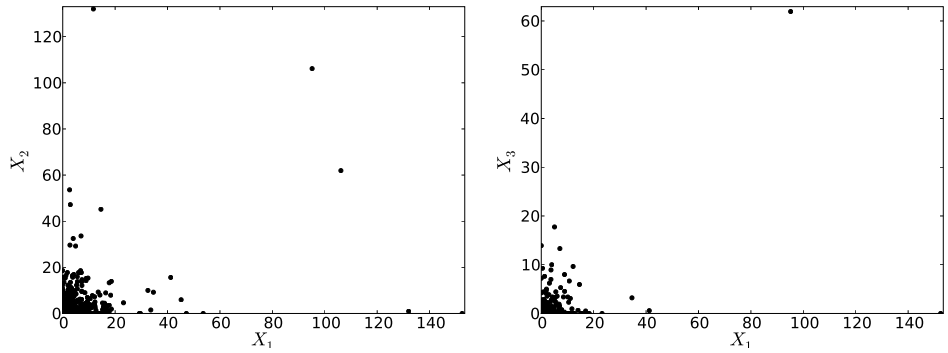
For descriptive purposes, we investigate the dependence structures for all pairs of compensatory damage categories. The Bayes estimates and 95% pointwise credibility sets for the cumulative distribution functions of the spectral measures are given in Figure 11. The thresholds u_1 and u_2 are chosen as the 90th percentiles of the respective margins.

We now focus on the first couple X_1 and X_2. Figure 11(a) clearly indicates that the two variables are not independent in the region $[u_1, \infty) \times [u_2, \infty)$. The joint predictive density (5.5) is shown in Figure 9. Finally, the mean of the conditional predictive density (5.6), that is the predicted value of the claim X_2 given the claim $X_1 = x_1$ and given $X_2 > u_2$, together with the 95% quantile of the predictive conditional distribution are shown in Figure 10.

8 Discussion

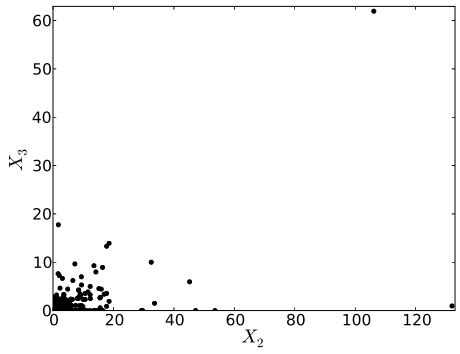
We have provided a nonparametric Bayesian framework for analysis of bivariate extremes. On the one hand, the nonparametric nature of the dependence structure (spectral measure) is fully respected. On the other hand, for purposes of prediction of future high levels (even

[1] `http://www.ma.hw.ac.uk/~mcneil/data.html`



(a) Damage to building (X_1) vs Damage to furniture and personal property (X_2)

(b) Damage to building (X_1) vs Loss of profits due to the incident (X_3)



(c) Damage to furniture and personal property (X_2) vs Loss of profits due to the incident (X_3)

Figure 8: Pairwise scatter plots of loss claims.

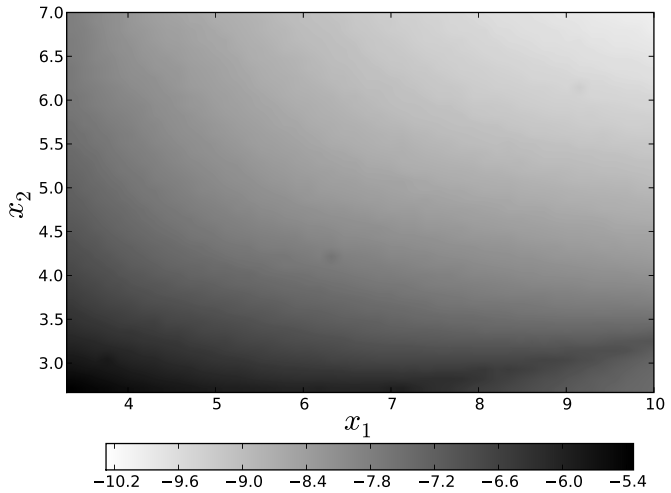


Figure 9: Image of the joint density (5.5), on log scale, in the region $[u_1, \infty) \times [u_2, \infty)$.

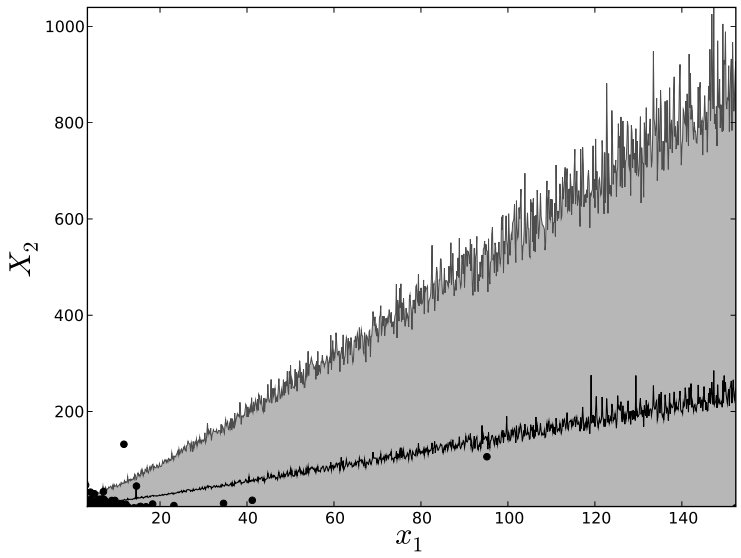
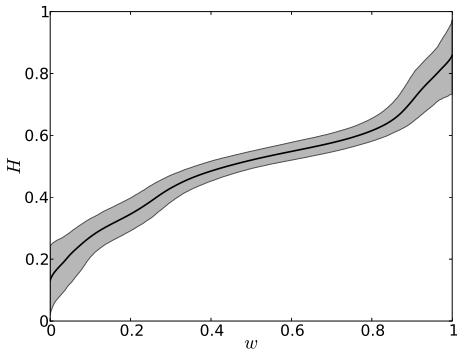


Figure 10: The data points, the predicted value of X_2 (black line) given the claim $X_1 = x_1$, along with the 95% pointwise quantiles (grey line) of the conditional predictive density (5.6).

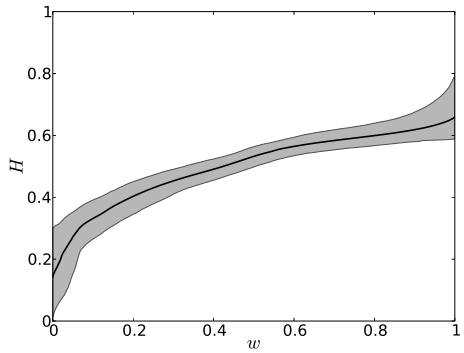
conditionally), the predictive distribution incorporates both process and estimation uncertainty. Actual computations are performed using a trans-dimensional MCMC algorithm. Software written in (parallel) C++ wrapped in a Python environment may be obtained from the authors.

Conceptually, it is not hard to see how to generalise the approach to arbitrary dimensions. Practically, however, there are some serious obstacles to be overcome. In dimension d the spectral measure may be an arbitrary probability measure on the $(d-1)$-dimensional unit simplex satisfying a certain number of moment constraints. It may have a density on each of the $2^d - 1$ faces of the unit simplex (Coles and Tawn, 1991). First, a prior needs to be specified on a manageable but still dense submodel of spectral measures, for instance by specifying the densities via splines or polynomials or some other set of basis functions. Second, efficient MCMC methodology should be proposed for numerical computations.

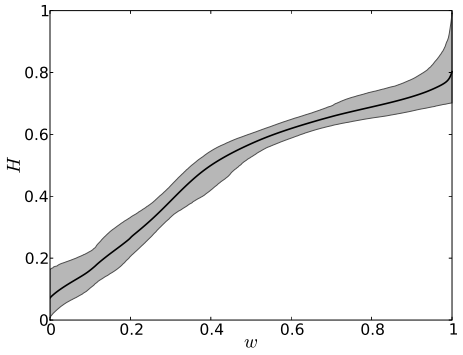
Finally, the bivariate tail approximation via extreme-value distributions is not well-suited to deal with asymptotic independence, in which case the tail approximation degenerates to exact independence, an approximation which may be unsatisfactory for instance in case of the bivariate Gaussian distribution with a high correlation. A distributional model encompassing both asymptotic independence and dependence has been proposed in Ramos and Ledford (2009), based on Ledford and Tawn (1996) and Ledford and Tawn (1997); see also Resnick (2003) and Beirlant et al. (2004).



(a) Damage to building (X_1) vs Damage to furniture and personal property (X_2)



(b) Damage to building (X_1) vs Loss of profits due to the incident (X_3)



(c) Damage to furniture and personal property (X_2) vs Loss of profits due to the incident (X_3)

Figure 11: Bayes estimates (thick line) and 95% pointwise credibility sets for the spectral measure for each couple of compensatory damage category are shown in (a), (b) and (c).

Acknowledgements

We are particularly grateful to the Co-Editor George Casella, to the Associate Editor and to the two referees for their comments and suggestions that have led to a much improved version of this paper. We also wish to thank Jonathan Taylor (Stanford University), Jeffrey Rosenthal (University of Toronto) for great advice and interesting discussions, and Alexander McNeil (Heriot-Watt University) for making available the Danish fire insurance data. Simon Guillotte and François Perron wish to acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC). Johan Segers gratefully acknowledges financial support from IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy) and from the contract "Projet d'Actions de Recherche Concertées" no. 07/12/002 of the Communauté française de Belgique, granted by the Académie universitaire Louvain. Finally, we want to acknowledge the Réseau Québécois de Calcul Haute Performance (RQCHP) for their computing facilities.

References

Aitchison, J. and I. R. Dunsmore (1975). *Statistical Prediction Analysis*. Cambridge: Cambridge University Press.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Ltd.

Boldi, M.-O. and A. C. Davison (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 69*(2), 217–229.

Chib, S. and E. Greenberg (1995). Understanding the Metropolis–Hastings algorithm. *The American Statistician 49*(4), 327–335.

Coles, S. and J. Tawn (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 53*, 377–392.

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. London: Springer-Verlag.

Coles, S. G. and J. A. Tawn (1994). Statistical methods for multivariate extremes: an application to structural design. *Applied Statistics 43*, 1–48.

Coles, S. G. and J. A. Tawn (1996). A Bayesian analysis of extreme rainfall data. *Applied Statistics 45*, 463–478.

Coles, S. G. and J. A. Tawn (2005). Bayesian modelling of extreme surges on the UK east coast. *Phil. Trans. R. Soc. A 363*, 1387–1406.

de Haan, L. and J. de Ronde (1998). Sea and wind: Multivariate extremes at work. *Extremes 1*(1), 7–45.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory*. Springer Series in Operations Research and Financial Engineering. New York: Springer.

de Haan, L., C. Neves, and L. Peng (2008). Parametric tail copula estimation and model testing. *Journal of Multivariate Analysis 99*(6), 1260–1275.

de Haan, L. and S. Resnick (1977). Limit theory for multidimensional sample extremes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete 40*, 317–337.

de Haan, L. and A. K. Sinha (1999). Estimating the probability of a rare event. *The Annals of Statistics 27*(2), 732–759.

Einmahl, J., L. de Haan, and V. Piterbarg (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *The Annals of Statistics 29*, 1401–1423.

Einmahl, J. H. J., L. de Haan, and D. Li (2006). Weighted approximations to tail copula processes with application to testing the bivariate extreme value condition. *The Annals of Statistics 34*(4), 1987–2014.

Einmahl, J. H. J., A. Krajina, and J. Segers (2008). A method of moments estimator of tail dependence. *Bernoulli 14*(4), 1003–1026.

Einmahl, J. H. J. and J. Segers (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *The Annals of Statistics 37*(5B), 2953–2989.

Fisher, R. (1929). Tests of significance in harmonic analysis. *Proceedings of the Royal Society of London, Series A 125*(796), 54–59.

Fremlin, D. H. (2003). *Measure Theory. Vol. 2.* Torres Fremlin, Colchester. Broad foundations, Corrected second printing of the 2001 original.

Fritsch, F. N. and J. Butland (1984). A method for constructing local monotone piecewise cubic interpolants. *SIAM J. Sci. Statist. Comput. 5*(2), 300–304.

Fritsch, F. N. and R. E. Carlson (1980). Monotone piecewise cubic interpolation. *SIAM J. Numer. Anal. 17*(2), 238–246.

Gamerman, D. and H. F. Lopes (2006). *Markov Chain Monte Carlo* (Second ed.). Texts in Statistical Science Series. Chapman & Hall/CRC, Boca Raton, FL. Stochastic simulation for Bayesian inference.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*(4), 711–732.

Guillotte, S. and F. Perron (2008). A Bayesian estimator for the dependence function of a bivariate extreme-value distribution. *The Canadian Journal of Statistics 36*(3), 383–396.

Hobert, J. P., C. P. Robert, and C. Goutis (1997). Connectedness conditions for the convergence of the Gibbs sampler. *Statistics & Probability Letters 33*(3), 235–240.

Joe, H., R. L. Smith, and I. Weissman (1992). Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 54*(1), 171–183.

Kotz, S. and S. Nadarajah (2000). *Extreme Value Distributions*. London: Imperial College Press. Theory and applications.

Ledford, A. W. and J. A. Tawn (1996). Statistics for near independence in multivariate extreme values. *Biometrika 83*(1), 169–187.

Ledford, A. W. and J. A. Tawn (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 59*, 475–499.

Marin, J.-M., K. L. Mengersen, and C. Robert (2005). Bayesian modelling and inference on mixtures of distributions. In D. Dey and C. Rao (Eds.), *Handbook of Statistics: Volume 25*. Elsevier.

McNeil, A. (1997). Estimating the tails of loss severity distributions using extreme value theory. *Astin Bulletin 27*, 117–137.

Meyn, S. and R. L. Tweedie (2009). *Markov Chains and Stochastic Stability* (Second ed.). Cambridge: Cambridge University Press. With a prologue by Peter W. Glynn.

Nadarajah, S., C. W. Anderson, and J. A. Tawn (1998). Ordered multivariate extremes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 60*(2), 473–496.

Ramos, A. and A. Ledford (2009). A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 71*(1), 219–241.

Resnick, S. (2003). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes 5*(4), 303–336.

Resnick, S. I. (1997). Discussion of the danish data on large fire insurance losses. *Astin Bulletin 27*, 139–151.

Robert, C. P. (2007). *The Bayesian Choice*. Springer Texts in Statistics. New York: Springer-Verlag.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods* (Second ed.). Springer Texts in Statistics. New York: Springer-Verlag.

Roberts, G. O. and J. S. Rosenthal (2006). Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains. *The Annals of Applied Probability 16*(4), 2123–2139.

Roberts, G. O. and A. F. M. Smith (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis–Hastings algorithms. *Stochastic Processes and their Applications 49*(2), 207–216.

Schmidt, R. and U. Stadtmüller (2006). Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics 33*(2), 307–335.

Smith, R. L. (1994). Multivariate threshold methods. In J. Galambos, J. Lechner, and E. Simiu (Eds.), *Extreme Value Theory and Applications*, pp. 225–248. Kluwer.

Tawn, J. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika 75*(3), 397–415.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics 22*(4), 1701–1762. With discussion and a rejoinder by the author.

Tierney, L. (1996). Introduction to general state-space Markov chain theory. In *Markov Chain Monte Carlo in Practice*, Interdisciplinary Statistics, pp. 59–74. London: Chapman & Hall.

A Appendix

A.1 Censored likelihood

We give explicit formulas for the censored density f^*. Let

$$F_j(x_j) = \exp\left\{-\zeta_j\left(1+\xi_j\frac{x_j-u_j}{\sigma_j}\right)^{-1/\xi_j}\right\}, \qquad j\in\{1,2\},$$

so that

$$f_j(x_j) = \frac{d}{dx_j}F_j(x_j) = \frac{\zeta_j}{\sigma_j}\left(1+\xi_j\frac{x_j-u_j}{\sigma_j}\right)^{-1/\xi_j-1}F_j(x_j), \qquad j\in\{1,2\},$$

for x_j such that $\sigma_j+\xi_j(x_j-u_j)>0$. Let

$$F(x_1,x_2) = \exp\left\{-\ell\big(-\log F_1(x_1), -\log F_2(x_2)\big)\right\},$$

where

$$\ell(s,t) = 2\int_{[0,1]}\max\big(ws,(1-w)t\big)\,dH(w), \qquad (s,t)\in[0,\infty)^2.$$

Write $h_a = H(\{a\})$ for $a\in\{0,1\}$. Assume H is absolutely continuous on $(0,1)$ with Radon-Nikodym derivative $h(\,\cdot\,)$. The mean constraint $\int_0^1 w\,dH(w)$ together with $H([0,1]) = 1$ implies

$$\int_0^1 w\,h(w)\,dw = 1/2-h_1 \quad\text{and}\quad \int_0^1(1-w)\,h(w)\,dw = 1/2-h_0.$$

Furthermore,

$$\frac{\partial}{\partial s}\ell(s,t) = 2\left(h_1 + \int_{\frac{t}{s+t}}^{1} w\,h(w)\,dw\right),$$
$$\frac{\partial}{\partial t}\ell(s,t) = 2\left(h_0 + \int_{0}^{\frac{t}{s+t}} (1-w)\,h(w)\,dw\right),$$
$$\frac{\partial^2}{\partial s\,\partial t}\ell(s,t) = -2\frac{st}{(s+t)^3}h\left(\frac{t}{s+t}\right),$$

and we have the representation

$$\ell(s,t) = s + t + \int_0^s \int_0^t \frac{\partial^2}{\partial\sigma\,\partial\tau}\ell(\sigma,\tau)\,d\sigma\,d\tau, \qquad (s,t)\in[0,\infty)^2.$$

Finally, recall $(x_1^*, x_2^*) = (x_1 \vee u_1, x_2 \vee u_2)$ and $d = (\mathbb{1}_{[u_1,\infty)}(x_1), \mathbb{1}_{[u_2,\infty)}(x_2))$. For $\{(x_1^*, x_2^*) : \sigma_j + \xi_j(x_j^* - u_j) > 0, j = 1, 2\}$, $f^*(x_1^*, x_2^*)$ is equal to:

$$\begin{cases} F(u_1,u_2) & \text{if } d=(0,0),\\ \frac{f_1(x_1)}{F_1(x_1)}\,\dfrac{\partial}{\partial s}\ell(s,-\log F_2(u_2))\Big|_{s=-\log F_1(x_1)}\,F(x_1,u_2) & \text{if } d=(1,0),\\ \frac{f_2(x_2)}{F_2(x_2)}\,\dfrac{\partial}{\partial t}\ell(-\log F_1(u_1),t)\Big|_{t=-\log F_2(x_2)}\,F(u_1,x_2) & \text{if } d=(0,1),\\ \left(\prod_{j=1}^2 \frac{f_j(x_j)}{F_j(x_j)}\right)\Delta\ell(-\log F_1(x_1),-\log F_2(x_2))F(x_1,x_2) & \text{if } d=(1,1), \end{cases}$$

where

$$\Delta\ell(s,t) = \frac{\partial}{\partial s}\ell(s,t)\frac{\partial}{\partial t}\ell(s,t) - \frac{\partial^2}{\partial s\partial t}\ell(s,t).$$

A.2 Construction of an interpolating monotone cubic spline

Let $0 = y_0 < y_1 < y_2 < \cdots < y_k < y_{k+1} = 1$ be a set of abscissas, and let $\varphi_0 \leqslant \varphi_1 \leqslant \varphi_2 \leqslant \cdots \leqslant \varphi_k \leqslant \varphi_{k+1}$ be a set of ordinates. We construct a piecewise cubic polynomial function $\varphi(\cdot)$ such that

$$\varphi(y_i) = \varphi_i, \qquad \text{for all} \quad i = 0,\ldots,k+1,$$

and such that $\varphi(\cdot)$ is nondecreasing and continuously differentiable on $(0,1)$.

On every interval $[y_i, y_{i+1}]$, $i = 0,\ldots,k$, we expand φ around y_i and we get

$$\varphi(y) = \frac{d_i + d_{i+1} - 2\Delta_i}{h_i^2}(y-y_i)^3 + \frac{-2d_i - d_{i+1} + 3\Delta_i}{h_i}(y-y_i)^2 + d_i(y-y_i) + \varphi_i,$$

for all $y \in [y_i, y_{i+1}]$, where $\Delta_i = (\varphi_{i+1} - \varphi_i)/h_i$, $h_i = y_{i+1} - y_i$, and where d_i and d_{i+1} are the endpoint derivatives. Thus, the construction of the spline depends only on the specification of the set of endpoint derivatives $0 \leqslant d_0, d_1, d_2, \cdots, d_k, d_{k+1}$. Fritsch and Carlson (1980) give necessary and sufficient conditions on these values in order to guaranty monotonicity throughout $[0,1]$. In particular, setting $d_i = 0$ for all $i = 0,\ldots,k+1$, always produces a

continuously differentiable nondecreasing interpolant, although the resulting curve is not very smooth. On the other hand, Fritsch and Butland (1984) propose a method for determining the endpoint derivatives which produces smoother curves. In fact, it suffices to set $d_0 = 0 = d_{k+1}$ and

$$d_i = \begin{cases} \dfrac{\Delta_{i-1}\Delta_i}{\alpha\Delta_i + (1-\alpha)\Delta_{i-1}} & \text{if } \Delta_{i-1}\Delta_i > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha = \frac{1}{3}(1 + \frac{h_i}{h_{i-1}+h_i})$, for $i = 1, \ldots, k$.

A.3 Proof of Lemma 4.1

Write Θ_{m,h_0,h_1} as

$$\Theta_{m,h_0,h_1} = m\frac{1/2 - h_1}{1 - h_0 - h_1}\, S_{m,h_0,h_1},$$

where, in view of (4.3),

$$S_{m,h_0,h_1} = \left\{(x_1, \ldots, x_{m-1}) : 0 < x_1, \ldots, x_{m-1}, x_m < \frac{1 - h_0 - h_1}{m(1/2 - h_1)}, \text{where } x_m = 1 - \sum_{i=1}^{m-1} x_i\right\}.$$

We have $S_{m,h_0,h_1} \subset S_m$, with $S_m = \{(x_1, \ldots, x_{m-1}) \in [0,1]^{m-1} : \sum_i x_i \leqslant 1\}$ the $(m-1)$-dimensional unit simplex. Therefore

$$\lambda_{m-1}(\Theta_{m,h_0,h_1}) = \left(\frac{m(1/2 - h_1)}{1 - h_0 - h_1}\right)^{m-1} \lambda_{m-1}(S_{m,h_0,h_1}).$$

It is well known that $\lambda_{m-1}(S_m) = 1/(m-1)!$, which is the normalising constant of the Dirichlet$(1, \ldots, 1)$ distribution, with $(1, \ldots, 1) \in \mathbb{R}^m$. As a consequence,

$$\begin{aligned} \lambda_{m-1}(S_{m,h_0,h_1}) &= \int_{S_{m,h_0,h_1}} d\lambda_{m-1}, \\ &= \lambda_{m-1}(S_m)\frac{\int_{S_{m,h_0,h_1}} d\lambda_{m-1}}{\lambda_{m-1}(S_m)}, \\ &= \frac{1}{(m-1)!}\,\mathrm{P}\left(Y_{(m)} < \frac{1 - h_0 - h_1}{m(1/2 - h_1)}\right), \end{aligned}$$

where $Y_{(m)} = \max\{Y_1, \ldots, Y_m\}$, with $Y_m = 1 - (Y_1 + \cdots + Y_{m-1})$ and $(Y_1, \ldots, Y_{m-1})$ uniformly distributed on S_m, that is, $Y = (Y_1, \ldots, Y_m)$ follows a Dirichlet$(1, \ldots, 1)$ distribution. Finally, equation (4.8) is derived from a result which can be found in Fisher (1929), namely

$$\mathrm{P}(Y_{(m)} < y) = 1 - \sum_{k=1}^{\lfloor 1/y \rfloor} (-1)^{k-1} \binom{m}{k} (1 - ky)^{m-1}, \qquad \text{for all } y \in [1/m, 1].$$

Equation (4.7) follows.

A.4 Proof of Proposition 6.1

The proof of Proposition 6.1 rests upon two ideas: first, any two points in an open and path-connected set E can be connected by a finite number of overlapping cubes; second, within a d-dimensional cube, every destination set can be reached in at most d moves. These ideas are formalised in the following two lemmas. The proof of the proposition comes at the end.

Lemma A.1 (Chain of Cubes)**.** *Let $\varnothing \neq E \subset \mathbb{R}^d$ be open and path-connected, and let $C(x,r)$ and $C(y,s)$ be two cubes in E. Then there exists a finite number of cubes $C_0, \ldots, C_M$ in E such that $C_0 = C(x,r)$, $C_M = C(y,s)$, and $C_{i-1} \cap C_i \neq \varnothing$ for every $i \in \{1,\ldots,M\}$.*

Proof of Lemma A.1. Let $\varphi : [0,1] \to E$ be a continuous path from $\varphi(0) = x$ to $\varphi(1) = y$. Since $\varphi([0,1])$ is a compact subset of the open set E, there exists $0 < \varepsilon \leqslant \min(r,s)$ such that $C(z,\varepsilon) \subset E$ for every point z on $\varphi([0,1])$. Otherwise, for every integer $n \geqslant 1$, we could find a point $z_n \in \varphi([0,1])$ such that $C(z_n, 1/n)$ would not be contained in E, that is, such that there would exist $q_n \in \mathbb{R}^d \setminus E$ with $\|z_n - q_n\|_\infty < 1/n$. By compactness, there would exist a subsequence $(z_{n_k})_{k\geqslant 1}$ such that $z_{n_k} \to z \in \varphi([0,1])$ as $k \to \infty$. Then necessarily also $q_{n_k} \to z$ as $k \to \infty$. The limit point z would therefore belong to the boundary of E, contradicting the fact that E is open.

Since φ is uniformly continuous, there exists an integer $M \geqslant 2$ such that $\|\varphi(t) - \varphi(s)\|_\infty \leqslant \varepsilon/2$ for all $s,t \in [0,1]$ such that $|t-s| \leqslant 1/M$.

Let $C_i = C(\varphi(i/M), \varepsilon)$ for $i \in \{0,\ldots,M\}$. By construction, $C_i \subset E$ for all $i \in \{0,\ldots,M\}$ and $\varphi(i/M)$ belongs to both C_{i-1} and C_i, for every $i \in \{1,\ldots,M\}$. □

Lemma A.2 (Within-Cube Transitions)**.** *Let $\varnothing \neq E \subset \mathbb{R}^d$ be open and let $(Y_t)_{t\in 0}^{\infty}$ be a homogeneous Markov chain on E with transition mechanism given in Proposition 6.1. Let $C \subset E$ be a cube and let $B \subset C$ be a Borel set such that $\lambda_E(B) > 0$. Then*

$$\mathrm{P}[Y_d \in B \mid Y_0 = y] > 0, \qquad \textit{for all } y \in C.$$

Proof of Lemma A.2. For notational convenience, we treat the case $d = 2$. Write $C = I_1 \times I_2$ for some non-empty open intervals I_1 and I_2. From $y = (y_1, y_2)$ in C, we will make two moves within C:

$$y = (y_1, y_2) \xrightarrow{\text{coordinate } i=1} y' = (y_1', y_2) \xrightarrow{\text{coordinate } i=2} y'' = (y_1', y_2').$$

More specifically:

1. first, coordinate $i = 1$ is chosen with probability $p_1(y)$, a new value $y_1' \in I_1$ is proposed according to the proposal density $q_1(y_1' \mid y)$, and this new value is accepted with probability $\alpha_1(y_1' \mid y)$, the new point being $y' = (y_1', y_2)$;
2. next, coordinate $i = 2$ is chosen with probability $p_2(y')$, a new value $y_2' \in I_2$ is proposed according to the proposal density $q_2(y_2' \mid y')$, and this new value is accepted with probability $\alpha_2(y_2' \mid y')$, the final point being $y'' = (y_1', y_2')$.

The final point y'' may or may not belong to B. By the Chapman–Kolmogorov equations,

$$\mathrm{P}[Y_2 \in B \mid Y_0 = y] \geqslant p_1(y) \int_{y_1' \in I_1} q_1(y_1' \mid y)\, \alpha_1(y_1' \mid y) \\ p_2(y') \left(\int_{y_2' \in I_2} q_2(y_2' \mid y')\, \alpha_2(y_2' \mid y')\, \mathbb{1}(y'' \in B)\, \mathrm{d}y_2' \right) \mathrm{d}y_1'.$$

The right-hand side is positive, being the integral of a nonnegative function which is positive on a set of positive measure. □

Proof of Proposition 6.1. Since E is open, we can write E as a countable union of open cubes (for instance, cubes inside E with rational vertices). At least one of these cubes, say C, must be such that $\lambda_E(B \cap C) > 0$. Replacing B by $B \cap C$, we can without loss of generality assume that $B \subset C$.

By Lemma A.1, we can find a chain of cubes $C_0, \ldots, C_M$, such that $C_M = C$ and $R_i := C_{i-1} \cap C_i \neq \varnothing$ for every $i \in \{1, \ldots, M\}$; see Figure 4. From $y \in C_0$, the chain can move via the rectangles $R_1, \ldots, R_M$ into B in at most $k = (M+1)d$ steps:

$$y \xrightarrow{d \text{ steps}} R_1 \xrightarrow{d \text{ steps}} R_2 \xrightarrow{d \text{ steps}} \cdots \xrightarrow{d \text{ steps}} R_M \xrightarrow{d \text{ steps}} B,$$

each arrow being justified by Lemma A.2. For instance, if $M = 2$, then for every $y \in C_0$,

$$\mathrm{P}[Y_{3d} \in B \mid Y_0 = y] \geqslant \int_{y^{(1)} \in R_1} \mathrm{P}[Y_d \in \mathrm{d}y^{(1)} \mid Y_0 = y] \\ \int_{y^{(2)} \in R_2} \mathrm{P}[Y_{2d} \in \mathrm{d}y^{(2)} \mid Y_d = y^{(1)}]\, \mathrm{P}[Y_{3d} \in B \mid Y_{2d} = y^{(2)}],$$

which is positive, since for all $y \in C_0$, $y^{(1)} \in R_1$, and $y^{(2)} \in R_2$, each of the probabilities

$$\mathrm{P}[Y_d \in R_1 \mid Y_0 = y], \qquad \mathrm{P}[Y_{2d} \in R_2 \mid Y_d = y^{(1)}], \qquad \mathrm{P}[Y_{3d} \in B \mid Y_{2d} = y^{(2)}]$$

is positive by Lemma A.2. □

A.5 Path-connectedness

Lemma A.3. *For every integer $m \geqslant 1$, the set Θ_m is path-connected.*

Proof. As $\Theta_1 = (0, 1/2)^2$ is convex, only the case $m \geqslant 2$ needs consideration. For $h_0, h_1 \in (0, 1/2)^2$, write $\bar{y}(h_0, h_1) = (1/2 - h_1)/(1 - h_0 - h_1)$. Note that the point $(h_0, \bar{y}, \ldots, \bar{y}, h_1) \in \mathbb{R}^{m+1}$ belongs to Θ_m.

Let $\theta = (h_0, y_1, \ldots, y_{m-1}, h_1)$ and $\theta' = (h_0', y_1', \ldots, y_{m-1}', h_1')$ be two points inside Θ_m. Write $\bar{y} = \bar{y}(h_0, h_1)$ and $\bar{y}' = \bar{y}(h_0', h_1')$. A continuous path from θ to θ' is constructed by joining the following three pieces:

$$\begin{array}{ccc} (h_0, y_1, \ldots, y_{m-1}, h_1) & \xrightarrow{(1)} & (h_0, \bar{y}, \ldots, \bar{y}, h_1) \\ & & \downarrow (2) \\ & & (h_0', \bar{y}', \ldots, \bar{y}', h_1') \xrightarrow{(3)} (h_0', y_1', \ldots, y_{m-1}', h_1') \end{array}$$

The pieces (1) and (3) are just straight line segments, which belong to Θ_m by convexity of the set Θ_{m,h_0,h_1} in (4.5). For the second piece, let $[0,1] \ni t \mapsto (h_0(t), h_1(t))$ be the straight line segment connecting (h_0, h_1) and (h_0', h_1'), and write $\bar{y}(t) = \bar{y}(h_0(t), h_1(t))$. Then the second piece in the diagram above is given by the path $t \mapsto (h_0(t), \bar{y}(t), \ldots, \bar{y}(t), h_1(t))$. □

A.6 Models used in the simulation study

In the simulation study in Subsection 7.1, random samples were drawn from the following distributions:

- The bivariate *mixture* distribution function

$$F(x,y) = \left(1 - \frac{1}{x}\right)\left(1 - \frac{1}{y}\right)\left(1 + \frac{r}{x+y}\right), \qquad x \geqslant 1, y \geqslant 1,$$

  with parameter $r \in [0,1]$, see de Haan and Resnick (1977, Example 3) and Einmahl and Segers (2009, Example 5.2). Its margins are unit Pareto and its extreme-value attractor has spectral measure H_r given by

$$H_r(w) = \frac{1}{2}(1-r) + rw, \qquad w \in [0,1).$$

  In particular, it has atoms given by $H_r(\{0\}) = H_r(\{1\}) = (1-r)/2$. We set $r = 1/2$ in the simulation.

- The bivariate distribution function with unit Pareto margins and with extreme-value dependence structure given by the *asymmetric logistic model* in Tawn (1988), the spectral measure being

$$H_{r,\psi_1,\psi_2}(w) = \frac{1}{2}\big(1 + \psi_1 - \psi_2 - (\psi_1^r(1-w)^{r-1} - \psi_2^r w^{r-1})(\psi_1^r(1-w)^r + \psi_2^r w^r)^{1/r-1}\big),$$

  for $0 \leqslant w < 1$, in terms of parameters $r \geqslant 1$ and $\psi_1, \psi_2 \in [0,1]$. The atoms are given by $H_{r,\psi_1,\psi_2}(\{0\}) = (1-\psi_2)/2$ and $H_{r,\psi_1,\psi_2}(\{1\}) = (1+\psi_1)/2$. We look at a symmetric case with $\psi_1 = 1 = \psi_2$ and $r = 2$, also called the *logistic model*, while in the asymmetric case we consider $\psi_1 = 0.45$, $\psi_2 = 0.55$, and $r = 3$.