Bayesian Analysis (2024) **00**, Number 0, pp. 1

Supplementary material for "Heavy-Tailed NGG-Mixture Models"

Vianey Palacios Ramírez*, Miguel de Carvalho,* and Luis Gutiérrez†

1 Basics on regular variation and extreme value theory

1.1 Background

For completeness, this section briefly reviews background on the interface between regular variation and extreme value theory. We begin with the definition of *regularly varying* function.

Definition 1. *A measurable function $f : \mathbb{R}^+ \to \mathbb{R}$, that is eventually positive, is regularly varying at infinity if*

$$\frac{f(ty)}{f(t)} \underset{t\to\infty}{\longrightarrow} y^\alpha, \quad y > 0, \tag{1.1}$$

for some $\alpha \in \mathbb{R}$.

The parameter α is known as the *index of regular variation*, and it is common to write $f \in \mathrm{RV}_\alpha$ to denote that f is regularly-varying (at infinity). Straightforward examples of regularly varying functions include

$$f_1(y) = 1/y, \qquad f_2(y) = y^2 \log y,$$

with indices of regular variation $\alpha_1 = -1$ and $\alpha_2 = 2$, respectively. A function is called *slowly varying* if it satisfies (1.1) with $\alpha = 0$. If $f \in \mathrm{RV}_\alpha$, then $f(y)/y^\alpha$ is a slowly varying function; thus, by setting $\mathscr{L}(y) = f(y)/y^\alpha$, we can always represent a regularly varying function with index α as $f(y) = y^\alpha \mathscr{L}(y)$. Indeed, for the latter examples it follows that $f_1(y) = y^{-1}\mathscr{L}_1(y)$, with $\mathscr{L}_1(y) = 1$, and $f_2(y) = y^2\mathscr{L}_2(y)$, with $\mathscr{L}_2(y) = \log y$.

As mentioned in the main paper, heavy-tailed distributions can be characterized via regular variation. A random variable Y with distribution function $F(y) = P(Y \leq y)$ has a heavy (right) tail if there exist a constant $\alpha > 0$ such that

$$\lim_{y\to\infty} \frac{1 - F(yt)}{1 - F(y)} = t^{-\alpha}. \tag{1.2}$$

Hence, there exists a slowly varying function $\mathscr{L}(y)$ such that $1 - F(y) = y^{-\alpha}\mathscr{L}(y)$. The parameter α is the *tail index* and $\gamma = 1/\alpha$ is the *extreme value index*. A straightforward example of a distribution with a regularly varying tail is the standard Pareto distribution. The standard

*School of Mathematics, Statistics and Physics; Newcastle University, UK. kramirez@ed.ac.uk
†School of Mathematics; The University of Edinburgh, UK. miguel.decarvalho@ed.ac.uk
‡Department of Mathematics; Universidade de Aveiro, Portugal.
§Pontificia Universidad Católica de Chile, Santiago, Chile llgutier@mat.uc.cl

Pareto distribution with shape paremeter $\alpha > 0$ has survival function $1 - F(y) = y^{-\alpha}$ for $y > 1$; hence, $1 - F(y) \in \text{RV}_{-\alpha}$ with slowly varying function $\mathscr{L}(y) = 1$. Trivially, the tail of the standard normal distribution $Y \sim N(0, 1)$,

$$P(Y > y) \sim \frac{1}{y\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right), \quad \text{as } y \to \infty,$$

is not regularly varying; indeed it has much lighter tail than a Pareto distribution.

The following theorem characterizes the class of extreme value distribution and defines the *(maximum) domains of attraction*. Below $M_n = \max(Y_1, \ldots, Y_n)$ is the sample maximum of independent and identically distributed random variables $Y_1, \ldots, Y_n$.

Theorem 1 (Extremal types theorem)**.** *Suppose that there exists a sequence of constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that*

$$P\{(M_n - b_n)/a_n \leq y\} \to G(y), \quad \text{as } n \to \infty,$$

where $G(y)$ is non-degenerate. Then, $G(y) \equiv G_\xi(ay + b)$ with $a > 0, b \in \mathbb{R}$, where

$$G_\xi(y) = \exp\{-(1 + \xi y)^{-1/\xi}\}, \quad 1 + \xi y > 0, \tag{1.3}$$

with $\xi \in \mathbb{R}$; for $\xi = 0$ the right-hand side is interpreted as $\exp(-e^{-y})$.

The cases $\gamma > 0, \gamma = 0$, and $\gamma < 0$ respectively correspond to the Fréchet, Gumbel, and Weibull domains of attraction. To put it differently, the cases $\gamma > 0$, $\gamma = 0$, and $\gamma < 0$ respectively correspond to heavy-tailed, light-tailed, and short-tailed distributions. For example, the beta distribution is in the Weibull domain of attraction, the normal distribution is in the Gumbel domain of attraction, and the Pareto distribution is in the Fréchet domain of attraction.

As a final remark, the terminology of heavy-tails in the Fréchet sense is indeed coherent with (1.2). Indeed, for (1.3) with $\xi > 0$, if we set $b = -a = -1/\xi$, it follows that $1 - G(y) = 1 - G_\xi((y-1)/\xi) = 1 - \exp(-y^{-1/\xi})$, for $y > 0$; and hence $1 - F(y) \in \text{RV}_{-1/\xi}$, with slowly varying function $\mathscr{L}(y) = 1 - y^{-1/\xi}/2 + o(y^{-1/\xi})$.

1.2 Notes and Comments

An encyclopedic treatment on regular variation can be found in Bingham et al. (1989). Regular variation has a long tradition in probability theory, at least since the celebrated monograph by Feller (1971). The definitions and main results above can be found, for example, in de Haan and Ferreira (2006, Appendix B). The extremal types theorem follows from Fisher and Tippett (1928), and it was made precise by Gnedenko (1943).

2 Supporting information for Section 2

2.1 Details on the lower envelope for the NGG

Derivation of (2.7) in the paper follows from the following well-known result on the long-run behavior of subordinators.

Lemma 1 (Bertoin, 1996, Theorem 13)**.** *Let $\{S(t)\}$ be a subordinator with drift zero. Let $u(t)$ be an increasing function such that $u(t)/t$ is a real-valued function that is positive, continuous, and increasing, then*

$$\limsup_{t\to\infty} \frac{S(t)}{u(t)} = \begin{cases} 0, & \int_1^\infty \nu[u(t),\infty)\,\mathrm{d}t < \infty, \\ \infty, & \int_1^\infty \nu[u(t),\infty)\,\mathrm{d}t = \infty, \end{cases} \quad a.s.$$

The derivation of (2.7) of proceeds as follows. Let $D \in (0,1)$. From Theorem 2 in the paper, it follows that a.s. there exists $t > 0$ such that for all $y > t$,

$$1 - G(y) \geq l\{M(1 - G_0(y))\}D(1-D)^{(1-D)/D}/S(M).$$

Let $M^* \gg 0$ be such that $M^* > M$; it follows that $S(M^*) \geq S(M)$. We can now use Lemma 1 to find a lower bound for $1/S(M)$. First, consider the following family of functions for $r > 0$,

$$w_r(t) = t^{1/D}|\log t|^{r/D} = t^{1/D}(\log t)^{r/D}, \quad t \in (e^{-r}, \infty), \tag{2.1}$$

We start by checking if $w_r(t)$ verifies the assumptions of Lemma 1. It follows that $w_r(t)$ is increasing on (δ,∞), with $\delta = \mathrm{e}^{-r}$. Clearly, $t^{-1}w_r(t)$ is positive and continuous over $(1,\infty)$. In addition, $t^{-1}w_r(t)$ is increasing on $t \in (\delta_1,\infty)$, where $\delta_1 = e^{r/(D-1)}$. The function $w_r(t)$ is increasing in a sufficiently small neighborhood of infinity, and outside this neighborhood it can be extended linearly with slope 0 so that is nondecreasing on $(0,\infty)$, in a similar fashion as Doss and Sellke (1982); we denote this extended function by $u_r(t)$. Hence, $u_r(t)$ obeys the assumptions of Lemma 1. Finally, we derive necessary and sufficient conditions for $\int_1^\infty \nu[u_r(t),\infty)\,\mathrm{d}t < \infty$. First, note that

$$\nu[u,\infty) = \int_u^\infty \frac{D}{\Gamma(1-D)} x^{-1-D}\,\mathrm{d}x = \frac{D}{\Gamma(1-D)}\left\{\frac{u^{-D}}{D}\right\},$$

Thus,

$$\int_1^\infty \nu[u_r(t),\infty)\,\mathrm{d}t < \infty \quad \Longleftrightarrow \quad \int_1^\infty u_r^{-D}(t)\,\mathrm{d}t < \infty. \tag{2.2}$$

Then, $\int_1^\infty u_r^{-D}(t)\,\mathrm{d}t = \int_1^\delta w_r^{-D}(e^{-r})\,\mathrm{d}t + \int_\delta^\infty w_r^{-D}(t)\,\mathrm{d}t$, where the first integral is finite, which means $\int_1^\infty u_r^{-D}(t)\,\mathrm{d}t$ is finite if and only if $\int_\delta^\infty w_r^{-D}(t)\,\mathrm{d}t < \infty$, which is satisfied when $r > 1$. Hence, Lemma 1 implies that

$$\limsup_{t\to\infty} \frac{S(t)}{u_r(t)} = \begin{cases} 0, & r > 1, \\ \infty, & 0 < r \leq 1. \end{cases} \quad \text{a.s.} \tag{2.3}$$

This implies $S(M) \leq u_r(M^*)$ for large M^*, hence yielding (2.7) in the paper

$$\begin{aligned} 1 - G(y) &\geq D(1-D)^{(1-D)/D}/S(M)\, l\{M(1-G_0(y))\} \\ &\geq D(1-D)^{(1-D)/D}/u_r(M^*)\, l\{M(1-G_0(y))\} \\ &= D(1-D)^{(1-D)/D}/\{(M^*)^{1/D}\log(M^*)^{r/D}\}\, l\{M(1-G_0(y))\}. \end{aligned} \tag{2.4}$$

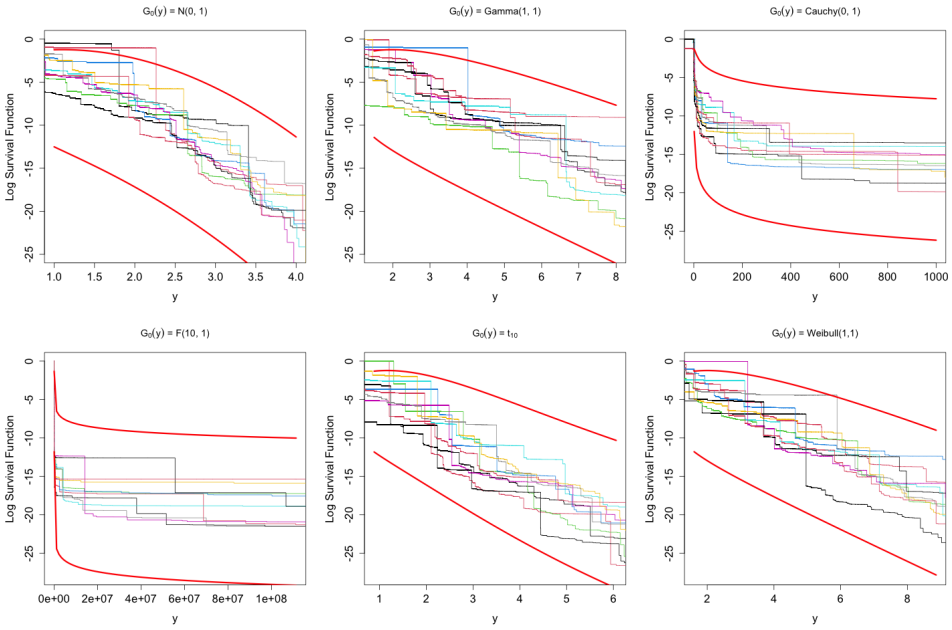

Figure 1: Asymptotic envelopes that follow from Theorem 2 in the main paper along with the log survival function of 10 random trajectories of a NGG$(1, 0, 0.5, G_0)$.

2.2 Numerical illustrations of asymptotic envelopes

Figure 1 presents the asymptotic envelopes that stem from Theorem 2 in the paper along with 10 random trajectories of a NGG$(1, 0, 0.5, G_0)$ process with parameter $D = 0.5$, centered at a variety of parametric models. The random trajectories were simulated using Algorithm 1 from Arbel et al. (2019), for exact simulation from an ε-stable process. As it can be seen from Figure 1, the random trajectories follow closely the asymptotic envelopes that stem from Theorem 2 in the main paper.

Figure 2 shows numerical illustrations of the asymptotic envelopes derived in Theorem 2 of the paper, in the case of NGG in $\mathcal{D}$ and in a particular case of NGG in $\mathcal{N}$ (the stable process), considering different values of M and D. As expected, choosing different values of D and M affects the tail's behavior.

2.3 Auxiliary results for the envelopes of Example 3

On the domain of attraction of Example 3

This section shows that the lower and upper envelopes in Example 3 based on a standard Pareto baseline are in the Gumbel domain of attraction—in the case of NGG in $\mathcal{D}$—and in the Fréchet domain of attraction—in the case of NGG in $\mathcal{N}$. The main result of this section is the following.

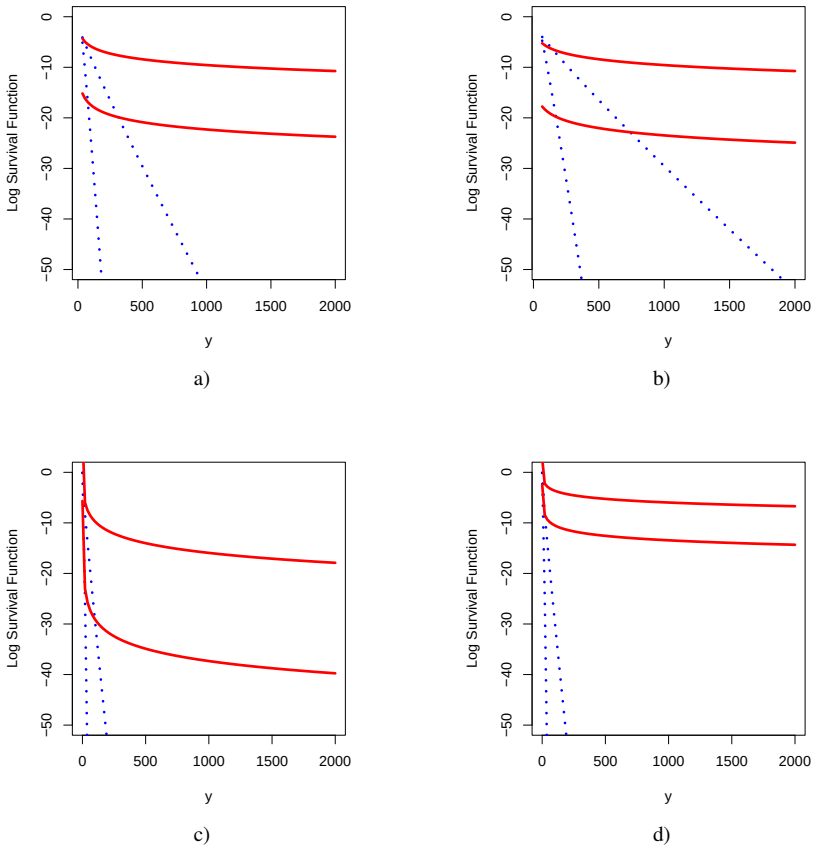

Figure 2: Asymptotic envelopes for Example 4 in the paper but for different values of M and D. The solid and dashed lines represent the stable process and the Dirichlet process, respectively. a): $D = 0.5, M = 5$. b): $D = 0.5, M = 10$. c): $D = 0.3, M = 1$. d): $D = 0.8, M = 1$.

Proposition 1. *Suppose G_0 is a standard Pareto distribution, $1-G_0(y)=y^{-1}$, for $y>1$. For large y:*

(a) If $G\sim NGG(M,\tau,D,G_0)$, with $(M,\tau,D,G_0)\in\mathcal{D}$ then $1-G$ is in the Gumbel domain of attraction, almost surely.

(b) If $G\sim NGG(M,\tau,D,G_0)$, with $(M,\tau,D,G_0)\in\mathcal{N}$, then $1-G$ is in the Fréchet domain of attraction, almost surely.

The proof of Proposition 1 is presented below, and it uses the following elementary lemma, which for completeness, we also prove below.

Lemma 2. *If $1-F(y)=\exp\{-y^{\tau}\mathscr{L}(y)\}$ where $\tau>0$ and $\mathscr{L}$ is a slowly varying function, then F belongs to the Gumbel domain of attraction.*

Proofs of Lemma 2 and Proposition 1

Proof of Lemma 2: Consider the auxiliary function $b(y)=y^{1-\tau}\mathscr{L}(y)^{-1}\tau^{-1}$, and note that

$$\begin{aligned}
\lim_{y\to\infty}\frac{b(y)}{b(y+vb(y))} &= \lim_{y\to\infty}\frac{y^{1-\tau}\mathscr{L}(y)^{-1}\tau^{-1}}{\{y+vb(y)\}^{1-\tau}\,\mathscr{L}\{y+vb(y)\}^{-1}\tau^{-1}}\\
&= \lim_{y\to\infty}\frac{y^{1-\tau}\mathscr{L}(y)^{-1}}{[y+v\{y^{1-\tau}\mathscr{L}(y)^{-1}\tau^{-1}\}]^{1-\tau}\mathscr{L}\{y+vb(y)\}^{-1}}\\
&= \lim_{y\to\infty}\frac{y^{1-\tau}\mathscr{L}(y)^{-1}}{y\{1+y^{-\tau}\mathscr{L}(y)^{-1}\tau^{-1}\}^{1-\tau}\,\mathscr{L}\{y+vb(y)\}^{-1}}\\
&= \lim_{y\to\infty}\frac{y^{1-\tau}\mathscr{L}(y)^{-1}}{y^{1-\tau}\{1+y^{-\tau}\mathscr{L}(y)^{-1}\tau^{-1}\}^{1-\tau}\,\mathscr{L}\{y+vb(y)\}^{-1}}\\
&= \lim_{y\to\infty}\frac{\mathscr{L}(y)^{-1}}{\{1+y^{-\tau}\mathscr{L}(y)^{-1}\tau^{-1}\}^{1-\tau}\,\mathscr{L}\{y+vb(y)\}^{-1}}\\
&= 1,
\end{aligned}$$

since,

$$\begin{aligned}
\lim_{y\to\infty}\frac{\mathscr{L}(y)^{-1}}{\mathscr{L}\{y+vb(y)\}^{-1}} &= \lim_{y\to\infty}\frac{\mathscr{L}[y+v\{y^{1-\tau}\mathscr{L}(y)^{-1}\tau^{-1}\}]}{\mathscr{L}(y)}\\
&= \lim_{y\to\infty}\frac{\mathscr{L}[y\{1+vy^{-\tau}\mathscr{L}(y)^{-1}\tau^{-1}\}]}{\mathscr{L}(y)}\\
&= 1.
\end{aligned}$$

Then, we can write the limit of interest as follows,

$$\lim_{y\to\infty}\frac{1-F(y+b(y)t)}{1-F(y)}=\lim_{y\to\infty}\exp\left\{-\int_y^{y+tb(y)}\frac{1}{b(u)}\mathrm{d}u\right\}.$$

Let $v = (u - y)/b(y)$ so that $u = y + vb(y)$; then, the integral above can be rewritten with the latter change of variables, and since the integrand converges locally uniformly to 1, $b(y)/b(y + vb(y)) \to 1$, it follows that

$$\begin{aligned}\lim_{y\to\infty} \frac{1 - F(y + b(y)t)}{1 - F(y)} &= \lim_{y\to\infty} \exp\left\{-\int_0^t \frac{b(y)}{b(y + vb(y))}\mathrm{d}v\right\} \\ &= \exp\left\{-\int_0^t \lim_{y\to\infty} \frac{b(y)}{b(y + vb(y))}\mathrm{d}v\right\} \\ &= \exp\left\{-\int_0^t 1\,\mathrm{d}v\right\} \\ &= \exp\{-t\}.\end{aligned}$$

This proves that F is in the Gumbel domain of attraction.

Proof of Proposition 1:

(a) If $G \sim \text{NGG}(M, \tau, D, G_0)$, with $(M, \tau, D, G_0) \in \mathcal{D}$, then for large y:

$$\exp\left\{-yr/M \log|\log(M/y)|\right\} \leq 1 - G(y) \leq \exp\left[-y/\{M|\log(M/y)|^r\}\right], \quad \text{a.s.}$$

For the upper bound, note that $\mathscr{L}_U(y) = 1/\{M|\log(M/y)|^r\}$ and $\mathscr{L}_L(y) = r/M \log|\log(M/y)|$ are slowly varying functions. We can also see that for very large y, $y > M$, $|\log(M/y)| = -\log(M/y) = \log(y/M)$, then

$$\begin{aligned}\lim_{y\to\infty} \frac{\mathscr{L}_U(ty)}{\mathscr{L}_U(y)} &= \lim_{y\to\infty} \frac{1/\{M \log(yt/M)^r\}}{1/\{M \log(y/M)^r\}} \\ &= \lim_{y\to\infty} \frac{\log(y/M)^r}{\log(yt/M)^r}.\end{aligned}$$

Let $\ell_U(y) = \log(y/M)$, then

$$\begin{aligned}\lim_{y\to\infty} \frac{\ell_U(ty)}{\ell_U(y)} &= \lim_{y\to\infty} \frac{\log(yt/M)}{\log(y/M)} = \frac{\infty}{\infty}, \quad \text{L'hopital,} \\ &= \lim_{y\to\infty} \frac{1/y}{1/y} \\ &= 1.\end{aligned}$$

Recall that if $\mathscr{L}$ is a slowly varying function, then $\mathscr{L}^\alpha$ is slowly varying as well for all $\alpha \in \mathbb{R}$, and in particular for $\alpha = r$. In addition, note that $\ell_U(y) = \log(y/M)$ is a slowly varying function, and hence $\lim_{y\to\infty} \mathscr{L}_U(ty)/\mathscr{L}_U(y) = 1$. For the lower bound consider,

$$\begin{aligned}\lim_{y\to\infty} \frac{\mathscr{L}_L(yt)}{\mathscr{L}_L(y)} &= \lim_{y\to\infty} \frac{\log\{\log(y/M)\}}{\log\{\log(yt/M)\}} = \frac{\infty}{\infty}, \quad \text{L'hopital,} \\ &= \lim_{y\to\infty} \frac{\mathscr{L}_L'(ty)}{\mathscr{L}_L'(y)}\end{aligned}$$

$$
\begin{aligned}
&= \lim_{y\to\infty} \frac{\log(yt/M)}{\log(y/M)} = \frac{\infty}{\infty}, \text{ L'hopital,} \\
&= \lim_{y\to\infty} \frac{1/y}{1/y} \\
&= 1. \qquad (2.5)
\end{aligned}
$$

Note that (2.5) is satisfied for any values of r and s. Then, for large y, we can rewrite the bounds in terms of upper and lower slowly varying functions as follows,

$$\exp\left\{-y\mathscr{L}_L(y)\right\} \leq 1 - G(y) \leq \exp\left\{-y\mathscr{L}_U(y)\right\}, \quad \text{a.s.}$$

Thus, the tail $1 - G(y)$ is bounded from both sides by $1 - F_l(y) = \exp\left\{-y\mathscr{L}_L(y)\right\}$ and $1 - F_u(y) = \exp\left\{-y\mathscr{L}_U(y)\right\}$, where by Lemma 2, F_l and F_u are in the Gumbel domain of attraction. Hence, there exist auxiliary functions $u(y)$ and $l(y)$ such that,

$$\lim_{y\to\infty} \frac{1 - G(y + u(y)t)}{1 - F_u(y)} \leq e^{-t},$$

and,

$$\lim_{y\to\infty} \frac{1 - G(y + l(y)t)}{1 - F_l(y)} \geq e^{-t},$$

where $u(y) = \mathscr{L}_U(y)^{-1}$ and $l(y) = \mathscr{L}_L(y)^{-1}$. Thus, we have that almost surely,

$$1 - G(y) = \exp\{-y^{1+o(1)}\}.$$

Hence, by defining an auxiliary function $b(y) = y^{-o(1)}$, satisfying $b(y)/b(y + vb(y)) \to 1$, it follows that

$$\lim_{y\to\infty} \frac{1 - G(y + b(y)t)}{1 - G(y)} = e^{-t},$$

and thus it finally follows from Proposition 2.1 in Beirlant et al. (2004) that $1 - G$ is in the Gumbel domain of attraction, almost surely.

(b) The result follows directly from Corollary 1 in the paper, as G_0 has a regularly varying tail. Indeed, G has a regularly varying tail with tail index $1/D$, and hence $1 - G$ is in the Fréchet domain of attraction.

2.4 Details on (A.8)

A step-by-step derivation of (A.8) is as follows. First, recall that if L varies slowly at infinity, it follows from Proposition 1.3.6 in Bingham et al. (1989), that

$$\lim_{x\to\infty} \log L(x)/\log x = 0 \iff \lim_{x\to 0} \log L(1/x)/\log x = 0. \qquad (2.6)$$

Equivalently, (2.6) can be rewritten using Landau's notation as

$$\log L(1/x) = o(1)\log x \iff \log L(1/x) = \log x^{o(1)} \iff L(1/x) = x^{o(1)}, \qquad (2.7)$$

as $x \to 0$. Secondly, Theorem 2 in the paper yields that, for $r > 1$, a.s. eventually for a large y,

$$\ell(M\{1 - G_0(y)\}) \leq 1 - G(y) \leq u_r(M\{1 - G_0(y)\}), \tag{2.8}$$

with $\ell(t) \equiv t^{1/D}\mathscr{L}(t)$ and $u_r(t) \equiv t^{1/D}\mathcal{U}_r(t)$. Here,

$$\mathscr{L}(t) = C\ \log|\log t|/\{(\log|\log t| + \tau^D t)^{1/D} - \tau t^{1/D}\}, \quad \mathcal{U}_r(t) = |\log t|^{r/D},$$

are slowly varying functions at zero;[†] equivalently, it holds that $L_1(x) \equiv \mathscr{L}(1/x)$ and $L_2(x) \equiv \mathcal{U}_r(1/x)$ are slowly varying functions at infinity (Bingham et al., 1989, p. 18). The latter claim, combined with (2.7) implies that

$$L_1(1/x) = \mathscr{L}(x) = x^{o(1)}, \quad L_2(1/x) = \mathcal{U}_r(x) = x^{o(1)}, \tag{2.9}$$

as $x \to 0$. Hence, it follows from (2.8) and (2.9) that, for $r > 1$, a.s. eventually for a large y,

$$[M\{1 - G_0(y)\}]^{1/D}\,[M\{1 - G_0(y)\}]^{o(1)} \leq 1 - G(y) \leq [M\{1 - G_0(y)\}]^{1/D}\,[M\{1 - G_0(y)\}]^{o(1)}.$$

Finally, this implies that $1 - G(y) = [M\{1 - G_0(y)\}]^{1/D+o(1)}$, as $y \to \infty$, or equivalently

$$1 - G(y) = [M\{1 - G_0(y)\}]^{\{1+o(1)\}/D}.$$

3 Posterior inference algorithms

Here we develop the posterior inference algorithm to learn about the proposed heavy-tailed mixture models. For generality, we will focus on the versions from Section 3, but trivially this algorithm can be updated to fit the other models on the paper For implementations it is computationally convenient to consider the conditional independence kernel, and hence we discuss this specification below. The numerical procedure to be discussed below is based on Gibbs sampling. Specifically, we propose to use the slice sampler algorithm for infinite mixtures proposed by Walker (2007), and further developed by Kalli et al. (2011). The proposal in Walker (2007) adapts the number of components in the mixture according to data complexity, and—conditional on the number of components at each iteration—the posterior inference is straightforward.

For concreteness, below we focus on the model in Equation (2.8), but similar comments readily apply to the multivariate heavy-tailed shape mixtures in Equation (3.1) of the paper. Let $\{\mathbf{y}_i\}_{i=1}^n$ be a random sample, with $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,d})^{\mathrm{T}}$. The joint likelihood for $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ is

$$f_{\pi,\sigma,\eta_\sigma}(\mathbf{y}) = \prod_{i=1}^{n}\sum_{h=1}^{\infty}\pi_h \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,h}}), \tag{3.1}$$

where $\pi = \{\pi_j\}_{j\geq 1}$ and $\sigma = \{\boldsymbol{\sigma}_j\}_{j\geq 1}$ are respectively the infinite collections of weights and d-dimensional atoms, and η_σ denotes the remainder parameters. Let s_i be a latent variable such that

$$(\mathbf{y}_i \mid s_i = j) \sim \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,j}}),$$

[†]Recall that, extending Definition 1, f is slowly varying at zero if $f(ty)/f(t) \to 1$ as $t \to 0^+$, for every $y > 0$.

where $j = 1, 2, \ldots$ and $i = 1, \ldots, n$. The joint likelihood of $\mathbf{y}$ and $\mathbf{s} = (s_1, \ldots, s_n)$ is

$$f_{\pi,\sigma,\eta_\sigma}(\mathbf{y}, \mathbf{s}) = \prod_{j=1}^{\infty} \pi_j^{n_j} \prod_{\{i:s_i=j\}} \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,j}}), \tag{3.2}$$

where $\pi_j = P(s_i = j)$ and $n_j = \sum_{i=1}^{n} I(s_i = j)$. To avoid the computation of infinite terms in (3.2), Walker (2007) proposes to use the latent variables $\{u_i\}_{i=1}^{n}$, such that:

$$f_{\pi,\sigma,\eta_\sigma}(\mathbf{y}_i, u_i) = \sum_{j=1}^{\infty} I(u_i < \pi_j) \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,j}}). \tag{3.3}$$

In (3.3) only a finite number N of π_j's satisfies the condition $\{u_i < \pi_j\}$. In particular, $N = \max_i\{N_i\}$ and N_i is the smallest integer such that $\sum_{j=1}^{N_i} \pi_j > 1 - u_i$, see Walker (2007) and Kalli et al. (2011) for more details. Note that this implies that now $\pi = \{\pi_j\}_{j=1}^{N}$ and $\sigma = \{\boldsymbol{\sigma}_j\}_{j=1}^{N}$. Considering the latent variables $\mathbf{s}$ and $\mathbf{u} = (u_1, \ldots, u_n)$, the likelihood for one observation is

$$f_{\pi,\sigma,\eta_\sigma}(\mathbf{y}_i, s_i, u_i) = I(u_i < \pi_{s_i}) \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,s_i}}),$$

and hence the likelihood for the n observations can be expressed as

$$f_{\pi,\sigma,\eta_\sigma}(\mathbf{y}, \mathbf{s}, \mathbf{u}) = \prod_{i=1}^{n} I(u_i < \pi_{s_i}) \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,s_i}}). \tag{3.4}$$

Using the likelihood in (3.4), the posterior inference via a Gibbs sampler is straightforward. At each iteration, it is necessary to update π, σ and η_σ, along with the latent variables $\mathbf{s}$ and $\mathbf{u}$. The weights π_j will be computed with the stick-breaking construction, and then the updates will be performed for the sticks V_j.

Following the recommendations in Section 3 of the paper (Remark 1), we suggest implementing the algorithm using as a baseline a multivariate heavy-tailed distribution with Pareto Type II margins; see Equation (3.6) on the paper. Some details on posterior inference for the extreme value index are in order. Assuming the Jeffrey's prior $\mathsf{p}(\alpha_k) = 1/\alpha_k$, it follows that the posterior distribution is

$$\begin{aligned}
\mathsf{p}(\alpha_k \mid \text{else}) &\propto f(\sigma \mid \alpha_k)\mathsf{p}(\alpha_k) \\
&= \frac{1}{\alpha_k} \prod_{h=1}^{N} \beta^{\alpha_k} \alpha_k (\beta + \sigma_h)^{-\alpha_k - 1} \\
&\propto \text{Gamma}\left\{ N, \sum_{h=1}^{N} \log\left(\frac{\beta + \sigma_h}{\beta}\right) \right\}.
\end{aligned}$$

Algorithm 1 shows how to learn from data about the stable process scale mixture models in Section 3.1 of the paper. Step 6 of Algorithm 1 is not conjugate, and hence a Metropolis–Hasting step is required, which is implemented with a random walk strategy

proposing candidates D^* from a Beta distribution. Algorithm 2 shows how to learn about the predictor dependent model (16) in Section 3.2 of the paper, from a random sample $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$, where the covariate $\mathbf{x}_i \in \mathbb{R}^p$ for all i. The algorithm is tantamount to Algorithm 1, except for an additional step to update parameters $\boldsymbol{\beta}_j$ and minor changes in the updating steps of the sticks and latent variables. To update $\boldsymbol{\beta}_j$ a Metropolis–Hasting step is required, which was implemented with a random walk proposing candidates $\boldsymbol{\beta}_j^*$ from multivariate Normal distribution in $\mathbb{R}^p$. For Step 3, a Metropolis–Hasting step might also be required—depending on the kernel to be used. The Erlang kernel is particularly convenient as Steps 3 and 8 can be performed using a Gibbs sampler. For example, the version of the model in Remark 1 of the paper can be fitted using Algorithms 1 and 2 with the following details for Steps 3 and 8:

3 Sample $\boldsymbol{\sigma}_j$ from $\mathsf{p}(\boldsymbol{\sigma}_j \mid \text{else}) \propto G_0(\boldsymbol{\sigma}_j) \prod_{\{i:s_i=j\}} \prod_{k=1}^d \text{Er}(y_{i,k}; \lceil \sigma_{k,s_i} \rceil, \sigma_{k,s_i}/\lambda)$.

8 Sample λ from $\text{Gamma}\big(a_\lambda + \sum_{i=1}^n \sum_{k=1}^d \lceil \sigma_{k,s_i} \rceil\,,\; b_\lambda + \sum_{i=1}^n \sum_{k=1}^d y_{i,k}/\sigma_{k,s_i}\big)$.

Here, $\text{Er}(y; a, b)$ is the density of the Erlang distribution with shape $a \in \mathbb{N}$ and scale $b > 0$, and (8) is implemented with a random walk strategy proposing candidates $\boldsymbol{\sigma}_j^*$ from a multivariate truncated Normal distribution in $\mathbb{R}_+^d$. Some final comments on notation are in order. In Algorithms 1 and 2, $\text{B}(\cdot \mid a, b)$ is the density of a $\text{Beta}(a, b)$ distribution, and $D_{i,j} \equiv \exp(\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_j)/\{1 + \exp(\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_j)\}$.

Algorithm 1: Slice sampler for stable process scale mixtures

1 Initialize $N, \boldsymbol{\sigma}_1^{(0)}, \ldots, \boldsymbol{\sigma}_N^{(0)}, \eta_\sigma^{(0)}, V_1^{(0)}, \ldots, V_N^{(0)}, \mathbf{u}^{(0)}$ and $\mathbf{s}^{(0)}$.

2 Sample α_k for each coordinate k, from $\text{Gamma}(N, \sum_{h=1}^N \log\{(1 + \sigma_{k,h})\})$.

3 Sample $\boldsymbol{\sigma}_j$ from $\mathsf{p}(\boldsymbol{\sigma}_j \mid \text{else}) \propto G_0(\boldsymbol{\sigma}_j) \prod_{\{i:s_i=j\}} \prod_{k=1}^d K(y_{i,k}; \eta_{\sigma_{k,s_i}})$.

4 Sample V_j from $\text{Beta}\big(1 - D + \sum_{i=1}^n I(s_i = j)\,, jD + \sum_{i=1}^n I(s_i > j)\big)$, set $\pi_j = V_j \prod_{\ell<j}(1 - V_\ell)$.

5 Sample u_i from $(u_i \mid \text{else}) \sim \text{Unif}(0, \pi_{s_i})$, for $i = 1, \ldots, n$. Then, set N as the smallest integer for which $\sum_{j=1}^N \pi_j > 1 - u^*$, where $u^* = \min_i\{u_i\}$.

6 Sample D from $\mathsf{p}(D \mid \text{else}) \propto \prod_{d=i}^N \text{B}(v_i \mid 1 - D, iD)\, \text{B}(D \mid a_D, b_D)$.

7 With probability $\mathsf{p}(s_i = j \mid \text{else}) \propto I(j : \pi_j > u_i) \prod_{k=1}^d K(y_{i,k}; \eta_{\sigma_{k,s_i}})$, set $s_i = j$ for $i = 1, \ldots, n$.

8 Update any remainder parameters in η_σ via Metropolis–Hastings.

9 Repeat Steps 2 through 7 until reaching stationarity.

Algorithm 2: Slice sampler for conditional stable process scale mixtures

1 Initialize $N, \boldsymbol{\sigma}_1^{(0)}, \ldots, \boldsymbol{\sigma}_N^{(0)}, \eta_\sigma^{(0)}, V_1^{(0)}, \ldots, V_N^{(0)}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N, \mathbf{u}^{(0)}$ and $\mathbf{s}^{(0)}$.

2 Sample α_k for each coordinate k, from $\text{Gamma}(N, \sum_{h=1}^{N} \log\{(1+\sigma_{k,h})\})$.

3 Sample $\boldsymbol{\sigma}_j$ from $\mathsf{p}(\boldsymbol{\sigma}_j \mid \text{else}) \propto G_0(\boldsymbol{\sigma}_j) \prod_{\{i:s_i=j\}} \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,s_i}})$.

4 Sample $V_j(\mathbf{x}_i)$ from $\text{Beta}\big(1 - D_{i,j} + I(s_i = j)\,, jD_{i,j} + I(s_i > j)\big)$ and set $\pi_j(\mathbf{x}_i) = V_j(\mathbf{x}_i) \prod_{\ell<j}\{1 - V_\ell(\mathbf{x}_i)\}$, for $i = 1, \ldots, n$ and $j = 1, \cdots, N$.

5 Sample $\boldsymbol{\beta}_j$ from $\mathsf{p}(\boldsymbol{\beta}_j \mid \text{else}) \propto \exp\left(\boldsymbol{\beta}_j^{\mathrm{T}}[s^2\mathbf{I}_p]^{-1}\boldsymbol{\beta}_j\right) \prod_{\{i:s_i=j\}} \mathrm{B}(V_j(\mathbf{x}_i) \mid 1 - D_{i,j}, jD_{i,j})$.

6 Sample u_i from $(u_i \mid \text{else}) \sim \text{Unif}(0, \pi_{s_i}(\mathbf{x}_i))$, $\;i = 1, \ldots, n$. Then set N as the smallest integer for which $\sum_{j=1}^{N} \pi_j(\mathbf{x}_i) > 1 - u^*$, where $u^* = \min_i\{u_i\}$.

7 With probability $\mathsf{p}(s_i = j \mid \text{else}) \propto I(j : \pi_j(\mathbf{x}_i) > u_i) \prod_{k=1}^{d} K(y_{i,k}; \eta_{\sigma_{k,s_i}})$, set $s_i = j;\ i = 1, \ldots, n$.

8 Update any remainder parameters in η_σ via Metropolis–Hastings.

9 Repeat steps 2 through 7 until reaching stationarity.

4 Additional numerical results

This section provides additional numerical evidence. The one-shot experiment and Monte Carlo means of the fits for the bivariate and conditional scenarios from Table 2 in the paper are reported in Figures 3–5. Overall, the estimates accurately recover the true joint densities.

5 Additional empirical results

This section reports supplementary empirical analyses. Table 1 reports posterior mean estimates of the extreme value index of the generalized Pareto distribution (Coles, 2001, Section 4), for each stimuli. All estimates are above zero, confirming that the brainwave data from Section 5 in the paper are indeed heavy-tailed. Such estimates are fairly in line with the ones reported in Tables 2 and Table 3, which respectively contain the posterior mean extreme value index obtained using the unconditional and conditional stable process scale mixture models. Table 4 presents the p-values for the Ljung–Box test for independence; according to the Ljung–Box test, the observations for all the stimuli can be regarded as independent over time. Figure 6 depicts the qq-plots of the randomized quantile residuals Dunn and Smyth (1996), of the stable process scale mixture fitted in Section 5.3 on the paper. Finally, Figure 7 shows the q-q boxplot (Rodu and Kafadar, 2022) of the randomized quantile residuals for the shape mixture model of a Pareto Type II kernel, which should be compared against Figure 5 in the paper.

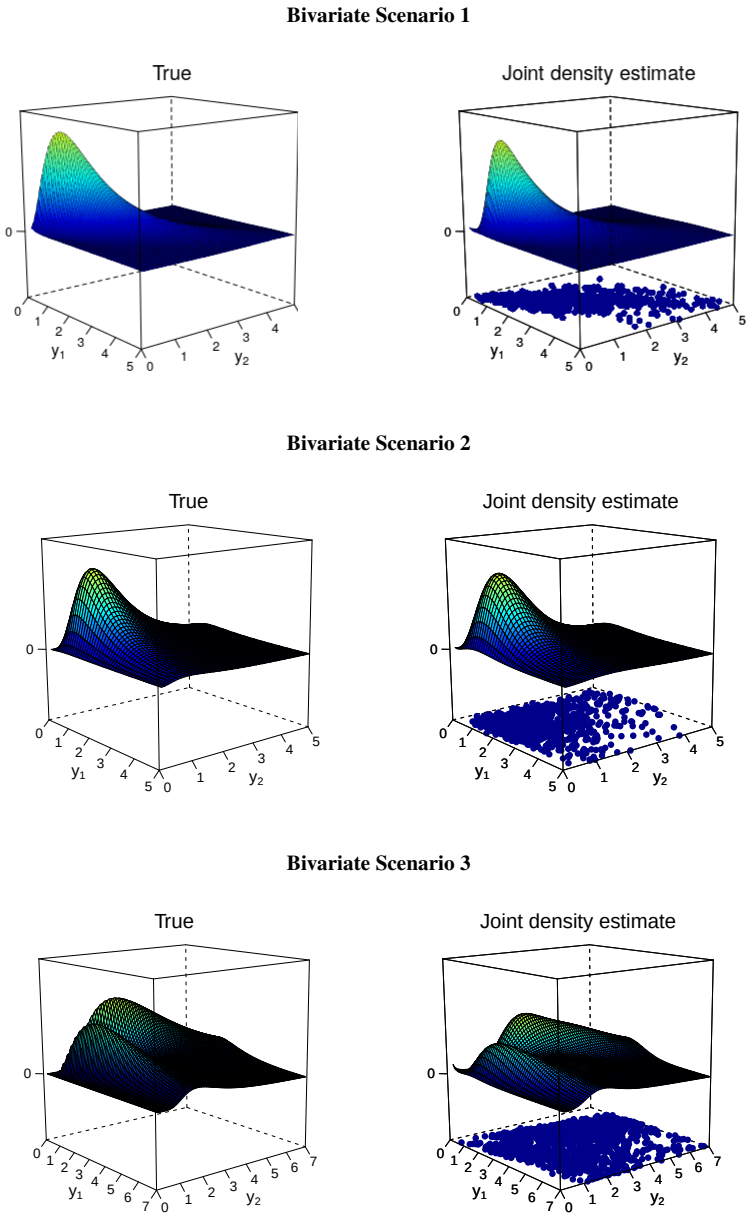

Figure 3: True (left) and estimated posterior densities (right) for Bivariate Scenarios 1–3 (top to bottom) obtained using the proposed stable process scale mixture model from Section 3.1 for a one-shot experiment. The simulated data for each scenario are overlaid at the bottom of the box.

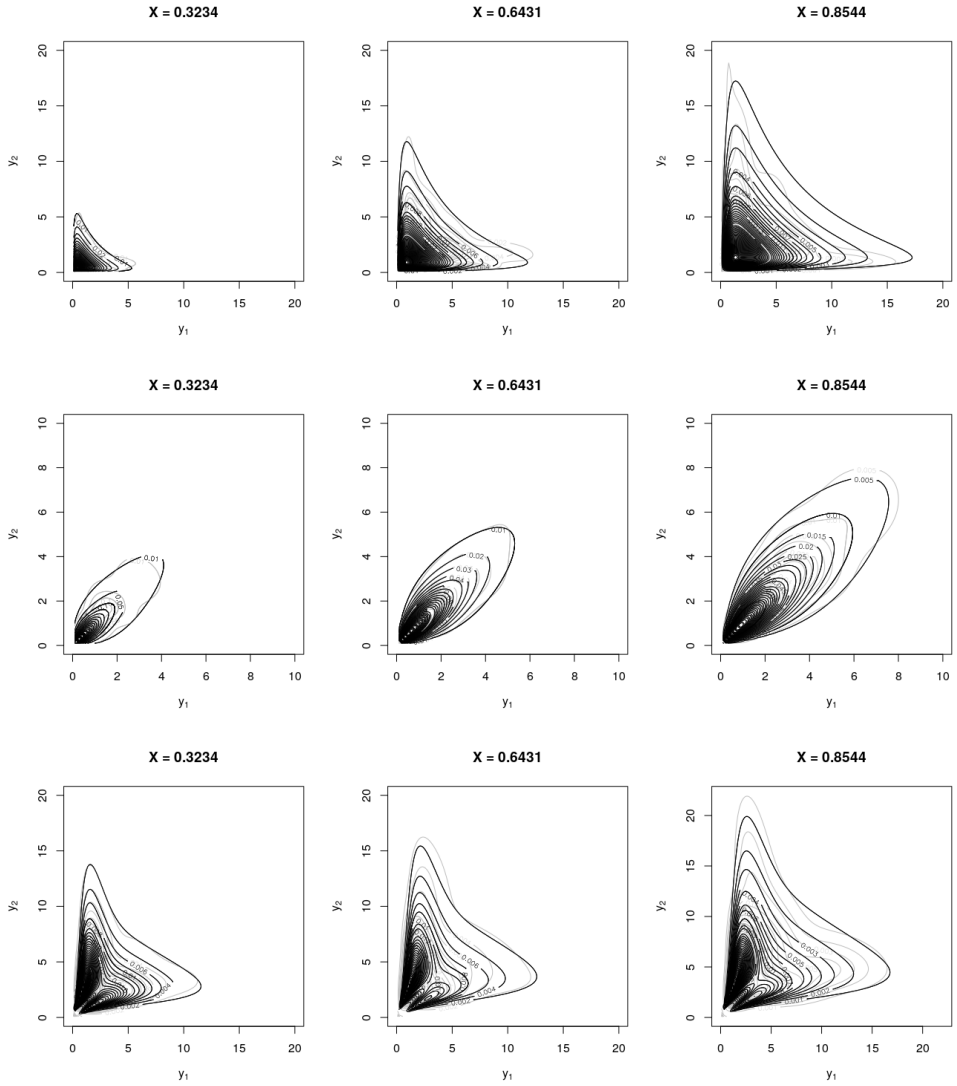

Figure 4: One-Shot Experiment for the Conditional Bivariate Scenarios 1–3: Contours of the conditional joint density estimates (gray)—given three levels of the covariate—obtained with proposed conditional stable process scale mixture model from Section 3.2 in the paper, for a one-shot experiment, plotted against the true (black).

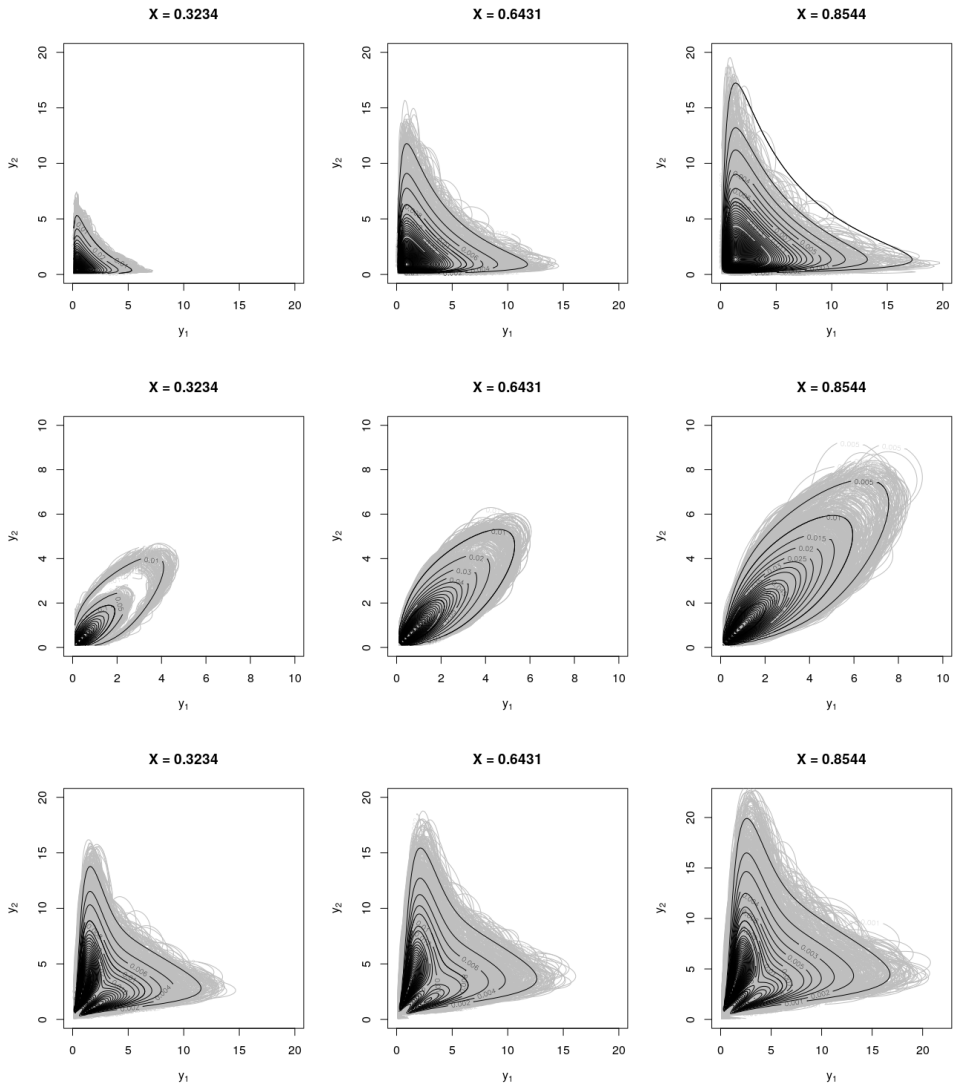

Figure 5: Monte Carlo Simulation for the Conditional Bivariate Scenarios 1–3: Contours of the conditional joint density estimates (gray)—given three levels of the covariate—obtained with proposed conditional stable process scale mixture model from Section 3.2 in the paper, for the 100 simulated datasets, plotted against the true (black).

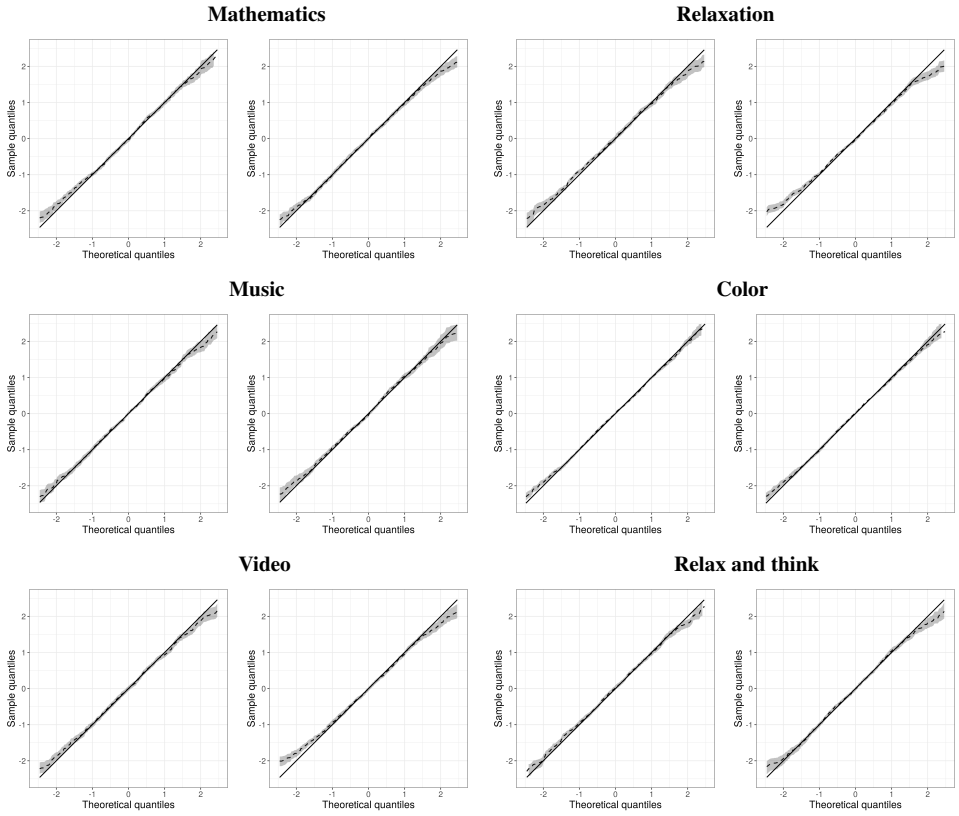

Figure 6: q-q plot of randomized quantile residuals for the proposed stable process scale mixture model from Section 2.2 of the paper, along with credible bands, against the Normal theoretical quantiles.

Stimulus	**Alpha Power**	**Beta Power**
Mathemathics	0.7938 (0.6819, 0.9015)	0.5167 (0.4355, 0.6034)
Relaxation	0.6140 (0.5083, 0.7226)	0.3306 (0.2581, 0.4078)
Music	0.7230 (0.6231, 0.8292)	0.3732 (0.3025, 0.4463)
Color	0.5160 (0.4551, 0.5977)	0.4138 (0.3598, 0.4705)
Video	0.6871 (0.5856, 0.7954)	0.3262 (0.2512, 0.4056)
Relax and think	0.7591 (0.6598, 0.8669)	0.2046 (0.1389, 0.2760)

Table 1: Posterior mean extreme value index of a generalized Pareto distribution along 95% with credible interval.

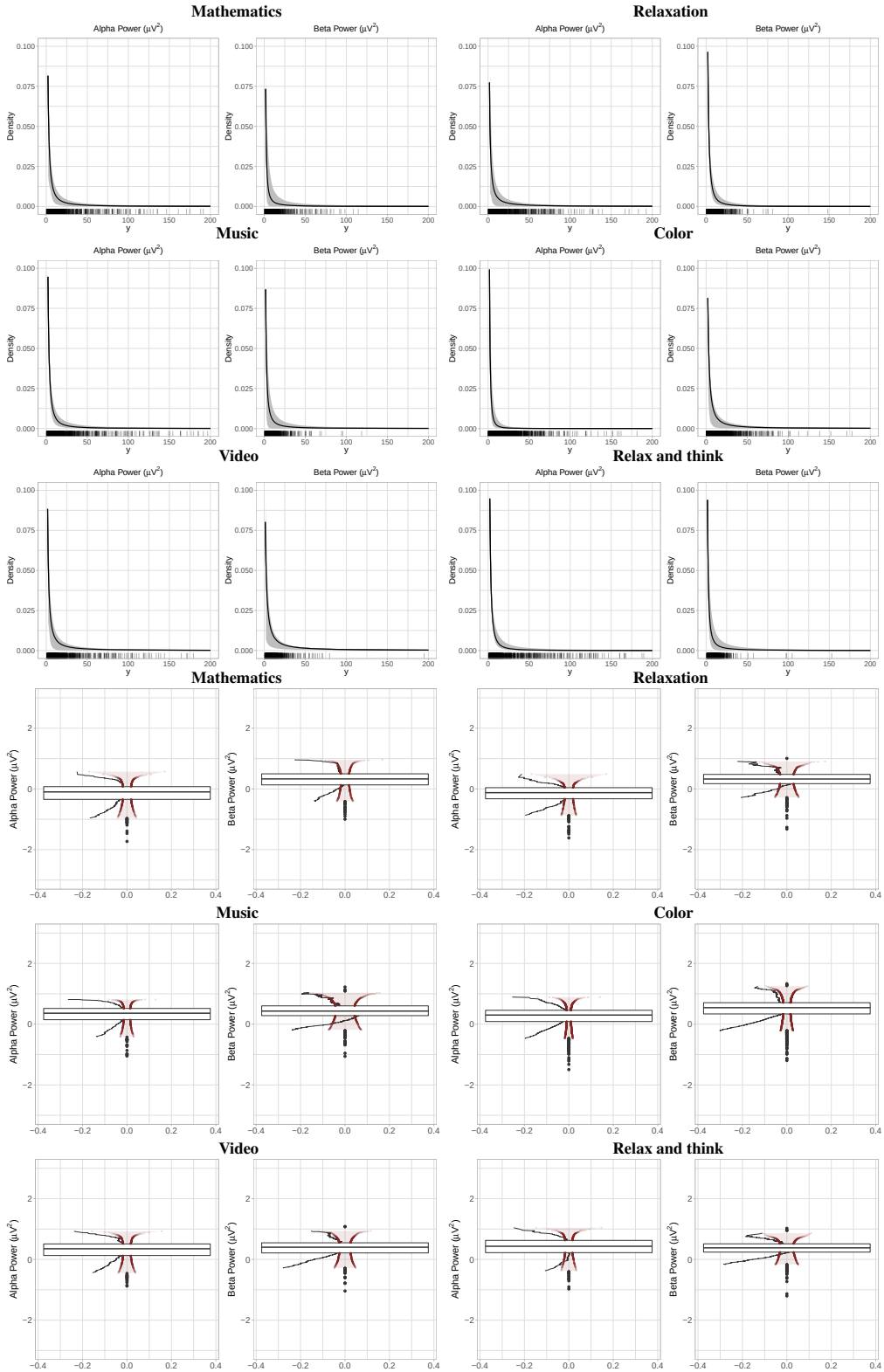

Figure 7: Top: Marginal density estimates of alpha and beta power for each stimulus, obtained with the shape mixture model from Section 2.2 of the paper, along with 95% credible bands. Bottom: Corresponding q-q boxplot of randomized quantile residuals.

Stimulus	**Alpha Power**	**Beta Power**
Mathemathics	0.8234 (0.6951, 0.9613)	0.4892 (0.3961, 0.5942)
Relaxation	0.6076 (0.4899, 0.7297)	0.3306 (0.2422, 0.4216)
Music	0.7419 (0.6292, 0.8614)	0.4056 (0.3222, 0.4976)
Color	0.4813 (0.4040, 0.5607)	0.4276 (0.3631, 0.4970)
Video	0.6953 (0.5773, 0.8109)	0.3284 (0.2432, 0.4172)
Relax and think	0.7715 (0.6628, 0.8987)	0.2215 (0.1445, 0.3025)

Table 2: Posterior mean extreme value index under the conditional stable process scale mixture model from Section 3.2 in the paper along 95% with credible interval.

Stimulus	**Alpha Power**	**Beta Power**
Mathemathics	0.7498 (0.6708, 0.8541)	0.4820 (0.4218, 0.5679)
Relaxation	0.6068 (0.5174, 0.7017)	0.3224 (0.2581, 0.3862)
Music	0.7040 (0.6162, 0.7916)	0.3503 (0.2801, 0.4021)
Color	0.4826 (0.4388, 0.5269)	0.3822 (0.3434, 0.4219)
Video	0.6648 (0.5772, 0.7557)	0.3531 (0.2862, 0.4213)
Relax and think	0.7317 (0.6481, 0.8184)	0.2208 (0.1645 , 0.2782)

Table 3: Posterior mean extreme value index under the conditional stable process scale mixture model from Section 3.2 in the paper along 95% with credible interval.

Stimulus	**Ljung–Box p-value (Alpha Power)**	**Ljung–Box p-value (Beta Power)**
Mathemathics	0.9362	0.3965
Relaxation	0.5891	0.5017
Music	0.0979	0.7899
Color	0.3651	0.3180
Video	0.2189	0.1446
Relax and think	0.0751	0.4035

Table 4: Ljung–Box test for all the stimuli.

Estimates of alpha power and beta power using model NGG scale mixtures over $\mathbb{N}$ are presented in Figure 8. We can see that the estimates using the conditional version of this mocel, are better, as it allows to borrow strength information between different predictor levels.

Table 5 provides a general criterion of comparison. The LPML (logarithm of the pseudo marginal likelihood) is presented for all the different stimuli, under the two models. For more details, see, for instance, Christensen et al. (2011, Section 4.9.2).

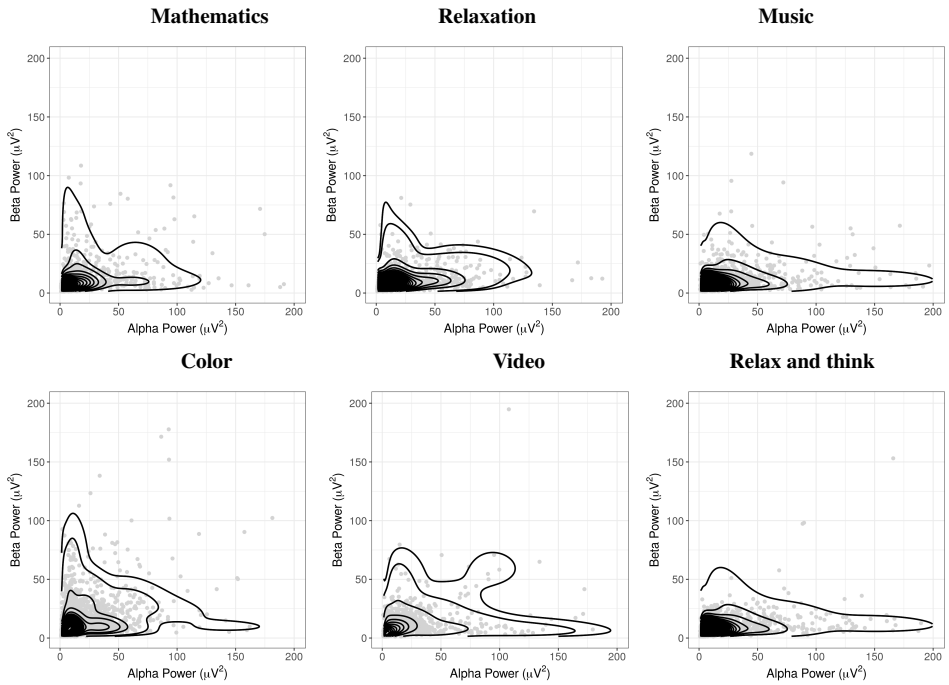

Figure 8: Contours of the posterior conditional joint density estimate of alpha and beta power for each specific stimulus along with raw data; the fit was obtained using the unconditional stable process scale mixture from Section 3.1 in the paper.

Stimuli	LPML (conditional model)	LPML (unconditional model)
Relaxation	-6 702.55	-6 817.55
Mathematics	-6 798.07	-7 234.69
Music	-6 979.77	-7 149.72
Color	-19 972.80	-20 004.76
Video	-6 827.99	-6 950.57
Relax & think	-6 809.74	-6 973.04

Table 5: LPML for all stimuli considering for conditional and unconditional models.

6 The NGGR Package

In this section, we sketch some details on the R package `NGGR`. Instances of the scale mixture models from Sections 2 and 3 in the paper can be fitted using the following functions:

- `SPmix`: Fits the univariate scale stable mixture model with an Erlang kernel and centered on a Pareto Type II distribution; see Example 4 in the paper.
- `SPmix_multi`: Function to compute the multivariate scale stable mixture model of Erlang kernel, centered on a multivariate heavy-tailed distribution with Pareto Type II margins; see Section 3.1 in the paper.
- `SPmix_cond`: Fits the conditional multivariate scale stable conditional mixture model, centered on a multivariate heavy-tailed distribution with Pareto Type II margins; see Section 3.2 in the paper.

As noted in p. 11 of this supplement the Erlang kernel presents some computational advantages; it also follows from p. 11 that the codes made available in the package can be readily adapted to other kernels.

The package also contains commands for fitting the shape mixture model presented in Section 5 of the paper, and a command to simulate trajectories from the instances of the NGG process along with the asymptotic envelopes that stem from Theorem 2; namely:

- `DPmix`: Fits the univariate shape Dirichlet mixture model, centered on a Pareto Type II distribution; see Example 5 on the paper.
- `rSP`: Simulates trajectories from the stable process, NGG$(1, 0, D, G_0)$, along with the corresponding envelopes given in Theorem 2 of the paper. Our computational experience suggests that ε has to be set rather small, and clearly as a function of D (with larger values of D implying a considerably smaller ε).

Below we will illustrate the function `SPmix_multi` so to illustrate how to fit one of the simulation scenarios presented in Section 5 of the paper (Bivariate Scenario 2); although below we focus on the bivariate case, the same command can be used to fit multivariate data. Before showing how to use `SPmix_multi`, we first load the required packages; as it will be

illustrated below, any copula from the `copula` package can be used as a baseline by using the command `mvdc`.

```
## Load required packages
packages <- c("copula", "NGGR")
sapply(packages, require, character.only = TRUE)
```

Before fitting the model, we set up the MCMC and specify the baseline as well as the prior information.

```
## MCMC settings
mcmc <- list(nburn = 200, nsave = 200)
## Prior information
centering <-  mvdc(copula = archmCopula(family = "frank", param = 1),
                   margins = c("pareto", "pareto"),
                   paramMargins = list(list(shape = .5, scale = 1),
                                       list(shape = .5, scale = 1)))
prior <- list(a1 = .1, b1 = .1, aD = .5, bD = .5, centering = centering)
```

Next, we simulate data from the bivariate distribution from Bivariate Scenario 2 using the command `rvdc` from the package `copula`. For reproducibility reasons, we fix the seed using `set.seed`.

```
set.seed(8452)
myMvd <- mvdc(copula = archmCopula(family = "frank", param = 1),
              margins = c("lgamma", "lgamma"),
              paramMargins = list(list(shapelog = 5, ratelog = 5),
                                  list(shapelog = 5, ratelog = 5)))
data <- rMvdc(1000, myMvd)
Y <- cbind(data[,1]-1, data[,2]-1)
```

Next, we learn about the proposed model from data using the function `SPmix_multi` from the package `NGGR`.

```
fit <- SPmix_multi(Y = Y, prior, mcmc)
```

To plot the contour of the posterior distribution one can use `plot` for the object class `SPmix_multi`, that is:

```
plot(fit, which = "contour", xlim = c(0.1, 10), ylim = c(0.1, 10), data = TRUE)
```

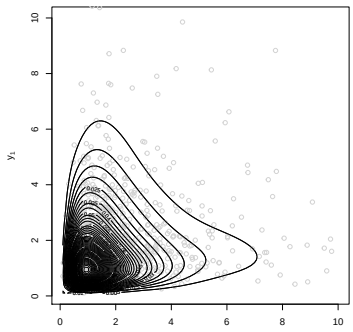

Figure 9: Contours of joint density estimate obtained with the proposed stable process scale mixture model.

To compare the obtained fit against the true joint density, use the following code:

```
gridy1 <- gridy2 <- seq(0.1, 10, length.out = 100)
gridtrue <- expand.grid(gridy1, gridy2)
gg <- as.matrix(gridtrue)
true <- dMvdc(gg + 1, myMvd)
z <- matrix(true, ncol = 100, nrow = 100)
lvls <- pretty(range(z), 50)
par(pty = "s")
contour(gridy1, gridy2, z, levels = lvls, ylab = expression(y[2]),
        xlab = expression(y[1]), main = "")
```

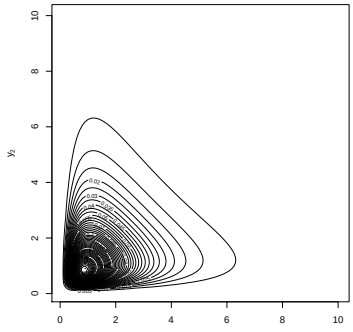

Figure 10: True joint density.

The extreme value index estimated using the proposed model along with 90% credible intervals can be readily obtained as follows:

```
fit$evi.hat

## [1] 0.2039750 0.2891748

fit$evi.CI

##             95%        5%
## [1,] 0.1564305 0.2594535
## [2,] 0.2442174 0.3406467
```

For comparison, the true extreme value index of each LG$(5, 5)$ marginal is $1/5 = 0.20$.

Some final comments on other visualizations available from the `NGGR` package are in order. For the marginal distributions, the function `plot`, from the object class `SPmix_multi`, contains the following options:

- `which = "density"` plots the fitted marginal density for a specific component.
- `which = "logsurvival"` plots the log-survival for a specific the marginal component.
- `which = "qqplot"` shows the q-q plot of the marginal estimates for a specific component.
- `which = "qqboxplot"` shows the q-q boxplot (Rodu and Kafadar, 2022) of the randomized quantile residuals for the marginal estimates for a specific component.

Hence, for example:

```
plot(fit, which = "logsurvival", marginal = 1, bands = TRUE, xlim = c(0, 10))
```

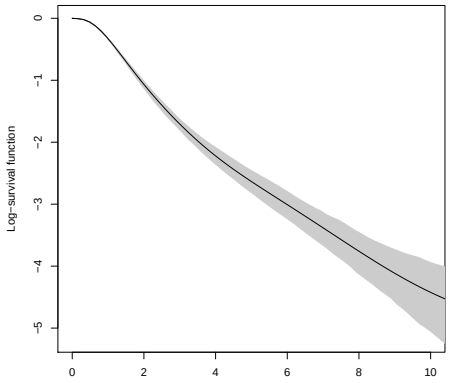

Figure 11: Log-survival estimate for the marginal Y_1 obtained with the proposed stable process scale mixture.

As another example:

```
plot(fit, which = "qqboxplot", marginal = 1, bands = TRUE)
```

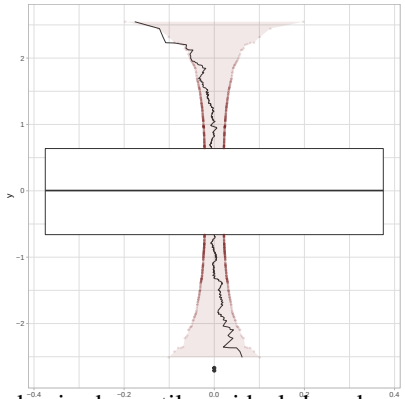

Figure 12: q-q boxplot of randomized quantile residuals based on marginal density estimate of Y_1.

References

Arbel, J., De Blasi, P., and Prünster, I. (2019). "Stochastic approximations to the Pitman–Yor process." *Bayesian Analysis*, 14(4): 1201–1219. 4

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Hoboken, NJ: Wiley. 8

Bertoin, J. (1996). *Lévy processes*, volume 121. Cambridge: CUP.

Bingham, N. H., Goldie, C. M., Teugels, J. L., and Teugels, J. (1989). *Regular Variation*. Cambridge: Cambridge University Press. 2, 8, 9

Christensen, R., Johnson, W., Branscum, A., and Hanson, T. E. (2011). *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*. Boca Raton, FL: Chapman & Hall/CRC. 18

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer. 12

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. New York: Springer. 2

Doss, H. and Sellke, T. (1982). "The tails of probabilities chosen from a Dirichlet prior." *The Annals of Statistics*, 10(4): 1302–1305. 3

Dunn, P. K. and Smyth, G. K. (1996). "Randomized quantile residuals." *Journal of Computational and Graphical Statistics*, 5(3): 236–244. 12

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*. New York: Wiley. 2

Fisher, R. A. and Tippett, L. H. C. (1928). "Limiting forms of the frequency distribution of the largest or smallest member of a sample." In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, 180–190. Cambridge Univ Press. 2

Gnedenko, B. (1943). "Sur la distribution limite du terme maximum d'une serie aleatoire." *Annals of Mathematics*, 423–453. 2

Kalli, M., Griffin, J. E., and Walker, S. G. (2011). "Slice sampling mixture models." *Statistics and Computing*, 21(1): 93–105. 9, 10

Rodu, J. and Kafadar, K. (2022). "The q–q boxplot." *Journal of Computational and Graphical Statistics*, 31(1): 26–39. 12, 23

Walker, S. G. (2007). "Sampling the Dirichlet mixture model with slices." *Communications in Statistics—Simulation and Computation*, 36(1): 45–54. 9, 10