Heavy-Tailed NGG-Mixture Models

Vianey Palacios Ramírez[*], Miguel de Carvalho[†,‡], and Luis Gutiérrez[§]

Abstract. Heavy tails are often found in practice, and yet they are an Achilles heel of a variety of mainstream random probability measures such as the Dirichlet process (DP). The first contribution of this paper focuses on characterizing the tails of the so-called normalized generalized gamma (NGG) process. We show that the right tail of an NGG process is heavy-tailed provided that the centering distribution is itself heavy-tailed; the DP is the only member of the NGG class that fails to obey this convenient property. A second contribution of the paper rests on the development of two classes of heavy-tailed mixture models and the assessment of their relative merits. Multivariate extensions of the proposed heavy-tailed mixtures are devised here, along with a predictor-dependent version, to learn about the effect of covariates on a multivariate heavy-tailed response. The simulation study suggests that the proposed method performs well in various scenarios, and we showcase the application of the proposed methods in a neuroscience dataset.

Keywords: Bayesian nonparametrics, bulk, normalized generalized gamma process, random probability measure, stick-breaking prior, tail index.

1 Introduction

Thousands of heavy-tailed signals are produced on a day-to-day basis across the globe in fields as diverse as engineering, finance, and medicine. And yet, despite the widespread need for modeling these, heavy tails remain a weak spot of several established random probability measures such as the Dirichlet process (DP) (e.g., Ghosal and Van der Vaart, 2017, Section 4.3).

Prior to introducing the main problems to be addressed, and the main contributions of this paper, we first lay the groundwork. The normalized generalized gamma (NGG) process is a random probability measure that was introduced and investigated by Lijoi et al. (2007) and that has received considerable attention in recent years (e.g., James et al., 2009; Lijoi et al., 2010; Barrios et al., 2013; Favaro et al., 2016). The NGG class includes the Dirichlet process, the stable process, and the normalized inverse Gaussian process as particular cases. NGGs are built by normalizing the generalized gamma process, and can be understood as a particular case of a normalized random measure with independent increments (NRMI) (Regazzini et al., 2003; Lijoi et al., 2007; James et al., 2009; Lijoi et al., 2010; Barrios et al., 2013). Some Bayesian nonparametric approaches, such as the NGG process, can be understood as an extension of standard

---

[*]School of Mathematics, Statistics and Physics; Newcastle University, UK, Vianey.Palacios-Ramirez@newcastle.ac.uk

[†]School of Mathematics; The University of Edinburgh, UK, Miguel.deCarvalho@ed.ac.uk

[‡]Department of Mathematics, CIDMA; Universidade de Aveiro, Portugal

[§]Pontificia Universidad Católica de Chile, Santiago, Chile, LlGutier@mat.uc.cl

parametric methods in the sense that they are centered a priori around a parametric model, $\{G_0 \equiv G_{0,\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q\}$, but assign positive mass to a variety of alternatives. Thus, a recurring theme in much of the Bayesian nonparametric literature is to regard a parametric approach—known as the baseline or centering distribution—as a reference, while allowing for deviations from it. See the monographs of Müller et al. (2015) and Ghosal and Van der Vaart (2017) or the review paper of Müller and Mitra (2013) for an introduction to Bayesian nonparametric inference.

It is well-known that the tails of the Dirichlet process are exponentially much thinner than those of the baseline (Ghosal and Van der Vaart, 2017, Section 4.2.3). Motivated by this, this paper opens with the question of whether this is simply a property of the Dirichlet process or is more generally an attribute of the NGG process. Hence, the first contribution of this paper will focus on the characterization of the tails of the NGG process, and we will derive envelopes for the trajectories of the tail of the process. As will be discussed below (Section 2.1), the latter envelopes combined with those of Doss and Sellke (1982) offer a complete portrait of the tails of the NGG process. In addition, we then show that the tail of the NGG process is only moderately thinner than that of the baseline, except for the Dirichlet process. In particular, our results imply that, with the exception of the Dirichlet process, the tail of the NGG process is heavy-tailed, provided that the baseline is itself heavy-tailed. This result sharply contrasts with the Dirichlet process, given that, as mentioned earlier, its tail is exponentially much lighter than that of the centering. Such property of the DP may seem unexpected, given that the process is centered around G_0. Yet, a notable example illustrates this: the tails of the DP centered around a Cauchy distribution are almost exponential, but G_0 has no mean (Ghosal and Van der Vaart, 2017, Example 4.24).

A second contribution of the paper rests on the study of two classes of heavy-tailed mixture models and the assessment of their relative merits. The heavy-tailed mixture models devised here have links with the phase-type scale mixtures of Bladt and Rojas-Nandayapa (2018) and the infinite mixtures of Pareto distributions of Tressou (2008). Our focus differs, however, from these papers in several important ways. Some key differences are that we take a general view of heavy-tailed NGG-mixtures and take advantage of our novel characterization of its tail. In addition, by keeping a general focus in mind, our theoretical and numerical analyses will reveal that there are some good reasons for preferring NGG scale mixtures over NGG-mixtures built from heavy-tailed kernels. Finally, motivated by the fact that heavy-tailed data are frequently multivariate—and since covariates are often available—we further extend the proposed heavy-tailed scale mixture models to model these as well. In other words, multivariate extensions of the proposed heavy-tailed mixtures are also devised below, along with a predictor-dependent version to learn about the effect of covariates on a multivariate heavy-tailed response. Our theoretical and numerical analyses pinpoint a clear preference for NGG scale mixtures over Dirichlet process mixtures of heavy-tailed kernels; overall, NGG scale mixtures tend to have superior numerical performance in the bulk and tails.

A final comment on the jargon of heavy tails is in order. Following the standard convention in the literature on heavy tails (e.g., Resnick, 2007), here we will characterize

these via regular variation (Bingham et al., 1989). A distribution function $F(y) = P(Y \leq y)$, or its density $f = \mathrm{d}F/\mathrm{d}y$ in case it exists, is said to have a regularly varying tail, with tail index $\alpha \equiv \alpha(F) > 0$, if

$$\lim_{y \to \infty} \frac{P(Y > yt)}{P(Y > y)} = t^{-\alpha}. \tag{1.1}$$

The smaller the tail index, the slower the decay of the tail, $1 - F(y)$, to 0 as $y \to \infty$, and thus the more heavy-tailed is the distribution. Throughout, the notation $1 - F \in \mathrm{RV}_{-\alpha}$ is used to denote that F verifies (1.1).

The remainder of this paper unfolds as follows. In Section 2, we study the tails of the NGG process and construct two classes of heavy-tailed NGG mixture models. In Section 3, we expand the proposed toolbox to the multivariate setting as well as to a conditional framework. Section 4 illustrates the performance of the proposed methods and reports the main findings of our numerical studies. An application of the proposed methods to a neuroscience case study is given in Section 5. Finally, in Section 6, we present closing remarks. Proofs are available in the Appendix, and further technical details and supporting numerical evidence can be found in the online supplementary material (Palacios Ramírez et al., 2024); the R package `NGGR`, available from the supplementary material, implements instances of the methods proposed herein.

2 Heavy-tailed NGG-mixture models

2.1 On the tails of the NGG process

Subordinator representations: We start by studying the tails of the NGG process since its properties will be vital for constructing our class of mixture models. The main result of this section is Theorem 2, but before we can discuss its implications, we first lay the groundwork. A subordinator, $\{S(t) : t \geq 0\}$, is an increasing stochastic process over the positive real line that has independent and homogeneous increments (e.g., Applebaum, 2009, Chapter 1). By the so-called Lévy–Khintchine representation (e.g., Bertoin, 1999, Section 1.2), a subordinator is fully characterized by its Laplace exponent,

$$\Phi(\lambda) = \mathtt{k} + \mathtt{d}\lambda + \int_0^\infty (1 - e^{-\lambda u})\nu(\mathrm{d}u), \quad \lambda \geq 0,$$

that obeys $\mathrm{E}[\exp\{-\lambda S(t)\}] = \exp\{-t\Phi(\lambda)\}$, for $t \geq 0$; here, $\mathtt{k} > 0$ is the killing rate, $\mathtt{d} > 0$ is the drift coefficient, and ν is a measure on $(0, \infty)$, known as Lévy measure, that governs the law of the increments and which obeys the constraint $\int_0^\infty \min(1, u)\,\nu(\mathrm{d}u) < \infty$.

It is well known that the NGG process, introduced in Lijoi et al. (2007), inherently admits a subordinator representation due to its construction through the normalization of the generalized gamma process. A random probability measure G follows an NGG process, here denoted as $G \sim \mathrm{NGG}(M, \tau, D, G_0)$, if

$$G(y) = \frac{S\{MG_0(y)\}}{S(M)}, \quad y \in \mathbb{R}. \tag{2.1}$$

Here, S is a generalized gamma subordinator, that is, $\mathtt{k} = \mathtt{d} = 0$ and

$$\nu(\mathrm{d}u) = D/\Gamma(1 - D)u^{-1-D}\exp(-\tau u)\,\mathrm{d}u,$$

for $u > 0$, where $\Gamma(z) = \int_0^\infty u^{z-1}\exp(-u)\,\mathrm{d}u$ is the gamma function. With a slight abuse of notation throughout, we use G to denote both the random measure of an NGG process and its corresponding distribution function.

The parameter G_0 of the NGG is known as the centering distribution function, and the other parameters are subject to the constraints $M > 0$, $\tau \geq 0$, and $0 \leq D < 1$. Particular cases include the Dirichlet process, NGG $(M, 1, 0, G_0)$, the stable process, NGG $(1, 0, D, G_0)$, and the normalized inverse Gaussian NGG $(1, \tau, 1/2, G_0)$ (Lijoi et al., 2005, 2007); the Dirichlet process can also be obtained for any $\tau > 0$ and not just for $\tau = 1$, provided $D = 0$. Keeping this in mind, and the fact that in our context it will be important to separate the Dirichlet process from the other members of the NGG class, we introduce the following notation:

$$\begin{cases} \mathcal{D} = \{(M, \tau, D) : M > 0, \tau > 0, D = 0\}, \\ \mathcal{N} = \{(M, \tau, D) : M > 0, \tau \geq 0, D \in (0, 1)\}. \end{cases}$$

The general subordinator representation of the NGG process is given in (2.1). In the particular case $G \sim \mathrm{NGG}(M, \tau, D, G_0)$ with $(M, \tau, D) \in \mathcal{D}$, then

$$G(y) = \frac{\gamma\{MG_0(y)\}}{\gamma(M)}, \quad y \in \mathbb{R}, \tag{2.2}$$

where γ is a gamma process, that is, $\mathtt{k} = \mathtt{d} = 0$ and $\nu(\mathrm{d}u) = u^{-1}\exp(-u)\,\mathrm{d}u$, for $u > 0$.

Finally, we recall that if G is a NGG process, then it admits a stick-breaking representation

$$G = \sum_{h=1}^\infty \pi_h \delta_{Y_h}, \quad Y_h \overset{\mathrm{iid}}{\sim} G_0, \tag{2.3}$$

with a stick-breaking sequence V_h as in Favaro et al. (2016, Proposition 3), so that $\pi_1 = V_1$ and $\pi_h = V_h \prod_{k<h}(1 - V_k)$, for $h = 2, 3, \ldots$, and where δ_Y denotes a point mass at Y.

On the tails of the NGG process: We now examine the behavior of $1 - G(y)$, as y approaches the right endpoint, $y^* = \sup\{y : G(y) < 1\}$. Recall that both $G \sim \mathrm{NGG}(M, \tau, D, G_0)$ and G_0 are supported over the same set, and thus the right endpoints of G and G_0 coincide. Following the standard convention in extreme value analysis, we focus on the right tail, but all claims below apply to the left tail with minor adjustments.

The tails of the Dirichlet process are much lighter than those of the centering distribution, with probability one, a fact that can be shown using the subordinator representation in 2.2. Theorem 1 is well known since Doss and Sellke (1982), it formalizes the latter claim, and it is only included here for completeness.

Theorem 1 (Tails of NGG in $\mathcal{D}$)**.** *Let $G(y)$ be the distribution of a $NGG(M, \tau, D, G_0)$ process with $(M, \tau, D) \in \mathcal{N}$ and with non-atomic G_0. Then,*

$$\lim \inf_{y \to y^*} \frac{1 - G(y)}{g_r\{1 - G_0(y)\}} = \begin{cases} 0, & \text{if } s < 1, \\ \infty, & \text{if } s > 1, \end{cases} \quad a.s.,$$

$$\lim \sup_{y \to y^*} \frac{1 - G(y)}{h_r\{1 - G_0(y)\}} = \begin{cases} 0, & \text{if } r > 1, \\ \infty, & \text{if } r \leq 1, \end{cases} \quad a.s.,$$

with $g_r(t) = \exp\{-r \log |\log t|/t\}$ and $h_r(t) = \exp\{-1/(t|\log t|^r)\}$, for $0 < t < 1$.

Proof. See Doss and Sellke (1982) or Ghosal and Van der Vaart (2017, Theorem 4.22). □

The takeaway from Theorem 1 can be loosely summarized in bounds, for $r > 1$, a.s. eventually for a large y,

$$\exp\left[-\frac{r \log |\log M\{1 - G_0(y)\}|}{M\{1 - G_0(y)\}}\right] \leq 1 - G(y) \leq \exp\left[-\frac{1}{M\{1 - G_0(y)\}|\log M\{1 - G_0(y)\}|^r}\right], \tag{2.4}$$

where $G \sim \text{NGG}(M, \tau, D, G_0)$ for $(M, \tau, D) \in \mathcal{D}$. Hence, the tails of the Dirichlet process are almost exponential, even if G_0 is heavy-tailed. Despite being one of the most popular Bayesian nonparametric priors, such deficiency of the DP rules out its use when the goal is to model heavy tails, extreme values, and risk. As shown next, what happens with the NGG process over $\mathcal{N}$ is substantially different as its tails are close to those of the baseline.

Theorem 2 (Tails of NGG in $\mathcal{N}$)**.** *Let $G(y)$ be the distribution of an $NGG(M, \tau, D, G_0)$ process with $(M, \tau, D) \in \mathcal{N}$. Then,*

$$\lim \inf_{y \to y^*} \frac{1 - G(y)}{l\{M(1 - G_0(y))\}} = D(1 - D)^{(1-D)/D}/S(M) \quad a.s.,$$

$$\lim \sup_{y \to y^*} \frac{1 - G(y)}{u_r\{M(1 - G_0(y))\}} = \begin{cases} 0, & r > 1, \\ \infty, & r \leq 1, \end{cases} \quad a.s.,$$

with $l(t) = t^{1/D} \log |\log t|/\{(\log |\log t| + \tau^D t)^{1/D} - \tau t^{1/D}\}$, for $0 < t < e^{-1}$, and $u_r(t) = t^{1/D}|\log t|^{r/D}$, for $0 < t < e^{-r}$.

Theorem 2 warrants some remarks. A key conclusion is that the tails of a random distribution following a NGG over $\mathcal{N}$ are almost as heavy as those of the centering, G_0. Indeed, the rough takeaway of Theorem 2 is that a.s. eventually for a large y,

$$\frac{1}{S(M)} D(1 - D)^{(1-D)/D} l(M\{1 - G_0(y)\}) \leq 1 - G(y) \leq u_r(M\{1 - G_0(y)\}), \tag{2.5}$$

for $r > 1$ with $y > G_0^{-1}(1 - e^{-r})$. Since $S(M)$ is random, the naïve lower bound in (2.5) would vary with each realization of the subordinator. To mitigate this, we use well-known results on long-run behavior of subordinators (e.g., Bertoin, 1996, p. 92) which yield a.s. eventually for a large y,

$$\frac{D(1-D)^{(1-D)/D}}{(M^*)^{1/D} \log(M^*)^{r/D}} \, l(M\{1 - G_0(y)\}) \leq 1 - G(y) \leq u_r(M\{1 - G_0(y)\}), \qquad (2.6)$$

for $r > 1$, $M^* \gg e^{-r}$ and $M^* > M$, with $y > G_0^{-1}(1 - e^{-r})$. Technical details on the derivation of (2.6) are in the supplementary material (Section 2); this refined lower bound follows by determining an asymptotic upper bound for $S(M)$, for a large $M^* > M$ (e.g., in our experience setting $M^* = 10$ already gives a reasonable bound, for any $M < 10$). Numerical illustrations of the asymptotic envelopes in (2.6) are presented in Figure 1 for a specific instance of the NGG process in $\mathbb{N}$, the stable process, and further examples are included in the supplementary material. Another implication of Theorem 2 is that if the tail of the centering distribution of NGG in $\mathbb{N}$ is heavy-tailed, then so will be that of the corresponding process, though with a lighter tail.

Corollary 3 (Stability of the heavy-tail property in $\mathbb{N}$)**.** *If $G \sim NGG(M, \tau, D, G_0)$, with $(M, \tau, D) \in \mathbb{N}$, and G_0 has a regularly varying tail, with tail index $\alpha_0 \equiv \alpha(G_0) > 0$, then G has a regularly varying tail, with tail index $\alpha(G) = \alpha(G_0)/D$, almost surely.*

Corollary 3 does not conflict with the full support of NGGs (e.g., Bissiri and Ongaro, 2014). Indeed, while the NGG centered on a regularly varying tail is heavy-tailed with probability one, the support of the NGG remains the space of all probability measures— provided the baseline measure satisfies the requirements stated in Bissiri and Ongaro (2014).

Next, we compare Theorems 1 and 2 in one example.

Example 1 (Pareto centering distribution: $\mathcal{D}$ *versus* $\mathbb{N}$)**.** Suppose first that $G \sim$ NGG$(M, 1, 0, G_0)$ where G_0 is a standard Pareto distribution, that is, $1 - G_0(y) = 1/y$, for $y > 1$. Then (2.4) yields

$$\exp[-ry/M \log |\log(M/y)|] \leq 1 - G(y) \leq \exp[-y/\{M|\log(M/y)|^r\}], \qquad (2.7)$$

for $r > 1$, and hence the tails of the Dirichlet process $G \sim$ NGG$(M, 1, 0, G_0)$ are almost exponential, despite the fact that G_0 is heavy-tailed. Indeed, as shown in the supplementary material (Section 2), the two bounds in (2.7) are in the Gumbel maximum domain of attraction, although G_0 is in the Fréchet maximum domain of attraction. Suppose now that $G \sim$ NGG(M, τ, D, G_0), with $(M, \tau, D) \in \mathbb{N}$. Then,

$$\frac{C \, y^{-1/D} M^{1/D} \log |\log My^{-1}|}{(\log |\log My^{-1}| + \tau^D My^{-1})^{1/D} - \tau(My^{-1})^{1/D}} \leq 1 - G(y) \leq y^{-1/D} M^{1/D} |\log My^{-1}|^{r/D},$$

for $C = 1/\{(M^*)^{1/D} \log(M^*)^{r/D}\}$, $r > 1$, $M^* \gg 0$ and $M^* > M$. It follows from Corollary 3 that $1 - G(y)$ is regularly varying at infinity with tail index $1/D$. Figure 1 illustrates that, for example, as predicted by Theorem 2, the trajectories of the stable process, NGG$(1, 0, 0.5, G_0)$, follow the asymptotic envelopes in (1); in addition, the same figure illustrates that the envelopes of the Dirichlet process in (2.4) fall abruptly in comparison with those of the stable process.
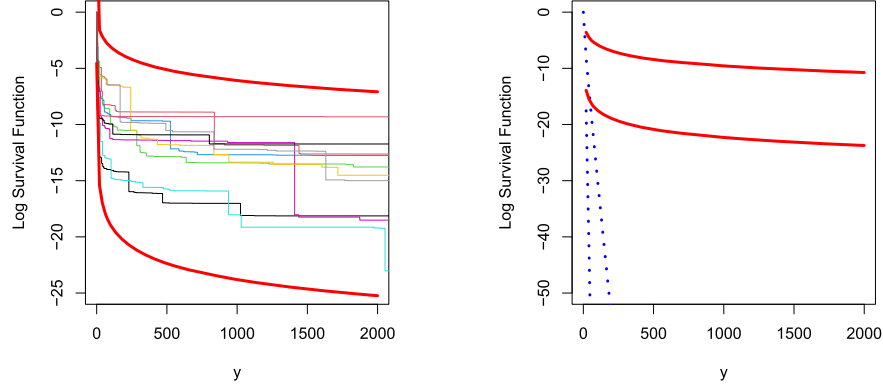
Figure 1: Asymptotic envelopes for Example 1. Left: Asymptotic envelopes that follow from Theorem 2 along with random trajectories of the log survival functions from a stable process $\mathrm{NGG}(1, 0, 0.5, G_0)$. Right: The same envelopes for log survival function of a stable process $\mathrm{NGG}(1, 0, 0.5, G_0)$ (solid) against those of a Dirichlet process $\mathrm{NGG}(1, 1, 0, G_0)$ (dashed); G_0 is the standard unit Pareto distribution.

2.2 Heavy-tailed NGG-mixture models

Empowered by the main findings from Section 2.1, this section shows that two classes of heavy-tailed NGG mixture models can be devised, and it discusses the relative merits of each option. Below, we focus on univariate mixtures; comments on multivariate and conditional extensions are given in Section 3.

Heavy-tailed NGG-$\mathcal{N}$ scale mixtures: The first class of mixture models to be considered is

$$\begin{cases} f(y) = \int_0^\infty K_\sigma(y; \eta_\sigma) \, \mathrm{d}G(\sigma), \\ G \sim \mathrm{NGG}(M, \tau, D, G_0), \quad (M, \tau, D) \in \mathcal{N} \\ 1 - G_0(\sigma) = \frac{\mathscr{L}(\sigma)}{\sigma^{\alpha_0}}, \end{cases} \tag{2.8}$$

with $y \in \mathbb{R}$ and $\alpha_0 \equiv \alpha(G_0) > 0$. Here, $K_\sigma(\cdot) = K(\cdot/\sigma; \eta_\sigma)/\sigma$ with K being a kernel, $\sigma > 0$ is a scale parameter, η_σ denotes additional parameters (possibly related with σ), and $\mathscr{L}$ is a slowly varying function, that is, $\mathscr{L}(yt)/\mathscr{L}(y) \to 1$ as $y \to \infty$ for any $t > 0$.

Heavy-tailed NGG shape mixtures: The second class of mixture models to be considered is

$$\begin{cases} f(y) = \int_0^\infty \Bbbk(y; \alpha, \eta_\alpha) \, \mathrm{d}G(\alpha), \\ G \sim \mathrm{NGG}(M, \tau, D, G_0), \quad (M, \tau, D) \in \mathcal{N} \cup \mathcal{D}, \\ 1 - \mathbb{K}(y; \alpha, \eta_\alpha) = \frac{\mathscr{L}(y)}{y^\alpha}, \end{cases} \tag{2.9}$$

with $y \in \mathbb{R}$ and $\alpha \geq 0$. Here, $\Bbbk$ is a Pareto-type kernel with distribution function $\mathbb{K}$, η_α denotes additional parameters (possibly related with α), and $\mathscr{L}$ is a slowly varying function.

The following theorem implies that NGG mixture models in (2.8) and (2.9) are indeed heavy-tailed. In addition, it shows how their tail indices relate to that of the centering. Below, $(a)_+ = \max(a, 0)$ denotes the positive part function.

Theorem 4 (Heavy-tailed NGG-mixtures)**.** *The following results hold for F the distribution function of f in* (2.8) *and* (2.9):

a) *If* (2.8) *holds, with $U \sim K_\sigma(\cdot; \eta_\sigma)$, $\mathrm{E}(U_+^{\alpha_0}) < \infty$, $P(U_+ > \sigma) = o\{1 - G_0(\sigma)\}^{1/D}$ and $\liminf_{\sigma \to \infty} \mathscr{L}(\sigma) > 0$, then the tail of f is regularly varying with tail index $\alpha(F) = \alpha_0/D$, almost surely.*

b) *If* (2.9) *holds, then the tail of f is regularly varying with tail index $\alpha(F) = \inf\{\alpha : G_0(\alpha) > 0\}$, for any $G \sim NGG(D, M, G_0(\alpha))$, almost surely.*

Some comments on Theorem 4 are in order:

- Theorem 4 a) shows that if the centering of the NGG process is heavy-tailed, and if the tail of the kernel is not heavier than that of the mixing, then f in (2.8) is heavy-tailed—with the same tail index as that of the mixing; that is, $\alpha(F) = \alpha(G) = \alpha_0/D$ with probability one, where $\alpha_0 = \alpha(G_0)$. Interestingly, this result offers a partial answer to an insightful open problem raised by Li et al. (2019, Theorem 3.5) on the range of $\alpha(F)$ under a polynomial decay of G_0. Indeed, Theorem 4 a) illustrates that under a polynomial decay of G_0, it holds that $\alpha(F) = \alpha_0/D$ differs from 0 and ∞, almost surely, and thus $\alpha(F)$ for (2.8) does not concentrate on a singleton when assigned hyperpriors on D an α_0. The range of $\alpha(F)$ can be the positive real line, provided that prior densities of D and α_0 put positive mass on all D and α_0, that is, $\mathsf{p}(D) > 0$ or $\mathsf{p}(\alpha_0) > 0$.

- Theorem 4 b) is a folklore result that formalizes the idea that infinite shape mixtures of heavy-tailed kernels are themselves heavy-tailed, with a tail index equal to that of the heaviest component (i.e., equal to the left endpoint of the centering in the case of (2.9)). While this would seem like a standard result, we could not find a formal statement of the result in the literature. As evident from the proof, the argument generally holds for any stick-breaking process G and not just for NGG processes.

For simplification, throughout, we refer to (2.8) as scale mixtures, but as it will be shown below, (2.8) also includes scale–shape mixtures of light-tailed kernels. Next, we present instances of the specifications in (2.8) and (2.9) that showcase the generality of the latter and how they relate to some mainstream approaches.

Example 2 (NGG scale mixtures with an Erlang kernel)**.** As an example of (2.8), consider

$$f(y) = \sum_{h=1}^{\infty} \pi_h \, \mathrm{Er}(y; a, \sigma_h), \tag{2.10}$$

where $\mathrm{Er}(y; a, \sigma_h)$ is the density of the Erlang distribution with shape $a \in \mathbb{N}$ and scale $\sigma_h > 0$, and $\pi_h = V_h \prod_{k<h}(1 - V_k)$ with $V_h \overset{\mathrm{ind}}{\sim} \mathrm{Beta}(1 - D, hD)$, for $h \in \mathbb{N}$. In the

Distribution	Kernel ($\Bbbk$)	Slowly varying function ($\mathscr{L}$)	Tail index ($\alpha_{\Bbbk}$)
Burr	$\propto y^{c-1}/(1+y^c)^{a+1}$	$\propto (y^{-c}+1)^{-(a+1)}$	ca
F	$\propto y^{a/2-1}(a+by)^{-(a+b)/2}$	$\propto (a/y+b)^{-(a+b)/2}$	$b/2$
Generalized Pareto	$\propto (1-ay/\sigma)^{1/a-1}$	$\propto (1/y-k/\sigma)^{1/a-1}$	$-1/a$
Pareto	$\propto y^{-(a+1)}$	$\propto 1$	a
Student-t	$\propto (1+y^2/a)^{-(a+1)/2}$	$\propto (1/y^2+1/a)^{-(a+1)/2}$	a

Table 1: Instances of Pareto-type kernels $\Bbbk$ following (2.9).

notation of (2.8), with $K(y; \eta_\sigma) = \mathrm{Er}(y; a, 1)$ and $\eta_\sigma = a$. For this kernel, the infinite mixture in (2.8) shows a connection with the phase-type scale mixtures of Bladt and Rojas-Nandayapa (2018) and Ayala et al. (2022). Two key differences are that: *i*) since the mixing in (2.8) is over an NGG process, inference can be conducted by standard Bayesian nonparametric samplers, whereas fitting phase–type scale mixtures is far from straightforward; *ii*) Theorem 4 a) allows for a general kernel with a lighter tail than that of the mixing, whereas Bladt and Rojas-Nandayapa (2018) consider phase-type kernels. Another version of (2.8), along the same lines as (2.10), which we have found to perform well in practice, is the following scale–shape mixture of NGG,

$$f(y) = \sum_{h=1}^{\infty} \pi_h \, \mathrm{Er}(y; \lceil \sigma_h \rceil, \sigma_h/\lambda), \tag{2.11}$$

where $\lceil \cdot \rceil$ is the ceiling function, and $K(y; \eta_\sigma) = \mathrm{Er}(y; \lceil \sigma \rceil, 1/\lambda)$ with $\eta_\sigma = (\lceil \sigma \rceil, 1/\lambda)^{\mathrm{T}}$ in the notation of (2.8). We will revisit (2.11) later in the paper.

Example 3 (NGG shape mixtures with a Pareto-type kernel)**.** Since the Burr, F, and generalized Pareto distributions are particular cases of Pareto-type kernels (Table 1), the mixture model in (2.9) includes as particular cases infinite mixtures of such distributions with a NGG mixing. In addition, (2.9) also includes the Pareto kernel Dirichlet process mixtures of Tressou (2008).

There are some reasons for preferring the NGG scale mixtures in (2.8) over (2.9). In particular, scale mixtures of stable processes offer a more natural link between the tail of the centering and the tail of F than (2.9). For example, if in (2.8) the centering is a Pareto Type II distribution over $(0, \infty)$ (i.e., $1 - G_0(\sigma) = (1+\sigma)^{-\alpha_0}$) then $\alpha(F) = \alpha_0/D$, and hence both the centering and F are heavy-tailed. Yet, if the centering in (2.9) is a Pareto Type II distribution over $(0, \infty)$ (i.e., $1 - G_0(\alpha) = (1+\alpha)^{-\beta}$), then $\alpha(F) = 0$ and hence while the centering is heavy-tailed, F is super-heavy tailed. Motivated by this, in the next sections, we will emphasize scale mixtures of NGG processes in $\mathcal{N}$, that is, (2.8); the only exception is Section 4, where we will consider once more (2.9) for comparison purposes.

3 Consequences and extensions

3.1 Multivariate variants

We now discuss how Section 2.2 can be extended to define priors on the space of multivariate heavy-tailed distributions, that is, the class of joint distributions with heavy-tailed marginals. This yields the following multivariate versions of (2.8) and (2.9) for $\mathbf{y} \in \mathbb{R}^d$.

First, extending (2.8) consider the following multivariate heavy-tailed NGG scale mixture model:

$$\begin{cases} f(\mathbf{y}) = \int_{\mathbb{R}_+^d} \mathbf{K}(\mathbf{y}; \boldsymbol{\eta_\sigma}) \, \mathrm{d}G(\boldsymbol{\sigma}), \\ G \sim \mathrm{NGG}(M, \tau, D, G_0), \\ (M, \tau, D) \in \mathcal{N}, \\ 1 - G_{0,k}(\sigma) = \frac{\mathscr{L}_k(\sigma)}{\sigma^{\alpha_{0,k}}}, \end{cases} \tag{3.1}$$

where $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_d) \in \mathbb{R}_+^d$, $\boldsymbol{\eta_\sigma} = (\eta_{\sigma_1}, \ldots, \eta_{\sigma_d})$ and $G_{0,k}(\sigma)$ is the kth marginal distribution of $G_0(\boldsymbol{\sigma})$, for $k = 1, \ldots, d$. In terms of the kernel for the mixture model in (3.1), $\mathbf{K}(\mathbf{y}; \boldsymbol{\eta_\sigma})$, we assume that its margins are given by a scale kernel, $K_{\sigma_k}(y_k; \eta_{\sigma_k})$, that is

$$\int_{\mathbb{R}_+^{d-1}} \mathbf{K}(\mathbf{y}; \boldsymbol{\eta_\sigma}) \, \mathrm{d}\mathbf{y}_{-k} = K_{\sigma_k}(y_k; \eta_{\sigma_k}), \tag{3.2}$$

with $\mathrm{d}\mathbf{y}_{-k} = \mathrm{d}y_1 \ldots \mathrm{d}y_{k-1} \mathrm{d}y_{k+1} \ldots \mathrm{d}y_d$. An example of a kernel obeying (3.2) is the conditional independence kernel, that assumes conditional independence among components along with a common parameter shared by all components, and which is given by

$$\mathbf{K}(\mathbf{y}; \boldsymbol{\eta_\sigma}) = \prod_{k=1}^{d} K_{\sigma_k}(y_k; \eta_{\sigma_k});$$

the latter kernel follows from the principles in Sarabia Alegría et al. (2008, Section 3.1).

Second, extending (2.9), consider the following multivariate heavy-tailed NGG shape mixtures:

$$\begin{cases} f(\mathbf{y}) = \int_{\mathbb{R}_+^d} \boldsymbol{\mathcal{K}}(\mathbf{y}; \boldsymbol{\alpha}; \boldsymbol{\eta_\alpha}) \, \mathrm{d}G(\boldsymbol{\alpha}), \\ G \sim \mathrm{NGG}(M, \tau, D, G_0), \\ (M, \tau, D) \in \mathcal{D}, \end{cases} \tag{3.3}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}_+^d$, and $G_{0,k}(\alpha)$ will denote the kth marginal distribution of $G_0(\boldsymbol{\alpha})$, for $k = 1, \ldots, d$. In terms of the kernel, for the mixture model in (3.3), we assume that its margins are heavy-tailed, that is

$$\int_{\mathbb{R}_+^{d-1}} \boldsymbol{\mathcal{K}}(\mathbf{y}; \boldsymbol{\alpha}, \boldsymbol{\eta_\alpha}) \, \mathrm{d}\mathbf{y}_{-k} = \Bbbk(y_k; \alpha_k, \eta_{\sigma_k}), \quad 1 - \Bbbk(y; \alpha_k, \eta_{\alpha_k}) = \frac{\mathscr{L}_k(y)}{y^{\alpha_k}}, \tag{3.4}$$

where $\mathscr{L}_k$ is a sequence of slowly varying functions. An example of a kernel obeying (3.4) is the following product of standard Pareto distributions $\prod_{k=1}^{d} \alpha_k / y^{\alpha_k + 1}$, also constructed according to Sarabia Alegría et al. (2008, Section 3.1).

Theorem 5 (Multivariate heavy-tailed NGG-mixtures)**.** *The following results hold for F_k the distribution function of the kth marginal distribution of f in* (3.1) *and* (3.3):

a) *If* (3.1) *and* (3.2) *hold, with $U_k \sim K_{\sigma_k}(\cdot; \eta_{\sigma_k})$, $\mathrm{E}(U_{k+}^{\alpha_{0,k}}) < \infty$, $P(U_{k+} > \sigma) = o\{1 - G_0(\sigma)\}^{1/D}$ and $\liminf_{\sigma \to \infty} \mathscr{L}(\sigma) > 0$, then the kth marginal of f has a regularly varying tail with tail index $\alpha(F_k) = \alpha_{0,k}/D$, almost surely for $k = 1, \dots, d$.*

b) *If* (3.3) *and* (3.4) *hold, then the kth marginal of f has a regularly varying tail with tail index $\alpha(F_k) = \inf\{\alpha : G_{0,k}(\alpha) > 0\}$, for any $G \sim NGG(D, M, G_0(\boldsymbol{\alpha}))$, almost surely.*

The previous result naturally extends Theorem 4 to the multivariate setting.

Remark 1. Some comments on the multivariate heavy-tailed stable process scale mixtures in (3.1) are in order:

- A concrete instance of (3.1) and (3.2) that we have found to work well in practice is the following extension of Example 2,

$$f(\mathbf{y}) = \sum_{h=1}^{\infty} \pi_h \prod_{k=1}^{d} \mathrm{Er}(y_k; \lceil \sigma_{k,h} \rceil, \sigma_{k,h}/\lambda), \tag{3.5}$$

  where once more $\mathrm{Er}(y; a, b)$ is the density of the univariate Erlang distribution and $\pi_h = V_h \prod_{k<h}(1 - V_k)$ with $V_h \overset{\mathrm{ind}}{\sim} \mathrm{Beta}(1 - D, hD)$, for $h \in \mathbb{N}$.

- The specification in (3.1) requires that the centering is itself a multivariate heavy-tailed distribution. These can be easily constructed from copulas (Nelsen, 2006), and we recommend opting for Pareto Type II margins; that is, for $k = 1, \dots, d$,

$$1 - G_{0,k}(\sigma) = \left(1 + \frac{\sigma}{\beta}\right)^{-\alpha_{0,k}}, \tag{3.6}$$

  with $\sigma \geq 0$, $\alpha_{0,k} > 0$, and $\beta > 0$. Such margins for the centering are convenient since they lead to a closed-form posterior of the extreme value index, as can be seen from the supplementary material (Section 3).

- To complete (3.1) we recommend a Jeffrey's prior on the tail index $\mathsf{p}(\alpha_{0,k}) \propto 1/\alpha_{0,k}$ and a Beta prior on the discount parameter, $D \sim \mathrm{Beta}(a_D, b_D)$. Finally, one may set a prior on the remainder parameters of the kernel, and for instance, in the Erlang kernel example in (3.5), we will opt for $\lambda \sim \mathrm{Gamma}(a_\lambda, b_\lambda)$. This implies a prior for the tail index α_0/D, with infinite expectation. Alternatively, with a proper prior on α_0, e.g., $\alpha_0 \sim \Gamma(a_{\alpha_0}, b_{\alpha_0})$, the induced prior on the tail index would have prior expectation $E(\alpha_0)E(1/D) = \{a_{\alpha_0}(a_D + b_D - 1)\}/\{b_{\alpha_0}(a_D - 1)\}$.

- The specifications in (3.1) and (3.2)—as well as (3.3) and (3.4)—can be used to model heavy-tailed partially exchangeable signals by considering a single parameter in the mixing. Given the prominence of partial exchangeability in Bayesian Nonparametric literature, we believe this to be an interesting feature of this specification (Camerlenghi et al., 2019a,b; Catalano et al., 2021).

3.2   Modeling conditional joint densities

We now show how to extend the proposed models to include the effect of covariates by using a particular NGG process—the stable process. Specifically, a single-atoms dependent stable process is constructed following the principles from Barrientos et al. (2012, Definition 3) and Quintana et al. (2022, Section 2.3). For conciseness, we focus on multivariate heavy-tailed NGG scale mixtures in (3.1), but the principles discussed below can be easily adapted for the multivariate shape mixtures from Section 3.1 as well as to the univariate methods from Section 2. Consider the following predictor-dependent model,

$$f(\mathbf{y} \mid \mathbf{x}) = \int_{\mathbb{R}^d_+} \prod_{k=1}^{d} K_{\sigma_k}(y_k; \eta_{\sigma_k}) \, \mathrm{d}G_{\mathbf{x}}(\boldsymbol{\sigma}), \tag{3.7}$$

where $\mathbf{y}, \boldsymbol{\sigma} \in \mathbb{R}^d$ and $\{G_{\mathbf{x}}\}$ is a family of random probability measures indexed by a covariate $\mathbf{x} \in \mathbb{R}^p$. Specifically, we consider the following dependent stable process

$$G_{\mathbf{x}} = \sum_{h=1}^{\infty} \pi_h(\mathbf{x})\delta_{\boldsymbol{\sigma}_h}, \qquad \boldsymbol{\sigma}_h \overset{\mathrm{iid}}{\sim} G_0(\boldsymbol{\sigma}). \tag{3.8}$$

Here, the weights of the stick-breaking representation and the discount parameter D of the NGG process are indexed over the covariate as follows, $\pi_h(\mathbf{x}) = V_h(\mathbf{x}) \prod_{k<h}\{1 - V_k(\mathbf{x})\}$, and

$$V_h(\mathbf{x}) \sim \mathrm{Beta}(1 - D_h(\mathbf{x}), hD_h(\mathbf{x})), \qquad D_h(\mathbf{x}) = \frac{e^{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_h}}{1 + e^{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_h}}, \tag{3.9}$$

where $\boldsymbol{\beta}_h$ is a parameter in $\mathbb{R}^p$. Clearly, since (3.8) is a NGG for every $\mathbf{x}$, Theorem 5 a) implies that the joint density mixture model in (3.7) yields a multivariate heavy-tailed distribution, for every $\mathbf{x}$. The model is completed with a prior distribution for $\boldsymbol{\beta}_h$, given by $\boldsymbol{\beta}_h \overset{\mathrm{iid}}{\sim} N_p(\mathbf{0}, s^2\boldsymbol{I})$, for $h \in \mathbb{N}$. A specific embodiment of the approach discussed in this section, that will be revisited later in the paper, is the following extension of Example 2,

$$f(\mathbf{y} \mid \mathbf{x}) = \sum_{h=1}^{\infty} \pi_h(\mathbf{x}) \prod_{k=1}^{d} \mathrm{Er}(y_k; \lceil \sigma_{k,h} \rceil, \sigma_{k,h}/\lambda). \tag{3.10}$$

4   Simulation study

4.1   Simulation scenarios and preparations

This section describes the true data-generating processes and the settings used over the Monte Carlo simulation study from Section 4.2.

Data-generating processes: We consider one scenario for the univariate version of the model from Section 2.2, three scenarios for the multivariate version from Section 3.1, and three scenarios for the multivariate conditional version from Section 3.2. The univariate

scenario is a standard unit Pareto distribution with tail function $1 - F(y) = 1/y$, for $y > 1$, and its main aim will be to highlight that for heavy-tailed data, NGG-$\mathcal{N}$ mixing leads to much better fits at the tails than Dirichlet process mixing. Beyond the univariate scenario, we also considered bivariate and conditional scenarios that contemplate different dependence levels and complexities of the marginals. Table 2 summarizes the bivariate and conditional scenarios, which are marginally characterized by

$$\begin{cases} f_k(y) = w \, f_{\mathrm{LG}}\{y \mid a_1, b_1\} + (1 - w) \, f_{\mathrm{LG}}\{y \mid a_2, b_2\}, \\ f_k(y \mid x) = w \, f_{\mathrm{LG}}\{y \mid a_1(x), b_1(x)\} + (1 - w) \, f_{\mathrm{LG}}\{y \mid a_2(x), b_2(x)\}, \end{cases} \quad (4.1)$$

for $k = 1, 2$, where $f_{\mathrm{LG}}(y; a, b) = b^a \log(y)^{a-1} y^{-(b+1)}/\Gamma(a)$ is the density of a log-gamma distribution with shape $a > 0$ and rate $b > 0$; parenthetically, we note that the log-gamma distribution is in the Fréchet domain of attraction with tail index b (Beirlant et al., 2004, Table 2.1).

The dependence is modeled via a Gumbel copula, so that data for the bivariate and conditional scenarios are respectively simulated from

$$\begin{cases} F(y_1, y_2) = C_\theta\{F_1(y_1), F_2(y_2)\}, \\ F(y_1, y_2 \mid x) = C_\theta\{F_1(y_1 \mid x), F_2(y_2 \mid x)\}. \end{cases} \quad (4.2)$$

Here, $C_\theta(u, v) = \exp[-\{(-\log u)^\theta + (-\log v)^\theta\}^{1/\theta}]$, for $(u, v) \in (0, 1)^2$, whereas $\theta \geq 1$ is the parameter controlling dependence, and F_1 and F_2 are the distribution functions of f_1 and f_2. For all scenarios, we have simulated $n = 1\,000$ observations, and for the conditional scenarios, covariates were drawn from a standard uniform distribution.

Scenario	Marginal (f_1)	Marginal (f_2)	Copula (θ)
Bivariate 1	$a_1 = a_2 = 5; w = 1$	$a_1 = a_2 = 5; w = 1$	3
2	$a_1 = a_2 = 5; w = 1$	$a_1 = a_2 = 5; w = 1$	1
3	$a_1 = 13; b_1 = 7;$ $a_2 = 10; b_2 = 8; w = .4$	$a_1 = 8; b_1 = 7;$ $a_2 = 15; b_2 = 8; w = .4$	1
Conditional 1	$a_1 = 1 + 4x; a_2 = 3; w = 1$	$a_1 = 1 + 4x; a_2 = 3; w = 1$	1
2	$a_1 = 1 + 4x; a_2 = 3; w = 1$	$a_1 = 1 + 4x; a_2 = 3; w = 1$	3
3	$a_1 = 11 + 5x; b_1 = 8 + 5x;$ $a_2 = b_2 = 7; w = .4$	$a_1 = 6 + 5x; b_1 = 12 + 5x;$ $a_2 = b_2 = 8; w = .4$	1

Table 2: Bivariate and Conditional Simulation Scenarios. The marginals are mixtures of log-gamma distributions as in (4.1), and dependence is set by a Gumbel copula with parameter θ.

MCMC and model specification: All models were fitted using an efficient slice sampler (Kalli et al., 2011) available from the supplementary material (Section 3). We considered a burn-in period of $5\,000$ iterations, and after that, scanned $5\,000$ samples from the posterior targets of interest. For the univariate scenario, we fitted a particular

NGG-$\mathcal{N}$ mixture; that is, we fit a stable process scale mixture with an uninformative gamma prior and an Erlang kernel (i.e., (2.11) with prior $\lambda \sim \text{Gamma}(0.1, 0.1)$); for the bivariate and conditional scenarios we fitted stable process scale mixtures with an Erlang kernel based on Remark 1 and Equation (3.10), respectively. For the latter, a Gumbel copula was used for the centering, and empirical Bayes was used to set the hyperparameter for θ via maximum likelihood. For the conditional version of the model in the regression parameters in (3.9), we consider the prior $\boldsymbol{\beta}_h \overset{\text{iid}}{\sim} N_p(\mathbf{0}, s^2 \boldsymbol{I})$, and set the hyperparameter to be $s^2 = 100$. For the hyperparameters of the marginals Pareto for the base measure, we set $\beta_k = 1$ and tail index $\alpha_{0,k} = 2$, for $k = 1, 2$, which implies that both margins are apriori heavy-tailed but have a finite expected value. Finally, we have assigned the prior $D \sim \text{Beta}(0.5, 0.5)$ to the discount parameter for all instances of the model. Keeping in mind space constraints and the preference for stable process scale mixtures noted in Section 2.2, here we mainly concentrate on assessing the performance of the latter. The posterior inference algorithms available from the supplementary material (Section 3) can, however, be used for fitting multivariate as well as conditional heavy-tailed NGG shape mixtures, and some instances of the latter are available from the `NGGR` package.

One-shot experiments: One-shot experiments for the bivariate and conditional scenarios are presented in the supplementary material (Section 3). All in all, the resulting fits suggest that the proposed methods accurately recover the true distribution for all scenarios being examined. Such findings should, of course, be regarded as tentative, as they are the outcome of a single-run experiment and will be subject to the scrutiny of the Monte Carlo simulation study in the next section.

4.2   Monte Carlo simulation study

We now report the main findings of a Monte Carlo simulation study. For each scenario from Section 4.1, we simulated 100 data sets, each containing $n = 1\,000$ observations. All models have been fitted using stable process scale mixture models with the same specifications and MCMC settings described in Section 4.1. We start with the univariate unit Pareto scenario, which will reinforce our preference for NGG-$\mathcal{N}$ scale mixture models.

In Figure 2, we present the posterior Monte Carlo means of the log-survival estimates for the tail of the univariate scenario and compare it with the corresponding Monte Carlo mean for a Dirichlet process mixture based on the same kernel. As can be seen from Figure 2, a stable process mixture accurately estimates the tail, whereas a Dirichlet process mixture markedly underestimates it; this numerical evidence showcases that the proposed stable process scale mixtures are a natural option for modeling risk and extremes in a heavy-tailed framework. Such numerical performance of the proposed methods finding is not surprising in light of Theorem 4 a); the performance over the bulk (not shown) is comparable for both forms of mixing. Interestingly, Figure 2 also reveals that the Monte Carlo mean based on fitting the Dirichlet process mixture with a Pareto kernel and a gamma-centering distribution overestimates the tail of the distribution. Such numerical finding is not surprising, keeping in mind Theorem 4 b), given that the
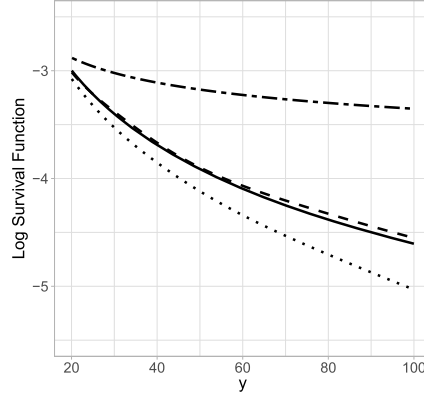
Figure 2: Mean of the Monte Carlo fits (dashed line) for the log-survival function for the univariate scenario obtained using a particular NGG-$\mathcal{N}$ mixture model (i.e., stable process scale mixture) from Section 2.2, focusing from the 95% to the 99% quantile, plotted against the true (solid line). The dotted line shows the Monte Carlo mean of the fits from an NGG-$\mathcal{D}$ mixture with the same Erlang kernel, whereas the dashed–dotted line shows the Monte Carlo mean of the fits from an NGG-$\mathcal{D}$ mixture with a Pareto kernel and a gamma centering distribution.

left endpoint of the gamma centering distribution function is 0 and hence the resulting mixture is super heavy-tailed (i.e., $\alpha(F) = 0$). Next, we move to the bivariate and conditional scenarios from Section 4.1. Figure 3 shows 100 posterior estimated contours for the three bivariate scenarios. As can be seen from the latter figure, the proposed stable process scale mixture model can capture the true contours over different levels of dependence (Scenarios 1–2) and even with challenging marginals such as mixtures (Scenario 3). The results for the conditional scenarios are presented in the supplementary material (Section 4) and also suggest an overall good performance of the proposed methods.

5 Application to heavy-tailed brain data

5.1 Applied context and data description

We now showcase the application of the proposed methods to a neuroscience case study. Brain rhythm signals are key for understanding how the human brain works; loosely speaking, they consist of patterns of neuronal activity that are believed to be linked with certain behaviors, arousal intensity, and sleep states (Frank, 2009). These signals are typically measured using an electroencephalogram (EEG), which records electrical activity in the brain via electrodes attached to the scalp. An EEG signal tracks the activity of billions of neurons, and such signals cover a broad spectrum of frequency bands. Say, the alpha band typically refers to 8–13Hz, while beta refers to 13–20Hz; for a primer on brain rhythms and EEG signals, see, for instance, Buzsaki (2006) and
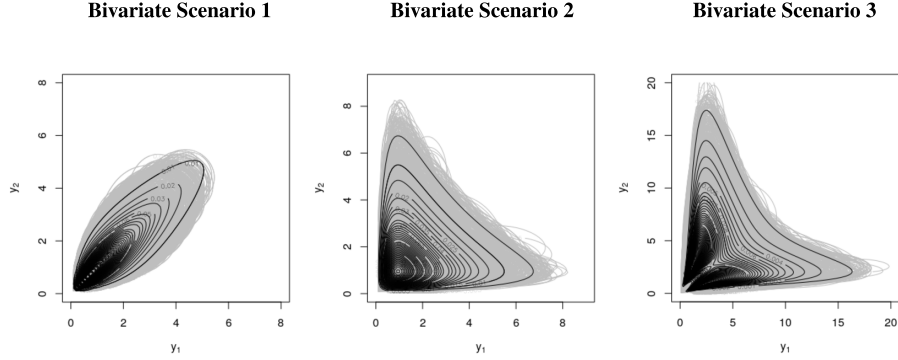
Figure 3: Monte Carlo Simulation for the Bivariate Scenarios 1–3: Contours of the joint density estimates (gray) obtained with proposed stable process scale mixture model from Section 3.1, for the 100 simulated data sets, plotted against the true (black).

Ombao et al. (2016, Chapter 7). Alpha and beta rhythms are believed to be heavy-tailed (e.g., Roberts et al., 2015). Hence, the main goal of our analysis will be to learn about the marginal and joint distribution of these heavy-tailed oscillations, given a variety of stimuli to be described below. In the supplementary material (Section 5), we report evidence supporting the claim that in line with Roberts et al. (2015) our alpha and beta brainwave data are indeed heavy-tailed. We assess this by learning from data about the so-called extreme value index of a generalized Pareto distribution—which is known to be positive for heavy-tailed data (Coles, 2001, Section 4). The data to be analyzed are available from the R package `NGGR`, and were gathered from a UC Berkeley study that involved 30 participants who were subject to several audio-visual activities and stimuli, namely: *mathematics*, *relaxation*, *music*, *color*, *video* as well as *relax and think*.

Figure 4 shows the spectral power in (micro)-Volts squared (μV^2) for alpha and beta waves for all participants; roughly speaking, higher peaks indicate higher neural activity at a certain point in time on a frequency band of interest. The Ljung–Box test results reported in the supplementary material (Section 5) suggest that the recorded trajectories of spectral power can be regarded as independent over time across all stimuli. Figure 4 illustrates some aspects of alpha and beta bands that help to build intuition on their signatures for the different stimuli. For instance, when the stimulus is *mathematics* we can notice high activity for both alpha and beta waves, similar behavior to watching a *video* or finding the *color*—all these being tasks that relate to immediate attention. Indeed, it has been suggested that alpha bands tend to be associated with 'attention' as well as with 'information processing' (Klimesch, 2012). Interestingly, the patterns of alpha and beta waves for both *mathematics* and *music* are reasonably similar—which might not be surprising in light of what has been claimed elsewhere (e.g., Boettcher et al., 1994, and references therein). Whether that similarity of *mathematics* and *music* also holds for the *joint* distribution of alpha and beta waves is something to be examined below in Section 5.3. Next, we learn from the heavy-tailed brainwave data discussed above using the methods proposed in Sections 2–3. Except where mentioned otherwise,
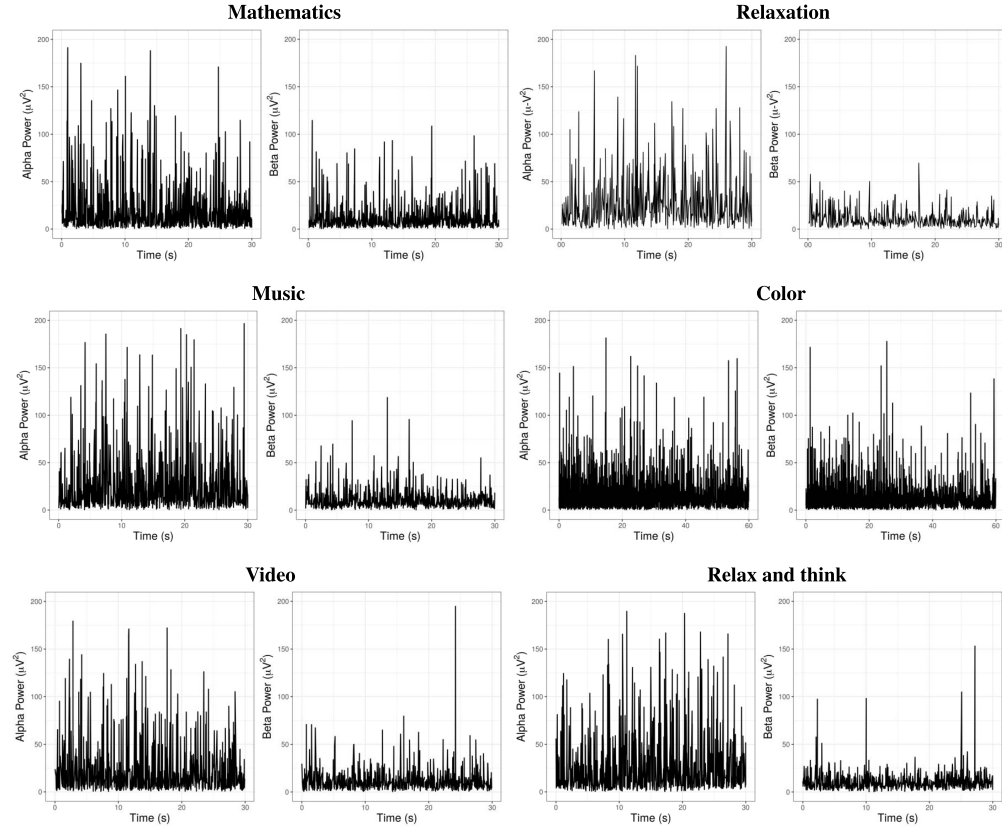
Figure 4: Spectral power (μV^2) of alpha and beta brainwave data plotted against time in seconds (s).

all fits have been conducted using the same model specifications and MCMC settings as in Section 4.

5.2 Marginal brainwave analysis

Figure 5 shows the marginal density estimates of alpha and beta power pooling all subjects for each stimulus that were obtained using the proposed stable process scale mixture model from Section 2.2. Specifically, the fits from Figure 5 were obtained using the specification in (2.11) along with an uninformative gamma prior and an Erlang kernel. To assess the quality of the obtained fits, we depict in Figure 5 q-q boxplots (Rodu and Kafadar, 2022) of random quantile residuals (Dunn and Smyth, 1996).

The obtained q-q boxplots provide evidence that the stable process scale mixture adjusts well both the bulk and the right tail of the data. We have also compared the fitted stable process scale mixture model against the DP shape mixtures in (2.9) with a Pareto
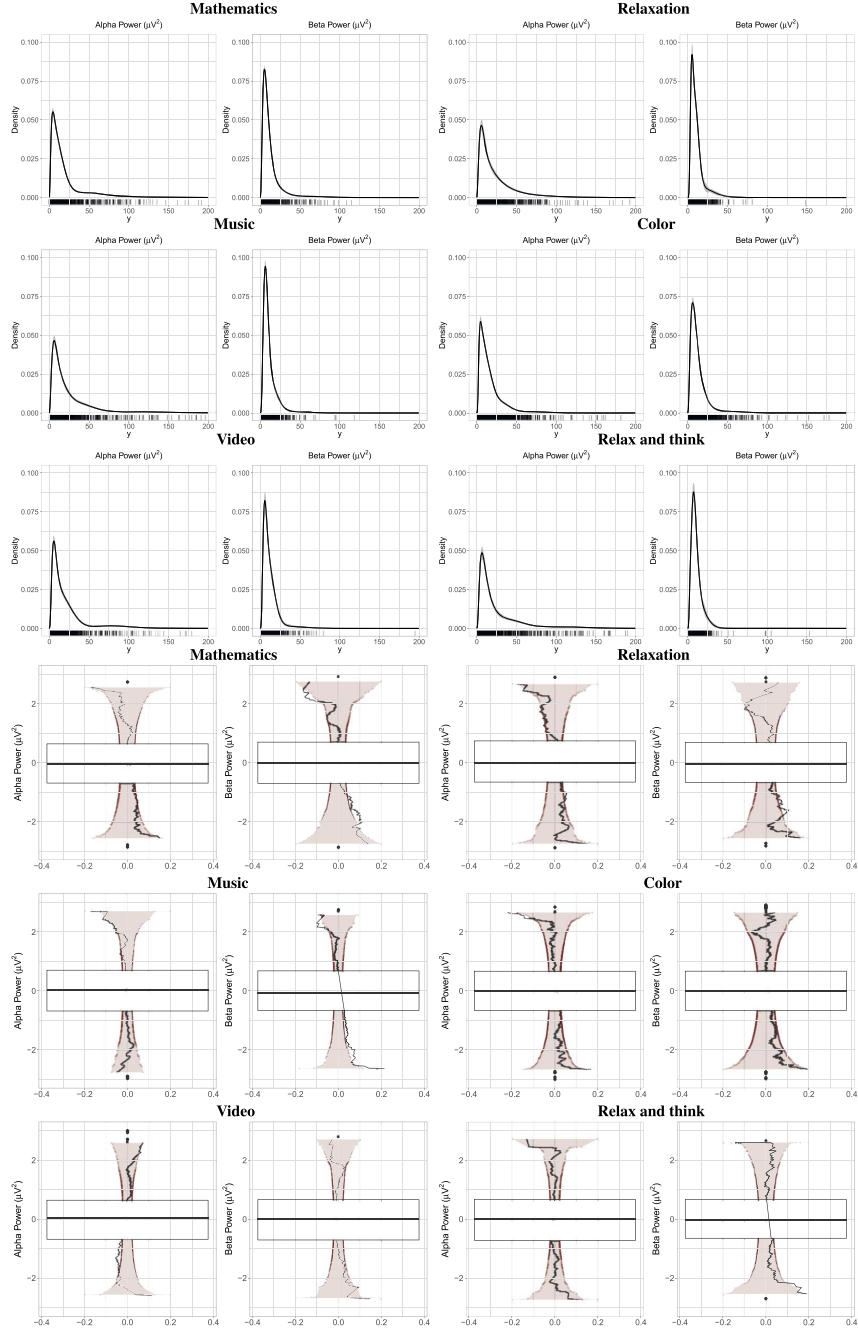
Figure 5: Top: Marginal density estimates of alpha and beta power for each stimulus, obtained with proposed stable process scale mixture model from Section 2.2, along with 95% credible bands. Bottom: Corresponding q-q boxplot of randomized quantile residuals.
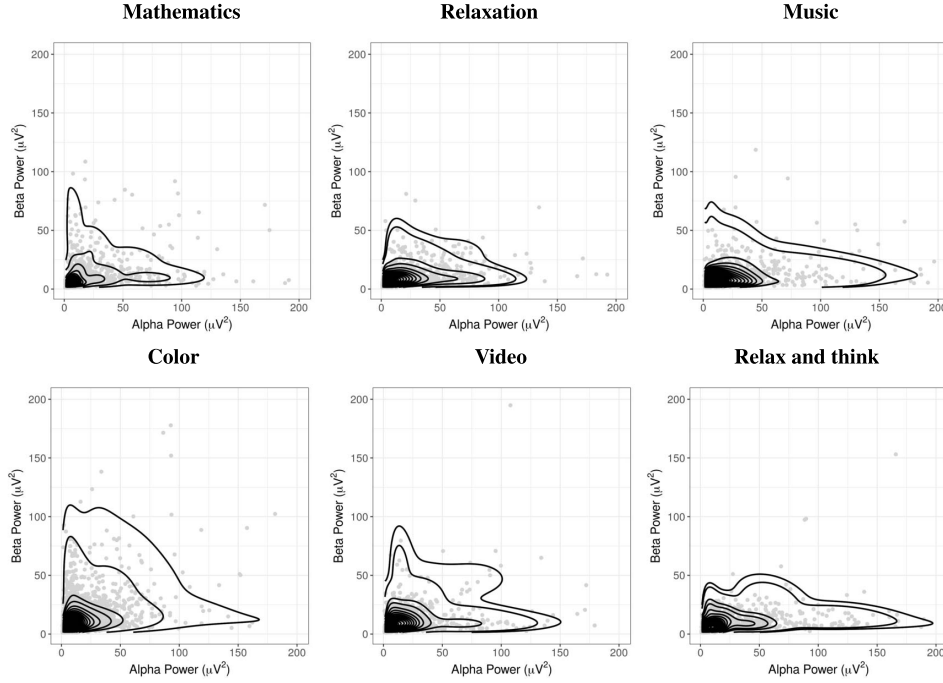
Figure 6: Contours of the posterior conditional joint density estimate of alpha and beta power for each specific stimulus along with raw data; the fit was obtained using the stable process scale mixture from Section 3.2.

kernel and a gamma centering distribution. In line with the findings from the univariate scenario of the simulation study in Section 4.2, we again found evidence in favor of a far more sensible behavior of the stable process scale mixture in comparison with the shape mixture of heavy-tailed kernels. The comparison of the q-q boxplots in Figure 5—for the stable process scale mixture—against those available from the supplementary material (Section 5)—for the DP shape mixture of Pareto kernels—clearly indicates a better performance of the former over the latter over both the left and right tails.

5.3   Stimulus-specific joint brainwave analysis

While Section 5.2 offered a one-dimensional snapshot across different stimuli, we now apply the proposed methods to learn about the joint distribution of the power of brainwaves on alpha and frequency bands, conditional on the activities and stimuli discussed in Section 5.1. To put it differently, we now apply the multivariate regression framework from Sections 3.1–3.2 to learn about the conditional dependence structure governing alpha and beta rhythms, and to borrow strength across stimuli, rather than just fitting each density individually as in Section 5.2.

Figure 6 shows the contours of the fitted conditional joint densities, given the stimulus

under analysis, and it sheds light on the dynamics governing the joint behavior of the alpha and beta brain rhythms. First, the joint densities for some stimuli look similar—such as, for example, *music* and *relax and think*—which suggests a similar joint behavior of the rhythms of alpha and beta bands for these stimuli. Second, *mathematics* and *music*—which looked similar just by examining the raw data in Figure 4 and the marginal fits in Figure 4—have a clearly different dependence structure as can be seen from Figure 6. In other words, while marginally the alpha- and beta-band oscillations for *mathematics* and *music* do look similar, their 'synchronization' or joint behavior looks markedly different.

6   Concluding remarks

This paper studied the tails of the NGG process, and it has shown in particular that the tail of the NGG in $\mathcal{N}$ is tantamount to that of the baseline. This result is in clear contrast to what is known to hold for the DP, whose tails are exponentially much thinner than those of the centering; in addition, we have also derived for the first time envelopes on which the tail of the NGG-$\mathcal{N}$ must lie. We then devised two classes of heavy-tailed NGG mixture models, along with their extensions to a multivariate heavy-tailed setting—as well as to a regression framework. Equipped with the above-mentioned characterization of the tails of the NGG process, we have shown that not all heavy-tailed NGG process mixture models are alike. To put it differently, our theoretical and numerical analyses pinpoint a clear preference for NGG-$\mathcal{N}$ scale mixtures over shape mixtures of heavy-tailed kernels. Particularly, we have shown that shape mixtures of Pareto-type kernels can be super heavy-tailed even though the centering is 'only' heavy-tailed; this implies that a naïve application of the latter mixture models might lead to an overestimation of the mass at the tail—along with poor inferences at the bulk of the distribution. On the contrary, we have found NGG-$\mathcal{N}$ scale mixtures to obey natural properties—such as the stability of the heavy-tail from Corollary 3—and to perform well numerically in both the bulk as well as in the right tail. Keeping in mind the scope of our case study as well as space constraints, we have concentrated the numerical illustrations on the right tail as well as on kernels supported over the positive real line. Yet, we underscore that the theory and methodologies from Sections 2–3 hold more generally over the entire real line as well as for left tails. Finally, the fact that other statistical functionals, such as tail indices, can be readily inferred from the proposed methods implies that, as a byproduct, the conditional version from Section 3 may be used as a tail index regression model in the same vein as Wang and Tsai (2009).

We close the paper with some final comments on open challenges. It is conceivable that similar bounds to the ones derived in this paper could be obtained for the class of processes that can be represented using subordinators. Preliminary derivations lead us to conjecture that a version of Theorem 1 might be possible to obtain for such standardized subordinators with a slowly varying Laplace exponent, and that a version of Theorem 2 might be possible to obtain for such processes with a regularly varying Laplace exponent (with a non-zero index of regular variation). We leave such an open problem for future analysis. As we have shown here, NGG processes obey the stability of the heavy-tail property, and it would be interesting to have a broader understanding

on how large is the class of random probability measures obeying that property. While here the focus has been on the NGG processes and on their mixtures, the potential for modeling heavy-tailed data of other classes of random measures—such as those of Ayed et al. (2019)—remains highly unexplored. Finally, by keeping in mind the importance of modeling rare but catastrophic events in a variety of fields—such as climatology, geology, insurance, risk analysis, and extreme value theory—the methodologies proposed herein may pave the way for further applications and developments at the interface between heavy-tails and Bayesian nonparametrics.

Appendix

A.1 Technical details and auxiliary lemmata

In addition to the auxiliary results below, we first recall a basic fact on subordinators that will be handy for the proof of Theorem 2. If $S(t)$ is a subordinator, then $S(t+h) - S(t)$ has the same distribution as $S(h)$, for every $t, h \geq 0$ (Bertoin, 1999, p. 5). This implies that $\{S(M) - S(t)\}_{t \in [0,1]}$ is equal in distribution to $\{S(M-t)\}_{t \in [0,1]}$, and hence the following subordinator representation holds for the tail of $G \sim \mathrm{NGG}(M, \tau, D, G_0)$:

$$1 - G(y) = 1 - \frac{S\{MG_0(y)\}}{S(M)} \stackrel{\mathrm{d}}{=} \frac{S\{M(1 - G_0(y))\}}{S(M)}. \tag{A.1}$$

Lemma 1 gathers two well-known results on lower and upper envelopes of stochastic processes over the short-run which can be found in Bertoin (1999, Theorem 11) and Bertoin (1999, Theorem 9). Lemma 2 is a well-known result in regular variation (e.g., Embrechts et al., 1997, Theorem A.33) and for the extended Breiman's lemma see Denisov and Zwart (2007, Proposition 2.1). For completeness, we include in the supplementary material (Section 1) some additional comments on regular variation and on heavy-tails.

Lemma 1. *The following results hold:*

a) *If $\{S(t) : t \geq 0\}$ is a subordinator with Laplace exponent $\Phi \in RV_D$, with $D \in (0, 1)$, then*

$$\lim \inf_{t \to 0^+} |S(t)|/l(t) = D(1 - D)^{(1-D)/D}, \quad a.s.,$$

where $l(t) = \log|\log t|/\Phi^{-1}(t^{-1} \log|\log t|)$ for $0 < t < e^{-1}$, and Φ^{-1} is the inverse function of Φ.

b) *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*

$$\lim \sup_{t \to 0^+} \frac{S(t)}{u(t)} = \begin{cases} 0, & \int_0^1 \nu[u(t), \infty) \, \mathrm{d}t < \infty, \\ \infty, & \int_0^1 \nu[u(t), \infty) \, \mathrm{d}t = \infty, \end{cases} \quad a.s.$$

Lemma 2 (Representation theorem)**.** *If $h \in RV_\alpha$ for some $\alpha \in \mathbb{R}$, then*

$$h(y) = c(y) \exp \left\{ \int_z^y \frac{a(u)}{u} \, \mathrm{d}u \right\}, \quad y > z,$$

for some $z > 0$ with $c(y) \to c \in (0, \infty)$, $a(y) \to \alpha$ as $y \to \infty$. The converse also holds.

Lemma 3 (Extended Breiman's lemma)**.** *Let X and Y be random variables and suppose X has a regularly varying tail, $P(X > x) = x^{-\alpha} \mathscr{L}(x)$, with tail index $\alpha \geq 0$, and $Y \geq 0$ with $E(Y^\alpha) < \infty$. Then, if*

$$\lim \inf_{x \to \infty} \mathscr{L}(x) > 0 \quad and \quad P(Y > x) = o\{P(X > x)\},$$

it follows that XY has regularly varying tail with tail index α.

A.2 Proofs of main results

Proof of Theorem 2: We start with the lower envelope. The Laplace exponent of a generalized gamma process with parameters D, and τ, $\Phi(\lambda) = (\lambda + \tau)^D - \tau^D$, is regularly varying at ∞ with index $D \in (0, 1)$, and note also that $\Phi^{-1}(y) = (y + \tau^D)^{1/D} - \tau$. Hence, Lemma 1 a) implies that

$$\lim \inf_{t \to 0^+} \frac{S(t)}{l(t)} = D(1-D)^{(1-D)/D}, \text{with} \quad l(t) = t^{1/D} \left\{ \frac{\log |\log t|}{(\log |\log t| + \tau^D t)^{1/D} - \tau t^{1/D}} \right\}. \tag{A.2}$$

Combining (A.2) with the representation of the NGG-$\mathcal{N}$ process in (2.1) yields

$$\lim \inf_{G_0(y) \to 0^+} \frac{S\{MG_0(y)\}/S(M)}{l\{MG_0(y)\}} = \lim \inf_{G_0(y) \to 0^+} \frac{G(y)}{l\{MG_0(y)\}} = D(1-D)^{(1-D)/D}/S(M).$$

Hence, (A.1) yields

$$\lim \inf_{G_0(y) \to 1^-} \frac{1 - G(y)}{l\{M(1 - G_0(y))\}} = D(1 - D)^{(1-D)/D}/S(M),$$

from where the final result follows. Next, we focus on the upper envelope. First, consider the following family of functions for $r > 0$,

$$w_r(t) = t^{1/D} |\log t|^{r/D} = t^{1/D} \{\log(1/t)\}^{r/D}, \qquad t \in (0, e^{-r}), \tag{A.3}$$

where $D \in (0, 1)$. We start by checking if $w_r(t)$ verifies the assumptions of Lemma 1 b). It follows that $w_r(t)$ is increasing on $(0, \delta)$, with $\delta = \mathrm{e}^{-r}$. Indeed,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \{w_r(t)\} &= D^{-1} t^{1/D-1} \{\log(1/t)\}^{r/D} + t^{1/D} r D^{-1} \{\log(1/t)\}^{r/D-1} (-1/t) \\
&= D^{-1} t^{1/D-1} \{\log(1/t)\}^{r/D} - t^{1/D-1} r D^{-1} \{\log(1/t)\}^{r/D-1} \\
&= D^{-1} t^{1/D-1} [\{\log(1/t)\}^{r/D} - r \{\log(1/t)\}^{r/D-1}] > 0, \quad t \in (0, \delta).
\end{aligned} \tag{A.4}$$

In addition, $t^{-1}w_r(t)$ is positive and continuous over $(0,1)$, and increasing on $t \in (0, e^{r/(D-1)})$ as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{w_r(t)}{t}\right) = t^{1/D-2}D^{-1}\{\log(1/t)\}^{r/D-1}\{-r - (D-1)\log(1/t)\} > 0,$$

when $t < e^{r/(D-1)}$.

The function $w_r(t)$ is increasing in a sufficiently small neighborhood of 0, 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 b). Finally, we derive necessary and sufficient conditions for $\int_0^1 \nu[u_r(t),\infty)\,\mathrm{d}t < \infty$. First, note that

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

and

$$\begin{aligned}
\nu[u,\infty) &= \int_u^\infty \frac{D}{\Gamma(1-D)}x^{-1-D}e^{-\tau x}\,\mathrm{d}x \\
&= \frac{D}{\Gamma(1-D)}\left\{\int_u^1 x^{-1-D}e^{-\tau x}\,\mathrm{d}x + \int_1^\infty x^{-1-D}e^{-\tau x}\,\mathrm{d}x\right\} \\
&\geq \frac{D}{\Gamma(1-D)}e^{-\tau}\int_u^1 x^{-1-D}\,\mathrm{d}x \\
&= \frac{D}{\Gamma(1-D)}e^{-\tau}\left\{\frac{1}{-D} + \frac{u^{-D}}{D}\right\}.
\end{aligned}$$

Thus,

$$\int_0^1 \nu[u_r(t),\infty)\,\mathrm{d}t < \infty \quad \Longleftrightarrow \quad \int_0^1 u_r^{-D}(t)\,\mathrm{d}t < \infty. \qquad (A.5)$$

Hence, applying Lemma 1 b) to the NGG-$\mathcal{N}$ process, with the family of functions on (A.3), yields

$$\lim_{G_0(y)\to 0^+}\sup \frac{G(y)}{u_r\{G_0(y)\}} = \begin{cases} 0, & \int_0^1 u_r^{-D}(t)\,\mathrm{d}t < \infty, \\ \infty, & \int_0^1 u_r^{-D}(t)\,\mathrm{d}t = \infty. \end{cases} \quad \text{a.s.} \qquad (A.6)$$

Then, $\int_0^1 u_r^{-D}(t)\,\mathrm{d}t = \int_0^\delta w_r^{-D}(t)\,\mathrm{d}t + \int_\delta^1 w_r^{-D}(e^{-r})\,\mathrm{d}t$, where the second integral is finite, which means $\int_0^1 u_r^{-D}(t)\,\mathrm{d}t$ is finite if and only if $\int_\delta^1 w_r^{-D}(t)\,\mathrm{d}t < \infty$, which holds if $r > 1$. This implies the final result,

$$\lim_{G_0(y)\to 0^+}\sup \frac{G(y)}{u_r\{G_0(y)\}} = \begin{cases} 0, & r > 1, \\ \infty, & 0 < r \leq 1. \end{cases} \quad \text{a.s.} \qquad (A.7)$$

Proof of Corollary 3: It follows from Theorem 2 that as $y \to y^*$, then

$$1 - G(y) = [M\{1 - G_0(y)\}]^{\{1+o(1)\}/D}, \quad \text{a.s.} \qquad (A.8)$$

By assumption, $1 - G_0 \in \mathrm{RV}_{-\alpha_0}$ and hence it follows by the representation theorem that $1 - G_0(y) = c(y) \exp\{\int_z^y a(u)/u \, du\}$, for some $z > 0$ with $c(y) \to c \in (0, \infty)$, $a(y) \to -\alpha_0$ as $y \to \infty$. This combined with (A.8) yields that

$$1 - G(y) = [M\{1 - G_0(y)\}]^{\{1+o(1)\}/D} = c^*(y) \exp\left\{ \int_z^y \frac{a^*(u)}{u} \, du \right\}, \quad y > z,$$

for some $z > 0$, with

$$c^*(y) = \{Mc(y)\}^{\{1+o(1)\}/D} \to Mc^{1/D} \in (0, \infty), \quad a^*(y) = \left(\frac{1 + o(1)}{D} \right) a(y) \to -\alpha_0/D,$$

as $y \to \infty$. The final result follows from the representation theorem. $\qquad\square$

Proof of Theorem 4

a) Let $U \mid V = \sigma \sim K(\,\cdot\,; \eta_\sigma)$ and $V \sim G$, and consider the decomposition $U = U_+ - U_-$, where $U_+ = \max(U, 0)$ and $U_- = \max(-U, 0)$. Since the focus is on the right tail, we concentrate on U_+, and note below that for $y > 0$ the scale mixture in (2.8) can be written as the density of the product of U_+ and V. In detail, it follows from Rohatgi's well-known result on the product of random variables (e.g., Glen et al., 2004), that

$$f_{U_+V}(y) = \int_0^\infty f_{U_+,V}\left(\sigma, \frac{y}{\sigma}\right) \frac{1}{\sigma} \, d\sigma. \qquad (A.9)$$

Hence, combining the fact that

$$U_+ \mid V = \sigma \sim K(y; \eta_\sigma) I(y > 0) + P(U_+ = 0 \mid V = \sigma) I(y = 0),$$

along with Bayes theorem implies that (A.9) can be rewritten as

$$\begin{aligned}
f_{U_+V}(y) &= \int_0^\infty f_{U_+|V}\left(\frac{y}{\sigma}, \sigma\right) \frac{dG(\sigma)}{d\sigma} \frac{1}{\sigma} \, d\sigma \\
&= \int_0^\infty K\left(\frac{y}{\sigma}; \eta_\sigma\right) \frac{1}{\sigma} \, dG(\sigma) \\
&= f(y),
\end{aligned} \qquad (A.10)$$

for $y > 0$. Since by assumption G_0 has a regularly varying tail with tail index α_0, it follows from Corollary 3 that V has a regularly varying tail with tail index α_0/D. Next, let $V_0 \sim G_0$ and $\mathscr{L}^*(\sigma) = \{\mathscr{L}(\sigma)\}^{1/D}$ and note that the assumptions along with Corollary 3 and the representation theorem imply $E(U_+^{\alpha_0}) < \infty$,

$$P(V > \sigma) = \sigma^{-\alpha_0/D} \mathscr{L}^*(\sigma), \qquad P(U_+ > \sigma) = o\{P(V_0 > \sigma)\}^{1/D} = o\{P(V > \sigma)\},$$

as well as that

$$\lim_{\sigma \to \infty} \mathscr{L}^*(\sigma) = \{ \lim_{\sigma \to \infty} \mathscr{L}(\sigma)\}^{1/D} > 0.$$

In other words, the assumptions of the extended Breiman's lemma apply from where it readily follows that U_+V is regularly varying at infinity with tail index α_0/D, and hence the same claim can be made about $f_{U_+V}(y) = f(y)$. This proves the result. $\qquad\square$

b) Let $i(h)$ be a permutation such that $\inf\{\alpha : G_0(\alpha) > 0\} \equiv \alpha_{i(1)} \leq \alpha_{i(2)} \leq \cdots$. Then,

$$1 - F(y) = \sum_{h=1}^{\infty} \pi_h \frac{\mathscr{L}(y)}{y^{\alpha_h}} = \sum_{j=1}^{\infty} \frac{\pi_{i(j)} \mathscr{L}(y)}{y^{\alpha_{i(j)}}} = \frac{\mathscr{L}^*(y)}{y^{\alpha_{i(1)}}},$$

and it can be easily shown that $\mathscr{L}^*(y) = \mathscr{L}(y)\{\pi_{i(1)} + \sum_{j=2}^{\infty} \pi_{i(j)}/(y^{\alpha_{i(j)} - \alpha_{i(1)}})\}$ is a slowly varying function, from where the final result follows. □

Proof of Theorem 5: We only present the proof of Theorem 5 a) as that of claim b) follows a similar line of attack. We start by showing that the marginal distributions F_k are univariate NGG-mixtures, and then using Theorem 4 a) it follows that their tails, $1 - F_k$, are regularly varying, for $k = 1, \ldots, d$. Let $d\mathbf{y}_{-k} = dy_1 \ldots dy_{k-1}dy_{k+1} \ldots dy_d$ and note that (3.1) and (3.2) implies that

$$f_k(y_k) = \int_{\mathbb{R}_+^{d-1}} f(\mathbf{y}) \, d\mathbf{y}_{-k} = \int_{\mathbb{R}_+^{d-1}} \sum_{h=1}^{\infty} \pi_h \mathbf{K}(\mathbf{y}; \boldsymbol{\eta}_{\boldsymbol{\sigma}_h}) \, d\mathbf{y}_{-k} = \sum_{h=1}^{\infty} \pi_h K_{\sigma_k}(y_k; \eta_{\sigma_{h,k}}).$$

Since by assumption $G_{0,k}(\sigma_k)$ has a regularly varying tail with tail index $\alpha_{0,k}$, it follows from Theorem 4 a) that $1 - F_k$ is regularly varying with tail index $\alpha(F_k) = \alpha_{0,k}/D$, for $k = 1, \ldots, d$, from where the final result follows. □

Acknowledgments

Funding

Supplementary Material

Supplementary material for "Heavy-Tailed NGG-Mixture Models" (DOI: 10.1214/24-BA1420SUPP; .pdf). The online supplementary materials contain further technical details and proofs, supporting numerical evidence, as well as the R-package NGGR which implements instances of the methods proposed herein, and includes the dataset from Section 5.

References

Applebaum, D. (2009). *Lévy Processes and Stochastic Calculus*. Cambridge: Cambridge University Press. MR2512800. doi: https://doi.org/10.1017/CBO9780511809781. 3

Ayala, D., Jofré, L., Gutiérrez, L., and Mena, R. H. (2022). "On a Dirichlet process mixture representation of phase-type distributions." *Bayesian Analysis*, 17(3): 765–790. MR4483238. doi: https://doi.org/10.1214/21-ba1272. 9

Ayed, F., Lee, J., and Caron, F. (2019). "Beyond the Chinese restaurant and Pitman-Yor processes: Statistical models with double power-law behavior." In *International Conference on Machine Learning*, 395–404. PMLR. 21

Barrientos, A. F., Jara, A., and Quintana, F. A. (2012). "On the support of MacEachern's dependent dirichlet processes and extensions." *Bayesian Analysis*, 7: 277–310. MR2934952. doi: https://doi.org/10.1214/12-BA709. 12

Barrios, E., Lijoi, A., Nieto-Barajas, L. E., and Prünster, I. (2013). "Modeling with normalized random measure mixture models." *Statistical Science*, 28(3): 313–334. MR3135535. doi: https://doi.org/10.1214/13-STS416. 1

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Hoboken, NJ: Wiley. MR2108013. doi: https://doi.org/10.1002/0470012382. 13

Bertoin, J. (1996). *Lévy processes*, volume 121. Cambridge: CUP. MR1406564. 6

Bertoin, J. (1999). "Subordinators: Examples and applications." In *Lectures on Probability Theory and Statistics*. New York: Springer. 3, 21

Bingham, N. H., Goldie, C. M., Teugels, J. L., and Teugels, J. (1989). *Regular Variation*. Cambridge: Cambridge University Press. MR0898871. doi: https://doi.org/10.1017/CBO9780511721434. 3

Bissiri, P. G. and Ongaro, A. (2014). "On the topological support of species sampling priors." *Electronic Journal of Statistics*, 8: 861–882. MR3229100. doi: https://doi.org/10.1214/14-EJS912. 6

Bladt, M. and Rojas-Nandayapa, L. (2018). "Fitting phase–type scale mixtures to heavy–tailed data and distributions." *Extremes*, 21(2): 285–313. MR3800300. doi: https://doi.org/10.1007/s10687-017-0306-4. 2, 9

Boettcher, W. S., Hahn, S. S., and Shaw, G. L. (1994). "Mathematics and music: A search for insight into higher brain function." *Leonardo Music Journal*, 53–58. 16

Buzsaki, G. (2006). *Rhythms of the Brain*. Oxford: Oxford University Press. MR2270828. doi: https://doi.org/10.1093/acprof:oso/9780195301069.001.0001. 15

Camerlenghi, F., Dunson, D. B., Lijoi, A., Prünster, I., and Rodríguez, A. (2019a). "Latent nested nonparametric priors (with Discussion)." *Bayesian Analysis*, 14(4): 1303. MR4044854. doi: https://doi.org/10.1214/19-BA1169. 11

Camerlenghi, F., Lijoi, A., Orbanz, P., and Prünster, I. (2019b). "Distribution theory for hierarchical processes." *The Annals of Statistics*, 47(1): 67–92. MR3909927. doi: https://doi.org/10.1214/17-AOS1678. 11

Catalano, M., Lijoi, A., and Prünster, I. (2021). "Measuring dependence in the Wasserstein distance for Bayesian nonparametric models." *The Annals of Statistics*, 49(5): 2916–2947. MR4338898. doi: https://doi.org/10.1214/21-aos2065. 11

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer. MR1932132. doi: https://doi.org/10.1007/978-1-4471-3675-0. 16

Denisov, D. and Zwart, B. (2007). "On a theorem of Breiman and a class of random difference equations." *Journal of Applied Probability*, 44(4): 1031–1046. MR2382943. doi: https://doi.org/10.1239/jap/1197908822. 21

Doss, H. and Sellke, T. (1982). "The tails of probabilities chosen from a Dirichlet prior." *The Annals of Statistics*, 10(4): 1302–1305. MR0673666. 2, 4, 5, 23

Dunn, P. K. and Smyth, G. K. (1996). "Randomized quantile residuals." *Journal of Computational and Graphical Statistics*, 5(3): 236–244. 17

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. New York: Springer. MR1458613. doi: https://doi.org/10.1007/978-3-642-33483-2. 21

Favaro, S., Lijoi, A., Nava, C., Nipoti, B., Prünster, I., and Teh, Y. (2016). "On the stick-breaking representation for homogeneous NRMIs." *Bayesian Analysis*, 11(3): 697–724. MR3498043. doi: https://doi.org/10.1214/15-BA964. 1, 4

Frank, M. G. (2009). *Brain Rhythms*, 482–483. Berlin, Heidelberg: Springer. 15

Ghosal, S. and Van der Vaart, A. W. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge: Cambridge University Press. MR3587782. doi: https://doi.org/10.1017/9781139029834. 1, 2, 5

Glen, A. G., Leemis, L. M., and Drew, J. H. (2004). "Computing the distribution of the product of two continuous random variables." *Computational Statistics & Data Analysis*, 44(3): 451–464. MR2026756. doi: https://doi.org/10.1016/S0167-9473(02)00234-7. 24

James, L. F., Lijoi, A., and Prünster, I. (2009). "Posterior analysis for normalized random measures with independent increments." *Scandinavian Journal of Statistics*, 36(1): 76–97. MR2508332. doi: https://doi.org/10.1111/j.1467-9469.2008.00609.x. 1

Kalli, M., Griffin, J. E., and Walker, S. G. (2011). "Slice sampling mixture models." *Statistics and Computing*, 21: 93–105. MR2746606. doi: https://doi.org/10.1007/s11222-009-9150-y. 13

Klimesch, W. (2012). "Alpha-band oscillations, attention, and controlled access to stored information." *Trends in Cognitive Sciences*, 16(12): 606–617. 16

Li, C., Lin, L., and Dunson, D. B. (2019). "On posterior consistency of tail index

for Bayesian kernel mixture models." *Bernoulli*, 25(3): 1999–2028. MR3961238. doi: https://doi.org/10.3150/18-BEJ1043. 8

Lijoi, A., Mena, R. H., and Prünster, I. (2005). "Hierarchical mixture modeling with normalized inverse-Gaussian priors." *Journal of the American Statistical Association*, 100(472): 1278–1291. MR2236441. doi: https://doi.org/10.1198/016214505000000132. 4

Lijoi, A., Mena, R. H., and Prünster, I. (2007). "Controlling the reinforcement in Bayesian non-parametric mixture models." *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(4): 715–740. MR2370077. doi: https://doi.org/10.1111/j.1467-9868.2007.00609.x. 1, 3, 4

Lijoi, A., Pruenster, I., et al. (2010). "Models beyond the Dirichlet process." In *Bayesian Nonparametrics*, volume 28, 80–136. Cambridge, MA: Cambridge University Press. MR2730661. 1

Müller, P. and Mitra, R. (2013). "Bayesian nonparametric inference—why and how." *Bayesian Analysis*, 8: 269–302. MR3066939. doi: https://doi.org/10.1214/13-BA811. 2

Müller, P., Quintana, F. A., Jara, A., and Hanson, T. (2015). *Bayesian Nonparametric Data Analysis*. New York: Springer. MR3309338. doi: https://doi.org/10.1007/978-3-319-18968-0. 2

Nelsen, R. B. (2006). *An Introduction to Copulas*. New York: Springer, 2nd ed edition. MR2197664. doi: https://doi.org/10.1007/s11229-005-3715-x. 11

Ombao, H., Lindquist, M., Thompson, W., and Aston, J. (2016). *Handbook of Neuroimaging Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC. MR3618824. 16

Quintana, F. A., Müller, P., Jara, A., and MacEachern, S. N. (2022). "The dependent Dirichlet process and related models." *Statistical Science*, 37(1): 24–41. MR4371095. doi: https://doi.org/10.1214/20-sts819. 12

Palacios Ramírez, V., de Carvalho, M., and Gutiérrez, L. (2024). Supplementary material for "Heavy-Tailed NGG-Mixture Models". doi: https://doi.org/10.1214/24-BA1420SUPP. 3

Regazzini, E., Lijoi, A., and Prünster, I. (2003). "Distributional results for means of normalized random measures with independent increments." *The Annals of Statistics*, 31(2): 560–585. MR1983542. doi: https://doi.org/10.1214/aos/1051027881. 1

Resnick, S. (2007). *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer. MR2271424. 2

Roberts, J. A., Boonstra, T. W., and Breakspear, M. (2015). "The heavy tail of the human brain." *Current Opinion in Neurobiology*, 31: 164–172. 16

Rodu, J. and Kafadar, K. (2022). "The q–q boxplot." *Journal of Computational and Graphical Statistics*, 31(1): 26–39. MR4387208. doi: https://doi.org/10.1080/10618600.2021.1938586. 17

Sarabia Alegría, J. M., Gómez Déniz, E., et al. (2008). "Construction of multivariate distributions: A review of some recent results." *SORT*, 32: 3–36. MR2440766. 10

Tressou, J. (2008). "Bayesian nonparametrics for heavy tailed distribution. Application to food risk assessment." *Bayesian Analysis*, 3(2): 367–391. MR2407431. doi: https://doi.org/10.1214/08-BA314. 2, 9

Wang, H. and Tsai, C.-L. (2009). "Tail index regression." *Journal of the American Statistical Association*, 104(487): 1233–1240. MR2750246. doi: https://doi.org/10. 1198/jasa.2009.tm08458. 20