Supplementary Material for
Beyond the Chinese Restaurant and Pitman-Yor processes: Statistical Models with double power-law behavior

Fadhel Ayed [*1] **Juho Lee** [*1 2] **François Caron** [1]

1. Background on regular variation

The material in this section is from the book of Bingham et al. (1989). In the following, U denotes a regularly varying function and ℓ denotes a slowly varying function, locally bounded on $(0, \infty)$.

Theorem 1 (Karamata's theorem) *(Bingham et al., 1989, Propositions 1.5.8 and 1.5.10). Suppose $\rho > -1$ and $U(t) \sim t^\rho \ell(t)$ as t tends to infinity. Then*

$$\int_0^x U(t)dt \sim \frac{1}{\rho+1} x^{\rho+1}\ell(x)$$

as x tends to infinity.
Suppose $\rho < -1$. Then $U(t) \sim t^\rho \ell(t)$ as t tends to infinity implies

$$\int_x^\infty U(t)dt \sim -\frac{1}{\rho+1} x^{\rho+1}\ell(x)$$

as x tends to infinity.

Corollary 2 *Suppose $\rho < -1$ and $U(y) \sim y^\rho \ell(1/y)$ as y tends to 0. Then*

$$\int_x^\infty U(y)dy \sim \frac{-1}{\rho+1} x^{\rho+1}\ell(1/x)$$

as x tends to 0.

Proof. $U(t) = t^\rho \ell_1(1/t)$ where $\ell_1(t) \sim \ell(t)$ as $t \to \infty$.

$$\begin{aligned}\int_x^\infty U(y)dy &= \int_0^{1/x} t^{-2-\rho}\ell_1(t)dt \\ &\sim \frac{-1}{1+\rho} x^{1+\rho}\ell(1/x)\end{aligned}$$

as x tends to 0 by Theorem 1. ■

*Equal contribution [1]Department of Statistics, University of Oxford, Oxford, United Kingdom [2]AITRICS, Seoul, Republic of Korea. Correspondence to: Fadhel Ayed <fadhel.ayed@stats.ox.ac.uk>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

2. Proofs

2.1. Proof of Equations (18) and (19)

For any $s > 0$, the function $x \to \gamma(s, x)$ is both regularly varying at 0 and infinity with

$$\gamma(s,x) \sim \begin{cases} \frac{x^s}{s} & \text{as } x \to 0 \\ \Gamma(s) & \text{as } x \to \infty \end{cases}$$

we have therefore for the generalized BFRY process

$$\rho(w) \sim \begin{cases} \frac{c^{\tau-\sigma}}{\Gamma(1-\sigma)(\tau-\sigma)} w^{-1-\sigma} & \text{as } x \to 0 \\ \frac{\Gamma(\tau-\sigma)}{\Gamma(1-\sigma)} w^{-1-\tau} & \text{as } x \to \infty \end{cases}$$

and Equations (18) and (19) follow from Theorem 1 and Corollary 2.

2.2. Proof of Equations (21) and (22)

We have for the beta prime process

$$\rho(w) \sim \begin{cases} \frac{c^{\sigma-\tau}\Gamma(\tau-\sigma)}{\Gamma(1-\sigma)} w^{-1-\sigma} & \text{as } x \to 0 \\ \frac{\Gamma(\tau-\sigma)}{\Gamma(1-\sigma)} w^{-1-\tau} & \text{as } x \to \infty \end{cases}$$

Equations (21) and (22) then follow from Theorem 1 and Corollary 2.

2.3. Proof of Proposition 1

Lemma 1 *Let $(X_k)_{k\geq 1}$ be a sequence of Poisson random variables such that*

$$\frac{\log k}{\mathbb{E}X_k} \to 0.$$

Then

$$\frac{X_k}{\mathbb{E}X_k} \to 1 \textit{ a.s.}$$

Proof. Let X be a Poisson random variable with parameter λ. Using the Chernoff bound, it comes that for any $t > 0$

$$\mathbb{P}(|X-\lambda| \geq \lambda t) \leq 2e^{-\frac{\lambda t^2}{2(1+t)}}.$$

Let $0 < \epsilon < 1/2$. We deduce from previous inequality that

$$\mathbb{P}\left(\left|\frac{X_k}{\mathbb{E}X_k} - 1\right| \geq \epsilon\right) \leq 2e^{-\frac{\epsilon^2 \mathbb{E}X_k}{4}} = 2k^{-\frac{\epsilon^2 \mathbb{E}X_k}{4\log k}}$$

Using the assumption, we have that $-\frac{\epsilon^2\mathbb{E}X_k}{4\log k} \to -\infty$. Therefore, the RHS is summable. The almost sure result follows from Borel-Cantelli lemma. ■

Now we can prove Proposition 1. Let $\nu = \sum_k \delta_{w_k}$. Then, for all $x > 0$, $\nu([x, +\infty))$ is a Poisson random variable with mean $\overline{\rho}(x)$. Let us show that,

$$\nu([1/x, +\infty)) \overset{x\to+\infty}{\sim} x^\alpha \ell(x) \quad a.s.$$

Using Lemma 1 on the sequence $(\nu([1/k, +\infty)))_{k\geq 1}$, we find that

$$\nu([1/k, +\infty)) \overset{k\to+\infty}{\sim} k^\alpha \ell_1(k) \quad a.s.$$

Now, since $x \mapsto \nu([1/x, +\infty))$ is almost surely non decreasing, it comes that

$$\nu([1/\lfloor x\rfloor, +\infty)) \leq \nu([1/x, +\infty)) \leq \nu([1/\lfloor x+1\rfloor, +\infty))$$

We get the desired result by noticing that

$$(\lfloor x+1\rfloor)^\alpha \ell_1(\lfloor x+1\rfloor) \sim (\lfloor x\rfloor)^\alpha \ell_1(\lfloor x\rfloor) \sim x^\alpha \ell_1(x).$$

Now, pick K_0 such that $\sum_{k\geq K_0} w_{(k)} < 1$, and define $p_k = w_{(k)}$ if $k \geq K_0$ and $p_k = \frac{1-\sum_{j\geq K_0} w_{(j)}}{K_0 - 1}$ otherwise. Notice that $\sum p_k = 1$ and for $x \leq w_{(K_0)}$,

$$\#\{p_k | p_k \leq x\} = \nu([x, +\infty)).$$

We can therefore apply Proposition 23 of Gnedin et al. (2007), leading to

$$p_k \sim k^{-1/\alpha} \ell_1^*(k)$$

with $k \to +\infty$, where ℓ_1^* is a slowly varying function defined by

$$\ell_1^*(x) = \frac{1}{\{\ell_1^{1/\alpha}(x^{1/\alpha})\}^\#},$$

where $\ell^\#$ denotes a de Bruijn conjugate (Bingham et al., 1989, Definition 1.5.13) of the slowly varying function ℓ. Therefore, since $w_{(k)} = p_k$ for k large enough, it comes that

$$w_{(k)} \sim k^{-1/\alpha} \ell_1^*(k)$$

almost surely as $k \to \infty$.

2.4. Proof of Proposition 2

The proof of this proposition follows the line of the proof of Theorem 1.2 of Kevei & Mason (2014). Let

$$\overline{\rho}^{-1}(y) = \sup\{x \mid \overline{\rho}(x) > y\}$$

denote the inverse tail Lévy intensity. Let $(w_{(k)})_{k\geq 1}$ be the ordered jumps of a CRM with Lévy measure $\eta\rho(dw)$. From the inverse Lévy measure representation of a real valued Poisson point process, we know that

$$(w_{(k)})_{k\geq 1} \overset{d}{=} (\overline{\rho}^{-1}(\Gamma_k/\eta))_{k\geq 1},$$

where $(\Gamma_k)_{k\geq 1}$ are the points of a unit-rate Poisson point process on $(0, \infty)$, sorted in increasing order. In particular, we have that

$$(w_{(k_1)}, w_{(k_1+k_2)}) \overset{d}{=} \left(\overline{\rho}^{-1}\left(\frac{X_1}{\eta}\right), \overline{\rho}^{-1}\left(\frac{X_1+X_2}{\eta}\right)\right),$$

where X_1 and X_2 are independent Gamma random variables, with respective parameters $(k_1, 1)$ and $(k_2, 1)$. Therefore,

$$\frac{w_{(k_1+k_2)}}{w_{(k_1)}} \overset{d}{=} \frac{\overline{\rho}^{-1}(X_1/\eta)}{\overline{\rho}^{-1}((X_1+X_2)/\eta)}.$$

Since $\overline{\rho}^{-1}$ is the generalized inverse of $\overline{\rho}_1$, which is regularly varying at ∞ with parameter τ, it follows from Lemma 22 of (Gnedin et al., 2007) that $\overline{\rho}^{-1}$ is regularly varying at 0 with parameter $1/\tau$. Therefore, the right-hand side expression of the last equation converges almost surely to $\frac{X_1^{1/\tau}}{(X_1+X_2)^{1/\tau}}$ as $\eta \to \infty$. From which we conclude that

$$\frac{w^\tau_{(k_1+k_2)}}{w^\tau_{(k_1)}} \overset{d}{\to} \frac{X_1}{X_1+X_2} \overset{d}{=} \mathrm{Beta}(k_1, k_2),$$

as $\eta \to +\infty$

2.5. Proof of Proposition 3

In order to prove this proposition, we need to introduce some notations and results on generalized-kernel based Abelian theorems. Interested reader can refer to Chapter 4 of Bingham et al. (1989) for more details. Given a measurable kernel $k : (0, \infty) \to \infty$, let

$$\check{k}(z) = \int_0^\infty t^{-z-1} k(t)dt = \int u^{z-1} k(1/u)du$$

be its Mellin transform, for $z \in \mathbb{C}$ such that the integral converges. We will use Theorem 4.1.6 page 201 in (Bingham et al., 1989) (that we recall here after) to derive the behaviour at $+\infty$.

Theorem 3 (Theorem 4.1.6 page 201 in Bingham et al.) *Suppose that k converge at least in the strip*

$\sigma \le \mathrm{Re}(z) \le \Sigma$, where $-\infty < \sigma < \Sigma < \infty$. Let $\xi \in (\sigma, \Sigma)$, ℓ a slowly varying function, $c \in \mathbb{R}$. If f is measurable, $f(x)/x^{\sigma}$ is bounded on every interval $(0,a]$ and

$$f(x) \sim cx^{\xi}\ell(x) \text{ as } x \to \infty$$

then

$$\int_0^{\infty} k(x/t)f(t)t^{-1}dt \sim c\check{k}(\xi)x^{\rho}\ell(x) \text{ as } x \to \infty$$

To get the behaviour at 0, we will use the following corollary.

Corollary 4 *Let the Mellin transform $\check{k}$ of k converge at least in the strip $\tau_1 \le \mathrm{Re}(z) \le \tau_2$, where $-\infty < \tau_1 < \tau_2 < \infty$. Let $\xi \in (\tau_1, \tau_2)$, ℓ a slowly varying function, $c \in \mathbb{R}$. If f is measurable, $f(x)x^{\tau_2}$ is bounded on every interval $[a, \infty)$ and*

$$f(x) \sim cx^{\xi}\ell(1/x) \text{ as } x \to 0$$

then

$$\int_0^{\infty} k(x/t)f(t)t^{-1}dt \sim c\check{k}(\xi)x^{\xi}\ell(1/x) \text{ as } x \to 0$$

Proof.

$$\begin{aligned}\int_0^{\infty} k(x/t)f(t)t^{-1}dx &= \int_0^{\infty} k(xu)f(1/u)u^{-1}du \\ &= \int_0^{\infty} \widetilde{k}(1/(xu))\widetilde{f}(u)u^{-1}du\end{aligned}$$

where $\widetilde{f}(x) = f(1/x)$, $\widetilde{f}(x)/x^{-\tau_2}$ bounded on every interval $(0, 1/a]$ with

$$\widetilde{f}(x) = f(1/x) \sim cx^{-\rho}\ell(x) \text{ as } x \to \infty$$

and $\widetilde{k}(x) = k(1/x)$ is such that its Mellin transform converges in the strip $-\tau_2 \le \mathrm{Re}(z) \le -\tau_1$. Theorem 4.1.6 above therefore gives the result.

■

We can now proceed with the proof of Proposition 3

Proof. Let ρ_0 and f_Z, both regularly varying at 0 such that

$$\overline{\rho}_0(x) \sim x^{-\alpha}\ell_1(1/x) \quad (1)$$
$$f_Z(z) \sim \tau z^{\tau-1}\ell_2(1/z) \quad (2)$$

with $\alpha < \tau$. Suppose that ρ_0 and f_Z are bounded on a set of the form $[a,b]$ for $0 < a < b$ and that there exists $\beta > \tau$ such that $\mu_\beta = \int_0^\infty w^\beta \rho_0(w)dw < +\infty$. Let

$$\rho(w) = \int_0^\infty z f_Z(z)\rho_0(wz)dz.$$

Using the change of variables $Y = 1/Z$, we can equivalently write

$$\rho(w) = \int_0^\infty f_Y(y)\rho_0(w/y)y^{-1}dz,$$

with $f_Y(y) = y^{-2}f_Z(1/y)$. From Equation (2), $f_Y(y) \sim \tau y^{-1-\tau}\ell_2(y)$ when $y \to +\infty$. Let $\xi = -1-\tau$, $\sigma = -1-\beta$ and $\Sigma \in (-1-\tau, -1-\alpha)$ (since $\alpha < \tau$). We notice that for any $\delta \in [\sigma, \Sigma]$,

$$\begin{aligned} t^{-\delta-1}\rho_0(t) &= O(t^{-\Sigma-1}\rho_0(t)) \text{ as } t \to 0 \\ t^{-\delta-1}\rho_0(t) &= O(t^{\beta}\rho_0(t)) \text{ as } t \to +\infty \end{aligned}$$

Since ρ_0 is bounded on any set of the form $[a,b]$ and $-\Sigma - 1 - 1 - \alpha > -1$, it comes that $\check{\rho}_0(\delta) < +\infty$. Besides, $f_Y(y)y^{-\sigma} \to 0$ as $y \to 0$. Therefore we can apply the previous theorem from which we deduce that

$$\rho(w) \sim \check{\rho}_0(-1-\tau)w^{-1-\tau}\ell_2(w),$$

which give the required asymptotic behaviour noticing that $\check{\rho}_0(-1-\tau) = \mu_\tau$. For the behaviour at 0, we write

$$\rho(w) = \int \rho_0(w_0)f_Y(w/w_0)w_0^{-1}dw_0,$$

and take $\tau_1 \in (-1-\tau, -1-\alpha)$, $\tau_2 = -1$ and $\xi = -1-\alpha$. Similarly as before, we can show that the conditions of the corollary are satisfied, which gives the expected result. ■

2.6. Proof of Corollary 2

Denote $f_{(k)} = \frac{w_{(k)}}{W(\Theta)}$. From Equation (24) of Theorem 1, and Proposition 23 of Gnedin et al. (2007), we have that almost surely the discrete probability measure $(f_{(k)})_{k\ge1}$ satisfies Equation (17) of (Gnedin et al., 2007) (which is simply an equivalent way of writing the regularly varying property). We conclude by noticing that Corollary 21 of the same paper gives Equation (27).

3. Useful properties

$$\gamma(1,x) = 1 - e^{-x}$$

$$\begin{aligned}\gamma(s,x) &= \int_0^x u^{s-1}e^{-u}du \\ &= x^s \int_0^1 v^{s-1}e^{-vx}dv\end{aligned}$$

$$\gamma(s,x) \sim \frac{x^s}{s}$$

as $x \to 0$. We have

$$\int_{w_0}^{\infty} w^{m-1-\tau}e^{-wt}dw = t^{\tau-m}\Gamma(m-\tau, tw_0)$$

4. Generalized BFRY distribution

The BFRY random variable (Bertoin et al., 2006; Devroye & James, 2014) is a positive random variable W with density

$$f_W(w) = \frac{\alpha}{\Gamma(1-\alpha)} w^{-1-\alpha}(1-e^{-w}), \quad \alpha \in (0,1).$$

W is a heavy tailed random variable with infinite mean, and is known to have a close connection to the stable and generalized gamma processes (Lee et al., 2016). W can be simulated as $W = X/Y$ where $X \sim \mathrm{Gamma}(1-\alpha, 1)$ and $Y \sim \mathrm{Beta}(\alpha, 1)$.

Now let $W = X/Y$, $X \sim \mathrm{Gamma}(\kappa, 1)$ and $Y \sim \mathrm{Beta}(\tau, 1)$, with parameters $\kappa, \tau > 0$. Then the density of W is computed as

$$\begin{aligned} f_W(w) &= \int_0^1 y f_X(wy) f_Y(y) dy \\ &= \frac{\tau}{\Gamma(\kappa)} \int_0^1 y (wy)^{\kappa-1} e^{-wy} y^{\tau-1} dy \\ &= \frac{\tau}{\Gamma(\kappa)} w^{\kappa-1} \int_0^1 y^{\kappa+\tau-1} e^{-wy} dy \\ &= \frac{\tau}{\Gamma(\kappa)} w^{-\tau-1} \gamma(\kappa+\tau, w). \end{aligned} \tag{3}$$

The resulting distribution, which we call as the *generalized BFRY distribution*, contains the BFRY as its special case when $\kappa = 1-\tau \in (0,1)$ and has potentially heavier tail than the BFRY distribution. Like the BFRY distribution has a close connection with the stable and generalized gamma process, the generalized BFRY distribution has a close connection with the generalized BFRY process we described in the main text. Indeed, the generalized BFRY process can be thought as a process version of the generalized BFRY random variable, and the name generalized BFRY process was coined after this connection.

For $m < \tau$, the moments are given by

$$\mathbb{E}(W^m) = \frac{\tau \Gamma(m+\kappa)}{(\tau - m)\Gamma(\kappa)}, \tag{4}$$

and $\mathbb{E}(W^m) = \infty$ for $m \geq \tau$.

5. Additional details on the inference

Here we describe detailed inference procedures for Generalized BFRY process and Beta-prime process.

5.1. Generalized BFRY process

The Lévy density of generalized BFRY process is written as

$$\rho(w) = \frac{1}{\Gamma(1-\sigma)} w^{-1-\tau} \gamma(\tau - \sigma, w), \tag{5}$$

where we fixed $c = 1$. The quantities required for the evaluation of the joint likelihood are

$$\psi(t) = \frac{\eta}{\sigma} \int_0^1 ((y+t)^\sigma - y^\sigma) y^{\tau-\sigma-1} dy \tag{6}$$

$$\kappa(m,t) = \frac{\eta \Gamma(m-\sigma)}{\Gamma(1-\sigma)} \int_0^1 \frac{y^{\tau-\sigma-1}}{(y+t)^{m-\sigma}} dy. \tag{7}$$

As explained in the main text, we introduce a set of latent variables $(Y_j)_{j=1}^{K_n}$ with

$$p(y_j \mid \text{rest}) \propto \frac{y_j^{\tau-\sigma-1}}{(y_j+t)^{m_j-\sigma}} \mathbb{1}_{0<y_j<1}. \tag{8}$$

The joint log-likelihood is then written as

$$\begin{aligned} &p((m_j)_{j=1,\dots,K_n}, y, u | \eta, \sigma, \tau) \propto u^{n-1} e^{-\psi(u)} \\ &\times \prod_{j=1}^{K_n} \frac{\eta \Gamma(m_j - \sigma)}{\Gamma(1-\sigma)} \frac{y_j^{\tau-\sigma-1}}{(y_j+u)^{m_j-\sigma}}. \end{aligned} \tag{9}$$

Since $y_j \in (0,1)$, we take a transformation

$$y_j = \frac{1}{1+e^{-\tilde{y}_j}}, \tag{10}$$

which yields

$$\begin{aligned} &p((m_j)_j, \tilde{y}, u | \eta, \sigma, \tau) \propto u^{n-1} e^{-\psi(u)} \\ &\times \prod_{j=1}^{K_n} \frac{\eta \Gamma(m_j - \sigma)}{\Gamma(1-\sigma)} \frac{y_j^{\tau-\sigma}(1-y_j)}{(y_j+u)^{m_j-\sigma}}. \end{aligned} \tag{11}$$

Sampling $\tilde{y}$ We update $\tilde{y}$ via HMC (Duane et al., 1987; Neal et al., 2011). The gradient of $\log p((m_j)_j, \tilde{y}, u | \eta, \sigma, \tau)$ w.r.t. $\tilde{y}_j$ is given as

$$\left(\frac{\tau - \sigma}{y_j} - \frac{1}{1-y_j} - \frac{m_j - \sigma}{y_j + u} \right) \cdot y_j(1-y_j). \tag{12}$$

For all experiments, we used step size $\epsilon = 0.05$ and number of leapfrog steps $L = 30$.

Sampling u We take a transform $u = e^{\tilde{u}}$ and update $\tilde{u}$ via Metropolis-Hastings with proposal distribution $q(\tilde{u}'|\tilde{u}) = \mathrm{Normal}(\tilde{u}, 0.05)$.

Sampling η We place a prior $\eta \sim \mathrm{Lognormal}(0,1)$, and updated η via Metropolis-Hastings with proposal distribution $q(\hat{\eta}|\eta) = \mathrm{Lognormal}(\log \eta, 0.05)$.

Sampling σ We place a prior $\sigma \sim \mathrm{Logitnormal}(0,1)$, and updated σ via Metropolis-Hastings with proposal distribution $q(\hat{\sigma}|\sigma) = \mathrm{Logitnormal}(\mathrm{logit}(\sigma), 0.05)$.

Sampling τ Since $\tau > \sigma$, instead of directly sampling τ, we sampled $\delta = \tau - \sigma > 0$. Then we place a prior $\delta \sim \mathrm{Lognormal}(0,1)$ and update δ via Metropolis-Hastings with proposal distribution $q(\hat{\delta}|\delta) = \mathrm{Lognormal}(\log\delta, 0.05)$.

5.2. Beta prime process

The Lévy density of Beta prime process is

$$\rho(w) = \frac{\Gamma(\tau-\sigma)}{\Gamma(1-\sigma)} w^{-1-\sigma}(1+w)^{\sigma-\tau}, \tag{13}$$

where we fixed $c = 1$. Then we have

$$\psi(t) = \frac{\eta}{\sigma}\int_0^\infty ((y+t)^\sigma - y^\sigma) y^{\tau-\sigma-1} e^{-y} dy, \tag{14}$$

$$\kappa(m,t) = \frac{\eta\Gamma(m-\sigma)}{\Gamma(1-\sigma)}\int_0^\infty \frac{y^{\tau-\sigma-1}e^{-y}}{(y+t)^{m-\sigma}} dy. \tag{15}$$

As for the generalized BFRY process, we augment the joint likelihood with a set of latent variables $(Y_j)_{j=1}^{K_n}$ with density

$$p(y_j \mid \text{rest}) \propto \frac{y_j^{\tau-\sigma-1}}{(y_j+u)^{m_j-\sigma}} \mathbb{1}_{y_j<0}, \tag{16}$$

which yields

$$\begin{aligned} p((m_j)_j, y, u|\eta,\sigma,\tau) &\propto u^{n-1}e^{-\psi(u)} \\ &\times \prod_{j=1}^{K_n} \frac{\eta\Gamma(m_j-\sigma)}{\Gamma(1-\sigma)} \frac{y_j^{\tau-\sigma-1}}{(y_j+u)^{m_j-\sigma}}. \end{aligned} \tag{17}$$

Since $y_j > 0$, we take a transformation $y_j = e^{\tilde{y}_j}$ to have

$$\begin{aligned} p((m_j)_j, y, u|\eta,\sigma,\tau) &= \frac{u^{n-1}e^{-\psi(u)}}{\Gamma(n)} \\ &\times \prod_{j=1}^{K_n} \frac{\eta\Gamma(m_j-\sigma)}{\Gamma(1-\sigma)} \frac{y_j^{\tau-\sigma}}{(y_j+u)^{m_j-\sigma}}. \end{aligned} \tag{18}$$

Sampling $\tilde{y}$ We update $\tilde{y}$ via HMC. The gradient required for $\tilde{y}$ is computed as

$$\tau - \sigma - y_j - (m_j - \sigma)\frac{y_j}{y_j+u}.$$

Sampling u, η, σ, τ Same as for the generalized BFRY process.

6. Results of experiments

Synthetic data

As explained in the main text, we sample simulated datasets from the GBFRY and the BP models with parameters $\sigma = 0.1$, $\tau = 2$, $c = 1$ and $\eta = 4000$. We run the MCMC algorithm described in Section 4.2 with 100 000 iterations. The 95% credible intervals are $\sigma \in (0.09, 0.12)$, $\tau \in (1.6, 2.2)$ for the BFRY and $\sigma \in (0.08, 0.11)$, $\tau \in (1.8, 2.3)$ for the BP model. The MCMC algorithm is therefore able to recover the true parameters. Trace plots are reported in Figures 1 and 2.

Real data

Here we report the results for the 5 datasets described in the main text. We report the 95% credible intervals of the posterior predictive for the proportion of occurrences and ranked frequencies of the Generalized BFRY, BP, normalized GGP and PY models for each dataset in Figures 3 to 12. We can see that as predicted the GGP and PY do not manage to capture the behavior of the large clusters (which are on the right of the figures displaying the proportion of clusters of a given size, and on the left on the figures displaying the ordered sizes of the clusters).

References

Bertoin, J., Fujita, T., Roynette, B., and Yor, M. On a particular class of self-decomposable random variables: the durations of Bessel excursions straddling independent exponential times. 2006.

Bingham, N. H., Goldie, C. M., and Teugels, J. L. *Regular variation*, volume 27. Cambridge university press, 1989.

Devroye, L. and James, L. On simulation and properties of the stable law. *Statistical methods & applications*, 23(3): 307–343, 2014.

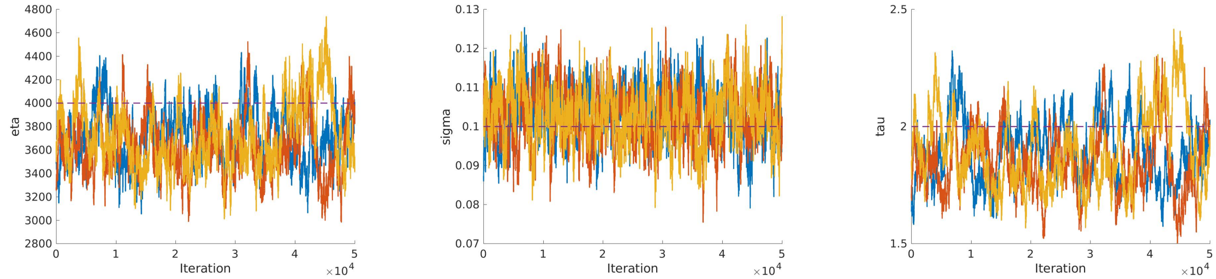
Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. Hybrid monte carlo. *Physics letters B*, 195(2):216–222, 1987.

Gnedin, A., Hansen, B., and Pitman, J. Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Probability surveys*, 4:146–171, 2007.

Kevei, P. and Mason, D. M. The limit distribution of ratios of jumps and sums of jumps of subordinators. *ALEA*, 11 (2):631–642, 2014.

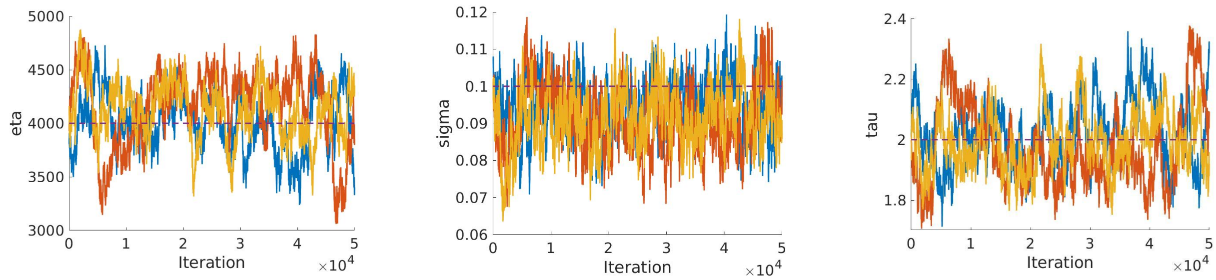
Lee, J., James, L. F., and Choi, S. Finite-dimensional BFRY priors and variational Bayesian inference for power law models. In *Advances in Neural Information Processing Systems*, pp. 3162–3170, 2016.

Neal, R. M. et al. Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2(11):2, 2011.



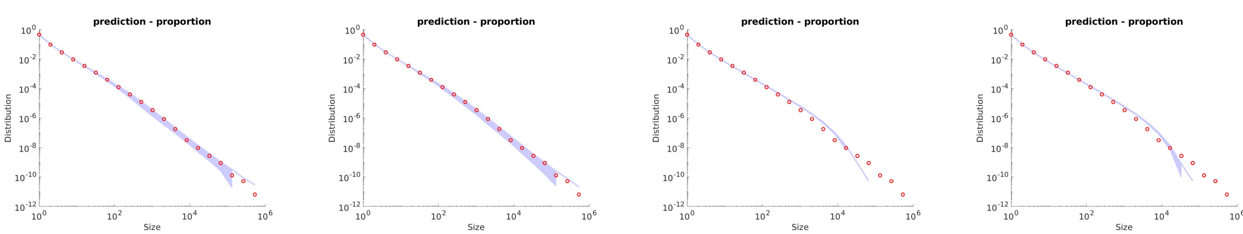
(a) Posterior η (b) Posterior σ (c) Posterior τ

Figure 1. Trace plots of the parameter samples for the Generalized BFRY model. Dashed line represents true value of the parameter.



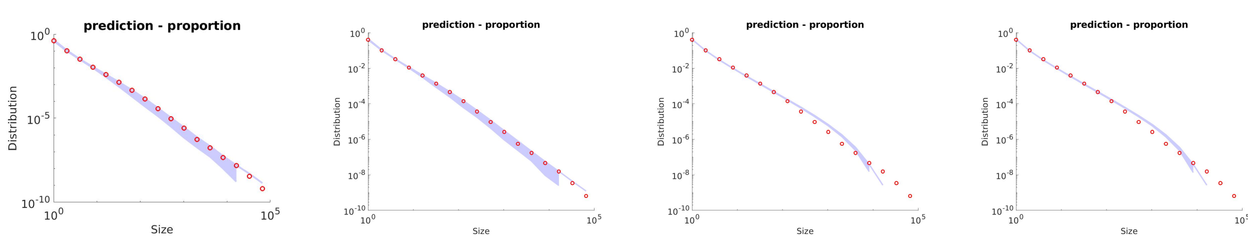
(a) Posterior η (b) Posterior σ (c) Posterior τ

Figure 2. Trace plots of the parameter samples for the Beta prime process model. Dashed line represents true value of the parameter.



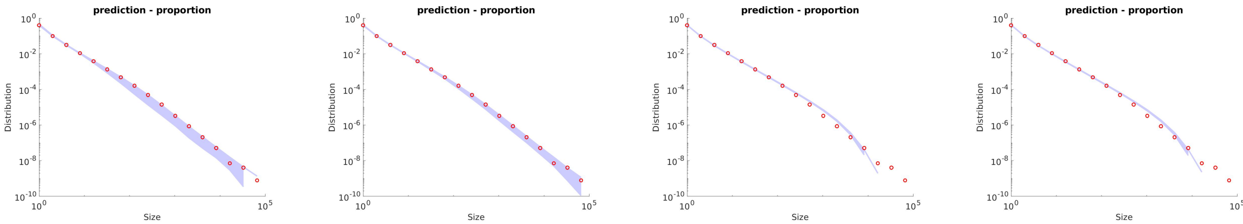
(a) Generalized BFRY (b) Beta prime (c) GGP (d) PY

Figure 3. Proportion of clusters of a given size in the ANC dataset: 95% credible interval of the posterior predictive in blue, real values in red



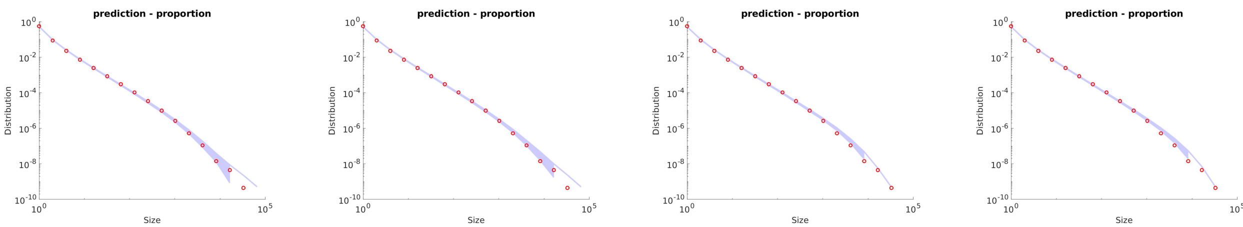
(a) Generalized BFRY (b) Beta prime (c) GGP (d) PY

Figure 4. Proportion of clusters of a given size in the English books dataset: 95% credible interval of the posterior predictive in blue, real values in red



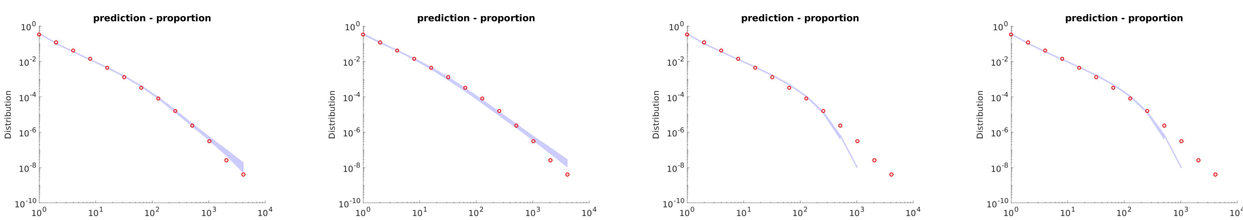
(a) Generalized BFRY (b) Beta prime (c) GGP (d) PY

Figure 5. Proportion of clusters of a given size in the French books dataset: 95% credible interval of the posterior predictive in blue, real values in red



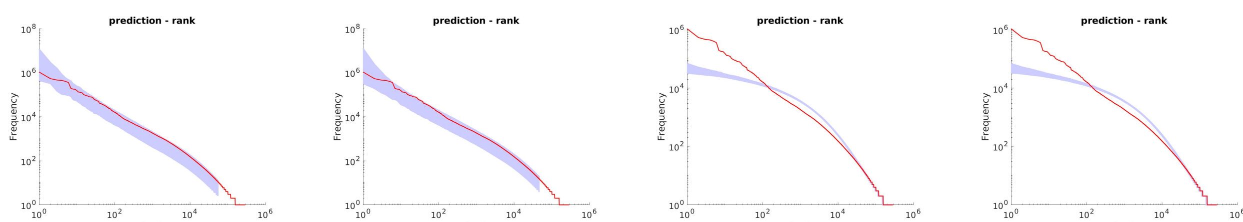
(a) Generalized BFRY (b) Beta prime (c) GGP (d) PY

Figure 6. Proportion of clusters of a given size in the nips dataset: 95% credible interval of the posterior predictive in blue, real values in red



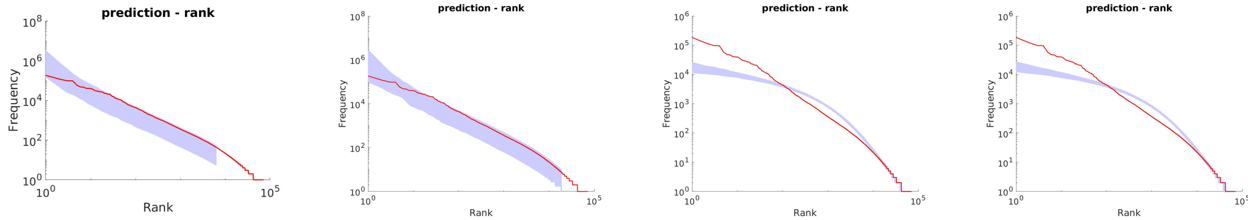
(a) Generalized BFRY (b) Beta prime (c) GGP (d) PY

Figure 7. Proportion of nodes with a given degree in the Twitter dataset: 95% credible interval of the posterior predictive in blue, real values in red



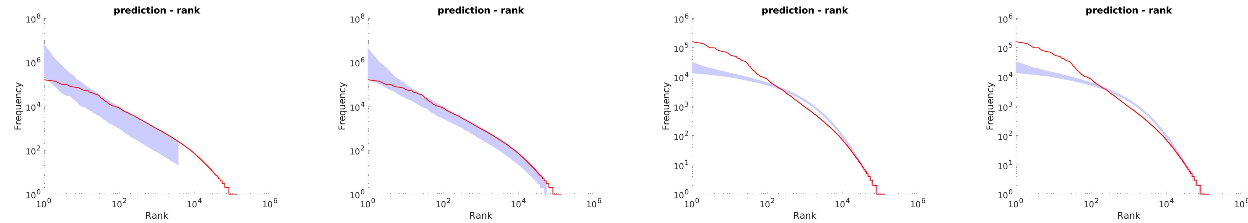
(a) Generalized BFRY (b) Beta prime (c) GGP (d) PY

Figure 8. Ordered size of the clusters in the ANC dataset: 95% credible interval of the posterior predictive in blue, real values in red



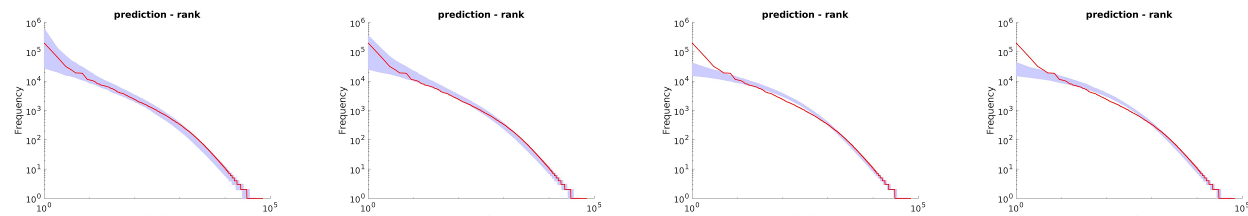
(a) Generalized BFRY (b) Beta prime (c) GGP (d) PY

Figure 9. Ordered size of the clusters in the English books dataset: 95% credible interval of the posterior predictive in blue, real values in red



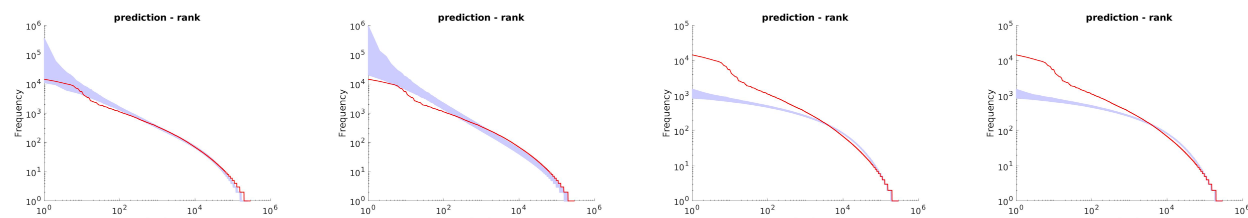
(a) Generalized BFRY (b) Beta prime (c) GGP (d) PY

Figure 10. Ordered size of the clusters in the French books dataset: 95% credible interval of the posterior predictive in blue, real values in red



(a) Generalized BFRY (b) Beta prime (c) GGP (d) PY

Figure 11. Ordered size of the clusters in the nips dataset: 95% credible interval of the posterior predictive in blue, real values in red



(a) Generalized BFRY (b) Beta prime (c) GGP (d) PY

Figure 12. Ordered degrees of the nodes in the Twitter dataset: 95% credible interval of the posterior predictive in blue, real values in red