A The CMS

For any $m \geq 1$ let $X_{1:m} = (X_1, \ldots, X_m)$ be a data stream of tokens taking values in a measurable space of symbols $\mathcal{V}$. A point query over $X_{1:m}$ asks for the estimation of the frequency f_v of a token of type $v \in \mathcal{V}$ in $X_{1:m}$, i.e. $f_v = \sum_{1\leq i\leq m} \mathbf{1}_{X_i}(v)$. The goal of CMS of (Cormode and Muthukrishnan, 2005b,a) consists in estimating f_v based on a compressed representation of $X_{1:m}$ by random hashing. In particular, let J and N be positive integers such that $[J] = \{1, \ldots, J\}$ and $[N] = \{1, \ldots, N\}$, and let $h_1, \ldots, h_N$, with $h_n : \mathcal{V} \to [J]$, be a collection of hash functions drawn uniformly at random from a pairwise independent hash family $\mathcal{H}$. That is, a random hash function $h \in \mathcal{H}$ has the property that for all $v_1, v_2 \in \mathcal{H}$ such that $v_1 \neq v_2$, the probability that v_1 and v_2 hash to values $j_1, j_2 \in [J]$, respectively, is

$$\Pr[h(v_1) = j_1, h(v_2) = j_2] = \frac{1}{J^2}.$$

Hashing $X_{1:m}$ through $h_1, \ldots, h_N$ creates N vectors of J buckets $\{(C_{n,1}, \ldots, C_{n,J})\}_{n\in[N]}$, with $C_{n,j}$ obtained by aggregating the frequencies for all x where $h_n(x) = j$. Every $C_{n,j}$ is initialized at zero, and whenever a new token X_i is observed we set $C_{n,h_n(X_i)} \leftarrow 1 + C_{n,h_n(X_i)}$ for every $n \in [N]$. After m tokens, $C_{n,j} = \sum_{1\leq i\leq m} \mathbf{1}_{h_n(X_i)}(j)$ and $f_v \leq C_{n,j}$ for any $v \in \mathcal{V}$. Under this setting, the CMS of (Cormode and Muthukrishnan, 2005a) estimates f_v with the smallest hashed frequency among $\{C_{n,h_n(v)}\}_{n\in[N]}$, i.e.,

$$\hat{f}_v^{(\mathrm{CMS})} = \min_{n\in[N]} \{C_{n,h_n(v)}\}_{n\in[N]}.$$

That is, $\hat{f}_v^{(\mathrm{CMS})}$ returns the count associated with the fewest collisions. This provides an upper bound on the true count. For an arbitrary data stream with m tokens, the CMS satisfies the following guarantee.

Theorem 1. *(Cormode and Muthukrishnan, 2005a) Let $J = \lceil e/2 \rceil$ and let $N = \lceil \log 1/\delta \rceil$, with $\varepsilon > 0$ and $\delta > 0$. Then, the estimate $\hat{f}_v^{(CMS)}$ satisfies $\hat{f}_v^{(CMS)} \geq f_v$ and, with probability at least $1 - \delta$, the estimate $\hat{f}_v^{(CMS)}$ satisfies $\hat{f}_v^{(CMS)} \leq f_v + \varepsilon m$.*

B CRMs and hNCRMs

Let $\mathcal{V}$ be a measurable space endowed with its Borel σ-field $\mathcal{F}$. A CRM μ on $\mathcal{V}$ is defined as a random measure such that for any $A_1, \ldots, A_k$ in $\mathcal{F}$, with $A_i \cap A_j = \emptyset$ for $i \neq j$, the random variables $\mu(A_1), \ldots, \mu(A_k)$ are mutually independent (Kingman, 1993). Any CRM μ with no fixed point of discontinuity and no deterministic drift is represented as $\mu = \sum_{j\geq 1} \xi_j \delta_{v_j}$, where the ξ_j's are positive random jumps and the v_j's are $\mathcal{V}$-valued random locations. Then, μ is characterized by the Lévy–Khintchine representation

$$\mathbb{E}\left[\exp\left\{-\int_{\mathcal{V}} f(v)\mu(\mathrm{d}v)\right\}\right] = \exp\left\{-\int_{\mathbb{R}^+\times\mathcal{V}} [1 - \mathrm{e}^{-\xi f(v)}]\right\} \rho(\mathrm{d}\xi, \mathrm{d}v),$$

where $f : \mathcal{V} \to \mathbb{R}$ is a measurable function such that $\int |f| \mathrm{d}\mu < +\infty$ and ρ is a measure on $\mathbb{R}^+ \times \mathcal{V}$ such that $\int_B \int_{\mathbb{R}^+} \min\{\xi, 1\} \rho(\mathrm{d}\xi, \mathrm{d}v) < +\infty$ for any $B \in \mathcal{F}$. The measure ρ, referred to as Lévy intensity measure, characterizes μ: it contains all the information on the distributions of jumps and locations of μ. For our purposes it will often be useful to separate the jump and location part of ρ by writing it as

$$\gamma(\mathrm{d}\xi, \mathrm{d}v) = \rho(\mathrm{d}\xi; v)\nu(\mathrm{d}v),$$

where ν denotes a measure on $(\mathcal{V}, \mathcal{F})$ and ρ denotes a transition kernel on $\mathcal{B}(\mathbb{R}^+) \times \mathcal{V}$, with $\mathcal{B}(\mathbb{R}^+)$ being the Borel σ-field of $\mathbb{R}^+$, i.e. $v \mapsto \rho(A; v)$ is $\mathcal{F}$-measurable for any $A \in \mathcal{B}(\mathbb{R}^+)$ and $\rho(\cdot; v)$ is a measure on $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ for any $v \in \mathcal{V}$. In particular, if $\rho(\cdot; v) = \rho(\cdot)$ for any v then the jumps of μ is independent of their locations and γ and μ are termed homogeneous. See (Kingman, 1993) and references therein.

CRMs are closely connected to Poisson processes. Indeed μ can be represented as a linear functional of a Poisson process Π on $\mathbb{R}^+ \times \mathcal{V}$ with mean measure γ. To stated this precisely, Π is a random subset of $\mathbb{R}^+ \times \mathcal{V}$ and if $N(A) = \mathrm{card}\{\Pi \cap A\}$ for any $A \subset \mathcal{B}(\mathbb{R}^+) \otimes \mathcal{F}$ such that $\gamma(A) < +\infty$, then

$$\Pr[N(A) = k] = \mathrm{e}^{-\gamma(A)} \frac{(\gamma(A))^k}{k!}$$

for $k \geq 0$. Then, for any $A \in \mathcal{F}$

$$\mu(A) = \int_A \int_{\mathbb{R}^+} N(\mathrm{d}v, \mathrm{d}\xi)$$

See (Kingman, 1993) and references therein. An important property of CRMs is their almost sure discreteness (Kingman, 1993), which means that their realizations are discrete measures with probability 1. This fact essentially entails discreteness of random probability measures obtained as transformations of CRMs, such as hNCRMs.

hNCRMs (James, 2002; Prünster, 2002; Regazzini et al., 2003; Pitman, 2006; Lijoi and Prünster, 2010) are defined in terms of a suitable normalization of CRMs. Let μ be a homogeneous CRM on $\mathcal{V}$ such that $0 < \mu(\mathcal{V}) < +\infty$ almost surely. Then, the random probability measure

$$P = \frac{\mu}{\mu(\mathcal{V})} \tag{1}$$

is termed hNCRM. Because of the almost sure discreteness of μ, the P is discrete almost surely. That is,

$$P = \sum_{j \geq 1} p_j \delta_{v_j},$$

where $p_j = \xi_j / \mu(\mathcal{V})$ for $j \geq 1$ are random probabilities such that $p_j \in (0, 1)$ for any $j \geq 1$ and $\sum_{j \geq 1} p_j = 1$ almost surely. Both the conditions of finiteness and positiveness of $\mu(\mathcal{V})$ are clearly required for the normalization (1) to be well-defined, and it is natural to express these conditions in terms of the Lévy intensity measure γ of the CRM μ. It is enough to have $\rho = +\infty$ and $0 < \mu(\mathcal{V}) < +\infty$. In particular, the former is equivalent to requiring that μ has infinitely many jumps on any bounded set: in this case μ is also called an infinite activity process. The previous conditions can also be strengthened to necessary and sufficient conditions but we do not pursue this here. See (Kingman, 1993).

C NGGP priors, and proof of Proposition 1

Let $\mathcal{V}$ be a measurable space endowed with its Borel σ-field $\mathcal{F}$. For any $m \geq 1$, let $X_{1:m}$ be a random sample of tokens from $P \sim \mathrm{NGGP}(\alpha, \sigma, \nu)$. Because of the discreteness of P, the random sample $X_{1:m}$ induces a random partition of the set $\{1, \dots, m\}$ into $K_m = k \leq m$ partition subsets, labelled by distinct symbols $\mathbf{v} = \{v_1, \dots, v_{K_m}\}$ in $\mathcal{V}$, with frequencies $\mathbf{N}_n = (N_1, \dots, N_{K_m}) = (n_1, \dots, n_k)$ such that $N_i > 0$ and $\sum_{1 \leq i \leq K_m} N_i = m$. Distributional properties of the random partition induced by $X_{1:m}$ induced by $X_{1:m}$ have been investigated in, e.g., (James, 2002), Pitman (2003), (Lijoi et al., 2007), (De Blasi et al., 2013) and (Bacallado et al., 2017). In particular,

$$\Pr[K_m = k, \mathbf{N}_m = (n_1, \dots, n_k)] = \frac{1}{k!} \binom{m}{n_1, \dots, n_k} V_{m,k} \prod_{i=1}^{k} (1 - \sigma)_{(n_i - 1)}, \tag{2}$$

where

$$V_{m,k} = \frac{(\alpha 2^{\sigma - 1})^k}{\Gamma(m)} \int_0^{+\infty} \frac{x^{m-1}}{(2^{-1} + x)^{m - k\sigma}} \exp\left\{ -\frac{\alpha 2^{\sigma - 1}}{\sigma} [(2^{-1} + x)^{\sigma} - 2^{-\sigma}] \right\} \mathrm{d}x. \tag{3}$$

Now, let $\mathcal{P}_{m,k} = \{(n_1, \dots, n_k) : n_i \geq 0 \text{ and } \sum_{1 \leq i \leq k} n_i = m\}$ denote the set of partitions of m into $k \leq m$ blocks. Then, the distribution of K_m follows my marginalizing (2) on the set $\mathcal{P}_{m,k}$, that is

$$\begin{aligned} \Pr[K_m = k] &= \sum_{(n_1, \dots, n_k) \in \mathcal{P}_{m,k}} \frac{1}{k!} \binom{m}{n_1, \dots, n_k} V_{m,k} \prod_{i=1}^{k} (1 - \sigma)_{(n_i - 1)} \\ &= \frac{V_{m,k}}{\sigma^k} C(m, k; \sigma), \end{aligned} \tag{4}$$

where $C(m, k; \sigma)$ denotes the (central) generalized factorial coefficient (Charalambides, 2005), which is defined as $C(m, k; \sigma) = (k!)^{-1} \sum_{1 \leq i \leq k} \binom{k}{i} (-1)^i (i\sigma)_{(m)}$, with the proviso $C(0, 0; \sigma) = 1$ and $C(m, 0; \sigma) = 0$ for any $m \geq 1$. For any $1 \leq r \leq m$, let $M_{r,m} \geq 0$ denote the number of distinct symbols with frequency r in $X_{1:m}$, i.e. $M_{r,m} = \sum_{1 \leq i \leq K_m} \mathbf{1}_{N_i}(r)$ such that $\sum_{1 \leq r \leq m} M_{r,m} = K_m$ and $\sum_{1 \leq r \leq m} r M_{r,m} = m$. Then, the distribution of $\mathbf{M}_m = (M_{1,m}, \dots, M_{m,m})$ follows directly form (2), i.e.

$$\Pr[\mathbf{M}_m = \mathbf{m}] = V_{m,k} m! \prod_{i=1}^{m} \left(\frac{(1 - \sigma)_{(i-1)}}{i!} \right)^{m_i} \frac{1}{m_i!} \mathbf{1}_{\mathcal{M}_{m,k}}(\mathbf{m}), \tag{5}$$

where $\mathcal{M}_{m,k} = \{(m_1, \ldots, m_n) : m_i \geq 0, \text{ and } \sum_{1 \leq i \leq m} m_i = k, \sum_{1 \leq i \leq m} i m_i = m\}$. The distribution (5) is the referred to as the sampling formula of the random partition with distribution (2).

For any $m \geq 1$, let $X_{1:m}$ be a random sample from $P \sim \text{NGGP}(\alpha, \sigma, \nu)$ featuring $K_m = k$ partition subsets, labelled by distinct symbols $\mathbf{v} = \{v_1, \ldots, v_{K_m}\}$ in $\mathcal{V}$, with frequencies $\mathbf{N}_n = (n_1, \ldots, n_k)$. The predictive distributions of P provides the conditional distribution of X_{m+1} given $X_{1:m}$. That is, for $A \in \mathcal{F}$

$$\Pr[X_{m+1} \in A \,|\, X_{1:m}] = \frac{V_{m+1,k+1}}{V_{m,k}} \nu(A) + \frac{V_{m+1,k}}{V_{m,k}} \sum_{i=1}^{k} (n_i - \sigma) \delta_{v_i}(A) \tag{6}$$

for any $m \geq 1$. We refer to Bacallado et al. (2017) for a characterization of (6) in terms of a meaningful Pólya like urn scheme. The predictive distributions (6) provides the fundamental ingredient of the proof of Proposition 1.

Proof of Proposition 1. The proof follows from the predictive distributions (6) by setting $A = \mathbf{v}_0$ and $A = \mathbf{v}_r$. $\square$

We conclude by showing that the distributional property of a random sample from $P \sim \text{DP}(\alpha, \nu)$ follows from the distributional property of a random sample from $P \sim \text{NGGP}(\alpha, \sigma, \nu)$ by letting $\sigma \to 0$. For any $m \geq 1$, let $X_{1:m}$ be a random sample from $P \sim \text{DP}(\alpha/2, \nu)$ featuring $K_m = k$ partition subsets, labelled by distinct symbols $\mathbf{v} = \{v_1, \ldots, v_{K_m}\}$ in $\mathcal{V}$, with frequencies $\mathbf{N}_n = (n_1, \ldots, n_k)$. The distribution of the random partition induced by $X_{1:m}$ follows from (2) by letting $\sigma \to 0$. Indeed,

$$\begin{aligned}
\lim_{\sigma \to 0} V_{m,k} &= \lim_{\sigma \to +0} \frac{(\alpha 2^{\sigma - 1})^k}{\Gamma(m)} \int_0^{+\infty} \frac{x^{m-1}}{(2^{-1} + x)^{m - k\sigma}} \exp\left\{ -\frac{\alpha 2^{\sigma-1}}{\sigma} [(2^{-1} + x)^{\sigma} - 2^{-\sigma}] \right\} \mathrm{d}x \\
&= \frac{(\alpha/2)^k}{\Gamma(m)} 2^{-\alpha/2} \int_0^{+\infty} \frac{x^{m-1}}{(2^{-1} + x)^{m + \alpha/2}} \mathrm{d}x \\
&= \frac{\left(\frac{\alpha}{2}\right)^k}{\left(\frac{\alpha}{2}\right)_{(m)}}.
\end{aligned} \tag{7}$$

Therefore, by combining the distribution (2) with (7), and letting $\sigma \to 0$, it follows directly the distribution of the random partition induced by a random sample $X_{1:m}$ from $P \sim \text{DP}(\alpha/2, \nu)$. That is,

$$\Pr[K_m = k, \mathbf{N}_m = (n_1, \ldots, n_k)] = \frac{1}{k!} \binom{m}{n_1, \ldots, n_k} \frac{\left(\frac{\alpha}{2}\right)^k}{\left(\frac{\alpha}{2}\right)_{(m)}} \prod_{i=1}^{k} (n_i - 1)!\,.$$

The distribution of K_m follows by combining the distribution (4) with (7), and from the fact that $\lim_{\sigma \to 0} \sigma^{-k} C(m, k; \sigma) = |s(m, k)|$, where $|s(m, k)|$ denotes the signless Stirling number of the first type (Charalambides, 2005). That is,

$$\Pr[K_m = k] = \frac{\left(\frac{\alpha}{2}\right)^k}{\left(\frac{\alpha}{2}\right)_{(m)}} |s(m, k)|.$$

In a similar manner, the distribution of $\mathbf{M}_m$ under the DP prior, which is referred to as Ewens sampling formula Ewens (1972), follows by combining the sampling formula (5) with (7), and letting $\sigma \to 0$.

Finally, the predictive distributions of $P \sim \text{DP}(\alpha, \nu)$. For any $m \geq 1$, let $X_{1:m}$ be a random sample from $P \sim \text{DP}(\alpha/2, \nu)$ featuring $K_m = k$ partition subsets, labelled by distinct symbols $\mathbf{v} = \{v_1, \ldots, v_{K_m}\}$ in $\mathcal{V}$, with frequencies $\mathbf{N}_n = (n_1, \ldots, n_k)$. The predictive distributions of P follows by combining the predictive distributions (6) with (7), and letting $\sigma \to 0$. That is, for $A \in \mathcal{F}$

$$\Pr[X_{m+1} \in A, \,|\, X_{1:m}] = \frac{\frac{\alpha}{2}}{\frac{\alpha}{2} + m} \nu(A) + \frac{1}{\frac{\alpha}{2} + m} \sum_{i=1}^{k} n_i \delta_{v_i}(A) \tag{8}$$

for any $m \geq 1$. The predictive distributions (8) is at the basis of the CMS-DP proposed in Cai et al. (2018). In particular, Equation 4 in Cai et al. (2018) follows from the predictive distributions (8) by setting $A = \mathbf{v}_0$ and $A = \mathbf{v}_r$.

D The NIGP prior

For $\sigma = 1/2$ the NGGP prior reduces to the NIGP prior (Prünster, 2002; Lijoi et al., 2005). Al alternative definition of the NIGP prior is given through its family of finite-dimensional distributions. This alternative definition relies on the IG distribution (Seshadri, 1993). In particular, a random variable W has IG distribution with shape parameter $a \geq 0$ and scale parameter $b \geq 0$ if it has the density function, with respect to the Lebesgue measure, given by

$$f_W(w; a, b) = \frac{a\mathrm{e}^{ab}}{\sqrt{2\pi}} w^{-\frac{3}{2}} \exp\left\{-\frac{1}{2}\left(\frac{a^2}{w} + b^2 w\right)\right\} \mathbf{1}_{\mathbb{R}^+}(w).$$

Let $(W_1, \ldots, W_k)$ be a collection of independent random variables such that W_i is distributed according to the IG distribution with shape parameter a_i and scale parameter 1, for $i = 1, \ldots, k$. The normalized IG distribution with parameter $(a_1, \ldots, a_k)$ is the distribution of the following random variable

$$(P_1, \ldots, P_k) = \left(\frac{W_1}{\sum_{i=1}^k W_i}, \ldots, \frac{W_k}{\sum_{i=1}^k W_i}\right).$$

The distribution of the random variable $(P_1, \ldots, P_{k-1})$ is absolutely continuous with respect to the Lebesgue measure on R^{k-1}, and its density function on the $(k-1)$-dimensional simplex coincides with

$$\begin{aligned} &f_{(P_1,\ldots,P_{k-1})}(p_1, \ldots, p_{k-1}; a_1, \ldots, a_k) \qquad (9)\\ &= \left(\prod_{i=1}^k \frac{a_i \mathrm{e}^{a_i}}{\sqrt{2\pi}}\right) \prod_{i=1}^{k-1} p_i^{-3/2} \left(1 - \sum_{i=1}^{k-1} p_i\right)^{-3/2} \\ &\times 2\left(\sum_{i=1}^{k-1} \frac{a_i^2}{p_i} + \frac{a_k^2}{1 - \sum_{i=1}^{k-1} p_i}\right)^{-k/4} K_{-k/2}\left(\sqrt{\sum_{i=1}^{k-1} \frac{a_i^2}{p_i} + \frac{a_k^2}{1 - \sum_{i=1}^{k-1} p_i}}\right), \end{aligned}$$

where $K_{-k/2}$ denotes the modified Bessel function of the second type, or Macdonald function, with parameter $-k/2$. If the random variable $(P_1, \ldots, P_k)$ is distributed according to a normalized IG distribution with parameter $(a_1, \ldots, a_k)$, and if $m_1 < m_2 < \cdots < m_r < k$ are positive integers, then

$$\left(\sum_{i=1}^{m_1} P_i, \sum_{i=m_1+1}^{m_2} P_i, \ldots, \sum_{i=m_{r-1}+1}^{k} W_i\right)$$

is a random variable distributed as a normalized inverse Gaussian distribution with parameter $(\sum_{1\leq i\leq m_1} a_i, \sum_{m_1+1\leq i\leq m_2} a_i, \ldots, \sum_{m_{r-1}+1\leq i\leq k}^{k} W_i)$. This projective property of the normalized inverse Gaussian distribution follows from the additive property of the inverse Gaussian distribution (Seshadri, 1993).

To define the NIGP prior through its family of finite-dimensional distributions, let $\mathcal{V}$ be a measurable space endowed with its Borel σ-field $\mathcal{F}$. Let $\mathcal{P} = \{Q_{B_1,\ldots,B_k} : B_1, \ldots, B_k \in \mathcal{F} \text{ for } k \geq 1\}$ be a family of probability distributions, and let $\tilde{\nu} = \alpha\nu$ be a diffuse (base) measure on $\mathcal{V}$ with $\tilde{\nu}(\mathcal{V}) = \alpha$. If $\{B_1, \ldots, B_k\}$ denotes a measurable k-partition of $\mathcal{V}$ and Δ_{k-1} is the $(k-1)$-dimensional simplex, then set

$$Q_{B_1,\ldots,B_k}(C) = \int_{C\cap\Delta_{k-1}} f_{(P_1,\ldots,P_{k-1})}(p_1, \ldots, p_{k-1}; a_1, \ldots, a_k)\mathrm{d}p_1 \cdots \mathrm{d}p_{k-1}$$

for any C in the Borel σ-field of $\mathbb{R}^k$, where $f_{(P_1,\ldots,P_{k-1})}$ is the normalized IG distribution with density function (9) with $a_i = \tilde{\nu}(B_i)$, for $i = 1, \ldots, k$. According to Proposition 3.9.2 of Regazzini (2001), the NIGP is the unique random probability measure admitting $\mathcal{P}$ as its family of finite-dimensional distributions.

The projective property of $P \sim \text{NIGP}(\alpha, \nu)$ follows directly from: i) the definition of P through its family of finite-dimensional distributions; ii) the projective property of the normalized IG distribution. In particular, for any finite family of sets $\{A_1, \ldots, A_k\}$ in $\mathcal{F}$, let $\{B_1, \ldots, B_h\}$ be a measurable h-partition of $\mathcal{V}$ such that it is finer then the partition generated by the family of sets $\{A_1, \ldots, A_k\}$. Then,

$$Q_{A_1,\ldots,A_k}(C) = Q_{B_1,\ldots,B_h}(C')$$

for any C in the Borel σ-field of $\mathbb{R}^k$, with $C' = \{(x_1, \ldots, x_h) \in [0, 1]^h : (\sum_i x_i, \ldots, \sum_i x_i) \in C\}$. See (Lijoi et al., 2005).

E Proof of Proposition 2, and proof of Theorem 3

To prove Proposition 2, we start with the following lemma under the assumption that $X_{1:m}$ is a random sample from $P \sim \text{NGGP}(\alpha, \sigma, \nu)$. The proof of Proposition 2 then follows by setting $\sigma = 1/2$. Let

$$p_{f_v}(\ell; m, \alpha, \sigma) = \sum_{\mathbf{m} \in \mathcal{M}_{k,m}} \Pr[X_{m+1} \in \mathbf{v}_\ell \mid \mathbf{M}_m = \mathbf{m}] \Pr[\mathbf{M}_m = \mathbf{m}], \quad \ell = 0, 1, \dots, m,$$

where the predictive distributions $\Pr[X_{m+1} \in \mathbf{v}_\ell \mid \mathbf{M}_m = \mathbf{m}]$ are displayed in Equation 5, and the distribution $\Pr[\mathbf{M}_m = \mathbf{m}]$ is displayed in Equation 4. For $\sigma \in (0, 1)$, let f_σ denote the density function of the positive σ-stable random variable X_σ, i.e. $\mathbb{E}[\exp\{-tX_\sigma\}] = \exp\{-t^\sigma\}$ for any $t > 0$.

Lemma 1. *For any $m \geq 1$, let $X_{1:m}$ be a random sample from $P \sim NGGP(\alpha, \sigma, \nu)$. Then, for $\ell = 0, 1, \dots, m$*

$$p_{f_v}(\ell; m, \alpha, \sigma) \begin{cases} \frac{\sigma(\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)}}{\Gamma(1-\sigma+\ell)} \\ \quad \times \int_0^{+\infty} \int_0^1 \frac{1}{h^\sigma} f_\sigma(hp) e^{-h\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}} p^{m-\ell}(1-p)^{1-\sigma+\ell-1} \mathrm{d}p\mathrm{d}h & \ell < m \\ \frac{\alpha 2^m (1-\sigma)_{(m)}}{\Gamma(m+1)} \\ \quad \times \int_0^{+\infty} \frac{x^m}{(1+2x)^{m+1-\sigma}} \exp\left\{-\frac{\alpha 2^{\sigma-1}}{\sigma}[(2^{-1}+x)^\sigma - 2^{-\sigma}]\right\} \mathrm{d}x & \ell = m. \end{cases} \tag{10}$$

Proof. We start by considering the case $\ell = 0$. The probability $p_{f_v}(0; m, \alpha, \sigma)$ follows by combining Proposition 1 with the distribution of K_m displayed in (4). Indeed, we can write the following expression

$$\begin{aligned} p_{f_v}(0; m, \alpha, \sigma) &= \sum_{\mathbf{m} \in \mathcal{M}_{m,k}} \Pr[X_{m+1} \in \mathbf{v}_0 \mid \mathbf{M}_m = \mathbf{m}] \Pr[\mathbf{M}_m = \mathbf{m}] & (11) \\ &= \sum_{\mathbf{m} \in \mathcal{M}_{m,k}} \frac{V_{m+1,k+1}}{V_{m,k}} \Pr[\mathbf{M}_m = \mathbf{m}] \\ &= \sum_{k=1}^m \frac{V_{m+1,k+1}}{\sigma^k} C(m, k; \sigma). & (12) \end{aligned}$$

Then, the expression of $p_{f_v}(0; m, \alpha, \sigma)$ in (10) follows by combining (11) with $V_{m+1,k+1}$ displayed in (3), i.e.,

$$\begin{aligned} p_{f_v}(0; m, \alpha, \sigma) &= \frac{(\alpha 2^{\sigma-1})}{\Gamma(m+1)} \int_0^{+\infty} \frac{u^m}{(2^{-1}+u)^{m+1-\sigma}} \exp\left\{-\frac{\alpha 2^{\sigma-1}}{\sigma}[(2^{-1}+u)^\sigma - 2^{-\sigma}]\right\} \\ &\quad \times \sum_{k=1}^m \left(\frac{\alpha 2^{\sigma-1}}{\sigma(2^{-1}+u)^{-\sigma}}\right)^k C(m, k; \sigma) \mathrm{d}u \end{aligned}$$

[Equation 13 of Favaro et al. (2015)]

$$\begin{aligned} &= \frac{(\alpha 2^{\sigma-1})}{\Gamma(m+1)} \int_0^{+\infty} \frac{u^m}{(2^{-1}+u)^{m+1-\sigma}} \exp\left\{-\frac{\alpha 2^{\sigma-1}}{\sigma}[(2^{-1}+u)^\sigma - 2^{-\sigma}]\right\} \\ &\quad \times \exp\left\{\frac{\alpha 2^{\sigma-1}}{\sigma(2^{-1}+u)^{-\sigma}}\right\} \left(\frac{\alpha 2^{\sigma-1}}{\sigma(2^{-1}+u)^{-\sigma}}\right)^{m/\sigma} \int_0^{+\infty} x^m \exp\left\{-x\left(\frac{\alpha 2^{\sigma-1}}{\sigma(2^{-1}+u)^{-\sigma}}\right)^{1/\sigma}\right\} f_\sigma(x)\mathrm{d}x\mathrm{d}u \end{aligned}$$

[Identity $(2^{-1}+u)^{-1+\sigma} = \frac{1}{\Gamma(1-\sigma)} \int_0^{+\infty} y^{1-\sigma-1} \exp\{-y(2^{-1}+u)\} \mathrm{d}y$]

$$\begin{aligned} &= \frac{(\alpha 2^{\sigma-1})^{1+m/\sigma}}{\sigma^{m/\sigma}\Gamma(m+1)} \\ &\quad \times \int_0^{+\infty} u^m \left(\frac{1}{\Gamma(1-\sigma)} \int_0^{+\infty} y^{1-\sigma-1} \exp\{-y(2^{-1}+u)\} \mathrm{d}y\right) \\ &\quad \times \left(\int_0^{+\infty} x^m \exp\left\{-x\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} u\right\} \exp\left\{-x\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\} f_\sigma(x)\mathrm{d}x\right) \mathrm{d}u \end{aligned}$$

$$
\begin{aligned}
&= \frac{(\alpha 2^{\sigma-1})^{1+m/\sigma}}{\sigma^{m/\sigma}\Gamma(1-\sigma)} \\
&\quad \times \int_0^{+\infty} x^m f_\sigma(x) \exp\left\{-x\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\} \\
&\qquad \times \int_0^{+\infty} y^{1-\sigma-1} \exp\left\{-y 2^{-1}\right\} \left[x\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} + y\right]^{-n-1} \mathrm{d}y\mathrm{d}x \\
&\text{[Change of variable } p = \frac{y}{x\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} + y}\text{]} \\
&= \frac{(\alpha 2^{\sigma-1})^{1+m/\sigma}}{\sigma^{m/\sigma}\Gamma(1-\sigma)} \\
&\quad \times \int_0^{+\infty} x^m f_\sigma(x) \exp\left\{-x\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\} \\
&\qquad \times \int_0^1 \left(\frac{\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} xp}{1-p}\right)^{1-\sigma-1} \frac{\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} x}{(1-p)^2} \\
&\qquad\quad \times \exp\left\{-\frac{\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} xp}{1-p} 2^{-1}\right\} \left[x\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} xp}{1-p}\right]^{-m-1} \mathrm{d}p\mathrm{d}x \\
&\text{[Change of variable } h = x/(1-p)\text{]} \\
&= \frac{(\alpha 2^{\sigma-1})^{1+m/\sigma}}{\sigma^{m/\sigma}\Gamma(1-\sigma)} \\
&\quad \times \int_0^{+\infty}\int_0^1 (h(1-p))^m f_\sigma(h(1-p)) \exp\left\{-h(1-p)\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\} \\
&\qquad \times \left(\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} hp\right)^{1-\sigma-1} \left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} h \\
&\qquad\quad \times \exp\left\{-\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} hp 2^{-1}\right\} \left[h(1-p)\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} hp\right]^{-m-1} \mathrm{d}p\mathrm{d}h \\
&= \frac{(\alpha 2^{\sigma-1})^{1+m/\sigma}}{\sigma^{m/\sigma}\Gamma(1-\sigma)} \left(\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}}\right)^{-\sigma-m} \\
&\quad \times \int_0^{+\infty}\int_0^1 (h(1-p))^n f_\sigma(h(1-p)) \exp\left\{-h\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\} p^{1-\sigma-1} h^{-m-\sigma} \mathrm{d}p\mathrm{d}h \\
&= \frac{\sigma}{\Gamma(1-\sigma)} \int_0^{+\infty}\int_0^1 f_\sigma(hp) \exp\left\{-h\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\} h^{-\sigma} p^m (1-p)^{1-\sigma-1} \mathrm{d}p\mathrm{d}h.
\end{aligned}
$$

This complete the case $\ell = 0$. Now, we consider $\ell > 0$. The probability $p_{f_v}(\ell; m, \alpha, \sigma)$ follows by combining Proposition 1 with the distribution of $(K_m, \mathbf{N}_n)$ displayed in (2). In particular, we can write

$$
\begin{aligned}
p_{f_v}(\ell; m, \alpha, \sigma) &= \sum_{\mathbf{m} \in \mathcal{M}_{m,k}} \Pr[X_{m+1} \in \mathbf{v}_\ell \mid \mathbf{M}_m = \mathbf{m}] \Pr[\mathbf{M}_m = \mathbf{m}] \\
&= \sum_{\mathbf{m} \in \mathcal{M}_{m,k}} \frac{V_{m+1,k}}{V_{m,k}} (\ell - \sigma) m_l \Pr[\mathbf{M}_m = \mathbf{m}]
\end{aligned}
$$

$$
\begin{aligned}
&= (\ell-\sigma)\sum_{k=1}^{m}\sum_{(n_1,\ldots,n_k)\in\mathcal{P}_{m,k}}\frac{1}{k!}\binom{m}{n_1,\ldots,n_k}V_{m,k}\prod_{i=1}^{k}(1-\sigma)_{(n_i-1)}\frac{V_{m+1,k}}{V_{m,k}}\sum_{j=1}^{k}\mathbf{1}_{n_j}(\ell)\\
&= (\ell-\sigma)\sum_{k=1}^{m}\frac{V_{m+1,k}}{V_{m,k}}\sum_{j=1}^{k}\Pr[K_m=k,N_j=\ell]\\
&= (\ell-\sigma)\sum_{k=1}^{m}\frac{V_{m+1,k}}{V_{m,k}}\sum_{j=1}^{k}\frac{V_{m,k}}{k}\binom{n}{\ell}(1-\sigma)_{(\ell-1)}\frac{C(m-\ell,k-1;\sigma)}{\sigma^{k-1}}\\
&= (\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)}\sum_{k=1}^{m}V_{m+1,k}\frac{C(m-\ell,k-1;\sigma)}{\sigma^{k-1}}. \qquad (13)
\end{aligned}
$$

Then, the expression of $p_{f_v}(\ell;m,\alpha,\sigma)$ in (10) follows by combining (13) with $V_{m+1,k}$ displayed in (3), i.e.,

$$
\begin{aligned}
p_{f_v}(\ell;m,\alpha,\sigma) &= (\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)}\sum_{k=1}^{m}V_{m+1,k}\frac{C(m-\ell,k-1;\sigma)}{\sigma^{k-1}}\\
&= (\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)}\\
&\quad\times\frac{\sigma}{\Gamma(m+1)}\int_0^{+\infty}u^m\exp\left\{-\frac{\alpha 2^{\sigma-1}}{\sigma}[(2^{-1}+u)^{\sigma}-2^{-\sigma}]\right\}(2^{-1}+u)^{-m-1}\mathrm{d}u\\
&\quad\quad\times\sum_{k=1}^{m}C(m-\ell,k-1;\sigma)\left(\frac{\alpha 2^{\sigma-1}}{\sigma(2^{-1}+u)^{-\sigma}}\right)^{k}.
\end{aligned}
$$

If $\ell=m$, then

$$
\begin{aligned}
p_{f_v}(m;m,\alpha,\sigma) &= (1-\sigma)_{(m)}\sum_{k=1}^{m}V_{m+1,k}\frac{C(0,k-1;\sigma)}{\sigma^{k-1}}\\
&= (1-\sigma)_{(m)}V_{m+1,1}\\
&= (1-\sigma)_{(m)}\frac{\alpha 2^{\sigma-1}}{\Gamma(m+1)}\int_0^{+\infty}\frac{x^m}{(2^{-1}+x)^{m+1-\sigma}}\exp\left\{-\frac{\alpha 2^{\sigma-1}}{\sigma}[(2^{-1}+x)^{\sigma}-2^{-\sigma}]\right\}\mathrm{d}x.
\end{aligned}
$$

If $\ell<m$, then

$$
\begin{aligned}
p_{f_v}(\ell;m,\alpha,\sigma) &= (\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)}\sum_{k=1}^{m}V_{m+1,k}\frac{C(m-\ell,k-1;\sigma)}{\sigma^{k-1}}\\
&= (\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)}\\
&\quad\times\frac{\alpha 2^{\sigma-1}}{\Gamma(m+1)}\int_0^{+\infty}u^m\exp\left\{-\frac{\alpha 2^{\sigma-1}}{\sigma}[(2^{-1}+u)^{\sigma}-2^{-\sigma}]\right\}(2^{-1}+u)^{-m-1+\sigma}\mathrm{d}u\\
&\quad\times\sum_{k=1}^{m-\ell}C(m-\ell,k;\sigma)\left(\frac{\alpha 2^{\sigma-1}}{\sigma(2^{-1}+u)^{-\sigma}}\right)^{k}\\
&\quad\text{[Equation 13 of Favaro et al. (2015)]}\\
&= (\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)}\\
&\quad\times\frac{\alpha 2^{\sigma-1}}{\Gamma(m+1)}\int_0^{+\infty}u^m\exp\left\{-\frac{\alpha 2^{\sigma-1}}{\sigma}[(2^{-1}+u)^{\sigma}-2^{-\sigma}]\right\}(2^{-1}+u)^{-m-1+\sigma}\\
&\quad\times\exp\left\{\frac{\alpha 2^{\sigma-1}}{\sigma(2^{-1}+u)^{-\sigma}}\right\}\left(\frac{\alpha 2^{\sigma-1}}{\sigma(2^{-1}+u)^{-\sigma}}\right)^{\frac{m-\ell}{\sigma}}\\
&\quad\times\int_0^{+\infty}x^{m-\ell}\exp\left\{-x\left(\frac{\alpha 2^{\sigma-1}}{\sigma(2^{-1}+u)^{-\sigma}}\right)^{\frac{1}{\sigma}}\right\}f_{\sigma}(x)\mathrm{d}x\mathrm{d}u
\end{aligned}
$$

[Identity $(2^{-1}+u)^{-1+\sigma} = \dfrac{1}{\Gamma(1-\sigma+\ell)} \int_0^{+\infty} y^{1-\sigma+\ell-1} \exp\left\{-y(2^{-1}+u)\right\} \mathrm{d}y$]

$$
\begin{aligned}
&= (\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)} \\
&\quad\times \frac{(\alpha 2^{\sigma-1})(\alpha 2^{\sigma-1})^{\frac{m-\ell}{\sigma}}}{\sigma^{\frac{m-\ell}{\sigma}}\Gamma(m+1)} \int_0^{+\infty} u^m \left(\frac{1}{\Gamma(1-\sigma+\ell)} \int_0^{+\infty} y^{1-\sigma+\ell-1} \exp\{-y(2^{-1}+u)\} \mathrm{d}y\right) \\
&\qquad\times \int_0^{+\infty} x^{m-\ell} \exp\left\{-xu\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}}\right\} \exp\left\{-x\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\} f_\sigma(x)\mathrm{d}x\mathrm{d}u \\
&= (\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)} \\
&\times \frac{(\alpha 2^{\sigma-1})^{1+\frac{m-\ell}{\sigma}}}{\sigma^{\frac{m-\ell}{\sigma}}\Gamma(1-\sigma+\ell)} \int_0^{+\infty} x^{m-\ell} f_\sigma(x) \exp\left\{-x\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\} \\
&\qquad\times \int_0^{+\infty} y^{1-\sigma+\ell-1} \exp\{-y 2^{-1}\} \left[x\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} + y\right]^{-m-1} \mathrm{d}y\mathrm{d}x
\end{aligned}
$$

[Change of variable $p = \dfrac{y}{x\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} + y}$]

$$
\begin{aligned}
&= (\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)} \\
&\quad\times \frac{(\alpha 2^{\sigma-1})^{1+\frac{m-\ell}{\sigma}}}{\sigma^{\frac{m-\ell}{\sigma}}\Gamma(1-\sigma+\ell)} \int_0^{+\infty} x^{m-\ell} f_\sigma(x) \exp\left\{-x\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\} \\
&\qquad\times \int_0^1 \left(\frac{\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} xp}{1-p}\right)^{1-\sigma+\ell-1} \frac{\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} x}{(1-p)^2} \\
&\qquad\quad\times \exp\left\{-\frac{\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} xp}{1-p} 2^{-1}\right\} \left[x\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} xp}{1-p}\right]^{-m-1} \mathrm{d}p\mathrm{d}x
\end{aligned}
$$

[Change of variable $h = x/(1-p)$]

$$
\begin{aligned}
&= (\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)} \\
&\quad\times \frac{(\alpha 2^{\sigma-1})^{1+\frac{m-\ell}{\sigma}}}{\sigma^{\frac{m-\ell}{\sigma}}\Gamma(1-\sigma+\ell)} \int_0^{+\infty}\int_0^1 (h(1-p))^{m-\ell} f_\sigma(h(1-p)) \exp\left\{-h(1-p)\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\} \\
&\qquad\times \left(\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} hp\right)^{1-\sigma+\ell-1} \left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} h \\
&\qquad\quad\times \exp\left\{-\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} hp 2^{-1}\right\} \left[h(1-p)\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}} hp\right]^{-m-1} \mathrm{d}p\mathrm{d}h \\
&= (\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)} \\
&\quad\times \frac{(\alpha 2^{\sigma-1})^{1+\frac{m-\ell}{\sigma}}}{\sigma^{\frac{m-\ell}{\sigma}}\Gamma(1-\sigma+\ell)} \left(\left(\frac{\alpha 2^{\sigma-1}}{\sigma}\right)^{\frac{1}{\sigma}}\right)^{-m-\sigma+\ell} \\
&\qquad\times \int_0^{+\infty}\int_0^1 (h(1-p))^{m-\ell} f_\sigma(h(1-p)) \exp\left\{-h\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\} p^{1-\sigma+\ell-1} h^{-m-\sigma+\ell} \mathrm{d}p\mathrm{d}h
\end{aligned}
$$

$$= (\ell - \sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)}$$

$$\times \frac{\sigma}{\Gamma(1-\sigma+\ell)} \int_0^{+\infty}\int_0^1 f_\sigma(hp)\exp\left\{-h\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\} h^{-\sigma} p^{m-\ell}(1-p)^{1-\sigma+\ell-1}\mathrm{d}p\mathrm{d}h.$$

□

Remark 2. *Here we present an alternative representation of* $p_{f_v}(\ell; m, \alpha, \sigma)$ *in* (10)*. It provides a useful tool for implementing a straightforward Monte Carlo evaluation of* $p_{f_v}(\ell; m, \alpha, \sigma)$*. For* $\ell = m$*,*

$$\begin{aligned}
p_{f_v}(m; m, \alpha, \sigma) &= \frac{\sigma(\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)}}{\Gamma(1-\sigma+\ell)} \\
&\quad \times \int_0^{+\infty}\int_0^1 \frac{1}{h^\sigma} f_\sigma(hp) e^{-h\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}} p^{m-\ell}(1-p)^{1-\sigma+\ell-1}\mathrm{d}p\mathrm{d}h \\
&= \frac{1}{\Gamma(m+1)}\int_0^{+\infty} \exp\left\{-h\left(\frac{\alpha}{2\sigma}\right)^{1/\sigma} + \frac{\alpha}{2\sigma}\right\} \frac{\sigma\Gamma(m+1)}{\Gamma(m+1-\sigma)} h^{-\sigma} \\
&\quad \times \int_0^1 (1-p)^{m+1-\sigma-1} f_\sigma(hp)\mathrm{d}p\mathrm{d}h \\
&= \frac{(1-\sigma)_{(m)}}{\Gamma(m+1)}\mathbb{E}\left[\exp\left\{-\frac{X}{Y}\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\}\right],
\end{aligned}$$

where Y *is a Beta random variable with parameter* $(m - \ell + \sigma, 1 - \sigma + \ell)$ *and* X *is a random variable, independent of* Y*, distributed according to a polynomially tilted* σ*-stable distribution of order* σ*, i.e.*

$$Pr[X \in \mathrm{d}x] = \frac{\Gamma(\sigma+1)}{\Gamma(2)} x^{-\sigma} f_\sigma(x)\mathrm{d}x.$$

For $\ell < m$*,*

$$\begin{aligned}
p_{f_v}(\ell; m, \alpha, \sigma) &= \frac{\sigma(\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)}}{\Gamma(1-\sigma+\ell)} \\
&\quad \times \int_0^{+\infty}\int_0^1 \frac{1}{h^\sigma} f_\sigma(hp) e^{-h\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}} p^{m-\ell}(1-p)^{1-\sigma+\ell-1}\mathrm{d}p\mathrm{d}h \\
&= (\ell-\sigma)\binom{m}{\ell}(1-\sigma)_{(\ell-1)} \\
&\quad \times \frac{\Gamma(m-\ell+\sigma)}{\Gamma(\sigma)\Gamma(m+1)} \\
&\qquad \times \int_0^{+\infty} f_\sigma(hp)\exp\left\{-h\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\} \frac{\Gamma(\sigma+1)}{\Gamma(2)} h^{-\sigma} \\
&\qquad\quad \times \frac{\Gamma(m+1)}{\Gamma(1-\sigma+\ell)\Gamma(m-\ell+\sigma)} \int_0^1 p^{m-\ell}(1-p)^{1-\sigma+\ell-1}\mathrm{d}p\mathrm{d}h \\
&= \frac{\Gamma(m-\ell+\sigma)}{\Gamma(\sigma)\Gamma(m+1)}\mathbb{E}\left[\exp\left\{-\frac{X}{Y}\left(\frac{\alpha 2^{-1}}{\sigma}\right)^{\frac{1}{\sigma}} + \frac{\alpha 2^{-1}}{\sigma}\right\}\right].
\end{aligned}$$

According to this alternative representation, $p_{f_v}(\ell; m, \alpha, \sigma)$ *allows for a Monte Carlo evaluation by sampling from a Beta random variable and from a polynomially tilted* σ*-stable random variable of order* σ*. See, e.g., (Devroye, 2009).*

Proof of Proposition 2. The proof follows by a direct application of Lemma (1) by setting $\sigma = 1/2$. First, let recall that the density function of the (1/2)-stable positive random variable coincides with the IG density function (Seshadri, 1993) with shape parameter $a = 2^{-1/2}$ and scale parameter $b = 0$. That is, we write

$$f_{1/2}(x) = \frac{1}{2\sqrt{\pi}} w^{-\frac{3}{2}} \exp\left\{-\frac{1}{4w}\right\}.$$

For $\ell = m$,

$$p_{f_v}(m; m, \alpha, \sigma) = \frac{\alpha 2^m \left(\frac{1}{2}\right)_{(m)}}{\Gamma(m+1)} \times \int_0^{+\infty} \frac{x^m}{(1+2x)^{m+\frac{1}{2}}} \exp\left\{-\alpha 2^{1/2}[(2^{-1}+x)^{1/2} - 2^{-1/2}]\right\} \mathrm{d}x.$$

For $\ell < m$,

$$\begin{aligned} p_{f_v}(\ell; m, \alpha, \sigma) &= \frac{2^{-1}(\ell - 2^{-1})\binom{m}{\ell}\left(\frac{1}{2}\right)_{(\ell-1)}}{\Gamma(2^{-1}+\ell)} \\ &\quad \times \int_0^{+\infty}\int_0^1 \frac{1}{\sqrt{h}} f_{1/2}(hp) \mathrm{e}^{-h\alpha^2+\alpha} p^{m-\ell}(1-p)^{\frac{1}{2}+\ell-1} \mathrm{d}p\mathrm{d}h \\ &= (\ell - 2^{-1})\binom{m}{\ell}(1-2^{-1})_{(\ell-1)} \\ &\quad \times \frac{2^{-1}}{2\pi^{1/2}} \mathrm{e}^{\alpha} \int_0^1 \int_0^{+\infty} h^{-1-1} \exp\left\{-h\alpha^2 - \frac{\frac{1}{4p}}{h}\right\} \\ &\quad \times \frac{1}{\Gamma(1-2^{-1}+\ell)} p^{m-\ell-\frac{1}{2}-1}(1-p)^{1-\frac{1}{2}+\ell-1} \mathrm{d}p\mathrm{d}h \\ &\quad \text{[Equation 3.471.9 of Gradshteyn and Ryzhic (2007)]} \\ &= (\ell - 2^{-1})\binom{m}{\ell}(1-2^{-1})_{(\ell-1)} \\ &\quad \times \frac{\mathrm{e}^{\alpha}\alpha}{\pi^{1/2}\Gamma(1-2^{-1}+\ell)} \int_0^1 K_{-1}\left(\frac{\alpha}{p^{1/2}}\right) p^{m-\ell-1}(1-p)^{1-\frac{1}{2}+\ell-1} \mathrm{d}p, \end{aligned}$$

where K_{-1} is the modified Bessel function of the second type, or Macdonald function, with parameter -1. □

Remark 3. *Here we present an alternative representation of $p_{f_v}(\ell; m, \alpha, \sigma)$ in Proposition 2. It provides a useful tool for implementing a straightforward Monte Carlo evaluation of $p_{f_v}(\ell; m, \alpha, \sigma)$. For $\ell = m$,*

$$\begin{aligned} p_{f_v}(m; m, \alpha, \sigma) &= \frac{\alpha 2^m \left(\frac{1}{2}\right)_{(m)}}{\Gamma(m+1)} \\ &\quad \times \int_0^{+\infty} \frac{x^m}{(1+2x)^{m+\frac{1}{2}}} \exp\left\{-\alpha 2^{1/2}[(2^{-1}+x)^{1/2} - 2^{-1/2}]\right\} \mathrm{d}x \\ &= \frac{1}{\Gamma(m+1)} \int_0^{+\infty} \exp\left\{-h\alpha^2 + \alpha\right\} \frac{2^{-1}\Gamma(m+1)}{\Gamma(m+1-1/2)} \frac{1}{\sqrt{h}} \\ &\quad \times \int_0^1 (1-p)^{m+1-\frac{1}{2}-1} \frac{1}{2\sqrt{\pi}} (hp)^{-\frac{3}{2}} \exp\left\{-\frac{1}{4hp}\right\} \mathrm{d}p\mathrm{d}h \\ &= \frac{\left(\frac{1}{2}\right)_{(m)}}{\Gamma(m+1)} \mathbb{E}\left[\exp\left\{-\frac{X}{Y}\alpha^2 + \alpha\right\}\right], \end{aligned}$$

where Y is a Beta random variable with parameter $(1/2, m+1/2)$ and X is a random variable, independent of Y, distributed according to a polynomially tilted IG distribution of the order $1/2$, that is

$$Pr[X \in \mathrm{d}x] = \frac{\Gamma(3/2)}{\Gamma(2)} x^{-\frac{1}{2}} \frac{x^{-\frac{3}{2}}}{2\sqrt{\pi}} \exp\left\{-\frac{1}{4x}\right\} \mathrm{d}x.$$

For $\ell < m$,

$$p_{f_v}(\ell; m, \alpha, \sigma) = (\ell - 2^{-1})\binom{m}{\ell}(1-2^{-1})_{(\ell-1)} \times \frac{e^{\alpha}\alpha}{\pi^{1/2}\Gamma(1-2^{-1}+\ell)} \int_0^1 K_{-1}\left(\frac{\alpha}{p^{1/2}}\right) p^{m-\ell-1}(1-p)^{1-\frac{1}{2}+\ell-1} \mathrm{d}p$$

$$
\begin{aligned}
&= (\ell - 2^{-1})\binom{m}{\ell}(1-2^{-1})_{(\ell-1)} \\
&\quad \times \frac{\Gamma(m-\ell)}{\pi^{1/2}\Gamma(m+1-2^{-1})} e^{\alpha}\alpha \\
&\qquad \times \int_0^1 K_{-1}\left(\frac{\alpha}{p^{1/2}}\right)\frac{\Gamma(m+1-2^{-1})}{\Gamma(1-2^{-1}+\ell)\Gamma(m-\ell)} p^{m-\ell-1}(1-p)^{1-\frac{1}{2}+\ell-1}\mathrm{d}p \\
&= (l-2^{-1})\binom{n}{l}(1-2^{-1})_{(l-1)} \\
&\quad \times \frac{\Gamma(m-\ell)}{\pi^{1/2}\Gamma(m+1-2^{-1})} e^{\alpha}\alpha\mathbb{E}\left[K_{-1}\left(\frac{\alpha}{Y^{1/2}}\right)\right],
\end{aligned}
$$

where Y is a Beta random variable with parameter $(m-\ell, 1/2+\ell)$. According to this alternative representation, $p_{f_v}(\ell; m, \alpha, \sigma)$ allows for a straightforward Monte Carlo evaluation by sampling from a Beta random variable and from a polynomially tilted IG random variable of order $1/2$. See, e.g., (Devroye, 2009).

Proof of Theorem 3. Because of the assumption of independence of the hash family, we can factorize the marginal likelihood of $(\mathbf{c}_1, \ldots, \mathbf{c}_N)$, i.e. of hash functions $h_1, \ldots, h_N$, into the product of the marginal likelihoods of $\mathbf{c}_n = (c_{n,1}, \ldots, c_{n,J})$, i.e. of each hash function. This, combined with Bayes theorem, leads to

$$
\begin{aligned}
&\Pr[f_v = \ell \mid \{C_{n,h_n(v)}\}_{n\in[N]} = \{c_{n,h_n(v)}\}_{n\in[N]}] \\
&\quad \text{[Bayes theorem and independence of the hash family]} \\
&\quad = \frac{1}{\Pr[\{C_{n,h_n(v)}\}_{n\in[N]} = \{c_{n,h_n(v)}\}_{n\in[N]}]}\Pr[f_v=\ell]\prod_{n=1}^{N}\Pr[C_{n,h_n(v)} = c_{n,h_n(v)} \mid f_v = \ell] \\
&\quad = \frac{1}{\Pr[\{C_{n,h_n(v)}\}_{n\in[N]} = \{c_{n,h_n(v)}\}_{n\in[N]}]}\Pr[f_v=\ell]\prod_{n=1}^{N}\frac{\Pr[C_{n,h_n(v)} = c_{n,h_n(v)},\, f_v = \ell]}{\Pr[f_v=\ell]} \\
&\quad = \frac{1}{\Pr[\{C_{n,h_n(v)}\}_{n\in[N]} = \{c_{n,h_n(v)}\}_{n\in[N]}]}(\Pr[f_v=\ell])^{1-N}\prod_{n=1}^{N}\Pr[C_{n,h_n(v)} = c_{n,h_n(v)}]\Pr[f_v = \ell \mid C_{n,h_n(v)} = c_{n,h_n(v)}] \\
&\quad = (\Pr[f_v=\ell])^{1-N}\prod_{n=1}^{N}\Pr[f_v = \ell \mid C_{n,h_n(v)} = c_{n,h_n(v)}] \\
&\quad \propto \prod_{n=1}^{N}\Pr[f_v = \ell \mid C_{n,h_n(v)} = c_{n,h_n(v)}] \\
&\quad \text{[Proposition 2 and Equation 9]} \\
&\quad = \prod_{n\in[N]}\begin{cases} \frac{\binom{c_{n,h_n(v)}}{\ell}e^{\frac{\alpha}{J}}\alpha}{J\pi}\int_0^1 K_{-1}\left(\frac{\alpha}{J\sqrt{x}}\right)x^{c_{n,h_n(v)}-\ell-1}(1-x)^{\frac{1}{2}+\ell-1}\mathrm{d}x & \ell = 0,1,\ldots,c_{n,h_n(v)}-1 \\ \frac{2^{c_{n,h_n(v)}}\alpha\left(\frac{1}{2}\right)_{(c_{n,h_n(v)})}}{J\Gamma(c_{n,h_n(v)}+1)}\int_0^{+\infty}\frac{x^{c_{n,h_n(v)}}}{(1+2x)^{c_{n,h_n(v)}+1/2}}e^{-\frac{\alpha}{J}(\sqrt{1+2x}-1)}\mathrm{d}x & \ell = c_{n,h_n(v)}, \end{cases}
\end{aligned}
$$

where $K_{-1}(\cdot)$ is the modified Bessel function of the second type, or Macdonald function, with parameter -1.

F Estimation of α

We start by deriving the marginal likelihood corresponding to the hashed frequencies $(\mathbf{c}_1, \ldots, \mathbf{c}_N)$ induced by the collection of hash functions $h_1, \ldots, h_N$. In particular, according to the definition of $P \sim \text{NIGP}(\alpha, \nu)$ through its family of finite-dimensional distributions, for a single hash function h_n the marginal likelihood of $\mathbf{c}_n = (c_{n,1}, \ldots, c_{n,J})$ is obtained by integrating the normalized IG distribution with parameter $(\alpha/J, \ldots, \alpha/J)$ against the multinomial counts $\mathbf{c}_n$. In particular, by means of the normalized IG distribution (9), the marginal likelihood of $\mathbf{c}_n$ has the following expression

$$p(\mathbf{c}_n; \alpha)$$

$$
\begin{aligned}
&= \frac{m!}{\prod_{i=1}^{J} c_{n,i}!} \\
&\quad \times \int_{\{(p_1,\dots,p_{J-1})\,:\, p_i\in(0,1) \text{ and } \sum_{i=1}^{J-1} p_i \le 1\}} \prod_{i=1}^{J-1} p_i^{c_{n,i}} \left(1-\sum_{i=1}^{J-1} p_i\right)^{c_{n,J}} f_{(P_1,\dots,P_{J-1})}(p_1,\dots,p_{J-1})\mathrm{d}p_1\cdots\mathrm{d}p_{J-1} \\
&= \frac{m!}{\prod_{i=1}^{J} c_{n,i}!} \\
&\quad \times \int_{\{(p_1,\dots,p_{J-1})\,:\, p_i\in(0,1) \text{ and } \sum_{i=1}^{J-1} p_i \le 1\}} \left(\prod_{i=1}^{J} \frac{(\alpha/J)\mathrm{e}^{\alpha/J}}{\sqrt{2\pi}}\right) \prod_{i=1}^{J-1} p_i^{c_{n,i}-3/2} \left(1-\sum_{i=1}^{J-1} p_i\right)^{c_{n,J}-3/2} \\
&\quad\quad \times \int_0^{+\infty} z^{-3J/2+J-1} \exp\left\{-\frac{1}{2z}\left(\sum_{i=1}^{J-1}\frac{(\alpha/J)^2}{p_i} + \frac{(\alpha/J)^2}{1-\sum_{i=1}^{J-1} p_i}\right) - \frac{z}{2}\right\} \mathrm{d}z\mathrm{d}p_1\cdots\mathrm{d}p_{J-1}
\end{aligned}
$$

[Change of variable $p_i = \frac{x_i}{\sum_{i=1}^{k} x_i}$, for $i = 1,\dots,J-1$, and $z = \sum_{i=1}^{J} x_i$]

$$
\begin{aligned}
&= \frac{m!}{\prod_{i=1}^{J} c_{n,i}!} \\
&\quad \times \left(\frac{(\alpha/J)\mathrm{e}^{\alpha/J}}{\sqrt{2\pi}}\right)^J \int_{(0,+\infty)^J} \prod_{i=1}^{J} x_i^{c_{n,i}-3/2} \left(\sum_{i=1}^{J} x_i\right)^{-\sum_{i=1}^{J} c_{n,i}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{J}\frac{(\alpha/J)^2}{x_i} - \frac{1}{2}\sum_{i=1}^{J} x_i\right\}\mathrm{d}x_1\cdots\mathrm{d}x_J \\
&= \frac{m!}{\prod_{i=1}^{J} c_{n,i}!} \\
&\quad \times \left(\frac{(\alpha/J)\mathrm{e}^{\alpha/J}}{\sqrt{2\pi}}\right)^J \frac{1}{\Gamma(m)} \int_{(0,+\infty)^J} \prod_{i=1}^{J} x_i^{c_{n,i}-3/2} \left(\int_0^{+\infty} y^{m-1} \exp\left\{-y\sum_{i=1}^{J} x_i\right\}\mathrm{d}y\right) \\
&\quad\quad \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{J}\frac{(\alpha/J)^2}{x_i} - \frac{1}{2}\sum_{i=1}^{J} x_i\right\}\mathrm{d}x_1\cdots\mathrm{d}x_J \\
&= \frac{m!}{\prod_{i=1}^{J} c_{n,i}!} \\
&\quad \times \left(\frac{(\alpha/J)\mathrm{e}^{\alpha/J}}{\sqrt{2\pi}}\right)^J \frac{1}{\Gamma(m)} \int_0^{+\infty} y^{m-1} \left(\prod_{i=1}^{J}\int_0^{+\infty} x_i^{c_{n,i}-3/2} \exp\left\{-\frac{(\alpha/J)^2}{2x_i} - x_i\left(y+\frac{1}{2}\right)\right\}\mathrm{d}x_i\right)\mathrm{d}y
\end{aligned}
$$

[Equation 3.471.9 of Gradshteyn and Ryzhic (2007)]

$$
\begin{aligned}
&= \frac{m!}{\prod_{i=1}^{J} c_{n,i}!} \\
&\quad \times \left(\frac{(\alpha/J)\mathrm{e}^{\alpha/J}}{\sqrt{2\pi}}\right)^J \frac{1}{\Gamma(m)} \int_0^{+\infty} y^{m-1} \left(\prod_{i=1}^{J} 2\left(\frac{(\alpha/J)^2}{1+2y}\right)^{c_{n,i}/2-1/4} K_{c_{n,i}-1/2}\left(\sqrt{c_{n,i}^2(1+2y)}\right)\right)\mathrm{d}y \\
&= \frac{m\left(\frac{\alpha}{J}\right)^{m+\frac{J}{2}}\mathrm{e}^{\alpha}}{(\pi/2)^{\frac{J}{2}}\prod_{j=1}^{J} c_{n,j}!} \int_0^{+\infty} \frac{y^{m-1}}{(1+2y)^{m/2-J/4}} \left(\prod_{i=1}^{J} K_{c_{n,i}-1/2}\left(\sqrt{(\alpha/J)_i^2(1+2y)}\right)\right)\mathrm{d}y.
\end{aligned}
$$

Because of the independence of the hash family, $h_1,\dots,h_N$ leads to the following marginal likelihood of $\{c_{n,j}\}_{n\in[N]\, j\in[J]}$

$$
\begin{aligned}
&p(\mathbf{c}_1,\dots,\mathbf{c}_N;\alpha) \qquad (14) \\
&\quad = \prod_{n\in[N]} \frac{m\left(\frac{\alpha}{J}\right)^{m+\frac{J}{2}}\mathrm{e}^{\alpha}}{(\pi/2)^{\frac{J}{2}}\prod_{j=1}^{J} c_{n,j}!} \int_0^{+\infty} \frac{x^{m-1}}{(1+2x)^{\frac{m}{2}-\frac{J}{4}}} \left(\prod_{j=1}^{J} K_{c_{n,j}-\frac{1}{2}}\left(\sqrt{\left(\frac{\alpha}{J}\right)^2(1+2x)}\right)\right)\mathrm{d}x.
\end{aligned}
$$

The marginal likelihood of $\{c_{n,j}\}_{n\in[N]\,j\in[J]}$ in (14) is applied to estimate the mass parameter α. This is the empirical Bayes approach to the estimation of α. In particular, we consider the following problem

$$\arg\max_{\alpha}\left\{\prod_{n\in[N]} V_{n,m,\alpha,J}\int_0^{+\infty} F_{n,m,\alpha,J}(y)\mathrm{d}y\right\},$$

where

$$V_{n,m,\alpha,J}=\frac{m\left(\frac{\alpha}{J}\right)^{m+\frac{J}{2}}\mathrm{e}^{\alpha}}{(\pi/2)^{\frac{J}{2}}\prod_{j=1}^{J} c_{n,j}!}$$

and

$$F_{n,m,\alpha,J}(y)=\frac{y^{m-1}}{(1+2y)^{\frac{m}{2}-\frac{J}{4}}}\left(\prod_{j=1}^{J} K_{c_{n,j}-\frac{1}{2}}\left(\sqrt{\left(\frac{\alpha}{J}\right)^2(1+2y)}\right)\right)$$

under the constraint that $\alpha>0$. To avoid overflow/underflow issues in the above optimization problem, here we work in log-space. That is, we consider the following equivalent optimization problem[1]

$$\begin{aligned}&\arg\max_{\alpha}\left\{\sum_{n\in[N]}\log(V_{n,m,\alpha,J})+\log\left(\int_0^{+\infty} F_{n,m,\alpha,J}(y)\mathrm{d}y\right)\right\}\\&=\arg\max_{\alpha}\left\{\sum_{n\in[N]} v_{n,m,\alpha,J}+\log\left(\int_0^{+\infty}\exp\{f_{n,m,\alpha,J}(y)\}\mathrm{d}y\right)\right\},\end{aligned}$$

with $v_{n,m,\alpha,J}=\log(V_{n,m,\alpha,J})$ and $f_{n,m,\alpha,J}(y)=\log(F_{n,m,\alpha,J}(y))$. For the computation of the integral we use double exponential quadrature (Takahasi and Mori, 1974), which approximates $\int_{-1}^{+1} f(y)\mathrm{d}y$ with $\sum_{j=1}^{m} w_j f(y_j)$ for appropriate weights $w_j\in\mathcal{W}$ and coordinates $y_j\in\mathcal{Y}$. Integrals of the form $\int_a^b f(y)\mathrm{d}y$ for $-\infty\le a\le b\le+\infty$ are handled via change of variable formulas. To avoid underflow/overflow issues it is necessary to apply the "log-sum-exp" trick to the above integral. That is,

$$\log\left(\int_0^{+\infty}\exp\{f_{n,m,\alpha,J}(y)\}\mathrm{d}y\right)=f^*+\log\left(\int_0^{+\infty}\exp\{f_{n,m,\alpha,J}(y)-f^*\}\mathrm{d}y\right)$$

and

$$f^*=\arg\max_{y\in\mathcal{Y}}\{f_{n,m,\alpha,J}(y)\}.$$

The computation of $\log(K_{c_{n,j}-\frac{1}{2}}(x))$ is performed via the following finite-sum representation of $K_{c_{n,j}-\frac{1}{2}}(x)$, which holds for $K_v(x)$ when v is an half-integer. Recall that $K_v(x)$ is symmetric in v. In particular,

$$\begin{aligned}&K_{c_{n,i}-1/2}\left(\sqrt{(\alpha/J)(1+2y)}\right)\\&=\sqrt{\frac{\pi}{2}}\frac{\exp\left\{-((\alpha/J)(1+2y))^{1/2}\right\}}{((\alpha/J)(1+2y))^{1/4}}\sum_{j=0}^{c_{n,i}-1}\frac{(j+c_{n,i}-1)!}{j!\,(c_{n,i}-j-1)!}(2((\alpha/J)(1+2y))^{1/2})^{-j}.\end{aligned}$$

In order to increase efficiency in the our optimization, we cache the log-factorials and, anew for each α and y the values of $\log(K_{c_{n,j}-\frac{1}{2}}(\sqrt{(\alpha/J)^2(1+2y)}))$ across j. In particular, as the dependency on j goes through $c_{n,j}$ only we can exploit the fact that many duplicates exists, i.e. the complexity scales in the number of unique $c_{n,j}$. All code is implemented in `LuaJIT`[2] by using the `scilua`[3] library.

[1] the computation of log-factorials is done via the specialized implementation of the log-gamma function
[2] https://luajit.org
[3] https://scilua.org

G Additional experiments

We present additional experiments on the application of the CMS-NIGP on synthetic and real data. First, we recall the synthetic and real data to which the CMS-NIGP is applied. As regards synthetic data, we consider datasets of $m = 500.000$ tokens from a Zipf's distributions with parameter $s = 1.3$, 1.6, 1.9, $2, 2$, 2.5. As regards real data, we consider: i) the 20 Newsgroups dataset, which consists of $m = 2.765.300$ tokens with $k = 53.975$ distinct tokens; ii) the Enron dataset, which consists of $m = 6412175$ tokens with $k = 28102$ distinct tokens. Tables 1, 2, 3 and 4 report the MAE (mean absolute error) between true frequencies and their corresponding estimates via: i) the CMS-NIGP estimate $\hat{f}_v^{(\mathrm{NIGP})}$; ii) the CMS estimate $\hat{f}_v^{(\mathrm{CMS})}$; iii) the CMS-DP estimate$\hat{f}_v^{(\mathrm{DP})}$, the CMM estimate $\hat{f}_v^{(\mathrm{CMM})}$.

References

Bacallado, S., Battiston, M., Favaro, S., and Trippa, L. (2017). Sufficientness postulates for gibbs-type priors and hierarchical generalizations. *Statistical Science*, 32:487–500.

Cai, D., Mitzenmacher, M., and Adams, R. P. (2018). A Bayesian nonparametric view on count-min sketch. In *Advances in Neural Information Processing Systems*.

Charalambides, C. A. (2005). *Combinatorial methods in discrete distributions*, volume 600. John Wiley & Sons.

Cormode, G. and Muthukrishnan, S. (2005a). An improved data stream summary: the count-min sketch and its applications. *Journal of Algorithms*, 55:58–75.

Cormode, G. and Muthukrishnan, S. (2005b). Summarizing and mining skewed data streams. In *Proceedings of the 2005 SIAM International Conference on Data Mining*.

De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Prünster, I., and Ruggiero, M. (2013). Are Gibbs-type priors the most natural generalization of the Dirichlet process? *IEEE transactions on pattern analysis and machine intelligence*, 37:212–229.

Devroye, L. (2009). Random variate generation for exponentially and polynomially tilted stable distributions. *ACM Transactions on Modeling and Computer Simulation*, 19.

Ewens, W. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 3:87–112.

Favaro, S., Nipoti, B., and Teh, Y. (2015). Random variate generation for laguerre-type exponentially tilted alpha-stable distributions. *Electronic Journal of Statistics*, 9:1230–1242.

Gradshteyn, I. and Ryzhic, I. (2007). *Table of integrals, series and products*. Academic Press.

James, L. F. (2002). Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics. *arXiv preprint arXiv:math/0205093*.

Kingman, J. (1993). *Poisson processes*. Wiley Online Library.

Lijoi, A., Mena, R. H., and Prünster, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *Journal of the American Statistical Association*, 100:1278–1291.

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*, 69:715–740.

Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics, Hjort, N.L., Holmes, C.C. Müller, P. and Walker, S.G. Eds. Cambridge University Press.*

Pitman, J. (2003). Poisson-Kingman partitions. In *Science and Statistics: A Festschrift for Terry Speed. Institute of Mathematical Statistics.*

Pitman, J. (2006). *Combinatorial stochastic processes*. Lecture Notes in Mathematics. Springer-Verlag.

Prünster, I. (2002). *Random probability measures derived from increasing additive processes and their application to Bayesian statistics*. Ph.D Thesis, University of Pavia.

Regazzini, E. (2001). *Foundations of Bayesian statistics and some theory of Bayesian nonparametric methods*. Lecture Notes, Stanford University.

Regazzini, E., Lijoi, A., and Prünster, I. (2003). Distributional results for means of normalized random measures with independent increments. *The Annals of Statistics*, 31:560–585.

Seshadri, V. (1993). *The inverse Gaussian distribution.* Oxford University Press.

Takahasi, H. and Mori, M. (1974). Double exponential formulas for numerical integration. *Publications of the Research Institute for Mathematical Sciences*, 9(3):721–741.

Table 1: Synthetic data: MAE for $\hat{f}_v^{(\mathrm{NIGP})}$, $\hat{f}_v^{(\mathrm{CMM})}$ and $\hat{f}_v^{(\mathrm{CMS})}$, case $J = 320, N = 2$

	$\mathcal{Z}_{1.3}$			$\mathcal{Z}_{1.6}$			$\mathcal{Z}_{1.9}$			$\mathcal{Z}_{2.2}$			$\mathcal{Z}_{2.5}$		
Bins v	$\hat{f}_v^{(\mathrm{CMS})}$	$\hat{f}_v^{(\mathrm{CMM})}$	$\hat{f}_v^{(\mathrm{NIGP})}$	$\hat{f}_v^{(\mathrm{CMS})}$	$\hat{f}_v^{(\mathrm{CMM})}$	$\hat{f}_v^{(\mathrm{NIGP})}$	$\hat{f}_v^{(\mathrm{CMS})}$	$\hat{f}_v^{(\mathrm{CMM})}$	$\hat{f}_v^{(\mathrm{NIGP})}$	$\hat{f}_v^{(\mathrm{CMS})}$	$\hat{f}_v^{(\mathrm{CMM})}$	$\hat{f}_v^{(\mathrm{NIGP})}$	$\hat{f}_v^{(\mathrm{CMS})}$	$\hat{f}_v^{(\mathrm{CMM})}$	$\hat{f}_v^{(\mathrm{NIGP})}$
(0,1]	1,061.3	161.72	231.31	629.40	62.19	134.75	308.11	81.10	65.71	51.65	1.04	12.91	32.65	1.02	7.16
(1,2]	1,197.9	169.74	287.43	514.31	102.42	119.22	154.20	2.00	37.03	289.50	2.04	61.87	48.15	2.01	9.88
(2,4]	1,108.3	116.37	262.18	474.82	52.10	95.78	2,419.51	2215.85	353.73	134.05	3.40	26.90	54.34	10.50	10.09
(4,8]	1,275.9	378.04	302.89	786.73	214.46	175.10	460.13	258.90	83.30	118.40	6.44	21.58	69.85	6.03	14.28
(8,16]	1,236.1	230.32	257.08	719.84	232.24	136.66	380.05	139.50	66.44	413.13	129.03	77.39	80.80	13.10	20.15
(16,32]	1,256.8	221.98	248.41	831.70	79.73	190.05	288.59	23.90	41.99	503.60	364.30	90.29	9.86	22.39	15.36
(32,64]	1,312.8	235.87	284.12	783.90	184.99	139.52	415.58	54.82	67.30	217.81	82.92	48.00	10.22	30.90	28.90
(64,128]	1,721.7	766.29	312.59	950.31	304.36	125.07	1,875.50	1762.20	353.10	64.01	97.40	65.91	13.75	96.98	66.18
(128,256]	1,107.7	334.57	97.91	1,727.19	1488.38	273.50	202.09	163.61	110.32	46.80	156.71	130.94	17.51	181.38	125.75

Table 2: Synthetic data: MAE for $\hat{f}_v^{(\mathrm{NIGP})}$, $\hat{f}_v^{(\mathrm{CMM})}$ and $\hat{f}_v^{(\mathrm{CMS})}$, case $J = 160, N = 4$

	$\mathcal{Z}_{1.3}$			$\mathcal{Z}_{1.6}$			$\mathcal{Z}_{1.9}$			$\mathcal{Z}_{2.2}$			$\mathcal{Z}_{2.5}$		
Bins v	$\hat{f}_v^{(\mathrm{CMS})}$	$\hat{f}_v^{(\mathrm{CMM})}$	$\hat{f}_v^{(\mathrm{NIGP})}$	$\hat{f}_v^{(\mathrm{CMS})}$	$\hat{f}_v^{(\mathrm{CMM})}$	$\hat{f}_v^{(\mathrm{NIGP})}$	$\hat{f}_v^{(\mathrm{CMS})}$	$\hat{f}_v^{(\mathrm{CMM})}$	$\hat{f}_v^{(\mathrm{NIGP})}$	$\hat{f}_v^{(\mathrm{CMS})}$	$\hat{f}_v^{(\mathrm{CMM})}$	$\hat{f}_v^{(\mathrm{NIGP})}$	$\hat{f}_v^{(\mathrm{CMS})}$	$\hat{f}_v^{(\mathrm{CMM})}$	$\hat{f}_v^{(\mathrm{NIGP})}$
(0,1]	212.1	590.48	0.94	262.00	146.11	0.25	424.8	130.90	0.18	154.79	47.10	0.32	56.7	1.01	0.38
(1,2]	339.8	359.57	0.56	332.75	63.21	0.70	552.0	65.00	0.82	182.72	2.01	1.24	48.2	2.03	1.45
(2,4]	270.9	69.42	1.33	277.80	301.89	2.47	487.3	163.55	2.53	184.70	97.15	2.66	57.8	14.35	2.74
(4,8]	234.6	339.95	4.69	375.74	579.94	4.67	545.2	243.08	5.28	252.53	62.70	5.96	51.1	8.30	5.42
(8,16]	213.3	313.37	10.57	165.73	152.53	10.68	493.2	196.20	10.86	247.33	29.70	10.28	24.1	14.11	11.75
(16,32]	283.0	23.30	20.72	217.20	22.94	19.21	535.5	154.30	22.08	295.90	190.92	21.57	25.0	23.20	23.37
(32,64]	305.7	133.09	42.66	284.61	209.13	43.14	637.8	150.05	42.64	120.62	71.86	44.49	31.7	40.40	44.03
(64,128]	244.5	102.43	92.26	120.21	118.42	94.43	425.1	198.60	95.19	180.30	113.75	95.10	29.2	94.73	93.34
(128,256]	237.4	294.43	170.09	141.30	573.12	173.87	525.9	267.15	185.83	129.70	176.50	180.41	32.1	119.19	179.51

Table 3: Real data ($J = 12000$ and $N = 2$): MAE for $\hat{f}_v^{(\mathrm{NIGP})}$, $\hat{f}_v^{(\mathrm{DP})}$ and $\hat{f}_v^{(\mathrm{CMS})}$

	20 Newsgroups			Enron		
Bins for true v	$\hat{f}_v^{(\mathrm{CMS})}$	$\hat{f}_v^{(\mathrm{DP})}$	$\hat{f}_v^{(\mathrm{NIGP})}$	$\hat{f}_v^{(\mathrm{CMS})}$	$\hat{f}_v^{(\mathrm{DP})}$	$\hat{f}_v^{(\mathrm{NIGP})}$
(0,1]	46.4	46.39	11.34	12.2	12.20	3.00
(1,2]	16.6	16.60	3.53	13.8	13.80	3.06
(2,4]	38.4	38.40	7.71	61.5	61.49	12.55
(4,8]	59.4	59.39	10.40	88.4	88.39	17.36
(8,16]	54.3	54.29	11.34	23.4	23.40	4.58
(16,32]	17.8	17.80	9.85	55.1	55.09	11.58
(32,64]	40.8	40.79	25.65	128.5	128.48	39.46
(64,128]	26.0	25.99	57.95	131.1	131.08	54.42
(128,256]	13.6	13.59	126.07	50.7	50.68	119.04

Table 4: Real data ($J = 8000$ and $N = 4$): MAE for $\hat{f}_v^{(\mathrm{NIGP})}$, $\hat{f}_v^{(\mathrm{DP})}$ and $\hat{f}_v^{(\mathrm{CMS})}$

	20 Newsgroups			Enron		
Bins for true v	$\hat{f}_v^{(\mathrm{CMS})}$	$\hat{f}_v^{(\mathrm{DP})}$	$\hat{f}_v^{(\mathrm{NIGP})}$	$\hat{f}_v^{(\mathrm{CMS})}$	$\hat{f}_v^{(\mathrm{DP})}$	$\hat{f}_v^{(\mathrm{NIGP})}$
(0,1]	53.4	53.39	0.39	71.0	70.98	0.41
(1,2]	30.5	30.49	1.40	47.4	47.38	1.47
(2,4]	32.5	32.49	2.70	52.5	52.49	3.25
(4,8]	38.7	38.69	5.97	53.1	53.08	6.17
(8,16]	25.3	25.29	11.97	57.0	56.98	11.28
(16,32]	25.0	24.99	21.25	90.0	89.98	19.82
(32,64]	39.7	39.69	42.81	108.4	108.37	47.07
(64,128]	22.1	22.09	91.06	55.7	55.67	87.32
(128,256]	25.8	25.79	205.58	80.8	80.76	178.23