Inner spike and slab Bayesian nonparametric models

Antonio Canale [a], Antonio Lijoi [b], Bernardo Nipoti [c,*], Igor Prünster [b]

[a] *Department of Statistical Sciences, University of Padova, Via Cesare Battisti, Padova 241/243 35121, Italy*
[b] *Department of Decision Sciences and BIDSA, Bocconi University, via Röntgen 1, Milan 20136, Italy*
[c] *Department of Economics, Management and Statistics, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, Milan, 20126, Italy*

ARTICLE INFO

ABSTRACT

Discrete Bayesian nonparametric models whose expectation is a convex linear combination of a point mass at some point of the support and a diffuse probability distribution allow to incorporate strong prior information, while still being extremely flexible. Recent contributions in the statistical literature have successfully implemented such a modelling strategy in a variety of applications, including density estimation, nonparametric regression and model-based clustering. A thorough study is presented on a large class of nonparametric models, named *inner spike and slab hNRMI models* and obtained by considering homogeneous normalized random measures with independent increments (hNRMI) with base measure given by a convex linear combination of a point mass and a diffuse probability distribution. In turn, the distributional properties of these models are investigated, with focus on: i) the exchangeable partition probability function they induce, ii) the distribution of the number of distinct values in an exchangeable sample, iii) the posterior predictive distribution, and iv) the distribution of the number of elements that coincide with the only point of the support with positive probability. These theoretical findings represent the main building block for an actual implementation of Bayesian inner spike and slab hNRMI models by means of a generalized Pólya urn scheme.

1. Introduction

One of the most appealing aspects of Bayesian nonparametric modelling is its ability to flexibly account for a rich variety of patterns beyond those described by specific parametric models. In many applications, however, the need of flexibility comes along with the availability of valuable prior information on some attributes of the data. For example, when monitoring a process which varies over time, interest typically lies in discriminating between a known baseline behaviour and departures from the same. Similarly, when modelling functional data, it is often the case that some of the observations are expected to exhibit specific regular features while others might display more erratic traits. Both these situations call for a model which might be able to accurately capture the expected baseline behaviour while, at the same time, being flexible in dealing with the more irregular observations. Many recent contributions have formalized this idea by considering nonparametric mixture models defined so that the expected baseline behaviour is assigned positive prior probability. We review here two simple strategies that allow to achieve this goal, henceforth referred to as *inner* and *outer* spike and slab models, following a terminology introduced in Canale et al. (2017).

---

* Corresponding author.:
  E-mail address: bernardo.nipoti@unimib.it (B. Nipoti).

We consider an observation X taking values in some space $\mathbb{X}$ and we suppose that the baseline behaviour is described by a point $x_0 \in \mathbb{X}$. The outer spike and slab model is a convex linear combination of a Dirac delta at x_0 and a discrete nonparametric measure on $\mathbb{X}$. More specifically, the model assumes that $X \sim \tilde{Q}$, where $\tilde{Q}$ on $\mathbb{X}$ is defined as

$$\tilde{Q} = \zeta \delta_{x_0} + (1 - \zeta) Q^*, \tag{1}$$

where $\zeta \in (0, 1)$ represents the probability of the event $\{X = x_0\}$, and Q^* is a discrete random probability measure such that $\mathbb{E}[Q^*] = P^*$, with P^* a diffuse probability measure. Consequently, one has $\mathbb{E}[\tilde{Q}] = \zeta \delta_{x_0} + (1 - \zeta) P^*$. Alternatively, the inner spike and slab model is a discrete random probability measure $\tilde{P} = \sum_j \tilde{p}_j \delta_{Z_j}$ with random locations Z_j's on $\mathbb{X}$ specified in such a way that

$$\mathbb{E}[\tilde{P}] = \zeta \delta_{x_0} + (1 - \zeta) P^*, \tag{2}$$

where $\zeta \in (0, 1)$ can be interpreted, again, as the probability that X coincides with x_0 and P^* is a diffuse probability distribution on $\mathbb{X}$. The two models are characterized by analogous spike and slab structures (Mitchell and Beauchamp, 1988) and are such that $\mathbb{E}[\tilde{Q}] = \mathbb{E}[\tilde{P}] = \zeta \delta_{x_0} + (1 - \zeta) P^*$. Nonetheless, the outer and inner specifications differ in that the spike at x_0 is external to the nonparametric component of the model, in the former, as displayed by (1), while it is included in the nonparametric part, in the latter, with x_0 being the only atomic component of the otherwise nonatomic expected value (2). Inner and outer spike and slab specifications have been largely adopted in the recent literature. Cassese et al. (2019) model pneumonia and influenza mortality in space and time by adopting the outer spike and slab mixture approach described in (1), with Q^* distributed as a Dirichlet process (DP) (Ferguson, 1973). Similarly, D'Angelo et al. (2021) use the inner spike and slab approach to Bayesian nonparametric mixtures to analyze data on neural activity of animals, measured via calcium imaging. Another example is provided by Scarpa and Dunson (2009) and Canale et al. (2017), where the same functional data set on women fertility, is analysed, respectively, by means of outer and inner spike and slab mixture models defined on functional spaces. Similar approaches have been used also in the context of variable selection. Dunson et al. (2008), MacLehose et al. (2007), Yang (2012) and Barcella et al. (2016), for example, adopt a DP specified such that (2) holds with $x_0 = 0$, to simultaneously allow for variable selection and clustering of the variables. The same construction is adopted in Suarez and Ghosal (2016) to model wavelet coefficients of functional data so to induce sparsity. Models to infer differential gene expression based on a DP with inner spike and slab base measure, can be found in Do et al. (2005) and Guindani et al. (2014), while applications to multiple testing problems are proposed by Bogdan et al. (2008) and Kim et al. (2009).

Most of the aforementioned contributions consider inner spike and slab models based on a DP. In fact, when assuming $\tilde{P}$ is distributed as a DP, specifications leading to a spike at x_0 in (2) do not change the structure of the resulting predictive distribution thanks to the conjugacy of the DP (Ferguson, 1973). However, when $\tilde{P}$ is not a DP, specifications implying an expected value with atoms, such as the spike at x_0 in (2), considerably change the posterior predictive structure of the process (see Sangalli, 2006; Canale et al., 2017), thus implying challenging technical issues that need to be addressed in order to perform Bayesian posterior inference. On the other hand, working with outer spike and slab models is in general less cumbersome as the nonparametric component of the model has diffuse base measure, for which, thus, standard techniques can be used.

The inner spike and slab model has been thoroughly studied by Canale et al. (2017) in the case of $\tilde{P}$ being distributed as a Pitman–Yor (PY) process (Perman et al., 1992; Pitman and Yor, 1997). More recently, Bassetti and Ladelli (2020) investigated the asymptotic behaviour of species sampling models with non-diffuse base measure. In this paper we focus on the flexible class of nonparametric priors obtained by normalizing homogeneous completely random measures (CRMs) (Regazzini et al., 2003), henceforth called homogeneous normalized random measures with independent increments (hNRMIs). Their stick-breaking representation has been derived in Favaro et al. (2016). We study the distributional properties of hNRMIs with spike and slab specification (2), thus providing the essential building block for carrying out posterior inference with inner spike and slab hNRMI models. Our investigations also highlight that, besides the apparent similarities, as far as hNRMIs are concerned, the inner and the outer spike and slab models are structurally different. While having coinciding prior expectations, the two model specifications differ in terms of prior variability, with the variance turning out to be larger for the inner spike and slab model. In this sense, the inner spike and slab prior is less informative than the outer one. Our investigation on this point is underpinned by a numerical study, whose results are in line with the findings of Canale et al. (2017) for the PY case, where the inner approach is showed to appear more robust than the outer one, when dealing with misspecified priors.

The paper is organised as follows. In Section 2 we provide a succinct introduction to hNRMIs and compare the prior variance of functionals of inner and outer spike and slab models. Section 3 presents a thorough investigation of the distributional properties of the inner spike and slab hNRMI model. As notable examples, in Section 4, the main results of the paper are displayed for the special cases of σ-stable and normalized inverse Gaussian hNRMIs. Section 5 provides a generalised Pólya urn scheme for the inner spike and slab hNRMI model. This is then implemented, for the σ-stable case, to allow for a numerical comparison of inner and outer spike and slab models, both *a priori* and *a posteriori*. Finally, the proofs of the main results are reported in Section 6.

2. Homogeneous normalized random measures and spike and slab models

We concisely recall the basics of CRMs and hNRMIs tailored to the present contribution. For an account on their role in Bayesian nonparametric statistics one can refer to Lijoi and Prünster (2010). A random measure $\tilde{\mu}$ on $\mathbb{X}$ is said *completely random* if for any collection of pairwise disjoint measurable subsets $A_1, \ldots, A_k$ of $\mathbb{X}$, and for any $k \geq 2$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_k)$ are mutually independent. If $\tilde{\mu}$ is without fixed points of discontinuity, which is assumed throughout the paper, for any function $f : \mathbb{X} \to \mathbb{R}^+$ one has

$$\mathbb{E}\left[e^{-\int_{\mathbb{X}} f(x)\tilde{\mu}(\mathrm{d}x)}\right] = \exp\left\{-\int_{\mathbb{R}^+ \times \mathbb{X}} \left(1 - e^{-sf(x)}\right)\nu(\mathrm{d}s, \mathrm{d}x)\right\},$$

where ν, a measure on $\mathbb{R}^+ \times \mathbb{X}$, is the Lévy intensity characterizing $\tilde{\mu}$. For the ease of illustration, henceforth we consider the class of almost surely finite homogeneous CRMs, corresponding to Lévy intensities that admit a factorization of the type $\nu(\mathrm{d}s, \mathrm{d}x) = \rho(s)\mathrm{d}s c P_0(\mathrm{d}x)$ for some measurable function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ and some constant $c > 0$. Noteworthy examples that will be considered in this paper are the σ-stable process and the inverse Gaussian process, characterized respectively by $\rho(s) = \sigma s^{-1-\sigma}/\Gamma(1-\sigma)$, for some $\sigma \in (0, 1)$, and $\rho(s) = s^{-3/2}e^{-\tau s}/(2\sqrt{\pi})$, for some $\tau > 0$. A random probability measure is then obtained by normalization as $\tilde{P} = \tilde{\mu}/\tilde{\mu}(\mathbb{X})$ and denoted by $\tilde{P} \sim \mathrm{hNRMI}(\rho; c, P_0)$. The study of this class of nonparametric priors was first considered in Regazzini et al. (2003). The special cases considered in this paper are two popular nonparametric priors, namely the σ-stable hNRMI and the normalized inverse Gaussian (N-IG) process (Lijoi et al., 2005), which are obtained by normalizing a σ-stable CRM and an inverse Gaussian process, respectively. See e.g. Lijoi and Prünster (2010) and Barrios et al. (2013) for a review of their inferential properties; and Corradin et al. (2021) for a recent example of their use in applications.

We are now in the position to highlight a structural difference between inner and outer hNRMI spike and slab models. Specifically, we let $\tilde{P} \sim \mathrm{hNRMI}(\rho; c, P_0)$ with spike and slab base measure

$$P_0 = \zeta \delta_{x_0} + (1 - \zeta)P^*. \tag{3}$$

This defines an inner spike and slab model and implies $\mathbb{E}[\tilde{P}] = P_0$ (James et al., 2006). Moreover, we consider an outer spike and slab model $\tilde{Q}$, defined as in (1) with $Q^* \sim \mathrm{hNRMI}(\rho; c, P^*)$. While the two random probability measures have the same expectation, namely $\zeta \delta_{x_0} + (1 - \zeta)P^*$, they differ in terms of variance, as displayed in the next proposition. Before stating the result, we define $\tau_q(u) = \int_0^\infty s^q e^{-us}\rho(s)\mathrm{d}s$, for any integer $q \geq 1$, and $\psi(u) = \int_0^\infty (1 - e^{-us})\rho(s)\mathrm{d}s$, for any $u > 0$. Moreover, for any probability measure P, we use the notation $P(f) = \int f \, \mathrm{d}P$.

Proposition 1. *Let $f : \mathbb{X} \to \mathbb{R}$ any measurable function such that $P^*(f^2) < \infty$. Then*

$$\mathrm{var}(\tilde{P}(f)) - \mathrm{var}(\tilde{Q}(f)) = p\,\zeta\,(1-\zeta)\,P^*\left(\{f - f(x_0)\}^2\right) \geq 0,$$

where $p = c \int_0^\infty u e^{-c\psi(u)}\tau_2(u)\mathrm{d}u \in (0, 1)$.

Proposition 1 indicates that, for the class of hNRMIs, the inner spike and slab model is characterized by larger prior uncertainty than the outer spike and slab model with the same expectation. In other terms, while both models are centered at the same prior guess, the inner specification of the model is less informative. This aspect will be further investigated in the numerical study presented in Section 5.

3. Main results

We now investigate the distributional properties of $\tilde{P} \sim \mathrm{hNRMI}(\rho; c, P_0)$, where P_0 is a spike and slab base measure defined as in (3). Our results include: i) the exchangeable partition probability function (EPPF) induced by $\tilde{P}$, that is the probability of observing a specific sample displaying k distinct values, henceforth also referred to as clusters, with corresponding frequencies summarized by the vector $(n_1, \ldots, n_k)$; ii) the distribution of K_n, the number of distinct values in an exchangeable sample $X^{(n)} = (X_1, \ldots, X_n)$ such that $X_i \mid \tilde{P} \overset{\mathrm{iid}}{\sim} \tilde{P}$; iii) the predictive distribution for one observation in the exchangeable sample, conditionally on the observation of the others; iv) the distribution of $N_0^{(n)}$, that is the number of elements of $X^{(n)}$ that coincide with x_0. Henceforth we use the notation $\Pi_k^{(n)}(n_1, \ldots, n_k; \zeta)$ to denote the EPPF induced by a hNRMI with spike and slab base measure (3) parametrized by $\zeta \in (0, 1)$. Similarly, we will write $\mathrm{Pr}(\cdot; \zeta)$ to stress the fact that the probability of an event depends on the parameter ζ. Accordingly, the notation $\Pi_k^{(n)}(n_1, \ldots, n_k; 0)$ and $\mathrm{Pr}(\cdot; 0)$ will refer to the EPPF and the probability measure induced by a hNRMI with diffuse base measure P^*. Finally, for any positive integer n, and any $i = 1, \ldots, n$, we define the function

$$\xi_{n,i}(u) = \frac{1}{i!}\sum_{j=0}^{i}(-1)^{n-j}\binom{i}{j}\psi^{i-j}(u)\frac{\mathrm{d}^n}{\mathrm{d}u^n}\left[\psi^j(u)\right]. \tag{4}$$

Theorem 1. *If $X_i \mid \tilde{P} \overset{\mathrm{iid}}{\sim} \tilde{P}$, for $i = 1, \ldots, n$ and for any $n \geq 1$, and $\tilde{P} \sim \mathrm{hNRMI}(\rho; c, P_0)$, with P_0 as in (3), the EPPF induced by $\tilde{P}$ is*

$$\Pi_k^{(n)}(n_1, n_2, \ldots, n_k; \zeta) = \frac{1}{\Gamma(n)} \left\{ c^k (1-\zeta)^k \int_0^\infty u^{n-1} e^{-c\psi(u)} \prod_{m=1}^k \tau_{n_m}(u) du \right.$$

$$\left. + c^{k-1} (1-\zeta)^{k-1} \sum_{l=1}^k \sum_{i=1}^{n_l} c^i \zeta^i \int_0^\infty u^{n-1} e^{-c\psi(u)} \xi_{n_{l,i}}(u) \prod_{m \neq l} \tau_{n_m}(u) du \right\}, \quad (5)$$

where we agree that $\prod_{m \neq \ell} \tau_{n_m}(u) \equiv 1$ when $k = 1$.

Out of the $k+1$ summands on the right-hand side of (5), the first one refers to the case in which none of the blocks of the partition coincide with x_0, while the remaining k terms account for the cases in which the l-th cluster is identified by the atom x_0, for $l = 1, 2, \ldots, k$.

We recall that the EPPF of a hNRMI with diffuse base measure P^* equals

$$\Pi_k^{(n)}(n_1, \ldots, n_k; 0) = \frac{c^k}{\Gamma(n)} \int_0^\infty u^{n-1} e^{-c\psi(u)} \prod_{j=1}^k \tau_{n_j}(u) du. \qquad (6)$$

Henceforth, for any $n_j \in \{1, \ldots, n-k+1\}$, let $\mathcal{X}_{k-1,0}^{(n-n_j)}(\mathbf{m}n_{-j})$ denote a sample of size $n - n_j$ clustered into $k-1$ groups with respective frequencies $\mathbf{m}n_{-j} = (n_1, \ldots, n_{j-1}, n_{j+1}, \ldots, n_k)$ such that: (i) none of its elements equals x_0 and (ii) when extended to a sample of size n, the overall number of observations that coincide with x_0 equals n_j, namely $N_0^{(n)} = \text{card}\{i : X_i = x_0\} = n_j$. One can then state the following result.

Corollary 1. *If $X_i \mid \tilde{P} \overset{\text{iid}}{\sim} \tilde{P}$, for $i = 1, \ldots, n$ and for any $n \geq 1$, and $\tilde{P} \sim \text{hNRMI}(\rho; c, P_0)$, with P_0 as in (3), the EPPF induced by $\tilde{P}$ is*

$$\Pi_k^{(n)}(n_1, \ldots, n_k; \zeta) = (1-\zeta)^k \Pi_k^{(n)}(n_1, \ldots, n_k; 0)$$

$$+ (1-\zeta)^{k-1} \sum_{j=1}^k \Pi_{k-1}^{(n-n_j)}(n_1, \ldots, n_{j-1}, n_{j+1}, \ldots, n_k; 0) \Pr(N_0^{(n)} = n_j \mid \mathcal{X}_{k-1,0}^{(n-n_j)}(\mathbf{m}n_{-j}); \zeta). \quad (7)$$

It is apparent that the first summand in (7) accounts for the case where none of the k clusters is identified by x_0, so one has the standard EPPF that corresponds to a diffuse base measure. On the other hand, the second summand takes into account the possibility that x_0 identifies one of the k clusters in the partition.

Starting from Theorem 1 we obtain the distribution of K_n, the number of distinct values in the sample $X^{(n)}$, as reported in the next result. To this end, for any $a, u > 0$ we will denote with $X_{a,u}$ a random variable such that its density function is $f_{a,u}(s) \propto s^a e^{-us} \rho(s) \mathbb{1}_{(0,+\infty)}(s)$. It can be seen that if the following condition holds true

(H1) ρ is such that for any finite collection of independent random variables $X_{a_1,u}, \ldots, X_{a_n,u}$, the distribution of $\sum_{i=1}^n X_{a_i,u}$ depends on $(a_1, \ldots, a_n)$ only through $\sum_{i=1}^n a_i$,

then $\Pr(N_0^{(n)} = r \mid \mathcal{X}_{k-1,0}^{(n-r)}(\mathbf{m}n_{-j}); \zeta) = \Pr(N_0^{(n)} = r \mid K_{n-r} = k-1; \zeta)$. This is used to prove the following theorem.

Theorem 2. *Let $X_i \mid \tilde{P} \overset{\text{iid}}{\sim} \tilde{P}$, for $i = 1, \ldots, n$ and for any $n \geq 1$, and $\tilde{P} \sim \text{hNRMI}(\rho; c, P_0)$, with P_0 as in (3). Moreover, ρ is such that (H1) holds true. Then, for any $k \in \{1, \ldots, n\}$, the distribution of K_n, parametrized by $\zeta \in (0, 1)$, is given by*

$$\Pr(K_n = k; \zeta) = (1-\zeta)^k \Pr(K_n = k; 0)$$

$$+ (1-\zeta)^{k-1} \sum_{r=1}^{n-k+1} \binom{n}{r} \Pr(K_{n-r} = k-1; 0) \Pr(N_0^{(n)} = r \mid K_{n-r} = k-1; \zeta).$$

It is worth stressing that for all the examples we consider henceforth condition (H1) holds true. In view of these findings, the predictive distributions associated to the sequence $(X_i)_{i \geq 1}$ can now be easily determined. We suppose that the observed sample $X^{(n)}$ displays k distinct values $x_1^*, \ldots, x_k^*$, with respective frequencies $n_1, \ldots, n_k$, and state the following theorem.

Theorem 3. *If $X_i \mid \tilde{P} \overset{\text{iid}}{\sim} \tilde{P}$, for $i = 1, \ldots, n$ and for any $n \geq 1$, and $\tilde{P} \sim \text{hNRMI}(\rho; c, P_0)$, with P_0 as in (3), the predictive distribution of X_{n+1} conditionally on $X^{(n)}$ is*

(i) *if $x_0 \notin \{x_1^*, \ldots, x_k^*\}$,*

$$\Pr(X_{n+1} \in A \mid X^{(n)}; \zeta) = \frac{c}{n} \frac{\int_0^\infty u^n e^{-c\psi(u)} \tau_1(u) \prod_{m=1}^k \tau_{n_m}(u) du}{\int_0^\infty u^{n-1} e^{-c\psi(u)} \prod_{m=1}^k \tau_{n_m}(u) du} P_0(A)$$

$$+ \frac{1}{n} \sum_{l=1}^k \frac{\int_0^\infty u^n e^{-c\psi(u)} \tau_{n_l+1}(u) \prod_{m \neq l} \tau_{n_m}(u) du}{\int_0^\infty u^{n-1} e^{-c\psi(u)} \prod_{m=1}^k \tau_{n_m}(u) du} \delta_{x_l^*}(A); \qquad (8)$$

(ii) if $x_0 = x_j^*$ for some $j = 1, \ldots, k$,

$$
\begin{aligned}
\Pr(X_{n+1} \in A \mid X^{(n)}; \zeta) &= \frac{c(1-\zeta)}{n} \frac{\sum_{i=1}^{n_j} c^i \zeta^i \int_0^\infty u^n e^{-c\psi(u)} \tau_1(u) \xi_{n_j,i}(u) \prod_{m \neq j} \tau_{n_m}(u) du}{\sum_{i=1}^{n_j} c^i \zeta^i \int_0^\infty u^{n-1} e^{-c\psi(u)} \xi_{n_j,i}(u) \prod_{m \neq j} \tau_{n_m}(u) du} P^*(A) \\
&+ \frac{1}{n} \frac{\sum_{i=1}^{n_j+1} c^i \zeta^i \int_0^\infty u^n e^{-c\psi(u)} \xi_{n_j+1,i}(u) \prod_{m \neq j} \tau_{n_m}(u) du}{\sum_{i=1}^{n_j} c^i \zeta^i \int_0^\infty u^{n-1} e^{-c\psi(u)} \xi_{n_j,i}(u) \prod_{m \neq j} \tau_{n_m}(u) du} \delta_{x_0}(A) \\
&+ \frac{1}{n} \sum_{l \neq j} \frac{\sum_{i=1}^{n_j} c^i \zeta^i \int_0^\infty u^n e^{-c\psi(u)} \tau_{n_l+1}(u) \xi_{n_j,i}(u) \prod_{m \neq l,j} \tau_{n_m}(u) du}{\sum_{i=1}^{n_j} c^i \zeta^i \int_0^\infty u^{n-1} e^{-c\psi(u)} \xi_{n_j,i}(u) \prod_{m \neq j} \tau_{n_m}(u) du} \delta_{x_l^*}(A).
\end{aligned}
\tag{9}
$$

Note that, if $x_0 \notin \{x_1^*, \ldots, x_k^*\}$, the form of the predictive distribution in (8) coincides with that one of the predictive distribution of a hNRMI with diffuse base measure (see James et al., 2006, Corollary 1).

We complete this section by studying the distribution of $N_0^{(n)}$, the number of elements, in an exchangeable sample $X^{(n)}$ such that $X_i \mid \tilde{P} \overset{\text{iid}}{\sim} \tilde{P}$, that coincide with x_0.

Theorem 4. Let $X_i \mid \tilde{P} \overset{\text{iid}}{\sim} \tilde{P}$, for $i = 1, \ldots, n$ and for any $n \geq 1$. If $\tilde{P} \sim \text{hNRMI}(\rho; c, P_0)$, with P_0 as in (3), then the distribution of $N_0^{(n)}$, parametrized by $\zeta \in (0, 1)$, is given, for any $j \in \{0, 1, \ldots, n\}$, by

$$
\Pr(N_0^{(n)} = j; \zeta) = \frac{1}{\Gamma(n)} \sum_{k=1}^{n-j+1-\delta_{0j}} k^{1-\delta_{0j}} c^{k+\delta_{0j}-1} (1-\zeta)^{k+\delta_{0j}-1}
$$

$$
\times \sum_{i=0}^{j} c^i \zeta^i \int_0^\infty u^{n-1} e^{-c\psi(u)} \xi_{j,i}(u) \frac{1}{(k+\delta_{0j}-1)!} \sum_{\mathbf{m}n \in \mathcal{N}_{k+\delta_{0j}-1}^{(n-j)}} \binom{n}{n_1 \cdots n_{k+\delta_{0j}-1}} \prod_{m=1}^{k+\delta_{0j}-1} \tau_{n_m}(u) du,
\tag{10}
$$

where $\mathcal{N}_k^{(n)} = \{\mathbf{m}n = (n_1, \ldots, n_k) : n_m \in \mathbb{N} \text{ and } \sum_{m=1}^k n_m = n\}$, and $\delta_{0j} = 1$ if $j = 0$ and $\delta_{0j} = 0$ otherwise.

The expressions obtained for the case of general hNRMIs with spike and slab base measure showcase how the techniques developed in the paper work in great generality for a large class of inner spike and slab models. In order to make such expressions amenable of direct application, special cases of the class of hNRMIs must be considered, as illustrated in the next section.

4. Notable special cases

Here we specialize the results of Section 3 to two popular special cases, namely the σ-stable hNRMI and the N-IG process. This allows us to display that the expressions we obtained for the EPPF, the predictive distributions, and the distribution of $N_0^{(n)}$, for the general case of hNRMIs, reduce to tractable expressions when special cases within the same family are considered. Moreover, the σ-stable hNRMI is the only random probability measure, along with the DP, which is at the same time a special case of the Pitman-Yor process and an element of the class of NRMIs. The last observation allows us to link the results of this paper with those in Canale et al. (2017).

4.1. σ-stable hNRMI

Assume that $\tilde{P}$ is a σ-stable hNRMI, with $\sigma \in (0, 1)$, with spike and slab base measure. In other terms, $\tilde{P} \sim \text{hNRMI}(\rho; 1, P_0)$ with $\rho(s) = \sigma s^{-1-\sigma}/\Gamma(1-\sigma)$, c set equal to 1, and P_0 defined as in (3). Such choice implies that $\psi(u) = u^\sigma$, $\tau_q(u) = u^{-q+\sigma} \sigma(1-\sigma)_{q-1}$, and $\xi_{n,i}(u) = u^{\sigma i - n} \mathscr{C}(n, i; \sigma)$, where $\mathscr{C}(n, i; \sigma) = \frac{1}{i!} \sum_{r=0}^i (-1)^r \binom{i}{r} (-r\sigma)_n$ is the generalized factorial coefficient (Charalambides, 2005), and $(a)_n = \Gamma(a+n)/\Gamma(a)$. These expressions for $\psi(u)$, $\tau_n(u)$ and $\xi_{n,i}(u)$, when plugged into (5), (8), (9) and (10), provide the EPPF, the prediction rule, and the distribution of $N_0^{(n)}$, implied by the σ-stable hNRMI $\tilde{P}$. For the sake of compactness we introduce the numbers $\varphi_{m,q}$, defined, for any positive integers m and q, as

$$
\varphi_{m,q}(\zeta) = \sum_{i=1}^m \zeta^i \Gamma(q+i-1) \mathscr{C}(m, i; \sigma).
$$

In the following, for the sake of simplicity, we will omit the dependence on ζ and write $\varphi_{m,q}$ instead of $\varphi_{m,q}(\zeta)$. The EPPF can then be written as

$$
\Pi_k^{(n)}(n_1, \ldots, n_k; \zeta) = \frac{\sigma^{k-2}(1-\zeta)^{k-1}}{\Gamma(n)} \prod_{m=1}^k (1-\sigma)_{n_m-1} \left(\sigma(1-\zeta)\Gamma(k) + \sum_{l=1}^k \frac{\varphi_{n_l,k}}{(1-\sigma)_{n_l-1}} \right).
$$

The predictive distributions of Theorem 3 reduce to:

(i) if $x_0 \notin \{x_1^*, \ldots, x_k^*\}$,

$$\Pr(X_{n+1} \in A \mid X^{(n)}; \zeta) = \frac{k\sigma}{n} P_0(A) + \frac{1}{n} \sum_{l=1}^{k} (n_l - \sigma) \delta_{x_l^*}(A); \tag{11}$$

(ii) if $x_0 = x_j^*$ for some $j = 1, \ldots, k$,

$$\Pr(X_{n+1} \in A \mid X^{(n)}; \zeta) = \frac{(1 - \zeta)\sigma}{n} \frac{\varphi_{n_j, k+1}}{\varphi_{n_j, k}} P^*(A) + \frac{1}{n} \frac{\varphi_{n_j+1, k}}{\varphi_{n_j, k}} \delta_{x_0}(A) + \frac{1}{n} \sum_{l \neq j} (n_l - \sigma) \delta_{x_l^*}(A). \tag{12}$$

We observe that (11) has the exact same structure of the predictive distribution of a σ-stable hNRMI with diffuse base measure. The same does not hold for the predictive distribution in (12), which accounts for the fact that X_{n+1} can coincide with x_0 either when it is drawn from the spike component of the base measure P_0 or when it is tied to x_j^*. It is interesting to observe that, by setting $A = \mathbb{X}$ in (12), the next triangular identity is obtained

$$\varphi_{m,q}(m + (q - 1)\sigma) = \varphi_{m+1, q} + (1 - \zeta)\sigma \varphi_{m, q+1},$$

for any positive integers m and q. This, combined with the fact that $\varphi_{1, q} = \zeta \sigma \Gamma(q)$, for any positive integer q, suggests a recursive strategy for efficiently evaluate the numbers $\varphi_{m,q}$, for the desired values of m and q.

Finally, the distribution of $N_0^{(n)}$ provided in Theorem 4 becomes

$$\Pr(N_0^{(n)} = j; \zeta) = \frac{1}{\sigma \Gamma(n)} \sum_{k=1}^{n-j+1-\delta_{0j}} k^{1-\delta_{0j}} (1 - \zeta)^{k+\delta_{0j}-1} \mathscr{C}(n - j, k + \delta_{0j} - 1; \sigma) \sum_{i=0}^{j} \zeta^i \Gamma(k + \delta_{0j} + i - 1) \mathscr{C}(j, i; \sigma), \tag{13}$$

for any $j \in \{0, 1, \ldots, n\}$.

4.2. Normalized inverse Gaussian process

Assume that $\tilde{P}$ is a N-IG process, with $\tau > 0$ and spike and slab base measure. In other terms, $\tilde{P} \sim \text{hNRMI}(\rho; c, P_0)$ with $\rho(s) = s^{-3/2} e^{-\tau s}/(2\sqrt{\pi})$, $c > 0$ and P_0 defined as in (3). Such choice implies that $\psi(u) = \sqrt{\tau + u} - \sqrt{\tau}$, $\tau_q(u) = (\tau + u)^{1/2-q} \Gamma(q - 1/2)/(2\sqrt{\pi})$, and $\xi_{n,i}(u) = (\tau + u)^{i/2-n} \mathscr{C}(n, i; 1/2)$. These expressions for $\psi(u)$, $\tau_n(u)$ and $\xi_{n,i}(u)$, when plugged into (5), (8), (9), and (10), provide the EPPF, the prediction rule, and the distribution of $N_0^{(n)}$, implied by a normalized inverse Gaussian process. For the sake of compactness, we introduce the notation $\beta = c\sqrt{\tau}$ and define the numbers $\varrho_{m,q}^{(n)}$ as follows. For any $n \geq q$ positive integers,

$$\varrho_{0,q}^{(n)} = \sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^r \beta^{2r} \Gamma(q - 2r, \beta),$$

where $\Gamma(n, z) = \int_z^\infty x^{n-1} e^x dx$ denotes the upper incomplete gamma function; moreover, for any m, q and n positive integers such that $n \geq q$ and $m \leq n - (q - 1)$,

$$\varrho_{m,q}^{(n)}(\zeta) = \sum_{i=1}^{m} (2\zeta)^i \frac{\Gamma(2m - i)}{\Gamma(m + 1 - i)\Gamma(i)} \varrho_{0, q+i-1}^{(n)}.$$

With the aim of simplifying the notation, in the following we will omit the dependence on ζ and write $\varrho_{m,q}^{(n)}$ instead of $\varrho_{m,q}^{(n)}(\zeta)$. The EPPF can then be written as

$$\Pi_k^{(n)}(n_1, n_2, \ldots, n_k; \zeta) = \frac{e^\beta (1 - \zeta)^{k-1}}{2^{k-1} \pi^{k/2} \Gamma(n)} \prod_{m=1}^{k} \Gamma(n_m - 1/2) \left((1 - \zeta)\varrho_{0,k}^{(n)} + \sqrt{\pi} \sum_{l=1}^{k} \frac{2^{1-2n_l}}{\Gamma(n_l - 1/2)} \varrho_{n_l, k}^{(n)} \right).$$

The predictive distribution of Theorem 3 simplifies to:

(i) if $x_0 \notin \{x_1^*, \ldots, x_k^*\}$,

$$\Pr(X_{n+1} \in A \mid X^{(n)}; \zeta) = \frac{1}{2n} \frac{\varrho_{0,k+1}^{(n+1)}}{\varrho_{0,k}^{(n)}} P_0(A) + \frac{1}{n} \frac{\varrho_{0,k}^{(n+1)}}{\varrho_{0,k}^{(n)}} \sum_{l=1}^{k} (n_l - 1/2) \delta_{x_l^*}(A); \tag{14}$$

(ii) if $x_0 = x_j^*$ for some $j = 1, \ldots, k$,

$$\Pr(X_{n+1} \in A \mid X^{(n)}; \zeta) = \frac{(1 - \zeta)}{2n} \frac{\varrho_{n_j,k+1}^{(n+1)}}{\varrho_{n_j,k}^{(n)}} P^*(A) + \frac{1}{4n} \frac{\varrho_{n_j+1,k}^{(n+1)}}{\varrho_{n_j,k}^{(n)}} \delta_{x_0}(A) + \frac{\varrho_{n_j,k}^{(n+1)}}{\varrho_{n_j,k}^{(n)}} \sum_{l \neq j} \frac{(n_l - 1/2)}{n} \delta_{x_l^*}(A). \tag{15}$$

We observe that (14) displays the standard structure of the predictive distribution of a N-IG process with diffuse base measure (see Lijoi et al., 2005). The same does not hold for the predictive distribution in (15), which accounts for the fact that X_{n+1} can coincide with x_0 either when it is drawn from the spike component of the base measure P_0 or when it is tied to x_j^*.

It is interesting to observe that, by setting $A = \mathbb{X}$ in (14), the next triangular identity is obtained

$$2n\varrho_{0,q}^{(n)} = \varrho_{0,q+1}^{(n+1)} + (2n-q)\varrho_{0,q}^{(n+1)},$$

for any positive integers $n \geq q$. In a similar fashion, by setting $A = \mathbb{X}$ in (15), one can show that the identity

$$4n\varrho_{m,q}^{(n)} = 2(1-\zeta)\varrho_{m,q+1}^{(n+1)} + \varrho_{m+1,q}^{(n+1)} + 2(2(n-m)-q+1)\varrho_{m,q}^{(n+1)}$$

holds true for any positive integers n, m, q such that $q \leq n$ and $1 \leq m \leq n - q + 1$.

Finally, the distribution of $N_0^{(n)}$ provided in Theorem 4 becomes

$$\Pr(N_0^{(n)} = j; \zeta) = \frac{2e^\beta}{\Gamma(n)} \sum_{k=1}^{n-j+1-\delta_{0j}} k^{1-\delta_{0j}} (1-\zeta)^{k+\delta_{0j}-1} \mathscr{C}(n-j, k+\delta_{0j}-1; 1/2) \sum_{i=0}^{j} \zeta^i \mathscr{C}(j, i; 1/2) \varrho_{0,k+i-1}^{(n)}, \tag{16}$$

for any $j \in \{0, 1, \ldots, n\}$.

5. Generalized Pólya urn scheme and numerical study

In this section we describe the steps of a generalized Pólya urn scheme which, under the assumption that $X_i \mid \tilde{P} \overset{\text{iid}}{\sim} \tilde{P}$, with $\tilde{P}$ distributed as a hNRMI with spike and slab base measure (3), allows us to generate a sample $X_{n+1}, \ldots, X_{n+m}$, conditionally on the observation of $(x_1, \ldots, x_n)$. The steps are summarized in Algorithm 1.

Algorithm 1: Generalised Pólya urn scheme for hNRMI inner spike and slab model.

Set $r = 1$;
while $r \leq m$ **do**

    let k be the number of distinct values $(x_1^*, \ldots, x_k^*)$ in $(x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+r-1})$, and let $(n_1, \ldots, n_k)$ be the corresponding frequencies;

    if $x_0 \neq x_j^*$ *for every* $j \in \{1, \ldots, k\}$ **then**

        generate a realization x_{n+r} of X_{n+r}, conditionally on the observation of $X^{(n+r-1)} = (x_1, x_2, \ldots, x_{n+r-1})$, from the predictive distribution (**??**);

    else

        generate a realization x_{n+r} of X_{n+r}, conditionally on the observation of $X^{(n+r-1)} = (x_1, x_2, \ldots, x_{n+r-1})$, from the predictive distribution (**??**);

   increase r by 1;

return $(x_{n+1}, \ldots, x_{n+m})$

While such a scheme can be adopted, in line of principle, for any specification of the Lévy intensity ρ, the results displayed in Section 4 allow for a direct implementation of this scheme for $\tilde{P}$ distributed as a σ-stable hNRMI or a N-IG process.

Next we implement the scheme we devised, in order to complement with a simulation study the findings of Proposition 1 on the different features of inner and outer spike and slab models. For the purpose of illustration, we focus on the case of σ-stable hNRMIs. In this study we investigate the variability of both the prior and the posterior distributions obtained by considering inner and outer spike and slab models, $\tilde{P}$ and $\tilde{Q}$ respectively, for different specifications of the parameters ζ and σ.

5.1. Inner vs outer models a priori

When analysing data, prior predictive checks represent a very effective, yet easy to implement, strategy to get a handle on the implications of a specific prior choice on functionals of interest. With this in mind, we study and compare the prior behaviour of inner and outer spike and slab models by generating, under a range of settings, realizations $(x_1, \ldots, x_m)$ of the random vector $(X_1, \ldots, X_m)$, modelled with either $\tilde{P}$ or $\tilde{Q}$. We then consider the corresponding empirical cumulative distribution function $F_m(x) = m^{-1} \sum_{i=1}^{m} \mathbb{1}_{(-\infty, x_i]}(x)$, realization of the random cumulative distribution function $\tilde{F}_m(x) = m^{-1} \sum_{i=1}^{m} \mathbb{1}_{(-\infty, X_i]}(x)$. We replicate the same experiment 100 000 times, which allows us to quantify, for any $x' \in \mathbb{R}$, the uncertainty associated with the random variable $\tilde{F}_m(x')$, by means of a Monte Carlo estimate of a quantile-based 95% credible interval. While both inner and outer spike and slab model are such that $\mathbb{E}[\tilde{F}_m(x)] = F_0(x)$, where F_0 denotes the cumulative distribution function corresponding to the spike and slab base measure (3), our simulation compares the associated uncertainties and investigates the role played by the parameters ζ and σ.
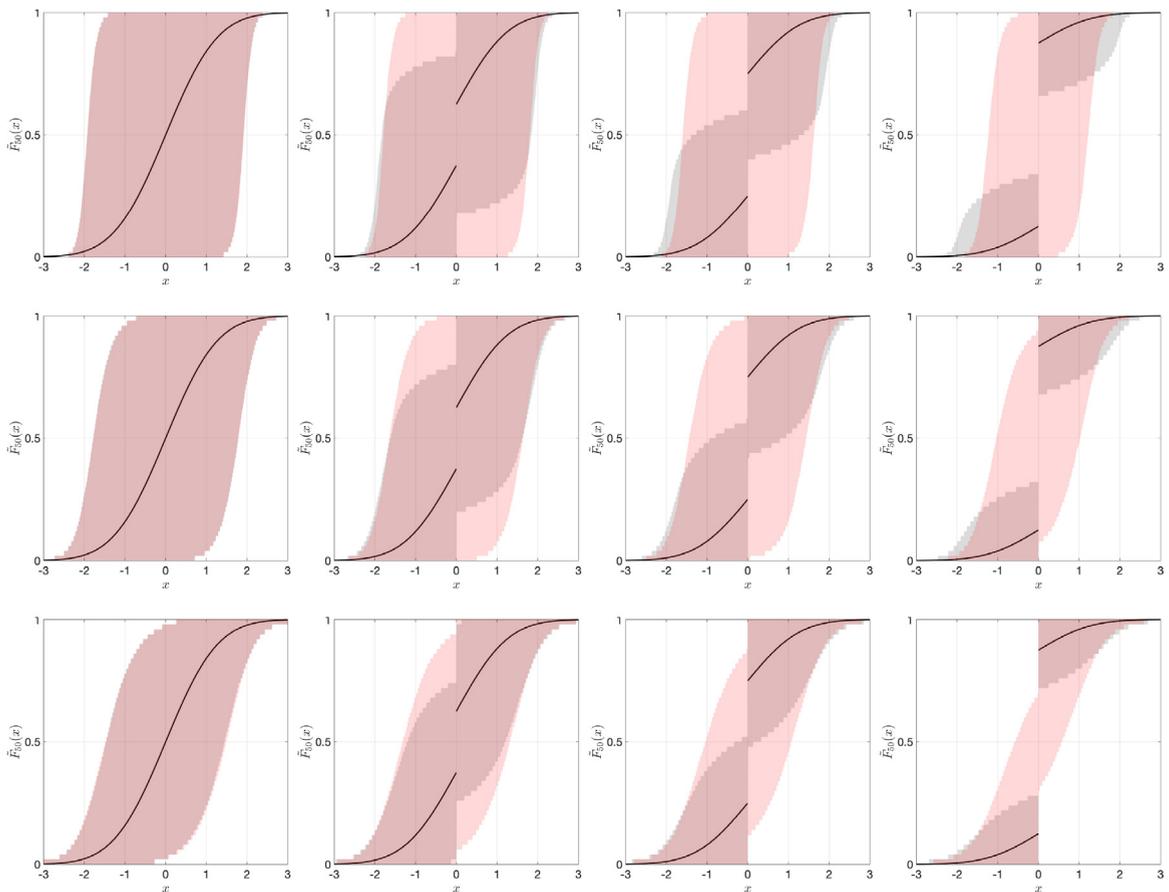
Fig. 1. Comparison of the estimated quantile-based 95% credible bands for the prior distributions of $\tilde{F}_{50}$ induced by inner (red shades) and outer (grey shades) spike and slab models for σ-stable hNRMIs, along with its expected value (black curves) $\mathbb{E}[\tilde{F}_{50}] = F_0$. Different rows refer to different values of $\sigma \in \{0.25, 0.5, 0.75\}$ (from top to bottom), different columns refer to different values of $\zeta \in \{0, 0.25, 0.5, 0.75\}$ (from left to right).

Table 1
Length of the estimated quantile-based 95% credible intervals for the distributions of M_{50} and med_{50}, induced by the inner and out spike and slab models for σ-stable hNRMIs, for different values of ζ and σ.

σ		0.25				0.5				0.75			
ζ		0	0.25	0.5	0.75	0	0.25	0.5	0.75	0	0.25	0.5	0.75
M_{50}	inner	3.48	3.22	2.86	2.14	2.93	2.67	2.31	1.65	2.20	1.95	1.59	1.10
	outer	3.48	2.60	1.77	0.92	2.94	2.21	1.49	0.78	2.17	1.64	1.12	0.59
med_{50}	inner	3.84	3.58	3.22	2.44	3.55	3.23	2.86	2.04	2.96	2.63	2.10	1.14
	outer	3.84	3.66	2.54	0.00	3.56	3.27	1.67	0.00	2.91	2.39	0.51	0.00

Let P^* be a standard normal distribution, $x_0 = 0$, and set $m = 50$. We consider $\zeta \in \{0, 0.25, 0.5, 0.75\}$ and $\sigma \in \{0.25, 0.5, 0.75\}$. For each of the resulting 12 combinations of parameter values, Figure 1 displays the expected value F_0 of $\tilde{F}_{50}$, as well as the estimated quantile-based 95% credible bands for the inner and the outer spike and slab models. First we observe that, when $\zeta = 0$, the two models coincide and boil down to a σ-stable hNRMI with diffuse base measure P^*. On the other hand, it can be appreciated that, when $\zeta > 0$, the variability of $\tilde{F}_{50}$ appears larger for the inner specification of the spike and slab model, such difference being more evident for larger values of ζ. The same behaviour can be appreciated across different values of the parameter σ, with larger values of σ leading to an overall smaller prior variability for both models. The structural difference between $\tilde{P}$ and $\tilde{Q}$ is further highlighted by looking at the distribution of two functionals of $\tilde{F}_{50}$, namely the mean $M_{50} = \sum_{i=1}^{50} X_i / 50$ and the median $\mathrm{med}_{50} = \mathrm{median}(X_1, \ldots, X_{50})$. Table 1 reports the length of the estimated 95% quantile-based credible intervals for M_{50} and for med_{50}, for inner and outer spike and slab models for the same combination of values for ζ and σ considered above. While all the estimated intervals are approximately centered at $x_0 = 0$, their lengths vary considerably. For both M_{50} and med_{50}, it can be appreciated that, for a fixed ζ, larger values of σ correspond to smaller intervals; similarly, for a fixed σ, larger values of ζ display smaller intervals. Intervals corresponding to inner and outer spike and slab models coincide when $\zeta = 0$, while the latter tend to be smaller for larger values of ζ.
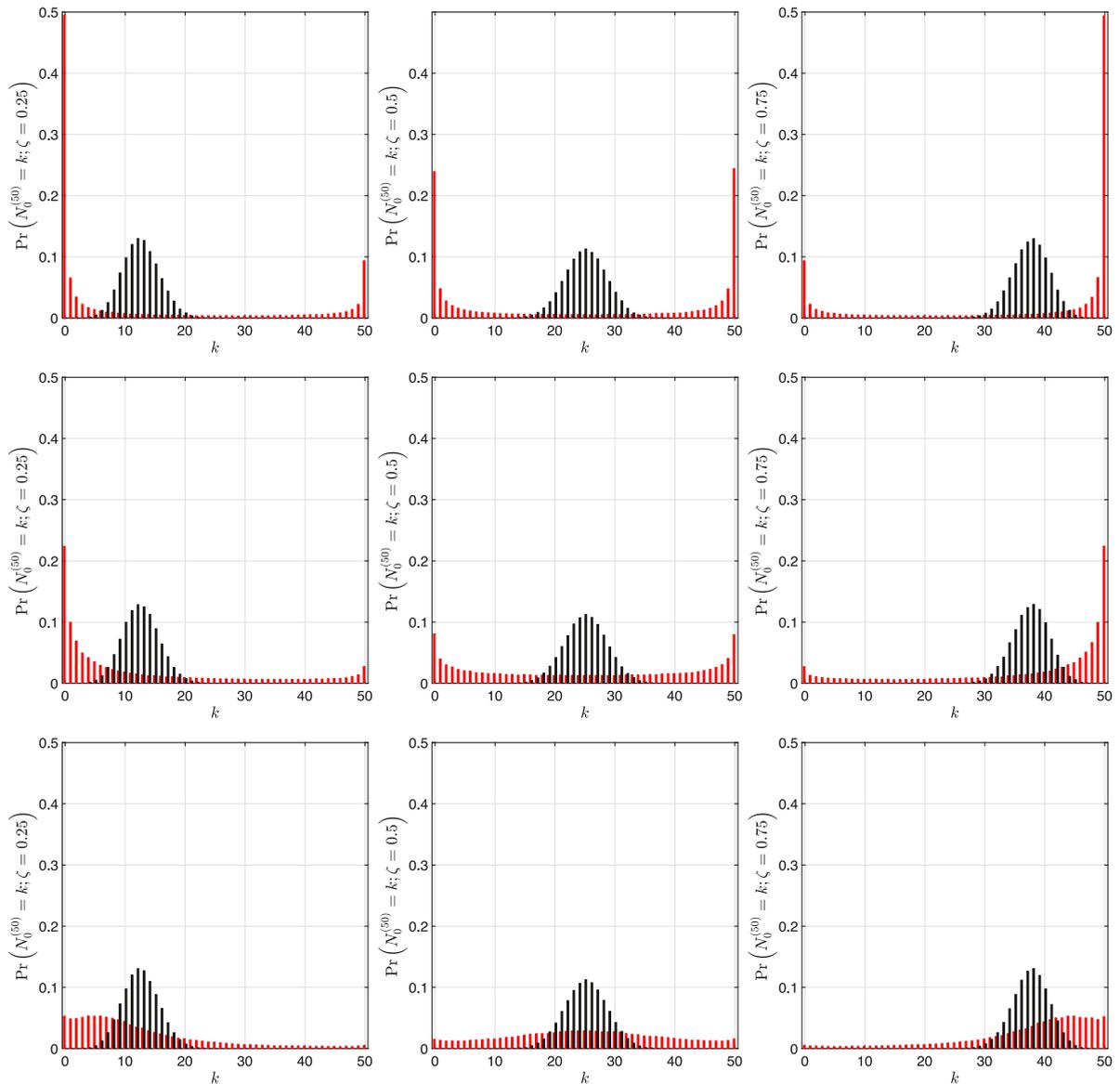
Fig. 2. Comparison of the distributions of $N_0^{(m)}$, for sample of size $m = 50$, induced by inner (in red) and outer (in black) spike and slab models for σ-stable hNRMIs. Different rows refer to different values of $\sigma \in \{0.25, 0.5, 0.75\}$ (from top to bottom), different columns refer to different values of $\zeta \in \{0.25, 0.5, 0.75\}$ (from left to right).

Such difference is particularly sizeable when $\zeta = 0.75$, even more so if the median is considered: while the length of the estimated interval for the distribution of med_{50} induced by $\tilde{P}$ is equal to 2.44, 2.04 and 1.14 when σ takes value 0.25, 0.5 and 0.75, respectively, the corresponding intervals for the distribution of med_{50} induced by $\tilde{Q}$ appear degenerate at $x_0 = 0$, which implies that $\Pr(\mathrm{median}(X_1, \ldots, X_{50}) = 0) > 0.95$.

We complete the first part of the study by investigating the distributions of the number of observations in a sample coinciding with x_0, induced by the inner and the outer spike and slab models. The notation $N_0^{(m)}$, already introduced for the inner spike and slab model, will be henceforth used also to denote the same quantity induced by the outer spike and slab model. While for the outer model specification, it is easily verified that $N_0^{(m)}$ is a Binomial with parameters m and ζ, the distribution of $N_0^{(m)}$ induced by the inner model is provided in Theorem 4 for the general case of hNRMIs, and coincides with (13) when a σ-stable hNRMI spike and slab model is considered. Alternatively, the same distributions can be estimated by resorting to the samples generated in the first part of this section. We follow the latter approach and exclude from our analysis the case $\zeta = 0$, given that it corresponds to $N_0^{(m)} \equiv 0$.

Figure 2 displays a comparison between the distributions of $N_0^{(m)}$ induced by the two models, for $\zeta \in \{0.25, 0.5, 0.75\}$ and $\sigma \in \{0.25, 0.5, 0.75\}$, and with $m = 50$. It is apparent, across all the combinations of the parameter values we considered,

Table 2

Estimated proportion of observations coinciding with x_0 in the sample $X^{(100)}$, conditional on the observation of $x_1^{(50)}$ ($\pi_1 = 0.2$), $x_2^{(50)}$ ($\pi_2 = 0.5$), and $x_3^{(50)}$ ($\pi_3 = 0.8$), for values of $\sigma \in \{0.25, 0.5, 0.75\}$ and $\zeta \in \{0.25, 0.5, 0.75\}$.

σ		0.25			0.5			0.75		
ζ		0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
$\pi_1 = 0.2$	inner	0.20	0.20	0.21	0.20	0.21	0.22	0.22	0.20	0.22
	outer	0.23	0.35	0.48	0.22	0.35	0.48	0.22	0.35	0.47
$\pi_2 = 0.5$	inner	0.50	0.50	0.51	0.50	0.50	0.51	0.51	0.49	0.51
	outer	0.38	0.50	0.63	0.37	0.50	0.62	0.37	0.50	0.63
$\pi_3 = 0.8$	inner	0.80	0.80	0.81	0.80	0.80	0.80	0.80	0.79	0.80
	outer	0.53	0.65	0.77	0.52	0.65	0.78	0.52	0.65	0.78

that the distribution of $N_0^{(m)}$ induced by $\tilde{P}$ is characterized by a larger variability than the one induced by $\tilde{Q}$. While for the latter most of the probability mass is allocated in a neighbourhood of the expected value, for the former the probability mass appears more spread over the whole support. Moreover, the study suggests that, for small values of σ, the inner spike and slab model, unlike its outer counterpart, assigns large probabilities to the extreme events $\{N_0^{(m)} = 0\}$ and $\{N_0^{(m)} = m\}$. These distributional differences can be relevant when performing Bayesian classification by means of either spike and slab hNRMI prior model, and are in line with the findings of Canale et al. (2017) for the PY case.

Our numerical investigation of the distribution of $\tilde{F}_m$, its functionals M_m and med_m, and the random variable $N_0^{(m)}$, although limited to the case of σ-stable hNRMIs, underpins the interpretation we gave of Proposition 1. Our study indicates that, if compared to the inner model, the outer one assigns less prior probability to realizations of $\tilde{F}_m$ that deviate considerably from F_0, as well as to values of $N_0^{(m)}$ which deviate considerably from its expected value. The inner spike and slab model appears characterized by a larger variability and thus can be interpreted as less informative than the outer one. In most contexts this represents a desirable feature and encapsulates the advantage of using a fully nonparametric specification as the inner spike and slab model is.

5.2. Inner vs outer models a posteriori

We next investigate the posterior behaviour of inner and outer spike and slab models, with a focus on the posterior distribution of $N_0^{(n+m)}$, conditional on the observation of a sample $x^{(n)}$. Such investigation allows us to better understand how the prior specification for the two models affects the posterior probability of observations coinciding with x_0. As in Section 5.1, we focus on the case of σ-stable hNRMI models. We set $n = 50$ and consider three observed samples, denoted as $x_1^{(50)}$, $x_2^{(50)}$ and $x_3^{(50)}$. In $x_1^{(50)}$, 10 observations coincide with x_0 (sample proportion $\pi_1 = 0.2$) and the remaining 40 display 7 distinct values with frequencies (25,5,3,3,2,1,1); in $x_2^{(50)}$, 25 observations coincide with x_0 (sample proportion $\pi_2 = 0.5$) and the remaining 25 display 7 distinct values with frequencies (10,5,3,3,2,1,1); in $x_3^{(50)}$, 40 observations coincide with x_0 (sample proportion $\pi_3 = 0.8$) and the remaining 10 display 5 distinct values with frequencies (3,3,2,1,1). Conditionally on each of the three samples, we generate realizations $(x_{n+1}, \ldots, x_{n+m})$ of $(X_{n+1}, \ldots, X_{n+m})$ by considering $\sigma \in \{0.25, 0.5, 0.75\}$ and $\zeta \in \{0.25, 0.5, 0.75\}$. Specifically, we set $m = 50$ and, for each $x_\ell^{(n)}$, with $\ell = 1, 2, 3$, and each combination of values for σ and ζ, we generate 100 000 samples and use them to estimate the posterior distribution of $N_0^{(n+m)}$, conditional on $x_\ell^{(n)}$. As for the outer spike and slab model, samples are generated by using Algorithm 1, as done in Section 5.1. Figure 3 refers to the case $\sigma = 0.25$ and displays a structural difference between inner and outer spike and slab models. Although not displayed here, similar conclusions can be drawn by investigating the cases $\sigma = 0.5$ and $\sigma = 0.75$. The posterior distribution of $N_0^{(n+m)}$, conditional on the sample $x_\ell^{(n)}$, for $\ell = 1, 2, 3$, appears concentrated around $\pi_\ell(n + m)$ when induced by the inner spike and slab model, regardless of the value of ζ. The same does not happen for the posterior distribution of $N_0^{(n+m)}$ induced by the outer spike and slab model, which is instead highly sensitive to the value of ζ. In other terms, as reported in Table 2, the expected proportion of observations coinciding with x_0 in the enlarged sample of size $n + m$, is close to the observed proportion π_ℓ when an inner spike and slab model is adopted; on the contrary, the outer spike and slab model leads to an expected proportion of observations coinciding with x_0 in $X^{(n+m)}$ which is shrunk towards the expected prior proportion ζ. Table 2 shows that this behaviour is common across all the considered values of σ. The empirical results of our study indicates that the prior induced on $N_0^{(m)}$ by the inner spike and slab model is less informative than the one induced by the outer spike and slab model. Our findings also indicate that, while a hyperprior for ζ seems needed when the outer spike and slab model is adopted, the inner specification appears structurally robust to any specification of ζ.

6. Proofs

This section contains the proofs of the main results of Sections 2 and 3.
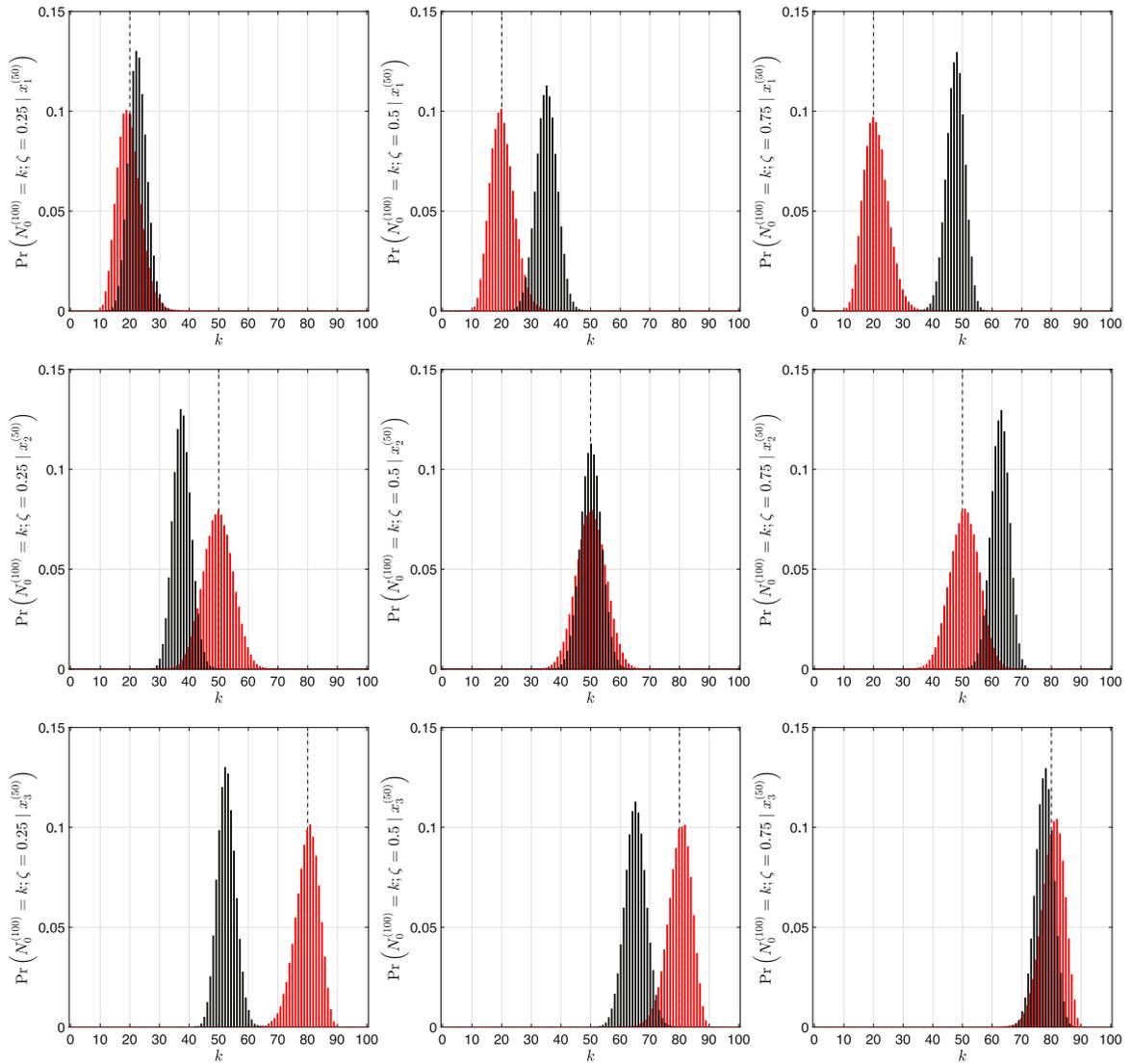
Fig. 3. Comparison of the distributions of $N_0^{(n+m)}$, with $n = m = 50$, induced by inner (in red) and outer (in black) spike and slab models for σ-stable hNRMIs, with $\sigma = 0.25$. Different rows refer to different samples $x_\ell^{(50)}$, with $\ell \in \{1, 2, 3\}$ (from top to bottom), different columns refer to different values of $\zeta \in \{0.25, 0.5, 0.75\}$ (from left to right). The vertical dashed black line indicates, in each panel, the value $\pi_\ell(n + m)$, that is the number of observations in the enlarged sample corresponding to the sample proportion π_ℓ.

6.1. Proof of Proposition 1

The proof follows from a simple application of Theorem 1. Indeed, let $p = \Pi_1^{(2)}(2; 0) = 1 - \Pi_2^{(2)}(1, 1; 0)$ and note that

$$\mathbb{E}[\tilde{P}^2(f)] = \mathbb{E}\left[\int_\mathbb{X} f^2(x)\, \tilde{P}^2(\mathrm{d}x)\right] + \mathbb{E}\left[\int_{\mathbb{X}_*^2} f(x_1) f(x_2)\, \tilde{P}(\mathrm{d}x_1)\, \tilde{P}(\mathrm{d}x_2)\right],$$

where $\mathbb{X}_*^2 = \{(x_1, x_2) \in \mathbb{X}^2 : x_1 \neq x_2\}$. Since $\xi_{2,2}(u) = \tau_2(u)$ and $\xi_{2,1}(u) = \tau_1^2(u)$, from (5) one has

$$\mathbb{E}\left[\int_\mathbb{X} f^2(x)\, \tilde{P}^2(\mathrm{d}x)\right] = p(1 - \zeta) P^*(f^2) + p\zeta f^2(x_0) + (1 - p)\zeta^2 f^2(x_0)$$

and

$$\mathbb{E}\left[\int_{\mathbb{X}_*^2} f(x_1) f(x_2)\, \tilde{P}(\mathrm{d}x_1)\, \tilde{P}(\mathrm{d}x_2)\right] = (1 - p)(1 - \zeta)^2 \left(P^*(f)\right)^2 + 2(1 - p)\zeta(1 - \zeta) f(x_0) P^*(f).$$

By combining these two, one obtains

$$\mathrm{var}(\tilde{P}(f)) = p(1 - \zeta) P^*(f^2) + p\zeta(1 - \zeta) f^2(x_0) - p(1 - \zeta)^2 (P^*(f))^2 - 2p\zeta(1 - \zeta) f(x_0) P^*(f).$$

Moreover, using (6) it can be easily seen that

$$\mathrm{var}(\tilde{Q}(f)) = (1-\zeta)^2 \left\{ p P^*(f^2) + (1-p)(P^*(f))^2 \right\} - (1-\zeta)^2 (P^*(f))^2$$

and the result follows.

6.2. Proof of Theorem 1

The proof relies on techniques similar to those used in the proof of Theorem 2 in Canale et al. (2017). By definition, we have

$$\Pi_k^{(n)}(n_1, n_2, \ldots, n_k; \zeta) = \mathbb{E}\left[\int_{\mathbb{X}^k} \prod_{j=1}^{k} \tilde{P}(\mathrm{d}x_j) \right],\tag{17}$$

for any vector of positive integers $(n_1, n_2, \ldots, n_k)$ such that $\sum_{j=1}^{k} n_k = n$, and where the integrating variables are such that $x_1 \neq x_2 \neq \ldots \neq x_k$. The right-hand side of (17) can be written as

$$\frac{1}{\Gamma(n)} \int_{\mathbb{X}^k} \int_0^\infty u^{n-1} \mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}(\mathbb{X})} \prod_{j=1}^{k} \tilde{\mu}^{n_j}(\mathrm{d}x_j) \right] \mathrm{d}u,\tag{18}$$

where we used Fubini's theorem. Henceforth we let $\mathrm{d}x_j$ denote, for any $j = 1, \ldots, k$, a neighbourhood of radius ε of x_j for which, thus, $\mathrm{d}x_j \downarrow \{x_j\}$ as $\varepsilon \downarrow 0$. We then start by focusing on the expected value appearing in (18), for which we consider two cases: i) none of the infinitesimal intervals $\mathrm{d}x_j$ contains x_0, ii) $x_0 \in \mathrm{d}x_l$ for one $l \in \{1, \ldots, k\}$. Note that, given the infinitesimal nature of the intervals $\mathrm{d}x_j$, with $j = 1, 2, \ldots, k$, and the fact that $x_1 \neq x_2 \neq \ldots \neq x_k$, x_0 is not contained in more than one interval $\mathrm{d}x_j$ with probability one.

i) In this case, only the diffuse component P^* in P_0 plays a role and it can be showed by standard techniques that

$$\mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}(\mathbb{X})} \prod_{j=1}^{k} \tilde{\mu}^{n_j}(\mathrm{d}x_j) \right] \cong c^k (1-\zeta)^k \mathrm{e}^{-c\psi(u)} \prod_{j=1}^{k} \tau_{n_j}(u) P^*(\mathrm{d}x_j),\tag{19}$$

where (19) provides a first-order approximation of the expected value on the left-hand side, which is everything we need as the higher order terms vanish when computing the integral over $\mathbb{X}^k$ in (18).

ii) We define $\mathbb{X}^* = \mathbb{X} \setminus \{x_1, \ldots, x_k\}$ and exploit the independence of increments of $\tilde{\mu}$ to write

$$\mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}(\mathbb{X})} \prod_{j=1}^{k} \tilde{\mu}^{n_j}(\mathrm{d}x_j) \right] = \mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}(\mathbb{X}^*)} \right] \prod_{m \neq l} \mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}(\mathrm{d}x_m)} \tilde{\mu}^{n_m}(\mathrm{d}x_m) \right] \mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}(\{x_0\})} \tilde{\mu}^{n_l}(\{x_0\}) \right].\tag{20}$$

Again, in the first two terms of the right-hand side of (20), only the diffuse component P^* of P_0 contributes to the integral. Thus, it can be showed by standard techniques that

$$\mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}(\mathbb{X}^*)} \right] = \mathrm{e}^{-c(1-\zeta)P^*(\mathbb{X}^*)\psi(u)}\tag{21}$$

and

$$\prod_{m \neq l} \mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}(\mathrm{d}x_m)} \tilde{\mu}^{n_m}(\mathrm{d}x_m) \right] \cong c^{k-1}(1-\zeta)^{k-1} \prod_{m \neq l} \mathrm{e}^{-c(1-\zeta)P^*(\mathrm{d}x_m)\psi(u)} P^*(\mathrm{d}x_m) \tau_{n_m}(u),\tag{22}$$

where the approximation can be interpreted as the one in (19) As for the last term of (20), we apply Feynman's technique for integration and write

$$\begin{aligned}
\mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}(\{x_0\})} \tilde{\mu}^{n_l}(\{x_0\}) \right] &= (-1)^{n_l} \frac{\mathrm{d}^{n_l}}{\mathrm{d}u^{n_l}} \mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}(\{x_0\})} \right] \\
&= (-1)^{n_l} \frac{\mathrm{d}^{n_l}}{\mathrm{d}u^{n_l}} \mathrm{e}^{-c\zeta\psi(u)} \\
&= \sum_{i=1}^{n_l} c^i \zeta^i \mathrm{e}^{-c\zeta\psi(u)} \frac{1}{i!} \sum_{j=0}^{i} (-1)^{n_l-j} \psi^{i-j}(u) \frac{\mathrm{d}^{n_l}}{\mathrm{d}u^{n_l}} \psi^j(u),
\end{aligned}$$

where the last identity is obtained by resorting to Hoppe's formula, a convenient variant of the more popular Faà di Bruno's formula (see, e.g., Johnson, 2002). Using the notation introduced in (4) we can write

$$\mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}(\{x_0\})} \tilde{\mu}^{n_l}(\{x_0\}) \right] = \sum_{i=1}^{n_l} c^i \zeta^i \mathrm{e}^{c\zeta\psi(u)} \xi_{n_l,i}(u).\tag{23}$$

By plugging (21), (22) and (23) into (20) we get

$$\mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}(\mathbb{X})} \prod_{j=1}^{k} \tilde{\mu}^{n_j}(\mathrm{d}x_j) \right] \cong c^{k-1}(1-\zeta)^{k-1} \mathrm{e}^{-c\psi(u)} \prod_{m \neq l} \tau_{n_m}(u) P^*(\mathrm{d}x_m) \sum_{i=1}^{n_l} c^i \zeta^i \xi_{n_l,i}(u).\tag{24}$$

A first-order approximation of the expected value $\mathbb{E}[\exp\{-u\tilde{\mu}(\mathbb{X})\}\prod_{j=1}^{k}\tilde{\mu}^{n_j}(dx_j)]$ can then be written as linear combination of (19) and (24), that is

$$
\left(1 - \sum_{l=1}^{k}\delta_{x_0}(dx_l)\right)c^k(1-\zeta)^k e^{-c\psi(u)}\prod_{m=1}^{k}P^*(dx_m)\tau_{n_m}(u)
$$

$$
+c^{k-1}(1-\zeta)^{k-1}\sum_{l=1}^{k}\delta_{x_0}(dx_l)\left(\prod_{m\neq l}P^*(dx_m)\tau_{n_m}(u)\right)\sum_{i=1}^{n_l}c^i\zeta^i\xi_{n_l,i}(u).
$$

The proof is completed by letting ε go to 0, by replacing the expected value in (18) with the last expression, and by computing the integral over $\mathbb{X}^k$.

6.3. Proof of Corollary 1

From Theorem 1 we obtain the following representation for $\Pi_k^{(n)}(n_1,\ldots,n_k;\zeta)$

$$
\Pi_k^{(n)}(n_1,\ldots,n_k;\zeta) = (1-\zeta)^k \Pi_k^{(n)}(n_1,\ldots,n_k;0) + (1-\zeta)^{k-1}\sum_{j=1}^{k}\frac{1}{\Gamma(n-n_j)\Gamma(n_j)}
$$

$$
\times \int_{\mathbb{X}_0^{k-1}}\mathbb{E}\left[\frac{\tilde{\mu}^{n_j}(\{x_0\})}{\tilde{\mu}^{n_j}(\mathbb{X})}\prod_{\ell\neq j}\frac{\tilde{\mu}^{n_\ell}(dx_\ell)}{\tilde{\mu}^{n_\ell}(\mathbb{X})}\right],
$$

where $\mathbb{X}_0^{k-1} = (\mathbb{X}\setminus\{x_0\})^{k-1}$. We do now focus on the integral appearing in the second summand and note that

$$
\int_{\mathbb{X}_0^{k-1}}\mathbb{E}\left[\frac{\tilde{\mu}^{n_j}(\{x_0\})}{\tilde{\mu}^{n_j}(\mathbb{X})}\prod_{\ell\neq j}\frac{\tilde{\mu}^{n_\ell}(dx_\ell)}{\tilde{\mu}^{n_\ell}(\mathbb{X})}\right]
$$

$$
= \int_0^\infty\int_0^\infty u^{n-n_j-1}v^{n_j-1}\int_{\mathbb{X}^{k-1}}\mathbb{E}\left[e^{-(u+v)\tilde{\mu}(\mathbb{X})}\tilde{\mu}^{n_j}(\{x_0\})\prod_{\ell\neq j}\tilde{\mu}^{n_\ell}(dx_\ell)\right]dv\,du
$$

$$
= \int_0^\infty\int_0^\infty u^{n-n_j-1}v^{n_j-1}e^{-c\psi(u+v)}\left(\prod_{\ell\neq j}\tau_{n_\ell}(u+v)\right)
$$

$$
\times \sum_{i=1}^{n_j}c^i\zeta^i\frac{1}{i!}\sum_{\Delta_{i,n_j}}\binom{n_j}{q_1\,\cdots\,q_i}\prod_{r=1}^{i}\tau_{n_r}(u+v)\,du\,dv,
$$

where Δ_{i,n_j} is the set of all vectors of positive integers $(q_1,\ldots,q_i)$ such that $\sum_{r=1}^{i}q_r = n_j$, and we use the fact that

$$
\xi_{n_j,i}(u+v) = \frac{1}{i!}\sum_{\Delta_{i,n_j}}\binom{n_j}{q_1\cdots q_i}\prod_{r=1}^{i}\tau_{q_r}(u+v).
$$

If one recalls the representation of the exchangeable partition probability function in (5), then

$$
\int_{\mathbb{X}_0^{k-1}}\mathbb{E}\left[\frac{\tilde{\mu}^{n_j}(\{x_0\})}{\tilde{\mu}^{n_j}(\mathbb{X})}\prod_{\ell\neq j}\frac{\tilde{\mu}^{n_\ell}(dx_\ell)}{\tilde{\mu}^{n_\ell}(\mathbb{X})}\right]
$$

$$
= \Pi_k^{(n)}(n_1,\ldots,n_{j-1},n_{j+1},\ldots,n_k;0)\int_0^\infty\frac{u^{n-n_j-1}e^{-\psi(u)}\prod_{\ell\neq j}\tau_{n_\ell}(u)}{\int_0^\infty u^{n-n_j}e^{-\psi(u)}\prod_{\ell\neq j}\tau_{n_\ell}(u)\,du}
$$

$$
\times \int_0^\infty v^{n_j-1}e^{-c\psi(u+v)}\left(\prod_{\ell\neq j}\tau_{n_\ell}(u+v)\right)\sum_{i=1}^{n_j}c^i\zeta^i\frac{1}{i!}\sum_{\Delta_{i,n_j}}\binom{n_j}{q_1\,\cdots\,q_i}\prod_{r=1}^{i}\tau_{n_r}(u+v)\,du\,dv
$$

$$
= \Pi_k^{(n)}(n_1,\ldots,n_{j-1},n_{j+1},\ldots,n_k;0)\int_0^\infty f_j(u)\int_0^\infty v^{n_j-1}e^{-c\psi(u+v)}\left(\prod_{\ell\neq j}\tau_{n_\ell}(u+v)\right)
$$

$$
\times \sum_{i=1}^{n_j}c^i\zeta^i\frac{1}{i!}\sum_{\Delta_{i,n_j}}\binom{n_j}{q_1\,\cdots\,q_i}\prod_{r=1}^{i}\tau_{n_r}(u+v)\,du\,dv,
$$

where, following James et al. (2009),

$$f_j(u) \propto u^{n-n_j-1} e^{-c\psi(u)} \prod_{\ell \neq j} \tau_{n_\ell}(u)$$

is the density function of a latent random variable U_{n-n_j}, conditional on on the $n - n_j$ observations $X^{(n)}_{-j}$ with $k-1$ distinct values that do not include x_0 and, hence, are not in cluster j. Henceforth we denote these distinct values as $(x^*_{1,j}, \ldots, x^*_{k-1,j})$. At this point, we benefit from the posterior representation of $\tilde{\mu}$ given in Theorem 1 of James et al. (2009), which entails that, conditional on $X^{(n)}_{-j}$ and on U_{n-n_j}, the distribution of $\tilde{\mu}$ equals the distribution of

$$\tilde{\mu}^{(u)}_j + \sum_{i=1}^{k-1} J^{(u)}_{i,j} \delta_{x^*_{i,j}},$$

where $\tilde{\mu}^{(u)}_j$ is a completely random measure without fixed discontinuities and with intensity $e^{-us}\rho(s)\,ds\,c\,P_0(dx)$ and the jumps $J^{(u)}_{i,j}$ are independent with respective distributions having density $f_{i,j}(s \mid u) \propto s^{n_i} e^{-us} \rho(s)$. Hence, a straightforward application of Theorem 2 of James et al. (2009) shows that, for any $u > 0$,

$$\mathbb{E}\left[\frac{\tilde{\mu}^{(u)}_j(\{x_0\})}{T^{(u)}_j + \sum_i J^{(u)}_{i,j}} \right]^{n_j} = \int_0^\infty v^{n_j-1} e^{-c\psi(u+v)} \left(\prod_{\ell \neq j} \tau_{n_\ell}(u+v) \right) \sum_{i=1}^{n_j} c^i \zeta^i \frac{1}{i!} \sum_{\Delta_{i,n_j}} \binom{n_j}{q_1 \cdots q_i} \prod_{r=1}^i \tau_{n_r}(u+v)\,dv,$$

where $T^{(u)}_j = \tilde{\mu}^{(u)}_j(\mathbb{X})$. Thus we have

$$\Pi^{(n)}_k(n_1, \ldots, n_k; \zeta) = (1-\zeta)^k \Pi^{(n)}_k(n_1, \ldots, n_k; 0) + (1-\zeta)^{k-1} \sum_{j=1}^k \Pi^{(n-n_j)}_{k-1}(n_1, \ldots, n_{j-1}, n_j, \ldots, n_k; 0)$$

$$\times \int_0^\infty f_j(u) \mathbb{E}\left[\frac{\tilde{\mu}^{(u)}_j(\{x_0\})}{T^{(u)}_j + \sum_{i=1}^{k-1} J^{(u)}_{i,j}} \right]^{n_j} du. \quad (25)$$

The proof is completed upon noting that the integral appearing in right-hand side of (25) is the conditional probability that n_j observations in $X^{(n)}$ equal x_0, given the remaining $n - n_j$ in the sample $X^{(n)}_{-j}$ of $n - n_j$ all differ from x_0 and clustered into $k-1$ groups with respective frequencies $n_1, \ldots, n_{j-1}, n_{j+1}, \ldots, n_k$.

6.4. Proof of Theorem 2

The proof follows by combining (7) in Theorem 1 with the fact that

$$\Pr(K_n = k; \zeta) = \frac{1}{k!} \sum_{\Delta_{k,n}} \binom{n}{q_1 \cdots q_k} \Pi^{(n)}_k(q_1, \ldots, q_k; \zeta).$$

6.5. Proof of Theorem 3

For the purpose of the proof, we introduce the quantity

$$\Pi^{(n)}_{k,j}(n_1, \ldots, n_k; \zeta),$$

for $j = 0, 1, \ldots, k$, to denote the probability of observing a partition of n observations into k distinct blocks such that the frequency of the block coinciding with x_0 is given by the j-th argument of the function. The case $j = 0$ indicates that no block in the partition coincides with x_0. From (5) we have that

$$\Pi^{(n)}_{k,0}(n_1, \ldots, n_k; \zeta) = \frac{1}{\Gamma(n)} c^k (1-\zeta)^k \int_0^\infty u^{n-1} e^{-c\psi(u)} \prod_{m=1}^k \tau_{n_m}(u)\,du,$$

$$\Pi^{(n)}_{k,j}(n_1, \ldots, n_k; \zeta) = \frac{1}{\Gamma(n)} c^{k-1}(1-\zeta)^{k-1} \sum_{i=1}^{n_j} c^i \zeta^i \int_0^\infty u^{n-1} e^{-c\psi(u)} \xi_{n_j,i}(u) \prod_{m \neq j}^k \tau_{n_m}(u)\,du.$$

The predictive distribution takes the form

$$\Pr(X_{n+1} \in A \mid X^{(n)}; \zeta) = w^{(*)}_{k,n} P^*(A) + w^{(0)}_{k,n} \delta_{x_0}(A) + \sum_{l: x^*_l \neq x_0} w^{(l)}_{k,n} \delta_{x^*_l}(A), \quad (26)$$

where the weights in (26) are as follows. If $x_0 \notin \{x_1^*, \ldots, x_k^*\}$,

$$w_{k,n}^{(*)} = \frac{\Pi_{k+1,0}^{(n+1)}(n_1, \ldots, n_k, 1; \zeta)}{\Pi_{k,0}^{(n)}(n_1, \ldots, n_k; \zeta)},$$

$$w_{k,n}^{(0)} = \frac{\Pi_{k+1,k+1}^{(n+1)}(n_1, \ldots, n_k, 1; \zeta)}{\Pi_{k,0}^{(n)}(n_1, \ldots, n_k; \zeta)},$$

$$w_{k,n}^{(l)} = \frac{\Pi_{k,0}^{(n+1)}(n_1, \ldots, n_l + 1, \ldots, n_k; \zeta)}{\Pi_{k,0}^{(n)}(n_1, \ldots, n_l, \ldots, n_k; \zeta)}.$$

On the other hand, if $x_0 = x_j^*$ for some $j = 1, \ldots, k$, then

$$w_{k,n}^{(*)} = \frac{\Pi_{k+1,j}^{(n+1)}(n_1, \ldots, n_k, 1; \zeta)}{\Pi_{k,j}^{(n)}(n_1, \ldots, n_k; \zeta)},$$

$$w_{k,n}^{(0)} = \frac{\Pi_{k,j}^{(n+1)}(n_1, \ldots, n_j + 1, \ldots, n_k; \zeta)}{\Pi_{k,j}^{(n)}(n_1, \ldots, n_j, \ldots, n_k; \zeta)},$$

$$w_{k,n}^{(l)} = \frac{\Pi_{k,j}^{(n+1)}(n_1, \ldots, n_l + 1, \ldots, n_k; \zeta)}{\Pi_{k,j}^{(n)}(n_1, \ldots, n_l, \ldots, n_k; \zeta)}.$$

Simple algebra completes the proof.

6.6. Proof of Theorem 4

As in the proof of Theorem 3, we use the notation

$$\Pi_{k,j}^{(n)}(n_1, \ldots, n_k; \zeta), \quad j \in \{0, 1, \ldots, k\}$$

to indicate the probability that an inner spike and slab hNRMI model induce a partition of n observations into k blocks of size $n_1, \ldots, n_k$, where, if $j \in \{1, \ldots, k\}$, the observations of the j-th block coincide with x_0, while if $j = 0$ then no observation coincides with x_0. Along similar lines as Theorem 5, we can show that

- if $j = 0$

$$\Pi_{k,0}^{(n)}(n_1, \ldots, n_k; \zeta) = \frac{1}{\Gamma(n)} c^k (1 - \zeta)^k \int_0^\infty u^{n-1} e^{-c\psi(u)} \prod_{m=1}^k \tau_{n_m}(u) \mathrm{d}u;$$

- if $j \in \{1, \ldots, k\}$

$$\Pi_{k,j}^{(n)}(n_1, \ldots, n_k; \zeta) = \frac{1}{\Gamma(n)} c^{k-1} (1 - \zeta)^{k-1} \sum_{i=1}^{n_j} c^i \zeta^i \int_0^\infty u^{n-1} e^{-c\psi(u)} \xi_{n_j,i}(u) \prod_{m \neq j} \tau_{n_m}(u) \mathrm{d}u.$$

Next we observe that

$$\Pr(N_0^{(n)} = 0) = \sum_{k=1}^n \frac{1}{k!} \sum_{\mathbf{m}n \in \mathcal{N}_k^{(n)}} \binom{n}{n_1 \cdots n_k} \Pi_{k,0}^{(n)}(n_1, \ldots, n_k; \zeta)$$

$$= \frac{1}{\Gamma(n)} \sum_{k=1}^n c^k (1 - \zeta)^k \int_0^\infty u^{n-1} e^{-c\psi(u)} \frac{1}{k!} \sum_{\mathbf{m}n \in \mathcal{N}_k^{(n)}} \binom{n}{n_1 \cdots n_k} \prod_{m=1}^k \tau_{n_m}(u) \mathrm{d}u. \tag{27}$$

Similarly, if $j \in \{1, \ldots, n\}$,

$$\Pr(N_0^{(n)} = j) = \sum_{k=1}^{n-j+1} \frac{1}{(k-1)!} \sum_{\mathbf{m}n \in \mathcal{N}_{k-1}^{(n-j)}} \binom{n}{n_1 \cdots n_{k-1}} k \, \Pi_{k,k}^{(n)}(n_1, \ldots, n_{k-1}, j; \zeta)$$

$$= \frac{1}{\Gamma(n)} \sum_{k=1}^{n-j+1} k c^{k-1} (1 - \zeta)^{k-1}$$

$$\times \sum_{i=1}^j c^i \zeta^i \int_0^\infty u^{n-1} e^{-c\psi(u)} \xi_{j,i}(u) \frac{1}{(k-1)!} \sum_{\mathbf{m}n \in \mathcal{N}_{k-1}^{(n-j)}} \binom{n}{n_1 \cdots n_{k-1}} \prod_{m=1}^{k-1} \tau_{n_m}(u) \mathrm{d}u, \tag{28}$$

with $\sum_{\mathbf{m}n \in \mathcal{N}_{k-1}^{(n-j)}} \equiv 1$ when $k = 1$. The two expressions in (27) and (28) can be summarized in one, leading to (10), where we observe that $\xi_{j,0} = 0$ if $j \geq 1$, and $\xi_{0,0} = 1$.

Acknowledgments

References

Barcella, W., De Iorio, M., Baio, G., Malone-Lee, J., 2016. Variable selection in covariate dependent random partition models: an application to urinary tract infection. Statistics in Medicine 35, 1373–1389.

Barrios, E., Lijoi, A., Nieto-Barajas, L.E., Prünster, I., 2013. Modeling with normalized random measure mixture models. Statistical Science 28 (3), 313–334.

Bassetti, F., Ladelli, L., 2020. Asymptotic number of clusters for species sampling sequences with non-diffuse base measure. Statistics & Probability Letters 108749.

Bogdan, M., Ghosh, J.K., Tokdar, S.T., 2008. A comparison of the Benjamini-Hochberg procedure with some Bayesian rules for multiple testing. In: Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen. In: Inst. Math. Stat. Collect., 1. Inst. Math. Statist., Beachwood, OH, pp. 211–230.

Canale, A., Lijoi, A., Nipoti, B., Prünster, I., 2017. On the Pitman–Yor process with spike and slab base measure. Biometrika 104 (3), 681–697.

Cassese, A., Zhu, W., Guindani, M., Vannucci, M., 2019. A Bayesian nonparametric spiked process prior for dynamic model selection. Bayesian Analysis 14 (2), 553–572.

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

Corradin, R., Nieto-Barajas, L.E., Nipoti, B., 2021. Optimal stratification of survival data via bayesian nonparametric mixtures. Econometrics and Statistics in press.

D'Angelo, L., Canale, A., Guindani, M., Yu, Z., 2021. Bayesian nonparametric analysis for the detection of spikes in noisy calcium imaging data. arXiv:2102.09403.

Do, K.-A., Müller, P., Tang, F., 2005. A Bayesian mixture model for differential gene expression. Journal of the Royal Statistical Society. Series C. Applied Statistics 54 (3), 627–644.

Dunson, D.B., Herring, A.H., Engel, S.M., 2008. Bayesian selection and clustering of polymorphisms in functionally related genes. Journal of the American Statistical Association 103 (482), 534–546.

Favaro, S., Lijoi, A., Nava, C., Nipoti, B., Pruenster, I., Teh, Y.W., 2016. On the stick-breaking representation for homogeneous NRMIs. Bayesian Analysis 11 (3), 697–724.

Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. Annals of Statistics 1, 209–230.

Guindani, M., Sepúlveda, N., Paulino, C.D., Müller, P., 2014. A Bayesian semi-parametric approach for the differential analysis of sequence counts data. Journal of the Royal Statistical Society. Series C, Applied Statistics 63 (3), 385.

James, L.F., Lijoi, A., Prünster, I., 2006. Conjugacy as a distinctive feature of the Dirichlet process. Scandinavian Journal of Statistics 33 (1), 105–120.

James, L.F., Lijoi, A., Prünster, I., 2009. Posterior analysis for normalized random measures with independent increments. Scandinavian Journal of Statistics 36 (1), 76–97.

Johnson, W.P., 2002. The curious history of Faà di Bruno's formula. The American mathematical monthly 109 (3), 217–234.

Kim, S., Dahl, D.B., Vannucci, M., 2009. Spiked Dirichlet process prior for Bayesian multiple hypothesis testing in random effects models. Bayesian Analysis 4 (4), 707–732.

Lijoi, A., Mena, R.H., Prünster, I., 2005. Hierarchical mixture modeling with normalized inverse-Gaussian priors. Journal of the American Statistical Association 100 (472), 1278–1291.

Lijoi, A., Prünster, I., 2010. Models beyond the Dirichlet process. In: Bayesian nonparametrics. Cambridge Univ. Press, Cambridge, pp. 80–136.

MacLehose, R.F., Dunson, D.B., Herring, A.H., Hoppin, J.A., 2007. Bayesian methods for highly correlated exposure data. Epidemiology 18 (2), 199–207.

Mitchell, T.J., Beauchamp, J.J., 1988. Bayesian variable selection in linear regression. Journal of the American Statistical Association 83 (404), 1023–1036. With comments by James Berger and C. L. Mallows and with a reply by the authors

Perman, M., Pitman, J., Yor, M., 1992. Size-biased sampling of Poisson point processes and excursions. Probability Theory and Related Fields 92 (1), 21–39.

Pitman, J., Yor, M., 1997. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. Ann. Probab. 25 (2), 855–900.

Regazzini, E., Lijoi, A., Prünster, I., et al., 2003. Distributional results for means of normalized random measures with independent increments. The Annals of Statistics 31 (2), 560–585.

Sangalli, L.M., 2006. Some developments of the normalized random measures with independent increments. Sankhyā 68 (3), 461–487.

Scarpa, B., Dunson, D.B., 2009. Bayesian hierarchical functional data analysis via contaminated informative priors. Biometrics 65 (3), 772–780.

Suarez, A.J., Ghosal, S., 2016. Bayesian clustering of functional data using local features. Bayesian Anal. 11 (1), 71–98.

Yang, M., 2012. Bayesian variable selection for logistic mixed model with nonparametric random effects. Computational Statistics & Data Analysis 56 (9), 2663–2674.