New Prior Near-ignorance Models on the Simplex

Francesca Mangili and Alessio Benavoli

IDSIA, Galleria 2, CH-6928 Manno (Lugano), Switzerland
email: francesca@idsia.ch, alessio@idsia.ch

Abstract

The aim of this paper is to derive new near-ignorance models on the probability simplex, which do not directly involve the Dirichlet distribution and, thus, are alternative to the Imprecise Dirichlet Model (IDM). We focus our investigation on a particular class of distributions on the simplex which is known as the class of Normalized Infinitely Divisible (NID) distributions; it includes the Dirichlet distribution as a particular case. For this class it is possible to derive general formulae for prior and posterior predictive inferences, by exploiting the Lévy-Khintchine representation theorem. This allows us to generally characterize the near-ignorance properties of the NID class. After deriving these general properties, we focus our attention on three members of this class. We will show that one of these near-ignorance models satisfies the representation invariance principle and, for a given value of the prior strength, always provides inferences that encompass those of the IDM. The other two models do not satisfy this principle, but their imprecision depends linearly or almost linearly on the number of observed categories; we argue that this is sometimes a desirable property for a predictive model.

Keywords: Prior near-ignorance, Normalized Infinitely Divisible distribution, Imprecise Dirichlet Model.

1. Introduction

Consider a variable Y taking values in a finite set $\mathscr{Y} = \{1, 2, \ldots, M\}$ of cardinality $M \geq 2$ and assume that we have a sample of size N of independent and identically distributed outcomes Y_j, $j = 1, \ldots, N$, of Y. Our aim is to estimate the probabilities P_i for $i = 1, \ldots, M$, that is the probability that Y takes the i-th value. In other words, by defining the M-dimensional simplex

$$\Delta_M(c) = \left\{ (w_1, \ldots, w_M) : w_i \in \mathbb{R}^+, \ \sum_{i=1}^{M} w_i = c \right\}, \tag{1}$$

we want to estimate a vector P belonging to the unit M-dimensional simplex $\Delta_M(1)$. A common way to tackle this problem within the Bayesian framework consists in assuming a conjugate prior for the vector of variables $P = (P_1, \ldots, P_M)$. Since, given P, the likelihood is a multinomial distribution, the conjugate prior is a Dirichlet distribution with parameters $\alpha_i > 0$, i.e., $\mathrm{Dir}(\alpha)$ with $\alpha = (\alpha_1, \ldots, \alpha_M)$. Then, by conjugacy, the posterior distribution is still Dirichlet, $\mathrm{Dir}(\alpha + n)$, with updated parameters $\alpha_i + n_i$ where n_i is the number of observations such that $Y_j = i$ and $n = (n_1, \ldots, n_M)$.

To compute prior and posterior inferences, we must choose the values of the prior parameters in α. This can be done based on the prior information or, when no prior information is available, these parameters must be selected to reflect this condition of prior ignorance. In the latter case, there are two main avenues that we can follow. The first assumes that ignorance can be modelled satisfactorily by a so-called noninformative prior, by which is meant a prior that contains no information. The problem is how to define the meaning of "containing no prior information". To guarantee that a simple reparametrization does not produce information, a noninformative prior is often defined as a prior that is invariant under certain reparametrizations of the parameter space [1, Sec. 3.3]. For instance, for the multinomial-Dirichlet model, this criterion leads to the following choices: Bayes and Laplace's uniform prior $\alpha_i = 1$, Jeffreys' prior $\alpha_i = 1/2$ [2], and Haldane's prior $\alpha_i \to 0$ [3]. These priors satisfy different properties of invariance. A problem with noninformative priors is that often they are not unique and that they do not really model lack of prior information, but only invariance. An alternative is to use a set of prior distributions, $\mathscr{M}$, rather than a single distribution, to model prior ignorance about statistical parameters. Each prior distribution in $\mathscr{M}$ is updated by Bayes' rule, producing a set of posterior distributions. In fact there are two distinct approaches of this kind, which have been compared by Walley [4]. The first approach, known as Bayesian sensitivity analysis or Bayesian robustness [5], assumes that there is an ideal prior distribution that we are unable to accurately determine because of limited time or resources; thus, we include in $\mathscr{M}$ any plausible candidate. The resulting set of priors is in general a neighborhood model, i.e., the set of all distributions that are close (w.r.t. some criterion) to the ideal distribution. Examples of neighbourhood models are: ϵ-contamination models [1, 6]; restricted ϵ-contamination models [7]; intervals of measures [6]; the density ratio class [4], etc. However, this approach can be unsatisfactory when there is (almost) no prior information or the information is of doubtful relevance. Then there is no ideal prior distribution, because no single prior distribution can adequately model the limited prior information. Therefore, in this case, also a neighbourhood model would be inadequate. To address this issue, Walley [4] has proposed a different approach, known as the theory of imprecise probabilities or coherent lower (and upper) previsions. This approach revises Bayesian sensitivity analysis by directly emphasizing the upper and lower expectations (also called previsions) that are generated by $\mathcal{M}$. The set $\mathcal{M}$ is chosen in such a way as to model prior ignorance for a function (or set of functions) of interest f; this means that the only information the model gives about the expected value of a function $E[f]$ is that it belongs to $[\inf f, \sup f]$, which is equivalent to stating a condition of complete prior ignorance about the value of f. The result in the context of multinomial distributions is the so-called *Imprecise Dirichlet Model* (IDM) [8, 9], which consists of the set of Dirichlet distributions:

$$\mathcal{M} = \left\{ \mathrm{Dir}(\alpha) \colon \ \alpha_i > 0, \ \sum_{i=1}^{M} \alpha_i = s \right\}, \tag{2}$$

where $\Gamma(\cdot)$ is the Gamma function and s is the prior strength. For a fixed value s, this is the set of all Dirichlet distributions obtained by allowing α to vary freely in the interior of the simplex $\varDelta_M(s)$. IDM is a model of prior ignorance about $E[P_i]$ for $i = 1, \ldots, M$, since $E[P_i] = \alpha_i / s$, with $0 < \alpha_i < s$, and thus

$$\underline{E}[P_i] = 0, \ \ \overline{E}[P_i] = 1, \tag{3}$$

where $\underline{E}, \overline{E}$ denote, respectively, the lower and upper expectations.

Prior ignorance cannot be imposed to all inferences otherwise the model would not be able to learn from data [4]. The IDM model is a prior near-ignorance model in the sense that it assures prior ignorance about many of the inferences of interest in statistical analysis, and at the same time is capable of learning from data and converges to the "truth" as the number of observations increases (is consistent in the terminology of Bayesian asymptotic analysis). Another important characteristic of the IDM is its computational tractability, which follows from the conjugacy between the categorical and Dirichlet distributions for i.i.d. observations. Finally, by suitably choosing

s, we can also guarantee that the IDM encompasses all Bayesian inferences derived from the aforementioned noninformative priors, thus bypassing the problem of uniqueness [8].

The question we aim to address in this paper is whether there are other models of prior near-ignorance on the simplex that are not directly derived from a Dirichlet distribution. For this, let us recall how the Dirichlet distribution can be obtained via normalization from a set of Gamma distributed independent variables divided by their sum. Consider a collection of variables $X_1, \dots, X_M$ which are assumed to be independent and distributed according to Gamma distributions with parameters $(\alpha_1, 1), \dots, (\alpha_M, 1)$, where $(\alpha_i, 1)$ is the shape and scale parameter of the Gamma distribution for the variable X_i. Define $W = X_1 + \cdots + X_M$ and $P_i = X_i / W$ for $i = 1, \dots, M$, then it can be shown that $P \sim \mathrm{Dir}(\alpha)$.

In analogy with the Dirichlet distribution, many distributions on the simplex can be obtained by normalization of a collection of independent positive random variables. For our purpose, an interesting class is that of the Infinitely Divisible (ID) distributions [10]. There are many examples of such distributions, including the inverse Gaussian distribution, the right-skewed stable distributions, etc. A counter-example is, for instance, the uniform distribution. The class of distributions that are obtained through a normalization of positive ID distributed variables, which is known as the class of *Normalized Infinitely Divisible* (NID) distributions, has been studied by Favaro in [10], and is described in details in Section 2. Favaro derives general expressions for the moments of any order of a NID distribution (similar to those presented in Section 2.2 of this paper for the mixed moments) and focuses on special cases of NID that lead to explicit expressions of the PDF. More extensive studies have been carried out in the context of Bayesian nonparametric inference, considering the class of normalized random measures with independent increments, which contains the Dirichlet process as a particular case (see in particular [11, 12, 13, 14, 15, 16]). When considering finite partitions of the support, these processes reduces to NID distributions (such as the Dirichlet process reduces to a Dirichlet distribution when considering a finite partition). Following the approach used in these studies (and in particular in [16]) for the posterior analysis of normalized random measures, we have derived general posterior results for the NID distributions.

The main contribution of this paper is to use the properties of the NID distributions to derive a class of prior near-ignorance models. In analogy with IDM, we call any member of this class an *Imprecise NID (INID)* model. We will show in Section 3 that an INID model can be obtained from any NID by letting some of its parameters vary in the interior of the simplex $\Delta_M(s)$, as in the IDM. After discussing some general properties for the prior near-ignorance models obtained from the NID class, we focus our attention on three new near-ignorance prior models on the simplex. These models are derived by normalization of a collection of variables distributed according to, respectively, a γ-stable distribution (Section 4), an Inverse Gaussian distribution (Section 5) and a Gamma distribution (Section 6). We will show that, although these three INID models are not conjugate with the multinomial distribution, the posterior inferences drawn from them are still computationally tractable: for the first model the lower and upper expectations of P_i can be computed by means of simple algebraic expressions. Assuming the validity of a conjecture, this is also true for the third model, while for the second the lower and upper expectations can be computed efficiently by numerically solving one-dimensional integrals.

Furthermore, in Section 7 we will study the invariance properties satisfied by all these new prior near-ignorance models, with particular attention to the representation invariance principle (RIP) (that is, the posterior inferences of an event are independent w.r.t. refinements and coarsenings of the possibility space) which was pointed out by Walley as an important feature of IDM [8]. We will show that the model based on the normalized Gamma distribution always provides, for a given s, inferences that are more conservative than those of IDM and, moreover, satisfies the RIP. On the other hand, the other two models, which do not satisfy RIP, have a posterior imprecision which increases linearly or almost linearly with the number of observed categories. All the properties and features of these three models are summarized and discussed in Section 8.

2. The class of Normalized Infinitely Divisible distributions

The aim of this section is to introduce the class of Normalized Infinitely Divisible (NID) distributions and review some general properties that allow to characterize them [10]. An NID distribution is obtained by normalizing a collection of Infinitely Divisible (ID) positive random variables $X_1, \ldots, X_M$. Therefore, we start the discussion by introducing the class of ID distributions.

2.1. Infinitely Divisible distributions

Definition 1. A random variable X_i is Infinitely Divisible (ID) if for any $k \in \mathbb{N}$ there exist i.i.d. variables $Z_1,\dots,Z_k$ such that $X_i \stackrel{d}{=} Z_1+\cdots+Z_k$ or, alternatively, the variable X_i can be separated into the sum of an arbitrary number of i.i.d. variables.

The notion of infinite divisibility can be equivalently introduced by means of the characteristic function of the variable X_i:

$$\varphi_{X_i}(u) = E[e^{\iota u X_i}] = \int_0^\infty e^{\iota u X_i} dF(X_i), \tag{4}$$

where $u > 0$ and ι denotes the imaginary unit. The variable X_i is infinitely divisible if, for all integers k, there exists a random variable Z such that

$$\varphi_{X_i}(u) = \left(\varphi_Z(u)\right)^k.$$

Example 1. Consider the case where X_i is Gamma-distributed with parameters $(1,1)$; the characteristic function of the Gamma distribution is

$$\varphi_{X_i}(u) = (1-\iota u)^{-1} = (1-\iota u)^{-\frac{1}{k}k} = \left(\varphi_Z(u)\right)^k,$$

where Z is Gamma distributed with parameters $(\frac{1}{k},1)$.

The question is then which (positive) variables are ID. The answer is provided by the Lévy-Khintchine representation theorem for non-negative variables [17] which states that there is a one-to one correspondence between the set of ID Laplace transforms $\varphi(\iota u)$ on $\mathbb{R}^+$ and the pair (X_i, ν_i) where $X_i \in \mathbb{R}^+$ and ν_i is a Lévy measure ν_i. Note that a non-negative Borel measure ν_i on $\mathbb{R}^+$ is called a Lévy measure if $\int_0^\infty \min(1,x)\nu_i(dx) < \infty$ [17, Sec. 16.4]. In particular, based on the Lévy-Khintchine representation theorem it holds:

$$\varphi_{X_i}(\iota u) = E[e^{-uX_i}] = e^{-\psi_i(u)}, \tag{5}$$

where

$$\psi_i(u) = \int_0^\infty (1-e^{-ux})\nu_i(dx), \tag{6}$$

is typically referred to as Laplace exponent of X_i. This result implies that the Laplace transform of a positive ID variable is completely characterized by its Lévy measure ν_i. Note that if ν_i is a well-defined Lévy measure, so it is $\alpha_i \nu_i$ for any $\alpha_i > 0$ and thus $\alpha_i \psi_i(u)$ is a well-defined Laplace exponent of an ID variable.

Sometimes the moment generating function $E[e^{uX_i}]$ is used instead of the Laplace transform. However, for a variable X_i the moment generating function may not exist, while the Laplace transform of a positive variable always exists (since $0 \le e^{-ux} \le 1$) and it can be shown that it is a completely monotonic function, i.e., $(-1)^r \frac{d^r}{du^r} \varphi_{X_i}(\iota u) \ge 0$ for any $r = 1,2,\ldots$ and for any $u > 0$ [17, Sec. 14.9]. The moment generating function $E[e^{uX_i}]$, when it exists, can be obtained from the Laplace exponent as $e^{-\psi_i(-u)}$.

Example 1 (continued). Consider again the case where X_i is Gamma-distributed with parameters $(1,1)$; the Gamma distribution is ID and has Lévy measure $\nu_i'(dx) = x^{-1} e^{-x} dx$. By multiplying ν_i' by the positive parameter α_i we obtain the well defined Lévy measure $\nu_i(dx) = \alpha_i x^{-1} e^{-x} dx$ which is the Lévy measure of a Gamma-distributed variable with parameters $(\alpha_i, 1)$. Its Laplace exponent is:

$$\psi_i(u) = \int_0^\infty (1 - e^{-ux}) \nu_i(dx) = \alpha_i \ln(1+u),$$

and so $\varphi_{X_i}(\iota u) = E[e^{-uX_i}] = e^{-\psi_i(u)} = (u+1)^{-\alpha_i}$. The function

$$(-1)^r \frac{d^r}{du^r} e^{-\psi_i(u)} = (-1)^r \frac{d^r}{du^r} (u+1)^{-\alpha_i} = \alpha_i(\alpha_i+1)\ldots(\alpha_i+r-1)(u+1)^{-\alpha_i - r},$$

is completely monotonic for any $r = 1,2,\ldots$ and gives, for $u = 0$, the non-central moments of a Gamma distribution with parameters $(\alpha_i, 1)$.[1]

2.2. Normalized Infinitely Divisible distributions

Based on the Lévy-Khintchine representation theorem, it has been shown in Section 2.1 that a collection of ID positive variables X_i, $i = 1,\ldots,M$, is completely characterized by the corresponding collection of Lévy measures $\nu_1, \ldots, \nu_M$. Hereafter, by deriving the moments of P_i, we will show that this holds also for

[1] This follows from the relation $E[X_i^r] = E[(-1)^r \frac{d^r}{du^r} e^{-uX_i}|_{u=0}]$ after passing the derivative through the integral, which is allowed by the fact that $x^r e^{-ux}$ is continuous and bounded.

the normalized variables $P_i = X_i/W,\ i = 1,\dots,M$, with $W = X_1 + \cdots + X_M$. In [10] moments of an arbitrary order are obtained for a NID variable P_i; here we extend those results to mixed moments and to the posterior distribution of P_i given the vector of counts n. Similar results are presented in [16] for the normalization of Lévy processes.

By exploiting the equality

$$\int_0^\infty u^{r-1} e^{-uW} du = \Gamma(r) W^{-r},$$

valid for any positive integer r, we can obtain the following general result about the mixed non-central moments of an NID distribution:

$$\begin{aligned}
E[P_1^{r_1}\cdots P_k^{r_k}] &= E[X_1^{r_1}\cdots X_k^{r_k} W^{-r}] \\
&= \frac{1}{\Gamma(r)}\int_0^\infty u^{r-1} E\left[e^{-uW}\prod_{j=1}^{k} X_j^{r_j}\right] du \\
&= \frac{1}{\Gamma(r)}\int_0^\infty u^{r-1} \prod_{i=k+1}^{M} E\left[e^{-uX_i}\right]\prod_{j=1}^{k} E\left[X_j^{r_j} e^{-uX_j}\right] du \\
&= \frac{1}{\Gamma(r)}\int_0^\infty u^{r-1} \prod_{i=k+1}^{M} e^{-\psi_i(u)}\prod_{j=1}^{k}(-1)^{r_j} E\left[\frac{d^{r_j}}{du^{r_j}} e^{-uX_j}\right] du, \\
&= \frac{1}{\Gamma(r)}\int_0^\infty u^{r-1} e^{-\sum_{i=k+1}^{M}\psi_i(u)}\prod_{j=1}^{k}(-1)^{r_j} \frac{d^{r_j}}{du^{r_j}} e^{-\psi_j(u)} du,
\end{aligned} \tag{7}$$

where $r_1,\dots,r_k$ are non-negative integers such that $r = \sum_{i=1}^{k} r_i$. This result extends to mixed moments the one in [10, Proposition 2] and is analogous to the one obtained in [16, Appendix A] for normalized Lévy processes. Notice that, (i) passing the expectation through the integral (second equality) is justified by Fubini's theorem; (ii) passing the derivatives through the integrals (fifth equality) is justified by the fact that $x^r e^{-ux}$ is continuous and bounded. We can simplify the above equation by exploiting the Faà di Bruno's formula:

$$\frac{d^r}{du^r} e^{f(u)} = e^{f(u)} \sum_{k=1}^{r} B_{r,k}\left(f^{(1)}(u),\dots,f^{(r)}(u)\right) = e^{f(u)} B_r\left(f^{(1)}(u),\dots,f^{(r)}(u)\right), \tag{8}$$

where $B_{r,k}$ and B_r are the partial and complete Bell polynomial. The complete

Bell polynomial is defined as [18]:

$$B_r(x_1,\ldots,x_r)=\det\begin{bmatrix} x_1 & \binom{r-1}{1}x_2 & \binom{r-1}{2}x_3 & \binom{r-1}{3}x_4 & \cdots & x_r \\ -1 & x_1 & \binom{r-2}{1}x_2 & \binom{r-2}{2}x_3 & \cdots & x_{r-1} \\ 0 & -1 & x_1 & \binom{r-3}{1}x_2 & \cdots & x_{r-2} \\ 0 & 0 & -1 & x_1 & \cdots & x_{r-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & x_1 \end{bmatrix} \tag{9}$$

for any $r=1,2,\ldots$. It can alternatively be defined in terms of the partial Bell polynomial:

$$B_r(x_1,\ldots,x_r)=\sum_{k=1}^{r}B_{r,k}(x_1,x_2,\ldots,x_{r-k+1}), \tag{10}$$

where

$$B_{r,k}(x_1,x_2,\ldots,x_{r-k+1})=\sum\frac{r!}{j_1!j_2!\cdots j_{r-k+1}!}\left(\frac{x_1}{1!}\right)^{j_1}\left(\frac{x_2}{2!}\right)^{j_2}\cdots\left(\frac{x_{r-k+1}}{(r-k+1)!}\right)^{j_{r-k+1}} \tag{11}$$

and the sum is taken over all sequences $j_1,j_2,\ldots,j_{r-k+1}$ of non-negative integers such that $j_1+j_2+\cdots=k$ and $j_1+2j_2+3j_3+\cdots=r$.

Using the Bell polynomials

$$B_r^i=B_r\left(-\psi_i^{(1)}(u),\ldots,-\psi_i^{(r)}(u)\right),$$

where $\psi_i^{(k)}$ denotes the k-th derivative of ψ_i, (7) becomes

$$E[P_1^{r_1}\cdots P_k^{r_k}]=\frac{1}{\Gamma(r)}\int_0^\infty u^{r-1}e^{-\sum_{i=1}^{M}\psi_i(u)}\prod_{j=1}^{k}B_{r_i}^i\,du. \tag{12}$$

This result extends to mixed moment the one in [10, Proposition 3]. Notice that for $r_i=r$ ($r_j=0$ for all $j\neq i$), (12) gives the r-th prior non-central moment of P_i, i.e., $E[P_i^r]$, while for $r_i=r_j=1$ and $r=2$ it gives $E[P_iP_j]$ and so on.

Proposition 1. *Given a NID prior with Laplace transforms $e^{-\psi_i(u)}$, $i=1,\ldots,M$, for the vector of probabilities $P=(P_1,\ldots,P_M)$ and the vector of counts $n=(n_1,\ldots,n_M)$ the posterior expectation of P_i^r is*

$$
\begin{aligned}
E[P_i^r|n] &= \frac{E[P_i^r P_1^{n_1}\cdots P_M^{n_M}]}{E[P_1^{n_1}\cdots P_M^{n_M}]} \\
&= \frac{(N-1)!}{(N+r-1)!}\frac{\displaystyle\int_0^\infty u^{N+r-1}e^{-\sum_j \psi_j(u)}\prod_{j=1}^M B^j_{n_j+rI_{j=i}}\,du}{\displaystyle\int_0^\infty u^{N-1}e^{-\sum_j \psi_j(u)}\prod_{j=1}^M B^j_{n_j}\,du}
\end{aligned}
\tag{13}
$$

where $B_0^j = 1$ and $I_{j=i}$ is the indicator function which is one when $j = i$ and zero otherwise.

The proof of this proposition and of the next theorems and corollaries can be found in the appendix. Proposition 1, whose result is analogous to that in [16, Proposition 2] for normalized Lévy processes, states that the posterior moments of an NID distribution on the simplex can be obtained by solving two one-dimensional integrals as it is shown in the next example.

Example 2. Assume that we have $N = 5$ observations taking values in $M = 3$ categories. The vector of counts is $n = (2,2,1)$. For all probabilities P_i in P we consider the NID priors with Laplace exponents $\alpha_i\psi(u) = \alpha_i\sqrt{2u}$ (this is the Laplace exponent of the 1/2-stable distribution with scale parameter α_i). Hence, it follows that

$$
\psi^{(N)}(u) = (-1)^{N-1}\frac{\Gamma(N-1/2)}{\sqrt{2\pi}}u^{1/2-N},
$$

and thus one has

$$
\begin{aligned}
B_1^i &= -\alpha_i(2u)^{-1/2}, \\
B_2^i &= \det\begin{bmatrix} -\alpha_i(2u)^{-1/2} & \alpha_i(2u)^{-3/2} \\ -1 & -\alpha_i(2u)^{-1/2} \end{bmatrix} \\
&= \alpha_i(2u)^{-3/2} + \alpha_i^2(2u)^{-1}, \\
B_3^i &= \det\begin{bmatrix} -\alpha_i(2u)^{-1/2} & 2\alpha_i(2u)^{-3/2} & -\alpha_i 3(2u)^{-5/2} \\ -1 & -\alpha_i(2u)^{-1/2} & \alpha_i(2u)^{-3/2} \\ 0 & -1 & -\alpha_i(2u)^{-1/2} \end{bmatrix} \\
&= -3\alpha_i(2u)^{-5/2} - 3\alpha_i^2(2u)^{-2} - \alpha_i^3(2u)^{-3/2}.
\end{aligned}
$$

We can now compute the posterior expectation of P_1 which, from (13), is equal to:

$$E[P_1|n] = \frac{E[P_1^3 P_2^2 P_3]}{E[P_1^2 P_2^2 P_3]} = -\frac{1}{5}\frac{\int_0^\infty u^5 e^{-s\sqrt{2u}} B_3^1 B_2^2 B_1^3 du}{\int_0^\infty u^4 e^{-s\sqrt{2u}} B_2^1 B_2^2 B_1^3 du}$$
$$= -\frac{1}{5}\frac{\int_0^\infty u^5 e^{-s\sqrt{2u}}[(2u)^{-2}+\alpha_2(2u)^{-3/2}][-3(2u)^{-5/2}-3\alpha_1(2u)^{-2}-\alpha_1^2(2u)^{-3/2}]du}{\int_0^\infty u^4 e^{-s\sqrt{2u}}[(2u)^{-2}+\alpha_2(2u)^{-3/2}][(2u)^{-3/2}+\alpha_1(2u)^{-1}]}$$
$$= \frac{1}{10}\frac{3(20\alpha_1^2\alpha_2+4\alpha_1(\alpha_1+3\alpha_2)s+3(\alpha_1+\alpha_2)s^2+s^3)}{s(12\alpha_1\alpha_2+3(\alpha_1+\alpha_2)s+s^2)}. \tag{14}$$

Equation (12) and Proposition 1 allow us to compute the prior and posterior moments of polynomial functions of P_i. As it will be shown in Section 7, this is all we need to compute predictive inferences. Thus, using the Lévy-Khintchine representation theorem, it is possible to derive distributional properties of NID distributions. This result is important because many NID do not admit a closed form expression for the probability density function. There are only three members of the NID class that have a density: the Gamma distribution, the Inverse Gaussian distribution, and the 1/2-stable distribution. When the PDF of X_i, denoted by g_i, admits a closed-form expression for every $i = 1,\dots,M$, the PDF of the vector P is:

$$g(p) = \int_0^\infty \prod_{i=1}^{M} g_i(p_i w) w^{M-1} dw. \tag{15}$$

where $p = (p_1,\dots,p_{M-1})$ and $p_M = 1-\sum_{i=1}^{M-1} p_i$. This can be proven by applying the change of variable theorem for PDFs.

Example 1 (continued). If X_i is Gamma-distributed with parameters $(\alpha_i, 1)$, then $g_i(x_i) \propto x_i^{\alpha_i-1}\exp(-x_i)$, and thus, neglecting the normalization constant, one derives from (15):

$$\begin{aligned}
&\int_0^\infty \prod_{i=1}^{M-1} (p_i w)^{\alpha_i-1}\exp(-p_i w)\\
&\cdot(w - w\textstyle\sum p_i)^{\alpha_M-1}\exp(-w(1-\sum_{i=1}^{M-1} p_i))w^{M-1}dw\\
&\propto p_1^{\alpha_1-1} p_2^{\alpha_2-1}\cdots p_M^{\alpha_M-1},
\end{aligned} \tag{16}$$

which is the PDF of the Dirichlet distribution.

3. The Imprecise NID models

We have already introduced the concept of prior near-ignorance in the introduction. A formal definition of a state of ignorance about a function f is now given.

Definition 2. A set of priors $\mathscr{M}$ models a state of ignorance about a real-valued bounded function f if the lower and upper envelopes of the expectation of f w.r.t. this class satisfy $\underline{E}[f]=\inf f$ and, respectively, $\overline{E}[f]=\sup f$.

This means that the range of $E(f)$ under the class $\mathscr{M}$ is the same as the original range of f. Walley has shown that prior ignorance can only be imposed to a subset of the possible functions of P (for this reason the model is called near-ignorant) otherwise it produces vacuous posterior inferences [4, Ch. 5], which means that we do not learn from data. However, near-ignorance guarantees prior ignorance for many of the inferences of interest in statistical analysis and, at the same time, allows learning and consistency.

The NID class can be employed to derive new prior near-ignorance models on the simplex. For this purpose we write the Laplace exponent of the M ID variables $X_1,\ldots,X_M$ as the product $\alpha_i\psi_i$ of a positive parameter α_i times the rescaled Laplace exponent ψ_i. For instance, in Example 2 the Laplace exponents of the variables X_i are written as the product of the positive parameter α_i and the Laplace exponent $\psi_i(u)=\sqrt{2u}$ of a 1/2-stable distribution with scale parameter 1. The NID distributions obtained by normalizing $X_1,\ldots,X_M$ is identified by the vector of Laplace exponents $(\psi_1,\ldots,\psi_M)$ and the vector of parameters $\alpha=(\alpha_1,\ldots,\alpha_M)$. A near-ignorance set of priors, called imprecise NID (INID) set, is obtained from such NID distributions by allowing α to vary in the interior of the simplex $\Delta_M(s)$. Notice that the choice of α_i and ψ_i is arbitrary (for instance, the same Laplace exponents of Example 2 could be written as the product $\alpha'_i\psi'_i(u)$ of the positive parameter $\alpha'_i=\alpha_i\sqrt{2}$ times the Laplace exponent $\psi'_i=\sqrt{u}$ of a 1/2-stable distribution with scale parameter $1/\sqrt{2}$). However, the simple transformation $\alpha'_i=c\,\alpha_i, \psi'_i=\psi_i/c$, with $c\in\mathbb{R}^+$, leads from a representation $\alpha_i\psi_i$ of the Laplace exponent of X_i to any other possible representation $\alpha'_i\psi'_i$; thus an INID model obtained using a representation $\alpha'_i\psi'_i$ can automatically be derived from the model using $\alpha_i\psi_i$.

Theorem 1. *The INID set of priors is a model of prior ignorance for all the non-central moments of arbitrary order of the variables $P_1,\ldots,P_M$, i.e.,*

$$\underline{E}[P_i^r]=0, \quad \overline{E}[P_i^r]=1, \qquad \forall\ i=1\ldots,M \text{ and } r=1,2,\ldots$$

We will distinguish two kinds of INID models depending on whether the Laplace exponents $\psi_i(u)$ are fixed and equal for all the ID variables X_i or are let free to vary.

Definition 3. An INID set of priors is said to be *homogeneous* if $\psi_i(u) = \psi(u)$ for all $i = 1, \ldots, M$. A *non-homogeneous* INID model is obtained, instead, by allowing the Laplace exponent $\psi_i(u)$ to be different for each i and free to vary in the non-degenerate set of Laplace exponents:

$$\mathscr{L} = \left\{ (\psi_1, \ldots, \psi_M) : \psi_i = \int_0^\infty (1 - e^{-ux}) \nu_i(dx) \text{ with } \nu_i \in \mathscr{V} \right\},$$

where $\mathscr{V}$ is a set of well-defined Lévy measures.

In the homogeneous model the Laplace transforms of the different variables X_i differ only for the constant α_i. We will show that the IDM belongs to this class of models with $\psi(u) = \ln(1+u)$. The non-homogeneous INID models introduce additional degrees of freedom to the Laplace exponents and allow us to obtain models that are more general than the IDM. For instance, the Gamma-INID model discussed in Section 6 includes the IDM. In this case, the upper and lower bounds for the inferences of interest have to be found by considering not only all possible values of $\alpha \in \Delta_M(s)$ but also all possible Laplace exponents $\psi_i(u) \in \mathscr{L}$.

The values of the parameters in α and of the Laplace exponents that give the lower and upper posterior expectations for P_i cannot be determined in general, i.e., they depend on the particular NID distributions we are considering. In the models studied in this work, the lower bound of the posterior expectation of P_i is often found for $\alpha_i \to 0$ and the upper bound for $\alpha_j \to s$ (i.e., $\alpha_j \to 0$ for all $j \neq i$). Note that, in these cases the expression of the posterior moments (13) can be simplified, since, when $\alpha_j \to 0$, one has [2]

$$B_{n_j}(-\alpha_j \psi_j^{(1)}(u),\ldots,-\alpha_j \psi_j^{(n_j)}(u)) = -\alpha_j \psi_j^{(n_j)}(u) + o(\alpha_j). \tag{17}$$

We will exploit this simplification in the proof of the next theorems.

Example 2 (continued). An imprecise model for the NID stable prior in Example 2 is obtained by letting the vector of parameters α vary in the interior of the simplex $\Delta_M(s)$. In this case we are interested in the upper and lower expectations of P_1 which are found by letting $\alpha_i \to 0$ for all but one of the parameters α_i (this follows from Theorem 3 in the next section, but can also be verified based on the posterior expectation in (14)). For example the lower expectation of P_1 is found for $\alpha_1, \alpha_2 \to 0$ and is given by:

$$\underline{E}[P_1|n] = \frac{1}{10} \frac{\int_0^\infty 3(2u)^{1/2} e^{-s\sqrt{2u}} du}{\int_0^\infty (2u)^{1/2} e^{-s\sqrt{2u}} du} = 3/10,$$

whereas the upper is found for $\alpha_1 \to s$:

$$\overline{E}[P_1|n] = \frac{1}{10} \frac{\int_0^\infty [3(2u)^{1/2} + 3s(2u)^1 + s^2(2u)^{3/2}] e^{-s\sqrt{2u}} du}{\int_0^\infty [(2u)^{1/2} + s(2u)^1] e^{-s\sqrt{2u}} du} = \frac{3}{5}.$$

In section 4 we give a closed form expression for these posterior expectations.

A general discussion about the properties of the INID models is limited by the fact that most properties depend on the Lévy measure of the distributions considered. Therefore, in the next sections we derive specific models using normalized γ-stable, inverse Gaussian and Gamma distributions and study them individually.

---

[2]This is obtained by substituting $\alpha_j \psi_j^{(i)}$ to x_i, $i = 1,\ldots,n_j - k + 1$ in (11) and observing that each term $B_{n_j,k}(\cdot)$ of $B_{n_j}(\cdot)$ is proportional to α_j^k so that for all terms with $k > 1$ it holds

$$\lim_{\alpha_j \to 0} \frac{B_{n_j,k}(\cdot)}{\alpha_j} = 0,$$

while for $k = 1$ we have $B_{n_j,1}(\cdot) = -\alpha_j \psi^{(n_j)}$.

4. γ-stable INID model

Consider now the case where the ID variables $X_1,\ldots,X_M$ have positive stable distribution $X_i \sim St(\gamma,\beta,\alpha_i,\mu)$ with characteristic exponent $\gamma\in(0,1)$, skewness parameter $\beta=1$, scale parameter $\alpha_i>0$, and a location parameter $\mu=0$ [19]. In the following, we will refer to this distribution as the γ-stable distribution. Its Lévy density and Laplace exponent can be written as the product of the scale parameter α_i times the Lévy density and Laplace exponent

$$\nu(dx)=\frac{x^{-1-\gamma}}{\sqrt{2\pi}}dx, \quad \text{and} \quad \psi(u)=\frac{\Gamma(1-\gamma)}{\sqrt{2\pi}\gamma}u^{\gamma} \tag{18}$$

of the γ-stable distribution with scale parameter 1. Then, one has

$$\psi^{(N)}(u)=(-1)^{N-1}\frac{\Gamma(N-\gamma)}{\sqrt{2\pi}}u^{\gamma-N}. \tag{19}$$

In general, the PDF of a stable distribution does not admit a closed-form expression, with an exception for $\gamma=1/2$, for which we have the so called Lévy distribution with PDF:

$$f(x|\alpha_i)=\frac{\alpha_i}{(2\pi)^{1/2}}x^{-3/2}\exp\left(\frac{\alpha_i^2}{2x}\right), \quad x\in\mathbb{R}^+. \tag{20}$$

Consequently, the normalized 1/2-stable distribution has PDF:

$$g(p)=\Gamma\left(\frac{M}{2}\right)\pi^{-\frac{M}{2}}\left(\prod_{i=1}^{M}\alpha_i p_i^{-\frac{3}{2}}\right)\left(\sum_{i=1}^{M}\frac{\alpha_j^2}{p_i}\right)^{-\frac{M}{2}}. \tag{21}$$

When the vector of probabilities P is modeled by a NID γ-stable distribution with γ fixed and equal for all variables $X_1,\ldots,X_M$ and the vector of parameters α is free to vary in the simplex $\Delta_M(s)$, we obtain an imprecise homogeneous NID model. We will refer to this model as γ-stable INID. A priori we can state the following result.

Theorem 2. *For the set of priors defined by the γ-stable INID model, it holds:*

$$\underline{E}[P_i^r]=0, \quad \overline{E}[P_i^r]=1, \tag{22}$$

$$\underline{E}[P_iP_j]=0, \quad \overline{E}[P_iP_j]=\frac{1}{4}\gamma, \tag{23}$$

for any i,j and $r=1,2,\ldots$ and $\gamma\in(0,1)$.

Note that, for $\gamma \to 1$ one has $\overline{E}[P_i P_j] \to \frac{1}{4}$ which means that the model is prior ignorant even with respect to the second order mixed moments. In fact, for $\gamma \to 1$ all priors in the set tend to atomic measures concentrated on the expected value $(\alpha_1/s, \ldots, \alpha_m/s)$ (this follows from the fact that the variance $\mathrm{Var}[P_i] = (1-\gamma)\alpha_i(s-\alpha_i)/s^2$ goes to zero) and thus the model becomes fully ignorant and all inferences derived from it are vacuous. A posteriori the following result holds.

Theorem 3. *Given $\gamma \in (0,1)$, if the following inequality holds for all $k = 2, \ldots, n_i$:*

$$\Gamma(1-\gamma)C_{n_i,k-1} - \gamma(k-1)C_{n_i,k} \geq 0, \tag{24}$$

where $C_{n_i,k} = B_{n_i,k}(\Gamma(1-\gamma), \ldots, \Gamma(n_i - k + 1 - \gamma))$ or, equivalently,

$$kS_{n_i,k-1}(\gamma) - (k-1)S_{n_i,k}(\gamma) \geq 0, \tag{25}$$

where

$$S_{n_i,k}(\gamma) = \sum_{a_1=k-1}^{n_i-1} \sum_{a_2=k-2}^{a_1-1} \cdots \sum_{a_{k-1}=1}^{a_{k-2}-1} \binom{n_i}{a_1}\binom{a_1}{a_2}\cdots\binom{a_{k-2}}{a_{k-1}} \times$$
$$\times \prod_{i=1}^{n_i-a_1-1}\left(\frac{i}{\gamma}i-1\right) \prod_{i=1}^{a_1-a_2-1}\left(\frac{i}{\gamma}-1\right) \cdots \prod_{i=1}^{a_k-1}\left(\frac{i}{\gamma}-1\right),$$

then, for the γ-stable INID model, the lower and upper posterior expectation of P_i are found, respectively, for $\alpha_i \to 0$ (the values of α_j for $j \neq i$ are irrelevant) and $\alpha_i \to s$ and are equal to:

$$\underline{E}[P_i|n] = \max\left(0, \frac{n_i - \gamma}{N}\right), \tag{26}$$

and

$$\overline{E}[P_i|n] = \begin{cases} \min\left(\dfrac{n_i + (m-1)\gamma}{N}, 1\right), & n_i > 0, \\ \dfrac{m\gamma}{N}, & n_i = 0, \end{cases} \tag{27}$$

where m is the number of observed categories, i.e. $m = \sum_{i=1}^{M} I_{n_i>0}$

Conjecture 1. *Condition (24) holds if and only if $\gamma \leq 1/2$.*

The conjecture has been confirmed by many numerical evaluations, but unfortunately we could not prove it even in the simpler case $\gamma = 1/2$ for which

$$\prod_{i=1}^{a}\left(\frac{i}{\gamma}-1\right)=(2a-1)!!,$$

where $(2a-1)!!$ is the double factorial (i.e., the product of all the odd integers up to $2a-1$). However, the inequality (24) can be easily verified numerically (especially for small sample sizes) once the value n_i for the specific problem is given.

If (24) holds, the posterior lower and upper expectations are given by (26)-(27) and thus, they can be computed by means of simple algebraic expressions. Note that, in both cases, the posterior expectations do not depend on the value of s. The upper expectation in (27) obtained for $\alpha_i \to s$ increases linearly with the number of observed categories m whereas the lower expectation in (26) is constant with m. Then the imprecision $\overline{E}[P_i|n]-\underline{E}[P_i|n]$ increases linearly with m. If we interpret m as a measure of the complexity of our representation of the data, then we can say that the imprecision of the inferences provided by the γ-stable INID model increases with the complexity. Since we have assumed that the total number of categories M is bounded, and thus $m < M$ is also bounded, the upper and lower expectations of P_i converge both to the frequency $\frac{n_i}{N}$ as $N \to \infty$.

When $\gamma > 1/2$ it is possible to find counter-examples which show that the condition in Theorem 3 is not verified. In these cases the lower posterior expectation can be obtained for a value $\alpha_i > 0$ which has to be found by minimizing (13) numerically.

Example 3. Consider a set of γ-stable INID priors. From [10, Proposition 1] we can derive the CDF of P_1 as

$$E[I_{[0,p]}(P_1)]=\frac{1}{2}-\frac{1}{\pi\gamma}\arctan\left[\frac{\alpha_1(1-p)^\gamma-(1-\alpha_1)p^\gamma}{\alpha_1(1-p)^\gamma+(1-\alpha_1)p^\gamma}\tan\frac{\pi\gamma}{2}\right]. \tag{28}$$

Assuming that there are only $M = 2$ categories, the PDF of the prior can be obtained by differentiating (28) with respect to p. Figure 1 shows the prior PDF for $\gamma = 0.5$ (left) and $\gamma = 0.8$ (right) when $\alpha_1 = 0.1$ and $\alpha_1 = 0.01$. When $\gamma = 0.8$ the prior PDF has a local maximum around $p = 0.05$ for $\alpha_1 = 0.1$, and around $p = 0.002$ for $\alpha = 0.01$ (in this case the maximum reaches a value of almost 200 and thus falls outside the scale of the figure). Instead, there are no local maxima in the prior PDF when $\gamma = 0.5$. Assume that after $N = 3$ observations we have $n_1 = 2$ and $n_2 = 1$. The posterior means of P_1 in the four cases of Figure 1 are listed in Table 1. Notice that when $\gamma = 0.8$ the posterior mean does not increase

	$\alpha_1 = 0.01$	$\alpha_1 = 0.1$
$\gamma = 0.5$	0.500	0.508
$\gamma = 0.8$	0.386	0.347

Table 1: Posterior mean of P_1 in the four cases of Figure 1.

with α_1. This means that the lower expectation is not found for $\alpha_1 \to 0$. An intuition about the reasons behind this result can be obtained by considering the posterior PDFs shown in figure 2 for the four cases considered. Let us focus on the case $\gamma = 0.8$: when $\alpha_1 = 0.01$ the local maximum of the prior is very close to 0 and thus it is strongly reduced by the small value of the likelihood around zero. Thus, the posterior tends to spread across all possible values of p_1 more than for $\alpha_1 = 0.1$, since in this second case the mass concentrates more in the neighborhood of the local maximum of the prior, which is close to 0.

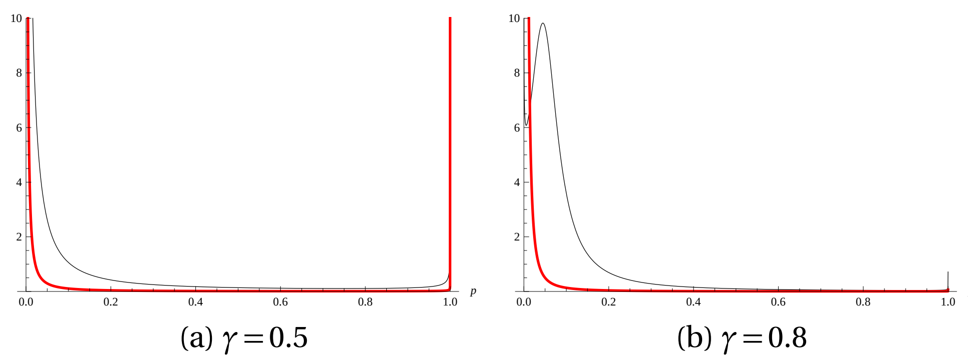
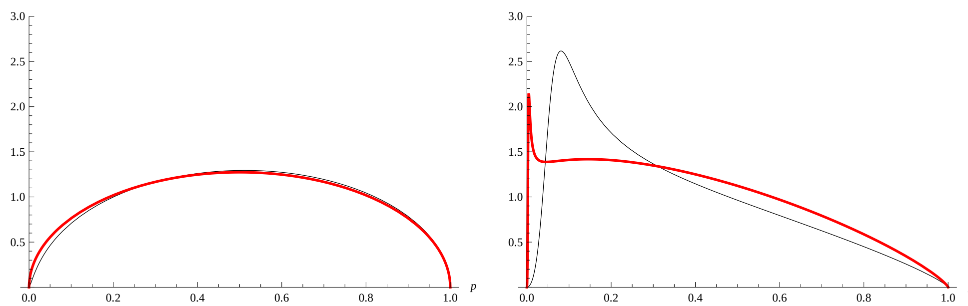


Figure 1: Prior distribution of P_1 for $\alpha_1 = 0.1$ (black thin line) and $\alpha_1 = 0.01$ (red thick line)

Figure 3 shows the posterior lower and upper expectations of P_1 as a function of γ. It can be noticed that $\underline{E}[P_1]$ decreases linearly for $\gamma \leq 1/2$, since (26) holds, and decreases faster for larger values of γ. The upper, instead, increases linearly for all values of γ, since, for the particular values of n_1 and n_2 used in this example, the upper is always found for $\alpha_1 \to 1$ and thus (27) always holds. For $\gamma \to 1$ the model becomes vacuous (cannot learn from data) and thus one has $\underline{E}[P_1] \to 0$ and $\overline{E}[P_1] \to 1$ regardless of the number of observations available.



(a) $\gamma = 0.5$ (b) $\gamma = 0.8$

Figure 2: Posterior distribution of P_1 for $\alpha_1 = 0.1$ (black thin line) and $\alpha_1 = 0.01$ (red thick line)

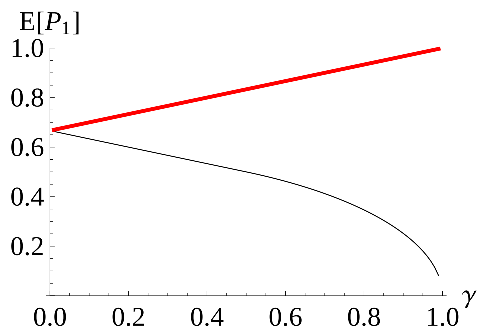


Figure 3: Posterior lower (black thin line) and upper (red thick line) expectations of P_1

5. Inverse Gaussian INID model

Consider now M ID variables $X_1, \ldots, X_M$ having Inverse-Gaussian (IG) distribution $X_i \sim IG(\alpha_i, \lambda)$ with shape parameter $\alpha_i \geq 0$ and scale parameter $\lambda > 0$. We can write their Lévy measures and Laplace exponents as the product of the shape parameter α_i and the Lévy measures and Laplace exponent

$$\nu(dx) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\lambda^2 x} x^{-\frac{3}{2}} \quad \text{and} \quad \psi(u) = \sqrt{2u + \lambda^2} - \lambda \tag{29}$$

of a IG distribution with shape parameter $(1, \lambda)$. Then, one has

$$\psi^{(N)} = (-1)^{N-1} \frac{2^{N-1}}{\sqrt{\pi}} \Gamma\left(N - \frac{1}{2}\right)(2u + \lambda^2)^{-N+\frac{1}{2}}. \tag{30}$$

A closed form expression for the PDF of the IG distribution exists [10]:

$$f(x|\alpha_i) = \frac{\alpha_i}{(2\pi)^{1/2}} \exp\left[-\frac{1}{2}\left(\frac{\alpha_i^2}{x} + \lambda^2 x\right) + \lambda\alpha_i\right], \quad x \in \mathbb{R}^+. \tag{31}$$

Therefore the PDF of the Inverse Gaussian NID distribution is:

$$g(p)=\frac{2e^{\sum_{i=1}^{M}\alpha_i}}{(2\pi)^{M/2}}\mathrm{K}_{-M/2}\left(\sqrt{\sum_{i=1}^{M}\frac{\alpha_j^2}{p_i}}\right)\prod_{i=1}^{M}\alpha_i p_i^{-3/2}\left(\sum_{i=1}^{M}\frac{\alpha_j^2}{p_i}\right)^{-\frac{M}{4}}, \tag{32}$$

where $\mathrm{K}_{-M/2}$ is the modified Bessel function of the second kind of order $-M/2$.

We assume λ fixed and identical for all ψ_i and we obtain a homogeneous INID model (which we call IG-INID model) by letting α free to vary in the interior of simplex $\Delta_M(s)$. A priori we can state the following result.

Theorem 4. *Given the set of priors defined by the IG-INID model, for any i, j and $r = 1,2,\ldots,$ it holds:*

$$E[P_i^r]=0, \quad \overline{E}[P_i^r]=1, \tag{33}$$

$$\underline{E}[P_iP_j]=0, \quad \overline{E}[P_iP_j]=\frac{1}{4}(1-s^2\mathrm{e}^s\Gamma(-2;s)), \tag{34}$$

where $\Gamma(a;b)=\int_b^\infty t^{a-1}e^{-t}\,dt$, with $a\in\mathbb{R}$ and $b>0$, is the extended incomplete gamma function.

A posteriori we have the following result.

Theorem 5. *If the following inequality holds for all $k=2,\ldots,n_i$:*

$$\sqrt{\pi}C_{n_i,k-1}-\frac{k-1}{2}C_{n_i,k}\geq 0, \tag{35}$$

or, equivalently,

$$kS_{n_i,k-1}(1/2)-(k-1)S_{n_i,k}(1/2)\geq 0,$$

then, for the IG-INID set of priors, the lower posterior expectation of P_i is found for $\alpha_i\to 0$. When $\alpha_i\to 0$ the expectation depends on the values of α_j for $j\neq i$. The lower is found for $\alpha_{j^*}\to s$, where the category $j^*\neq i$ is the one with the minimum number of observations. The lower expectation is

$$\underline{E}[P_i|n]=\begin{cases} 0, & n_i=0,\\[2ex] \dfrac{n_i-\frac{1}{2}}{N}\dfrac{-(s\lambda)^2\sum\limits_{r=1}^{n_{j*}}W(N+1,s\lambda,M+r-1)C_{n_{j*},r}(2\sqrt{\pi})^{-r}}{\sum\limits_{r=1}^{n_{j*}}W(N,s\lambda,M+r-1)C_{n_{j*},r}(2\sqrt{\pi})^{-r}}, & n_i,n_{j*}>0,\\[2ex] \dfrac{n_i-\frac{1}{2}}{N}\dfrac{-(s\lambda)^2W(N+1,s\lambda,m)}{W(N,s\lambda,m)}, & n_i>0,n_{j*}=0, \end{cases} \tag{36}$$

where

$$W(N,s\lambda,m)=\sum_{k=0}^{N-1}\binom{N-1}{k}[-(s\lambda)^2]^{-k}\Gamma(2+m+2k-2N;s\lambda).$$

Notice that if $n_{j*}=0$ then the number of observed categories is $m<M$; if instead $n_i,n_{j*}>0$, then $M=m$. If the upper expectation is found for $\alpha_i\to s$, then it is given by

$$\overline{E}[P_i|n]=\begin{cases}\dfrac{-(s\lambda)^2}{N}\dfrac{\sum\limits_{r=1}^{n_i+1}W(N+1,s\lambda,m+r-1)(2\sqrt{\pi})^{-r}C_{n_i+1,r}}{\sum\limits_{r=1}^{n_i}W(N,s\lambda,m+r-1)(2\sqrt{\pi})^{-r}C_{n_i,r}}, & n_i>0,\\[2ex] \dfrac{-(s\lambda)^2}{2N}\dfrac{W(N+1,s\lambda,m+1)}{W(N,s\lambda,m)}, & n_i=0.\end{cases} \tag{37}$$

Note that, despite the minus sign, the right hand sides of (36) and (37) are non-negative since the numerator and denominator always have opposite sign. This result is analogous to the one obtained in [13, Proposition 3] for the normalized inverse Gaussian process. The upper and lower expectations are only influenced by the product $s\lambda$, and not by the value of the individual parameters. Then, without loss of generality, in the homogeneous model we can take $\lambda=1$. In the IG-INID model, both the lower and upper expectations depend on the number of observed categories. The condition for the lower in Theorem 5 is the same found for the γ-stable INID model with $\gamma=1/2$, which we have conjectured to be satisfied. Concerning the upper we can state the following conjecture

Conjecture 2. *The upper expectation of P_i for the IG-INID model is always found for $\alpha_i\to s$.*

We believe that this conjecture is true, since it has been confirmed by several numerical examples.

Example 4. Consider a IG-INID model with $s=1$ and $\lambda=1$ and assume that, after $N=5$ samples, we have observed $m=3$ categories with $n_1=1$, $n_2=3$ and

$n_3 = 1$ counts. We are interested in estimating P_1. As a verification of Conjecture 2 Figure 4 shows the posterior expectation of P_1 when α_1 spans the interval $[0,1]$ and $\alpha_2 \to 1-\alpha 1$ (then $\alpha_3 \to 0$). Figure 4 shows that by taking only two parameters α_1 and α_2 different from 0, the supremum of $E[P_i|n]$ is clearly found for $\alpha_1 \to s$.

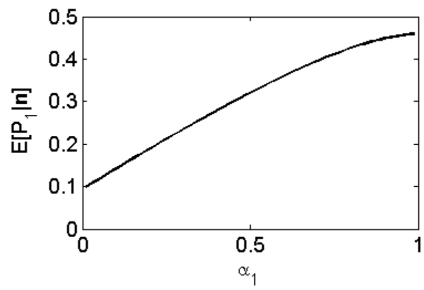


Figure 4: Posterior expectation of P_1.

If Conjectures 1 and 2 are true, we can see from (36) and (37) that the upper and lower expectations of P_i depend on m. From numerical examples (see for instance Figure 5) it can be seen that the lower $\underline{E}[P_i|n]$ is almost constant with m whereas, for sufficiently large values of m, $\overline{E}[P_i|n]$ grows almost linearly with m with slope close to that of the upper expectation obtained from the 1/2-stable INID model, that is $1/(2N)$ (see equation (27)). Thus, the imprecision increases almost linearly with m.

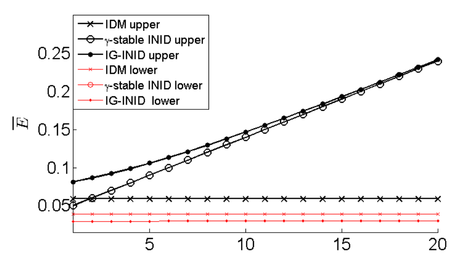


Figure 5: Posterior expectation of P_1 for the IDM, the γ-stable and the IG-INID models when $n_1 = 2$, $N = 50$, $s = 1$, and m ranging from 1 to 20.

Assume now that we are interested in making predictive inferences about the probability of observing at the next draw a category not yet observed in the previous trials, that is $P_{\text{new}} = \sum_{i=1}^{M} P_i I_{n_i=0}$, where the indicator function $I_{n_i=0}$ is 1 if no observation has fallen in the i-th category. From (36) we can see that for each category i such that $n_i = 0$ the lower expectation $\underline{E}[P_i|n] = 0$ is found for $\alpha_i \to 0$ regardless of the value of all other parameters α_j, $j \neq i$. The lower expectation for P_{new} can then be found by letting $\alpha_i \to 0$ for all categories for which $n_i = 0$, thus obtaining

$$\underline{E}[P_{\text{new}}|n] = \sum_{i=1}^{M} \underline{E}[P_i|n] I_{n_i=0} = 0.$$

The upper expectation of P_{new} can be found by considering that

$$\overline{E}\left[\sum_{i=1}^{M} P_i I_{n_i=0} \middle| n\right] = 1 - \underline{E}\left[\sum_{i=1}^{M} P_i I_{n_i>0} \middle| n\right].$$

From Theorem 5 we can see that for each category i such that $n_i > 0$ the lower expectation is found by letting $\alpha_i \to 0$ and $\alpha_j \to s$ for one of the categories j for which $n_j = 0$. We can then find the upper expectation of P_{new} by letting $\alpha_j \to s$ for an arbitrary unobserved category j. Then we obtain

$$\overline{E}[P_{\text{new}}|n] = E\left[\sum_{i=1}^{M} P_i I_{n_i=0} \middle| n, \alpha_j \to s\right] = \overline{E}[P_j|n]$$

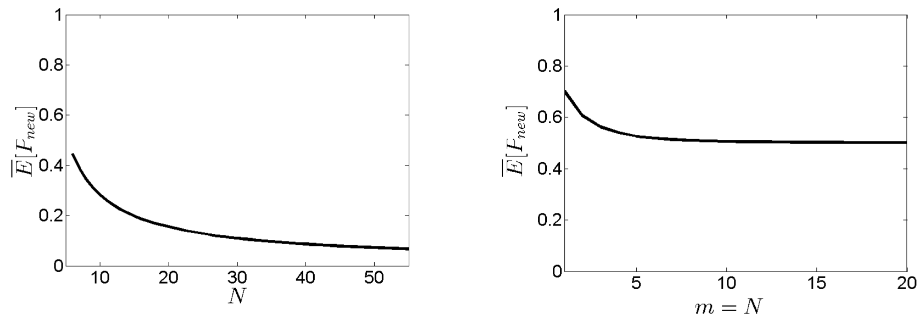
since for such choice of the parameters we have that $\alpha_i = 0$ for all unobserved categories $i \neq j$ and thus the posterior expectation of their probabilities P_i is 0. We can see empirically in the next example that the upper expectation $\overline{E}[P_{\text{new}}|n]$ decreases with the total number of samples N.

Example 5. Consider a situation where we have collected, for instance, $N > 5$ draws. Figure 6 shows how $\overline{E}[P_{\text{new}}|n]$ varies with N when (a) $m = 5$ and (b) when $m = N$, i.e. at each draw a new category is observed.

Proposition 2. *Assuming that condition (35) holds, the function*

$$\frac{-(s\lambda)^2}{2N} \frac{W(N+1, s\lambda, m+1)}{W(N, s\lambda, m)}$$

that gives the upper probability of observing a new category at the next draw converges, for $N \to \infty$, to 0 if m is constant and to $1/2$ if $m = N$.



(a) $m=5$,

(b) $m=N$

Figure 6: Upper posterior expectation of P_{new}.

Let $\hat{m}$ be the *true* number of different categories that can be observed. Assume that after the first N' draws we have observed all $\hat{m}$ categories. If $\hat{m}=M$, then $\overline{E}[P_{\text{new}}|n]=0$ since all categories have already been observed. If $\hat{m}<M$ (that is, if our model includes more categories than needed), then for $N>N'$ the number of observed categories remains constant ($m=\hat{m}$) and thus it follows from Proposition 2 that $\overline{E}[P_{\text{new}}|n]$ can be made arbitrarily closed to 0 by increasing N. This is not true, instead, if new categories keep appearing as more observations are collected (this can happen for any finite value of N provided that $\hat{m}$ and M are sufficiently large). For instance, we can consider the case $m=N$: from Proposition 2, we have that $\overline{E}[P_{\text{new}}|n]$ can be made arbitrarily closed to $1/2$ (and not to 0) by increasing the number of observations N. Admitting the possibility that $m=N$ makes sense only in a situation where the total number M of categories included in the model is much larger than N; this is a reasonable model , for example, if we are a priori ignorant about the process that generates the data. This situation, as well as the results obtained for $\overline{E}[P_{\text{new}}|n]$ will be discussed in Section 8.

6. Gamma INID model

Suppose that all the variables X_i are Gamma distributed with parameters $\alpha_i>0$ (shape) and $\beta_i>0$ (scale). The corresponding PDF of X_i is

$$f(x|\alpha_i,\beta_i)=\frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}x^{\alpha_i-1}e^{-\beta_i x}.$$

We can write its Lévy density and Laplace exponent as the product of shape parameter α_i and the Lévy density is $\nu_i(dx)=x^{-1}e^{-\beta_i x}dx$ and Laplace exponent

$$\psi_i(u) = \log\left(1 + \frac{u}{\beta_i}\right) \tag{38}$$

of a Gamma distribution with shape parameter $(1,\beta_i)$. The Gamma-NID distribution admits a closed form for the PDF of P, that is

$$g(p) \;=\; \Gamma(s)\prod_{i=1}^{M}\frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}p_i^{\alpha_i-1}\left(\sum_{i=1}^{M}\beta_i p_i\right)^{-s}. \tag{39}$$

From (39) it can be noticed that if $\beta_i = \beta$ for $i = 1,\dots,M$ the Gamma-NID distribution reduces to to the Dirichlet distribution. This corresponds to the homogeneous case, i.e., $\psi_i(u) = \psi(u)$ for $i = 1,\dots,M$. In the following, when it is not explicitly mentioned, we will assume that the parameters β_i are not equal, i.e, by default we will consider the non-homogeneous case.

Our aim is to derive the prior and posterior moments of the Gamma-NID. First of all, note that from (38) it follows:

$$(-1)^r \frac{d^r}{du^r} e^{-\alpha_i\psi_i(u)} = \frac{\Gamma(\alpha_i + r)}{\Gamma(\alpha_i)\beta_i^r}\left(1 + \frac{u}{\beta_i}\right)^{-\alpha_i - r}. \tag{40}$$

Hence, from (7) and (38)–(40), we can derive the expression for the prior moments:

$$E[P_1^{r_1}P_2^{r_2}\cdots P_M^{r_M}] = \frac{1}{\Gamma(r)}\prod_{k=1}^{M}\frac{\beta_k^{\alpha_k}\Gamma(\alpha_k + r_k)}{\Gamma(\alpha_k)}\int_0^\infty \frac{u^{r-1}}{\prod_{k=1}^{M}\left(\beta_k + u\right)^{\alpha_k + r_k}}du. \tag{41}$$

The posterior expectations of P_i can be computed from (13) and (38)–(40):

$$\begin{aligned} E[P_i|n] \;&=\; \frac{E[P_i P_1^{n_1}\cdots P_M^{n_M}]}{E[P_1^{n_1}\cdots P_M^{n_M}]} \\ &=\; \frac{\alpha_i + n_i}{N}\,\frac{\displaystyle\int_0^\infty u^N\left(\beta_i + u\right)^{-1}\prod_{i}^{M}\left(\beta_j + u\right)^{-\alpha_j - n_j}du}{\displaystyle\int_0^\infty u^{N-1}\prod_{j=1}^{M}\left(\beta_j + u\right)^{-\alpha_j - n_j}du}. \end{aligned} \tag{42}$$

Since the constant $B = \sum_{i=1}^{M}\beta_i$ simplifies a-posteriori, as can be seen from (42) by the change of variables $Bu' = u$, w.l.o.g. we can take $\sum_{i=1}^{M}\beta_i = 1$. Then, to model prior near-ignorance, we consider the set of Gamma-NID distributions obtained by letting α and β vary in the interior of, respectively, $\Delta_M(s)$ and $\Delta_M(1)$. Notice that this is a non-homogeneous INID model, since the functions ψ_i are different and free to vary in

$$\mathscr{L} = \left\{ \begin{array}{ll} (\psi_1,\ldots,\psi_M) : \psi_i = \log\left(1+\frac{u}{\beta_i}\right), & \text{with } i=1,\ldots,M, \\ & \beta_i > 0 \text{ and } \sum_{i=1}^M \beta_i = 1 \end{array} \right\}.$$

We call this imprecise model Gamma-INID model. By fixing $\beta_i = 1/M$ for $i = 1,\ldots,M$ (homogeneous case) we obtain the IDM.

Theorem 6. *A-priori the homogeneous and non-homogeneous Gamma-INID models satisfy:*

$$\underline{E}[P_i^r] = 0, \quad \overline{E}[P_i^r] = 1, \tag{43}$$

$$\underline{E}[P_i P_j] = 0, \quad \overline{E}[P_i P_j] \geq \frac{1}{4}\frac{s}{s+1}, \tag{44}$$

for any i, j and $r = 1,2,\ldots$.

From (43), which follows directly from Theorem 1, we see that both the Gamma-INID and the IDM are prior ignorance models for inferences about P_i^r. This is not surprising, the IDM being included in the Gamma-INID (therefore when the IDM is vacuous so is the Gamma-INID). Note that for the upper expectation of $P_i P_j$ in (44) the equality holds in the homogeneous case (IDM). We were not able to prove it, but we conjecture that the upper bound of $E[P_1 P_2]$ is the same as for the IDM, that is $\overline{E}[P_i P_j] = s/4(s+1)$. Then, as the IDM, the Gamma-INID would not be a prior ignorance model for inferences about $P_i P_j$ and a-priori, would give the same upper expectation for $P_i P_j$ as the IDM (homogeneous). Notice however that this does not hold for inferences about any polynomial function of P_i and P_j; for instance, if we consider the function $P_1 P_2^2$, we obtain for $s = 1$

$$\overline{E}_{IDM}[P_1 P_2^2] = 0.1283, \quad \overline{E}_{\Gamma}[P_1 P_2^2] \approx 0.1301,$$

where the uppr expectation is found for $\alpha_1 \approx 0.42$, $\alpha_2 \approx 0.58$, $\beta_1 \approx 0.5894$, $\beta_2 \approx 0.4106$. Note that the inequality $\overline{E}_{IDM}[P_1 P_2^2] \leq \overline{E}_{\Gamma}[P_1 P_2^2]$ follows from the fact that the IDM is always included in the Gamma-INID model, and thus the inequalities

$$\underline{E}_{\Gamma}[f] \leq \underline{E}_{IDM}[f] \leq \overline{E}_{IDM}[f] \leq \overline{E}_{\Gamma}[f]$$

hold for any real valued bounded function f of P. This is obviously true also for the posterior inferences.

Next theorem gives the posterior expectations of P_i both for the homogeneous and non-homogeneous Gamma-INID model. Although posterior results for the homogeneous model, i.e. the IDM, are well known [8], we believe that it can be helpful to derive them using the NID formalism also in this case, to better understand and compare with the non-homogeneous case.

Theorem 7. *In the homogeneous Gamma-INID model (i.e., $\beta_i = 1/M$ for all $i = 1,\ldots,M$) the lower and upper posterior expectations of P_i are obtained for $\alpha_i \to 0$ (lower) and, respectively, $\alpha_i \to s$ (upper) and are equal to:*

$$\underline{E}[P_i|n] = \frac{n_i}{N+s}, \quad \overline{E}[P_i|n] = \frac{n_i+s}{N+s}. \tag{45}$$

In the non homogeneous Gamma-INID model the lower and upper posterior expectations are obtained for $\alpha_i \to 0, \beta_i \to 1$ (lower) and, respectively, $\alpha_i \to s, \beta_i \to 0$ (upper) and are:

$$\underline{E}[P_i|n] = \max\left(0, \frac{n_i - s}{N}\right), \quad \overline{E}[P_i|n] = \min\left(1, \frac{n_i+s}{N}\right). \tag{46}$$

By looking at (45)–(46), we can highlight the following difference between the homogeneous (IDM) and non-homogeneous case. In the homogeneous case, the lower probability for the second observation to be equal to the first, is $1/(1+s)$, i.e., $1/2$ for $s=1$. In the non-homogeneous case with $s=1$, this lower probability is zero. Walley has shown that, in case $M=2$, the IDM with $s=2$ encompasses all Bayesian inferences derived from the Jeffreys ($\alpha_i = 0.5$), uniform ($\alpha_i = 1$) and Haldane ($\alpha_i \to 0$) priors [8]. For the non-homogeneous Gamma-INID this is already true with $s=1$. Another difference is that in the non-homogeneous case the lower and upper expectations derived in (46) are symmetric w.r.t. the sample mean n_i/N whenever $n_i - s \geq 0$ and $n_i + s \leq N$. Furthermore, the denominator in (46) depends only on N and not on s. Thus, for $n_i - s \geq 0$ and $n_i + s \leq N$, the imprecision $2s/N$ should not be interpreted as additional counts that are added to the observations (as for the IDM) but as a swing scenario in which s counts among the N are moved from a category to another. It should be pointed out that the lower and upper expectations in (46) coincide with those derived in [20, Sec. 5.2] for a near-ignorance model based on finitely additive priors obtained as limits of truncated exponential priors. Moreover, the inferences drawn from the non-homogeneous Gamma-INID with $s=1$ coincide with those of the *Nonparametric Predictive Inference* model [21], although this latter is derived in a very different, non-Bayesian way.

7. Invariance properties

Consider a real valued function h on $\mathscr{Y}^t$, i.e., $h(Y^t)$, where $Y^t = (Y_1, \ldots, Y_t)$ and Y_i represents the i-th observation variable taking the value $y_i \in \mathscr{Y} = \{1, 2, \ldots, M\}$. We are interested in computing the expectation of $h(Y^t)$ given a set of N observations summarized by the vector of counts n:

$$E[h|n] = \sum_{y_1=1}^{M} \cdots \sum_{y_t=1}^{M} h(y^t) P(Y_1 = y_1, \ldots, Y_t = y_t | n), \tag{47}$$

where $y^t = (y_1, \ldots, y_t)$. We call the posterior expectation $E[h|n]$ and the prior expectation $E[h]$ (this corresponds to the case in which $N = 0$) a *predictive inference.*

By exploiting the assumption of independence of $Y_1, \ldots, Y_t$ we can write

$$\begin{aligned}
P(Y_1 = y_1, \ldots, Y_t = y_t | n) = & \int_{\Delta_M(1)} P(Y_1 = y_1, \ldots, Y_t = y_t | p) P(p|n) dp \\
= & \int_{\Delta_M(1)} \prod_{i=1}^{t} P(Y_i = y_i | p) P(p|n) dp \\
= & \int_{\Delta_M(1)} \prod_{i=1}^{t} p_{y_i} \, P(p|n) dp.
\end{aligned}$$

where $p = (p_1, \ldots, p_{M-1})$ and $p_{y_i} = p_j$ whenever $y_i = j$. Therefore, we have that

$$\begin{aligned}
E[h|n] &= \int_{\Delta_M(1)} \sum_{y_1=1}^{M} \cdots \sum_{y_t=1}^{M} h(y^t) \prod_{i=1}^{t} p_{y_i} \, P(p|n) dp \\
&= E[f(P)|n],
\end{aligned} \tag{48}$$

with

$$f(P) = \sum_{y_1=1}^{M} \cdots \sum_{y_t=1}^{M} h(y^t) \prod_{i=1}^{t} P_{y_i}. \tag{49}$$

Note that, by defining $r_j = \sum_{k=1}^{t} I_{y_k=j}$, then $\prod_{i=1}^{t} P_{y_i} = \prod_{j=1}^{M} P_j^{r_j}$. This means that a predictive inference corresponds to computing the expectation of a polynomial function of the parameters $P_1, \ldots, P_{M-1}$ with coefficients $h(y^t) \in \mathbb{R}$. A special case of interest is when $t = 1$, and thus

$$f(P) = \sum_{y_1=1}^{M} h(y_1) P_{y_1};$$

we will then talk about *immediate prediction*.

We can now define a set of invariance principles that may be desirable for a probability model on the simplex.

- **Symmetry principle (SP):** Prior expectation of any function f should be invariant w.r.t. permutations of the categories.

- **Representation Invariance Principle (RIP):** The RIP states that posterior expectations of any function f should not depend on refinement or coarsening of categories, provided that f represents the same predictive inference.

- **Likelihood principle (LP):** Posterior inferences should depend on the data through the likelihood function only.

- **Coherence principle (CP):** Prior and posterior inferences should be strongly coherent. Coherence is a self-consistency requirement, ensuring that combinations of several bets cannot lead to a sure loss and have consistent behavioral implications [4, Sec. 7.1.4(b)].

Frequentist methods typically violate LP and CP, whereas SP and RIP are mutually exclusive for Bayesian models using proper priors [4, Sec. 5.5]. It is known that the IDM satisfies all these principles jointly. The question we aim to answer in this section is if there are other prior near-ignorance models based on NID distributions that satisfy these principles. It is clear that all prior near-ignorance models based on NID distributions satisfy SP, LP and CP (for the same reasons as IDM). Furthermore, since in Theorems 3 and 5 we have shown that the posterior upper expectation of P_i for the γ-stable INID and Inverse Gaussian INID models depend on the number of observed categories, it follows that these two INID models do not satisfy RIP. Conversely, we will show in the following theorems that the (homogeneous and non-homogeneous)

Gamma-INID models satisfies RIP. The result for the homogeneous case was already known [8] because the Gamma-INID model coincides with IDM in this case. However, before proceeding with the proof for the non-homogeneous case, we prove RIP by means of a transformation of coordinates also for the homogeneous case. This will be helpful to understand the proof for the non-homogeneous case and to highlight the differences between the homogeneous and non-homogeneous case.

To simplify the notation, we will denote by $g_\alpha(p)$ the unnormalized density function of the Dirichlet distribution with parameter vector α, i.e.,

$$g_\alpha(p) = \prod_{i=1}^{M} p_i^{\alpha_i - 1}.$$

Then the density function of a Gamma-NID distributions with parameter vectors α and β is proportional to

$$g_\alpha(p)\left(\sum_{i=1}^{M} \beta_i p_i\right)^{-s}.$$

Theorem 8 (Homogeneous case). *Let f be the polynomial function in (49), n a vector of counts for M categories in N observations and*

$$\underline{E}[f|n] = \inf_{\alpha_i \in \Delta_M(s)} \frac{\displaystyle\int_{\Delta_M(1)} f(p)\, g_{\alpha+n}(p)\, dp}{\displaystyle\int_{\Delta_M(1)} g_{\alpha+n}(p)\, dp} \tag{50}$$

the lower posterior expectation of f computed with respect to the homogeneous Gamma-INID model (i.e., IDM). Let C_i be the i-th category and consider the refinement of the categories

$$C_i = \bigcup_{j \in \{1,\dots,l_i\}} C_{i,j}, \tag{51}$$

which divides each category C_i into l_i non-overlapping subcategories $C_{i,j}$ so to obtain $M' = l_1 + l_2 + \cdots + l_M$ categories with the associated vectors of counts $n' = (n_{1,1}, \dots, n_{M,l_M})$ and probabilities $P' = (P_{1,1}, \dots, P_{M,l_M})$ where $n_{i,j}$ and $P_{i,j}$ are the counts and the probability associated to the subcategory $C_{i,j}$ and where

$P_j = \sum_{i=1}^{l_j} P_{j,i}$ and $n_j = \sum_{i=1}^{l_j} n_{j,i}$. By modeling P' by a homogeneous Gamma-INID model with parameters vector $\alpha' = (\alpha_{1,1}, \ldots, \alpha_{M,l_M})$ one obtains that the lower posterior expectation

$$\underline{E}[f|n'] = \inf_{\alpha' \in \Delta_{M'}(s)} \frac{\displaystyle\int_{\Delta_M(1)} f(p)\, g_{\alpha'+n'}(p')\, dp'}{\displaystyle\int_{\Delta_M(1)} g_{\alpha'+n'}(p)\, dp'}, \tag{52}$$

with $p_j = \sum_{i=1}^{l_j} p_{j,i}$ and $p' = (p_{1,1}, \ldots, p_{M,l_M})$, coincides with that in (50) and vice-versa (coarsening). The same holds for the upper expectation.

Now let us consider the non-homogeneous case.

Theorem 9 (Non-Homogeneous case). *Let f be the polynomial function in (49), n a vector of counts for M categories in N observations and*

$$\underline{E}[f|n] = \inf_{\alpha \in \Delta_M(s), \beta \in \Delta_M(1)} \frac{\displaystyle\int_{\Delta_M(1)} f(p)\, g_{\alpha+n}(p) \left(\sum_{i=1}^{M} \beta_i p_i\right)^{-s} dp}{\displaystyle\int_{\Delta_M(1)} g_{\alpha+n}(p) \left(\sum_{i=1}^{M} \beta_i p_i\right)^{-s} dp} \tag{53}$$

the lower posterior expectation of f computed with respect to the non-homogeneous Gamma-INID model. Consider the refinement of categories in (51); the lower posterior expectation for P' given a homogeneous Gamma-INID prior with parameter vectors $\alpha' = (\alpha_{1,1}, \ldots, \alpha_{M,l_M})$ and $\beta' = (\beta_{1,1}, \ldots, \beta_{M,l_M})$

$$\underline{E}[f|n'] = \inf_{\alpha' \in \Delta_{M'}(s), \beta' \in \Delta_{M'}(1)} \frac{\displaystyle\int_{\Delta_M(1)} f(p)\, g_{\alpha'+n'}(p) \left(\sum_{i=1}^{M} \sum_{j=1}^{l_i} \beta'_{i,j} p_{i,j}\right)^{-s} dp}{\displaystyle\int_{\Delta_M(1)} g_{\alpha'+n'}(p) \left(\sum_{i=1}^{M'} \sum_{j=1}^{l_i} \beta'_{i,j} p_{i,j}\right)^{-s} dp} \tag{54}$$

coincides with that in (53) and vice-versa (coarsening). The same holds for the upper expectation.

Comparing the proofs of Theorems 8–9, it can be noticed that in the homogeneous case, due to the aggregation property of the Dirichlet distribution, RIP holds for any pair of parameter vectors α and α', provided that $\sum_j \alpha_{i,j} = \alpha_i$, and so it holds also for the parameters that give the upper and lower expectation of f. Conversely, in the non-homogeneous case, RIP holds only for the pair of parameter vectors (α, β) and (α', β') that gives the lower and upper posterior expectation of f. This means that RIP is not a property of the specific Gamma-NID distribution, but a property of the lower and upper envelopes. This is the main difference with respect to the IDM. Since Gamma-INID is derived by normalization of Gamma distributed variables, both in the homogeneous and non-homogeneous case, it seems that RIP is a property that is somehow related to the functional form of the Laplace exponent of the Gamma distribution.

8. Models comparison

Here, we discuss some interesting differences between the new prior near-ignorance models proposed in this paper and IDM. The aspects discussed in this section are particularly relevant in a situation where a priori we are completely ignorant about the process that generates the data, and thus also about the type and total number M of categories which have to be learned from the data. Therefore, as we do not know a priori the number of categories M, it is reasonable to take M very large.

8.1. Lower probability of the second observation to be equal to the first

Let us consider the example of a bag of marbles containing colored marbles of an unknown number of different colours [8]. Each color represents a different category. Assume to draw a red ball from the bag of marbles. A characteristic of IDM, which has been criticized, is that the lower probability for the second observation to be red (i.e., equal to the first) is equal to $1/(1+s)$ [8]. The values $s = 1$ or $s = 2$ lead to high values for this lower probability. However, it seems reasonable to assume that the lower probability of observing a second time a category already observed is significantly larger than 0 only if we have a strong prior belief that the number of categories is low. Instead, under complete prior ignorance, we may not want to bet on a category after we have seen it only once, but we would preferably wait until we see it for the second time before starting betting on it. If, for example, the observations come from a continuous distribution, the probability of observing a second time an outcome already observed would be 0 (see also [8, pages 43-44] and [21] for further discussion on this point).

Non-homogeneous Gamma-INID. From (46), the lower probability of the second observation to be equal to the first is given by:

$$\max(0, 1-s).$$

Thus, if $s \geq 1$, this lower probability is equal to 0. More generally, the lower probability of observing a specific category after N trials is equal to 0 until we have observes more than s realizations in that category. In this view, the parameter s can be interpreted as a threshold on the number of observations in a given category below which we would never bet on it, regardless of the reward. Thus, the non-homogeneous Gamma-INID model satisfies RIP but is also able to account for our prior ignorance about the number of categories. Note that a similar result is obtained by [22] using a bounded derivative model.

γ-stable INID. From (26), the lower probability for the second observation to be equal to the first is:

$$1-\gamma$$

so it goes to 0 as $\gamma \to 1$. This condition is analogous to the condition $s \to \infty$ for the IDM, since in both cases we obtain a vacuous model.

Inverse Gaussian INID. From (36), the lower probability for the second observation to be equal to the first is obtained by setting $m = 1$ and is found to be equal to 0.2982 for $s = 1$ and to 0.2227 for $s = 2$. This probability is smaller than for the IDM, but it goes to 0 only for $s \to \infty$, as for the IDM.

Indeed, among the models developed, only the non-homogenous one can give a probability for the second observation to equal to the first equal to zero without being vacuous.

8.2. Upper probability of seeing a new category

Consider a predictive model about the probability of observing a new category not yet observed in any of the N trials already observed. Let m be the number of different categories observed after N trials and $\hat{m} < M$ the *true* number of observable categories; we are interested in the probability that the event of observing any category $j > m$ occurs at the $(N+1)$-th trial. In particular, we focus on the case where all $\hat{m}$ categories are observed in the first

$N' < N$ draws (and thus $m = \hat{m}$) and the case where a new category is observed at each draw (and thus $m = N$). Clearly, to be consistent with the observations, when $m = N$ the prior model has to include $M \geq N$ categories. For the IG-INID model the lower and upper posterior expectations of P_{new} have been derived in Section 5. By similar reasoning one can show that also for the remaining three models, the lower expectation is found by letting $\alpha_j \to 0$ for all unobserved categories (i.e., such that $n_j = 0$) and is equal to 0. The upper posterior expectation of P_{new} can, instead, be found by letting $\alpha_j \to s$ for a single (arbitrarily chosen) unobserved category j and is $\overline{E}[P_{\text{new}}|n] = \overline{E}[P_j|n]$. The upper posterior expectations (assumed that Conjectures 1 and 2 are true) are equal to:

$$\text{IDM: } \frac{s}{N+s} \overset{N\to\infty}{=} 0, \quad \text{Gamma-INID: } \min\left(\frac{s}{N}, 1\right) \overset{N\to\infty}{=} 0,$$

$$\gamma\text{-stable INID: } \frac{m\gamma}{N} \overset{N\to\infty}{=} \begin{cases} 0, & \textit{if } m = \hat{m}, \\ \gamma, & \textit{if } m = N, \end{cases}$$

and

$$\text{IG-INID: } \frac{-s^2}{2N} \frac{W(N+1, s\lambda, m+1)}{W(N, s\lambda, m)} \overset{N\to\infty}{=} \begin{cases} 0, & \textit{if } m = \hat{m}, \\ \frac{1}{2}, & \textit{if } m = N. \end{cases}$$

In this context, another criticism to IDM is the following: after N observations, the upper probability of observing a new category goes to zero as $s/(s+N)$. This upper probability does not depend on how much variety there has been in the previous observations, i.e., the upper probability in case we have observed the same category in all the N previous observations or N different categories is the same. However, if N different categories have been observed in N trials we may not want to bet against seeing a new category at the next trial, regardless of the reward. In this case, we would like the upper probability of observing a new category to be equal to 1. This weak point is also discussed by Walley in [8, page 51].

Non-homogeneous Gamma-INID. It gives in practice the same upper probability as IDM for large N.

γ-stable INID. We can see that the upper probability of observing a new category depends on how much variety there has been in the previous observations. In fact, if we observe N different categories in all the previous trials this upper probability is equal to γ, while if we observe the same categories this upper probability can be made arbitrarily closed to 0 as

N increases. The result provided by the γ-stable INID model in the first case seems more appropriate than that provided by IDM.

Inverse Gaussian INID. The upper probability of observing a new category depends on how much variety there has been in the previous observations. As N increases, the upper probability can be made arbitrarily close to $1/2$ if $m = N$ and to 0 if, instead, $m = \hat{m}$.

Notice that, we have found an upper probability of observing a new category having observed N different categories in N previous trials equal to 1 only for vacuous models (e.g., the γ-stable INID with $\gamma \to 1$ or the Gamma-INID for $s \to \infty$). However, the peculiarity of the γ-stable INID is that the upper probability of observing a new category having observed N different categories in N previous trials can be made to be arbitrarily close to 1 by taking $\gamma = 1 - \epsilon$ and still the model can learn from data. Instead, for the Gamma-INID models the only case where this probability does not go to zero for $N \to \infty$ is when $s \to \infty$ and the model is vacuous.

This difference between IDM and Gamma-INID from one-side and γ-stable INID and Inverse Gaussian INID on the other side seems to be due to the RIP property. IDM and Gamma-INID satisfy RIP, while γ-stable INID and Inverse Gaussian INID do not. This means that in the complete absence of prior knowledge about the process that generates the data, the dependency of the lower and upper posterior expectations on the number of observed categories can lead to more intuitive inferences than those derived from models that satisfy RIP.

8.3. Discussion

In this section we recall the properties of the three new prior near-ignorance models developed and suggest possible scenarios where such properties could prove valuable. The analyses of these models carried out in this paper are of a theoretical nature; the necessary investigation of their performance on numerical case studies is left for future work.

One important difference between the models considered in this work is that the γ-Stable INID and the IG-INID models do not satisfy RIP whereas the homogeneous and non-homogeneous Gamma-INID models do. It has already been argued in [23, Sec. 4] that the RIP principle is not always a desirable property. In this paper, the authors stress that from the perspective of interval probability theory, the difference between lower and upper probabilities should depend on the amount of information available and the data representation. We think that this is especially true for predictive models in which the set of all possible categories is not known a priori and is updated as new categories are observed. Situations where the number of classes is unknown arise in numerous applications. One of the most important examples is the "species problem" in biology; other examples can be found in linguistics (e.g., the estimation of the frequency of words in an author's vocabulary) or software development (e.g., estimating the probabilities of different types of software failures). In such inferential problems we are interested in the species composition of an unknown population. In particular, given a sample from the population, it is interesting to know the species richness, which can be quantified by the probability of sampling new species in future trials. In principle, one would expect the population richness to be larger if we have already observed many different species rather than few. Therefore, models that do not respect RIP may be more suitable for modeling prior ignorance in this kind of problems.

On the other hand, the RIP property seems to be desirable for a prior ignorance model when the classes are clearly defined and known a priori (although this is not unanimously agreed, see for instance [21, 24]). In objective Bayesian analysis, a common practice is to impose invariance principles to derive non-informative priors. In this respect, the fact that IDM and Gamma-INID satisfy EP, SP and RIP, while the commonly used precise non-informative priors do not, is valuable. In [25], the authors show that IDM satisfies general invariance principles, among which exchangeability and representation insensitivity (which is similar to RIP). This result reinforces the importance of IDM as a model of prior ignorance. In [25], the authors conclude the paper listing several open questions about representation insensitivity for predictive systems. One of these questions was if there exist other models which satisfy RIP besides IDM. With the Gamma-INID model derived in this paper, we have shown that this is the case. The inferences provided by the non-homogeneous Gamma-INID model are more robust, in the sense that, for a given strength of the prior, they always encompass those of the IDM. Moreover, its lower and upper expectations cannot be interpreted as the result of adding s additional counts to the available observations, but rather as a swing scenario in which s counts among the N observed are moved from a category to another. Situations where such swing scenarios are of particular interest are, for example, election polls (or other types of opinion surveys) where it is important to evaluate the effects of swing votes, i.e. votes that are seen as potentially going to any of a number of candidates. In those cases, the Gamma-INID model could be more appropriate than the IDM.

9. Conclusions

In this paper, we have derived new near-ignorance models based on the class of Normalized Infinitely Divisible (NID) distributions. For this class, by exploiting the Lévy-Khintchine representation theorem, it is possible to derive general formulae for prior and posterior predictive inferences. By using these results we have been able to generally characterize the near-ignorance properties of the NID class. For three members of this class, we have shown that, although these three near-ignorance NID models are not conjugate with the likelihood, the posterior inferences drawn from them can still be computationally tractable. In particular, for the Gamma INID model the lower and upper expectations of the P_i can be computed by means of simple algebraic expressions. Assuming the validity of a conjecture, this is also true for the γ-stable (with $\gamma \leq 1/2$) model, while for the inverse Gaussian model, the lower and upper expectations can be computed efficiently by numerically solving one-dimensional integrals. On the other side, when the lower and upper expectations are not found for $\alpha \to 0$ and, respectively, $\alpha \to 1$, as for the γ-stable INID model with $\gamma > 1/2$, numerical optimization is also necessary. We have shown that one of the models developped, the Gamma-INID model, always provides, for a given s, inferences that are more conservative than those of the Imprecise Dirichlet Model and, moreover, satisfies the representation invariance principle (RIP), i.e. the posterior inferences of an event are invariant w.r.t. refinements and coarsening of the possibility space. On the other hand, the other two models, which do not satisfy RIP, have a posterior imprecision which increases linearly or almost linearly with the number of observed categories. As future work, we plan to investigate the properties of other members of the near-ignorance NID class of priors.

Moreover, the theoretical work carried out in this paper needs to be supported by application studies that we have left for future work. In particular, we plan to apply our models to solve classification and prediction problems and compare the results with those obtained by precise models and by the Imprecise Dirichlet Model.

Acknowledgements

This work has been partially supported by the Swiss NSF grant n. 200020-137680/1. The authors would like to thank Marco Zaffalon, Gert de Cooman, Jasper De Bock, Antonio Lijoi, Cassio P. de Campos and Enrique Miranda for their useful comments.

Appendix A. Proofs

Appendix A.1. Proof of Proposition 1

The posterior expectation $E[P_i^r|n]$ is obtained as the ratio

$$\frac{E[P_i^r P_1^{n_1}\cdots P_m^{n_m}]}{E[P_1^{n_1}\cdots P_m^{n_m}]},$$

which, using (12), is given by (13).

Appendix A.2. Proof of Theorem 1

For $\alpha_i \to 0$ the distribution of X_i tends to the Dirac delta function $\delta(x)$, because its Laplace transform tends to the unit constant, and so does the distribution of P_i. Then, one has that $\underline{E}[P_i^r] \to 0$.

For $\alpha_i \to s$, $\alpha_j \to 0$ for all $j \neq i$, we have that

$$\overline{E}[P_i^r] = \overline{E}\Big[\Big(1 - \sum_{i\neq j} P_j\Big)^r\Big] \to 1,$$

because the distribution of X_j for all $j \neq i$ are Dirac's delta functions in zero.

Appendix A.3. Proof of Theorem 2

The lower and upper expectations of P_i^r follow by the prior near-ignorance properties of NID distributions (Theorem 1). We can thus focus on (23). $E[P_i P_j]$ can be obtained from (7) with $r_1, r_2 = 1$ and $r_k = 0$ for $k > 2$, i.e.:

$$\begin{aligned} E[P_i P_j] &= \alpha_i \alpha_j \int_0^\infty u \left[\frac{d\psi(u)}{du}\right]^2 e^{-s\psi(u)} du \\ &= \alpha_i \alpha_j \frac{\Gamma(1-\gamma)^2}{2\pi} \int_0^\infty u^{2\gamma-1} e^{-s\frac{\Gamma(1-\gamma)}{\sqrt{2\pi\gamma}} u^\gamma} du \qquad \text{(A.1)} \\ &= \alpha_i \alpha_j \frac{\Gamma(1-\gamma)^2}{2\pi} \frac{2\pi\gamma}{s^2 \Gamma(1-\gamma)^2} = \alpha_1 \alpha_2 \frac{\gamma}{s^2}. \end{aligned}$$

Finally, we get

$$0 < E[P_i P_j] < \frac{1}{4}\gamma.$$

Since 0 is obtained for $\alpha_i \to 0$ and $\frac{\gamma}{4}$ for $\alpha_i = \alpha_j \to \frac{s}{2}$, these are the upper and lower expectations of $E[P_i P_j]$.

Appendix A.4. Proof of Theorem 3

We prove the result for P_1. Using (19), the partial Bell polynomial $B_{n_i,k}$ for the γ-stable model can be written as

$$B_{n_i,k}(-\alpha\psi^{(1)}(u),\dots,-\alpha\psi^{(n_i-k+1)}(u)) = (-1)^{n_i}\sum \frac{n_i!}{j_1!j_2!\cdots j_{n_i-k+1}!}\times$$
$$\times\left(\frac{\Gamma(1-\gamma)_1}{1!}\right)^{j_1}\left(\frac{\Gamma(2-\gamma)_2}{2!}\right)^{j_2}\cdots\left(\frac{\Gamma(n_i-k+1-\gamma)}{(r-k+1)!}\right)^{j_{r-k+1}}\left(\frac{\alpha}{\sqrt{2\pi}}\right)^k u^{k\gamma-n_i}.$$

Then the complete Bell polynomial is

$$B_{n_i}(-\alpha\psi^{(1)}(u),\dots,-\alpha\psi^{(n_i)}(u)) = (-1)^{n_i}\sum_{k=1}^{n_i} C_{n_i,k}\left(\frac{\alpha}{\sqrt{2\pi}}\right)^k u^{k\gamma-n_i}$$

where

$$C_{n_i,k} = B_{n_i,k}(\Gamma(1-\gamma),\dots,\Gamma(n_i-k+1-\gamma)).$$

From (8) it follows that

$$\begin{aligned}
&B_{n_i+1}(\cdot) = e^{\alpha\psi}\frac{d}{du}\left[B_{n_i}(\cdot)e^{-\alpha\psi}\right]\\
&= -\frac{\alpha\Gamma(1-\gamma)}{\sqrt{2\pi}}u^{\gamma-1}B_{n_i}(\cdot) + (-1)^{n_i}\sum_{k=1}^{n_i} C_{n_i,k}\left(\frac{\alpha}{\sqrt{2\pi}}\right)^k(\gamma k - n_i)u^{k\gamma-n_i-1}\\
&= -(n_i-\gamma)\frac{B_{n_i}(\cdot)}{u} + (-1)^{n_i}\sum_{k=2}^{n_i} C_{n_i,k}\left(\frac{\alpha}{\sqrt{2\pi}}\right)^k\gamma(k-1)u^{k\gamma-n_i-1}\\
&-(-1)^{n_i}\Gamma(1-\gamma)\left[\sum_{k=2}^{n_i} C_{n_i,k-1}\left(\frac{\alpha}{\sqrt{2\pi}}\right)^k u^{k\gamma-n_i-1} + \Gamma(1-\gamma)^{n_i}\left(\frac{\alpha}{\sqrt{2\pi}}\right)^{n_i+1} u^{(n_i+1)\gamma-n_i-1}\right]\\
&= -(n_i-\gamma)\frac{B_{n_i}(\cdot)}{u} - (-1)^{n_i}\frac{\alpha^2}{u}[T_1(n_i) + T_2(n_i)]
\end{aligned}$$

where

$$T_1(n_i) = \alpha^{n_i-1}\left(\frac{\Gamma(1-\gamma)}{\sqrt{2\pi}}\right)^{n_i+1} u^{(n_i+1)\gamma-n_i}$$

$$T_2(n_i) = \sum_{k=2}^{n_i} \alpha^{k-2}[\Gamma(1-\gamma)C_{n_i,k-1} - \gamma(k-1)C_{n_i,k}](\sqrt{2\pi})^{-k} u^{k\gamma-n_i}.$$

Then, from (13), if $n_1 > 0$ the posterior expectation of P_1 is

$$E[P_1|n] = \frac{n_1-\gamma}{N} + \alpha_1 \frac{\int u^{N-1} e^{-s\psi(u)}(T_1(n_1)+T_2(n_1)) \prod_{j=2}^{m} B_{n_j}(\cdot)}{\int u^{N-1} e^{-s\psi(u)} \sum_{k=1}^{n_1} C_{n_1,k}\alpha_1^{k-1}(\sqrt{2\pi})^{-k} u^{k\gamma-n_1} \prod_{j=2}^{m} B_{n_j}(\cdot)}. \tag{A.2}$$

Since $T_1 \geq 0$ and $T_2 \geq 0$ due to the assumption (24), the second term at the r.h.s. of (A.2) is always positive and thus the infimum of $E[P_1|n]$ is found for $\alpha_1 \to 0$ and is $(n_1-\gamma)/N$ whereas the supremum is found for $\alpha_1 \to s$ (and thus $\alpha_j \to 0$ for all $j \neq i$) and is

$$\overline{E}[P_1|n] = 1 - \sum_{j=2}^{M} \lim_{\alpha_j \to 0} E[P_j|n] = 1 - \frac{N-n_1-(m-1)\gamma}{N} = \frac{n_1+(m-1)\gamma}{N}$$

for $n_1 > 0$. If, instead, $n_1 = 0$, one has

$$B_1(-\alpha_1\psi(u)) = -\alpha_1\psi^{(1)}(u)$$

and

$$E[P_1|n] = \frac{\alpha_1}{N} \frac{\displaystyle\int_0^\infty u^N e^{-s\psi(u)}\psi^{(1)}(u) \prod_{j=2}^{m} B_{n_j}(\cdot)\,du}{\displaystyle\int_0^\infty u^{N-1} e^{-s\psi(u)} \prod_{j=2}^{m} B_{n_j}(\cdot)\,du} \to 0,$$

and thus the infimum is found again for $\alpha_1 \to 0$ and is $\underline{E}[P_1|n] = 0$. The upper is found, as before, for $\alpha_1 \to s$:

$$\overline{E}[P_1|n] = 1 - \sum_{j=2}^{M} \lim_{\alpha_j \to 0} E[P_j|n] = 1 - \frac{N-m\gamma}{N} = \frac{m}{N}\gamma$$

It remains to prove that the conditions in (24) and (25) are equivalent. This follows from the fact that the partial Bell polynomial $B_{n_i,k}(x_1,\ldots,x_{n_i})$ can be written as follows [18]:

$$\begin{aligned}
&B_{n_i,k+1}(x_1,\ldots,x_{n_i})\\
&=\frac{1}{(k+1)!}\sum_{a_1=k}^{n_i-1}\sum_{a_2=k-1}^{a_1-1}\cdots\sum_{a_k=1}^{a_{k-1}-1}\binom{n_i}{a_1}\binom{a_1}{a_2}\ldots\binom{a_{k-1}}{a_k}x_{n_i-a_1}x_{a_1-a_2}\ldots x_{a_k}.
\end{aligned}$$

In this case we have

$$x_a=\Gamma(a-\gamma)=\Gamma(1-\gamma)\prod_{i=1}^{a-1}(i-\gamma)=\gamma^{a-1}\Gamma(1-\gamma)\prod_{i=1}^{a-1}\left(\frac{i}{\gamma}-1\right),$$

where we have used the recurrence relation $\Gamma(a-\gamma)=(a-1-\gamma)\Gamma(a-1-\gamma)$ for $a-1$ times.

Then we obtain

$$\begin{aligned}
C_{n_i,k+1}=&\frac{\gamma^{n_i-k}\Gamma(1-\gamma)^k}{(k+1)!}\sum_{a_1=k}^{n_i-1}\sum_{a_2=k-1}^{a_1-1}\cdots\sum_{a_k=1}^{a_{k-1}-1}\binom{n_i}{a_1}\binom{a_1}{a_2}\ldots\binom{a_{k-1}}{a_k}\times\\
&\times\prod_{i=1}^{n_i-a_1-1}\left(\frac{i}{\gamma}-1\right)\prod_{i=1}^{a_1-a_2-1}\left(\frac{i}{\gamma}-1\right)\cdots\prod_{i=1}^{a_k-1}\left(\frac{i}{\gamma}-1\right)\\
&=\frac{\gamma^{n_i-k}\Gamma(1-\gamma)^k}{(k+1)!}S_{n_i,k+1},
\end{aligned}$$

having defined

$$S_{n_i,k+1}=\sum_{a_k=1}^{a_{k-1}-1}\binom{n_i}{a_1}\ldots\binom{a_{k-1}}{a_k}\prod_{i=1}^{n_i-a_1-1}\left(\frac{i}{\gamma}-1\right)\cdots\prod_{i=1}^{a_k-1}\left(\frac{i}{\gamma}-1\right).$$

Then, the condition in (24) becomes

$$\begin{aligned}
&\frac{\gamma^{n_i-k+2}\Gamma(1-\gamma)^{k-1}}{(k-1)!}S_{n_i,k-1}-(k-1)\frac{\gamma^{n_i-k+2}\Gamma(1-\gamma)^{k-1}}{k!}S_{n_i,k}>0\\
&\Longrightarrow kS_{n_i,k-1}-(k-1)S_{n_i,k}>0.
\end{aligned}$$

Appendix A.5. Proof of Theorem 4

The lower and upper expectations of P_i^r follows by the prior near-ignorance properties of NID distributions (Theorem 1). We can thus focus on (34). From (7) we derive $E[P_i P_j]$:

$$\begin{aligned} E[P_i P_j] &= \alpha_i \alpha_j \int_0^\infty u \left[\frac{d\psi(u)}{du}\right]^2 e^{-s\psi(u)} du \\ &= \alpha_i \alpha_j e^{s\lambda} \int_0^\infty \frac{u}{2u+\lambda^2} e^{-s\sqrt{2u+\lambda^2}} du. \end{aligned} \tag{A.3}$$

By the change of variables $y = s\sqrt{2u+\lambda^2}$, one obtains

$$\begin{aligned} E[P_i P_j] &= \frac{\alpha_i \alpha_j}{s^2} \frac{e^{s\lambda}}{2} \int_{s\lambda}^\infty y e^{-y} dy - \int_{s\lambda}^\infty (s\lambda)^2 \frac{e^{-y}}{y} dy \\ &= \frac{\alpha_i \alpha_j}{s^2} \frac{e^{s\lambda}}{2} \left[s\lambda e^{-s\lambda} + e^{-s\lambda} - (s\lambda)^2 \Gamma(0, s\lambda)\right]. \end{aligned} \tag{A.4}$$

By using the recurrence relations for the incomplete gamma function, one gets:

$$E[P_i P_j] = \frac{\alpha_i \alpha_j}{s^2} \left[1 - e^{s\lambda}(s\lambda)^2 \Gamma(-2, s\lambda)\right].$$

Finally, we obtain

$$0 \leq E[P_i P_j] \leq \frac{1}{4} \left[1 - e^{s\lambda}(s\lambda)^2 \Gamma(-2, s\lambda)\right].$$

Since the lower bound is reached for $\alpha_i \to 0$ or $\alpha_j \to 0$ and the upper bound for $\alpha_i = \alpha_j \to \frac{s}{2}$, these are the upper and lower expectations of $E[P_i P_j]$.

Appendix A.6. Proof of Theorem 5

By applying the same reasoning used in Theorem 3 (that, therefore, we omit here) one obtains that if $n_1 = 0$ the lower expectation of P_1 is 0 (for $\alpha_1 \to 0$) and if $n_1 > 0$

$$\begin{aligned} E[P_1|n] = {} & \frac{n_1 - \frac{1}{2}}{N} \frac{\int u^N (2u+\lambda^2)^{-1} B^1_{n_1} \cdots B^m_{n_m} du}{\int u^{N-1} B^1_{n_1} \cdots B^m_{n_m} du} \\ & + \alpha_1 \frac{\int u^N (2u+\lambda^2)^{-1} e^{-s\psi(u)} (T_1(n_1) + T_2(n_1)) \prod_{j=2}^m B^i_{n_j}}{\int u^{N-1} e^{-s\psi(u)} \sum_{k=1}^{n_1} C_{n_1,k} \alpha_1^{k-1} (2\sqrt{\pi})^{-k} (2u+\lambda^2)^{k/2-n_1} \prod_{j=2}^m B^j_{n_j}}, \end{aligned} \tag{A.5}$$

where

$$T_1(n_i) = \alpha^{n_i-1} 2^{-n_i-1} (2u+\lambda^2)^{(n_i+1)/2-n_i}$$
$$T_2(n_i) = \sum_{k=2}^{n_i} \alpha^{k-2} [\sqrt{\pi} C_{n_i,k-1} - \frac{k-1}{2} C_{n_i,k}] (2\sqrt{\pi})^{-k} (2u+\lambda^2)^{k/2-n_i},$$

and

$$B^i_{n_i} = (-2)^{n_i} \sum_{k=1}^{n_i} C_{n_i,k} \left(\frac{\alpha_i}{2\sqrt{\pi}} \right)^k (2u+\lambda^2)^{k/2-n_i}.$$

If $n_1 = 1$, the first term at the r.h.s. of (A.5) does not depend on α_1 and the second term is always positive; then, the infimum is found for $\alpha_1 \to 0$. If $n_1 > 1$ the first term in (A.5) becomes

$$\frac{n_1 - \frac{1}{2}}{N} \frac{\int \frac{u^N}{2u+\lambda^2} \left(\sum_{k=1}^{n_1} C_{n_1,k} \left(\frac{\alpha_1}{2\sqrt{\pi}} \right)^k (2u+\lambda^2)^{k/2-n_1} \right) B^2_{n_2} \cdots B^m_{n_m} \, du}{\int u^{N-1} \left(\sum_{k=1}^{n_1} C_{n_1,k} \left(\frac{\alpha_1}{2\sqrt{\pi}} \right)^k (2u+\lambda^2)^{k/2-n_1} \right) B^2_{n_2} \cdots B^m_{n_m} \, du} \tag{A.6}$$

Each term at the numerator is proportional to

$$\int u^N (2u+\lambda^2)^{k/2-n_1-1} B^2_{n_2} \cdots B^m_{n_m} \, du,$$

and each term at the denominator is proportional to

$$\int u^{N-1} (2u+\lambda^2)^{k/2-n_1} B^2_{n_2} \cdots B^m_{n_m} \, du;$$

The proportionality constant for a given k is the same for the numerator and denominator. Since

$$\frac{w_1 A_1 + w_2 A_2}{w_1 B_1 + w_2 B_2} \geq \frac{A_1}{B_1}$$

holds if

$$\frac{A_1}{B_1} \leq \frac{A_2}{B_2},$$

by proving that

$$\frac{\displaystyle\int u^{N}(2u+\lambda^{2})^{-n_1-1/2}B_{n_2}^{2}\cdots B_{n_m}^{m}\,du}{\displaystyle\int u^{N-1}(2u+\lambda^{2})^{-n_1+1/2}B_{n_2}^{2}\cdots B_{n_m}^{m}\,du}\le\frac{\displaystyle\int u^{N}(2u+\lambda^{2})^{-n_1-1+k/2}B_{n_2}^{2}\cdots B_{n_m}^{m}\,du}{\displaystyle\int u^{N-1}(2u+\lambda^{2})^{-n_1+k/2}B_{n_2}^{2}\cdots B_{n_m}^{m}\,du}\tag{A.7}$$

we can show that the infimum of (A.6) is found when only the term with $k=1$ survives. This is obtained for $\alpha_1\to 0$. Let us define

$$h(u)=u^{N-1}(2u+\lambda^{2})^{-n_1+k/2}B_{n_2}^{2}\cdots B_{n_m}^{m}.$$

We can then re-write the condition in (A.7)

$$\frac{\displaystyle\int\frac{u}{(2u+\lambda^{2})}\frac{h(u)}{(2u+\lambda^{2})^{k/2}}du}{\displaystyle\int\frac{h(u)}{(2u+\lambda^{2})^{k/2}}du}\le\frac{\displaystyle\int\frac{u}{(2u+\lambda^{2})}h(u)du}{\displaystyle\int h(u)du}\tag{A.8}$$

Let us define

$$K_1=\int\frac{h(u)}{(2u+\lambda^{2})^{k/2}}du\quad\text{and}\quad K_2=\int h(u)du.$$

The two sides of the inequality can be seen as the expected values of the positive variable $u/(2u+\lambda^2)$ when its distribution is defined by the CDF

$$P_1\left(\frac{u}{(2u+\lambda^{2})}\le x\right)=\frac{\displaystyle\int_0^{u_x}\frac{h(u)}{(2u+\lambda^{2})^{k/2}}du}{K_1}$$

for the l.h.s. of the inequality (with $u_x=x(x+\sqrt{x^2+\lambda^2})$) and

$$P_k\left(\frac{u}{(2u+\lambda^{2})}\le x\right)=\frac{\displaystyle\int_0^{u_x}h(u)du}{K_2}$$

for the r.h.s. Since the two curves

$$\frac{h(u)}{(2u+\lambda^{2})^{k/2}K_1}\quad\text{and}\quad\frac{h(u)}{K_2}$$

do not coincide and their integral over the support is 1 (so that one cannot be always larger than the other), there must be at least one crossing point $u^* > 0$ for which it holds

$$\frac{h(u^*)}{(2u^* + \lambda^2)^{k/2} K_1} = \frac{h(u^*)}{K_2}$$

Then for all $u > u^*$ it holds

$$\frac{h(u)}{K_2} > \frac{h(u)}{K_1(2u+\lambda^2)^{k/2}}, \tag{A.9}$$

since

$$\frac{h(u^*)}{K_2} \geq \frac{h(u^*)}{K_1(2u^*+\lambda^2)^{k/2}} \implies \frac{1}{K_2} \geq \frac{1}{K_1(2u^*+\lambda^2)^{k/2}} > \frac{1}{K_1(2u+\lambda^2)^{k/2}}$$

(where we have used the fact that $h(u^*) > 0$) from which (A.9) follows.

By similar reasoning we can show that for all $u < u^*$

$$\frac{h(u)}{K_2} < \frac{h(u)}{K_1(2u+\lambda^2)^{k/2}}. \tag{A.10}$$

As a consequence, we have that

$$P_1\left(\frac{u}{(2u+\lambda^2)} \geq x\right) < P_k\left(\frac{u}{(2u+\lambda^2)} \geq x\right) \tag{A.11}$$

since if $u_x > u^*$

$$\frac{\displaystyle\int_{u_x}^{\infty} \frac{h(u)}{(2u+\lambda^2)^{k/2}} du}{K_1} < \frac{\displaystyle\int_{u_x}^{\infty} h(u) du}{K_2}$$

and if $u_x < u^*$

$$1 - \frac{\displaystyle\int_{0}^{u_x} \frac{h(u)}{(2u+\lambda^2)^{k/2}} du}{K_1} < 1 - \frac{\displaystyle\int_{0}^{u_x} h(u) du}{K_2}.$$

Since the expectation of a positive random variable X is given by the integral over its support of $P(X > x)$, by (A.11) it follows that (A.7) holds for all k and thus the infimum of the expectation of P_1 is found for $\alpha_1 \to 0$.

By similar reasoning one can also show that the infimum is obtained by setting $\alpha_j \to 0$ for all categories j for which $n_j > 1$ or, if there are no categories $j \neq 1$ with $n_j \leq 1$, by setting $\alpha_j \to s$ for the category with the minimum number of counts.

For simplicity, let M be the category with the minimum number of count. To compute the posterior expectation of an Inverse Gaussian NID distribution when $\alpha_M \to s$ we use the following result:

$$\int_0^\infty \frac{u^{N-1} e^{-s\sqrt{2u+\lambda^2}}}{(2u+\lambda^2)^{N-\frac{M}{2}}} du = 2^{-N+1} \int_\lambda^\infty \frac{(y^2-\lambda^2)^{N-1} e^{-sy}}{y^{2N-M-1}} dy =$$

$$= 2^{-N+1} \sum_{k=0}^{N-1} \binom{N-1}{k} (-\lambda^2)^{N-1-k} s^{2N-2-M-2k} \int_{s\lambda}^\infty \frac{e^{-t}}{t^{2N-M-2k-2}} dt =$$

$$= 2^{-N+1} \sum_{k=0}^{N-1} \binom{N-1}{k} (-\lambda^2)^{N-1-k} s^{2N-2-M-2k} \Gamma(2+M+2k-2N; s\lambda).$$

From (7) and (17) the mixed non-central moments are

$$E[P_1^{n_1} \cdots P_M^{n_M}] = (-1)^N \int_0^\infty u^{N-1} e^{-s(\sqrt{2u+\lambda^2}-\lambda)} \prod_{i=1}^{M-1} (-\alpha_i) \psi^{(n_i)} \sum_{r=1}^{n_M} C_{n_M,r} \left(\frac{-\alpha_M}{2\sqrt{\pi}}\right)^r \psi^{(r)} du$$

$$= \frac{2^{N-M+1} e^{s\lambda}}{\sqrt{\pi}^{M-1}} \prod_{i=1}^{M-1} \alpha_i \Gamma(n_i - \frac{1}{2}) \sum_{r=1}^{n_M} C_{n_M,r} \left(\frac{\alpha_M}{2\sqrt{\pi}}\right)^r I_N(r)$$

with

$$I_N(k) = \int_0^\infty \frac{u^{N-1} e^{-s\sqrt{2u+\lambda^2}}}{(2u+\lambda^2)^{N-\frac{M+r-1}{2}}} du$$

$$= 2^{-N+1} \sum_{k=0}^{N-1} \binom{N-1}{k} (-\lambda^2)^{N-1-k} s^{2N-1-M-r-2k} \Gamma(1+M+r+2k-2N; s\lambda).$$

For $\alpha_M \to s$ we obtain

$$E[P_1|n] = \frac{n_1 - \frac{1}{2}}{N} \frac{-(s\lambda)^2 \sum_{r=1}^{n_M} \frac{C_{n_M,r}}{(2\sqrt{\pi})^r} \sum_{k=0}^{N} \binom{N}{k} \frac{\Gamma(-1+M+r+2k-2N; s\lambda)}{[-(s\lambda)^2]^{-k}}}{\sum_{r=1}^{n_M} \frac{C_{n_M,r}}{(2\sqrt{\pi})^r} \sum_{k=0}^{N-1} \binom{N-1}{k} \frac{\Gamma(1+M+r+2k-2N; s\lambda)}{[-(s\lambda)^2]^{-k}}}$$

$$= \frac{n_1 - \frac{1}{2}}{N} \frac{-(s\lambda)^2 \sum_{r=1}^{n_M} \frac{C_{n_M,r}}{(2\sqrt{\pi})^r} W(N+1, s\lambda, M+r-1)}{\sum_{r=1}^{n_M} \frac{C_{n_M,r}}{(2\sqrt{\pi})^r} W(N, s\lambda, M+r-1)}.$$

Similarly it can be shown that if $n_M = 0$ and there are m observed categories the lower (found again for $\alpha_M \to s$) is

$$E[P_i|n] = \frac{n_1 - \frac{1}{2}}{N} \frac{-(s\lambda)^2 W(N+1, s\lambda, m)}{W(N, s\lambda, m)}.$$

The posterior expectation for $\alpha_1 \to s$ can be obtained by straightforward computations similar to those used for $\alpha_M \to s$.

Appendix A.7. Proof of Proposition 2

The function that gives the upper expectation of P_{new} can be written as

$$\frac{-(s\lambda)^2}{2N} \frac{W(N+1, s\lambda, m+1)}{W(N, s\lambda, m)} = \frac{-s^2}{2N} \frac{\sum_{k=0}^{N} \binom{N}{k} (-s^2)^{-k} \Gamma(1+m+2k-2N; s)}{\sum_{k=0}^{N-1} \binom{N-1}{k} (-s^2)^{-k} \Gamma(2+m+2k-2N; s)}.$$

The incomplete gamma function $\Gamma(2k-N; s)$ with argument $2k-N<0$ is bounded if $s>1$ and goes to infinity as s^{2k-N} if $s<1$. Moreover, for positive integer arguments, the incomplete gamma function can be written as $\Gamma(N; s) = (N-1)! e^{-s} \sum_{j=0}^{N-1} \frac{s^j}{j!}$. Then, $\lim_{N\to\infty} \frac{1}{k!(N-k)!} \Gamma(2k-N; s) = 0$ for all $k \le N$. As a consequence, for $m = N$ and $N \to \infty$ one has

$$\lim_{N\to\infty} \frac{-s^2}{2N} \frac{N! \sum_{k=0}^{N} \frac{(-s^2)^{-k}}{k!(N-k)!} \Gamma(1+2k-N; s)}{(N-1)! \sum_{k=0}^{N-1} \frac{(-s^2)^{-k}}{k!(N-k-1)!} \Gamma(2+2k-N; s)}$$

$$= \lim_{N\to\infty} \frac{-s^2}{2N} \frac{N(-s^2)^{-N} \frac{\Gamma(N+1; s)}{N!}}{(-s^2)^{-N+1} \frac{\Gamma(N; s)}{(N-1)!}} = \frac{1}{2}.$$

When the number of observed categories is constantly equal to m, the upper posterior probability of observing a new category P_{new} can equivalently be derived from the posterior PDF of the vector of probabilities P obtained by letting $\alpha_i \to 0$ for all $i = 1,\ldots,m$ and $\alpha_j \to s$ for an arbitrary unobserved category (for simplicity we refer to it as category M). The upper expectation of P_{new} does not depend on the number of categories assumed by the model and therefore we take $M = m+1$ (that is, we assume that there is only 1 unobserved category). The posterior PDF of P is then:

$$g(p) = \frac{\prod_{i=1}^{m} p_i^{-3/2+n_i} p_M^{-3/2+m/4} \mathrm{K}_{-m/2}\left[\left(\frac{s}{p_M}\right)^{1/2}\right]}{\int \prod_{i=1}^{m} p_i^{-3/2+n_i} p_M^{-3/2+m/4} \mathrm{K}_{-m/2}\left[\left(\frac{s}{p_M}\right)^{1/2}\right]}. \tag{A.12}$$

The integral at the denominator is bounded since $n_i \geq 1$ for all $i = 1,\ldots,m$, and thus the Bernstein-von Mises theorem [26, Theorem 20.1] holds and the expected value of any P_i for $N \to \infty$ is asymptotically equivalent to the frequency n_i/N. Then, since $n_M = 0$, we have that $E[P_{\text{new}}|n] = E[P_M|n] \to 0$.

Appendix A.8. Proof of Theorem 6

The lower and upper expectations of P_i^r follows by the general prior near-ignorance properties of NID (Theorem 1). The value of $\underline{E}[P_1P_2]$ and the lower bound for the value of $\overline{E}[P_1P_2]$ in (44) follows from the fact that the Gamma-INID set of priors contains the set of priors of the IDM as a subset. Thus, the upper and lower bounds of the Gamma-INID model necessarily encompass those of the IDM.

Appendix A.9. Proof of Theorem 7

In the proof we exploit the following result:

$$\int_0^{\infty} \frac{u^{\alpha_1-1}}{(u+1)^{\alpha_1+\alpha_2}} du = \int_0^1 v^{\alpha_1-1}(1-v)^{\alpha_2-1} dv = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}, \tag{A.13}$$

for any $\alpha_1, \alpha_2 > 0$ (it can be derived from the transformation $v = u/(1+u)$ and the properties of the Beta function). The homogeneous case can be derived from (42) by setting $\beta_i = \beta$ for all $i = 1, \dots, M$:

$$
\begin{aligned}
E[P_i|n] &= \frac{E[P_i P_1^{n_1} \cdots P_M^{n_M}]}{E[P_1^{n_1} \cdots P_M^{n_M}]} \\
&= \frac{\alpha_i + n_i}{N} \int_0^\infty u^N (1+u)^{-(s+N+1)} du \cdot \left(\int_0^\infty u^{N-1} (1+u)^{-(s+N)} du \right)^{-1} \\
&= \frac{\alpha_i + n_i}{N} \frac{\Gamma(N+1)\Gamma(s)}{\Gamma(s+N+1)} \frac{\Gamma(s+N)}{\Gamma(N)\Gamma(s)} = \frac{\alpha_i + n_i}{s+N}.
\end{aligned} \tag{A.14}
$$

Thus, for $\alpha_i \to 0$ and $\alpha_i \to s$ we obtain the lower and, respectively, upper expectations. For the non-homogeneous case it can be noticed that

$$
\begin{aligned}
&\frac{\alpha_i + n_i}{N} \int_0^\infty u^N \left(\beta_i + u\right)^{-\alpha_i - n_i - 1} \prod_{i \neq j=2}^{M} \left(\beta_j + u\right)^{-\alpha_j - n_j} du \\
&\cdot \left(\int_0^\infty u^{N-1} \prod_{j=1}^{M} \left(\beta_j + u\right)^{-\alpha_j - n_j} du \right)^{-1} \\
&\leq \frac{\alpha_i + n_i}{N} \int_0^\infty u^N u^{-1} \left(\beta_i + u\right)^{-\alpha_i - n_i} \prod_{i \neq j=2}^{M} \left(\beta_j + u\right)^{-\alpha_j - n_j} du \\
&\cdot \left(\int_0^\infty u^{N-1} \prod_{j=1}^{M} \left(\beta_j + u\right)^{-\alpha_j - n_j} du \right)^{-1} \leq \frac{\alpha_i + n_i}{N}.
\end{aligned} \tag{A.15}
$$

If $\alpha_i + n_i < N$, the upper bound $(\alpha_i + n_i)/N$ can be obtained by setting $\beta_i \to 0$ otherwise, from the coherence condition $\overline{E}[P_i|n] \leq \max(P_i) = 1$, the upper bound is 1. Hence, for $\alpha_i \to s$, we have $\overline{E}[P_i|n] = \min(1, (n_i + s)/N)$. The lower posterior expectation can be obtained by

$$
\sum_{j \neq i} E[P_j|n] \leq \sum_{j \neq i} \frac{n_j + \alpha_j}{N} \leq \frac{N - n_i + s}{N}.
$$

Thus, we have that

$$
E[P_i|n] = 1 - \sum_{j \neq i} E[P_j|n] \geq 1 - \frac{N - n_i + s}{N} = \frac{n_i - s}{N},
$$

provided that $n_i - s \geq 0$. The lower bound is tight, i.e., it can be obtained for $\beta_i \to 1, \alpha_i \to 0$:

$$
\begin{aligned}
&\frac{\alpha_i+n_i}{N}\frac{\displaystyle\int_0^\infty u^N\left(\beta_i+u\right)^{-\alpha_i-n_i-1}\prod_{i\neq j=2}^{M}\left(\beta_j+u\right)^{-\alpha_j-n_j}du}{\displaystyle\int_0^\infty u^{N-1}\prod_{j=1}^{M}\left(\beta_j+u\right)^{-\alpha_j-n_j}du}\\
&\overset{\beta_i\to1,\alpha_i\to0}{\longrightarrow}\frac{n_i}{N}\int_0^\infty\frac{u}{u+1}\frac{u^{N-1}(1+u)^{-n_i}(u)^{-s-N+n_i}du}{\displaystyle\int_0^\infty u^{N-1}(1+u)^{-n_i}(u)^{-s-N+n_i}du}\\
&=\frac{n_i}{N}\frac{n_i-s}{n_i}=\frac{n_i-s}{N},
\end{aligned}
\tag{A.16}
$$

where the first equality in the last line is well-defined provided that $n_i - s > 0$, otherwise the lower prevision is zero. Thus, we have $\underline{E}[P_i|n] = \max(0, (n_i - s)/N)$.

Appendix A.10. Proof of Theorem 8

Consider the integrals in (52) and for any $i = 1, \ldots, M$ the change of variables: $p_{i,j} = u_i v_{i,j}$ for $j = 1, \ldots, l_i$, with $v_{i,j} > 0$, $\sum_{j=1}^{l_i} v_{i,j} = 1$ and $u_i > 0$, $\sum_{i=1}^{M} u_i = 1$. Note that $p_{i,l_i} = u_i(1 - \sum_{j=1}^{l_i-1} v_{i,j})$, $p_i = \sum_{j=1}^{l_i} p_{i,j} = u_i$ and thus $f(p) = f(u)$ with $u = (u_1, \ldots, u_{M-1})$. Then the numerator in (52) can be rewritten as

$$
\int_{\Delta_M(1)} f(u)\, g_{\alpha+n}(u)\, du \prod_{i=1}^{M-1} \int_{\Delta_{l_i}(1)} g_{\alpha_{i,\cdot}+n_{i,\cdot}}(v_{i,\cdot})\, dv_{i,\cdot},
$$

where $\alpha_{i,\cdot}, n_{i,\cdot}$ and $v_{i,\cdot}$ are the vectors of $\alpha_{i,j}, n_{i,j}$ (with $j = 1, \ldots, l_i$) and $v_{i,j}$ (with $j = 1, \ldots, l_i - 1$) and where we have used the fact that

$$
g_{\alpha+n}(u) = \prod_{i=1}^{M} u_i^{\sum_{j=1}^{l_i} n_{i,j} + \alpha_{i,j} - 1}
$$

since $\sum_{j=1}^{l_i} n_{i,j} + \alpha_{i,j} = \alpha_i + n_i$, $i = 1, \ldots, M$. A similar expression holds for the denominator in (52). Now, since f does not depend on the variables $v_{i,j}$ we can marginalize them out to obtain

$$E[f|n'] = \frac{\displaystyle\int_{\Delta_M(1)} f(u)\, g_{\alpha+n}(u)\, du}{\displaystyle\int_{\Delta_M(1)} g_{\alpha+n}(u)\, du}. \tag{A.17}$$

Therefore the functions to be minimizes in (50) and (52) are equal for any choice of α_i and thus, also for the values that give the lower expectation of f. □

Appendix A.11. Proof of Theorem 9

Consider the integrals in (54) and for any $i = 1,\ldots,M$ the change of variables: $p_{i,j} = u_i v_{i,j}$ for $j = 1,\ldots,l_i$, with $v_{i,j} > 0$, $\sum_{j=1}^{l_i} v_{i,j} = 1$ and $u_i > 0$, $\sum_{i=1}^{M} u_i = 1$. We look for the infimum of (54), which can be rewritten as

$$\frac{1}{K}\int_{\Delta_M(1)} \frac{f(u) g_{\alpha+n}(u) \prod_{i=1}^{M} g_{\alpha_{i,\cdot}+n_{i,\cdot}}(v_{i,\cdot})}{\left(\sum_{i=1}^{M} u_i \sum_{j=1}^{l_i} \beta_{i,j} v_{i,j}\right)^s} du, \tag{A.18}$$

where we have omitted the integration with respect to $v_{i,\cdot}$, $i = 1,\ldots,M$ and where K is the normalizing constant. Let β' and V be the vectors of all $\beta_{i,j}$ and $v_{i,j}$, respectively, then K can be written as

$$K(\alpha',\beta') = \int \prod_{i=1}^{M} g_{\alpha_{i,\cdot}+n_{i,\cdot}}(v_{i,\cdot}) K'(V,\alpha',\beta') dV.$$

where the integration is done over the simplexes $\Delta_{l_1}(1),\ldots,\Delta_{l_M}(1)$ and where

$$K'(V,\alpha',\beta') = \int_{\Delta_M(1)} \frac{g_{\alpha+n}(u)}{\left(\sum_{i=1}^{M} u_i \sum_{j=1}^{l_i} \beta_{i,j} v_{i,j}\right)^s} du. \tag{A.19}$$

Let $\beta^* = (\beta_1^*,\ldots,\beta_M^*)$ be the vector of parameters that gives the infimum of (53), we have that for any value of the variables in V, the infimum of

$$\frac{1}{K'(V,\alpha',\beta')}\int_{\Delta_M(1)} \frac{f(u) g_{\alpha+n}(u) \prod_{i=1}^{M} g_{\alpha_{i,\cdot}+n_{i,\cdot}}(v_{i,\cdot})}{\left(\sum_{i=1}^{M} u_i \sum_{j=1}^{l_i} \beta_{i,j} v_{i,j}\right)^s} du,$$

is found for

$$\sum_{j=1}^{l_i} \beta_{i,j} v_{i,j} = \beta_i^*.$$

Therefore, (A.18) is lower bounded by

$$\frac{\displaystyle\int \prod_{i=1}^{M} g_{\alpha_{i,\cdot}+n_{i,\cdot}}(v_{i,\cdot}) dV \int_{\Delta_M(1)} f(u) g_{\alpha+n}(u) \left(\sum_{i=1}^{M} u_i \beta_i^* \right)^{-s} du}{\displaystyle\int \prod_{i=1}^{M} g_{\alpha_{i,\cdot}+n_{i,\cdot}}(v_{i,\cdot}) dV \int_{\Delta_M(1)} g_{\alpha+n}(u) \left(\sum_{i=1}^{M} u_i \beta_i^* \right)^{-s} du}. \tag{A.20}$$

Then, we can marginalize out the variables $v_{i,j}$ both at the numerator and denominator and simplify them. Hence, the infimum does not depend on the variables $v_{i,j}$ which do not give us any additional degree of freedom in the optimization. The infimum in (A.20) is attained by taking $\beta_{i,j} = \beta_i^*/\beta_{\text{sum}}$ for all j, where the normalizing constant

$$\beta_{\text{sum}} = \sum_{i=1}^{M} l_i \beta_i^*$$

is used to impose the condition $\sum_{i=1}^{M} \sum_{j=1}^{l_i} \beta_{i,j} = 1$. Notice that, renormalizing the parameters β_i^* does not change the optimum, since the constant β_{sum} at the numerator and denominator of (A.20) simplifies.

References

[1] J. O. Berger, Statistical Decision Theory and Bayesian Analysis, 2nd edition, Springer Series in Statistics, New York, 1985.

[2] H. Jeffreys, Theory of Probability, 3rd edition, Clarendon Press, Oxford, 1983.

[3] J. B. S. Haldane, The precision of observed values of small frequencies, Biometrika 35 (3-4) (1948) 297–300.

[4] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman & Hall, 1991.

[5] J. O. Berger, E. Moreno, L. R. Pericchi, M. J. Bayarri, J. M. Bernardo, et al., An overview of robust Bayesian analysis, Test 3 (1) (1994) 5–124.

[6] P. Huber, Robust estimation of a location parameter, The Annals of Mathematical Statistics 35 (1) (1964) 73–101.

[7] S. Sivaganesan, J. Berger, Ranges of posterior measures for priors with unimodal contaminations, The Annals of Statistics 17 (2) (1989) 868–889.

[8] P. Walley, Inferences from multinomial data: learning about a bag of marbles (with discussion), Journal of the Royal Statistical Society. Series B (Methodological) 58 (1) (1996) 3–57.

[9] J.-M. Bernard, An introduction to the imprecise dirichlet model for multinomial data, International Journal of Approximate Reasoning 39 (2-3) (2005) 123 – 150.

[10] S. Favaro, G. Hadjicharalambous, I. Prünster, On a class of distributions on the simplex, Journal of Statistical Planning and Inference 141 (9) (2011) 2987–3004.

[11] J. F. C. Kingman, Completely random measures, Pacific Journal of Mathematics 21 (1) (1967) 59–78.

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

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

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

[15] A. Lijoi, I. Prünster, Models beyond the Dirichlet process, in: N. L. Hjort, C. C. Holmes, P. Müller, S. G. Walker (Eds.), Bayesian Nonparametrics, Cambridge University Press, Cambridge, 2010, pp. 80–136.

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

[17] B. Fristedt, L. Gray, A modern approach to probability theory, Birkhäuser Boston, 1996.

[18] D. CvijoviÄĞ, New identities for the partial Bell polynomials, Applied Mathematics Letters 24 (9) (2011) 1544 – 1547.

[19] J. P. Nolan, Stable Distributions - Models for Heavy Tailed Data, Birkhauser, Boston, 2015, in progress, Chapter 1 online at academic2.american.edu/∼jpnolan.

[20] A. Benavoli, M. Zaffalon, Prior near-ignorance for inferences in the k-parameter exponential family, available at http://www.idsia.ch/∼alessio/TR2011.pdf (2012).

[21] F. P. A. Coolen, T. Augustin, A nonparametric predictive alternative to the Imprecise Dirichlet Model: the case of a known number of categories, International Journal of Approximate Reasoning 50 (2) (2009) 217–230.

[22] P. Walley, A bounded derivative model for prior ignorance about a real-valued parameter, Scandinavian Journal of Statistics 24 (4) (1997) 463–483.

[23] F. P. Coolen, T. Augustin, Learning from multinomial data: a nonparametric predictive alternative to the imprecise dirichlet model., in: ISIPTA, Vol. 5, 2005, pp. 125–134.

[24] S. Moral, Imprecise probabilities for representing ignorance about a parameter, International Journal of Approximate Reasoning 53 (3) (2012) 347 – 362.

[25] G. de Cooman, E. Miranda, E. Quaeghebeur, Representation insensitivity in immediate prediction under exchangeability, International Journal of Approximate Reasoning 50 (2) (2009) 204 – 216.

[26] A. DasGupta, Asymptotic theory of statistics and probability, Springer, New York, 2008.