Finitary Bayesian Statistical Inference Through Partitions Tree Distributions

Federico Bassetti  
*Università degli Studi di Pavia, Pavia, Italy*

Pier Giovanni Bissiri  
*Università degli Studi di Cagliari, Cagliari, Italy*

**Abstract**

According to the Bayesian theory, observations are usually considered to be part of an infinite sequence of random elements that are conditionally independent and identically distributed, given an unknown parameter. Such a parameter, which is the object of inference, depends on the entire sequence. Consequently, the unknown parameter cannot generally be observed, and any hypothesis about its realizations might be devoid of any empirical meaning. Therefore it becomes natural to focus attention on finite sequences of observations. The present paper introduces specific laws for finite exchangeable sequences and analyses some of their most relevant statistical properties. These laws, assessed through sequences of nested partitions, are strongly reminiscent of Pólya-tree distributions and allow forms of conjugate analysis. As a matter of fact, this family of distributions, called partitions tree distributions, contains the exchangeable laws directed by the more familiar Polya-tree processes. Moreover, the paper gives an example of partitions tree distribution connected with the hypergeometric urn scheme, where negative correlation between past and future observations is allowed.


**Keywords and phrases.** de Finetti’s theorem, empirical distribution, finitary Bayesian inference, finite exchangeability, hypergeometric distribution, partitions tree distributions, Pólya tree distributions, predictive inference, random partitions, species sampling.

**1 Introduction**

In statistics, observations are often assumed to play a symmetric role with respect to prevision, meaning that the “chronological” order of observations is deemed irrelevant for all previsional purposes. For this reason,
frequentist statisticians consider observations as independent and identically distributed random variables with common distribution affected by unknown parameters. Bayesians assume instead that observations are conditionally independent and identically distributed given unknown parameters, which are the main objects of the inference.

This approach is challenged by de Finetti (see, e.g., de Finetti, 1937, 1974; and Roberts, 1965), who persuasively argues that it is not always possible to give concrete meaning to these parameters. He claims that it should be possible, at least in theory, to experimentally verify hypotheses about the parameters. We call any hypothesis having this property empirical. Bayesian statisticians very often ignore this precaution. They adopt the above hypothesis of conditional independence and indiscriminately draw inference from observations for both empirical and non empirical hypotheses.

Without assuming the existence of unobservable parameters, the above-mentioned conditional formulation does not have a clear interpretation anymore. Therefore it is natural to describe the symmetry of observations resorting to the notion of exchangeability, as used and studied by de Finetti. As a matter of fact, if the observation process is assumed to be infinitely extensible, the two formulations are mathematically equivalent according to the strong version of the well-known de Finetti’s representation theorem. On the other hand, in many situations, such as sampling from a finite population, the assumption of infinite extensibility of the observation process is not consistent with the real situation under study. Therefore, one might be forced to construct probability laws for the observations, without resorting to the usual conditional formulation. Hence it is necessary to find alternative methods for defining laws for any kind of exchangeable sequences.

Considering all the above remarks, we aim at (a) presenting specific forms of exchangeable laws defined according to the characteristics of actual situations, apart from the standard conditional formulation, and (b) working out some of their inherent statistical problems. All the proofs are deferred to the Appendix.

2 Inference Based on $N$-exchangeable Observations

To start with, let us mention some remarkable facts connected with the conditional form of the laws of infinite exchangeable random sequences.

Suppose that each observation takes values in some measurable space $(\mathcal{X}, \mathcal{B})$. Write $\mathcal{X}^N$ for the $N$-fold product $\mathcal{X} \times \cdots \times \mathcal{X}$ and $\mathcal{X}^\infty$ for $\mathcal{X} \times \mathcal{X} \times \cdots$, \[ \mathcal{X} \times \mathcal{X} \times \cdots, \]
and indicate by $\mathcal{B}^N$ and $\mathcal{B}^\infty$ the usual product $\sigma$-fields on $X^N$ and $X^\infty$, respectively. In the usual Bayesian framework, the observation process is assumed to be extendible to infinity, that is, observations $\xi_i$’s are viewed as coordinates of a random element taking values in $(X^\infty, \mathcal{B}^\infty)$. Recall also that the random elements $\xi_i$’s are said to be exchangeable if the distribution of $(\xi_{\sigma(i)})_{i \geq 1}$ is the same as the distribution of $(\xi_i)_{i \geq 1}$ for any finite permutation $\sigma$ of $(1, 2, \ldots)$. Finally, define as usual the empirical distribution of $\xi(n) := (\xi_1, \ldots, \xi_n)$, $(n = 1, 2, \ldots)$, as the random probability

$$
\tilde{e}_n(\cdot) = \frac{1}{n} \sum_{i=1}^n \delta_{\xi_i}(\cdot),
$$

where $\delta_{\xi_i}(A) = 1$ or $0$ depending on whether $\xi_i$ belongs to $A$ or not.

Given any random variable $V$, $L_V$ will denote its probability distribution. Moreover, for any other random element $U$, $L_{V|U}$ will stand for a conditional probability distribution for $V$ given $U$.

In this notation, de Finetti’s representation theorem can be stated as follows.

Let $X$ be a complete and separable metric space and $X$ its Borel sigma-field. If $(\xi_i)_{i \geq 1}$ is exchangeable, then there exists a random probability $\tilde{p}$ on $(X, \mathcal{B})$ such that:

$$
L_{\xi(n)|\tilde{p}}(A) = \tilde{p}^{(n)}(A) \quad (A \in \mathcal{B}^n, \ n = 1, 2, \ldots), \quad (2.1)
$$

where $p^{(n)}$ denotes the probability that makes $\xi_1, \ldots, \xi_n$ independent with the same distribution $p$, and $\tilde{p}$ is the weak limit of $\tilde{e}_N$.

In this general formulation, $\tilde{p}$ replaces the usual unknown parameter. In this case, it is customary to speak about Bayesian nonparametric representation and, consequently, about Bayesian nonparametric methods. In general, Bayesian inference concerns functions of $\tilde{p}$. Therefore, the just recalled de Finetti’s representation theorem highlights that those inferences usually deal with hypotheses that might be devoid of any empirical value, being related to limiting mathematical entities.

Assuming that observations form a finite exchangeable sequence $(\xi_1, \ldots, \xi_N)$, representation $(2.1)$ does not hold anymore. Therefore, to deal with the finitary approach, $(2.1)$ is to be replaced by its finite version. This version states that a finite random sequence $(\xi_1, \ldots, \xi_N)$ is exchangeable if and only if, for each $n \leq N$, conditionally on $\tilde{e}_N$, $\xi_1, \ldots, \xi_n$ are distributed as
n drawings without replacement from an urn with \( N \) balls, with \( N \approx e^N \{x\} \) balls having label \( x \), for each atom \( x \) of \( \tilde{e}_N \).

Going back to the Bayesian framework, notice that \( L_{\tilde{p}\xi(n)} \) denotes the posterior distribution, while the predictive distribution is \( L_{\xi(n,N)\xi(n)} \) with \( \xi(n,N) := (\xi_{n+1}, \ldots, \xi_N) \). Predictive distributions represent the sole aspect of the finitary approach that is taken into consideration by the usual conditional Bayesian standpoint, where, nevertheless, these distributions are viewed as functionals of posterior laws, namely

\[
L_{\xi(n,N)\xi(n)}(A) = \int_{\mathcal{P}} p^{(N-n)}(A)L_{\tilde{p}\xi(n)}(dp) \quad (A \in \mathcal{X}^{N-n}).
\]

It is clear that in a pure finitary setting this expression could be inadmissible, as it happens when \( \tilde{p} \) does not exist because of the finiteness of \( (\xi_n)_{n \geq 1} \). However, \( L_{\xi(n,N)\xi(n)} \) can always be assessed by resorting to the definition of conditional distribution.

3 Assessing the Law of an Exchangeable Sequence by the Law of the Empirical Process

The previous section discusses why, in many statistical problems, the usual directing measure \( \tilde{p} \) of the exchangeable sequence of observations should be replaced by the empirical measure, both as object of inference and as intermediate tool for assessing the joint distribution of the observations. In this section, we deal with a general method of assigning probability distributions for \( \tilde{e}_N \). According to the Kolmogorov’s consistency theorem, this method is based on the concept of projective family of probability measures. According to the nature of the domain of the probability law we have to assign, the indexing set of the above-mentioned family must be a class of events, more specifically a \( \sigma \)-algebra \( \mathcal{X} \) of subsets of \( \mathbb{X} \). In order to define \( \mathcal{X} \), we begin with a sequence \( (\pi_m)_{m \geq 0} \) of finite partitions of \( \mathbb{X} \), such that \( \pi_0 := \{\mathbb{X}\} \) and \( \pi_{m+1} \) is a refinement of \( \pi_m \) for every \( m \geq 0 \). Moreover, set \( \mathcal{G} := \cup_{m \geq 0} \pi_m \), and denote by \( \mathcal{A} \) the algebra generated by \( \mathcal{G} \) and by \( \mathcal{A}^N \) the algebra of all finite disjoint unions of cartesian products of sets in \( \mathcal{A} \). Finally, let \( \mathcal{X} \) being the \( \sigma \)-algebra generated by \( \mathcal{G} \). For each \( \pi_m = \{B_{m,1}, \ldots, B_{m,k_m}\} \), define a finite-dimensional distribution (element of the projective family) as follows. Assume there exists a random vector \( (\xi_1, \ldots, \xi_N) \) taking values in \( \mathbb{X}^N \), and, for each \( B_{m,j} \), define \( N(B_{m,j}) \) to be the (random) number of elements of \( (\xi_1, \ldots, \xi_N) \) contained in \( B_{m,j} \).
Federico Bassetti and Pier Giovanni Bissiri

\[ \tilde{N}(B) = N e_{B}(B) \]

i.e., \( \tilde{N}(B) = N e_{N}(B) \). Hence, the above-mentioned projective family corresponds to the set of the probability distributions of the random vectors \( \tilde{N}_0 := \tilde{N}(X) = N, \quad \tilde{N}_m := (\tilde{N}(B_m, 1), \ldots, \tilde{N}(B_m, k_m)) \) \( (m = 1, 2, \ldots) \). Of course, the following identity

\[ \sum_{B : \text{ge}(B) = C} \tilde{N}(B) = \tilde{N}(C) \quad (3.1) \]

holds for each \( C \) in \( \mathcal{G} \), where \( \text{ge}(B) \) stands for the most recent superset of \( B \) in \( \pi_{m} \), i.e., the set \( C \) in \( \pi_{m-1} \) that includes \( B \). Conversely, if a projective family \( \mathcal{L}_{\tilde{N}_m} \) \( (m = 1, 2, \ldots) \) of laws is assessed so that \( (3.1) \) holds, then there exists a unique finitely-additive probability \( P \) on the product measurable space \( (X^N, \mathcal{A}^N) \), whose coordinate functions are just \( (\xi_1, \ldots, \xi_N) \) such that

\[ P\{\xi_1 \in B_{m,i_1}, \ldots, \xi_N \in B_{m,i_N}\} = \frac{1}{(N_1 \ldots N_m)^{\mathcal{L}_{\tilde{N}_m}}(\{(N_1, \ldots, N_k_m)\})} \quad (3.2) \]

holds for any \( m \geq 1 \) and any \( n \)-tuple \( (i_1, \ldots, i_n) \) of elements from \( \{1, \ldots, k_m\} \) with \( N_j := |\{l = 1, \ldots, N : i_l = j\}| \) for \( j = 1, \ldots, k_m \). From now on, \( |A| \) denotes cardinality of set \( A \).

It is worth noticing that if \( X \) is a complete separable metric space, then, under the further condition that

\[ \tilde{N}(C_n) \to 0 \text{ in law} \quad (3.3) \]

for any decreasing sequence \( (C_n)_{n \geq 1} \) of events in \( \mathcal{G} \) such that \( C_n \downarrow \emptyset \), there is a unique probability measure on \( (X^N, \mathcal{A}^N) \) satisfying \( (3.2) \), \( \mathcal{A} \) being the \( \sigma \)-algebra generated by \( \mathcal{G} \). See Bissiri (2006) for a detailed discussion of the existence and uniqueness both of the additive law \( P \) and its completely additive extension.

Of course, in order to effectively assess the law of \( (\tilde{N}_m)_{m \geq 1} \), one can proceed in many ways. A direct procedure based on a particular allocation scheme is described by Bassetti and Bissiri (2008). In this paper, the law of interest will be deduced after assessing a sequence of conditional laws \( \mathcal{L}_{\tilde{N}_m+1|\tilde{N}_m} \) for every \( m \geq 0 \). Indeed, thanks to \( (3.1) \), it is clear that \( (\tilde{N}_m)_{m \geq 1} \) meets the Markov property in the sense that, for every \( m \geq 1 \), \( \tilde{N}_{m+2} \) and \( (\tilde{N}_1, \ldots, \tilde{N}_m) \) turn out to be stochastically independent given \( \tilde{N}_{m+1} \). Hence, to specify the law of \( (\tilde{N}_m)_{m \geq 1} \), it is enough to assess \( \mathcal{L}_{\tilde{N}_m+1|\tilde{N}_m} \). The specific procedure taken in consideration is based on the same arguments...
used to define the well-known Pólya-tree distributions. See, for example, Ferguson (1974), Mauldin et al. (1992), Lavine (1992), Schervish (1995) and Ghosh and Ramamoorthi (2003).

4 Partitions Tree Distribution

At this stage, we are in a position to describe the particular laws that will be considered henceforth. They can be characterized by means of a set of properties, collected in the following Condition 4.1, which, as already mentioned, connects these laws with the Pólya-tree process. Like this process, they allow a sort of conjugate analysis (see Section 7). In order to state Condition 4.1, it is worth introducing some additional notation. For each $m \geq 0$ and each $C$ in $\pi_m$, set $h_C := |\{B \in \pi_{m+1} : \text{ge}(B) = C\}|$ and let $\psi_C$ be a function from $\{0, \ldots, N\} \times \{0, \ldots, N\}^{h_C}$ into $[0, 1]$ such that

$$
\psi_C(\tilde{N}(C); N_1, \ldots, N_{h_C}) = P(\tilde{N}(D_1) = N_1, \ldots, \tilde{N}(D_{h_C}) = N_{h_C} \mid \tilde{N}(C))
$$

a.s. (with respect to $P$) if $D_1, \ldots, D_{h_C}$ are the subsets of $C$ belonging to $\pi_{m+1}$, which will be called the descendents of $C$.

Moreover, $\mathcal{H}_{m_1, \ldots, m_k}(n_1, \ldots, n_k)$ is the probability of getting $n_j$ balls marked with $j$ ($j = 1, \ldots, k$), when one draws $n$ balls without replacement from an urn containing $m_j$ balls marked by $j$ ($j = 1, \ldots, k$).

**Condition 4.1.** For each $m$ in $\mathbb{N}$, the following hold.

4.1.1. The collections of random variables $\{\tilde{N}(B) : \text{ge}(B) = C\}$, as $C$ varies in $\pi_m$, are conditionally independent given $\tilde{N}_m$.

4.1.2. For each $C$ in $\pi_m$, the collections $\{\tilde{N}(B) : \text{ge}(B) = C\}$ and $\{\tilde{N}(B) : B \in \pi_m \setminus \{C\}\}$ are conditionally independent given $\tilde{N}(C)$.

4.1.3. For each $C$ in $\mathcal{B}$, with $\tilde{N}(C)$ not degenerate at zero and $M^{*}_C := \max\{M = 0, \ldots, N : P(\tilde{N}(C) = M) > 0\}$, and for each $(N_1, \ldots, N_{h_C})$ such that $N_1 + \cdots + N_{h_C} = M \leq M^{*}_C$, the identity

$$
\psi_C(M; N_1, \ldots, N_{h_C}) = \sum_{M_1, \ldots, M_{h_C}} \mathcal{H}_{M_1, \ldots, M_{h_C}}(N_1, \ldots, N_{h_C}) \psi_C(M^{*}_C; M_1, \ldots, M_{h_C}), \quad (4.1)
$$

must hold with the sum running over all vectors $(M_1, \ldots, M_{h_C})$ such that $M_1 + \cdots + M_{h_C} = M^{*}_C$. 

From now on, any exchangeable law for \((\xi_1, \ldots, \xi_N)\) that satisfies Condition 4.1 will be called partitions tree distribution.

It should be noted that Conditions 4.1.1–4.1.2 are equivalent to assuming that

\[
\mathcal{L}_{\tilde{N}_{m+1} \mid \tilde{N}_m} = \times_{C \in \pi_m} \mathcal{L}_{(\tilde{N}(B) \mid \text{ge}(B) = C) \mid \tilde{N}(C)},
\]  

(4.2)

Moreover, they imply that, conditionally on the knowledge of the frequency of a set \(C\) in \(\pi_m\), any additional information on the frequencies of other elements of \(\pi_m\), or of their descendants in \(\pi_{m+1}\) does not affect the prevision of frequencies of the subsets of \(C\).

Condition 4.1.3 is the same as assuming that, for each \(C\) in \(\mathcal{G}\) such that \(P\{\tilde{N}(C) > 0\}\),

\[
\mathcal{L}_{(\tilde{N}(B_{m+1,j}) \mid \text{ge}(B_{m+1,j}) = C) \mid \tilde{N}(C)} = \mathcal{L}_{\left(\sum_{i=1}^{M_C} \mathbb{1}_{\{Y_i(C) = j\}} \mid \text{ge}(B_{m+1,j}) = C\right)},
\]  

(4.3)

holds for some exchangeable random vector \(Y(C) = (Y_1(C), \ldots, Y_{M_C^*}(C))\), where for \(1 \leq i \leq M_C^*\), \(Y_i(C)\) belongs to \(\{j = 1, \ldots, k_{m+1} : B_{m+1,j} \subset C\}\), and

\[
M_{B_{m+1,j}} := \max \left\{ M = 0, \ldots, N : \mathcal{L}_{\sum_{i=1}^{M_C^*} \mathbb{1}_{\{Y_i(C) = j\}}} (\{M\} > 0) \right\},
\]

for each \(j\) such that \(B_{m+1,j} \subset C\). For details see the Appendix. Notice that each vector \(Y(C)\) (with \(M_C^* > 0\)) can be taken to be the outcome of \(M_C^*\) drawings from an urn according to some particular scheme.

The marginal law of any element of \((\xi_1, \ldots, \xi_N)\) can be derived from \((\mathcal{L}_{\tilde{N}_{m+1} \mid \tilde{N}_m})_{m \geq 0}\), by exploiting Condition 4.1.3.

**Proposition 4.1.** If Condition 4.1 holds, then, for each \(B\) in \(\mathcal{G}\) such that \(P\{\tilde{N}(B) > 0\} > 0\), we have

\[
\mathbb{E}\left(\frac{\tilde{N}(B)}{M} \mid \tilde{N}(\text{ge}(B)) = M\right) = \frac{\mathbb{E}(\tilde{N}(B))}{\mathbb{E}(\tilde{N}(\text{ge}(B)))}
\]

for each positive \(M\) such that \(P\{\tilde{N}(\text{ge}(B)) = M\} > 0\) and

\[
P\{\xi_1 \in B\} = \mathbb{E}\left(\frac{\tilde{N}(B_1)}{N}\right) \prod_{j=2}^{m} \mathbb{E}\left(\frac{\tilde{N}(B_j)}{M_{j-1}} \mid \tilde{N}(B_{j-1}) = M_{j-1}\right), \quad (B \in \pi_m),
\]

(4.4)

where \(B_m = B\) and, for \(j < m\), \(B_j\) denotes the set in \(\pi_j\) that contains \(B_m\), and \(M_j\) is any positive value such that \(P\{\tilde{N}(B_j) = M_{j}\}\) is positive.
Two examples of partitions tree distributions will be described in the next sections.

5 Hypergeometric Partitions Tree Distributions

It is very easy to verify Condition 4.1 when \((\xi_1, \ldots, \xi_N)\) are independent and identically distributed. More interesting classes of distributions can be constructed following familiar urn schemes. A feature of the resulting schemes, which could be of some interest with respect to statistical inference, is that they allow negative correlation between past and future observations, in contrast to what happens, for example, in the presence of infinite exchangeable sequences. More precisely, they allow inverse relations between the predictive probability that a future observation belongs to a specific set \(A\) and the observed frequency of \(A\).

To see the point in assessing \(N\)-exchangeable laws of this kind, consider the following description of a concrete situation that seems to require forms of negative dependence between predictions and observed frequencies.

Example 5.1 (Species sampling). In the species sampling problem from a community of animals, one considers a finite community of \(N\) units and identifies each particular species with a real number in the unit interval \((0, 1]\), as it is usually done. Biologists classify each organism in a hierarchical way according to different taxonomic units or taxa: Phylum, Class, Order, Family, Genus, Species, etc. One can use the partitions tree structure, intrinsic to this classification process, to assign the probability distribution of \((\xi_1, \ldots, \xi_N)\), where \(\xi_i\) denotes the species of the \(i\)-th animal in the community. Now, if the animals share a common habitat, it seems reasonable to assess the above law by taking into account possible competitions between species belonging to the same taxonomic unit and within the same species. In situations in which there exists some a priori constraint on the number of units in the same taxon, one can resort to partitions tree distributions. Indeed, it is reasonable to assume such a constraint when competition is stronger among animals of the same species rather than between animals of species that are “far” from each other. One can argue that competition is reasonably stronger among animals of the same species since, for instance, they eat the same food, but it is present also among animals of similar species for the same reason. While a weaker competition is expected between species
which are “far” from each other in the tree structure, provided that the taxonomy takes into account specific features (e.g., habitat, eating habits) that can be seen as competition indexes. Therefore, one can resort to probability models that a priori set some constraints on the number of units in the same taxon. For instance, given that in a sample of \( n \) animals, \( n_1 \) mammals are detected, \( n_1 \) of which being carnivorous, it would make sense to assign the conditional probability that the animal detected at the stage \( (n + 1) \) is carnivorous, under the additional hypothesis that it is a mammal, in such a way that it turns out to be decreasing as \( n_1/n_1 \) increases. In other words, one presumes a sort of saturation effect in the taxon of carnivorous animals. It should be noted that it is unlikely that a sampling scheme without replacement is realistic in such a case, since one cannot remove the animals after observing them. However, one can mark the observed animals and not count them again in case they are re-observed.

5.1. Definition of hypergeometric partitions tree distributions. Let \( X \) be the interval \((0, 1]\), and let \( \mathcal{H} \) denote its Borel sigma-algebra. Put \( E = \{0, 1\} \) and \( E^0 := \emptyset \), \( E^* := \bigcup_{m=0}^{\infty} E^m \). Define \( \pi_m \) to be the set of all \( 2^m \) dyadic intervals of rank \( m \), i.e., \( \pi_m := \{I_\varepsilon : \varepsilon \in E^m\} \), where

\[
I_{\varepsilon_1 \ldots \varepsilon_m} := \left( \sum_{j=1}^{m} \varepsilon_j 2^{-j}, \sum_{j=1}^{m} \varepsilon_j 2^{-j} + 2^{-m} \right)
\]

if \( m \geq 1 \) and \( I_\emptyset = (0, 1] \).

In this case, \( \Pi := (\pi_m)_{m \geq 0} \) is a binary tree and, therefore, if Conditions 4.1.1 and 4.1.2 are in force, the law of \( \tilde{N} \) can be determined just by the assessment of the conditional distribution of \( \tilde{N}(I_{\varepsilon_1}) \) given \( \tilde{N}(I_\varepsilon) \) for every \( \varepsilon \in E^* \). Of course, such a distribution is supported by \( \{0, \ldots, \tilde{N}(I_\varepsilon)\} \). Now a practical way to assess the conditional distribution of \( \tilde{N}(I_{\varepsilon_1}) \) given \( \tilde{N}(I_\varepsilon) \) for each \( \varepsilon \in E^* \) will be presented.

Introduce a set of nonnegative integers \( \mathcal{S} := \{\alpha_\varepsilon : \varepsilon \in \mathcal{N} : \varepsilon \in E^*\} \) with \( \alpha_\emptyset := N \) satisfying

\[
\alpha_\emptyset + \alpha_{\varepsilon_1} \geq \alpha_\varepsilon \quad (5.1)
\]

for every \( \varepsilon \in E^* \). At this stage, assign the conditional distribution of the random variable \( \tilde{N}(I_{\varepsilon_1}) \) given \( \tilde{N}(I_\varepsilon) \), for \( \varepsilon \in E^* \), in such a way that it turns out to be the same as the distribution of the number of white balls in a sample without replacement of size \( \tilde{N}(I_\varepsilon) \) drawn from an urn containing \( \alpha_{\varepsilon_1} + \alpha_\emptyset \) balls, of which \( \alpha_{\varepsilon_1} \) are white and \( \alpha_\emptyset \) are black.

In concrete terms, the process \( \tilde{N} \) may be generated according to the following scheme.
1. Draw $N$ balls without replacement from an urn with $\alpha_1$ white balls and $\alpha_0$ black balls, and suppose that you get $N_1$ white balls and $N_0 := N - N_1$ black balls.

2. Now draw, without replacement, $N_1$ balls from an urn with $\alpha_{11}$ white balls and $\alpha_{10}$ black balls, and $N_0$ balls from an urn with $\alpha_{01}$ white balls and $\alpha_{00}$ black balls, respectively. Suppose the former sample contains $N_{11}$ white balls and $N_{01}$ black balls, while the latter contains $N_{01}$ white balls and $N_{00}$ black balls.

3. Continue this process in the following way: at the $m$-th step, draw $N_\varepsilon$ balls from an urn with $\alpha_{\varepsilon 1}$ white balls and $\alpha_{\varepsilon 0}$ black balls, for each $\varepsilon$ in $E^{m-1}$, and let $N_{\varepsilon 0}$ and $N_{\varepsilon 1}$ be the observed number of black balls and white balls respectively.

4. $N_\varepsilon$ is the number of observations that belong to $I_\varepsilon$ for each $\varepsilon$ in $E^*$. 

Note that at the $(m+1)$-th step the total number of balls in each urn must be greater than the number of trials. Since the number of trials $N_\varepsilon$ at the $m$-th step is less than or equal to the number of balls of the corresponding colour in the urn at the $(m-1)$-th step – which is $\alpha_\varepsilon$ – everything makes sense whenever (5.1) holds.

The above-mentioned procedure gives rise to a unique exchangeable finitely-additive probability on the algebra $\mathcal{A}^N$. Regarding the existence of a unique $\sigma$–additive extension $P$ of such a probability to $\mathcal{X}^N$, in this case, (3.3) becomes

$$\lim_{m \to \infty} \prod_{j=1}^{m} \frac{\alpha_{\varepsilon_1 \ldots \varepsilon_j 1}}{\alpha_{\varepsilon_1 \ldots \varepsilon_j 0} + \alpha_{\varepsilon_1 \ldots \varepsilon_j 1}} = 0 \quad (5.2)$$

for any zero-one sequence $\varepsilon_1, \varepsilon_2, \ldots$ satisfying $\varepsilon_m = 0$ and for every $m \geq k$ and for some $k$. See Proposition A.2.

If the set $\mathcal{S}$ satisfies (5.1) and (5.2), and the empirical process of the sequence $(\xi_1, \ldots, \xi_N)$ is generated by the above urn scheme, then we shall denote the distribution of this sequence by $\mathcal{H}(\mathcal{S})$.

Observe that if $\alpha_{\varepsilon 0} + \alpha_{\varepsilon 1} = \alpha_\varepsilon$ for each $\varepsilon$ in $E^*$, then $(\xi_1, \ldots, \xi_N)$ happens to be distributed according to a drawing scheme without replacement from an urn with $(\alpha_0 + \alpha_1)$ balls such that, for each $x$ in $(0, 1]$, the number of balls labelled with $x$, initially contained in the urn, is the limit of $\alpha_{\varepsilon_1 \ldots \varepsilon_m}$.
as \( m \to +\infty \), where \((\varepsilon_1, \varepsilon_2, \ldots)\) is the dyadic expansion of \( x \). Throughout the present paper, for any point \( x \) with two binary expansions, we take the nonterminating one.

It is clear that any sequence \((\xi_1, \ldots, \xi_N)\) distributed according to \( \mathcal{H}(\mathcal{G}) \) is not infinitely extendible, i.e., it is not distributed as the initial segment of any infinite exchangeable sequence. In fact, if it were extendible, the sequence \( \mathbb{I}_{\{\xi_1 \in (0, 1/2]\}}, \ldots, \mathbb{I}_{\{\xi_N \in (0, 1/2]\}} \) would also be extendible, whilst it happens to have the same distribution as the one of \( N \) drawings without replacement from an urn.

The following proposition provides useful expressions both for the law and for the expectation of each \( \xi_i \).

**Proposition 5.1.** Under \( \mathcal{H}(\mathcal{G}) \) the law of each \( \xi_i \) is characterized by

\[
P\{\xi_i \in I_\varepsilon\} = \frac{\alpha_{\varepsilon_1}}{\alpha_0 + \alpha_1} \cdot \frac{\alpha_{\varepsilon_1 \varepsilon_2}}{\alpha_{\varepsilon_1 0} + \alpha_{\varepsilon_1 1}} \cdots \frac{\alpha_{\varepsilon_1 \cdots \varepsilon_m}}{\alpha_{\varepsilon_1 \cdots \varepsilon_{m-1} 0} + \alpha_{\varepsilon_1 \cdots \varepsilon_{m-1} 1}},
\]

for each \( \varepsilon = (\varepsilon_1 \ldots \varepsilon_m) \) in \( E^* \), while the expectation of each \( \xi_i \) is

\[
E(\xi_i) = \sum_{\varepsilon \in E^m} \sum_{m=1}^{+\infty} \frac{1}{2^m} \prod_{k=1}^m \frac{\alpha_{\varepsilon_1 \cdots \varepsilon_k}}{\alpha_{\varepsilon_1 \cdots \varepsilon_{k-1} 0} + \alpha_{\varepsilon_1 \cdots \varepsilon_{k-1} 1}}.
\]

**5.2. A bivariate model.** In this section, we assume that another variable \( \zeta_i \) of interest is observed together with \( \xi_i \). For each \( 1 \leq i \leq N \), let \((\xi_i, \zeta_i)\) be a random vector taking value in \(([0, 1] \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{Y})\), \( \mathcal{Y} \) being a Polish space with Borel \( \sigma \)-field \( \mathcal{Y} \). Suppose, for instance, that \( \xi_i \) quantifies a potential risk factor in a clinical study and \( \zeta_i \) the response. One can imagine that each \( \xi_i \) is the result of a hierarchical classification process (for instance, sex, age, smoking, eating habits and so on). Hence, a tree structure naturally arises again and can be used to assess the law of \((\xi_1, \ldots, \xi_N)\). It is sometimes reasonable to assume a priori a dispersion effect in the population among the different classes. In other words, we do not expect that too many subjects will fall into the same class. Hence, we choose a hypergeometric partition tree distribution for the law of \((\xi_1, \ldots, \xi_N)\), and therefore we can assign an exchangeable law to \((\xi_1, \zeta_1), \ldots, (\xi_N, \zeta_N)\) by fixing the conditional distribution of \((\zeta_1, \ldots, \zeta_N)\) given \((\xi_1, \ldots, \xi_N)\). To this end, set for every \( \varepsilon \) in \( E^m \) (\( m \) being a fixed integer), \( H_\varepsilon = \{i = 1, \ldots, N : \xi_i \in I_\varepsilon\} \) and \( \Delta = \{\varepsilon \in E^m : H_\varepsilon \neq \emptyset\} \). Moreover, denote by \( Q^{(k)} \) the \( k \)-dimensional
marginal law of any exchangeable law $Q$. At this stage, set

$$\mathcal{L}(\zeta_1, \ldots, \zeta_N)(A_1 \times \cdots \times A_N) = \prod_{\varepsilon \in \Delta} q_\varepsilon^{[H_\varepsilon]}(\times_{i \in H_\varepsilon} A_i), \quad (5.5)$$

for every $N$-tuple $(A_1, \ldots, A_N)$ of sets in $\mathcal{Y}$, where $\{q_\varepsilon : \varepsilon \in E^m\}$ is a family of exchangeable probability measures on $\mathcal{Y}^N$.

The law of $((\xi_1, \zeta_1), \ldots, (\xi_N, \zeta_N))$ is exchangeable since $(\xi_1, \ldots, \xi_N)$ also is exchangeable, and

$$\mathcal{L}(\xi_1, \ldots, \xi_N)(A_1 \times \cdots \times A_N) = \mathcal{L}(\zeta_1, \ldots, \zeta_N)(A_{\sigma_1} \times \cdots \times A_{\sigma_N}) \quad (5.6)$$

for every $N$-tuple $(A_1, \ldots, A_N)$ of sets in $\mathcal{Y}$ and for every permutation $\sigma$ of $(1, \ldots, N)$.

Letting $A_{n+1} = \cdots = A_N = \mathbb{R}$, the conditional probability

$$\mathcal{L}(\zeta_1, \ldots, \zeta_n)(A_1 \times \cdots \times A_n) = \prod_{\varepsilon \in \Delta} q_\varepsilon^{[H_\varepsilon]}(\times_{i \in H_\varepsilon} A_i),$$

where $H'_\varepsilon = \{i = 1, \ldots, n : \xi_i \in I_\varepsilon\}$ and $\Delta' = \{\varepsilon \in E^m : H'_\varepsilon \neq \emptyset\}$. At this stage, notice that the right hand side of the last equation does not depend on $(\xi_{n+1}, \ldots, \xi_N)$. Therefore, for every $n \leq N$,

$$\mathcal{L}(\zeta_1, \ldots, \zeta_n)(\xi_1, \ldots, \xi_n) = \mathcal{L}(\zeta_1, \ldots, \zeta_n)(\zeta_1, \ldots, \zeta_n), \quad (5.7)$$

i.e., $(\zeta_1, \ldots, \zeta_n)$ and $(\xi_{n+1}, \ldots, \xi_N)$ are conditionally independent given $(\xi_1, \ldots, \xi_n)$.

### 6 Pólya-tree Processes and Partitions Tree Distributions

As recalled in Section 4, Condition 4.1 is inspired by the theory of Pólya-tree processes. If $(\xi_n)_{n \geq 1}$ is an infinite sequence of exchangeable random elements, with de Finetti’s representation directed by some Pólya-tree distribution – i.e., a law of a Pólya tree process – it will be shown that the law of the empirical distribution of $(\xi_1, \ldots, \xi_N)$ satisfies Condition 4.1 for every $N$. In order to understand this statement, it is worth recalling one of the well-known characterizations of the Pólya-tree processes, using the same notations as those in Section 4.

A random probability measure $\tilde{p}$ is said to be a Pólya tree process with parameter $\{\alpha_{m, j} \geq 0 : j = 1, \ldots, k_m; m = 0, 1, \ldots\}$ whenever the following conditions hold.
The random collections \( \{ p(C \mid \text{ge}(C)) : C \in \pi_m \} \), with \( m = 1, 2, \ldots \), are stochastically independent (independence between partitions), i.e., \( p \) is an F-neutral process (also called tail free process; see Ferguson, 1974).

The random collections \( \{ p(C \mid B) : \text{ge}(C) = B \} \), with \( B \) varying in \( \pi_m \), are stochastically independent for each \( m \) (independence within partitions).

For each \( B \) in \( \pi_m \) and for \( m \geq 0 \), the random vector \( (p(C \mid B) : \text{ge}(B) = C) \) has Dirichlet distribution with parameter \( (\alpha_{m+1,j} : B_{m+1,j} \subset B) \) (in order to correctly speak of random vectors or parameters vectors, it is necessary to introduce some order among the descendants of \( \text{ge}(\cdot) \), for example a natural left-to-right order).

**Proposition 6.1.** If \( \xi_1, \xi_2, \ldots \) is an (infinite) exchangeable sequence whose de Finetti’s measure is a Pólya-tree distribution with parameters \( \{ \alpha_{m,j} : j = 1, \ldots, k_m \} \), then, for each \( N \), \( (\xi_1, \ldots, \xi_N) \) satisfies Condition 4.1 with \( L(\tilde{N}(B) : \text{ge}(B) = C) \) given by the following form of Dirichlet-compound multinomial distribution

\[
L(\tilde{N}(B_{m,j}) : \text{ge}(B_{m,j}) = C) = \frac{\prod_{j \in \mathcal{T}(C)} (-\alpha_{m,j})}{\tilde{N}(C)} \frac{\prod_{j \in \mathcal{T}(C)} (\alpha_{m,j} + N_j - 1) \cdots \alpha_{m,j}}{\prod_{j \in \mathcal{T}(C)} N_j !} \frac{\tilde{N}(C)!}{(\sum_{j \in \mathcal{T}(C)} \alpha_{m,j} + \tilde{N}(C) - 1) \cdots \sum_{j \in \mathcal{T}(C)} \alpha_{m,j}}
\]

for \( l = 1, \ldots, k_{m-1} \) if \( m \geq 1 \), and \( \mathcal{T}(C) \) is the vector obtained by ordering the elements of the set \( \{ j = 1, \ldots, k_m : \text{ge}(B_{m,j}) = C \} \).

It should be noted that the scheme in Subsection 5.1 can be slightly modified to obtain the empirical process of the \( N \)-initial segment \( (N = 1, 2, \ldots) \) of an infinite exchangeable sequence directed by a Pólya-tree process, provided that \( X = (0, 1] \), and the sets in \( \pi_m \) \( (m \geq 1) \) are the dyadic intervals of rank \( m \). See Mauldin et al. (1992). To this end, it is enough, at each step \( m \) and for each \( \epsilon \) in \( E^{m-1} \), to draw \( N_\epsilon \) balls from an urn with \( \alpha_{\epsilon1} \) white
balls and $\alpha_{\epsilon_0}$ black balls according to the well-known Pólya scheme (i.e. the drawn ball is placed back in the urn along with one more ball of the colour drawn).

7 Partition Tree Posterior and Predictive Distributions

This section contains some results on predictive and posterior distributions relating to partitions tree distributions.

7.1. General case. The following proposition is useful to determine the posterior distribution for $\tilde{e}_N$, i.e., the conditional distribution of $\tilde{e}_N$ given $\tilde{\xi}(n)$, when Condition 4.1 is in force.

**Proposition 7.1.** If Condition 4.1 holds, then, for any $n \leq N$ and any vector $(N_1, \ldots, N_{k_m})$ of positive integers summing up to $N$,

$$P(\tilde{N}(B_{m,1}) = N_1, \ldots, \tilde{N}(B_{m,k_m}) = N_{k_m} \mid \xi_1 = x_1, \ldots, \xi_n = x_n)$$

$$= P^{a.s.}(\tilde{N}(B_{m,1}) = N_1, \ldots, \tilde{N}(B_{m,k_m}) = N_{k_m} \mid \xi_1 \in B_{m,1}^{x_1}, \ldots, \xi_n \in B_{m,n}^{x_n}),$$

(7.1)

where $B_{m,x}^x$ denotes the set of $\pi_m$ which contains $x$. Moreover,

$$P(\tilde{N}(B_{m,l_j}) = N_{m,l_j}, j = 1, \ldots, d \mid \tilde{N}(B_{m-1,l}) = M, \tilde{\xi}(n) = x(n))$$

$$= P(\tilde{N}(B_{m,l_j}) = N_{m,l_j}, j = 1, \ldots, d \mid \tilde{N}(B_{m-1,l}) = M, \xi_1 \in B_{m,1}^{x_1}, \ldots, \xi_n \in B_{m,n}^{x_n}),$$

(7.2)

where $l_1 \leq \cdots \leq l_d$ are such that $B_{m,l_j}$ is contained in $B_{m-1,l}$.

Proposition 7.1 says that the posterior distribution of $\tilde{N}_m$ given $(\xi_1, \ldots, \xi_n)$ is the same as the posterior distribution of $\tilde{N}_m$ given $\{\tilde{\xi}_j, (\xi_i) : j = 1, \ldots, k_m, i = 1, \ldots, m\}$. This property, which is shared by the process $\tilde{N}$ with F-neutral processes, makes calculations for the posterior easy. Now, applying Bayes’ theorem to the right hand side of (7.1), one gets

$$P(\tilde{N}(B_{m,1}) = N_1, \ldots, \tilde{N}(B_{m,k_m}) = N_{k_m} \mid \xi_1 \in B_{m,1}^{x_1}, \ldots, \xi_n \in B_{m,n}^{x_n})$$

$$= \frac{P(\xi_1 \in B_{m,1}^{x_1}, \ldots, \xi_n \in B_{m,n}^{x_n} \mid \tilde{N}(B_{m,1}) = N_1, \ldots, \tilde{N}(B_{m,k_m}) = N_{k_m})}{P(\xi_1 \in B_{m,1}^{x_1}, \ldots, \xi_n \in B_{m,n}^{x_n})} \cdot P(\tilde{N}(B_{m,1}) = N_1, \ldots, \tilde{N}(B_{m,k_m}) = N_{k_m}).$$
Therefore, resorting to the finite version of de Finetti’s representation theorem (already mentioned in Section 2), one obtains the following.

**Proposition 7.2.** If Condition 4.1 holds, then

\[
P(\tilde{N}(B_{m,1}) = N_1, \ldots, \tilde{N}(B_{m,k_m}) = N_{k_m} \mid \xi(n) = x(n))
\propto \mathcal{H}_{N_1, \ldots, N_{k_m}}(n_1, \ldots, n_{k_m}) \cdot P(\tilde{N}(B_{m,1}) = N_1, \ldots, \tilde{N}(B_{m,k_m}) = N_{k_m}),
\]

where \( n_j = |\{i = 1, \ldots, n : x_i \in B_{m,j}\}| \) with \( 1 \leq j \leq k_m \).

It is possible to show that the posterior for \((\tilde{N}(B): B \in \pi_m)\) is uniquely characterized by the conditional probability of \((\tilde{N}(B_{m,j}) = N_m,j: B_{m,j} \subset B_{m-1,l})\) given that \(\xi(n) = x(n)\) and \(\tilde{N}(B_{m-1,l}) = M\), for each \(l = 1, \ldots, k_{m-1}\). More precisely,

\[
P(\tilde{N}(B_{m,l_j}) = N_{m,l_j}, j = 1, \ldots, d \mid \tilde{N}(B_{m-1,l}) = M, \xi(n) = x(n))
\propto \mathcal{H}_{N_{l_1}, \ldots, N_{l_d}}(n_{l_1}, \ldots, n_{l_d}) \cdot P(\tilde{N}(B_{m,l_j}) = N_{m,l_j}, j = 1, \ldots, d \mid \tilde{N}(B_{m-1,l}) = M),
\]

where \(l_1 \leq \cdots \leq l_d\) are such that \(B_{m,l_j}\) is contained in \(B_{m-1,l}\).

This fact is based on the next proposition.

**Proposition 7.3.** If Condition 4.1 holds, then

\[
\mathcal{L}_{\tilde{N}_{m+1} | \tilde{N}_m, \xi_1, \ldots, \xi_n} = \prod_{j=1}^{k_m} \mathcal{L}_{\tilde{N}(B_{m+1,l}): B_{m+1,l} \subset B_{m,j} | \tilde{N}(B_{m,j}); \xi_i \in B_{m,j}}.
\]

At this point, to obtain (7.4), it is enough to apply Proposition 7.3 to (7.2) and argue as in the proof of Proposition 7.2.

### 7.2 Hypergeometric partitions tree posterior and predictive distributions

Proposition 7.1 yields explicit forms for posterior and predictive distributions relating to hypergeometric partitions tree distributions. Now, from Subsection 5.1, the conditional distribution of \(\tilde{N}(I_\varepsilon)\) (with \(\varepsilon \in E^*\)) given \(\tilde{N}(I_\varepsilon)\) turns out to be hypergeometric relative to \(\tilde{N}(I_\varepsilon)\) drawings, when the initial numbers of white and black balls are \(\alpha_{\varepsilon 1}\) and \(\alpha_{\varepsilon 0}\), respectively. Therefore, applying Proposition 7.1, it is straightforward to see that
the conditional law of \( \sum_{i=n+1}^{N} \delta_{\xi_i}(I_{\bar{\xi}_i}) \) given \( (\tilde{N}(I_e), \xi(n)) \) is the hypergeometric distribution for \( (\tilde{N}(I_e) - \tilde{n}_e) \) drawings when the initial numbers of white and black balls are \((\alpha_{\varepsilon_{1}} - n_{\varepsilon_{1}})\) and \((\alpha_{\varepsilon_{0}} - n_{\varepsilon_{0}})\), respectively, with \( \tilde{n}_e := \sum_{i=1}^{n} \delta_{\xi_i}(I_e) \). This yields the following.

**Proposition 7.4.** Let the distribution of \((\xi_1, \ldots, \xi_N)\) be \( \mathcal{H}(\mathcal{S}) \), where \( \mathcal{S} = \{\alpha_{\varepsilon} : \varepsilon \in E^*\} \). Then the conditional distribution of \((\xi_{n+1}, \ldots, \xi_N)\) given \((\xi_1, \ldots, \xi_n)\) is \( \mathcal{H}(\mathcal{S}^*) \), where \( \mathcal{S}^* := \{\alpha_{\varepsilon}^*: \varepsilon \in E^*\} \) and \( \alpha_{\varepsilon}^* := \alpha_{\varepsilon} - \sum_{i=1}^{n} \delta_{\xi_i}(I_e) \) for each \( \varepsilon \) in \( E^* \).

In particular, for the predictive distribution, one has

\[
P\{(\xi_{n+1} \in I_{\bar{\xi}_1 \ldots \bar{\xi}_m} \mid \xi(n))\} = \mathbb{E}\left(\frac{\sum_{i=n+1}^{N} \delta_{\xi_i}(I_{\bar{\xi}_1 \ldots \bar{\xi}_m})}{N - n} \mid \xi(n)\right)
\]

\[
= \frac{\alpha_{\varepsilon_1} - \tilde{n}_{\varepsilon_1}}{\alpha_0 + \alpha_1 - n} \cdot \frac{\alpha_{\varepsilon_1 \varepsilon_2} - \tilde{n}_{\varepsilon_1 \varepsilon_2}}{\alpha_{\varepsilon_1} \varepsilon_0 + \alpha_{\varepsilon_1 \varepsilon_2} - \tilde{n}_{\varepsilon_1 \varepsilon_2}} \cdots \frac{\alpha_{\varepsilon_1 \ldots \varepsilon_m 1} - \tilde{n}_{\varepsilon_1 \ldots \varepsilon_m 1}}{\alpha_{\varepsilon_1 \ldots \varepsilon_m 0} + \alpha_{\varepsilon_1 \ldots \varepsilon_m 1} - \tilde{n}_{\varepsilon_1 \ldots \varepsilon_m}}.
\]  

(7.6)

It is worth noticing that \( \mathcal{H}(\mathcal{S}) \) meets the requirements discussed at the beginning of Section 5 and, in particular, in Example 5.1, where the species sampling problem from a community of animals is dealt with.

**Example 1 (continued).** Here, the observations range in the class of all species. Assume that a sample with just one unit is taken, i.e., only \( \xi_1 \) is observed. Then the predictive distribution of \( \xi_2 \) should put a lower mass around \( \xi_1 \) than the one assigned by the (unconditional) distribution \( \mathcal{L}_{\xi_1} \) of the single observation. If \( \xi_1 \) belongs to \( I_1 = (1/2, 1] \), then the conditional probability that the second observation belongs to \( I_1 \), given \( \xi_1 \), should be lower (w.r.t. the unconditional distribution) on the set \( I_1 \) and higher on \( I_0 := (0, 1/2] \); the opposite occurs if \( \xi_1 \) belongs to \( I_0 \). In fact, in our model, such a probability can be written as

\[
P(\xi_2 \in I_1 \mid \xi_1) = p_1 P(\xi_1 \in I_1) - (1 - p_1)I_{\{\xi_1 \in I_1\}},
\]

where \( p_1 = (\alpha_0 + \alpha_1)/(\alpha_0 + \alpha_1 - I_{\{\xi_1 \in I_1\}}) \).

More generally, the conditional probability that the species \( \xi_{n+1} \) of the \((n + 1)\)-th animal belongs to \( I_{\bar{\xi}_1} \), given \( \xi_1, \ldots, \xi_n \) and that \( \xi_{n+1} \) belongs to
If \( I_\varepsilon \) turns out to be a linear combination of the conditional distribution of \( \xi_1 \) and the “conditional empirical measure” \( \tilde{n}_{\varepsilon1}/\tilde{n}_\varepsilon \), that is,

\[
P(\xi_{n+1} \in I_{\varepsilon1} \mid \xi(n), \xi_{n+1} \in I_\varepsilon) = p_{n+1}P(\xi_1 \in I_{\varepsilon1} \mid \xi_1 \in I_\varepsilon) - (1-p_{n+1})\frac{\tilde{n}_{\varepsilon1}}{\tilde{n}_\varepsilon},
\]

where \( \xi(n) := (\xi_1, \ldots, \xi_n) \),

\[
p_{n+1} = \frac{\alpha_{e0} + \alpha_{e1}}{\alpha_{e0} + \alpha_{e1} - \tilde{n}_\varepsilon} \quad \text{and} \quad P(\xi_1 \in I_{\varepsilon1} \mid \xi_1 \in I_\varepsilon) = \frac{\alpha_{e1}}{\alpha_{e0} + \alpha_{e1}}.
\]

7.3. Predictive distributions in the bivariate model. If \( ((\xi_1, \xi_1), \ldots, (\xi_N, \xi_N)) \) is distributed as in Section 5.2, then the predictive distributions can easily be obtained. In what follows, we denote by \( Q \) the law of \( ((\xi_1, \xi_1), \ldots, (\xi_N, \xi_N)) \).

For each \( \varepsilon \in E^m \), let \( \phi_m(x) = \varepsilon \) if \( x \) belongs to \( I_\varepsilon \) and set

\[
q_\varepsilon(A_{n+1}; A_1, \ldots, A_n) := q^{(n+1)}(\mathbb{R}^n \times A_{n+1} \mid A_1 \times \cdots \times A_n \times \mathbb{R})
\]

for \( 1 \leq n \leq N - 1 \) and any \((n+1)\)-tuple \( (A_1, \ldots, A_{n+1}) \) of sets in \( \mathcal{Y} \).

Combining (5.5) with (5.7), one obtains

\[
Q(\xi_{n+1} \in A_{n+1} \mid \xi(n+1), \xi(n) \in A_1 \times \cdots \times A_n)) = q_{\phi_m(\xi_{n+1})}(A_{n+1}; A_i, i \leq n : \xi_i \in I_\varepsilon),
\]

and if \( B_{n+1} \) belongs to \( \mathcal{X}_\varepsilon \), then

\[
Q(\xi_{n+1} \in A_{n+1}, \xi_{n+1} \in B_{n+1} \mid \xi(n), \xi(n) \in A_1 \times \cdots \times A_n)
= \int_{B_{n+1}} q_{\phi_m(x)}(A_{n+1}; A_i, i \leq n : \xi_i \in I_{\phi_m(x)}) \mathcal{L}_{\xi_{n+1}|\xi(n)}(dx).
\]

This assessment seems reasonable if one believes that the effect of \( \xi_i \) on \( \xi_i \) and the effect of \( \xi_j \) on \( \xi_j \) are the same when \( \xi_i \) and \( \xi_j \) are “close” enough to each other. In this specific case, \( \xi_i \) and \( \xi_j \) are considered close enough if they belong to the same dyadic interval, but their “closeness” can be evaluated considering \( |\xi_i - \xi_j| \) instead. To this end, one can resort to another assessment of the law of \( ((\xi_1, \xi_1), \ldots, (\xi_N, \xi_N)) \) directly assigning the predictive distributions as follows. Fix \( \delta > 0 \) and let \( q \) be an exchangeable probability distribution on \( \mathcal{Y}^N \). Set

\[
q(A_{n+1}; A_1, \ldots, A_n) := q^{(n+1)}(\mathbb{R}^n \times A_{n+1} \mid A_1 \times \cdots \times A_n \times \mathbb{R}).
\]
Predictive distributions can be assessed so that (5.7) is satisfied and
\[ Q(\zeta_{n+1} \in A_{n+1} | \xi(n+1), \zeta(n) \in A_1 \times \cdots \times A_n) = q(A_{n+1}; A_i, i \leq n : |\xi_i - \xi_{n+1}| < \delta). \quad (7.7) \]

In this way, one gets
\[ Q(\zeta_{n+1} \in A_{n+1}, \xi_{n+1} \in B_{n+1} | \xi(n), \zeta(n) \in A_1 \times \cdots \times A_n) = \int_{B_{n+1}} q(A_{n+1}; A_i, i \leq n : |\xi_i - x| < \delta) \mathcal{L}_{\xi_{n+1} | \xi(n)}(dx). \quad (7.8) \]

Notice that by (7.8),
\[ \mathcal{L}(\xi_{n+1}, \zeta_{n+1}) | (\xi_1, \ldots, \xi_n, \zeta_1, \ldots, \zeta_n) = \mathcal{L}(\xi_{n+1}, \zeta_{n+1}) | (\xi_{\sigma_1}, \ldots, \xi_{\sigma_n}, \zeta_{\sigma_1}, \ldots, \zeta_{\sigma_n}) \]
holds for every permutation \( \sigma \) of \( (1, \ldots, n) \). Moreover, using (5.7) and (7.7), one can prove that
\[ \mathcal{L}((\xi_{n+2}, \zeta_{n+2}), (\xi_{n+1}, \zeta_{n+1})) | (\xi(n), \zeta(n)) = \mathcal{L}((\xi_{n+2}, \zeta_{n+2}), (\xi_{n+1}, \zeta_{n+1})) | (\xi(n), \zeta(n)). \]

These two equalities imply that \( ((\xi_1, \zeta_1), \ldots, (\xi_N, \zeta_N)) \) is exchangeable. The proof of this fact is contained in the proof given by Forttini et al. (2000) for their Theorem 3.1.

8 Concluding Remarks and Future Work

Motivated by de Finetti’s criticism regarding inference about non-observable quantities, we focus our attention on a finitary Bayesian approach to statistical inference. This leads to the study of finite exchangeability from a statistical perspective. In particular, we propose to assess the law of observations by means of sequences of nested partitions.

Obviously, many other ways of assessing the law of a finite exchangeable sequence can be considered, see, for instance, Bassetti and Bissiri (2008). Moreover, it would be interesting to apply the proposed methodology – in particular the bivariate model – to some real problems.

Finally, we point out that infinite exchangeability gives rise to more tractable “a posteriori” quantities, which is, of course, very attractive in real applications. Apart from this practical advantage, the theoretical deficiency of the usual approach with respect to an empirical approach remains. Classical Bayesian inference can still be seen as a “limit” of the finitary approach, as the number of possible observations \( N \) diverges. From this point
Federico Bassetti and Pier Giovanni Bissiri

of view, it is interesting to study when posterior quantities computed under finite exchangeability can be approximated by those arising under the assumption of infinite exchangeability. Diaconis and Freedman (1980) provide an optimal bound for the total variation distance between the law of \((\xi_1, \ldots, \xi_n)\) and the law of \((\zeta_1, \ldots, \zeta_N)\), \((\xi_1, \ldots, \xi_N)\) being a given finite exchangeable sequence and \((\zeta_k)_{k \geq 1}\) a suitable infinite exchangeable sequence. However, such a result does not give a direct answer to this issue. It would be more appropriate to compare the conditional distribution of \(\tilde{\nu}_N\) given \(\bar{\nu}(\xi)\) with the conditional distribution of \(\tilde{\nu}\) given \(\bar{\nu}(\xi(n))\), when \((\xi_k)_{k \geq 1}\) is an infinite exchangeable sequence directed by \(\tilde{\nu}\). This kind of study will be the subject of a forthcoming paper.

Appendix

In order to prove (4.3), it is convenient to introduce a family of discrete r.v.’s \(W_{m,i} (m = 0, 2, \ldots, i = 1, \ldots, n)\) defined on \((\mathcal{X}^N, \mathcal{B}^N, P)\) such that
\[
W_{m,i} = \sum_{j=1}^{k_m} j \mathbb{I}_{\{\xi_i \in B_{m,j}\}},
\]
\(i.e., W_{m,i}\) is equal to \(j\) if \(\xi_i\) falls in \(B_{m,j}\).

Note that, by exchangeability, if \(D_1, \ldots, D_h\) are the descendants of \(C\), then
\[
P(\bar{\nu}(D_1) = N_1, \ldots, \bar{\nu}(D_h) = N_h \mid \bar{\nu}(C) = M) = \frac{P(\bar{\nu}(D_1) = N_1, \ldots, \bar{\nu}(D_h) = N_h, \bar{\nu}(C^c) = N - M)}{P(\bar{\nu}(C) = M)} = \left(\frac{M}{N_1 \cdots N_h}ight) \frac{P(F_M \cap E_M)}{P(E_M)} = \left(\frac{M}{N_1 \cdots N_h}\right) \frac{P(F_M \mid E_M)}{P(E_M)},
\]
\((A.1)\)

where \(M = N_1 + \cdots + N_h \leq N,\)
\(E_M = \{\xi_1 \in C, \ldots, \xi_m \in C, \xi_{M+1} \in C^c, \ldots, \xi_N \in C^c\}, \) and
\(F_M = \{\xi_1 \in D_1, \ldots, \xi_{N_1} \in D_1, \xi_{N_1+1} \in D_2, \ldots, \xi_M \in D_h\}.\)

Hence, (4.3) can be re-written as
\[
P(F_M \mid E_M) = P'(Y_1(C) = j_1, \ldots, Y_M(C) = j_M),
\]
\((A.2)\)

where \(P'\) is the probability defined on the space that supports all the \(Y(C)\)’s, and \((j_1, \ldots, j_M)\) is the vector such that \(B_{m+1,j_i} = D_i\) for \(i = 1, \ldots, M\), or equivalently
\[
P(W_{m+1,1} = j_1, \ldots, W_{m+1,M} = j_M \mid E_M) = P'(Y_1(C) = j_1, \ldots, Y_M(C) = j_M).
\]
\((A.3)\)
Therefore, (A.2) is equivalent to saying that, for any \((j_1, \ldots, j_M)\) such that 
\[\text{ge}(B_{m+1,j_i}) = C \ (i = 1, \ldots, M)\] 
and \(P(\bar{N}(C) = M) > 0\), and for any \(1 \leq n \leq M\),
\[P(W_{m+1,1} = j_1, \ldots, W_{m+1,n} = j_n \mid E_M) = P'(Y_1(C) = j_1, \ldots, Y_n(C) = j_n).\]
\[
\text{(A.4)}
\]
One can see that this fact is equivalent to Condition 4.1.3, by applying
the finite version of de Finetti’s representation theorem. Each vector \(Y(C)\) 
(with \(M^*_C > 0\)) can be taken to be the outcome of \(M^*_C\) drawings from an urn 
according to some particular scheme.

For each \(C \in \mathcal{G}\) and each \(0 \leq n \leq N\), denote
\[S_n(C) := \left\{ M = 0, \ldots, N - n : P(\bar{N}(C) = M + n) > 0 \right\}.
\]

**Proposition A.1.** Assume that Conditions 4.1.1–4.1.2 are satisfied.

The following facts are equivalent.

(i) Condition 4.1.3 hold.

(ii) For any \(C \in \mathcal{G}\), any \(n \leq N\) and any \(n\)-tuple \((A_1, \ldots, A_n)\) of sets 
such that \(\text{ge}(A_i) = C\), the conditional probability 
\[P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in C, \ldots, \xi_{n+M} \in C, \xi_{n+M+1} \in C^c, \ldots, \xi_N \in C^c)\]
does not depend on \(M\), as \(M\) varies in \(S_n(C)\).

(iii) Under the same assumptions of (ii),
\[P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in C, \ldots, \xi_{n+M} \in C, \xi_{n+M+1} \in C^c, \ldots, \xi_N \in C^c) = P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in C, \ldots, \xi_n \in C)\]
\[
\text{(A.5)}
\]
for any \(M\) in \(S_n(C)\). In other words, the random vectors \((\mathbb{I}_{\xi_i \in A_1}, \ldots, \mathbb{I}_{\xi_n \in A_n})\) \(\text{and} (\mathbb{I}_{\xi_{n+1} \in C}, \ldots, \mathbb{I}_{\xi_N \in C})\) are conditionally independent 
given \((\mathbb{I}_{\xi_1 \in C}, \ldots, \mathbb{I}_{\xi_n \in C})\).

**Proof.** We shall show that (i) is equivalent to (ii), and (ii) entails (iii).
It is trivial to prove that (iii) implies (ii).
At the beginning of this Appendix, it was already shown that, under Conditions 4.1.1–4.1.2, Condition 4.1.3 is satisfied if and only if, for each \( C \) in \( \mathcal{G} \) such that \( \tilde{N}(C) \) is not degenerate, and there exists an exchangeable vector \( Y(C) \) such that

\[
P(W_{m+1,1} = j_1, \ldots, W_{m+1,n} = j_n \mid \xi_1 \in C, \ldots, \xi_M \in C) = P'(Y_1(C) = j_1, \ldots, Y_n(C) = j_n),
\]

for any \((j_1, \ldots, j_M)\) such that \( \text{ge} \left( B_{m+1,j_i} \right) = C \) \((i = 1, \ldots, M)\) and \( P(\tilde{N}(C) = M) > 0 \), and for any \( 1 \leq n \leq M \).

Notice that, for any fixed \((j_1, \ldots, j_n)\), the right hand side of (A.6) does not depend on \((M-n)\) \((\text{for any } M \geq n \text{ such that } P(\tilde{N}(C) = M) \text{ is positive})\). Hence, if Conditions 4.1.1–4.1.3 hold, then, (ii) is also satisfied.

On the other hand, if Conditions 4.1.1–4.1.2 hold together with (ii), then, for each \( C \) in \( \mathcal{G} \), there exists a random vector \( Y(C) \) that satisfies (4.3). In fact, under (ii),

\[
P(W_{m+1,1} = j_1, \ldots, W_{m+1,n} = j_n \mid E_M)
= P(W_{m+1,1} = j_1, \ldots, W_{m+1,n} = j_n \mid E_{M^*_C})
= \sum_{j_{n+1}=1}^{k_{m+1}} \cdots \sum_{j_{M^*_C}=1}^{k_{m+1}} P(W_{m+1,1} = j_1, \ldots, W_{m+1,M^*_C} = j_{M^*_C} \mid E_{M^*_C})
= \sum_{j_{n+1}=1}^{k_{m+1}} \cdots \sum_{j_{M^*_C}=1}^{k_{m+1}} P(W_{m+1,1} = j_1, \ldots, W_{m+1,M^*_C} = j_{M^*_C} \mid \xi_1 \in C, \ldots, \xi_{M^*_C} \in C),
\]

if \( 1 \leq n \leq M \leq M^*_C \) and \( P(\tilde{N}(C) = M) > 0 \), letting

\[
M^*_C := \max \left\{ j \geq 0 : P(\tilde{N}(C) = j) > 0 \right\}.
\]

Therefore, (A.6) is satisfied if, for each \( C \) in \( \mathcal{G} \) such that \( P(\tilde{N}(C) = 0) < 1 \), i.e.,

\[
1 \leq M^*_C := \max \left\{ M \geq 0 : P(\tilde{N}(C) = M) > 0 \right\},
\]

\( Y(C) := (Y_1(C), \ldots, Y_{M^*_C}(C)) \) is such that

\[
P'(Y_1(C) = j_1, \ldots, Y_{M^*_C}(C) = j_{M^*_C})
= P(W_{m+1,1} = j_1, \ldots, W_{m+1,M^*_C} = j_{M^*_C} \mid \xi_1 \in C, \ldots, \xi_{M^*_C} \in C),
\]
for any \((j_1, \ldots, j_{M^*_C})\) such that \(\text{ge}(B_{m+1,j_i}) = C\) \((i = 1, \ldots, M^*_C)\).

Now, let us prove that (ii) implies (iii). Denote \(C^1\) to be \(C\) and \(C^0\) to be \(C^c\). Hence, by exchangeability, we can write:

\[
P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in C, \ldots, \xi_n \in C)
= \sum_{t_1, \ldots, t_{N-n}} \sum_{C^M \in \mathcal{G}} P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in C, \ldots, \xi_n \in C) \cdot P(\xi_{n+1} \in C^1, \ldots, \xi_N \in C^{N-n} \mid \xi_1 \in C, \ldots, \xi_n \in C),
\]

(A.7)

where the sum runs over all vectors \((t_1, \ldots, t_{N-n})\) in \(\{0, 1\}^{N-n}\). By (ii), (A.7) becomes:

\[
P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in C, \ldots, \xi_n \in C)
= P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid E_n(M)) \cdot \sum_{t_1, \ldots, t_{N-n}} P(\xi_{n+1} \in C^1, \ldots, \xi_N \in C^{N-n} \mid \xi_1 \in C, \ldots, \xi_n \in C),
\]

and the proof is complete. \(\square\)

**Proof of Proposition 4.1.** First, notice that by (5.3), for each \(C\) in \(\mathcal{G}\) and each descendant \(B\) of \(C\),

\[
\mathbb{E}(\tilde{N}(B) \mid \tilde{N}(C)) = \tilde{N}(C) \cdot P'(Y_1(C) = j)
\]

if \(B = B_{m,j}\), i.e.,

\[
\mathbb{E}(\tilde{N}(B) \mid \tilde{N}(C)) = \tilde{N}(C) \cdot \frac{\mathbb{E}(\tilde{N}(B))}{\mathbb{E}(\tilde{N}(C))}. \tag{A.8}
\]

In fact, by (A.5), taking \(n = 1\), we obtain

\[
P'(Y_1(C) = j) = \frac{P(\xi_1 \in B_{m,j})}{P(\xi_1 \in C)} = \frac{\mathbb{E}(\tilde{N}(B_{m,j}))}{\mathbb{E}(\tilde{N}(C))}.
\]
Federico Bassetti and Pier Giovanni Bissiri

By (A.8), for each positive \( M \) such that \( P\{\tilde{N}(C) = M\} > 0 \),

\[
\mathbb{E}\left( \frac{\tilde{N}(B)}{M} \mid \tilde{N}(C) = M \right) = \frac{\mathbb{E}(\tilde{N}(B))}{\mathbb{E}(\tilde{N}(C))}. \tag{A.9}
\]

Hence,

\[
P\{\xi_1 \in B\} = \frac{\mathbb{E}(\tilde{N}(B_m))}{N} = \frac{\mathbb{E}(\tilde{N}(B_1))}{N} \prod_{j=2}^{m} \frac{\mathbb{E}(\tilde{N}(B_j))}{\mathbb{E}(\tilde{N}(B_{j-1}))},
\]

which is equal to the right hand side of (4.4) by (A.9).

**Proposition A.2.** Let \( X = (0, 1] \), \( \mathcal{A} = \mathcal{B}([0, 1]) \), and denote \( \tilde{N}(\cdot) = \sum_{i=1}^{N} \delta_{\xi_i} \), where \( \xi_1, \ldots, \xi_N \) are the coordinate functions on \( \mathbb{R}^N \).

Assume that \( \rho \) is an exchangeable finitely additive probability on \( \mathcal{A}^N \). Then, \( \rho \) is countably additive if and only if \( \tilde{N}(I_{\varepsilon_1 \ldots \varepsilon_m}) \) goes to zero in mean (as \( m \) diverges to \( +\infty \)), for any zero-one sequence \( (\varepsilon_1, \varepsilon_2, \ldots) \) that is eventually zero, i.e., there exists \( m_0 \) such that for any \( m \geq m_0 \), \( \varepsilon_m = 0 \).

**Proof.** Let \( \mu = \rho \circ \xi_1^{-1} \). First, note that given an infinite zero-one sequence \( \varepsilon^* = (\varepsilon_1, \varepsilon_2, \ldots) \) that is eventually zero, the set \( \bigcap_{m \geq 1} I_{\varepsilon_1 \ldots \varepsilon_m} \) is empty and, therefore,

\[
\lim_{m \to \infty} \mathbb{E}(\tilde{N}(I_{\varepsilon_1 \ldots \varepsilon_m})) = \lim_{m \to \infty} N \cdot \mu(I_{\varepsilon_1 \ldots \varepsilon_m}) = 0
\]

if \( \rho \) is completely additive.

In order to prove the “if” part, it is convenient to apply a result of Sazonov (1965), since \( ([0, 1], \mathcal{B}([0, 1])) \) is a Polish space, \( \rho \) is sigma-additive if (and only if) \( \mu \) is so.

Since \( \mu \) is finitely additive, the function \( F(\cdot) = \mu(\cdot, \cdot] \) defined on the set of dyadic rationals is non-decreasing. Moreover, \( F(1) = \mu((0, 1]) = 1 \). Since the set of dyadic rationals is dense in \( \mathbb{R} \), it is sufficient to show that \( \lim_{x \to 0^+} F(x) = 0 \), and that \( F \) is continuous to the right. Let \( x \) be a dyadic rational. Hence, one can find \( m \in \mathbb{N} \) and a zero-one sequence \( (\varepsilon_1, \ldots, \varepsilon_m) \) of length \( m \) such that \( \varepsilon_m = 1 \) and \( x = \sum_{k=1}^{m-1} \varepsilon_k 2^{-k} + 2^{-m} \). Therefore if we let \( \varepsilon_k = 0 \) for \( k \geq m + 1 \)

\[
F(x + 2^{-n}) = F(x) + \mu(I_{\varepsilon_1 \ldots \varepsilon_n}) \quad \text{for } n > m. \tag{A.10}
\]
By hypothesis, \( \lim_{n \to \infty} \mu(I_{\varepsilon_1, \ldots, \varepsilon_n}) = 0 \) since \((\varepsilon_1, \varepsilon_2, \ldots)\) is eventually zero and equation (A.10) implies that \( \lim_{n \to \infty} F(x + 2^{-n}) = F(x) \) for any dyadic rational \( x \). When \( x = 0 \), one can see that \( \lim_{n \to \infty} F(2^{-n}) = 0 \) since \( F(2^{-n}) = \mu(I_{\varepsilon_1, \ldots, \varepsilon_n}) \), where \( \varepsilon_k = 0 \) for any \( k \).

**Remark A.1.** If the finitely-additive exchangeable probability \( \rho \) is defined by means of the urn scheme described in Section 5.1, then

\[
\mathbb{E}(\tilde{N}(I_{\varepsilon_1, \ldots, \varepsilon_m})) = \prod_{j=1}^{m} \frac{\alpha_{\varepsilon_1, \ldots, \varepsilon_{j+1}}}{\alpha_{\varepsilon_1, \ldots, \varepsilon_j} 0 + \alpha_{\varepsilon_1, \ldots, \varepsilon_{j+1}}}, \tag{A.11}
\]

for each zero-one vector \((\varepsilon_1, \ldots, \varepsilon_m)\). Hence, by virtue of Proposition A.2, \( \rho \) is completely additive if and only if, for any sequence \((\varepsilon_1, \varepsilon_2, \ldots)\) that is eventually zero, (A.11) goes to zero as \( m \) diverges.

**Proof of Proposition 5.1.** Equation (5.3) follows by equation (4.4). In order to prove (5.4), denote by \( d_k(x) \) the \( k \)-th binary digit of \( x \), for every \( x \in \mathbb{X} = (0, 1] \). Hence, for any \( x \in (0, 1] \),

\[
x = \sum_{m=1}^{+\infty} 2^{-m} d_m(x), \tag{A.12}
\]

and then

\[
\mathbb{E}(\xi_1) = \sum_{m=1}^{+\infty} 2^{-m} \mathbb{E}(d_m(\xi_1))
\]

\[
= \sum_{m=1}^{+\infty} 2^{-m} \sum_{\varepsilon \in E^{m-1}} P(\xi_1 \in I_{\varepsilon_1})
\]

\[
= \sum_{m=1}^{+\infty} 2^{-m} \sum_{\varepsilon \in E^{m-1}, \varepsilon_m = 1} \prod_{k=1}^{m} \frac{\alpha_{\varepsilon_1, \ldots, \varepsilon_k}}{\alpha_{\varepsilon_1, \ldots, \varepsilon_k} 0 + \alpha_{\varepsilon_1, \ldots, \varepsilon_k} 1}. \tag*{\Box}
\]

**Proof of Proposition 6.1.** Assume, as usual, that \( \xi_1, \ldots, \xi_N \) are the coordinate functions on \((\mathbb{X}^N, \mathcal{F}^N, P)\) so that \( P = \mathcal{L}(\xi_1, \ldots, \xi_N) \). In order to show that \( P \) is a partitions tree distribution, we need to find \( \mathcal{L}(\tilde{N}_{m+1} | \tilde{N}_m) \) for each \( m \). Let \((N_1, \ldots, N_{k_m})\) be a vector of nonnegative integers summing
up to \( N \) and \( i_1, \ldots, i_N \) are integers in \( \{1, \ldots, k_m\} \) such that \( N_j = |\{l = 1, \ldots, k_m : i_l = j\}| \). We can write:

\[
P\{\xi_1 \in B_{m,i_1}, \ldots, \xi_N \in B_{m,i_N}\} = \mathbb{E}\left[ \prod_{j=1}^{k_m} \tilde{p}(B_{m,j})^{N_j} \right]
\]

\[
= \mathbb{E}\left[ \prod_{j=1}^{k_m} \tilde{p}(B_{m,j} \mid \text{ge } (B_{m,j}))^{N_j} \prod_{j=1}^{k_m} \tilde{p}(\text{ge } (B_{m,j}))^{N_j} \right]
\]

by (P1)

\[
= \mathbb{E}\left( \prod_{j \in \mathcal{T}(C)} \tilde{p}(B_{m,j} \mid C)^{N_j} \right) \mathbb{E}\left( \prod_{C \in \pi_{m-1}} \tilde{p}(C)^{\Sigma_{j \in \mathcal{T}(C)} N_j} \right). \tag{A.13}
\]

Observing that in the last term of (A.13), the first expectation is the \((N_j : j \in \mathcal{T}(C))\)-th mixed moment of the singular Dirichlet distribution with parameters \((\alpha_{m,j} : j \in \mathcal{T}(C))\), we obtain:

\[
P\{\xi_1 \in B_{m,i_1}, \ldots, \xi_N \in B_{m,i_N}\} = P\{\xi_1 \in \text{ge } (B_{m,i_1}), \ldots, \xi_N \in \text{ge } (B_{m,i_N})\}
\]

\[
\cdot \prod_{C \in \pi_{m-1}} \left( \frac{1}{\left( \sum_{\alpha_{m,j}}^{N_j} \right)^{\Sigma_{j \in \mathcal{T}(C)} N_j}} \right), \tag{A.14}
\]

where \( a^{[b]} := a(a + 1) \ldots (a + b - 1) \).

Note that in general,

\[
P(\tilde{N}(B_{m,1}) = N_1, \ldots, \tilde{N}(B_{m,k_m}) = N_{k_m} \mid \tilde{N}(C) = \sum_{j \in \mathcal{T}(C)} N_j : C \in \pi_{m-1})
\]

\[
= \frac{P(\tilde{N}(B_{m,1}) = N_1, \ldots, \tilde{N}(B_{m,k_m}) = N_{k_m})}{P(\tilde{N}(C) = \sum_{j \in \mathcal{T}(C)} N_j : C \in \pi_{m-1})}
\]

\[
= \left( \frac{N}{N_1, \ldots, N_{k_m}} \right) P\{\xi_1 \in B_{m,i_1}, \ldots, \xi_N \in B_{m,i_N}\}
\]

\[
\cdot \left( \frac{1}{\sum_{j \in \mathcal{T}(C)} N_j : C \in \pi_{m-1}} \right)^{\Sigma_{j \in \mathcal{T}(C)} N_j} P\{\xi_1 \in \text{ge } (B_{m,i_1}), \ldots, \xi_N \in \text{ge } (B_{m,i_N})\}
\]

\[
= \frac{P\{\xi_1 \in B_{m,i_1}, \ldots, \xi_N \in B_{m,i_N}\}}{P\{\xi_1 \in \text{ge } (B_{m,i_1}), \ldots, \xi_N \in \text{ge } (B_{m,i_N})\}} \prod_{C \in \pi_{m-1}} \left( \frac{\sum_{j \in \mathcal{T}(C)} N_j}{N_j : j \in \mathcal{T}(C)} \right), \tag{A.15}
\]
and, therefore, combining (A.14) and (A.15) one realizes that the conditional law of \( \tilde{N}_m \) given \( \tilde{N}_{m-1} \) can be written as a product of measures:

\[
L_{\tilde{N}_m \mid \tilde{N}_{m-1}} = \prod_{C \in \pi_{m-1}} L_(\tilde{N}(B) \mid \text{ge}(B) = C) | \tilde{N}(C),
\]

where each factor is given by (6.1). Hence, Conditions 4.1.1–4.1.2 hold good.

In order to prove Condition 4.1.3, it is sufficient to verify (4.1) for any \( C \) in \( \mathcal{G} \) such that \( M^* \) is positive (the opposite case is trivial). By (6.1), for each positive \( M \) such that \( P(\tilde{N}(C) = M) \) is positive, \( \psi_C(M, \cdot) \) is the probability mass function of a mixture of multinomial distributions, where the mixing distribution is Dirichlet with parameter \( (M; \alpha_{m,j} : B_{m,j} \subset C) \) (see Johnson et al. (1997)).

At this stage, let \( (X_1, \ldots, X_{M^*_C}) \) be a \( \mathcal{T}(C)^N \) valued (exchangeable) vector such that

\[
L_{X_1, \ldots, X_{M^*_C}}((r_1, \ldots, r_{M^*_C})) = \psi_C(M; N_1 \ldots N_{h_C})
\]

\[
= \int_{[0,1]^{h_C-1}} \prod_{j=1}^{h_C} \theta_1^{N_j} \ldots \theta_{h_C}^{N_{h_C}} \text{Dir}_C(\alpha_{m,j} : j \in \mathcal{T}(C)),
\]

where \( N_j = |\{i = 1, \ldots, M^*_C : r_i = j\}| \) with \( j = 1, \ldots, h_C \), and \( \text{Dir}_C(\beta_1, \ldots, \beta_h) \) denotes the density of the Dirichlet distribution on \([0,1]^{h-1}\) with parameter \( (\beta_1, \ldots, \beta_h) \).

Hence, the law of \( (X_1, \ldots, X_{M^*_C}) \) is a mixture of distributions of finite i.i.d. sequences, and, for each \( M \leq M^*_C \), also the law of \( (X_1, \ldots, X_M) \) must be so (with the same mixing measure). Recalling the finite version of de Finetti’s theorem, this fact can be formulated by (4.3), which, therefore, must be satisfied. \( \square \)

**Lemma A.1.** If Conditions 4.1.1–4.1.2 hold, then, for any \( m \geq 0 \) and any \( N \)-tuple \( (A_1, \ldots, A_N) \) of events in \( \pi_{m+1} \),

\[
P(\xi_1 \in A_1, \ldots, \xi_N \in A_N \mid \xi_1 \in \text{ge}(A_1), \ldots, \xi_N \in \text{ge}(A_N))
= \prod_{C \in \pi_m : N_C > 0} P(\xi_1 \in A_{l_1(C)}, \ldots, \xi_{N_C} \in A_{l_{N_C}(C)} \mid \xi_1 \in C, \ldots, \xi_{N_C} \in C),
\]

\[
\xi_{N_C+1} \in C^c, \ldots, \xi_N \in C^c), \quad (A.16)
\]
where $N_C = |\{i = 1, \ldots, N : A_i \subset C\}|$, $l_1 < l_2 < \cdots < l_{N_C}$, and
$A_{l_1}(C), \ldots, A_{l_{N_C}}(C)$ are descendants of $C$ for each $C$ in $\pi_m$. In other words, for any $m$,

(i) if $(C_1, \ldots, C_N)$ is any $N$-tuple of sets in $\pi_m$, the random collections
$\{W_{m+1,i}, \ i : C_i = C\}$ (as $C$ varies in $\pi_m$) are conditionally independent given the event $\{\xi_1 \in C_1, \ldots, \xi_N \in C_N\}$; and

(ii) for any $C$ in $\pi_{m+1}$ and any for any $n \leq N$, $(W_{m+1,1}, \ldots, W_{m+1,n})$
and $(W_{m,n+1}, \ldots, W_{m,N})$ are conditionally independent given the event
$\{\xi_1 \in C, \ldots, \xi_n \in C\}$.

**Proof.** For each $C \in \mathcal{G}$, let $h_C = |\{B \in \mathcal{G} : \text{ge}(B) = C\}|$ and denote
by $D_1(C), \ldots, D_{h_C}(C)$ the descendants of $C$. Finally, set $N_j(C) = |\{i = 1, \ldots, n : A_i = D_j(C)\}|$ ($j = 1, \ldots, h_C$). Hence,

$$P(\xi_1 \in A_1, \ldots, \xi_N \in A_N \mid \xi_1 \in \text{ge}(A_1), \ldots, \xi_N \in \text{ge}(A_N))$$

$$= \frac{P(\xi_1 \in A_1, \ldots, \xi_N \in A_N)}{P(\xi_1 \in \text{ge}(A_1), \ldots, \xi_N \in \text{ge}(A_N))}$$

$$= \frac{P(\tilde{N}(D_j(C)) = N_j(C), C \in \pi_m, j = 1, \ldots, h_C) \prod_{C \in \pi_m : N_C > 0} \left(\frac{N_C}{N_j(C) \ldots h_C(C)}\right)}{P(\tilde{N}(D_j(C)) = N_j(C), C \in \pi_m, j = 1, \ldots, h_C) \prod_{C \in \pi_m : N_C > 0} \left(\frac{N_C}{N_j(C) \ldots h_C(C)}\right)}$$

$$= P(\tilde{N}(D_j(C)) = N_j(C), C \in \pi_m, j \leq h_C \mid \tilde{N}(C))$$

$$= N_C, C \in \pi_m / \prod_{C \in \pi_m : N_C > 0} \left(\frac{N_C}{N_j(C) \ldots h_C(C)}\right),$$

which, by virtue of (4.2) and (A.1), becomes

$$\prod_{C \in \pi_m : N_C > 0} P(\tilde{N}(D_j(C)) = N_j(C), j = 1, \ldots, h_C \mid \tilde{N}(C) = N_C)$$

$$= \prod_{C \in \pi_m : N_C > 0} P(\xi_1 \in A_{l_1}(C), \ldots, \xi_{N_C} \in A_{l_{N_C}}(C) \mid \xi_1 \in C, \ldots, \xi_{N_C} \in C, \xi_{N_C+1} \in C^c, \ldots, \xi_N \in C^c)$$

as desired. \qed
From now on, denote
\[ \text{ge}^{(0)}(\cdot) = \cdot, \quad \text{ge}^{(1)}(\cdot) = \text{ge}(\cdot), \quad \text{ge}^{(m+1)}(\cdot) = \text{ge}(\text{ge}^{(m)}(\cdot)) \quad (m \geq 1). \]

**Lemma A.2.** If Condition 4.1 holds, then, for any pair \((m, h)\) of positive integers and any \(N\)-tuple \((A_1, \ldots, A_N)\) of events in \(\pi_{m+h}\),
\[
\begin{align*}
P(\xi_1 \in A_1, \ldots, \xi_N \in A_N | \xi_1 \in \text{ge}^{(h)}(A_1), \ldots, \xi_N \in \text{ge}^{(h)}(A_N)) &= \prod_{C \in \pi_{m:N_C>0}} P(\xi_1 \in A_{l^{(h)}_1(C)}, \ldots, \xi_{N_C} \in A_{l^{(h)}_{N_C}(C)} | \xi_1 \in C, \ldots, \xi_{N_C} \in C),
\end{align*}
\]
where \(N_C = |\{i = 1, \ldots, N : A_i \subset C\}|, \ l^{(h)}_1 < l^{(h)}_2 < \cdots < l^{(h)}_{N_C}\), and \(A_{l^{(h)}_i(C)} \subset C\) for each \(i \leq N\) and each \(C\) in \(\pi_m\).

In other words,

(i) if \((C_1, \ldots, C_N)\) is any \(N\)-tuple of sets in \(\pi_m\), the random collections
\[ \{W_{m+h,i}, i : C_i = C\} \quad (C \text{ varies in } \pi_m) \]
are conditionally independent given the event \(\{\xi_1 \in C_1, \ldots, \xi_N \in C_N\}\); and

(ii) if \(C\) belongs to \(\pi_{m+h}\) and \(n \leq N\), then \((W_{m+h,1}, \ldots, W_{m+h,n})\) and \((W_{m,n+1}, \ldots, W_{m,N})\) are conditionally independent given the event \(\{\xi_1 \in C, \ldots, \xi_n \in C\}\).

**Proof.** The proof will be done by induction w.r.t. \(h\). By Lemma A.1 and Proposition A.1(iii), the thesis holds for \(h = 1\). Assume that
\[
P(\xi_1 \in A_1, \ldots, \xi_N \in A_N | \xi_1 \in \text{ge}^{(h-1)}(A_1), \ldots, \xi_N \in \text{ge}^{(h-1)}(A_N))
\]
\[
= \prod_{D \in \pi_{m:N_D>0}} P(\xi_1 \in A_{l^{(h-1)}_1(D)}, \ldots, \xi_{N_D} \in A_{l^{(h-1)}_{N_D}(D)} | \xi_1 \in D, \ldots, \xi_{N_D} \in D),
\]
for any \(m\) and any \(N\)-tuple \((A_1, \ldots, A_N)\) of sets in \(\pi_{m+h-1}\). Hence, by Lemma A.1 and the induction hypothesis (A.18) the left hand side of (A.17)
If Condition 4.1 holds, then, for any $N$ any $2$ which is equal to the right hand side of (A.17).

The last expression, by Lemma A.1, becomes

$$
\begin{align*}
& P(\xi_1 \in A_1, \ldots, \xi_N \in A_N \mid \xi_1 \in \text{ge}(A_1), \ldots, \xi_N \in \text{ge}(A_N)) \\
& \quad \cdot P(\xi_1 \in \text{ge}(A_1), \ldots, \xi_N \in \text{ge}(A_N) \mid \xi_1 \in \text{ge}(h)(A_1), \ldots, \xi_N \in \text{ge}(h)(A_N)) \\
& = \prod_{C \in \pi_{m+h-1}: N_C > 0} P\left(\xi_1 \in A_{I_{1}(C)}, \ldots, \xi_{N_C} \in A_{I_{N_C}(C)} \mid \xi_1 \in C, \ldots, \xi_{N_C} \in C, \xi_{N_C+1} \in C^c, \ldots, \xi_N \in C^c\right) \\
& \quad \cdot \prod_{D \in \pi_m} P(\xi_1 \in \text{ge}\left(\mathcal{A}_{(h)}(D)\right)), \ldots, \xi_{N_D} \in \text{ge}\left(\mathcal{A}_{(h)}(D)\right) \mid \xi_1 \in D, \ldots, \xi_{N_D} \in D),
\end{align*}
$$

(A.19)

which, by (iii) of Proposition A.1, turns out to be

$$
\begin{align*}
\prod_{D \in \pi_m} & \left(P\left(\xi_1 \in \text{ge}\left(\mathcal{A}_{(h)}(D)\right), \ldots, \xi_{N_D} \in \text{ge}\left(\mathcal{A}_{(h)}(D)\right) \mid \xi_1 \in D, \ldots, \xi_{N_D} \in D\right) \right. \\
& \left. \cdot \prod_{C \in \pi_{m+h-1}: N_C > 0, C \subseteq D} P(\xi_1 \in A_{I_{1}(C)}, \ldots, \xi_{N_C} \in A_{I_{N_C}(C)} \mid \xi_1 \in C, \ldots, \xi_{N_C} \in C)\right).
\end{align*}
$$

(A.20)

The last expression, by Lemma A.1, becomes

$$
\begin{align*}
& \prod_{D \in \pi_m} \left(P\left(\xi_1 \in \text{ge}\left(\mathcal{A}_{(h)}(D)\right), \ldots, \xi_{N_D} \in \text{ge}\left(\mathcal{A}_{(h)}(D)\right) \mid \xi_1 \in D, \ldots, \xi_{N_D} \in D\right) \right. \\
& \left. \cdot P\left(\xi_1 \in A_{I_{1}(D)}, \ldots, \xi_{N_D} \in A_{I_{N_D}(D)} \mid \xi_1 \in \text{ge}\left(\mathcal{A}_{(h)}(D)\right), \ldots, \xi_{N_C} \in \text{ge}\left(\mathcal{A}_{(h)}(D)\right)\right)\right),
\end{align*}
$$

which is equal to the right hand side of (A.17). \hfill \Box

Lemma A.2 yields the following.

**Lemma A.3.** If Condition 4.1 holds, then, for any $m \geq 0$, any $h \geq 0$, any $N$-tuple $(A_1, \ldots, A_N)$ of events in $\pi_{m+h}$ and any $n \leq N$,

$$
P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in \text{ge}(h)(A_1), \ldots, \xi_N \in \text{ge}(h)(A_N))
\quad = \quad P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in \text{ge}(h)(A_1), \ldots, \xi_n \in \text{ge}(h)(A_n)).
$$

(A.21)
Proof. Notice that by (i) of Lemma A.2

\[ P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in \text{ge}(h)(A_1), \ldots, \xi_N \in \text{ge}(h)(A_N)) \]

\[ = \prod_{C \in \pi_m : n_C > 0} P(\xi_{1(C)} \in A_{1(C)}, \ldots, \xi_{n(C)} \in A_{n(C)} \mid \xi_1 \in \text{ge}(h)(A_1), \ldots, \xi_N \in \text{ge}(h)(A_N)), \]  

(A.22)

where \( n_C = |\{i = 1, \ldots, n : A_i \subset C\}|, l_1 < l_2 < \cdots < l_{n_C}, \) and \( A_{l_i(C)}, \ldots, A_{l_{n_C}(C)} \) are subsets of \( C \) for each \( C \) in \( \pi_m \). Hence, by (ii) of Lemma A.2,

\[ P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in \text{ge}(h)(A_1), \ldots, \xi_N \in \text{ge}(h)(A_N)) \]

\[ = \prod_{C \in \pi_m : n_C > 0} P(\xi_{1(C)} \in A_{1(C)}, \ldots, \xi_{n(C)} \in A_{n(C)} \mid \xi_1 \in C, \ldots, \xi_N \in C), \]

which, applying again (i) and (ii) of Lemma A.2, is equal to the right hand side of (A.21). \( \square \)

Lemma A.4. Assume that Conditions 4.1.1–4.1.2 are satisfied. Let \( A_1, \ldots, A_N \) be \( N \) sets in \( \pi_{m+1} \), and denote

\[ N_j = |\{i = 1, \ldots, N : A_i = B_{m+1,j}\}| \quad (j = 1, \ldots, k_{m+1}). \]

If \( P(\tilde{N}(B_{m+1,1}) = N_1, \ldots, \tilde{N}(B_{m+1,k_{m+1}}) = N_{k_{m}}) > 0, \) then, for each \( n \leq N, \)

\[ P(\tilde{N}(B_{m+1,1}) = N_1, \ldots, \tilde{N}(B_{m+1,k_{m+1}}) = N_{k_{m+1}} \mid \xi_1 \in A_1, \ldots, \xi_n \in A_n, \]

\[ \tilde{N}(C) = N_C : C \in \pi_m \]

\[ = \prod_{C \in \pi_m} P(\tilde{N}(B_{m+1,j}) = N_j : j \in T(C) \mid \tilde{N}(C) = N_C, \]

\[ \xi_i \in A_i : i \leq n, \text{ge}(A_i) = C), \]

(A.23)

where \( N_C := \sum_{j \in T(C)} N_j \) for each \( C \in \pi_m. \)
PROOF. Notice that

\[
P(\xi_1 \in A_1, \ldots, \xi_N \in A_N \mid \tilde{N}(C) = N_C : C \in \pi_m)
= P\left(\tilde{N}(B_{m+1,1}) = N_1, \ldots, \tilde{N}(B_{m+1,k_m+1}) = N_{k_m+1} \mid \tilde{N}(C) = N_C : C \in \pi_m\right) \prod_{C \in \pi_m} \left(\frac{N_C}{N_j : j \in \mathcal{T}(C)}\right), \tag{A.24}
\]

which, by hypothesis, becomes

\[
\prod_{C \in \pi_m} \left[ P(\tilde{N}(B_{m+1,j}) = N_j : j \in \mathcal{T}(C) \mid \tilde{N}(C) = N_C) \right]
= \prod_{C \in \pi_m} P(\xi_i \in A_i : i \leq N, \text{ge}(A_i) = C \mid \tilde{N}(C) = N_C). \tag{A.25}
\]

Combining equations (A.24) and (A.25), one obtains

\[
P(\xi_1 \in A_1, \ldots, \xi_N \in A_N \mid \tilde{N}(C) = N_C : C \in \pi_m) = \prod_{C \in \pi_m} P(\xi_i \in A_i : i \leq N, \text{ge}(A_i) = C \mid \tilde{N}(C) = N_C). \tag{A.26}
\]

At this stage, denote

\[
n_j := \left|\{i = 1, \ldots, n : A_i = B_{m+1,j}\}\right| \quad (j = 1, \ldots, k_m+1), \text{ and}
\]

\[
n_C := \sum_{j \in \mathcal{T}(C)} n_j \quad \text{(for each } C \text{ in } \pi_m).\]

Hence, the left hand side of (A.23) is equal to

\[
P(\xi_{n+1} \in A_{n+1}, \ldots, \xi_N \in A_N \mid \xi_1 \in A_1, \ldots, \xi_n \in A_n, \tilde{N}(C) = N_C : C \in \pi_m) \prod_{C \in \pi_m} \left(\frac{N_C - n_C}{N_j - n_j : j \in \mathcal{T}(C)}\right). \tag{A.27}
\]
The conditional probability in (A.27) can be rewritten as

\[
P(\xi_1 \in A_1, \ldots, \xi_N \in A_N \mid \tilde{N}(C) = N_C : C \in \pi_m)
\]

and, by (A.26), it becomes

\[
\prod_{C \in \pi_m} \frac{P(\xi_i \in A_i : i \leq N, \text{ge}(A_i) = C \mid \tilde{N}(C) = N_C)}{P(\xi_i \in A_i : i \leq n, \text{ge}(A_i) = C, \mid \tilde{N}(C) = N_C)}
\]

\[
= \prod_{C \in \pi_m} P(\xi_i \in A_i : i \leq N, A_i \subset C \mid \tilde{N}(C) = N_C, \xi_i \in A_i : i \leq n, A_i \subset C).
\]

(A.28)

If one substitutes the conditional probability in (A.27) with the right hand side of (A.28), then the right hand side of (A.23) is obtained as desired. □

**Proof of Proposition 7.1.** By virtue of Lemma A.3, one realizes that, for each \( n \leq N \) and each pair \((h, m)\) of positive integers, the random vectors \((W_{m+h,1}, \ldots, W_{m+h,n})\) and \((W_{m,n+1}, \ldots, W_{m,N})\) are conditionally independent given \((W_{m,1}, \ldots, W_{m,n})\). Since \( W_{r,i} \) is a function of \( W_{A,i} \) for each \( 1 \leq i \leq N \) and each pair \((r, s)\) with \( r < s \), one can assert that for each \( 1 \leq n \leq N \) and each vector \((m, l_1, \ldots, l_n)\) of positive integers such that \( l_i \geq m \) for \( i = 1, \ldots, n \), the random vectors \((W_{l_1,1}, \ldots, W_{l_n,n})\) and \((W_{m,n+1}, \ldots, W_{m,N})\) are conditionally independent given \((W_{m,1}, \ldots, W_{m,n})\), i.e.,

\[
\mathcal{L}(W_{m,n+1}, \ldots, W_{m,N})|(W_{l_1,1}, \ldots, W_{l_n,n}) = \mathcal{L}(W_{m,n+1}, \ldots, W_{m,N})|(W_{m,1}, \ldots, W_{m,n}) . \quad (A.29)
\]

Since \( \tilde{N}_m \) is a function of \((W_{m,1}, \ldots, W_{m,N})\), (A.29) implies

\[
\mathcal{L}(\tilde{N}_m)(W_{l_1,1}, \ldots, W_{l_n,n}) = \mathcal{L}(\tilde{N}_m)(W_{m,1}, \ldots, W_{m,n}) . \quad (A.30)
\]

Notice that:

\[
\sigma(\xi(n)) = \sigma(\cup_{l_1, \ldots, l_n} \sigma(W_{l_1,1}, \ldots, W_{l_n,n})). \quad (A.31)
\]

By (A.30), for each \((l_1, \ldots, l_n)\) and each \( D \in \sigma(W_{l_1,1}, \ldots, W_{l_n,n})\),

\[
\mathbb{E} \left[ I_D \mathbb{E} \left[ I\{\tilde{N}_m = (N_1, \ldots, N_{km})\} | W_{m,1}, \ldots, W_{m,n} \right] \right] \\
= \mathbb{E} \left[ I_D \mathbb{E} \left[ I\{\tilde{N}_m = (N_1, \ldots, N_{km})\} | W_{m,l_1}, \ldots, W_{m,l_n} \right] \right] \\
= P(\{\tilde{N}_m = (N_1, \ldots, N_{km})\} \cap D),
\]
and therefore, by (A.31), $\mathcal{L}_{\hat{N}_m|\xi_1,\ldots,\xi_n} = \mathcal{L}_{\hat{N}_m|W_{m,1},\ldots,W_{m,n}}$, which the thesis follows from.

Proof of Proposition 7.3. By combining Proposition 7.1 and Lemma A.4, the proof follows.

Acknowledgments. The authors are grateful to Eugenio Regazzini for providing much of the inspiration behind this paper and helping to improve an earlier version of the paper. They also thank R.V. Ramamoorthi for his encouragement and an anonymous referee and a Co-editor for their helpful comments and suggestions.

References


Federico Bassetti
Università degli Studi di Pavia
Dipartimento di Matematica
Via Ferrata 1
27100 Pavia, Italy
E-mail: federico.bassetti@unipv.it

Pier Giovanni Bissiri
Università degli Studi di Cagliari
Dipartimento di Matematica ed Informatica
Via Ospedale 72, 09124 Cagliari, Italy
E-mail: bissiri@unica.it

Paper received January 2007; revised August 2007.