Università degli Studi di Pavia
Dottorato di Ricerca in Matematica e Statistica

Statistical issues connected with finitary exchangeable sequences

Relatore: Chiar.mo Prof. Eugenio Regazzini

Tesi di dottorato
Pier Giovanni Bissiri

Anno Accademico 2006 – 07

Contents

1 Introduction **7**

1.1 Inferences based on N–exchangeable observations . . . 9

1.2 Finitary approaches in statistical literature . . . 12

1.3 Peculiarities and outline of the present work . . . 18

2 Pólya-trees and species sampling sequences **21**

2.1 Pólya-tree distributions . . . 21

2.1.1 A very special case: the Dirichlet process . . . 23

2.1.2 Marginal, posterior, and predictive distributions . . . 24

2.1.3 Main differences between the Dirichlet process and the more general family of Pólya-tree processes . . . 25

2.1.4 Pólya-tree processes defined on the unit interval . . . 26

2.2 Species sampling sequences . . . 28

2.3 Normalized random measures with independent increments . . . 30

3 Representations for N-exchangeable sequences **33**

3.1 De Finetti's theorem and exchangeable prolongable sequences . . . 33

3.2 The law of the empirical measure as prior distribution . . . 35

3.2.1 Representation of a finite exchangeable sequence through the law of the empirical measure . . . 35

3.2.2 Representation of a finite exchangeable sequence through the finite-dimensional distributions of the empirical measure. 38
3.2.3 Constructing the law of a finite exchangeable sequences on a Polish space 43
3.3 Characterization through predictive distributions 46

4 Exchangeable laws based on partition trees 49
4.1 Definition of the model 51
4.2 Posterior and predictive distributions 54
4.3 Partition tree laws and absolute continuity 56
4.4 Construction of a partitions tree law 58
4.5 A couple of partitions tree distributions 61
4.5.1 Hypergeometric partitions tree distributions 61
4.5.2 Exchangeable sequences directed by Pólya-tree processes 68

5 Exchangeable laws based on random partitions 71
5.1 Introductory examples 71
5.2 Definition of the model 73
5.2.1 Marginal distribution and correlation between observations 77
5.3 Posterior and predictive distributions 79
5.3.1 Predictive distributions 80

6 Some applications 83
6.1 Decision theoretic formulation 83
6.2 Estimation of the mean 84
6.2.1 Random Partition model 84
6.2.2 Hypergeometric partitions tree distributions 87
6.3 Regression problems 90
6.3.1 First approach 91
6.3.2 Second approach 91
6.3.3 A numerical example 92

A Proofs **95**

A.1 Proof of Theorem 3.4 95

A.2 Properties of partitions tree distributions 100

A.2.1 Proof of Proposition 4.1 and Proposition 4.2 103

A.2.2 Some further properties of a partitions tree law 112

A.3 Random partitions 119

Chapter 1

Introduction

Statistics deals with logical foundations of inference from observed realizations of a given phenomenon to (yet) unobserved values of the same phenomenon, or to unknown parameters appearing in the probability law of the observation process. According to a classical view of statistical inference, observations are often assumed to play a symmetric role with respect to *prevision*, in the sense that, for all previsional purposes, the "chronological" order of observations is deemed irrelevant. In the frequentistic framework, the above idea of symmetry is translated into the hypothesis that observations are thought of as independent random variables, with the same distribution affected by unknown parameters. In the Bayesian theory, the same idea is captured by assuming that observations are *conditionally* independent and identically distributed, given unknown parameters. This way of behaving is inspired by the thought that what is unknown – parameters, in the present case – must be equipped with a probability distribution and, therefore, considered as a random element.

This approach is powerfully criticized by de Finetti [see for instance de Finetti (1937a, 1974)], who persuasively argues that one need not assume the existence of such things. He considers a simple example: the well-known Bayes-Laplace scheme, where observations are 0-1 random variables, the parameter $\tilde{\theta}$ is the "unknown probability" that the single observation is equal to one, and $\tilde{\theta}$ is uniformly distributed in the unit interval. To get an approximate but more realistic idea, one may consider the sequence of drawings without replacement from an urn chosen at random among $1,000,001$ urns, such that each one of them contains one million balls and the number of white balls is 0 in the first urn, 1 in the second urn, $\dots$, one million in the last urn, and the probability of each urn to be chosen is the same, i.e. $1/1,000,001$. The same probability distribution can be generated by the Pólya's urn scheme (of "contagious" probabilities): the urn initially contains two balls, one white and one black, and, after each drawing, the ball drawn is placed back in the urn together with another one of the same color. In the former urn scheme, the parameter $\tilde{\theta}$ is a factual, but unknown quantity: one could check its value if it were not forbidden to inspect the content of the urn. In the latter urn scheme, $\tilde{\theta}$ "is a merely fictitious, or 'mythical', pseudo-entity". De Finetti asserts that it is difficult to present $\tilde{\theta}$ as the unknown proportion of white balls in a "hidden urn". In fact, such proportion should be equal to the limit of the composition of the Pólya urn considered above as the number of drawings diverges, and this does not make sense since

> "*not even the Vestals would assure the continuation of such experiment for the eternity*", which would imply, incidentally, to get "*more balls than atoms in the world*", and, on the other hand, "*there is no reason to expect such limit to exist, since 'stochastic' (even if strong) convergence does not guarantee any conclusion on this point*".

As a consequence of de Finetti's argument, it is clear that one should acknowledge at least the theoretical possibility of experimentally verifying whether hypotheses about unknown entities are true or false. We will call *empirical* any hypothesis having this property. Only confining statistical inference to consider objective hypotheses (on observable elements), the phrase "to learn from experience" may have a real meaning, and it is possible to approach the problem of induction, as defined, for instance, by Hume (1748).

Bayesian statisticians very often ignore this precaution, they adopt the above hypothesis of conditional independence and indiscriminately draw inferences from observations both to empirical and to non empirical hypotheses. Diaconis (1988) is very clear on this point:

> *de Finetti's alarm to statisticians introducing realms of unobservable parameters have been repeatedly justified in the modern curve fitting exercises of today's big models. These seem to lose all contacts with scientific reality focusing attention on details of large programs and fitting instead of observation and understanding*

of basic mechanism. It is to be hoped that a fresh implementation of de Finetti's program based on observables will lead us out of this mess.

Without assuming the existence of unobservable parameters, the previous conditional formulation does not have a clear interpretation anymore. For this reason, it is natural to describe the symmetry between observations resorting, instead, to the notion of *exchangeability*, as used and studied by Bruno de Finetti. As a matter of fact, if the observation process is assumed to be *infinitely extensible*, the two formulations are equivalent, in view of the well-known de Finetti's representation theorem. On the other hand, in many situations, as sampling from a finite population, such assumption of infinite extensibility of the observation process need not be consistent with the real situation under study. In the latter case, one might be forced to construct probability laws for the observations without resorting to the usual conditional formulation. Thus, our initial problem boils down to the one of finding *alternative* methods to define laws for any kind of N*-exchangeable* or *infinite exchangeable* sequences.

It must be emphasized that the assessment of an exchangeable law, without resorting to the standard representation, forces to revise subjects and purposes of Bayesian statistical inferences and, consequently, to get the Bayesian statistical procedures to adapt to these new subjects and purposes.

Considering all the above remarks, we intend (a) to present specific forms of exchangeable laws defined, aside from the standard conditional formulation, according to the characteristics of actual situations and (b) to work out some of their inherent statistical problems.

1.1 Inferences based on N–exchangeable observations

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

Suppose that each observation takes value in some measurable space $(\mathbb{X}, \mathscr{X})$. Write $\mathbb{X}^N$ for the N-fold product $\mathbb{X} \times \cdots \times \mathbb{X}$ and $\mathbb{X}^\infty$ for $\mathbb{X} \times \mathbb{X} \times \cdots$, and indicate by $\mathscr{X}^N$ and $\mathscr{X}^\infty$ the usual product σ-fields on $\mathbb{X}^N$ and $\mathbb{X}^\infty$, respectively. In the usual Bayesian framework, the observation process is assumed to be extendible to infinity, that is, observations ξ_i are viewed as coordinates of a random element of $(\mathbb{X}^\infty, \mathscr{X}^\infty)$.

Denote by $\mathbb{P} = \mathbb{P}(\mathbb{X})$ the space of all probability measures on $(\mathbb{X}, \mathscr{X})$, and by $\mathscr{P}$ the σ-field induced by all evaluation maps $m_B : \mu \mapsto \mu(B)$, $\mu \in \mathbb{P}$ and $B \in \mathscr{X}$. Any random element from a probability space to $(\mathbb{P}, \mathscr{P})$ is said to be a *random probability measure* on $(\mathbb{X}, \mathscr{X})$.

Define the *empirical distribution* of

$$\xi(n) := (\xi_1, \dots, \xi_n),$$

$n = 1, 2, \dots$, to be 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 ξ_i belongs to A or not.

The random elements ξ_is 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 σ of $(1, 2, \dots)$.

Provided that $(\mathbb{X}, \mathscr{X})$ is a "nice" space (e.g., separable and complete metric space) and the ξ_ns form an infinite sequence of exchangeable elements, then $(\tilde{e}_n)_{n\geq 1}$ converges in distribution to a random probability $\tilde{p}$, with probability one. Moreover, these very same ξ_ns turn out to be conditionally independent given $\tilde{p}$, with the same distribution $\tilde{p}$. This is the statement of the celebrated representation theorem for infinite exchangeable sequences, provided in de Finetti (1930, 1937a) [see also Aldous (1985)].

At this stage, let us specify some further preliminary notation. Given any random variable V, $\mathfrak{L}_V$ will denote its probability distribution; moreover, for any other random element U, $\mathfrak{L}_{V|U}$ will stand for a conditional probability distribution for V given U. $U \overset{\mathfrak{L}}{=} V$ will be sometimes written in place of $\mathfrak{L}_U = \mathfrak{L}_V$.

In this notation, de Finetti's representation theorem can be enunciated in this way: *If $(\xi_i)_i$ is exchangeable, then there exists a random probability $\tilde{p}$ on $(\mathbb{X}, \mathscr{X})$ such that:*

$$\mathfrak{L}_{\xi(n)|\tilde{p}}(A) = \tilde{p}^{(n)}(A) \qquad (A \in \mathscr{X}^n,\ n = 1, 2, \dots), \tag{1.1}$$

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

According to the terminology introduced by Aldous (1985), the random probability $\tilde{p}$ is called the *directing measure of* $(\xi_i)_{i\geq 1}$.

Another equivalent version of de Finetti's representation theorem is given by the following assertion:

$(\xi_i)_{i\geq 1}$ *is exchangeable if and only if there is a probability measure* γ *on* $(\mathbb{P}, \mathscr{P})$ *such that*

$$\mathcal{L}_{\xi(n)}(A) = \int_{\mathbb{P}} p^{(n)}(A)\gamma(\mathrm{d}p) \qquad (A \in \mathscr{X}^n,\ n = 1, 2, \dots). \tag{1.2}$$

γ*, the so-called de Finetti's measure of* $(\xi_i)_{i\geq 1}$*, is uniquely determined and coincides with the distribution of* $\tilde{p}$*.*

In this general formulation, $\tilde{p}$ takes the place of the usual unknown parameter; in this case it is the custom to speak of *Bayesian nonparametric representation* and, consequently, of *Bayesian nonparametric methods*. The usual parametric formulation can be recovered by requiring that the ξ_ns must be conditionally independent given some random element θ with the following form:

$$\tilde{\theta} := t(\tilde{p}),$$

where t is a function defined on a subset $\mathbb{P}_0$ of $\mathbb{P}(\mathbb{X})$ containing the range of $\tilde{p}$. Such a function – a sort of sufficient statistic for $\tilde{p}$ – is called parameter of the conditional law of each ξ_n. For instance, think of t as a distinguished function of a vector of moments of $\tilde{p}$, letting $\mathbb{P}_0$ be the class of all probabilities in $\mathbb{P}$ such that those moments exist. Now, since the ordinary Bayesian inferences concern functions of $\tilde{p}$, the just recalled de Finetti's representation theorem highlights that those inferences generally deal with hypotheses that, being related to limiting mathematical entities, might be devoid of any empirical value.

In the parametric formulation, de Finetti's representation theorem can be rewritten as following:

$$\mathcal{L}_{\xi(n)}(A) = \int_{\Theta} p_{\theta}^{(n)}(A)\mathcal{L}_{t(\tilde{p})}(d\theta) \qquad (A \in \mathscr{X}^n,\ n = 1, 2, \dots) \tag{1.3}$$

being Θ a set containing all the realizations of $\tilde{\theta}$.

If observations are assumed to form a finite exchangeable sequence $(\xi_1, \dots, \xi_N)$, the representations (1.1)–(1.3) do not hold anymore. Therefore, to deal with the finitary approach, (1.1) is replaced by a finite version, which states that a finite random sequence $(\xi_1, \dots, \xi_N)$ 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 $N\tilde{e}_N(\{x\})$ balls having label x, for each atom x of $\tilde{e}_N$. This and other peculiarities of finite exchangeable sequence can be found, for instance, in Kingman (1978c), Aldous (1985), Diaconis and Freedman (1980), Schervish (1995), Spizzichino (1982), Wood (1992).

Going back to the Bayesian framemork, notice that $\mathcal{L}_{\tilde{\theta}|\xi(n)}$ denotes the *a posteriori distribution*, while the *predictive distribution* is $\mathcal{L}_{\xi(n,N)|\xi(n)}$ 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 *a posteriori* laws, namely

$$\mathcal{L}_{\xi(n,N)|\xi(n)}(\,A\,) = \int_{\Theta} p_\theta^{(N-n)}(A)\mathcal{L}_{t(\tilde{p})|\xi(n)}(d\theta) \qquad (A \in \mathscr{X}^{N-n}).$$

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 $(\xi_n)_{n\geq 1}$ and it does not have a clear meaning anyway. Vice versa, $\mathcal{L}_{\xi(n,N)|\xi(n)}$ can be always assessed in any case by resorting to the definition of conditional distribution.

Given that, in a finitary framework, inferences from $\xi(n)$ to $\tilde{p}$ are basically uninteresting even when $\tilde{p}$ can be defined. Therefore it is to be expected that statisticians focus on empirical versions

$$\tilde{\theta}_N = t(\tilde{e}_N)$$

of the usual limiting parameter $\tilde{\theta}$, and on inferences to them from $\xi(n)$ when $n < N$. This stance, which turns out to be notably significant when one is dealing with N–exchangeable sequences of observations, traces any inferential process back to a predictive problem.

1.2 Statistical methods based on a finitary approach in literature

The finitary approach to statistical inference have not been very often considered in literature, with some exceptions [see, for instance, de Finetti (1972, 1974) and Roberts (1965)]. A very concrete situation where it really seems appropriate to resort to the finitary approach arises when the statistician deals with a sample obtained without replacement from a finite population with known dimension.

Consider a finite population that consists of N units labelled $1, \ldots, N$. Attached to unit i let ξ_i be the unknown value of some character of interest. In this setting, $\tilde{e}_N$ is the frequency distribution of the character in the population. It is natural to assign an exchangeable distribution to the sequence $(\xi_1, \ldots, \xi_N)$. This reflects the assumption that labels carry no information about the units. For this reason, in this work we shall assume that $\xi_1, \ldots, \xi_n$ are the values of the character related to the sampled part of the population. Generally, the quantity to be estimated is some symmetric function (e.g. mean, median, variance) of $(\xi_1, \ldots, \xi_N)$, which, in this context, is often called *parameter* or *state of nature*. Of course, any hypotheses about such parameter is empirical since it can be always verified (at least theoretically) taking a census of the population. Most current Bayesian statistical methods applied in finite population sampling involve also other parameters, besides the state of nature, since $\xi_1, \ldots, \xi_N$ are given a joint law that makes them conditionally independent and identically distributed given an unknown parameter $\tilde{\theta}_\infty$. As explained by Ericson (1969),

> *the generation of a joint prior distribution* [for $(\xi_1, \ldots, \xi_N)$] *by this approach is, barring differences in probabilistic interpretation, equivalent to viewing the finite population as a sample from an infinite superpopulation having unknown parameter* $\tilde{\theta}_\infty$.

As already highlightened in the previous paragraphs, one can see that $\tilde{\theta}_\infty$ is not given any concrete interpretation. In the area of finite population sampling, scholars usually resort to the so called *superpopulation model.* However some papers related to specific statistical problems make use of exchangeable distributions according to a finitary approach. In this section, we shall focus on two specific proposals that are consistent with such approach: the model proposed by Hill (1968), and the *Pólya posterior.*

Hill's model. Hill considers a finite exchangeable sequence of real-valued observations such that: (1) Ties have probability zero; (2) Conditionally upon the first n observations, the next observation is equally like to fall in any of the open intervals between successive order statistics of the given sample. Under these assumptions, Hill calculates the posterior distribution of the number of distinct values in the whole population and of the percentiles. But then he realizes himself that there is no countably additive probability distribution on the space of observations such that all the assumptions are satisfied. Fortunately, Lane and Sudderth (1978) establish that the underlying probability evaluation of Hill (1968) is coherent, in the sense of de Finetti (1975), in a finitely additive framework. More precisely, they prove that it is possible to define a finitely additive exchangeable probability β on $\mathbb{R}^n$ $(n \geq 1)$ such that:

(A) For $1 \leq i < j \leq n$,

$$\beta\{x \in \mathbb{R}^n : x_i = x_j\} = 0$$

(B) For every $A \subset \mathbb{R}^{n-1}$ and $1 \leq i \leq n$,

$$\beta\big\{x \in \mathbb{R}^n : (x_1, \ldots, x_{n-1}) \in A \text{ and } x_n = x_{(i)}\big\} = \beta(A \times \mathbb{R})/n,$$

where $x_{(i)}$ is the i-th smallest coordinate of $(x_1, \ldots, x_n)$.

Hill (1968) argues that (B) means that the numerical characteristic under observation has an arbitrary or “rubbery” scale and other distinctions between observations are vague. In such case, the numerical values in the sample may be regarded as carrying only negligible information about the overall population values. On the basis of these considerations, Hill's model seems to be suitable to approach inferential problems in the context of *species sampling*, where the value of the character of interest has no numerical meaning, but it is just a label, which indicates that the unit belongs to a certain species. Hill (1979), in fact, resorts to the probability evaluation defined by (A) and (B) to obtain the posterior expectation and variance of the number of distinct species in the population and the exact posterior probability of finding a new species. Moreover, Berliner and Hill (1988) apply the very same model to the field of survival analysis, and Hill (1980) shows that (B) is reasonable to be assumed in other situations too, as in multidimensional contingency tables.

Hill (1993) introduces the *nested splitting process*, that is a concrete example, in a finitely additive framework, of an exchangeable sequence that satisfies (A) and (B) for each $n \geq 1$. A real-valued random sequence $(\xi_n)_n$ is a nested splitting process if the predictive distribution of each open interval I is

$$\mathcal{L}_{\xi_{n+1}|\xi(n)}(I) = \frac{1}{n+1}\,\mathcal{L}_{\xi_1}(I) + \frac{n}{n(n+1)}\big(C_n(I) + \tfrac{1}{2}D_n(I)\big) \quad n \geq 1,$$

where $C_n(I)$ denotes the number of observations among the first n that lie in I and $D_n(I)$ the number of ξ_i (with $i \leq n$) that are on the boundary of I.

The Pólya posterior. Meeden and Vardeman (1991) argue that a prior distribution cannot be always fully specified and that Bayes estimates are always based only on the posterior distribution, and therefore they assert that, in order to approach an inferential problem in a finite population setting, it is sufficient to create a *posterior distribution* "for the unseen given the seen", i.e. the conditional distribution of $(\xi_{n+1}, \ldots, \xi_N)$ given the sample $(\xi_1, \ldots, \xi_n)$. They claim that a good choice for such conditional distribution is the so-called *Pólya posterior*, introduced by Meeden and Ghosh (1983), i.e. the law of $(N - n)$ drawings from an urn containing $\xi_1, \ldots, \xi_n$ by a Pólya scheme. This means that each ball drawn is returned to the urn together with another one with the same label.

Note that, resorting to the Pólya posterior, the distribution of $(\xi_1, \ldots, \xi_N)$ is not fully determined (since $\mathcal{L}_{(\xi_1,\ldots,\xi_n)}$ is not assessed), and, moreover, $\mathcal{L}_{(\xi_1,\ldots,\xi_n)}$ cannot been assessed so that $\mathcal{L}_{(\xi_{i+1},\ldots,\xi_N)|(\xi_1,\ldots,\xi_i)}$ is the Pólya posterior for each $i = 1, \ldots, N$, unless one takes $\xi_1, \ldots, \xi_N$ all equal with probability one. In other words, the Pólya posterior does not arise as posterior distribution from any given (reasonable) assessment of $\mathcal{L}_{(\xi_1,\ldots,\xi_N)}$. For this reason, dealing with the Pólya posterior, Ghosh and Meeden (1997) talk about a "pseudo-posterior".

The Pólya posterior was proposed for the first time by Meeden and Ghosh (1983) approaching a particular problem: giving Bayesian justification for standard frequentistic methods by proving their admissibility. Let us recall that a decision rule δ for $\tilde{\theta}_N$ is called *admissible* if there exists no decision rule δ' dominating δ, i.e. such that $R(\theta, \delta) \geq R(\theta, \delta')$ for all θ in Θ, where $R(\theta, \delta)$ denotes the risk function

$$\theta \to \mathbb{E}(L(\tilde{\theta}_N, \delta(\xi(n))) \mid \tilde{\theta}_N = \theta).$$

Hsuan (1979) proves that an estimator is admissible if and only if it is stepwise-Bayes. Let $\Delta(\pi, \tilde{\mathbb{D}})$ denote the class of all Bayes rules against a prior π over $\tilde{\mathbb{D}} \subset \mathbb{D}$. A rule δ is said to be stepwise Bayes against the sequence $\pi_1, \pi_2, \ldots$ of priors on Θ if δ belongs to D_j for all $j = 1, 2, \ldots$, where $D_1 := \Delta(\pi_1, \mathbb{D})$ and, for $j \geq 2$, $D_j = \Delta(\pi_j, D_{j-1})$.

The Pólya posterior arises in the stepwise Bayesian argument that proves the admissibility of a variety of point estimators, in particular, the admissibility of the sample mean for estimating the population mean. In fact, assuming that $\mathbb{X}$ is finite, one can find a set of priors $\pi_1, \pi_2, \ldots, \pi_k$ for $(\xi_1, \ldots, \xi_N)$ such that the posterior of each π_j is the Pólya posterior and the sample mean is stepwise Bayes w.r.t. $\pi_1, \ldots, \pi_k$. For a detailed proof, see, for instance, Ghosh and Meeden (1997).

One notices that the choice of the Pólya posterior is equivalent to assess the predictive distribution of ξ_{n+1} given $(\xi_1, \ldots, \xi_n)$ to be equal to the empirical distribution $\sum_{i=1}^{n} \delta_{\xi_i}/N$ of the sample. For this reason, Meeden and Vardeman (1991) assert that the Pólya posterior "is a sensible predictive distribution for the unseen given the seen when the sample is assumed to be representative" and no prior informations about $(\xi_1, \ldots, \xi_N)$ is available.

A weighted version of the Pólya posterior is introduced in order to take into account prior believes about the population. In this case, $\mathbb{X}$ is assumed to be finite – say $\{b_1, \ldots, b_d\}$ – and one considers a vector of real numbers $w := (w_1, \ldots, w_d)$ such that $w_i > -1$ for each i. The *weighted Pólya posterior* for the sample $(\xi_1, \ldots, \xi_n)$ with weights given by the vector w is the probability distribution (on $\mathbb{X}^{N-n}$) of $(N - n)$ drawings from an Pólya-urn containing $\sum_{i=1}^{n} \mathbb{I}_{\{b_j\}}(\xi_i) + w_j$ balls with label b_j, for $j = 1, \ldots, d$. Roussanov (1999) proves the admissibility of the weighted Pólya posterior in finite population problems.

It should be pointed out that, in practice, the weighted Pólya posterior is the conditional law of $(\zeta_{n+1}, \ldots, \zeta_N)$ given $(\zeta_1, \ldots, \zeta_n)$ if $(\zeta_i)_{i\geq 1}$ is an exchangeable sequence directed by a Dirichlet process with parameter α being a discrete measure with finite support $\{b_1, \ldots, b_d\}$ and $\alpha(\{b_j\}) = w_j$ for $j = 1, \ldots, d$. Moreover, if $(\zeta_i)_{i\geq 1}$ is an exchangeable sequence directed by a Dirichlet process with parameter $a\bar{\alpha}$, being $\bar{\alpha}$ a probability measure and $a > 0$, then the unweighted Pólya posterior is the setwise limit of the conditional law $\mathcal{L}_{(\zeta_{n+1},\ldots,\zeta_N)|(\zeta_1,\ldots,\zeta_n)}$ as the total mass a goes to zero.

What we want to stress is that the weighted and the unweighted Pólya posteriors can be used to obtain point or interval estimations for some functions of $(\xi_1, \ldots, \xi_N)$, as it is shown by Ghosh and Meeden (1997). One is forced to resort to simulation procedures since closed forms for the estimators of interest are not available, except for the point estimator of the population mean, which is the sample mean. Therefore Ghosh and Meeden (1997) suggest to calculate an approximated estimate by simulation in this way: generate the $(N-n)$ unobserved values of the character by the Pólya posterior (that is by simulating $(N-n)$ drawings from a Pólya urn); do this a big number of times – 500 or 1000, say R – obtaining R simulated copies of the entire population; calculate the value of the quantity to be estimated for each simulated population and then consider the mean of the R values obtained. In this way, Ghosh and Meeden (1997) obtain point and interval estimates for the population median and interval estimates for the mean and for the ratio of the medians of two different characters in the same population.

Before concluding, let us point out the main differences between Pólya posterior and Hill's model, and the main features they share. The main difference, of course, is that the probability distribution of $(\xi_1, \ldots, \xi_N)$ is fully specified in Hill's model, but it is not if one resorts to Pólya posterior. Moreover, Berliner and Hill (1988) highlights that the use of the empirical measures "as predictive distribution forces one to assign a discrete distribution to a future observation with mass only at observed data points. Unreasonable statements, assigning probability 0 to a future observation larger (smaller) than the largest (smallest), are a consequence." Instead, the predictive distribution suggested by Hill spreads mass throughout each interval between two consecutive sample order statistics. (For the same reason, Hill's model differs from the probability distribution of a sequence directed by a Dirichlet process, which puts positive mass on the observed values.) On the other hand, Hill's model is suitable only "in the case of extremely vague a priori knowledge" (Hill (1993)), just like the (unweighted) Pólya posterior. Moreover, Hill's predictive distribution and the empirical measure "both give essentially the same mass to any interval that contains a moderate number of observations" (Berliner and Hill (1988)). In fact, as explained by Berliner and Hill (1988), for any open interval I containing k observations, $\tilde{e}_n(I) = k/n$, and in Hill's model the predictive probability of I is not less than $\frac{k-1}{n+1}$ and not greater than $\frac{k+1}{n+1}$, where n is the sample size.

1.3 Peculiarities and outline of the present work

In the previous section, we talked about Hill's model and Pólya posterior, which are very specific proposals that agree with a finitary approach, but consider only extensible N-exchangeable sequences. The aim of this work is to introduce, without resorting to de Finetti's representation theorem, some general classes of laws for N-exchangeable sequences, which seem appropriate in some real situations, and, moreover, need not to be infinitely exchangeable. Parts of this dissertation are based on a joint work in progress with Eugenio Regazzini and Federico Bassetti.

We shall define two families of distributions that consist in a natural extension to the finitary setting of two well-known families of infinite exchangeable sequences: sequences directed by a Pólya tree process and species sampling sequences. For this reason, we start with a review about their main features (Chapter 2).

Pólya-tree distributions were formally introduced by Lavine (1992) and Lavine (1994) and Mauldin et al. (1992), although they had been already described by Ferguson (1974) and some particular cases of Pólya-tree distributions were studied by Dubins and Freedman (1967) and Mauldin and Williams (1990). The popularity they gained in the last decades is due to the fact that they allow the possibility to put positive mass on the set of absolute continuous probability measures. Moreover, they distinguish by their versatility. In fact, they have been used in many different statistical fields, such as autoregressive modeling [Sarno (1998)], regression problems [Hanson and Johnson (2002)], statistical modeling of partially observed data [Paddock (2002)], and survival analysis [Muliere and Walker (1997), Neath (2003)].

Species sampling sequences were introduced for the first time by Pitman (1996) and studied by Hansen and Pitman (2000), Pitman (2003), Gnedin and Pitman (2005). They hinge upon the concept of random partition, which has been introduced by Kingman (1978b) and studied by Kingman (1978a), Aldous (1985) and Pitman (1995). For a systematic account on random partitions, see Pitman (2006).

Chapter 3 describes the possible representation forms of the law of a finite exchangeable sequence, providing also some mathematical tools, which will be useful in Chapter 4. It is known that the classical de Finetti's theorem cannot be applied in this case, but, as mentioned in Section 1.1, it may be replaced by a finite version, which indicates the relationship between the law of a finite exchangeable sequence $(\xi_1, \ldots, \xi_N)$ and the law of its empirical distribution $\tilde{e}_N$. The problem of assessing the law of $\tilde{e}_N$ is approached in Section 3.2.1. For each measurable partition $\{A_1, \ldots, A_k\}$ of $\mathbb{X}$, we consider a p.m.f. $\psi_{A_1,\ldots,A_k}$ that gives, for any vector of integers $(N_1, \ldots, N_k)$, the probability that the number of ξ_i in A_j is N_j $(j = 1, \ldots, k)$. We find some necessary conditions on the $\psi_{A_1,\ldots,A_k}$'s for the existence of $(\xi_1, \ldots, \xi_N)$, and we show that the very same conditions are also sufficient to provide a complete characterization of $\tilde{e}_N$, and, therefore, of the exchangeable probability measure $\mathcal{L}_{(\xi_1,\ldots,\xi_N)}$ as well. After showing how the $\psi_{A_1,\ldots,A_k}$'s look like in some common examples of finite exchangeable sequences, it is explained how such functions can be concretely assessed in order to determine an exchangeable probability measure on $\mathbb{X}^N$ if $\mathbb{X}$ is a Polish space.

Chapter 4 and Chapter 5 introduce two new classes of distributions of N-exchangeable sequences, which we shall call *partition tree distributions* and *random partition distributions*, respectively. In the construction of these laws we will follow two different strategies: the former based on partitions trees, the latter on random partitions. They enable us to consider forms of negative correlation between past and future observations, contrary to what happens in infinite exchangeable sequences, which permit positive correlation only. For the sake of clarity, the aforesaid strategies can give rise to laws that exhibit inverse relation between the conditional probability that ξ_{n+1} belongs to a specific set A given $\xi(n)$, and $\tilde{e}_N(A)$. Therefore, these laws need not be infinitely extensible. On the other hand, partition tree distributions include as specific cases laws of N-exchangeable sequences directed by a Pólya-tree process, and random partition distributions include the laws of the initial segments, with length N, of species sampling sequences.

Chapter 6 deals with applications of the distributions introduced in Chapter 4 and Chapter 5 to some standard statistical problems: we show how one can estimate the mean of the empirical measure, and we propose a bivariate model based on partition tree distributions in order to approach regression problems.

Acknowledgments

First, I wish to thank my supervisor, Eugenio Regazzini, for proposing me the interesting issues that originated this work and for his invaluable guidance on this dissertation. Moreover, I am thankful to Federico Bassetti for his meaningful suggestions and helpful advise. I also thank Pietro Rigo for his suggestions and encouragement. Finally, I thank my sister, Maria Paola, for helping me improve the English form of the present work.

Ἓν οἶδα ὅτι οὐδὲν οἶδα.

Plato, *The Apology of Socrates.*

Chapter 2

Sequences directed by a Pólya-tree process and species sampling sequences

This chapter is a review about two well-known families of infinite exchangeable sequences: sequences directed by a Pólya-tree process and species sampling sequences. The distributions of the initial N-segments of such sequences admit meaningful representation forms, different from de Finetti's theorem, which – as it will be shown in Chapters 4 and 5 – lead to define more general classes of laws of finite exchangeable sequences, which also include distributions of non-extensible sequences. This chapter focuses on the peculiarities of Pólya tree distributions and species sampling sequences that look more interesting according to a finitary point of view.

2.1 Pólya-tree distributions

Lavine (1992) and Mauldin et al. (1992) define Pólya-tree distributions only on particular spaces, such as the unit interval. Here Pólya-trees distributions will be presented in relation with more general spaces, according to the definition given, for example, by Schervish (1995). For a systematic account about Pólya-tree processes, see also Ghosh and Ramamoorthi (2003).

Let $(\mathbb{X}, \mathscr{X})$ be a measurable space such that $\mathscr{X}$ is countably generated, and let Π be a separating tree of partitions of $\mathbb{X}$. This means that Π is a sequence $(\pi_m)_{m=0}^{\infty}$ of ordered, finite, measurable partitions of $\mathbb{X}$ such that $\pi_0 := \{\mathbb{X}\}$; π_{m+1} is a refinement of π_m for every $m \geq 0$ and $\mathscr{G} := \cup_0^{\infty} \pi_m$ generates the measurable sets. Denote by $B_{m,\,1}, \ldots, B_{m,\,k_m}$ the elements of partition π_m. Moreover, it will be convenient to indicate the *most recent superset* of $B \in \pi_m$ by $\mathrm{ge}(B) \in \pi_{m-1}$, i.e. the set C in π_{m-1} that includes B.

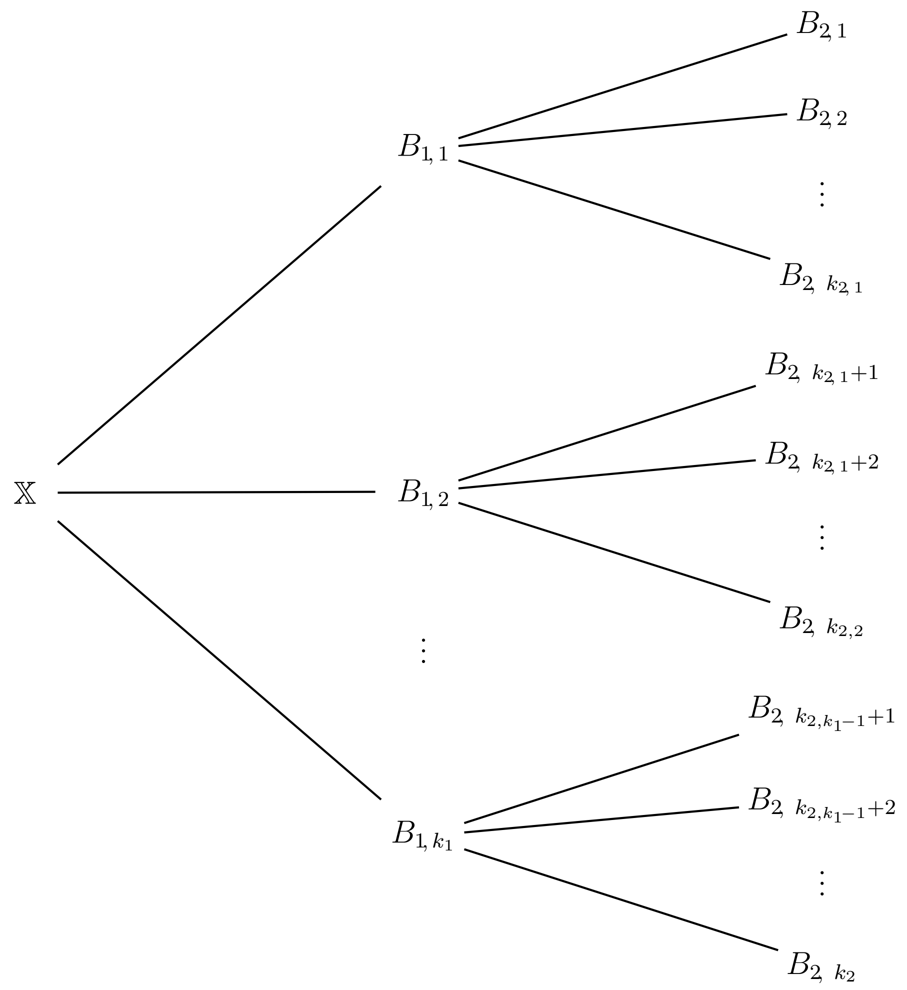


Figure 2.1: Partitions tree.

A random probability measure $\tilde{p}$ on $(\mathbb{X}, \mathscr{X})$ is said to be a *Pólya tree* process (or equivalently its law is called Pólya-tree distribution) with parameter $(\Pi, \mathcal{A})$, where $\mathcal{A} = \{\alpha_{m,j} \geq$

$0 : j = 1, \dots, k_m; m = 0, 1, \dots\}$ is a set of nonnegative numbers, if

- The collections $\{\tilde{p}(C \mid \mathrm{ge}(C)) : C \in \pi_m\}$, $m \geq 1$, are stochastically independent (*independence between partitions*), i.e. $\tilde{p}$ is an *F-neutral* process (also called *tail free* process; see Ferguson (1974)).
- The collections $\{\tilde{p}(C \mid B) : \mathrm{ge}(C) = B\}$, $B \in \pi_m$, are stochastically independent for each m (*independence within partitions*).
- For each B in π_m and for $m \geq 0$, the random vector $(\tilde{p}(C \mid B) : \mathrm{ge}(C) = B)$ has Dirichlet distribution with parameter $(\alpha_{m+1,j} : B_{m+1,j} \subset B)$ (the reference both to random vectors and to parameters vectors implies the introduction of some order among the descendants of $\mathrm{ge}(\cdot)$, for example a natural left-to-right order).

In order to parametrize the distribution of a Pólya tree process, it is sometimes useful to consider a sequence $(\alpha^{(m)})_{m\geq 1}$ of finite measures such that, for each $m \geq 1$, $\alpha^{(m)}$ is defined on the algebra $\mathscr{A}_m$ generated by π_m and $\alpha^{(m)}(B_{m,j}) = \alpha_{m,j}$ for $j = 1, \dots, k_m$.

Provided that $\mathbb{X}$ is a Polish (i.e. complete, separable and metric) space and $\mathscr{X}$ its Borel sigma-field, a necessary and sufficient condition for the existence of a Pólya tree process with parameter $\mathcal{A}$ is that if B_n is a union of elements of π_m for each n, $B_1 \supset B_2 \supset \dots$ and $\cap_{n=1}^{\infty} B_n = \emptyset$, then

$$\prod_{k=1}^{\infty} \left(\sum_{B \in \pi_n : B \subset B_n} \alpha^{(n)}(B \mid \mathrm{ge}(B)) \right) = 0.$$

For a systematic account about the construction and conditions for existence of Pólya-tree distributions and other probability measures on $(\mathbb{P}, \mathscr{P})$ see Regazzini (2004) or Ghosh and Ramamoorthi (2003).

2.1.1 A very special case: the Dirichlet process

Pólya-tree distributions are a generalization of the well-known Dirichlet process, which was firstly considered by Freedman (1963), but has become a basic component of nonparametric Bayesian statistics after the apparition of a celebrated paper by Ferguson (1973).

Before defining the Dirichelet process, we need to recall the notion of (finite-dimensional) Dirichlet distribution. Let $(a_1, \dots, a_h)$ be a vector of positive numbers. A random vector $(\zeta_1, \ldots, \zeta_{h-1})$ is said to have (finite-dimensional) Dirichlet distribution with parameter $(a_1, \ldots, a_h)$ if and only if its law has density (w.r.t. the Lebesgue measure on $\mathbb{R}^{h-1}$) given by

$$\frac{\Gamma(a_1 + \cdots + a_h)}{\Gamma(a_1) \cdots \Gamma(a_h)}\, x_1^{a_1 - 1} \cdots x_{h-1}^{a_{h-1}-1} (1 - x_1 - \cdots - x_{h-1})^{a_h - 1}\, \mathbb{I}_{T_{h-1}}(x_1, \ldots, x_{h-1}),$$

when $a_j > 0$ for each j, denoting:

$$T_{h-1} = \{(x_1, \ldots, x_{h-1}) \in \mathbb{R}^{h-1} : x_j \geq 0 \text{ for } 1 \leq j \leq h-1,\ x_1 + \cdots x_{h-1} < 1\}.$$

If some of the a_j is zero, we still define the Dirichlet distribution convening that those coordinates corresponding to $a_i = 0$ are equal to zero with probability one and the rest of the coordinates have the usual Dirichlet distribution.

Let α be a finite measure on some measurable space $(\Omega, \mathscr{F})$. A random probability measure $\tilde{p}$ on $(\Omega, \mathscr{F})$ is said to be a *Dirichlet process* with parameter α if and only if, for any finite measurable partition $\{A_1, \ldots, A_k\}$ of Ω, the probability distribution of the random vector $(\tilde{p}(A_1), \ldots, \tilde{p}(A_{k-1}))$ is the (finite-dimensional) Dirichlet distribution with parameters $(\alpha(A_1), \ldots, \alpha(A_k))$. The law of a Dirichlet process is called *Dirichlet distribution.*

Notice that no request was made about the measurable space $(\Omega, \mathscr{F})$. A good peculiarity of the Dirichlet process is indeed that it can be properly defined on any measurable space. This can be seen by resorting to its series representation presented by Sethuraman (1994). If one confines oneself to considering only countably genereted σ-fields, then we can say that the class of Pólya-tree distributions contains the class of Dirichlet distributions. More precisely, a Pólya-tree process on $(\mathbb{X}, \mathscr{X})$ with parameter $(\Pi, \mathcal{A})$ such that

$$\alpha_{m,j} = \sum_{l : B_{m+1,l} \subset B_{m,j}} \alpha_{m+1,\, l}$$

– or equivalently such that, for each $m \geq 1$, $\alpha^{(m)}$ is a restriction to $\mathscr{A}_m$ of $\alpha^{(m+1)}$ – is a Dirichlet process with parameter α, where α is that measure on $(\mathbb{X}, \mathscr{X})$ such that $\alpha(B_{m,j}) = \alpha_{m,j}$ for each $m \geq 0$ and each $1 \leq j \leq k_m$. In fact, Dirichlet process satisfies independence within and between partitions. For a detailed proof of this fact, see Regazzini (2004).

2.1.2 Marginal, posterior, and predictive distributions

The class of Pólya-tree distributions is conjugate, in the sense that for any Pólya-tree process $\tilde{p}$, its conditional distribution given $\xi_1, \ldots, \xi_n$ (i.e. its posterior distribution) is still Pólya-tree, for any n. More precisely, if $\tilde{p}$ is a Pólya-tree with parameters $(\Pi, \mathcal{A})$, then the conditional distribution of $\tilde{p}$ given $(\xi_1, \ldots, \xi_n)$ is Pólya-tree with parameters $(\Pi, \mathcal{A}^*)$, where

$$\mathcal{A}^* = \{\alpha^*_{m,j} \geq 0 : \ j = 1, \ldots, k_m; \ m = 0, 1, \ldots\},$$

and $\alpha^*_{m,j} = \alpha_{m,j} + \sum_{i=1}^n \delta_{\xi_i}(B_{m,j})$.

The predictive distribution of a Pólya-tree process with parameter $(\Pi, \mathcal{A})$ is given by (see Regazzini (2004)):

$$\begin{aligned}
&\mathcal{L}_{\xi_{n+1}|\xi(n)}(B) \\
&\quad = \frac{\alpha^{(1)}(B_1) + n\,\tilde{e}_n(B_1)}{\alpha^{(1)}(\mathbb{X}) + n} \cdot \frac{\alpha^{(2)}(B_2) + n\,\tilde{e}_n(B_2)}{\alpha^{(2)}(B_1) + n\,\tilde{e}_n(B_1)} \cdots \frac{\alpha^{(m)}(B_m) + n\,\tilde{e}_n(B_m)}{\alpha^{(m)}(B_{m-1}) + n\,\tilde{e}_n(B_{m-1})}
\end{aligned} \tag{2.1}$$

for each B in π_m and each $m \geq 1$, where $B_m = B$ and, for $j < m$, B_j denotes the set in π_j that contains B_m. Keeping the same notation, the marginal distribution can be expressed as following:

$$\mathcal{L}_{\xi_{n+1}}(B) = \frac{\alpha^{(1)}(B_1)}{\alpha^{(1)}(\mathbb{X})} \cdot \frac{\alpha^{(2)}(B_2)}{\alpha^{(2)}(B_1)} \cdots \frac{\alpha^{(m)}(B_m)}{\alpha^{(m)}(B_{m-1})} \tag{2.2}$$

for any $B \in \pi_m$ and $m \geq 1$.

Also the subclass of Dirichlet distributions is conjugate. In fact the posterior distribution of a Dirichlet process with parameter α is Dirichlet with parameter $\alpha + \sum_{i=1}^n \delta_{\xi_i}$.

Predictive distributions of Dirichlet process have a very appealing form, being a convex linear combination of the observed frequency and the marginal distribution, which is $\bar{\alpha}(\cdot) = \alpha(\cdot)/\alpha(\mathbb{X})$:

$$\mathcal{L}_{\xi_{n+1}|\xi_1,\ldots,\xi_n}(\cdot) = \frac{n}{a+n}\,\tilde{e}_n(\cdot) + \frac{a}{a+n}\,\bar{\alpha}(\cdot), \tag{2.3}$$

and, moreover, they characterize it [see Regazzini (1978), Lo (1991), Fortini et al. (2000)].

2.1.3 Main differences between the Dirichlet process and the more general family of Pólya-tree processes

The popularity that Pólya tree processes gained is mainly due to the fact that their paths can be continuous or even absolutely continuous with probability one, while a Dirichlet distribution always gives probability one to the set of discrete probability measures. This is considered a drawback of Dirichlet processes, according to the usual approach to inference, which assumes that observations are generated by an “unknown” probability distribution. According to a predictive approach, an important difference between Dirichlet distributions and other Pólya-tree processes is that the second ones can be constructed so that predictive distributions turn out to be absolutely continuous. Conditions under which the prior, the posterior, or the predictive distribution of a Pólya-tree process turn out to be absolutely continuous can be found in Schervish (1995) and Regazzini (2004), while Drăghici and Ramamoorthi (2000) give conditions for the prior and the posterior of a Pólya tree process to be mutually continuous, as well as conditions for the prior and the posterior to be mutually singular.

On the other hand, a drawback of Pólya-trees and F-neutral distributions is their dependence on the choice of Π: Doksum (1974) shows that, except for trivial special cases, the Dirichlet processes are the only tail-free processes in which Π does not play any role.

2.1.4 Pólya-tree processes defined on the unit interval

Initially, Pólya tree processes were defined only on the unit interval (Mauldin et al. (1992)). Let us focus on this spacial case, i.e. take $\mathbb{X} = (0, 1]$. Let E be $\{0, 1\}$ and $E^0 := \emptyset$, $E^* := \cup_{m=0}^{\infty} E^m$. Here π_m can be taken to be the set of all 2^m dyadic intervals of rank m, i.e.

$$\pi_m := \{I_\varepsilon : \varepsilon \in E^m\},$$

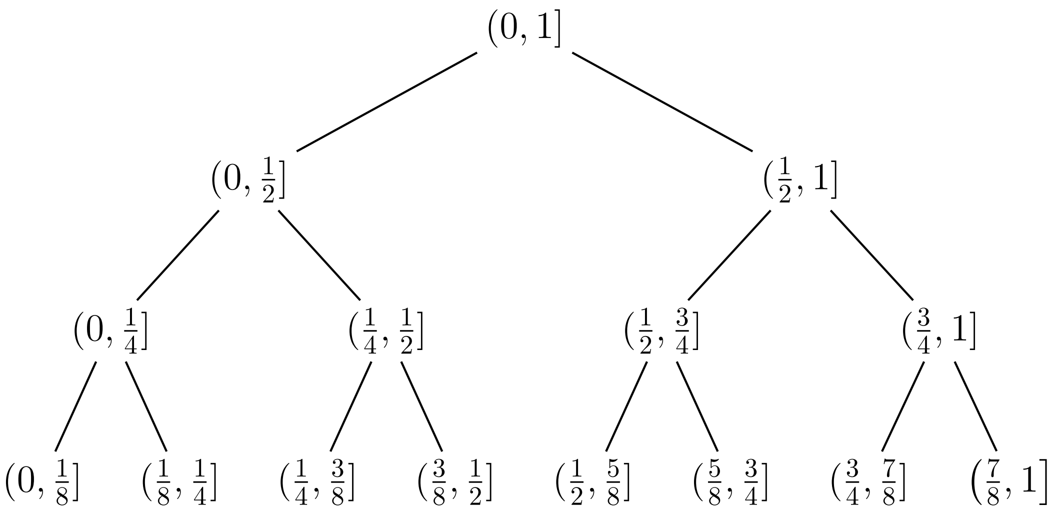
where

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

if $m \geq 1$ and $I_\emptyset = (0, 1]$. In fact, it is known that the class of dyadic intervals generates the Borel sigma-field $\mathscr{B}((0, 1])$ of $(0, 1]$.

In this case, Π is a binary tree, and therefore we can say that $\tilde{p}$ is a Pólya-tree on $((0, 1], \mathscr{B}((0, 1]))$ with parameter $(\Pi, \mathcal{N})$ if there exist nonnegative numbers $\mathcal{N} = \{\alpha_\varepsilon : \varepsilon \in E^*\}$ such that the random variables $\tilde{p}(I_{\varepsilon 1} \mid I_\varepsilon)$ with $\varepsilon \in E^*$ are stochastically independent and each one of them has the Beta distribution with parameter $(\alpha_{\varepsilon 0}, \alpha_{\varepsilon 1})$.

Mauldin et al. (1992) prove that a sequence of exchangeable $(0, 1]$-valued random variables directed by a Pólya-tree process, with such parameters, can be generated by an urn scheme. They define Pólya–tree distributions using the set $E = \{0, 1, \ldots, k\}$, but their results can be described taking $E = \{0, 1\}$ instead, without losing any important detail. In



the terminology of Mauldin et al. (1992), a "Pólya-tree" is a function that assigns to every ε in E^* an urn $u(\varepsilon)$ containing balls labeled as "one" and balls labeled as "zero". Let $\alpha_{\varepsilon 0}$ be the number of balls in urn $u(\varepsilon)$ with label "zero" and $\alpha_{\varepsilon 1}$ the number of balls in the same urn with label "one", for each $\varepsilon \in E^*$. For instance, urn $u(\emptyset)$ contains α_0 balls with label "zero" and α_1 balls with label "one", while urn $u(101)$ contains α_{1010} balls with label "zero" and α_{1011} balls with label "one". The Pólya-tree u can be used to generate a sequence of random variables $\xi_{1,1}, \xi_{1,2}, \ldots$ and a new tree $u^{(1)}$ as follows: (a) draw a ball at random from $u(\emptyset)$, replace it by two with the same label, and set $\xi_{1,1} = j_1$ if the ball is labeled with j_1 ($j_1 \in \{0,1\}$), (b) draw a ball from $u(j_1)$, replace it by two of the same color, and set $\xi_{1,2} = j_2$ if the ball just drawn has label j_2, (c) go on to $u(j_1, j_2)$, and continue in this fashion. Let $u^{(1)}$ be the new Pólya tree that was obtained in the construction. Iterate the entire process to obtain the sequences $(\xi_{1,1}, \xi_{1,2}, \ldots), (\xi_{2,1}, \xi_{2,2}, \ldots), \ldots$ and the Pólya-trees $u^{(1)}, u^{(2)}, \ldots$. Finally set

$$\xi_i = \sum_{k=1}^{\infty} 2^{-k} \xi_{k,i} \quad \text{for } i = 1, 2, \ldots \, .$$

Mauldin et al. (1992) show that the sequence $\xi_1, \xi_2, \ldots,$ generated by this scheme, is an exchangeable sequence directed by a Pólya-tree process with parameter $(\Pi, \mathcal{N})$.

The predictive distribution can be expressed as:

$$\mathcal{L}_{\xi_{n+1}|\xi(n)}(I_{\varepsilon_1\dots\varepsilon_m 1}) = \frac{\alpha_{\varepsilon_1}+n_{\varepsilon_1}}{\alpha_0+\alpha_1+n}\cdot\frac{\alpha_{\varepsilon_1\varepsilon_2}+n_{\varepsilon_1\varepsilon_2}}{\alpha_{\varepsilon_1 0}+\alpha_{\varepsilon_1 1}+n_{\varepsilon_1}}\cdots\frac{\alpha_{\varepsilon_1\dots\varepsilon_m 1}+n_{\varepsilon_1\dots\varepsilon_m 1}}{\alpha_{\varepsilon_1\dots\varepsilon_m 0}+\alpha_{\varepsilon_1\dots\varepsilon_m 1}+n_{\varepsilon_1\dots\varepsilon_m}}, \quad (2.4)$$

where $n_\varepsilon = \sum_{i=1}^n \delta_{\xi_i}(I_\varepsilon)$ for any ε in E^*; see Theorem 4.3 in Mauldin et al. (1992) and Walker and Muliere (1997).

2.2 Species sampling sequences

Hansen and Pitman (2000) consider a class of exchangeable sequences that represents another generalization of the Dirichlet process, since, as in (2.3), their predictive distribution is a linear combination of the empirical measure and the marginal law, but the coefficients are functions of the sample. In formula, a sequence $\xi_1,\dots,\xi_N$ in this class admits predictive distributions of the form:

$$\mathcal{L}_{\xi_{n+1}|\xi_1,\dots,\xi_n}(\cdot) = \sum_{i=1}^n r_{i,n}\delta_{\xi_i}(\cdot) + q_n\nu(\cdot) \qquad (n=2,3,\dots), \quad (2.5)$$

for some $r_{i,n}$ and q_n, which are non-negative product-measurable functions of $(\xi_1,\dots,\xi_n)$, and for some probability measure ν on $(\mathbb{X},\mathscr{X})$. Hansen and Pitman (2000) focus on the case in which ν is a diffuse measure, i.e. $\nu(\{x\})=0$ for each x in $\mathbb{X}$. The only requirement that $(\mathbb{X},\mathscr{X})$ needs to satisfy is to render singletons $\mathscr{X}$-measurable and the diagonals $\{(x,y):x=y\}$ $\mathscr{X}^2$-measurable.

Rule (2.5) can be rewritten as follows, by grouping terms with equal values of ξ_i:

$$\mathcal{L}_{\xi_{n+1}|\xi_1,\dots,\xi_n}(\cdot) = \sum_{j=1}^{K_n} p_{j,n}\delta_{\xi_j^*}(\cdot) + q_n\nu(\cdot) \qquad (n=2,3,\dots), \quad (2.6)$$

where the ξ_j^* for $1\le j\le K_n$ are the distinct values among $\xi_1,\dots,\xi_n$ in the order that they appear, and the $p_{j,n}$ and q_n are some non-negative product-measurable functions of $(\xi_1,\dots,\xi_n)$.

Pitman (1996) shows that the law of an exchangeable sequence satisfying (2.6) can be represented by means of the law of an exchangeable random partition of $\{1,2,\dots\}$.

Before presenting such result, it seems appropriate to recall the concept of exchangeable random partition, which have been introduced by Kingman (1978a), and studied by

Kingman (1978a); Aldous (1985) and Pitman (1995).

Given a measurable space $(\Omega, \mathscr{F})$, a random partition $\tilde{\pi}$ of $\{1, 2, \dots\}$ on Ω is a map from Ω into the set $\mathcal{P}_\infty$ of all partitions of $\{1, 2, \dots\}$ such that the sets

$$\tilde{\pi}_{i,j} := \{\omega \in \Omega : \ i \text{ and } j \text{ belong to the same component of } \tilde{\pi}(\omega)\} \qquad \text{for } i, j \in \{1, 2, \dots\}$$

are $\mathscr{F}$-measurable. $\tilde{\pi}$ is said to be exchangeable if $\mathcal{L}_{\{\mathbb{I}_{\tilde{\pi}_{i,j}} : 1 \leq i,j \leq n\}} = \mathcal{L}_{\{\mathbb{I}_{\tilde{\pi}_{\sigma_i,\sigma_j}} : 1 \leq i,j \leq n\}}$ for any permutation σ of $\{1, \dots, n\}$ and for any $n \geq 1$. Let $\tilde{\pi}|_n$ denote the restriction of $\tilde{\pi}$ to the finite set $\{1, \dots, n\}$ for each $n \geq 1$. A random partition $\tilde{\pi}$ on a probability space $(\Omega, \mathscr{F}, P)$ is exchangeable if and only if there exists a symmetric function p of sequences of positive integers, such that, for each n and for each particular partition $\{A_1, \dots, A_k\}$ of $\{1, \dots, n\}$,

$$P\{\tilde{\pi}|_n = \{A_1, \dots, A_k\}\} = p(n_1, \dots, n_k),$$

where $n_j = |A_j|$ for $1 \leq j \leq k$, $n_j \geq 1$, and $\sum_{i=1}^k n_i = n$. The function p is called the *exchangeable partition probability function* (EPPF) of $\tilde{\pi}$. An EPPF is subject to the following sequence of *addition rules*:

$$p(n_1, \dots, n_k) = \sum_{j=1}^{k} p(\dots, n_j + 1, \dots) + p(n_1, \dots, n_k, 1) \qquad (k = 1, 2, \dots), \tag{2.7}$$

where $(\dots, n_j + 1, \dots)$ is derived from $(n_1, \dots, n_k)$ by substituting $n_j + 1$ for n_j.

Given a random sequence $X_1, X_2, \dots$, let $\Pi(X_1, X_2, \dots)$ denote the random partition generated by $X_1, X_2, \dots$, i.e. such that two positive integers i and j belong to the same block of $\Pi(X_1, X_2, \dots)$ if and only if $X_i = X_j$.

Going back to the exchangeable sequences satisfying (2.6), Pitman (1996) states the following result:

Theorem 2.1 (Pitman (1996)). *Let ν be a diffuse probability measure. An exchangeable sequence $(\xi_n)_{n\geq 1}$ satisfies* (2.6) *if and only if there exist a random partition $\tilde{\pi}$ and a sequence of i.i.d. r.v.'s $(\xi_n^*)_{n\geq 1}$ with distribution ν such that*

1. *conditionally on $\{\tilde{\pi} = \{A_1, \dots, A_k\}\}$, $\xi_n = \xi_i^*$ for $n \in A_i$,*

2. *$\tilde{\pi}$ and $(\xi_n^*)_{n\geq 1}$ are stochastically independent,*

3. $\tilde{\pi}$ *is distributed as the partition* $\Pi(\xi_1, \xi_2, \dots)$ *generated by* $\xi_1, \xi_2, \dots,$ *and its EPPF* p *satisfies for each* k *and each* k*-sequence of integers* $(n_1, \dots, n_k)$ *the following:*

$$p_{j,\,n}(n_1, \dots, n_k) = \frac{p(\dots, n_{j+1}, \dots)}{p(n_1, \dots, n_k)} \qquad \textit{for} \quad 1 \le j \le k$$
$$q_n(n_1, \dots, n_k) = \frac{p(n_1, \dots, n_k, n_{k+1})}{p(n_1, \dots, n_k)}$$

provided $p(n_1, \dots, n_k) > 0$.

Pitman (1996) calls such an exchangeable sequence $(\xi_n)_{n\ge1}$ a *species sampling sequence.*

As it is explained by Hansen and Pitman (2000),

> *this terminology is used to suggest the interpretation of* $(\xi_n)_{n\ge1}$ *as the sequence of species of individuals in a process of sequential random sampling from some hypothetical infinite population of individuals of various species. The species of the first individual to be observed is assigned a random tag* ξ_1^* *distributed according to* ν*. Given the tags* $\xi_1, \dots, \xi_n$ *of the first* n *individuals observed, it is supposed that the next individual is one of the* j*-th species observes so far with probability* $p_{j,n}$ *and one of a new species with probability* q_n*. Each distinct species is assigned an independent random tag with distribution* ν *as it appears in the sampling process.*

2.3 Normalized random measures with independent increments

Species sampling sequences are related to the class of exchangeable sequences directed by normalized random measures with independent increments. These have been introduced by Regazzini et al. (2003) and studied by Prünster (2002), Nieto-Barajas et al. (2004), James (2005), and Sangalli (2006).

A random measure $\tilde{\mu}$ with independent increments on the real line $\mathbb{R}$ is a random measure such that, for any measurable collection $\{A_1, \dots, A_k\}$ $(k \ge 1)$ of pairwise disjoint measurable subsets of $\mathbb{R}$, the random variable $\tilde{\mu}(A_1), \dots, \tilde{\mu}(A_k)$ are stochastically independent. A systematic account of these random measures is given for example by Kingman (1967).

Random measures with independent increments are completely characterized by a measure ν on $\mathbb{R} \times \mathbb{R}^+$ via their Laplace functional, more precisely for every A in $\mathcal{B}(\mathbb{R})$ and every positive λ one has

$$\mathbb{E}(e^{-\lambda\tilde{\mu}(A)}) = \exp\left\{-\int_{A\times\mathbb{R}^+}(1-e^{-\lambda v})\nu(\mathrm{d}x\,\mathrm{d}v)\right\}.$$

Following Regazzini et al. (2003), if $\int_{\mathbb{R}\times\mathbb{R}^+}(1-e^{-\lambda v})\nu(\mathrm{d}x\,\mathrm{d}v) < +\infty$ for every positive λ and $\nu(\mathbb{R}\times\mathbb{R}^+) = +\infty$, one defines a *normalized random measure with independent increments* (NRMII) setting $\tilde{p}(\cdot) := \tilde{\mu}(\cdot)/\tilde{\mu}(\mathbb{R})$. In point of fact, under the previous assumptions, $P\{\tilde{\mu}(\mathbb{R}) = 0\} = 0$, see Regazzini et al. (2003).

Consider now a sequence $(\xi_i)_{i\geq 1}$ of exchangeable random variables driven by $\tilde{p}$. If

$$\nu(\mathrm{d}x\,\mathrm{d}v) = a\,\alpha(\mathrm{d}x)\,q(\mathrm{d}v), \tag{2.8}$$

given any $N < +\infty$, one can restate Corollary 2 in Sangalli (2006) saying that there is a random partition $\tilde{\pi}$, taking values in $\mathcal{P}_\infty$, such that, for each $n \geq 1$ and for each particular partition $\{A_1, \dots, A_k\}$ of $\{1, \dots, n\}$,

$$\mathcal{L}_{\tilde{\pi}|_n}(\{A_1, \dots, A_k\}) = \frac{a^k}{\Gamma(n)}\int_{\mathbb{R}^+}\lambda^{n-1}L(\lambda)\prod_{j=1}^{k}\int_{\mathbb{R}^+}v^{n_j}e^{-\lambda v}q(\mathrm{d}v)\mathrm{d}\lambda, \tag{2.9}$$

where $n_j = |A_j|$ for $1 \leq j \leq k$, $n_j \geq 1$, $\sum_{i=1}^k n_i = n$, $L(\lambda) = \exp\{-a\int_{\mathbb{R}^+}(1-e^{-\lambda v})q(\mathrm{d}v)\}$, and, moreover, the law of $\xi_1, \xi_2, \dots$ given $\tilde{\pi}$ satisfies Condition 1 of Theorem 2.1. Hence, if the law of ξ_1 is diffuse, $(\xi_1, \xi_2, \dots)$ is a species sampling sequence. NRMII is another generalization of the Dirichlet process. In fact, a NRMII with $\nu(\mathrm{d}x\mathrm{d}v) = \alpha(\mathrm{d}x)v^{-1}e^{-v}\mathrm{d}v$ is a Dirichlet process with parameter α (see Regazzini et al. (2003), Prünster (2002)).

Chapter 3

Forms of representation for laws of finite exchangeable sequences

On the basis of the considerations contained in Chapter 1, the sequence of observations will be assumed to be finite and exchangeable. This chapter will focus on the possible representations of laws of finite exchangeable sequences. In particular, we shall show how such laws can be determined by assessing the finite dimensional distributions of the empirical process. Theorem 3.4 will be used in Chapter 4 for defining new classes of exchangeable distributions for finite sequences.

3.1 De Finetti's theorem and exchangeable prolongable sequences

From now on, we shall consider a finite sequence $\xi := (\xi_1, \ldots, \xi_N)$ of observations instead of an infinite sequence. Such observations may refer to the values that a character of interest assumes on the N units of a finite population or to the results in a given experiment. In the latter case, the value N is the maximum number of trials that can be performed.

Suppose that observations take values in a measurable space $(\mathbb{X}, \mathscr{X})$. Hence, each observation ξ_i $(i = 1, \ldots, N)$ can be viewed as a measurable function from $\mathbb{X}^N$ into $\mathbb{X}$ according to the following definition:

$$\xi_i(x) = x_i \qquad \text{for every} \quad x = (x_1, \ldots, x_N) \text{ in } \mathbb{X}^N \quad \text{and} \quad i = 1, \ldots, N$$

Our goal is to assess a probability measure P on $(\mathbb{X}^N, \mathscr{X}^N)$, i.e. the probability distribution of $(\xi_1, \ldots, \xi_N)$.

We shall take P to be *exchangeable*, that is invariant under permutation, i.e. for any permutation σ of $(1, \ldots, N)$, the distribution of $(\xi_{\sigma_1}, \ldots, \xi_{\sigma_N})$ is the same as the distribution of $(\xi_1, \ldots, \xi_N)$. With such assessment of P, we can equivalently say that $(\xi_1, \ldots, \xi_N)$ is an *N-exchangeable sequence*.

After introducing the notation we shall refer to in this and in the next chapters, let us stress the fact that P need not satisfy representation (1.2). By de Finetti's theorem, such representation is satisfied for some random measure $\tilde{p}$ if and only if $(\xi_1, \ldots, \xi_N)$ is *(infinitely) extensible*, i.e. it is distributed as the initial segment of some infinite exchangeable sequence. A counter-example is easily found to show that a finite exchangeable sequence need not be extensible. For instance, if ξ_1 is a Bernoulli distributed r.v. with parameter 1/2 and $\xi_2 = 1 - \xi_1$ the 2-exchangeable sequence (ξ_1, ξ_2) is not extensible since the two components are negatively correlated. In fact, it is known that, given an infinite sequence $\eta_1, \eta_2, \ldots$ of random variables with the same variance, if they are equally correlated (in particular if they are exchangeable), then they cannot be negatively correlated. Indeed an easy algebraic calculation yields:

$$\mathrm{Var}(\textstyle\sum_{i=1}^{n} \eta_i) = n\,\mathrm{Var}(\eta_1) \, + \, n(n-1)\,\mathrm{Cov}(\eta_1, \eta_2), \tag{3.1}$$

and, since the variance must be nonnegative, by (3.1) the correlation coefficient between η_1 and η_2 must be grater or equal than $-1/(n-1)$ for any n, hence it must be greater than zero. This fact was already highlighted by de Finetti (1937b), who also proposed an interesting geometric interpretation.

In conclusion, in a finitary framework, it may be that the usual directing measure $\tilde{p}$ does not even exist, and, as already said in Chapter 1, it is destined to be replaced by the empirical measure, both as object of inference, and as base to assess the joint distribution of the observations.

3.2 The law of the empirical measure as prior distribution

From now on, let us denote by $\tilde{e}$ the *empirical measure* of $(\xi_1, \ldots, \xi_N)$:

$$\tilde{e} = \tfrac{1}{N} \sum_{i=1}^{N} \delta_{\xi_i}.$$

First let us stress that $\tilde{e}$ is a random probability measure on $(\mathbb{X}, \mathscr{X})$, i.e. a measurable map from $(\mathbb{X}^N, \mathscr{X}^N, P)$ into $(\mathbb{P}, \mathscr{P})$. In fact, for any A belonging to $\mathscr{X}$ and any B belonging to the Borel σ-algebra $\mathscr{B}(\mathbb{R})$ of $\mathbb{R}$, the set

$$\begin{aligned} C' &= \Big\{(x_1, \ldots, x_N) \in \mathbb{X}^N : \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \delta_{x_i} \in \{p \in \mathbb{P} : p(A) \in B\}\Big\} \\ &= \Big\{(x_1, \ldots, x_N) \in \mathbb{X}^N : \tfrac{1}{N} \textstyle\sum_{i=1}^{N} \delta_{x_i}(A) \in B\Big\} \end{aligned}$$

turns out to be a union of a finite class of finite intersections of measurable sets of the form $\big\{(x_1, \ldots, x_N) \in \mathbb{X}^N : x_k \in A\big\}$ or $\big\{(x_1, \ldots, x_N) \in \mathbb{X}^N : x_k \notin A\big\}$. Hence, C' belongs to $\mathscr{X}^N$, and measurability of $\tilde{e}$ follows from the fact that the sets of the form $\{p \in \mathbb{P} : p(A) \in B\}$ (with $A \in \mathscr{X}$ and $B \in \mathscr{B}(\mathbb{R})$) generate $\mathscr{P}$. Hence we can properly talk about the law of the empirical measure. Moreover, we can regard such a law as a prior distribution. In fact, it is known that we can characterize the law of an N-exchangeable sequence by the law of its empirical measure.

3.2.1 Representation of a finite exchangeable sequence through the law of the empirical measure

Before giving a precise formulation to the finite version of de Finetti's theorem, which was mentioned in Chapter 1, some more notation is necessary.

Let $|A|$ be the cardinality of set A, and denote by $\mathcal{H}_{m_1,\ldots,m_h}(n_1, \ldots, n_h)$ the probability to get n_j balls marked with j $(j = 1, \ldots, h)$ when one draws $(n_1 + \cdots + n_h)$ balls without replacement from an urn containing m_j balls marked by j. Define the *multinomial coefficient* as

$$\binom{m}{m_1 \ldots m_j} = \frac{m!}{m_1! \ldots m_j!} \quad \text{whenever } m = m_1 + \cdots + m_j.$$

Denote by $\mathbb{P}_N$ the class of all probability measures p on $(\mathbb{X}, \mathscr{X})$ such that $p = \sum_{i=1}^{N} \delta_{x_i}/N$ for some N-tuple $(x_1, \ldots, x_N)$, where $x_i \in \mathbb{X}$ for $i = 1, \ldots, N$, and finally, for

$n \leq N$ and for any $p \in \mathbb{P}_N$ with $p = \sum_{i=1}^{N} \delta_{x_i}/N$, let $\varphi_n(p)$ be the probability distribution of n drawings without replacement from an urn containing N balls labeled as $x_1, \ldots, x_N$.

At this stage, the representation theorem for finite exchangeable sequences due to Bruno de Finetti can be stated:

Theorem 3.1 (The finite version of de Finetti's representation theorem). *The following facts are equivalent:*

(*i*) $(\xi_1, \ldots, \xi_N)$ *is an* N*-exchangeable sequence;*

(*ii*) *for any* $n \leq N$, $\varphi_n(\tilde{e})$ *represents a regular conditional distribution for* $(\xi_1, \ldots, \xi_n)$ *given* $\tilde{e}$.

Proof. For the proof see either Schervish (1995) on pages 38–40 or Aldous (1985) on pages 37–38. □

Theorem 3.1 describes the one-to-one relationship between the law of a finite exchangeable sequence and the law of its empirical measure. Briefly, it says that a finite sequence of N random elements is exchangeable if and only if, conditionally on the empirical measure $\tilde{e}$, the first n components are distributed as n drawings without replacement from an urn with N balls and whose composition is given by $\tilde{e}$.

Another less abstract representation theorem can be stated, which relates the distribution of $(\xi_1, \ldots, \xi_N)$ to the finite-dimensional distributions of $\tilde{e}$, i.e. the laws of the random vectors $(\tilde{e}(A_1), \ldots, \tilde{e}(A_k))$ where $\{A_1, \ldots, A_k\}$ is a measurable partition of $\mathbb{X}$. Such distributions of course are discrete. For any measurable partition $\{A_1, \ldots, A_k\}$ of $\mathbb{X}$ and any vector of integers $(M_1, \ldots, M_k)$, denote the probability that exactly M_j observations fall in A_j by

$$\psi_{A_1 \ldots A_k}(M_1, \ldots, M_k) := P(\tilde{e}(A_1) = M_1/N, \ldots, \tilde{e}(A_k) = M_k/N). \tag{3.2}$$

Corollary 3.2. *The following facts are equivalent:*

(*i*) $(\xi_1, \ldots, \xi_N)$ *is an* N*-exchangeable sequence;*

(*ii*) *for any measurable partition* $\{A_1, \ldots, A_k\}$ *of* $\mathbb{X}$ *and any* n*-tuple* $(i_1, \ldots, i_n)$ *of elements of* $\{1, \ldots, k\}$ *with* $n \leq N$,

$$P(\xi_1 \in A_{i_1}, \ldots, \xi_n \in A_{i_n}) = \sum_{(M_1, \ldots, M_k)} \psi_{A_1 \ldots A_k}(M_1, \ldots, M_k) \cdot \frac{\mathcal{H}_{M_1, \ldots, M_k}(n_1, \ldots, n_k)}{\binom{N}{M_1 \ldots M_k}}, \tag{3.3}$$

where the sum runs over the finite set of all vectors $(N_1, \ldots, N_k)$ *whose components are nonnegative integers that sum up to* N *and* $n_j = |\{l = 1, \ldots, n : i_l = j\}|$ *for* $j = 1, \ldots, k$*;*

(*iii*) *under the same assumption as (ii),*

$$P(\xi_1 \in A_{i_1}, \ldots, \xi_n \in A_{i_n} \mid \tilde{e}(A_1), \ldots, \tilde{e}(A_k)) = \frac{\mathcal{H}_{\tilde{N}(A_1),\ldots,\tilde{N}(A_k)}(n_1, \ldots, n_k)}{\binom{N}{\tilde{N}(A_1),\ldots,\tilde{N}(A_k)}}, \tag{3.4}$$

where $\tilde{N}(A_j) := N\tilde{e}(A_j)$ *for* $j = 1, \ldots, k$.

Proof. Trivially (ii) implies (i) (take $n = N$), and (iii) implies (ii).

In order to prove that (i) implies (iii), we note that by Theorem 3.1,

$$P(\xi_1 \in A_{i_1}, \ldots, \xi_n \in A_{i_n} \mid \tilde{e}) = \frac{\mathcal{H}_{\tilde{N}(A_1),\ldots,\tilde{N}(A_k)}(n_1, \ldots, n_k)}{\binom{N}{\tilde{N}(A_1),\ldots,\tilde{N}(A_k)}} \quad P-\text{a.s.} \tag{3.5}$$

Since the right hand side of (3.5) depends on $\tilde{e}$ only through $(\tilde{e}(A_1), \ldots, \tilde{e}(A_k))$,

$$\begin{aligned}
&P(\xi_1 \in A_{i_1}, \ldots, \xi_n \in A_{i_n} \mid \tilde{e}(A_1), \ldots, \tilde{e}(A_k)) \\
&\quad = \mathbb{E}\left(P(\xi_1 \in A_{i_1}, \ldots, \xi_n \in A_{i_n} \mid \tilde{e}) \mid \tilde{e}(A_1), \ldots, \tilde{e}(A_k)\right) \\
&\quad = P(\xi_1 \in A_{i_1}, \ldots, \xi_n \in A_{i_n} \mid \tilde{e}) = \frac{\mathcal{H}_{\tilde{N}(A_1),\ldots,\tilde{N}(A_k)}(n_1, \ldots, n_k)}{\binom{N}{\tilde{N}(A_1),\ldots,\tilde{N}(A_k)}} \quad P-\text{a.s.}.
\end{aligned}$$

□

Remark 3.3. Equation (3.3) in Corollary 3.2 can be rewritten as:

$$P(\xi(n) \in A_{i_1} \times \cdots \times A_{i_n}) = \int_{[0,1]^k} \frac{\mathcal{H}_{N\theta_1,\ldots,N\theta_k}(n_1, \ldots, n_k)}{\binom{N}{N\theta_1 \ldots N\theta_k}} \mathcal{L}_{\tilde{e}(A_1),\ldots,\tilde{e}(A_k)}(\mathrm{d}\theta_1 \ldots \mathrm{d}\theta_k) \tag{3.6}$$

When $(\xi_1, \ldots, \xi_N)$ turns out to be infinitely extensible, one can heuristically derive (1.2) from (3.6), by taking the limit as N goes to infinity, and recalling that (a) the multivariate hypergeometric distribution converges (uniformly) to the multinomial distribution with parameters $(n; \tilde{p}(A_1), \ldots, \tilde{p}(A_k))$; (b) $(\tilde{e}(A_1), \ldots, \tilde{e}(A_k))$ converges in distribution to $(\tilde{p}(A_1), \ldots, \tilde{p}(A_k))$.

When $n = N$ (3.3) yields a trivial reformulation of the notion of exchangeability:

$$\psi_{A_1 \ldots A_k}(M_1, \ldots, M_k) = \binom{N}{M_1, \ldots, M_k} P(\xi_1 \in A_{i_1}, \ldots, \xi_N \in A_{i_N}) \tag{3.7}$$

holds for any n-tuple $(i_1, \ldots, i_n)$ of elements of $\{1, \ldots, k\}$ such that

$$M_j = |\{l = 1, \ldots, N : i_l = j\}| \qquad (j = 1, \ldots, k).$$

3.2.2 Representation of a finite exchangeable sequence through the finite-dimensional distributions of the empirical measure.

By Theorem 3.1, the law of a finite exchangeable sequence is determined by the specification of the probability distribution of its empirical measure. Our aim is to identify the latter through the finite-dimensional distributions of the empirical measure. It consists in singling out conditions that are necessary and sufficient so that a family of finite-dimensional distributions may provide a complete characterization of the law of the empirical measure of a random sequence. Such discrete distributions are fully specified by the functions $\psi_{A_1,\ldots,A_k}$ by (3.2).

First we shall enunciate some necessary conditions on the $\psi_{A_1,\ldots,A_k}$'s for the existence of P, which hold in general (without hypothesis of exchangeability).

Condition 3.1. For each measurable partition $\{A_1,\ldots,A_k\}$ of $\mathbb{X}$, let $(\tilde{N}(A_1),\ldots,\tilde{N}(A_k))$ be a random vector having non-negative integer coordinates whose p.m.f. is given by $\psi_{A_1,\ldots,A_k}$. Then, for each measurable partition $\{A_1,\ldots,A_k\}$ of $\mathbb{X}$:

3.1.1. $\tilde{N}(A_1),\ldots,\tilde{N}(A_k)$ are non-negative integers and sum up to N;

3.1.2. if $\{B_1,\ldots,B_m\}$ is a measurable partition not coarser than $\{A_1,\ldots,A_k\}$, then

$$\left(\sum_{l:B_l\subset A_1}\tilde{N}(B_l),\ldots,\sum_{l:B_l\subset A_k}\tilde{N}(B_l)\right)\stackrel{\mathcal{L}}{=}(\tilde{N}(A_1),\ldots,\tilde{N}(A_k))$$

3.1.3. If $\{C_n\}$ is a sequence of events belonging to $\mathscr{X}$ and such that $C_n\downarrow\emptyset$, then the sequence $\{(\tilde{N}(C_n),\tilde{N}(C_n^c))\}$ of random vectors must converge to $(0,N)$ in law.

Conditions 3.1.1 and 3.1.2 trivially follow by definition of $\psi_{A_1,\ldots,A_k}$, while Condition 3.1.3 is due to continuity from above of P. In fact, notice that if $M\neq 0$

$$P\{\tilde{N}(C_n)=M\}\leq P\{\tilde{N}(C_n)\neq 0\}=1-P\{\tilde{N}(C_n)=0\}=1-P(C_n^c\times\ldots\times C_n^c)$$

which goes to zero as $n\to\infty$ by continuity of P.

Condition 3.1 is necessary for the existence of an N-exchangeable sequence $(\xi_1,\ldots,\xi_N)$ that satisfies (3.7). It is known that, by Condition 3.1, the $\psi_{A_1,\ldots,A_k}$'s identify a consistent family of (discrete) finite-dimensional-probability distributions for a random probability measure [see either Regazzini (1991) or Regazzini and Petris (1992)]. It still need to be proved that such random process is distributed as the empirical measure of an N-exchangeable sequence. Hence, we shall prove that Condition 3.1 and (3.7) are sufficient to identify the exchangeable law P. Furthermore, the very same result can be obtained also when Conditions 3.1.1-3.1.2 are satisfied only by the sets in some semialgebra[1] that generates $\mathscr{X}$. More precisely we can state the following result:

Theorem 3.4 (A representation theorem for finite exchangeable sequences). *Let $\mathbb{X}$ be a Polish space, let $\mathscr{X}$ be its Borel σ-algebra and let $\mathscr{G}$ be a semi-algebra that generates $\mathscr{X}$. Denote by $\mathscr{A}$ the algebra generated by $\mathscr{G}$ and by $\mathscr{A}^N$ the algebra of all finite disjoint unions of cartesian products of sets in $\mathscr{A}$.*
If for any partition $\{A_1, \ldots, A_k\}$ of $\mathbb{X}$ with A_j belonging to $\mathscr{G}$ for each j, $(\tilde{N}(A_1), \ldots, \tilde{N}(A_k))$ is a random vector such that:

1. *its components are non-negative integers and sum up to N;*
2. *if $\{B_1, \ldots, B_m\}$ is a partition not coarser than $\{A_1, \ldots, A_k\}$ and B_l belongs to $\mathscr{G}$ for each l ($l = 1, \ldots, m$), then*
$$\left(\sum_{l: B_l \subset A_1} \tilde{N}(B_l), \ldots, \sum_{l: B_l \subset A_k} \tilde{N}(B_l) \right) \stackrel{\mathcal{L}}{=} (\tilde{N}(A_1), \ldots, \tilde{N}(A_k));$$

Then there exists a unique finitely-additive exchangeable probability measure ρ_N on $(\mathbb{X}^N, \mathscr{A}^N)$ such that
$$(\tilde{e}(A_1), \ldots, \tilde{e}(A_k)) \stackrel{\mathcal{L}}{=} \tfrac{1}{N}(\tilde{N}(A_1), \ldots, \tilde{N}(A_k)) \tag{3.8}$$
for each partition $\{A_1, \ldots, A_k\}$ of $\mathbb{X}$ with A_j belonging to $\mathscr{G}$ for each j.

Moreover, ρ_N can be uniquely extended to an (exchangeable) probability measure P on $(\mathbb{X}^N, \mathscr{X}^N)$ if and only if the following is true:

3. *if $C_n \downarrow \emptyset$ as $n \to +\infty$ and $C_n \in \mathscr{A}$ for each n, then the sequence $\{\tilde{e}(C_n)\}$ of random variables converges to zero in law as $n \to +\infty$.*

[1] a class $\mathscr{S}$ of sets is said to be a semi-algebra if and only if:

- $S, T \in \mathscr{S} \Longrightarrow T \cap S \in \mathscr{S}$ i.e. $\mathscr{S}$ is closed under intersection,
- $S \in \mathscr{S} \Longrightarrow S^c$ is a finite disjoint union of sets in $\mathscr{S}$.

Proof. See Appendix, page 95. □

Remark 3.5. As it is shown in the proof of the theorem, if one wishes to define a finitely-additive exchangeable probability, then it is not necessary for $\mathbb{X}$ to be Polish. That is: if $(\mathbb{X}, \mathscr{X})$ is any measurable space, and the hypotheses 1 and 2 of the theorem hold, then there exists a unique finitely-additive exchangeable probability measure ρ_N on $(\mathbb{X}^N, \mathscr{A}^N)$ such that (3.8) is satisfied.

Briefly, Theorem 3.4 says that it is possible to characterize the law of an N-exchangeable sequence through the $\psi_{A_1,\dots,A_k}$'s, i.e. the finite-dimensional probability mass functions of the process $\tilde{N}$.

We shall now see how the $\psi_{A_1,\dots,A_k}$'s look like in some common examples of finite exchangeable sequences.

Example 3.1 (independent observations)**.** If $\xi_1, \dots, \xi_N$ are N independent random elements and α is their common distribution, then

$$\psi_{A_1\dots A_k}(M_1, \dots, M_k) = \binom{N}{M_1, \dots, M_k} \alpha(A_1)^{M_1} \dots \alpha(A_k)^{M_k}.$$

Example 3.2 (extensible N-exchangeable sequences)**.** If $(\xi_1, \dots, \xi_N)$ are the first N coordinates of an exchangeable sequence with de Finetti's measure π, then

$$\psi_{A_1\dots A_k}(M_1, \dots, M_k) = \int_{\mathbb{P}} \binom{N}{M_1, \dots, M_k} \mu(A_1)^{M_1} \dots \mu(A_k)^{M_k} d\pi(\mu).$$

Before the next example, define the *factorial* of x, for x real number, by

$$x! = \begin{cases} \Gamma(x+1) & \text{if } x \text{ is not a negative integer} \\ \dfrac{(-1)^{n-1}}{(n-1)!} & \text{if } x = -n \text{ and } n \text{ is a positive integer,} \end{cases}$$

where Γ denotes the generalized gamma function. In this way, $(x+1)! = (x+1)\,x!$ still holds when $x \neq 0$ and the *binomial (coefficient) of a of order b* can be defined for any pair (a, b) of real numbers by: $\binom{a}{b} = \frac{a!}{b!(a-b)!}$.

Moreover, let

$$x^{[n]} := x(x+1)\dots(x+n-1).$$

$x^{[n]}$ is called the *ascending factorial of x of order n.*

Example 3.3 (a sample from a Dirichlet process)**.** Let α be a finite measure on $(\mathbb{X}, \mathscr{X})$, and set

$$\psi_{A_1\dots A_k}(M_1,\dots,M_k) = \binom{N}{M_1\dots M_k}\frac{\prod_{i=1}^k(\alpha_i+M_i)(\alpha_i+M_i-1)\cdots\alpha_i}{(a+N)(a+N-1)\cdots a} \tag{3.9}$$

that is equivalent to

$$\psi_{A_1\dots A_k}(M_1,\dots,M_k) = \frac{\prod_{i=1}^{k}\binom{-\alpha_i}{M_i}}{\binom{-a}{N}} = \frac{\prod_{i=1}^{k}\binom{\alpha_i+M_i}{M_i}}{\binom{a+N}{N}} = \frac{1}{a^{[N]}}\prod_{j=1}^{k}\alpha_j^{[N_j]}, \tag{3.10}$$

where we denote $\alpha(A_i)$ by α_i and $\alpha(\mathbb{X})$ by a.

The p.m.f. in (3.10) is called the *Dirichlet (or Beta)-compound multinomial distribution* with parameters $(N;\alpha_1,\dots,\alpha_k)$ [see Johnson et al. (1997), page 80] and is the natural multivariate version of the beta-binomial distribution, also known as negative (or inverse) hypergeometric distribution [see Johnson et al. (2005)]. In fact, the marginal p.m.f.'s of (3.10) are of this form.

Here P is the joint distribution of the first N coordinates of an exchangeable random sequence directed by a Dirichlet process with parameter α, whose law we shall denote by $\mathscr{D}_\alpha$. In fact, if $\xi_1,\xi_2,\dots$ is an exchangeable sequence and $\mathscr{D}_\alpha$ is its de Finetti's measure, then (3.10) gives the finite-dimensional p.m.f.'s of the empirical measure $N\tilde{e} = \sum_{i=1}^N \delta_{\xi_i}$. This follows by a simple algebraic argument considering (3.7) and knowing that the law of an N-dimensional sample from $\mathscr{D}_\alpha$ evaluated in $A_1^{M_1}\times\dots\times A_k^{M_k}$ (where M_j's are positive integers and sum up to N) coincides with the mixed $(M_1,\dots,M_k)$-th moment of a singular Dirichlet distribution with parameters $(\alpha_1,\dots,\alpha_k)$, that is:

$$\begin{aligned}
P((\xi_1,\dots,\xi_N)\in A_1^{M_1}\times\dots\times A_k^{M_k}) &= \int_{\mathbb{P}} p(A_1)^{M_1}\dots p(A_k)^{N_k}d\mathscr{D}_\alpha(p)\\
&= \int_{[0,1]^{k-1}} p_1^{M_1}\dots p_k^{N_k}\frac{\Gamma(a)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_k)}p_1^{\alpha_1}\dots p_k^{\alpha_k}\,\mathrm{d}p_1\dots\mathrm{d}p_{k-1}\\
&= \frac{\Gamma(a)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_k)}\int_{\mathbb{P}} p_1^{\alpha_1+M_1}\dots p_k^{\alpha_k+N_k}\,\mathrm{d}p_1\dots\mathrm{d}p_{k-1}\\
&= \frac{\Gamma(a)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_k)}\frac{\Gamma(\alpha_1+M_1)\dots\Gamma(\alpha_k+N_k)}{\Gamma(a+N)}\\
&= \frac{1}{a^{[N]}}\prod_{j=1}^{k}\alpha_j^{[N_j]},
\end{aligned} \tag{3.11}$$

where $p_k := 1 - \sum_{j=1}^{k-1} p_j$.

Suppose that α is concentrated on a finite subset of $\mathbb{X}$, say $\{b_1, \ldots, b_h\}$. In this case, it is known that (3.10) is the p.m.f. of a Pólya distribution [see Blackwell and MacQueen (1973)]. Consider an urn containing $\alpha(\{b_l\})$ balls labeled with l. Draw N balls at random from the urn, replacing each ball drawn with two balls of the same color. Then, (3.10) is the probability to get N_j balls labeled as l with b_l in A_j, if you draw N balls from such urn and you put back each ball drawn together with another ball with the same label.

Example 3.4 (A very simple not-extendible finite exchangeable sequence)**.** Put $\mathbb{X} = \mathbb{R}$ and $N = 2$. Let $\mathscr{B}$ denote the Borel σ-field of $\mathbb{R}$ and let μ be a not degenerate probability measure on $(\mathbb{R}, \mathscr{B})$ that is symmetric w.r.t. zero (i.e. such that $\mu((-\infty, t]) = \mu([-t, \infty))$ for each $t \in \mathbb{R}$; e.g.: a normal, a double exponential, a Cauchy distribution with scale parameter zero, or any other distribution that is absolutely continuous w.r.t. Lebesgue measure whose density function is an even function). If $M_1 + M_2 = 2$, set

$$\psi_{A,A^c}(M_1, M_2) = \begin{cases} \mu(A \cap (-A)) & \text{if } \ M_1 = 2, M_2 = 0 \\ 2\,\mu(A \cap (-A^c)) & \text{if } \ M_1 = 1, M_2 = 1 \\ \mu(A^c \cap (-A^c)) & \text{if } \ M_1 = 0, M_2 = 2, \end{cases}$$

$$\text{where} \quad -A = \{x \in \mathbb{R} : -x \in A\}.$$

It's easy to see that P is the distribution of the exchangeable vector $(\tilde{Z}, -\tilde{Z})$, where $\tilde{Z}$ is a random variable with distribution μ.

Put $f = \mathbf{1}_{(0,+\infty)}$. Notice that the correlation coefficient of $(f(\tilde{Z}), f(-\tilde{Z}))$ is -1, that is $-\frac{1}{N-1}$. Hence, the sequence is not extendible to a random vector of $\mathbb{R}^3$ [see Spizzichino (1982), Proposition 2.1, page 316].

In what follows, if $R = I_1 \times \cdots \times I_q$, $\mathbb{R} \supseteq I_m = (a_m, b_m]$ and x is a vertex of R (i.e. $x \in R$, and for each m the m^{th} coordinate x_m of x is either a_m or b_m), let $\operatorname{sgn}_R x$ be $+1$ or -1, according as the number of m's satisfying $x_m = a_m$ is even or odd. For a real function the difference of F around the vertices of R is $\Delta_R F = \sum_x \operatorname{sgn}_R x \cdot F(x)$, the sum extending over the 2^q vertices x of R.

Example 3.5 ($\mathbb{X} = \mathbb{R}$)**.** Let $\mathscr{G}$ be the class of intervals like $(a, b]$, $(a, +\infty)$, or $(-\infty, b]$, $a, b \in \mathbb{R}$. If F is a symmetric distribution function (s.d.f.) and $(A_1, \dots, A_k) \subset \mathscr{G}$, set

$$\psi_{A_1 \dots A_k}(M_1, \dots, M_k) = \binom{N}{M_1, \dots, M_k} \Delta_E F,$$

$$\text{where } E = A_{i_1} \times \cdots \times A_{i_N}$$

$$\text{and } M_j = |\{l = 1, \dots, N : i_l = j\}| \quad (j = 1, \dots, k).$$

Here P is the probability measure on $(\mathbb{R}, \mathscr{B})$ associated with the s.d.f. F and therefore $\Delta_E F = P(E)$.

Example 3.6 (Gaussian random variables)**.** In Example 3.5, F can be the distribution function (d.f.) of a $\mathbb{R}^N$-valued Gaussian random vector with mean vector $(\mu, \dots, \mu)$ and covariance matrix $\sigma^2 R$, where $\mu \in \mathbb{R}$, $\sigma^2 > 0$, R is a matrix whose elements are all equal to ρ , except those in the diagonal that are equal to 1, being $\rho > -\frac{1}{N-1}$. In this way, the matrix $\sigma^2 R$ is positive definite. In fact, $\sigma^2 R$ is positive definite if and only if R is so and the quadratic form associated with R is

$$\begin{aligned} \textstyle\sum_i x_i^2 + \sum_i \sum_{j \neq i} \rho\, x_i x_j &= N(1-\rho)[\tfrac{1}{N} \textstyle\sum_i x_i^2 - (\tfrac{1}{N} \sum_i x_i)^2] \\ &\quad + \tfrac{1}{N}[(N-1)\rho + 1](\textstyle\sum_i x_i)^2 \\ &\geq \tfrac{1}{N}[(N-1)\rho + 1](\textstyle\sum_i x_i)^2 > 0 \end{aligned}$$

whenever $(x_1, \dots, x_N) \in \mathbb{R}^N \backslash \{0\}$ if $\rho > -\frac{1}{N-1}$.

3.2.3 Constructing the law of a finite exchangeable sequences on a Polish space

It is not easy to have a guess about a possible choice of the $\psi_{A_1, \dots, A_k}$'s consistently with Condition 3.1.2 , except in the area of a family of distributions already studied in literature. In order to be able to construct new examples of extensible and not extensible finite exchangeable sequences on the basis of Theorem 3.4, some further considerations are necessary.

Take $\mathbb{X}$ to be a Polish space and let $\mathscr{X}$ be its Borel σ-field. Recall that in this case, $\mathscr{X}$ is countably generated, i.e. there is a countable class of sets $\{B_1, B_2, \dots\}$ generating $\mathscr{X}$.

Therefore, we can construct a separating tree of partition $\Pi = \{\pi_m\}$ of $\mathbb{X}$ by taking $\pi_0 = \{\mathbb{X}\}$ and π_m equal to the partition generated by $\{B_1, \ldots, B_m\}$ for $m = 1, 2, \ldots$. Notice that $\mathscr{G} = \cup_{m \geq 0} \pi_m$ is a semi-algebra of $\mathbb{X}$ that generates $\mathscr{X}$. Resorting to the notation introduced in Section 2.1, let $(\tilde{N}_m)_{m \geq 0}$ be a sequence of random vectors, defined on the same probability space, such that

$$
\begin{aligned}
\tilde{N}_0 &:= \tilde{N}(\mathbb{X}) = N \\
\tilde{N}_1 &:= (\tilde{N}(B_{1,\,1}), \ldots, \tilde{N}(B_{1,\,k_1})) \\
&\cdots \\
\tilde{N}_m &:= (\tilde{N}(B_{m,\,1}), \ldots, \tilde{N}(B_{m,\,k_m})) \\
&\cdots,
\end{aligned}
$$

where, for each $m \geq 0$, $\tilde{N}(B_{m,\,1}), \ldots, \tilde{N}(B_{m,\,k_m})$ are (almost surely) non-negative integers summing up to N, and for each C in $\mathscr{G}$:

$$
\sum_{B:\ \mathrm{ge}(B) = C} \tilde{N}(B) = \tilde{N}(C). \tag{3.12}
$$

Notice that the hypotheses 1 and 2 of Theorem 3.4 are satisfied and therefore the probability distribution of the random sequence $(\tilde{N}_m)_{m \geq 1}$ characterizes a finitely-additive exchangeable probability ρ_N on $(\mathbb{X}^N, \mathscr{X}^N)$. Since $(\tilde{N}_m)_{m \geq 1}$ satisfies (3.12) for each C in $\mathscr{G}$, $(\tilde{N}_m)_{m \geq 1}$ 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 $\tilde{N}_{m+1}$. So, to achieve our ends, it is enough to assess the conditional distribution of $\tilde{N}_{m+1}$ given $\tilde{N}_m$, for every $m \geq 0$, consistently with (3.12).

At this stage, we can define, for each $A \in \mathscr{A}$, the random variable $\tilde{N}(A)$ by

$$
\tilde{N}(A) = \sum_{j=1}^{h} \tilde{N}(B_j)
$$

if A is the finite disjoint union of the sets $B_1, \ldots, B_h$ in $\mathscr{G}$. Such definition is consistent in virtue of (3.12).

Finally, again by Theorem 3.4, the finitely-additive probability ρ_N can be (uniquely) extended to an (exchangeable) probability measure P on $(\mathbb{X}^N, \mathscr{X}^N)$, provided that

$$
\tilde{N}(C_n) \to 0 \text{ in law} \tag{3.13}
$$

for any decreasing sequence $(C_n)_{n\geq 1}$ of events in $\mathscr{A}$ such that $C_n \downarrow \emptyset$.

In this way, an exchangeable probability distribution P is assigned to $(\xi_1, \dots, \xi_N)$, so that

$$P\{\xi_1 \in B_{m,i_1}, \dots, \xi_N \in B_{m,i_N}\} = \frac{1}{\binom{N}{N_1 \dots N_m}} \psi_{B_{m,1},\dots,B_{m,k_m}}((N_1, \dots, N_{k_m}))$$

holds for any $m \geq 1$, any n-tuple $(i_1, \dots, i_n)$ of elements from $\{1, \dots, k_m\}$ with $N_j := |\{l = 1, \dots, N : i_l = j\}|$ for $j = 1, \dots, k_m$, denoting by $\psi_{B_{m,1},\dots,B_{m,k_m}}$ the p.m.f. of $(\tilde{N}(B_{m,1}), \dots, \tilde{N}(B_{m,k_m}))$.

Dealing with a general $\mathbb{X}$, it looks difficult to verify that a given assessment of the conditional laws $\mathcal{L}_{\tilde{N}_{m+1}|\tilde{N}_m}$ (for each m) satisfies (3.13). This is not the case if one considers some specific spaces. If for instance $\mathbb{X}$ is the unit interval, it is sufficient for sigma-additivity of P just that (3.13) holds only for dyadic intervals. It is appropriate to clarify this point since some models based on $(0, 1]$-valued observations will be presented later on (Chapter 4).

Hence, take $\mathbb{X} = (0, 1]$ and $\mathscr{X} = \mathscr{B}((0, 1])$, and, resorting to the notation introduced in Section 2.1.4, let π_m be the set of all dyadic intervals of rank m, for each $m \geq 1$, and let $\mathscr{A}$ be the algebra generated by the class $\mathscr{G}$ of dyadic intervals.

First, note that given an infinite zero-one sequence $\varepsilon^* = (\varepsilon_1, \varepsilon_2, \dots)$ that is definitely zero (i.e. there exists m_0 such that for any $m \geq m_0$, $\varepsilon_m = 0$), the set $\cap_{m\geq 1} I_{\varepsilon_1 \dots \varepsilon_m}$ is empty and therefore, by (3.13), $\tilde{N}(I_{\varepsilon_1 \dots \varepsilon_m})$ converges to zero in law, as m goes to infinity. We can prove that this condition is also sufficient for sigma-additivity. The proof of the following Proposition is similar to the one of Theorem 2.3.2 in Ghosh and Ramamoorthi (2003), which is about random measures.

Proposition 3.6. *Let ρ be an exchangeable finitely additive probability on $\mathscr{A}^N$.*

Hence, ρ is countably additive if and only if $\tilde{N}(I_{\varepsilon_1 \dots \varepsilon_m})$ goes to zero in law (as m diverges to $+\infty$), for any zero-one sequence $(\varepsilon_1, \varepsilon_2, \dots)$ that is definitely zero.

Proof. We already observed that the mentioned condition is necessary for countable additivity; therefore we have to prove only sufficiency.

Let $\mu = \rho \circ \xi_1^{-1}$, where ξ_1 is the first coordinate function on $\mathbb{X}^N$. As we already observed in analogous circumstances (in the proof of Theorem 3.4), applying Sazonov (1965)'s results, we can prove that ρ is sigma-additive just showing that μ is so.

Since μ is finitely additive, the function $F(\cdot) = \mu(0, \cdot]$ 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, we can find $m \in \mathbb{N}$ and a zero-one sequence $(\varepsilon_1, \dots, \varepsilon_m)$ 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 $k \geq m+1$

$$F(x + 2^{-n}) = F(x) + \mu(I_{\varepsilon_1 \dots \varepsilon_n}) \qquad \text{for } n > m. \tag{3.14}$$

Now recall that $\frac{1}{N}\,\mathbb{E}(\tilde{N}(\cdot)) = \mu(\cdot)$ and that weak convergence of $(\tilde{N}(I_{\varepsilon_1 \dots \varepsilon_m}))_m$ is equivalent to L^1-convergence since each random variable in the sequence has the same finite support. So by hypothesis, $\lim_{n\to\infty} \mu(I_{\varepsilon_1 \dots \varepsilon_n}) = 0$ since $(\varepsilon_1, \varepsilon_2, \dots)$ is definitely zero and equation (3.14) 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 $F(2^{-n}) = \mu(I_{\varepsilon_1 \dots \varepsilon_n})$, where $\varepsilon_k = 0$ for any k. □

Remark 3.7. Intuitively, this condition is due to the fact that the dyadic expansion establishes a one to one map ϕ from $(0,1]$ onto the subset D^c of $\{0,1\}^\infty$, where D is the set of all zero-one sequences that are definitely zero. This is another way to say that for dyadic rationals we consider the expansion that is definitely one and not the other possible one that is definitely zero. So ϕ associates a cylinder in $\{0,1\}^\infty$ to each dyadic interval in $(0,1]$. Now, recall that any finitely-additive probability on the algebra of the cylinders is sigma-additive [see Billingsley (1995) on page 29]. Therefore a finitely-additive probability on the algebra of the cylinders in $\{0,1\}^\infty$ that is concentrated on D^c corresponds to a (sigma-additive) probability on the algebra of the dyadic intervals in $(0,1]$.

3.3 Characterization of the law of the observations through predictive distributions

It is possible to find necessary and sufficient conditions so that a set of predictive distributions may be consistent with the law of a finite exchangeable sequence. In fact, the solution to the analogous problem found by Fortini et al. (2000) for infinite sequences of observations can be easily adapted to the case of finite sequences in the following way:

Theorem 3.8. *Let $(\mathbb{X}, \mathscr{X})$ be a Polish space endowed with its Borel sigma-algebra, and let $\xi_1, \ldots, \xi_N$ be the coordinate functions on $(\mathbb{X}^N, \mathscr{X}^N)$.*

Then the following are equivalent:

1. $\mathrm{P}^{(1)}$ *is a probability measure on* $(\mathbb{X}, \mathscr{X})$ *and* $\mathrm{P}^{(\mathrm{n})}$ *is a transition probability w.r.t.* $\mathbb{X}^{n-1} \times \mathbb{X}$ *for each* $n = 2, \ldots, N$ *such that:*

   (a) $\mathrm{P}^{(\mathrm{n})}(x(n-1), A) = \mathrm{P}^{(\mathrm{n})}((x_{\sigma(1)}, \ldots, x_{\sigma(n-1)}), A)$ *holds true for each* $n = 2, \ldots, N$, *every subset* A *of* $\mathscr{X}$, *and every permutation* σ *of* $(1, \ldots, n-1)$;

   (b) *for every* A, B *in* $\mathscr{X}$ *and for each* $n = 2, \ldots, N$,

$$\int_B \mathrm{P}^{(\mathrm{n+1})}(x(n), A)\ \mathrm{P}^{(\mathrm{n})}(x(n-1), \mathrm{d}x_n) = \int_A \mathrm{P}^{(\mathrm{n+1})}(x(n), B)\ \mathrm{P}^{(\mathrm{n})}(x(n-1), \mathrm{d}x_n);$$

2. *there exist a unique probability measure P on $(\mathbb{X}^N, \mathscr{X}^N)$ such that the random elements $\xi_1, \ldots, \xi_N$ are exchangeable,* $\mathrm{P}^{(1)}$ *is the distribution of* ξ_1, *and* $\mathrm{P}^{(\mathrm{n})}$ *is a version of the conditional distribution of* ξ_n *given* $\xi(n-1)$ *for each* $n = 2, \ldots, N$.

Proof. The proof is already contained in the proof given by Fortini et al. (2000) for their Theorem 3.1. □

Among other possible representations for finite exchangeable sequences, a result given by Dellacherie and Meyer (1980) deserves to be mentioned: any finite exchangeable sequence is a "mixture" of i.i.d. sequences if we allow the mixing measure to be a *signed* measure [see also Kerns and Szekely (2005)].

Chapter 4

Exchangeable laws based on partition trees

The present chapter contains the description of a class of laws for N–exchangeable sequences defined by means of Theorem 3.4, with the law of $\tilde{e}$ assessed according to the same idea of *partitions tree* as the one used to introduce the Pólya-tree distributions. See, for example, Ferguson (1974) and Mauldin et al. (1992). 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, contrary 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 associated to 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 4.1 (Species sampling)**.** In the species sampling problem from a community of animals, one can consider a finite community of N units and identify 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. (see Fig. 4.1).

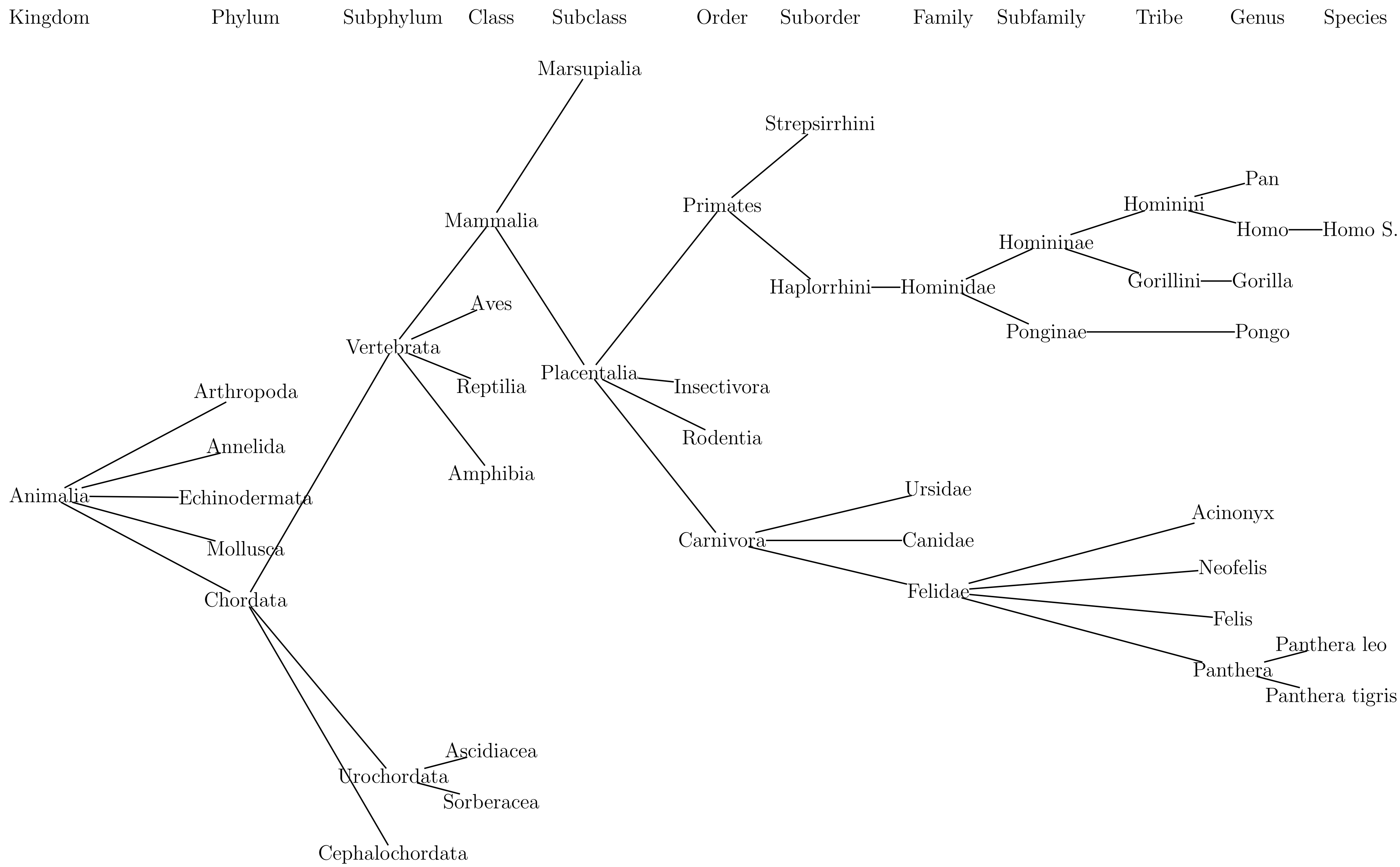


Figure 4.1: Part of the scientific taxonomy.

One can use the partitions tree structure intrinsic to this classification process to assign the probability distribution of $(\xi_1, \dots, \xi_N)$, where ξ_i denotes the species of the i–th animal in the community. After assessing $\mathcal{L}_{(\xi_1,\dots,\xi_N)}$, one can, for instance, estimate the number of distinct species, their frequencies, the possible existence of unknown species, etc. 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 species. So, for instance, given that in a sample of n animals, n_1 mammals are dectected, n_{11} of which being carnivourous, it could 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_{11}/n_1 increases. $\diamondsuit$

4.1 Definition of the model

As mentioned at the beginning of this chapter, the tree structure is used to assess the law of $\tilde{e}$. It is time to define, in general terms, the method sketched in the previous example. Consider a separating tree of partitions – say Π – of some space $\mathbb{X}$ that includes the range of each of the random variables $\xi_1, \dots, \xi_N$. This means that Π is a sequence $\{\pi_m\}_{m=0}^{\infty}$ of ordered, finite partitions of $\mathbb{X}$ such that $\pi_0 := \{\mathbb{X}\}$ and π_{m+1} is a refinement of π_m for every $m \geq 0$. Here $\mathscr{X}$ and $\mathscr{A}$ stand for the algebra and the σ-algebra, respectively, generated by $\mathscr{G} := \cup_{m\geq 0}\pi_m$. Resorting to the notation introduced in Section 2.1, denote by $B_{m,\,1}, \dots, B_{m,\,k_m}$ the elements of partition π_m. By the way, with reference to Example 4.1, the sets $B_{m,\,1}, \dots, B_{m,\,k_m}$ play the role of taxonomic units. As in Section 2.1, indicate the *most recent superset* of $B \in \pi_m$ by $\mathrm{ge}(B) \in \pi_{m-1}$, i.e. the set C in π_{m-1} that includes B. In addition, those sets in π_{m+1} that are included by $C \in \pi_m$ will be called *descendants of* C. For each B in $\mathscr{G}$, define $\tilde{N}(B)$ to be the (random) number of elements $(\xi_1, \dots, \xi_N)$ contained in B, i.e. $\tilde{N}(B) = N\tilde{e}_N(B)$, and write

$$\tilde{N}_0 := \tilde{N}(\mathbb{X}) = N$$
$$\tilde{N}_m := (\tilde{N}(B_{m,1}), \dots, \tilde{N}(B_{m,k_m})) \quad m = 1, 2, \dots .$$

The distribution of the random sequence $(\tilde{N}_m)_{m\geq 0}$ completely characterizes the exchangeable probability measure P, as it was shown in Section 3.2.3 applying Theorem 3.4. Moreover, recall that the sequence $(\tilde{N}_m)_{m\geq 0}$ meets the Markov property, since

$$\sum_{B\colon \mathrm{ge}(B)=C} \tilde{N}(B) = \tilde{N}(C) \tag{4.1}$$

for each C in $\mathscr{G}$, and therefore the law of $(\tilde{N}_m)_{m\geq 0}$ is determined by the sequence of conditional distributions $(\mathcal{L}_{\tilde{N}_{m+1}|\tilde{N}_m})_{m\geq 0}$. Countable additivity of P implies that

$$\tilde{N}(C_n) \to 0 \text{ in law} \tag{4.2}$$

holds for any decreasing sequence $(C_n)_{n\geq 1}$ of events in $\mathscr{A}$ such that $C_n \downarrow \emptyset$. In Chapter 3, it was proved that, under some suitable conditions for $(\mathbb{X}, \mathscr{X})$, (4.1) and (4.2) are sufficient for the existence of the exchangeable probability measure P on $(\mathbb{X}^N, \mathscr{X}^N)$.

At this stage, one is in a position to describe the particular laws $\mathcal{L}_{\tilde{N}_m}$ $(m = 0, 1, \dots)$ that will be considered in the rest of this chapter. In point of fact they are strongly reminescent of the Pólya–tree distributions and allow forms of conjugate analysis, being destiguished by the fact that they satisfy the following condition:

Condition 4.1. For each m in $\mathbb{N}$,

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

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

4.1.3. For every B in π_{m+1} and any $m \geq 0$,

a. $\mathbb{E}(\tilde{N}(B) \mid \tilde{N}(\mathrm{ge}(B)))$ is a linear function of $\tilde{N}(\mathrm{ge}(B))$, P-a.s.,

b. for any $n < N$ and $A_1, \ldots, A_n$ in π_{m+1},

$$\mathbb{E}\left(\sum_{i=n+1}^{N} \delta_{\xi_i}(B) \mid \sum_{i=n+1}^{N} \delta_{\xi_i}(\mathrm{ge}(B)), \xi_1 \in A_1, \ldots, \xi_n \in A_n\right)$$

is a linear function of $\sum_{i=n+1}^{N} \delta_{\xi_i}(\mathrm{ge}(B))$, P-a.s..

It should be noted that conditions 4.1.1 – 4.1.2 are tantamount to assuming that

$$\mathcal{L}_{\tilde{N}_{m+1}|\tilde{N}_m} = \bigtimes_{C \in \pi_m} \mathcal{L}_{\tilde{N}(B):\mathrm{ge}(B)=C|\tilde{N}(C)}. \tag{4.3}$$

One can explain the value of Condition 4.1 by means of Example 4.1. The adoption, in Example 4.1, of a law which satisfies Condition 4.1 entails supposing that, conditionally on the knowledge of the frequency of a taxon C in π_m, any additional information on the frequencies of other elements of π_m, or of their descendants in π_{m+1} does not affect the prevision of frequencies of the subsets of C.

About Condition 4.1.3, notice that it can be reformulated saying that, for every B in π_{m+1} and any $m \geq 0$,

a. $\mathbb{E}(\frac{\tilde{N}(B)}{M} \mid \tilde{N}(\mathrm{ge}(B)) = M)$ does not depend on M, as M varies in the following set:

$$\Big\{j = 1, \ldots, N : P(\tilde{N}(\mathrm{ge}(B)) = j) > 0\Big\};$$

b. for any $n < N$ and $A_1, \ldots, A_n$ in π_{m+1},

$$\mathbb{E}\left(\frac{\sum_{i=n+1}^{N} \delta_{\xi_i}(B)}{M} \mid \sum_{i=n+1}^{N} \delta_{\xi_i}(\mathrm{ge}(B)) = M, \xi_1 \in A_1, \ldots, \xi_n \in A_n\right)$$

does not depend on M, as M varies in the following set:

$$\Big\{j = 1, \ldots, N : P\left(\xi_1 \in A_1, \ldots, \xi_n \in A_n, \textstyle\sum_{i=n+1}^{N} \delta_{\xi_i}(\mathrm{ge}(B)) = j\right) > 0\Big\}.$$

In relation with Example 4.1, Condition 4.1.3 requires that, given that the number of units in a taxon C of π_m is $M > 0$, the expected proportion of units that belong to some descendant B of C, among those that are in C, does not really depend on M. Moreover, if one also knows which taxa in π_{m+1} the n units sampled from the population belong to, the conditional expectation of the proportion of unobserved units that belong to some descendant B of C, among those that are in C, does not really depend on the number of unobserved units in C.

From now on, any exchangeable law for $(\xi_1, \dots, \xi_N)$ that satisfies Condition 4.1 will be called *partitions tree* distribution. In particular, the marginal law of any element of $(\xi_1, \dots, \xi_N)$ can be derived from $(\mathcal{L}_{\tilde{N}_{m+1}|\tilde{N}_m})_{m\geq 0}$, by exploiting Condition 4.1.3.a, i.e., for each B in $\mathscr{G}$ such that $\tilde{N}(B)$ is not degenerate at zero,

$$P\{\xi_1 \in B\} = \mathbb{E}\left(\frac{\tilde{N}(B_1)}{N}\right) \prod_{j=2}^{m} \mathbb{E}\left(\frac{\tilde{N}(B_j)}{M_{j-1}} \mid \tilde{N}(B_{j-1}) = M_{j-1}\right) \qquad (B \in \pi_m), \tag{4.4}$$

where $B_m = B$ and, for $j < m$, B_j denotes the set in π_j that contains B_m, and M_j is any positive value such that $P(\tilde{N}(B_j) = M_j)$ is positive.

The class of exchangeable laws considered here satisfies a nice property: a partitions tree distribution w.r.t. Π is a partitions tree w.r.t. any subsequence of Π. For the proof of this fact and of (4.4) refer to Appendix, pages 112 and pages 102, respectively.

Two examples of partitions tree distributions will be explained in Section 4.5.

4.2 Posterior and predictive distributions

This section contains some results on predictive and a posteriori distributions relating to partitions tree distributions. The following propositions are useful to determine the posterior distributions for $\tilde{e}$, i.e. the conditional distribution of $\tilde{e}$ given $\xi(n)$, when Condition 4.1 is in force.

Proposition 4.1. *If Condition 4.1 holds, then*

$$\mathcal{L}_{\tilde{N}_{m+1}|\tilde{N}_m,\,\xi_1,\dots,\xi_n} = \bigtimes_{j=1}^{k_m} \mathcal{L}_{\tilde{N}(B_{m+1,\,l}):\,B_{m+1,\,l}\subset B_{m,j}|\tilde{N}(B_{m,j});\ \xi_i:\,\xi_i\in B_{m,j}} \tag{4.5}$$

In other words, if the *prior* distribution of $\tilde{e}$ satisfies Condition 4.1, so does its *posterior* distribution, i.e. the conditional distribution of $\tilde{e}$ given $(\xi_1, \dots, \xi_n)$. In that sense, we can say that Condition 4.1 defines a conjugate model.

Moreover:

Proposition 4.2. *If Condition 4.1 holds, then, for any $n \leq N$ and any vector $(N_1, \dots, N_{k_m})$ of positive integers summing up to N,*

$$\begin{aligned} &P(\tilde{N}(B_{m,\,1}) = N_1, \dots, \tilde{N}(B_{m,\,k_m}) = N_{k_m} \mid \xi_1 = x_1, \dots, \xi_n = x_n) \\ &\qquad \overset{P\text{-}a.s.}{=} P(\tilde{N}(B_{m,\,1}) = N_1, \dots, \tilde{N}(B_{m,\,k_m}) = N_{k_m} \mid \xi_1 \in B_m^{x_1}, \dots, \xi_n \in B_m^{x_n}), \end{aligned} \tag{4.6}$$

where B_m^x denotes the set of π_m which x belongs to. Moreover

$$
\begin{aligned}
&P(\tilde{N}(B_{m,l_j}) = N_{m,l_j}, j = 1,\ldots,d \mid \tilde{N}(B_{m-1,l}) = M,\ \xi(n) = x(n)) \\
&\quad = P(\tilde{N}(B_{m,l_j}) = N_{m,l_j}, j = 1,\ldots,d \mid \tilde{N}(B_{m-1,l}) = M, \xi_1 \in B_m^{x_1},\ldots,\xi_n \in B_m^{x_n}) \qquad (4.7)
\end{aligned}
$$

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

For the proof of these propositions, see the Appendix.

Proposition 4.2 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 $\{\mathbb{I}_{B_{m,\,j}}(\xi_i):\ j = 1,\ldots,k_m,\ i = 1,\ldots,m\}$. This property, which process $\tilde{N}$ share with F-neutral processes, makes calculations for the posterior easy.

Applying Bayes' theorem to the right hand side of (4.6), one gets

$$
\begin{aligned}
&P(\tilde{N}(B_{m,\,1}) = N_1,\ldots,\tilde{N}(B_{m,\,k_m}) = N_{k_m} \mid \xi_1 \in B_m^{x_1},\ldots,\xi_n \in B_m^{x_n}) \\
&\qquad = \frac{P(\xi_1 \in B_m^{x_1},\ldots,\xi_n \in B_m^{x_n} \mid \tilde{N}(B_{m,\,1}) = N_1,\ldots,\tilde{N}(B_{m,\,k_m}) = N_{k_m})}{P\{\xi_1 \in B_m^{x_1},\ldots,\xi_n \in B_m^{x_n}\}} \\
&\qquad\qquad \cdot P(\tilde{N}(B_{m,\,1}) = N_1,\ldots,\tilde{N}(B_{m,\,k_m}) = N_{k_m}).
\end{aligned}
$$

Therefore, by Theorem 3.1 one obtains the following

Proposition 4.3. *If Condition 4.1 holds, then*

$$
\begin{aligned}
&P(\tilde{N}(B_{m,\,1}) = N_1,\ldots,\tilde{N}(B_{m,\,k_m}) = N_{k_m} \mid \xi(n) = x(n)) \\
&\propto \mathcal{H}_{N_1,\ldots,N_{k_m}}(n_1,\ldots,n_{k_m}) \cdot P(\tilde{N}(B_{m,\,1}) = N_1,\ldots,\tilde{N}(B_{m,\,k_m}) = N_{k_m}),
\end{aligned} \qquad (4.8)
$$

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

Applying Proposition 4.1 to (4.7) and arguing as for the proof of Proposition 4.3, one can see that

$$
\begin{aligned}
&P(\tilde{N}(B_{m,l_j}) = N_{m,l_j}, j = 1,\ldots,d \mid \tilde{N}(B_{m-1,l}) = M,\ \xi(n) = x(n)) \\
&\propto \mathcal{H}_{N_{l_1},\ldots,N_{l_d}}(n_{l_1},\ldots,n_{l_d}) \cdot P(\tilde{N}(B_{m,l_j}) = N_{m,l_j}, j = 1,\ldots,d \mid \tilde{N}(B_{m-1,l}) = M), \qquad (4.9)
\end{aligned}
$$

where $l_1 \leq \cdots \leq l_d$ are such that B_{m,l_j} is contained by $B_{m-1,\,l}$. Therefore, the posterior for $(\tilde{N}(B)\colon B \in \pi_m)$ is uniquely identified by the conditional probability of $(\tilde{N}(B_{m,j}) = N_{m,j}\colon B_{m,j} \subset B_{m-1,l})$ given that $\xi(n) = x(n)$ and $\tilde{N}(B_{m-1,l}) = M$, for each $l = 1,\ldots,k_{m-1}$.

4.3 Partition tree laws with absolute continuous marginal and predictive distributions

By (4.4), one can assess the laws $\mathfrak{L}_{\tilde{N}_{m+1}|\tilde{N}_m}$ $(m = 1, 2, \dots)$ so that the law $\mathfrak{L}_{\xi_1}$ of the single observation turns out to be a desired distribution. Moreover, it is not difficult to find a sufficient condition under which $\mathfrak{L}_{\xi_1}$ is absolutely continuous w.r.t. some measure (as Lebesgue measure).

Lemma 4.4. *Let μ be a σ-finite measure on $(\mathbb{X}, \mathscr{X})$. Assume that each element of each π_m has positive μ measure. For each x in $\mathbb{X}$, define:*

$$f_{(m)}(x) = \begin{cases} \frac{1}{\mu(B_m^x)} \prod_{k=1}^m \mathbb{E}\left(\frac{\tilde{N}(B_k^x)}{\tilde{N}(B_{k-1}^x)} \mid \tilde{N}(B_{k-1}^x) = M_{k-1}\right) & \text{if} \quad P(\tilde{N}(B_m^x) = 0) < 1 \\ 0 & \text{otherwise,} \end{cases} \tag{4.10}$$

where, for each $k \geq 0$, B_k^x denotes the set of π_k which x belongs to, and $M_k > 0$ is such that the probability $P(\tilde{N}(B_k) = M_k)$ is positive.
If $\lim_{m\to\infty} f_{(m)}(x) = f(x)$ a.e.-μ, and $\int_{\mathbb{X}} f \mathrm{d}\mu = 1$, then $\mathfrak{L}_{\xi_1} \ll \mu$ and $f = \mathrm{d}\mathfrak{L}_{\xi_1}/\mathrm{d}\mu$.

The proof as well as the statement of this lemma is similar to that one of Lemma 1.113 in Schervish (1995) (about F-neutral processes).

Proof. We need to prove that for each $B \in \mathscr{G}$, $\mathfrak{L}_{\xi_1}(B) = \int_B f(x)\mathrm{d}\mu(x)$. The extension to $\mathscr{X}$ is straightforward.

Let $B \in \pi_m$. By (4.4), we have, for each $x \in B$,

$$\mathfrak{L}_{\xi_1}(B) = \mu(B) f_{(m)}(x) = \int_B f_{(m)}(x)\mathrm{d}\mu(x).$$

For $k > m$, write $B = \cup_{\alpha\in A} D_\alpha$ as the partition of B by elements of π_k. Since $f_{(k)}$ is constant on each D_α, we can write:

$$\int_B f_{(k)}(x)\mathrm{d}\mu(x) = \sum_{\alpha\in A} \int_{D_\alpha} f_{(k)}(x)\mathrm{d}\mu(x) = \sum_{\alpha\in A} \mathfrak{L}_{\xi_1}(D_\alpha) = \mathfrak{L}_{\xi_1}(B).$$

Hence we have

$$\int_B f_{(k)}(x)\mathrm{d}\mu(x) = \mathfrak{L}_{\xi_1}(B) \tag{4.11}$$

for all $k \geq m$. So,

$$\lim_{k\to\infty} \int_B f_{(k)}(x)\mathrm{d}\mu(x) = \mathcal{L}_{\xi_1}(B).$$

It is known that, by Scheffé's lemma [see Schervish (1995), page 634–635], this implies:

$$\lim_{k\to\infty} f_{(k)}(x)\mathrm{d}\mu(x) = \int_B f(x)\mathrm{d}\mu(x),$$

which the thesis follows from. □

The following proposition, which provides a rule to calculate the predictive density w.r.t. a diffuse measure, is a generalization of an analogous result stated about Pólya–tree processes by Lavine (1992).

Proposition 4.5. *Let μ be a σ-finite diffuse measure on $(\mathbb{X}, \mathscr{X})$. Assume that $\mathcal{L}_{\xi_1}$ and $\mathcal{L}_{\xi_{n+1}|\xi(n)}$ are absolutely continuous w.r.t. μ, for some $n \leq N$. Let $f_0 = \mathrm{d}\mathcal{L}_{\xi_1}/\mathrm{d}\mu$.*
Hence, (a version of) the density of $\mathcal{L}_{\xi_{n+1}|\xi(n)}$, for each $x \notin \{\xi_1, \ldots, \xi_n\}$, is given by:

$$f_n(\xi(n); x) = \frac{\mathcal{L}_{\xi_{n+1}|\xi(n)}(B_m^x)}{\mathcal{L}_{\xi_{n+1}}(B_m^x)} f_0(x),$$

where m is such that $B_m^{\xi_i} \neq B_m^x$ – i.e. x and ξ_i do not belong to the same element of π_m – for each i.

Proof. By Lemma 4.4, a density for the predictive distribution is given by:

$$f_{n+1}(\xi(n); x) = \lim_{h\to\infty} \frac{P(\xi_{n+1} \in B_h^x \mid \xi(n))}{\mu(B_h^x)}$$

that, by Proposition 4.2, is the same as:

$$\begin{aligned} f_n(x(n); x) &= \lim_{\substack{h\to\infty \\ h>m}} \frac{P(\xi_{n+1} \in B_h^x \mid \xi(n) \in B_h^{x_1} \times \ldots \times B_h^{x_n})}{\mu(B_h^x)} \\ &= P(\xi_{n+1} \in B_m^x \mid \xi(n) \in B_m^{x_1} \times \ldots \times B_m^{x_n}) \qquad (4.12) \\ &\quad \cdot \lim_{\substack{h\to\infty \\ h>m}} \frac{P(\xi_{n+1} \in B_h^x \mid \xi_{n+1} \in B_m^x, \, \xi(n) \in B_h^{x_1} \times \ldots \times B_h^{x_n})}{\mu(B_h^x)}. \end{aligned}$$

Proposition A.14 entails that:

$$\begin{aligned} f_n(x(n);x) &= P(\xi_{n+1} \in B_m^x \mid \xi(n) \in B_m^{x_1} \times \ldots \times B_m^{x_n}) \cdot \lim_{\substack{h\to\infty \\ h>m}} \frac{P(\xi_{n+1} \in B_h^x \mid \xi_{n+1} \in B_m^x)}{\mu(B_h^x)} \\ &= \frac{P(\xi_{n+1} \in B_m^x \mid \xi(n) \in B_m^{x_1} \times \ldots \times B_m^{x_n})}{P(\xi_{n+1} \in B_m^x)} \cdot \lim_{\substack{h\to\infty \\ h>m}} \frac{P(\xi_{n+1} \in B_h^x)}{\mu(B_h^x)} \\ &= \frac{P(\xi_{n+1} \in B_m^x \mid \xi(n) \in B_m^{x_1} \times \ldots \times B_m^{x_n})}{P(\xi_{n+1} \in B_m^x)} f_0(x). \end{aligned} \tag{4.13}$$

□

4.4 Construction of a partitions tree law

In this section, a general procedure is described to construct a partition tree distribution, assessing the conditional laws $\mathcal{L}_{\tilde{N}_{m+1}|\tilde{N}_m}$ for each $m \geq 0$.

Let $Y(\mathbb{X}) = (Y_1(\mathbb{X}), \ldots, Y_N(\mathbb{X}))$ be an exchangeable random vector such that each $Y_i(\mathbb{X})$ takes value into $\{1, \ldots, k_1\}$. For $j = 1, \ldots, k_m$, denote by $M^*_{B_{1,\,j}}$ the maximum value that the random variable $\sum_{i=1}^N \mathbb{I}_{\{Y_i(\mathbb{X})=j\}}$ assumes with positive probability. With a recursive procedure as $m = 1, 2, \ldots$, for each $C \in \pi_m$ such that $M^*_C > 0$, define $Y(C)$ to be another exchangeable random vector $(Y_1(C), \ldots, Y_{M^*_C}(C))$ such that each $Y_i(C)$ belongs to $\{j = 1, \ldots, k_{m+1} : B_{m+1,\,j} \subset C\}$, and define

$$M^*_{B_{m+1,\,j}} := \max\{M = 0, \ldots, N : P'\left(\textstyle\sum_{i=1}^{M^*_C} \mathbb{I}_{\{Y_i(C)=j\}} = M\right) > 0\},$$

for each j such that $B_{m+1,\,j} \subset C$, being P' the probability defined on the space that supports all the $Y(C)$'s. If $M^*_C = 0$, set $Y(C) = 0$ P'-a.s. and $M^*_D = 0$ for each descendant D of C. One can take, for instance, each vector $Y(C)$ (with $M^*_C > 0$) to be the outcome of M^*_C drawings from a urn according to some particular scheme.

Finally, for each m and each $C \in \pi_m$, set:

$$\mathcal{L}_{(\tilde{N}(B_{m+1,\,j}):\mathrm{ge}(B_{m+1,\,j})=C)|\tilde{N}(C)} = \mathcal{L}_{\left(\sum_{i=1}^{\tilde{N}(C)} \mathbb{I}_{\{Y_i(C)=j\}}:\,\mathrm{ge}(B_{m+1,\,j})=C\right)}. \tag{4.14}$$

In order to better explain how (4.14) can be interpreted, 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 $(\mathbb{X}^N, \mathscr{X}^N, P)$ 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 ξ_i falls in $B_{m,j}$.

Note that, by exchangeability, if $D_1, \dots, D_h$ are the descendants of C, then

$$\begin{aligned}
&P(\tilde{N}(D_1) = N_1, \dots, \tilde{N}(D_h) = N_h \mid \tilde{N}(C) = M) \\
&\qquad = \frac{P(\tilde{N}(D_1) = N_1, \dots, \tilde{N}(D_h) = N_h, \tilde{N}(C^c) = N - M)}{P(\tilde{N}(C) = M)} \\
&\qquad = \frac{\binom{N}{N_1 \dots N_h \; N-M} P(F_M \cap E_M)}{\binom{N}{M} P(E_M)} = \binom{M}{N_1 \dots N_h} P(F_M \mid E_M),
\end{aligned}$$

where $M = N_1 + \cdots + N_h \leq N$,

$$E_M = \{\xi_1 \in C, \dots, \xi_M \in C, \xi_{M+1} \in C^c, \dots, \xi_N \in C^c\},$$

$$F_M = \{\xi_1 \in D_1, \dots, \xi_{N_1} \in D_1, \xi_{N_1+1} \in D_2, \dots, \xi_M \in D_h\},$$

and therefore

$$E_M \cap F_M = \{\xi_1 \in D_1, \dots, \xi_{N_1} \in D_1, \xi_{N_1+1} \in D_2, \dots, \xi_M \in D_h, \xi_{M+1} \in C^c, \dots, \xi_N \in C^c\}.$$

Hence, (4.14) can be rewritten as

$$P(F_M \mid E_M) = P'(Y_1(C) = j_1, \dots, Y_M(C) = j_M), \tag{4.15}$$

where $(j_1, \dots, j_M)$ is the vector such that $B_{m+1, j_i} = D_i$ for $i = 1, \dots, M$, or equivalently:

$$P(W_{m+1,1} = j_1, \dots, W_{m+1,M} = j_M \mid E_M) = P'(Y_1(C) = j_1, \dots, Y_M(C) = j_M). \tag{4.16}$$

Therefore, (4.15) is tantamount to saying that, for any $(j_1, \dots, j_M)$ such that $\mathrm{ge}(B_{m+1,j_i}) = C$ $(i = 1, \dots, M)$ and $P(\tilde{N}(C) = M) > 0$, and for any $1 \leq n \leq M$,

$$P(W_{m+1,1} = j_1, \dots, W_{m+1,n} = j_n \mid E_M) = P'(Y_1(C) = j_1, \dots, Y_n(C) = j_n). \tag{4.17}$$

Notice that, for any fixed $(j_1, \dots, j_n)$, the right hand side of (4.17) does not depend on $(M-n)$(for any $M \geq n$ such that $P(E_M)$ is positive). Hence, by Proposition A.13 (ii), (4.14) implies Condition 4.1.3, as long as Conditions 4.1.1– 4.1.2 are also satisfied.

On the other hand, if Conditions 4.1.1– 4.1.3 hold, then, for each C in $\mathscr{G}$, there exists a random vector $Y(C)$ that satisfies (4.14). In fact, by Proposition A.13, under Conditions

4.1.1– 4.1.3,

$$\begin{aligned}
&P(W_{m+1,1} = j_1, \dots, W_{m+1,n} = j_n \mid E_M)\\
&\qquad = P(W_{m+1,1} = j_1, \dots, W_{m+1,n} = j_n \mid E_{M_C^*})\\
&\qquad = \sum_{j_{n+1}=1}^{k_{m+1}} \cdots \sum_{j_{M_C^*}=1}^{k_{m+1}} P(W_{m+1,1} = j_1, \dots, W_{m+1,M_C^*} = j_{M_C^*} \mid E_{M_C^*})\\
&\qquad = \sum_{j_{n+1}=1}^{k_{m+1}} \cdots \sum_{j_{M_C^*}=1}^{k_{m+1}} P(W_{m+1,1} = j_1, \dots, W_{m+1,M_C^*} = j_{M_C^*} \mid \xi_1 \in C, \dots, \xi_{M_C^*} \in C),
\end{aligned}$$

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

$$M_C^* := \max\Big\{j \ge 0 : P(\tilde{N}(C) = j) > 0\Big\}.$$

Therefore, (4.17) is satisfied if, for each C in $\mathscr{G}$ such that $P(\tilde{N}(C) = 0) < 1$, i.e.

$$1 \le M_C^* := \max\Big\{M \ge 0 : P(\tilde{N}(C) = M) > 0\Big\},$$

$Y(C) := (Y_1(C), \dots, Y_{M_C^*}(C))$ is such that

$$\begin{aligned}
&P'(Y_1(C) = j_1, \dots, Y_{M_C^*}(C) = j_{M_C^*})\\
&\qquad = P(W_{m+1,1} = j_1, \dots, W_{m+1,M_C^*} = j_{M_C^*} \mid \xi_1 \in C, \dots, \xi_{M_C^*} \in C),
\end{aligned}$$

for any $(j_1, \dots, j_{M_C^*})$ such that $\mathrm{ge}(B_{m+1,j_i}) = C$ $(i = 1, \dots, M_C^*)$.

Notice that in order to define the law of $\tilde{e}$ by (4.14), it is enough to assess the p.m.f. of $(\sum_{i=1}^{M_C^*} \mathbb{I}_{\{Y_i(C)=j\}} : \mathrm{ge}(B_{m+1,j}) = C)$, for each m and C in π_m (without any particular requirement about the joint distribution of vectors $Y(C)$). In this way, in fact, the distribution of each exchangeable vector $Y(C)$ (by means of the law of its empirical measure) is also determined. In conclusion, if we denote by $\psi(M;\cdot)$ the joint p.m.f. of $(\sum_{i=1}^{M} \mathbb{I}_{\{Y_i(C)=j\}} : \mathrm{ge}(B_{m+1,j}) = C)$, then (4.14) can be reformulated in terms of $\psi(M;\cdot)$, making the $Y_i(C)$ disappear, as in the following proposition, which summarizes some of the above considerations.

Proposition 4.6. *Assume that Conditions 4.1.1 and 4.1.2 hold.*

For each $m \ge 0$ and each $C \in \pi_m$, denote $h_C := |\{B \in \pi_m : \mathrm{ge}(B) = C\}|$, and let ψ_C be a function from $\{0, \dots, N\} \times \{0, \dots, N\}^{h_C}$ into $[0,1]$ such that

$$\psi_C(\tilde{N}(C); N_1, \dots, N_{h_C}) = P(\tilde{N}(D_1) = N_1, \dots, \tilde{N}(D_{h_C}) = N_{h_C} \mid \tilde{N}(C)) \quad P-a.s.$$

if $D_1, \dots, D_{h_C}$ are the descendants *of C, i.e. the sets in π_{m+1} contained by C.*

Condition 4.1.3 is satisfied, if and only if for each $C \in \mathscr{G}$ with $\tilde{N}(C)$ not degenerate at zero and for each $(N_1, \dots, N_{h_C})$ such that $N_1 + \cdots + N_{h_C} = M \leq M_C^*$,

$$\psi_C(M; N_1, \dots, N_{h_C}) = \sum_{M_1, \dots, M_{h_C}} \mathcal{H}_{M_1, \dots, M_{h_C}}(N_1, \dots, N_{h_C})\, \psi_C(M_C^*; M_1, \dots, M_{h_C}) \tag{4.18}$$

where the sum runs over all vectors $(M_1, \dots, M_{h_C})$ such that $M_1 + \cdots + M_{h_C} = M_C^*$ and $M_C^* := \max\{M = 0, \dots, N : P(\tilde{N}(C) = M) > 0\}$.

So, in order to specify a partition tree distribution, one needs to assess only the conditional distribution of $(\tilde{N}(B) : \mathrm{ge}(B) = C)$ given the event $\{\tilde{N}(C) = M_C^*\}$ for each C in $\mathscr{G}$.

Remark 4.7. Note that the $Y_i(C)$ can be used to define an algorithm to generate $\xi_1, \dots, \xi_N$, setting:

$$\begin{aligned}
J_{1,i} &= Y_i(\mathbb{X}) && \text{for } i = 1, \dots, N \\
J_{m+1,1} &= Y_1(B_{m,r}) && \text{if } J_{m,1} = r \\
J_{m+1,i} &= Y_{l+1}(B_{m,r}) && \text{if } J_{m,i} = r \quad \text{and} \quad l = |\{h < i : J_{m,h} = r\}| \qquad (i = 2, \dots, N)
\end{aligned}$$

for $m = 1, 2, \dots,$ and then putting $\xi_i \in \cap_m B_{m,\, J_{m,i}}$. If singletons are measurable and (4.2) holds, each intersection $\cap_m B_{m,\, J_{m,i}}$ reduces to one point (with P-probability one).

4.5 A couple of partitions tree distributions

It is immediate to verify Condition 4.1 when $(\xi_1, \dots, \xi_N)$ are independent and identically distributed. More interesting classes of distributions can be constructed according to familiar urn schemes.

4.5.1 Hypergeometric partitions tree distributions

Let $\mathbb{X}$ be the interval $(0, 1]$, and let $\mathscr{X}$ denote its Borel sigma-algebra. Put $E = \{0, 1\}$ and $E^0 := \emptyset$, $E^* := \cup_{m=0}^{\infty} E^m$. Define π_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

$$I_{\varepsilon_1 \dots \varepsilon_m} := \left(\sum_{j=1}^m \varepsilon_j 2^{-j}, \quad \sum_{j=1}^m \varepsilon_j 2^{-j} + 2^{-m} \right]$$

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

In this case, Π is a binary tree and, therefore, if we assume Conditions 4.1.1 and 4.1.2, the exchangeable law P of $(\xi_1, \dots, \xi_N)$ 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 $\{0, \dots, \tilde{N}(I_\varepsilon)\}$. As it was explained in Section 3.2.3, no other conditions are necessary for the existence of a finitely-additive (exchangeable) probability on $\mathscr{A}^N$, consistently with such assessment. In order to be able to define a completely additive probability on $\mathbb{X}^N$, one can resort to Proposition 3.6, and one realizes that for the existence of P it suffices that Condition 4.1.1 also holds together with

$$\prod_{k=2}^{\infty} \mathbb{E} \left(\frac{\tilde{N}(I_{\varepsilon_1 \dots \varepsilon_k})}{\tilde{N}(I_{\varepsilon_1 \dots \varepsilon_{k-1}})} \,\middle|\, \tilde{N}(I_{\varepsilon_1 \dots \varepsilon_{k-1}}) = M_{k-1} \right) = 0 \tag{4.19}$$

for any zero-one sequence $\varepsilon_1, \varepsilon_2, \dots$ that is definitely zero and such that $\tilde{N}(I_{\varepsilon_1 \dots \varepsilon_k})$ is not degenerate at zero for each $k \geq 1$ and for any sequence $M_1, M_2, \dots$ of positive integers such that $P(\tilde{N}(I_{\varepsilon_1 \dots \varepsilon_k}) = M_k) > 0$ for each $k \geq 1$.

Now it will be presented a possible assessment of the conditional distribution of $\tilde{N}(I_{\varepsilon 1})$ given $\tilde{N}(I_\varepsilon)$ for each $\varepsilon \in E^*$. Introduce a set of nonnegative integers $\aleph := \{\alpha_\varepsilon \in \mathbb{N} : \varepsilon \in E^*\}$ with $\alpha_\emptyset := N$ satisfying

$$\alpha_{\varepsilon 0} + \alpha_{\varepsilon 1} \geq \alpha_\varepsilon \tag{4.20}$$

for every $\varepsilon \in E^*$. At this stage, we propose to 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 of $\alpha_{\varepsilon 1} + \alpha_{\varepsilon 0}$ balls, of which $\alpha_{\varepsilon 1}$ are white and $\alpha_{\varepsilon 0}$ are black.

In concrete terms, the process $\tilde{N}$ may be generated according to the following scheme:

- Draw N balls without replacement from an urn of α_1 white balls and α_0 black balls, and suppose you get N_1 white balls and $N_0 := N - N_1$ black balls.

- Now draw without replacement N_1 balls from an urn of α_{11} white balls and α_{10} black balls and N_0 balls from an urn of α_{01} white balls and α_{00} black balls, respectively;

suppose the former sample contains N_{11} and N_{01} white balls, while the latter contains N_{10} and N_{00} black balls.

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

- N_ε gives the number of observations that belong to I_ε for each $\varepsilon \in E^*$.

Note that at the step $(m+1)$ the total number of balls in each urn must be greater than the number of trials. Since the number of trials N_ε at the m-th step is less than or equal to the number of balls of the corresponding color in the urn at the $(m-1)$-th step – which is α_ε – everything makes sense whenever (4.20) holds true.

The aforesaid procedure gives rise to a unique exchangeable finitely-additive probability on the algebra $\mathscr{A}^N$. As to the existence of a unique σ–additive extension P of such a probability to $\mathscr{X}^N$, note that (4.19) becomes

$$\lim_{m\to\infty} \prod_{j=1}^{m} \frac{\alpha_{\varepsilon_1 \ldots \varepsilon_{j+1}}}{\alpha_{\varepsilon_1 \ldots \varepsilon_j 0} + \alpha_{\varepsilon_1 \ldots \varepsilon_j 1}} = 0 \tag{4.21}$$

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

If the set $\aleph$ meets (4.20) and (4.21) 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 $\mathscr{H}(\aleph)$.

Observe that if $\alpha_{\varepsilon 0} + \alpha_{\varepsilon 1} = \alpha_\varepsilon$ for each $\varepsilon \in E^*$, then $(\xi_1, \ldots, \xi_N)$ is distributed like in a drawing without replacement from an urn of $(\alpha_0 + \alpha_1)$ 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, with the proviso that a point x that has two expansions takes the nonterminating one.

It should be noted that the urn scheme described above 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. To this end, it is enough, at each step m and for each ϵ in E^{m-1}, to draw N_ϵ balls from an urn of $\alpha_{\epsilon 1}$ 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 ball of the color drawn). This modification of the scheme introduced above is equivalent to the scheme introduced by Mauldin et al. (1992) to generate a Pólya–tree process (see Section 2.1.4). In point of fact, by (4.19), one can see that (4.21) is also a necessary and sufficient condition for the existence of a Pólya–tree distribution with parameters α_ε, $\varepsilon \in E^*$. The analogous condition for the case $\mathbb{X} = \mathbb{R}$ can be found in Ghosh and Ramamoorthi (2003).

It is clear that any sequence $(\xi_1, \dots, \xi_N)$ distributed according to $\mathscr{H}(\aleph)$ is not infinitely prolongable, i.e. it is not distributed as the initial segment of any infinite exchangeable sequence. In fact, if it was prolongable, the sequence $\mathbb{I}_{\{\xi_1 \in (0,1/2]\}}, \dots, \mathbb{I}_{\{\xi_N \in (0,1/2]\}}$ would also be prolongeable, whilst it proves 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 ξ_i.

Proposition 4.8. *Under $\mathscr{H}(\aleph)$ the law of each ξ_i is given by*

$$P(\xi_1 \in I_\varepsilon) = \frac{\alpha_{\varepsilon_1}}{\alpha_0 + \alpha_1} \cdot \frac{\alpha_{\varepsilon_1\varepsilon_2}}{\alpha_{\varepsilon_1 0} + \alpha_{\varepsilon_1 1}} \cdots \frac{\alpha_{\varepsilon_1 \dots \varepsilon_m}}{\alpha_{\varepsilon_1 \dots \varepsilon_{m-1} 0} + \alpha_{\varepsilon_1 \dots \varepsilon_{m-1} 1}}, \tag{4.22}$$

for each $\varepsilon = (\varepsilon_1 \dots \varepsilon_m)$ in E^*, while the expectation is

$$\mathbb{E}(\xi_1) = \sum_{m=1}^{+\infty} 2^{-m} \sum_{\substack{\varepsilon \in E^m: \\ \varepsilon_m = 1}} \prod_{k=1}^{m} \frac{\alpha_{\varepsilon_1 \dots \varepsilon_k}}{\alpha_{\varepsilon_1 \dots \varepsilon_k 0} + \alpha_{\varepsilon_1 \dots \varepsilon_k 1}}. \tag{4.23}$$

Proof. Equation (4.22) follows by equation (4.4). In order to prove (4.23), denote by $d_k(x)$ the k-th binary digit of x, for every $x \in \mathbb{X} = (0, 1]$. Hence, for any $x \in (0, 1]$,

$$x = \sum_{m=1}^{+\infty} 2^{-m} d_m(x), \tag{4.24}$$

and then

$$\begin{aligned}
\mathbb{E}(\xi_1) &= \sum_{m=1}^{+\infty} 2^{-m}\, \mathbb{E}(d_m(\xi_1)) \\
&= \sum_{m=1}^{+\infty} 2^{-m} \sum_{\varepsilon \in E^{m-1}} P(\xi_1 \in I_{\varepsilon 1}) \\
&= \sum_{m=1}^{+\infty} 2^{-m} \sum_{\substack{\varepsilon \in E^m: \\ \varepsilon_m = 1}} \prod_{k=1}^{m} \frac{\alpha_{\varepsilon_1 \dots \varepsilon_k}}{\alpha_{\varepsilon_1 \dots \varepsilon_k 0} + \alpha_{\varepsilon_1 \dots \varepsilon_k 1}}.
\end{aligned}$$

□

Posterior and predictive distributions

Proposition 4.2 yields explicit forms for posterior and predictive distributions relating to hypergeometric partitions tree distributions. Suppose $\xi(n) := (\xi_1, \ldots, \xi_n)$ has been observed and by $\tilde{n}_\varepsilon$ denote the number of observations of the sample that fall into I_ε, that is $\tilde{n}_\varepsilon := \sum_{i=1}^n \delta_{\xi_i}(I_\varepsilon)$. Now, recall that the conditional distribution of $\tilde{N}(I_{\varepsilon 1})$ (with ε in E^*) given $\tilde{N}(I_\varepsilon)$ is the hypergeometric distribution relating to $\tilde{N}(I_\varepsilon)$ drawings when the initial number of white and black balls is $\alpha_{\varepsilon 1}$ and $\alpha_{\varepsilon 0}$, respectively. Therefore, applying Proposition 4.2, it is straightforward to see that the conditional law of $\sum_{i=n+1}^N \delta_{\xi_i}(I_{\varepsilon 1})$ given $(\tilde{N}(I_\varepsilon), \xi(n))$ is the hypergeometric distribution relating to $(\tilde{N}(I_\varepsilon) - \tilde{n}_\varepsilon)$ drawings when the initial number of white and black balls is $(\alpha_{\varepsilon 1} - n_{\varepsilon 1})$ and $(\alpha_{\varepsilon 0} - n_{\varepsilon 0})$, respectively. Thus one gets

Proposition 4.9. *Let the distribution of $(\xi_1, \ldots, \xi_N)$ be $\mathscr{H}(\aleph)$, where $\aleph = \{\alpha_\varepsilon : \varepsilon \in E^*\}$. Then the conditional distribution of $(\xi_{n+1}, \ldots, \xi_N)$ given $(\xi_1, \ldots, \xi_n)$ is $\mathscr{H}(\aleph^*)$ where $\aleph^* := \{\alpha^*_\varepsilon : \varepsilon \in E^*\}$ and $\alpha^*_\varepsilon := \alpha_\varepsilon - \sum_{i=1}^n \delta_{\xi_i}(I_\varepsilon)$ for each ε in E^*.*

In particular, for the predictive distribution, one has

$$\begin{aligned} &P(\xi_{n+1} \in I_{\varepsilon_1 \ldots \varepsilon_m 1} \mid \xi(n)) = \mathbb{E}\Big(\frac{\sum_{i=n+1}^N \delta_{\xi_i}(I_{\varepsilon_1 \ldots \varepsilon_m 1})}{N-n} \mid \xi(n)\Big) \\ &= \frac{\alpha_{\varepsilon_1} - \tilde{n}_{\varepsilon_1}}{\alpha_0 + \alpha_1 - n} \cdot \frac{\alpha_{\varepsilon_1 \varepsilon_2} - \tilde{n}_{\varepsilon_1 \varepsilon_2}}{\alpha_{\varepsilon_1 0} + \alpha_{\varepsilon_1 1} - \tilde{n}_{\varepsilon_1}} \cdots \frac{\alpha_{\varepsilon_1 \ldots \varepsilon_m 1} - \tilde{n}_{\varepsilon_1 \ldots \varepsilon_m 1}}{\alpha_{\varepsilon_1 \ldots \varepsilon_m 0} + \alpha_{\varepsilon_1 \ldots \varepsilon_m 1} - \tilde{n}_{\varepsilon_1 \ldots \varepsilon_m}}. \end{aligned} \tag{4.25}$$

Let us stress that this assessment for the joint law of $\xi_1, \ldots, \xi_N$ satisfies the requirements discussed at the beginning of this chapter, where we have considered, in Example 4.1, the species sampling problem from a community of animals. In this case, the observations range in the class of all species, and animals of the same species - but also of a similar species - compete with each other. In order to better clarify this point, consider one more example:

Example 4.2 (Workers' income)**.** Let us suppose we are doing a survey in a community of people about monthly incomes of workers. On the basis of the salary of the workers in a sample, we could be concerned with estimating, for instance, the average income, and the variability of the incomes in the community. Moreover, if we obtain informations about other variables in the sample – such as age, gender, job, seniority – we could study their effects on the salary, etc.

Assume that N is the number of workers in the community where our sample comes from, and by ξ_i denote the income of the i-th worker in the community (for $i = 1, \ldots, N$). In order to assign a probability distribution for $(\xi_1, \ldots, \xi_N)$, we can first consider the (random) number of workers who receive an income bigger than a fixed threshold and the number of those who do not. Then we can consider separately each group and split each one into two subgroups with respect to another income threshold and so on. In general at the k-th stage we obtain in this way 2^k income intervals, and each one of them will be split into two subintervals in the next stage. Of course, the financial resources are limited and the workers compete with each other to get a better income. Therefore, an inverse relation between predictions and observed frequencies seems appropriate in this case, too. $\diamondsuit$

Keeping in mind both Example 4.1 and Example 4.2, assume that a sample with just one unit is taken (i.e. only ξ_1 is observed). Then the predictive distribution of ξ_2 should put a lower mass around ξ_1 than the (unconditional) distribution $P \circ \xi_1^{-1}$ of the single observation. If $\xi_1 \in I_1 := (1/2, 1]$, then the conditional probability that the second observation falls in I_1 after having observed ξ_1 should be lower (w.r.t. the unconditional distribution) on the set I_1 and higher on $I_0 := (0, 1/2]$; if $\xi_1 \in I_0$, viceversa. In fact, in our model, such probability can be written as

$$P(\xi_2 \in I_1 \mid \xi_1) = p_1 P(\xi_1 \in I_1) - (1 - p_1)\mathbb{I}_{\{\xi_1 \in I_1\}},$$

where $p_1 = (\alpha_0 + \alpha_1)/(\alpha_0 + \alpha_1 - \mathbb{I}_{\{\xi_1 \in I_1\}})$.

Suppose more generally that we observed $\xi_1, \ldots, \xi_n$ and that we do not know exactly the species ξ_{n+1} of the $(n+1)$-th animal sampled, but we know only that it belongs to I_ε. In this case, the conditional probability that the $(n+1)$-th observation belongs to $I_{\varepsilon 1}$ turns out to be a linear combination of the conditional probability $P \circ \xi_1^{-1}$ and the conditional empirical measure $\tilde{e}_n$ of the sample for the same event given I_ε, that is:

$$\begin{aligned} &P(\xi_{n+1} \in I_{\varepsilon 1} \,|\xi_{(n)}, \xi_{n+1} \in I_\varepsilon) \\ &\qquad\qquad = p_{n+1} P(\xi_1 \in I_{\varepsilon 1} \mid \xi_1 \in I_\varepsilon) - (1 - p_{n+1}) \frac{\tilde{n}_{\varepsilon 1}}{\tilde{n}_\varepsilon} \end{aligned}$$

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

$$\begin{aligned} p_{n+1} &= \frac{\alpha_{\varepsilon 0} + \alpha_{\varepsilon 1}}{\alpha_{\varepsilon 0} + \alpha_{\varepsilon 1} - \tilde{n}_\varepsilon} \\ P(\xi_1 \in I_{\varepsilon 1} \mid \xi_1 \in I_\varepsilon) &= \frac{\alpha_{\varepsilon 1}}{\alpha_{\varepsilon 0} + \alpha_{\varepsilon 1}}. \end{aligned}$$

Possible generalizations

The hypergeometric partition tree distributions differ from Pólya–tree distributions even because they require more heavy conditions on the parameters. Anyway, with some precautions, their parameters can also be non-integers. In fact, for the hypergeometric p.m.f. $\binom{a}{x}\binom{b}{m-x}/\binom{a+b}{m}$ it is not essential that all the parameters m, a, b are positive: with certain restrictions we can take any two of them negative and the remaining one positive, and we still obtain a probability mass function. The conditions under which $\binom{a}{x}\binom{b}{m-x}/\binom{a+b}{m}$ provides a honest distribution, with m, a, b taking real values, have been investigated, for instance, by Kemp and Kemp (1956).

As far as we are concerned, it can be noted that $\binom{a}{x}\binom{b}{m-x}/\binom{a+b}{m}$ is a probability mass function on $\{0, \ldots, m\}$ if a and b are two real numbers such that one of the followings is true:

- $a + b \geq m$, and
  - if $a \leq m$, then a is a non-negative integer;
  - if $b \leq m$, then b is a non-negative integer;
- a and b are both negative.

Anyway, the support of such generalized distribution is

$$\{m \wedge (0 \vee (m - b)), \ldots, 0 \vee (m \wedge a)\}.$$

These facts can be easily proved taking the binomial expansion in $(1+x)^{a+b} = (1+x)^a(1+x)^b$. In this way, one obtains: $\sum_{x=0}^{m} \binom{a}{x}\binom{b}{m-x}/\binom{a+b}{m} = 1$. Hence, we can let $\{\alpha_\varepsilon : \varepsilon \in E^*\}$ be any set of real numbers such that $\alpha_\emptyset := N$ and, for each $\varepsilon \in E^*$, one of the following is true:

- $\alpha_{\varepsilon 0} + \alpha_{\varepsilon 1} \geq N \wedge \alpha_\varepsilon$ and if $\alpha_{\varepsilon i} < N \wedge \alpha_\varepsilon$, then $\alpha_{\varepsilon i}$ is a non-negative integer $(i = 0, 1)$;
- $\alpha_{\varepsilon 0}$ and $\alpha_{\varepsilon 1}$ are both negative.

Of course, if all the α_ε are negative, one obtains the distribution of an N-exchangeable sequence directed by a Pólya tree process.

4.5.2 Exchangeable sequences directed by Pólya-tree processes

As recalled in Section 4, Condition 4.1 is inspired by the theory of Pólya-tree processes. In point of fact, 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, then we shall show that the law of the empirical distribution of $(\xi_1, \ldots, \xi_N)$ satisfies Condition 4.1 for every N.

Proposition 4.10. *If $(\xi_n)_{n\geq 1}$ 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 $\mathcal{L}_{(\tilde{N}(B):\mathrm{ge}(B)=C)|\tilde{N}(C)}$ given by the following form of Dirichlet-compound multinomial distribution*

$$\begin{aligned} \mathcal{L}_{(\tilde{N}(B_{m,j}):\mathrm{ge}(B_{m,j})=C)|\tilde{N}(C)}(\{(N_{m,j} : B_{m,j} \subset C)\}) &= \frac{\prod_{j\in\mathcal{T}(C)} \binom{-\alpha_{m,j}}{N_{m,j}}}{\binom{-\sum_{j\in\mathcal{T}(C)}\alpha_{m,j}}{\tilde{N}(C)}} \\ = \frac{\tilde{N}(C)!}{\prod_{j\in\mathcal{T}(C)} N_{m,j}!}\ \frac{\prod_{j\in\mathcal{T}(C)}(\alpha_{m,j}+N_{m,j}-1)\cdots\alpha_{m,j}}{(\sum_{j\in\mathcal{T}(C)}\alpha_{m,j}+\tilde{N}(C)-1)\cdots\sum_{j\in\mathcal{T}(C)}\alpha_{m,j}} \end{aligned} \tag{4.26}$$

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

Moreover, for each $n = 0, \ldots, N$ and each $m \geq 1$, if C belongs to π_m and $\mathrm{ge}(B_{m,j}) = C$, then

$$\mathbb{E}\left(\sum_{i=n+1}^{N}\delta_{\xi_i}(B_{m,j})\,\middle|\,\sum_{i=n+1}^{N}\delta_{\xi_i}(C),\xi(n)\right) = \frac{\alpha_{m,j}+n\,\tilde{e}_n(B_{m,j})}{\sum\limits_{r\in\mathcal{T}(C)}\alpha_{m,r}+n\,\tilde{e}_n(C)}\sum_{i=n+1}^{N}\delta_{\xi_i}(C), \tag{4.27}$$

where $\tilde{e}_0 \equiv 0$ and $\xi(0) \equiv 0$ by convention.

Then, in view of (4.26)–(4.27), the distribution of $(\xi_1, \ldots, \xi_N)$ is a partitions tree distribution.

Proof. Assume, as usual, that $\xi_1, \ldots, \xi_N$ are the coordinate functions on $(\mathbb{X}^N, \mathscr{X}^N, P)$, 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 $(i_1, \ldots, i_N)$ integers in $\{1, \ldots, k_m\}$ such that $N_j = |\{l = 1, \ldots, k_m : i_l = j\}|$. Recalling the definition of Pólya-tree processes (given in Section 2.1), we can write:

$$\begin{aligned}
&P\{\xi_1 \in B_{m,i_1}, \dots, \xi_N \in B_{m,i_N}\} \\
&\quad= \mathbb{E}\left[\prod_{j=1}^{k_m} \tilde{p}(B_{m,j})^{N_j}\right] \\
&\quad= \mathbb{E}\left[\prod_{j=1}^{k_m} \tilde{p}(B_{m,j} \mid \mathrm{ge}(B_{m,j}))^{N_j} \cdot \prod_{j=1}^{k_m} \tilde{p}(\mathrm{ge}(B_{m,j}))^{N_j}\right] \\
&\quad= \mathbb{E}\left(\prod_{j=1}^{k_m} \tilde{p}(B_{m,j} \mid \mathrm{ge}(B_{m,j}))^{N_j}\right) \cdot \mathbb{E}\left(\prod_{j=1}^{k_m} \tilde{p}(\mathrm{ge}(B_{m,j}))^{N_j}\right) \\
&\quad= \prod_{C \in \pi_{m-1}} \mathbb{E}\left(\prod_{j \in \mathcal{T}(C)} \tilde{p}(B_{m,j} \mid C)^{N_j}\right) \\
&\qquad\quad \cdot \mathbb{E}\left(\prod_{C \in \pi_{m-1}} \tilde{p}(C)^{\sum_{j \in \mathcal{T}(C)} N_j}\right).
\end{aligned} \tag{4.28}$$

Observing that in the last term of (4.28) 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:

$$\begin{aligned}
P\{\xi_1 \in B_{m,i_1}, \dots, \xi_N \in B_{m,i_N}\} &= P\{\xi_1 \in \mathrm{ge}(B_{m,i_1}), \dots, \xi_N \in \mathrm{ge}(B_{m,i_N})\} \\
&\quad \cdot \prod_{C \in \pi_{m-1}} \left(\frac{\prod_{j \in \mathcal{T}(C)} \alpha_{m,j}^{[N_j]}}{(\sum_{j \in \mathcal{T}(C)} \alpha_{m,j})^{[\sum_{l \in \mathcal{T}(C)} N_l]}}\right),
\end{aligned} \tag{4.29}$$

where $a^{[h]} := a(a+1)\dots(a+h-1)$.

Note that in general,

$$\begin{aligned}
&P(\tilde{N}(B_{m,1}) = N_1, \dots, \tilde{N}(B_{m,k_m}) = N_{k_m} \mid \tilde{N}(C) = \textstyle\sum_{j \in \mathcal{T}(C)} N_j : C \in \pi_{m-1}) \\
&\quad= \frac{P(\tilde{N}(B_{m,1}) = N_1, \dots, \tilde{N}(B_{m,k_m}) = N_{k_m})}{P(\tilde{N}(C) = \sum_{j \in \mathcal{T}(C)} N_j : C \in \pi_{m-1})} \\
&\quad= \frac{\binom{N}{N_1, \dots, N_{k_m}} P\{\xi_1 \in B_{m,i_1}, \dots, \xi_N \in B_{m,i_N}\}}{\binom{N}{\sum_{j \in \mathcal{T}(C)} N_j : C \in \pi_{m-1}} P\{\xi_1 \in \mathrm{ge}(B_{m,i_1}), \dots, \xi_N \in \mathrm{ge}(B_{m,i_N})\}} \\
&\quad= \frac{P\{\xi_1 \in B_{m,i_1}, \dots, \xi_N \in B_{m,i_N}\}}{P\{\xi_1 \in \mathrm{ge}(B_{m,i_1}), \dots, \xi_N \in \mathrm{ge}(B_{m,i_N})\}} \prod_{C \in \pi_{m-1}} \binom{\sum_{j \in \mathcal{T}(C)} N_j}{N_j :\ j \in \mathcal{T}(C)}
\end{aligned} \tag{4.30}$$

and, therefore, combining (4.29) and (4.30) one realizes that the conditional law of $\tilde{N}_m$ given $\tilde{N}_{m-1}$ can be written as a product of measures:

$$\mathcal{L}_{\tilde{N}_m|\tilde{N}_{m-1}} = \bigtimes_{C\in\pi_{m-1}} \mathcal{L}_{\tilde{N}(B):\mathrm{ge}(B)=C|\tilde{N}(C)},$$

where each factor is given by (4.26). Hence, Conditions 4.1.1–4.1.2 hold true.

At this stage, note that (4.26) is the Dirichlet-compound multinomial distribution with parameter $(\tilde{N}(C); \alpha_{m,j} : j \in \mathcal{T}(C))$ and therefore its j-th marginal is the beta-binomial distribution with parameter $(\tilde{N}(C); \alpha_{m,j}, \sum_{l\neq j:\mathrm{ge}(B_{m,\,l})=C} \alpha_{m,\,l})$. Hence, one obtains (4.27) for $n = 1$, recalling that the expectation of a beta-binomial distribution with parameters $(M; \alpha, \beta)$ is $M \cdot \alpha/(\alpha + \beta)$. For the other cases, (4.27) follows by the fact that Pólya-tree processes are conjugate. Since (4.27) is a linear function of $\sum_{i=n+1}^{N} \delta_{\xi_i}(C)$, Condition 4.1.3 also holds true. $\square$

Chapter 5

Exchangeable laws based on random partitions

We intend to present one specific form of (finitary) exchangeable laws defined, aside from the standard conditional formulation, according to the characteristics of actual situations, and to work out some of their inherent statistical problems. The exchangeable law we wish to consider, rests on the concept of exchangeable random partition.

5.1 Introductory examples

It seems suitable to begin with the illustration of a couple of examples which could lead, in a natural way, to the use of random partitions in order to define laws of exchangeable vectors. The precise definition of this model, which, from now on, will be called *Random Partition Model* (RP-Model), is contained in Section 5.2.

Example 5.1 (Stochastic price system)**.** Let N subjects, labeled by $1, \ldots, N$, be consumers (or users) of a specific item (or service). Note that N could be thought of as a realization of a random number. Each of these subjects has the right to choose his own provider among M companies $C_1, \ldots, C_M$. Denote the provider chosen by the subject j with γ_j, and group together, into the same class, all the subjects who refer to the same provider. This gives rise to a partition of $\{1, \dots, N\}$, say

$$\tilde{\pi} = (\tilde{\pi}_1, \dots, \tilde{\pi}_\nu),$$

defined as follows:

$$\tilde{\pi}_1 := \{i_{1,1}, i_{1,2}, \dots\},$$

where $i_{1,1} = 1$ and subsequent elements $i_{1,2}, \dots$ are the labels of the subjects who apply to the provider of 1;

$$\tilde{\pi}_2 := \{i_{2,1}, i_{2,2}, \dots\},$$

where $i_{2,1}$ is the smallest of the labels not contained in $\tilde{\pi}_1$ and labels $i_{2,2}, \dots$ denote the subjects who get their supplies from the same provider of $i_{2,1}$, and so on. Notice that $\tilde{\pi}$ could be expressed as a function of $(\gamma_1, \dots, \gamma_N)$.

At this stage, one pairs each subject j with the unit price ξ_j fixed by γ_j. This way, each subject belonging to the same block $\tilde{\pi}_k$ is bound to pay the unit price ξ_k^* that amounts to

$$\xi_i = \xi_k^* \qquad (i \text{ in } \tilde{\pi}_k,\ k = 1, \dots, \nu). \tag{5.1}$$

$\diamondsuit$

Example 5.2 (Distribution of a chemical agent in a given population)**.** Consider N subjects who are allowed to drink water from sources $C_1, \dots, C_M$. Arguing in the same way as in the previous example, we obtain a partition of $\{1, \dots, N\}$, say $(\tilde{\pi}_1, \dots, \tilde{\pi}_\nu)$, where $\tilde{\pi}_1$ contains all the subjects who draw water from the same spring as 1, and so on, as in Example 5.1. Moreover, like in Example 5.1, associate each subject j with the concentration ξ_j of a specific bacteriological or chemical agent contained in the source γ_j of the water j drinks. Thus, (5.1) holds with the sole change that ξ_k^* now represents the concentration characterizing block $\tilde{\pi}_k$ in $\tilde{\pi}$. $\diamondsuit$

Our aim is to assess a probability law both for the price system $(\xi_1, \dots, \xi_N)$ in Example 5.1 and for the concentration vector $(\xi_1, \dots, \xi_N)$ in Example 5.2. To this purpose, we try to take advantage of the organization of the above descriptions. First, we assign a conditional probability law $\rho = \rho(\cdot|N)$ for π, given N, in such a way that it depends only on $\{|\tilde{\pi}_1|, \dots, |\tilde{\pi}_\nu|\}$, where, given a set A, $|A|$ stands for its cardinality. Then, we assess a conditional probability distribution for $\xi_1^*, \ldots, \xi_\nu^*$, given $(N, \tilde{\pi})$, so that $\xi_1^*, \ldots, \xi_\nu^*$ turn out to be conditionally independent with distributions depending only on the cardinalities of their respective blocks. At the end of this process we get the characterization of the probability law of the random vector $(\xi_1, \ldots, \xi_N, \tilde{\pi})$, and we can deduce the law of $(\xi_1, \ldots, \xi_N)$ from it.

With this distribution in hand, in Example 5.1 one could, for instance, initiate any kind of econometric analysis of the demand of the item, or service, taken into consideration. Analogously, in Example 5.2, that very same distribution could be used to forecast the spread of any disease that depends on the bacteriological or chemical agent at issue. To this end, the original probabilistic framework needs to be extended to $(\xi_1, \ldots, \xi_N, \tilde{\pi}, \eta_1, \ldots, \eta_N)$ where, with reference to Example 5.1, η_j represents the quantity requested by the j–subject and, in connection with Example 5.2, the same symbol could denote the result of a specific medical test on the j–th subject.

Now we show that $(\xi_1, \ldots, \xi_N)$ is exchangeable with respect to the law defined according to the above steps. Moreover, we shall describe how to assess $(\eta_1, \ldots, \eta_N)$ in such a way that $((\xi_1, \eta_1), \ldots, (\eta_N, \xi_N))$ turns out to be exchangeable.

5.2 Definition of the model

In order to propose a suitable probability law for problems of the same type as those sketched in the previous examples, it is worth providing an accurate description of the statistical data to be processed. As primary data we consider the vectors

$$(j, \gamma_j, \xi_j, \eta_j) \qquad j = 1, \ldots, N,$$

where γ_j, in $\{C_1, \ldots, C_M\}$, denotes an entity (provider, spring, etc.) which j decides to interact with, ξ_j is a characteristic (price, concentration, etc.) of the interplay between j and γ_j, and, finally, η_j is another characteristic of interest of j (the demand, etc). In Example 5.1, (j, γ_j, ξ_j) specifies the terms of a contract, while, in Example 5.2, (j, γ_j, ξ_j) can be thought of as the exposure to risk of subject j. As explained in the previous subsection, $(\gamma_1, \ldots, \gamma_N)$ induces a partition $\tilde{\pi}$ of $\{1, \ldots, N\}$ presented in such a way that $\tilde{\pi}_1$ contains subject 1, i.e.

$$\tilde{\pi}_1 = \{i : i \in \{1, \ldots, N\}, \gamma_i = \gamma_1\}$$

and, inductively, for $k = 2, 3, \dots, \nu$,

$$\tilde{\pi}_k = \{i : i \in \{1, \dots, N\}, \gamma_i = \gamma_{m(k)}\},$$

where $m(k) = \min\{i : i \in \{1, \dots, N\}, i \notin \cup_{j=1}^{k-1} \tilde{\pi}_j\}$ and $\nu = \min\{i : i \in \{1, \dots, N \wedge M\}, \cup_{j=1}^{i} \tilde{\pi}_j = \{1, \dots, N\}\}$. Notice that, if M is smaller than N, the class of partitions induced by $(\gamma_1, \dots, \gamma_N)$ is strictly contained in the class $\mathcal{P}_N$ of all partitions of $\{1, \dots, N\}$. We assume that the range of each (ξ_j, η_j) is contained in a product of complete separable metric spaces $\mathbb{X} \times \mathbb{Y}$ equipped with its Borel σ–algebra $\mathscr{X} \otimes \mathscr{Y}$. This extra–condition of a topological nature is required because of measurability problems connected, for example, with (5.1).

Our main goal is the assessment of a probability distribution for

$$\zeta = (\xi_1, \eta_1, \dots, \xi_N, \eta_N, \tilde{\pi}),$$

i.e. a probability on the product measurable space $(Z^N \times \mathcal{P}_N, \mathcal{Z}^N \otimes \mathcal{U})$, where $\mathcal{U}$ stands for the power set of $\mathcal{P}_N$ and $(Z, \mathcal{Z}) = (\mathbb{X} \times \mathbb{Y}, \mathscr{X} \otimes \mathscr{Y})$. We identify the j–th coordinate of $\Omega := Z^N \times \mathcal{P}_N$ with observation $\zeta_j := (\xi_j, \eta_j)$, namely

$$(x_j, y_j) = (\xi_j, \eta_j)(x_1, y_1 \dots, x_N, y_N, \pi) \qquad (x_1, y_1, \dots, x_N, y_N, \pi) \in \Omega$$

for $j = 1, \dots N$. Moreover, we define $\tilde{\pi}$ to be the $(N+1)$-th coordinate of the product space, that is

$$\pi = \tilde{\pi}(x_1, y_1, \dots, x_N, y_N, \pi) \qquad (x_1, y_1, \dots, x_N, y_N, \pi) \in \Omega.$$

Now we are in a position to specify the law of ζ. To this end, we begin with the definition of a distribution for $\tilde{\pi}$, say ρ. Then, we fix a conditional law $p(\cdot|\tilde{\pi})$ for $(\xi_1, \dots, \xi_N)$ given $\tilde{\pi}$ consistently with (5.1), and, finally, we fix a conditional law $q(\cdot|\tilde{\pi}, \xi_1, \dots, \xi_N)$ for $(\eta_1, \dots, \eta_N)$ given $\tilde{\pi}$ and $(\xi_1, \dots, \xi_N)$. It is enough to specify $p(\cdot|\tilde{\pi})$ for any measurable rectangle $A_1 \times \cdots \times A_N$, and, for the sake of definiteness, we assume that

$$p\left(A_1 \times \cdots \times A_N | \tilde{\pi}\right) := \prod_{k=1}^{|\tilde{\pi}|} \alpha_{|\tilde{\pi}_k|}(\cap_{i \in \tilde{\pi}_k} A_i), \tag{5.2}$$

where $|\tilde{\pi}|$ stands for the number of the elements of $\tilde{\pi}$, $|\tilde{\pi}_j|$ is the number of elements of the j–th block of $\tilde{\pi}$, and, finally, $\alpha_1, \dots, \alpha_N$ are probabilities on $(\mathbb{X}, \mathscr{X})$. The right–hand side of (5.2) reveals that we are assuming that the conditional distribution of ξ_r, $1 \le r \le N$, given

$\tilde{\pi}$ is $\alpha_{|\tilde{\pi}_{\phi(\tilde{\pi},r)}|}$, where $\phi(\tilde{\pi},r)$ denotes the index of the block of $\tilde{\pi}$ that contains r. Thus, this conditional law depends only on the cardinality of the block containing r. Since it is easy to check that

$$p\left(\xi_i = \xi_j \text{ for every } j \text{ in } \tilde{\pi}_{\phi(\tilde{\pi},i)} \mid \tilde{\pi}\right) = 1$$

for every $i = 1, \dots, N$, we can conclude that (5.2) is consistent with (5.1). Now, given $\tilde{\pi}$ and $(\xi_1, \dots, \xi_N)$, with $\tilde{\pi} = ([i_{1,1}, i_{1,2}, \dots, i_{1,k_1}], [i_{2,1}, i_{2,2}, \dots, i_{2,k_2}], \dots)$, a conditional law $q(\cdot|\tilde{\pi}, \xi_1, \dots, \xi_N)$ for $(\eta_1, \dots, \eta_N)$ is specified in such a way that $(\eta_{i_{j,1}}, \dots, \eta_{i_{j,k_j}})_{j=1,\dots,|\pi|}$ are independent, and, for each j, $(\eta_{i_{j,1}}, \dots, \eta_{i_{j,k_j}})$ is exchangeable with de Finetti measure that may depend on $\xi_{i_{j,1}}$. More formally, for every $B_1, \dots, B_N$ in $\mathcal{Y}$, we set

$$q(B_1 \times \cdots \times B_N | \tilde{\pi}, \xi_1, \dots, \xi_N) = \prod_{i=1}^{|\tilde{\pi}|} \kappa_{\xi_i^*}(\times_{j \in \tilde{\pi}_i} B_j), \tag{5.3}$$

where $\kappa_{\cdot}(\cdot)$ is an (exchangeable) transition kernel on $(\mathbb{X}, \mathscr{X}) \times (\mathbb{Y}^N, \mathcal{Y}^N)$, that is: for every x in $\mathbb{X}$, $\kappa_x(\cdot)$ is a probability measure on $(\mathbb{Y}^N, \mathcal{Y}^N)$, for every $B_1, \dots, B_N$ in $\mathcal{Y}$, $x \mapsto \kappa_x(B_1 \times \cdots \times B_N)$ is $\mathscr{X}/\mathcal{B}([0,1])$–measurable and

$$\kappa_x(B_1 \times \cdots \times B_N) = \kappa_x(B_{\sigma(1)} \times \cdots \times B_{\sigma(N)})$$

for every x in $\mathbb{X}$ and every permutation σ of $\{1, \dots, N\}$. Note that, for simplicity, from now on we write $\kappa_x(B_1 \times \cdots \times B_k)$ for $\kappa_x(B_1 \times \cdots \times B_k \times \mathbb{Y} \times \cdots \times \mathbb{Y})$ for every $k \leq N$. The simplest example is to take a product kernel, which is

$$\kappa_x(B_1 \times \cdots \times B_N) = \prod_{i=1}^{N} k_x^*(B_i).$$

At this stage it turns out that the distribution P of ζ is defined through

$$\begin{aligned} P\{\xi_1 \in A_1, \eta_1 \in B_1 \dots, \xi_N \in A_N, \eta_N \in B_N, \tilde{\pi} = \pi\} \\ = \rho(\pi) \prod_{i=1}^{|\pi|} \int_{\mathbb{X}} \mathbb{I}_{\cap_{j \in \pi_i} A_j}(x_i) \kappa_{x_i}(\times_{j \in \pi_i} B_j) \alpha_{|\pi_i|}(\mathrm{d}x_i), \end{aligned} \tag{5.4}$$

assumed to be valid for any $A_1, \dots, A_N$ in $\mathscr{X}$, every $B_1, \dots, B_N$ in $\mathcal{Y}$ and every π in $\mathcal{P}_N$. As a consequence, one obtains

$$P\{\xi_1 \in A_1, \dots, \xi_N \in A_N\} = \sum_{\pi \in \mathcal{P}_N} \rho(\pi) \prod_{k=1}^{|\pi|} \alpha_{|\pi_k|}(\cap_{i \in \pi_k} A_i), \tag{5.5}$$

and, hence,

$$P\{\xi_j \in \cdot\} = \sum_{l=1}^{N} w_{l,j}\alpha_l(\cdot) \tag{5.6}$$

with

$$w_{l,j} := P\{|\tilde{\pi}_{\phi(\tilde{\pi},j)}| = l\} = \sum_{\{\pi\in\mathcal{P}_n \,:\, |\pi_{\phi(\pi,j)}|=l\}} \rho(\pi) \qquad (l = 1,\dots,N).$$

At this point, we specify ρ in such a way that $(\xi_1,\eta_1),\dots,(\xi_N,\eta_N)$ is proved to be exchangeable. As already said, ρ will be defined so that its value at π depend only on the set of cardinalities of the blocks forming π. In other words, if for any π in $\mathcal{P}_N$ we define

$$h(\pi) = (|\{j : |\pi_j| = 1\}|,\dots,|\{j : |\pi_j| = N\}|)$$

then ρ must satisfy

$$\pi,\pi^* \in \mathcal{P}_N \text{ and } h(\pi) = h(\pi^*) \Rightarrow \rho(\pi) = \rho(\pi^*). \tag{5.7}$$

We present a simple example of distributions on $\mathcal{P}_N$ that meet (5.7).

Example 5.3. Let

$$\{A_\lambda(k) \geq 0, k = 1,\dots,N : \lambda \in \mathbb{R}\}, \qquad \{\beta_\lambda(k) \geq 0, k = 1,\dots,N : \lambda \in \mathbb{R}\}$$

be families of real functions and μ_N be a σ–finite measure on the Borel subsets of $\mathbb{R}$, such that

$$K_N := \sum_{\pi\in\mathcal{P}_N} \int_{\mathbb{R}} A_\lambda(|\pi|) \prod_{i=1}^{|\pi|} \beta_\lambda(|\pi_i|)\mu_N(\mathrm{d}\lambda)$$

turns out to be strictly positive and finite. Then

$$\rho(\pi) = K_N^{-1} \int_{\mathbb{R}} A_\lambda(|\pi|) \prod_{i=1}^{|\pi|} \beta_\lambda(|\pi_i|)\mu_N(\mathrm{d}\lambda)$$

gives a probability on $\mathcal{P}_N$. Since ρ can be written as

$$\rho(\pi) = K_N^{-1} \int_{\mathbb{R}} A_\lambda(|b|) \prod_{i=1}^{N} \beta_\lambda(i)^{b_i}\mu_N(\mathrm{d}\lambda)$$

with $b := h(\pi)$, it is easy to check that ρ meets (5.7). $\diamondsuit$

Proposition 5.1. *The random vector* $((\xi_1,\eta_1),\dots,(\xi_N,\eta_N))$ *from* Ω *into* $\mathbb{X}^N \times \mathbb{Y}^N$ *turns out to be exchangeable if its law is defined by* (5.4) *with any* ρ *satisfying* (5.7).

Proof. See the Appendix, on page 120. $\square$

The authors mentioned in Chapter 1 handle partitions $\Pi = \Pi(\xi_1, \ldots, \xi_N)$ *generated by sampling* from an exchangeable (infinite) sequence $(\xi_1, \ldots, \xi_n, \ldots)$, namely random partitions defined by the circumstance that i and j in $\{1, \ldots, N\}$ belong to the same block of Π if and only if $\xi_i = \xi_j$. It is easy to see that if $(\xi_n)_{n\geq 1}$ is an exchangeable sequence of random elements, then for any N the probability law on $\mathcal{P}_N$ of the partition generated by sampling from $(\xi_1, \ldots, \xi_N)$ meets (5.7). Moreover, it is easy to check that Π and $\tilde{\pi}$ have the same distribution whenever the hypotheses of Proposition 5.1 are in force and $\alpha_1, \ldots, \alpha_N$ are diffuse probabilities.

The Random Partition Model is related to species sampling sequences and normalized random measures with independent increments, which we already talked about in Sections 2.2 and 2.3 respectively. In fact, it is straightforward that the vector of the first N coordinates of a species sampling sequences satisfies the hypothesis of Proposition 5.1 with $\alpha_1 = \cdots = \alpha_N = \alpha$, for any N. Moreover, the same is true for an N-exchangeable sequence directed by a normalized random measure with independent increments such that $\nu(\mathrm{d}x\mathrm{d}v) = a\alpha(\mathrm{d}x)q(\mathrm{d}v)$. In the latter case, by (2.9) we can write:

$$\rho(\pi) = \frac{a^{|\pi|}}{\Gamma(N)} \int_{\mathbb{R}^+} \lambda^{N-1} L(\lambda) \prod_{j=1}^{|\pi|} \int_{\mathbb{R}^+} v^{|\pi_j|} e^{-\lambda v} q(\mathrm{d}v)\mathrm{d}\lambda, \tag{5.8}$$

where $L(\lambda) = \exp\{-a \int_{\mathbb{R}^+} (1 - e^{-\lambda v}) q(\mathrm{d}v)\}$.

5.2.1 Marginal distribution and correlation between observations

The first properties of (5.4) that we want to present concerns the distribution of each (ξ_j, η_j) and the correlation between ξ_i and ξ_j for $i \neq j$.

Proposition 5.2. *Let the law of the N-exchangeable sequence $(\xi_1, \ldots, \xi_N)$ be the same as in the previous proposition. Then, for any $j = 1, \ldots, N$,*

$$P\{\xi_j \in A\} = \sum_{l=1}^{N} w_l \ \alpha_l(A) =: \alpha_0(A) \qquad (A \in \mathscr{X}) \tag{5.9}$$

with

$$w_l := P\{|\tilde{\pi}_1| = l\} = \sum_{\{\pi\in\mathcal{P}_N \,:\, |\pi_1|=l\}} \rho(\pi) \qquad (l = 1, \ldots, N),$$

and

$$P\{\xi_j \in A, \eta_j \in B\} = \int_A \kappa_x(B)\alpha_0(\mathrm{d}x) \qquad (A \in \mathscr{X}, B \in \mathscr{Y}). \tag{5.10}$$

Let f and g be real–valued measurable functions defined on $(X, \mathscr{X})$, such that $\int_X (|f(x)| + |g(x)|)\alpha_j(\mathrm{d}x)$ is finite for every $j = 1, \dots, N$, then

$$\mathbb{E}(f(\xi_1)) = \int_{\mathbb{X}} f(x)\alpha_0(\mathrm{d}x) = \sum_{l=1}^{N} w_l M_{1,l}(f), \tag{5.11}$$

where $M_{1,l}(f) := \int_X f(x)\alpha_l(\mathrm{d}x)$. Moreover, if $\int_X |f(x)g(x)|\alpha_j(\mathrm{d}x)$ is finite for every $j = 1, \dots, N$, then the covariance can be expressed as

$$Cov(f(\xi_1), g(\xi_2)) = \sum_{i=2}^{N} t_i M_{1,i}(fg) + \sum_{i,j} s_{ij} M_{1,i}(f) M_{1,j}(g) \tag{5.12}$$

with

$$t_i = P\{2 \in \tilde{\pi}_1, |\tilde{\pi}_1| = i\}, \quad r_{ij} = \sum_{k=2}^{N} P\{2 \in \tilde{\pi}_k, |\tilde{\pi}_1| = i, |\tilde{\pi}_k| = j\},$$
$$s_{ij} = r_{ij} - t_i t_j - t_i \sum_{k=1}^{N} r_{j,k} + t_j \sum_{k=1}^{N} r_{i,k}.$$

Finally, if h is a real–valued measurable function defined on $(Z, \mathcal{Z})$, then

$$\mathbb{E}[h(\xi_1, \eta_1)] = \int_Z h(x, y)\kappa_x(dy)\alpha_0(\mathrm{d}x)$$

whenever the last integral is well defined.

Proof. See the Appendix, on pages 120. □

If

$$\alpha_1 = \dots = \alpha_N = \alpha, \tag{5.13}$$

then α turns out to be the common probability distribution of the ξ_j's. Moreover,

$$Cov(f(\xi_1), f(\xi_2)) = Var(\xi_1) \sum_{i=2}^{N} t_i,$$

and, hence, $Cov(f(\xi_1), f(\xi_2)) \geq 0$.

It is well–known that if $(X_n)_{n \geq 1}$ is an infinite exchangeable sequence of real–valued random elements, then $Cov(X_1, X_2) \geq 0$. Hence, if $(X_1, \dots, X_N)$ is exchangeable with

$Cov(X_1, X_2) < 0$, one can argue that the sequence $(X_n)_{1 \le n \le N}$ cannot be extended to an infinite exchangeable sequence. As shown by the next example, there are models consistent with Proposition 5.1, which present negative correlation.

Example 5.4 (negatively correlated random variables)**.** Assume that (ξ_1, ξ_2, ξ_3) is an exchangeable random vector valued into $\mathbb{R}^3$ and distributed according to Proposition 5.1 with

$$\begin{aligned}
\rho([1][2][3]) &= p_1 \\
\rho([12][3]) &= \rho([13][2]) = \rho([1][23]) = p_2 \\
\rho([123]) &= 0 \\
\int_{\mathbb{R}} x\alpha_1(\mathrm{d}x) &= 0 \\
\int_{\mathbb{R}} x\alpha_2(\mathrm{d}x) &= K \\
\int_{\mathbb{R}} x^2\alpha_1(\mathrm{d}x) &= \int_{\mathbb{R}} x^2\alpha_2(\mathrm{d}x) = 1,
\end{aligned}$$

where $p_1 + 3p_2 = 1$ and $K^2 \le 1$. Simple computations show that

$$Cov(\xi_1, \xi_2) = p_2(1 - 4p_2K^2).$$

Hence, if one chooses a couple (p_2, K) for which $1/4 < p_2 < 1/3$ and $1/(4p_2) < K^2 < 1$, then the covariance turns out to be negative.

5.3 Posterior and predictive distributions

From a finitistic point of view, in order to make any kind of inference, it is essential to handle the predictive distribution of the observations and the conditional distribution of $\tilde{e}$ given the first n observations. Indeed, in a finitary setting, the conditional law of $\tilde{e}$ given $\xi(n) := (\xi_1, \dots, \xi_n)$ plays the role of the posterior distribution of $\tilde{p}$ given $\xi(n)$, in the notation introduced in Chapter 1.

It is clear that $\mathcal{L}_{\tilde{e}|\xi(n)}$ is easily deducible from $\mathcal{L}_{\xi(n,N)|\xi(n)}$, where $\xi(n, N) = (\xi_{n+1}, \dots, \xi_N)$. More precisely, for every measurable partition $(A_1, \dots, A_k)$ of $\mathbb{X}$ and every k-tuple of integers

$M_1, \dots, M_k$ such that $M_j \geq M_j^* := \sum_{i=1}^n \delta_{\xi_i}(A_j)$ $(j = 1, \dots, k)$ and $\sum_{j=1}^k M_j = N$, one has

$$\begin{aligned}\mathcal{L}_{\tilde{e}(A_1),\dots,\tilde{e}(A_k)|\xi(n)}(\{(M_1/N, \dots, M_k/N)\}) \\ = \frac{(N-n)!}{\prod_{i=1}^k (M_j - M_j^*)!} \mathcal{L}_{\xi(n,N)|\xi(n)}(\underbrace{A_1 \times \cdots \times A_1}_{M_1^* \text{ times}} \times \cdots \times \underbrace{A_k \times \dots A_k}_{M_k^* \text{ times}}).\end{aligned}$$

Hence we will restrict our attention to $\mathcal{L}_{\xi(n,N)|\xi(n)}$.

5.3.1 Predictive distributions

Consider an exchangeable random vector $(\xi_1, \dots, \xi_N)$ whose law is characterized by Proposition 5.1. For any $n < N$, given $\xi(n) = (\xi_1, \dots, \xi_n)$, denote by $\Pi(\xi(n))$ the partition of $\{1, \dots, n\}$ generated by $\xi(n)$. Moreover, denote the distinct elements of $\xi(n)$ by $\xi_1^*, \dots, \xi_{\bar{n}}^*$, where $\bar{n}$ is the number of blocks in $\Pi(\xi(n))$, i.e. $\bar{n} = |\Pi(\xi(n))|$. Finally, given any π in $\mathcal{P}_N$, $\pi|_n$ will stand for the element of $\mathcal{P}_n$ that coincides with the restriction of π to $\{1, \dots, n\}$. If $g_{n,N}(\cdot \mid \xi(n), \tilde{\pi})$ denotes the conditional distribution of $\xi(n, N)$ given $(\xi(n), \tilde{\pi})$, it is clear that

$$g_{n,N}(A_{n+1} \times \cdots \times A_N \mid \xi(n), \tilde{\pi}) = \prod_{i=1}^{\bar{n}} \delta_{\xi_i^*} \left(\cap_{j \in \pi_i \setminus \{1,\dots n\}} A_j\right) \prod_{i=\bar{n}+1}^{|\pi|} \alpha_{|\pi_i|} \left(\cap_{j \in \pi_i} A_j\right) \tag{5.14}$$

for every $A_{n+1}, \dots, A_N$ in $\mathscr{X}$, with the convention that $\cap_{j \in \emptyset} A_j = \mathbb{X}$. Moreover, it is not difficult to prove that, for every q in $\mathcal{P}_N$,

$$\mathcal{L}_{\tilde{\pi}|\xi(n)}(q) = \tau_{\xi(n)}(q) := \frac{\rho(q) \prod_{i=1}^{\bar{n}} a_{|q_i|}(\xi_i^*) \ \mathbb{I}_{\{q|_n = \Pi(\xi(n))\}}}{\sum_{\pi \in \mathcal{P}_N} \rho(\pi) \prod_{i=1}^{\bar{n}} a_{|\pi_i|}(\xi_i^*) \ \mathbb{I}_{\{\pi|_n = \Pi(\xi(n))\}}} \tag{5.15}$$

holds true if $\alpha_i(\mathrm{d}x) = a_i(x)\mu(\mathrm{d}x)$ $(i = 1, \dots, N)$, μ being a σ–finite diffuse measure. See Lemma A.16 in the Appendix. This paves the way to prove the following

Proposition 5.3. *Let* $(\xi_1, \dots, \xi_N)$ *be defined as in* Proposition 5.1*. If* $\alpha_i(\mathrm{d}x) = a_i(x)\mu(\mathrm{d}x)$ $(i = 1, \dots, N)$*,* μ *being a* σ*–finite diffuse measure, then*

$$\begin{aligned}\mathcal{L}_{\xi_{n+1},\dots,\xi_N|\xi(n)}(A_{n+1} \times \cdots \times A_N) \\ = \sum_{\pi \in \mathcal{P}_N} \tau_{\xi(n)}(\pi) \ g_{n,N}(A_{n+1} \times \cdots \times A_N | \xi(n), \pi)\end{aligned}$$

for every $A_{n+1}, \dots, A_N$ *in* $\mathscr{X}$*.*

In particular, if $\alpha_1 = \dots \alpha_N = \alpha$, α being a diffuse probability measure, then

$$\mathcal{L}_{\xi_{n+1},\dots,\xi_N|\xi(n)}(A_{n+1} \times \dots A_N)$$

$$= \sum_{\pi \in \mathcal{P}_N} \frac{\rho(\pi)\ \mathbb{I}_{\{\pi|_n = \Pi(\xi(n))\}}}{\sum_{q \in \mathcal{P}_N} \rho(q)\ \mathbb{I}_{\{q|_n = \Pi(\xi(n))\}}} \prod_{i=1}^{\bar{n}} \delta_{\xi_i^*} \left(\cap_{j \in \pi_i \setminus \{1,\dots n\}} A_j\right) \prod_{i=\bar{n}+1}^{|\pi|} \alpha\left(\cap_{j \in \pi_i} A_j\right)$$

for every $A_{n+1}, \dots, A_N$ in $\mathscr{X}$.

Now, let us consider the conditional law of $(\xi(n,N),\, \eta(n,N))$ given $(\xi(n),\, \eta(n))$. Let $h_{n,N}(\cdot \mid \xi(N),\, \eta(n),\, \tilde{\pi})$ be a conditional distribution of $\eta(n,N)$ given $(\xi(N),\, \eta(n),\, \tilde{\pi})$. Let $(\xi_1^*, \dots, \xi_{|\pi|}^*)$ be the distinct values of $(\xi_1, \dots, \xi_N)$, $\eta^*(\pi_i) := [\eta_j \ :\ 1 \leq j \leq n,\ j \in \pi_i]$, denote by $\kappa_x(\cdot \mid \tilde{y}_1, \dots, \tilde{y}_k)$ the conditional distribution of $(\tilde{y}_{k+1}, \dots, \tilde{y}_N)$ given $(\tilde{y}_1, \dots, \tilde{y}_k)$ when $(\tilde{y}_1, \dots, \tilde{y}_N)$ is an N-exchangeable sequence with distribution $\kappa_x(\cdot)$, and convenue that $\kappa_x(\cdot \mid \emptyset) = \kappa_x(\cdot)$. Hence, $h_{n,N}$ can be written as

$$h_{n,N}(B_{n+1} \times \dots \times B_N \mid \xi(N),\, \eta(n),\, \tilde{\pi}) = \prod_{i=1}^{|\tilde{\pi}|} \kappa_{\xi_i^*}\{\times_{j \in \tilde{\pi}_i, j > n} B_j \mid \eta^*(\tilde{\pi}_i)\}$$

for every $B_1, \dots, B_N$.

Now, it is easy to see that the conditional law of $\tilde{\pi}$ given $(\xi(n), \eta(n))$ is the same as the conditional law of $\tilde{\pi}$ given $\xi(n)$. Moreover, the conditional law of $\xi(n,N)$ given $(\xi(n), \eta(n), \tilde{\pi})$ turns out to be equal to the conditional law of $\xi(n,N)$ given $(\xi(n), \tilde{\pi})$. See Lemmas A.16 and A.17 in the Appendix. This leads immediately to the following

Proposition 5.4. *Let* $(\xi_1, \dots, \xi_N)$ *be defined as in* Section 5.2. *If* $\alpha_i(\mathrm{d}x) = a_i(x)\mu(\mathrm{d}x)$ $(i = 1, \dots, N)$*, then*

$$\mathcal{L}_{\xi(n,N),\eta(n,N)|\xi(n),\eta(n)}(A \times B) = \sum_{\pi \in \mathcal{P}_N} \tau_{\xi(n)}(\pi) \int_A h_{n,N}(B \mid (\xi(n), x),\, \eta(n),\, \pi)\ g_{n,N}(\mathrm{d}x \mid \xi(n),\, \pi)$$

for every A *in* $\mathscr{X}^{N-n}$ *and every* B *in* $\mathcal{Y}^{N-n}$.

As a consequence of Proposition 5.3, when $\alpha_1, \dots, \alpha_N$ are diffuse probability measures, one has

$$P(\{\xi_{n+1} \in A\} \mid \xi(n)) = \sum_{j=1}^{N} \Delta_{n,j}(\xi(n))\ \alpha_j(A)\ + \sum_{i=1}^{\bar{n}} D_{n,i}(\xi(n))\ \delta_{\xi_i^*}(A) \qquad (A \in \mathscr{X}),$$

where

$$D_{n,i}(\xi(n)) = D_i = \sum_{\pi \in \mathcal{P}_N :\, \phi(\pi, n+1) = i} \tau_{\xi(n)}(\pi) \qquad (i = 1, \dots, \bar{n})$$

and

$$\Delta_{n,j}(\xi(n)) = \Delta_{n,j} = \sum_{\pi \in \mathcal{P}_N} \tau_{\xi(n)}(\pi)\, \mathbb{I}_{\{\phi(\pi,n+1) \geq \bar{n}+1, |\pi_{\phi(\pi,n+1)}|=j\}} \qquad (j = 1, \dots, N).$$

In particular, if $\alpha_1 = \dots \alpha_N = \alpha$, α being a diffuse probability measure, one has

$$P(\{\xi_{n+1} \in A\}|\xi(n)) = D_0\, \alpha(A) + \sum_{i=1}^{\bar{n}} D_i\, \delta_{\xi_i^*}(A) \qquad (A \in \mathscr{X}),$$

where $D_0 = 1 - \sum_{i=1}^{\bar{n}} D_i$. In the same way, from Proposition 5.4 one can derive the predictive distribution for (ξ_{n+1}, η_{n+1}) given $(\xi(n), \eta(n))$, which is

$$\begin{aligned}\mathcal{L}_{\xi_{n+1},\eta_{n+1}|(\xi(n),\eta(n))}(A \times B) = &\sum_{i=1}^{\bar{n}} \sum_{\pi \in \mathcal{P}_N:\ \phi(\pi,n+1)=i} \tau_{\xi(n)}(\pi)\, \delta_{\xi_i^*}(A)\, \kappa_{\xi^*}(B|\eta^*(\pi_i)) \\ &+ \sum_{j=1}^{N} \Delta_{n,j}(\xi(n)) \int_A \kappa_x(B)\, \alpha_j(\mathrm{d}x)\end{aligned}$$

for any A in $\mathscr{X}^{N-1}$ and B in $\mathcal{Y}^{N-1}$. Hence, for any measurable real–valued function h on $\mathbb{X} \times \mathbb{Y}$, one has

$$\begin{aligned}\mathbb{E}[h(\xi_{n+1}, \eta_{n+1})|\xi(n),\, \eta(n)] = &\sum_{i=1}^{\bar{n}} \sum_{\pi \in \mathcal{P}_N:\ \phi(\pi,n+1)=i} \tau_{\xi(n)}(\pi) \int_{\mathbb{Y}} h(\xi_i^*, y)\, \kappa_{\xi_i^*}(\mathrm{d}y|\eta^*(\pi_i)) \\ &+ \sum_{j=1}^{N} \Delta_{n,j}(\xi(n)) \int_{\mathbb{X}\times\mathbb{Y}} h(x,y)\, \kappa_x(dy)\, \alpha_j(\mathrm{d}x)\end{aligned}$$

if $\mathbb{E}|h(\xi_{n+1}, \eta_{n+1})| < +\infty$.

Chapter 6

Some applications

The aim of this chapter is to apply the distributions introduced in Chapter 4 and Chapter 5 to some standard statistical problems: we show how one can estimate the mean of the empirical measure, and we propose a bivariate model based on partition tree distributions in order to approach regression problems.

6.1 Decision theoretic formulation

As explained in Chapter 1, we shall focus on inferences from the sample $\xi(n) := (\xi_1, \dots, \xi_n)$ $(n < N)$ to empirical versions of the usual parameter, i.e. on random elements with the following form:

$$\tilde{\theta} = t(\tilde{e}),$$

where t is a mapping from a subset $\mathbb{P}_0$ of the class $\mathbb{P}$ of all probabilities on $(\mathbb{X}, \mathscr{X})$ into Θ.

A decision theoretic approach will be followed. So, the statistician is assumed to have a set $\mathbb{D}$ of *decision rules* at his disposal, and these rules are defined, for any $n \leq N$, as measurable functions from $\mathbb{X}^n$ to some set $\mathbb{A}$ of actions. Then, one considers a *loss function* L, i.e. a positive real–valued function on $\Theta \times \mathbb{A}$. The function $L(\theta, a)$ represents the loss when the value of $\tilde{\theta}$ is θ and the statistician chooses action a. It is supposed that

$$r(\delta(\xi(n))) := \int_{\Theta} L(\theta, \delta(\xi(n)))\mathcal{L}_{\tilde{\theta}|\xi(n)}(d\theta)$$

is finite for any δ in $\mathbb{D}$. In this case, $r(\cdot)$ is said to be the *a posteriori Bayes risk* of $\delta(\xi(n))$. Finally, a *Bayes rule* is defined to be any element δ_0 of $\mathbb{D}$ such that

$$r(\delta_0(\xi(n))) = \min_{\delta\in\mathbb{D}} r(\delta(\xi(n)))$$

for any realization (observation) of $\xi(n)$.

6.2 Estimation of the mean

Dealing with an estimation problem, Θ and $\mathbb{A}$ are two subsets of $\mathbb{R}^k$, and the usual loss function is $L(\theta, a) = \|\theta - a\|^2$, i.e. the quadratic cost. Take $k = 1$, assume that $\mathbb{X}$ is $\mathbb{R}$, and consider the standard problem of estimating the mean: take $t(p) = \int_{\mathbb{X}} f(x)\mathrm{d}p(x)$, where f is some measurable function from $\mathbb{X}$ into itself and $\mathbb{P}_0$ the set of probabilites in $\mathbb{P}$ such that $t(p)$ exists.

In the usual setting, dealing with an infinite sequence of observations, the parameter to be estimated is $\tilde{\theta}_\infty = t(\tilde{p})$ and its estimator is simply $\mathbb{E}(\tilde{\theta}_\infty \mid \xi(n)) = \mathbb{E}(f(\xi_{n+1}) \mid \xi(n))$. In a finitary setting, one considers the parameter $\tilde{\theta} = \int_{\mathbb{X}} f(x)\,\mathrm{d}\,\tilde{e}(x) = \frac{1}{N}\sum_{i=1}^{N} f(\xi_i)$, instead. In virtue of linearity of expectation, a finitary Bayes rule is given by a a linear combination of the sample mean and the classical estimate for the mean:

$$\mathbb{E}(\tilde{\theta} \mid \xi(n)) = \frac{n}{N-n}\,\frac{1}{n}\sum_{i=1}^{n} f(\xi_i) + \frac{N-n}{N}\,\mathbb{E}(f(\xi_{n+1}) \mid \xi(n)) \tag{6.1}$$

Let us see how this estimate can be found both for the Random Partition Model and for the hypergeometric partitions tree distributions introduced in Chapter 5 and Chapter 4, respectively.

6.2.1 Random Partition model

Let $(\xi_1, \ldots, \xi_N)$ be distributed as in the Random Partition Model, introduced in Section 5.2. For simplicity, take $\alpha_1 = \cdots = \alpha_N = \alpha$, with α a diffuse probability on $(\mathbb{X}, \mathscr{X})$. Denote by $\bar{n}$ the number of distinct values in $(\xi_1, \ldots, \xi_n)$ and by $\xi_1^*, \ldots, \xi_{\bar{n}}^*$ their values. Moreover, let

$$D_{n,i}(\xi(n)) = D_i = \sum_{\pi\in\mathcal{P}_N:\ \phi(\pi,n+1)=i} \tau_{\xi(n)}(\pi) \qquad (i = 1, \ldots, \bar{n})$$

$$D_0 = 1 - \sum_{i=1}^{\bar{n}} D_i,$$

where $\mathcal{L}_{\tilde{\pi}|\xi(n)}(\{\pi\}) = \tau_{\xi(n)}(\pi)$. Our goal is to evaluate

$$\mathbb{E}[f(\xi_{n+1}) \mid \xi(n)] = \sum_{i=0}^{\bar{n}} D_i f_i,$$

where $f_i := f(\xi_i^*)$ for $i = 1, \dots, \bar{n}$ and $f_0 := \int_x f(x)\alpha(\mathrm{d}x)$.

Since an exact evaluation of $\mathbb{E}[f(\xi_{n+1}) \mid \xi(n)]$ can be computationally cumbersome even if N is small, we shall suggest a very simple algorithm to approximate such a sum in some particular situations.

Consider ρ as in Example 5.3 with $\mu(\mathrm{d}\lambda) = \delta_0$, $A_0 = A$, and $\beta_0 = \beta$. In other words, take

$$\rho(\pi) = K_N^{-1} A(|\pi|) \prod_{i=1}^{|\pi|} \beta(|\pi_i|) \tag{6.2}$$

with $K_N := \sum_{\pi\in\mathcal{P}_N} A(|\pi|)\prod_{i=1}^{|\pi|}\beta(|\pi_i|)$. Define $V_M^{(\bar{n})} := \{t \in \{0,\dots,M\}^{\bar{n}} : \sum_{i=1}^{\bar{n}} t_i = M\}$, and write $\mathfrak{X}_{\bar{n}}^M$ for $\{1,\dots,\bar{n}\}^M$. For any $x = (x_1,\dots,x_M)$ in $\mathfrak{X}_{\bar{n}}^M$, set

$$t(x) = (|\{i = 1,\dots,M : x_i = 1\}|, \dots, |\{i = 1,\dots,M : x_i = \bar{n}\}|), \quad \text{and denote}$$
$$t^* = t^*(\Pi(\xi(n))) := (|\Pi(\xi(n))|_1, \dots, |\Pi(\xi(n))|_{\bar{n}}),$$
$$\mathcal{W}_l := \Big\{(b_1,\dots,b_l) \in \mathbb{Z}^l : \textstyle\sum_{i=1}^l i\, b_i = l\Big\}.$$

Hence, by (6.2), letting $C := K_N \sum_{\pi\in\mathcal{P}_N} \rho(\pi)\,\mathbb{I}_{\{\pi|_n = \Pi(\xi(n))\}}$, one can write, for every $i = 1,\dots,\bar{n}$,

$$\begin{aligned}
D_i &:= \sum_{\pi\in\mathcal{P}_N:\,\phi(\pi,n+1)=i} \tau_{\xi(n)}(\pi) = \frac{K_N}{C}\sum_{\pi\in\mathcal{P}_N}\rho(\pi)\,\mathbb{I}_{\{\pi|_n=\Pi(\xi(n)),n+1\in\pi_i\}}\\
&= \frac{1}{C}\sum_{\pi\in\mathcal{P}_N} A(|\pi|)\prod_{j=1}^{|\pi|}\beta(|\pi_j|)\,\mathbb{I}_{\{\pi|_n=\Pi(\xi(n)),n+1\in\pi_i\}}\\
&= \frac{1}{C}\sum_{l=0}^{N-n-1}\binom{N-(n+1)}{l}\sum_{b\in\mathcal{W}_l}\frac{l!}{\prod_{j=1}^l b_j!\,(j!)^{b_j}}\,A(\bar{n}+|b|)\prod_{j=1}^l\beta(j)^{b_j}\\
&\qquad\cdot\sum_{t\in V^{(\bar{n})}_{N-(l+1+n)}}\frac{(N-(n+l+1))!}{\prod_{j=1}^{\bar{n}} t_j!}\prod_{j=1}^{\bar{n}}\beta(t_j^*+\delta_{ij}+t_j)\\
&= \frac{1}{C}\sum_{l=0}^{N-n-1}\binom{N-(n+1)}{l}Q(l)\sum_{z\in\mathfrak{X}_{\bar{n}}^{N-l-1-n}}\prod_{j=1}^{\bar{n}}\beta(t_j^*+\delta_{ij}+t(z)_j)\\
&= \frac{1}{C}\sum_{l=0}^{N-n-1}\binom{N-(n+1)}{l}Q(l)\sum_{z\in\mathfrak{X}_{\bar{n}}^{N-l-1-n}} F(t(z)\mid i,l),
\end{aligned}$$

where

$$Q(0) := A(\bar{n}),$$

$$Q(l) := \sum_{k=1}^{l} A(\bar{n}+k)\, B_{l,k}(\beta(1),\dots,\beta(l)) \qquad (l=1,\dots,N-n-1),$$

$B_{l,k}$ being the usual partial Bell partition polynomial and

$$F(t \mid i,l) := \prod_{j=1}^{\bar{n}} \beta(t_j^* + \delta_{ij} + t_j)$$

with the convention that $\mathfrak{X}_{\bar{n}}^0 := \emptyset$ and $\sum_{z\in\emptyset} F(t(z) \mid i,l) = F(\emptyset \mid i,l) = \prod_{j=1}^{\bar{n}} \beta(t_j^* + \delta_{ij})$. Recall that the partial Bell polynomial of degree (n,k) is defined by

$$B_{n,k}(x_1,\dots,x_n) = \sum \frac{n!}{k_1!\,(1!)^{k_1}\,\dots\,k_n!\,(n!)^{k_n}}\; x_1^{k_1} x_2^{k_2} \dots x_n^{k_n},$$

where the summation is extended over all partitions of n into k parts, i.e. over all nonnegative integer solutions $(k_1,\dots,k_n)$ of the equations $k_1 + 2k_2 + \dots + nk_n = n$, $k_1 + \dots + k_n = k$ and $B_{0,0} := 1$.

Analogously, it is plain to check that

$$\begin{aligned} D_0 &:= \frac{K_N}{C} \sum_{\pi\in\mathcal{P}_N} \rho(\pi)\, \mathbb{I}_{\{\pi|_n=\Pi(\xi(n)),n+1\notin\pi_i,\, i=1,\dots,\bar{n}\}} \\ &= \frac{1}{C} \sum_{l=0}^{N-n-1} \binom{N-(n+1)}{l} Q(l+1) \sum_{z\in\mathfrak{X}_{\bar{n}}^{N-l-1-n}} F(t(z) \mid 0,l). \end{aligned}$$

Hence,

$$\mathbb{E}[f(\xi_{n+1}) \mid \xi(n)] = \frac{\sum_{i=0}^{\bar{n}} \sum_{l=0}^{N-n-1} \sum_{z\in\mathfrak{X}_{\bar{n}}^{N-l-1-n}} \binom{N-(n+1)}{l} Q(l+\delta_{i,0})\, F(t(z) \mid i,l)\, f_i}{\sum_{i=0}^{\bar{n}} \sum_{l=0}^{N-n-1} \sum_{z\in\mathfrak{X}_{\bar{n}}^{N-l-1-n}} \binom{N-(n+1)}{l} Q(l+\delta_{i,0})\, F(t(z) \mid i,l)}$$

holds true. The last equation suggests an easy way to estimate $\mathbb{E}[f(\xi_{n+1}) \mid \xi(n)]$ by a sequential importance sampling method. That is: let $(i_M, l_M, z_M)_{M\geq 1}$ be a sequence of i.i.d random variables taking values in $\{0,\dots,\bar{n}\} \times \{0, N-n-1\} \times \cup_{l=0}^{N-n-1} \mathfrak{X}_{\bar{n}}^{N-(l+1+n)}$, with common distribution given by

$$q(i,l,z) = q_1(i)\, q_2(l \mid i)\, q_3(z \mid i,l)$$

and $q(\mathfrak{X}_{\bar{n}}^{N-(l+1+n)} \mid i,l) = 1$, that is to say that z_M belongs to $\mathfrak{X}_{\bar{n}}^{N-(l+1+n)}$ with conditional probability one given $l_M = l$. Next, set

$$w(i,l,z) := \frac{\binom{N-(n+1)}{l} Q(l+\delta_{i,0})\, F(t(z) \mid i,l)}{q_1(i)\, q_2(l \mid i)\, q_3(z \mid i,l)}.$$

It follows that

$$E_{\bar{M}} = \frac{\sum_{M=1}^{\bar{M}} f_{i_M}\ w(i_M, l_M, z_M)}{\sum_{M=1}^{\bar{M}} w(i_M, l_M, z_M)}$$

is an asymptotically unbiased and strong consistent estimator of $\mathbb{E}[f(\xi_{n+1}) \mid \xi(n)]$.

It should be emphasized that the partial Bell partition polynomials are easily computable by recursion, indeed

$$B_{n+1,k+1}(x_1, \dots, x_{n+1}) = \sum_{r=0}^{n-k} \binom{n}{r}\, x_{r+1}\, B_{n-r,k}(x_1, \dots, x_{n-r})$$

for $k = 0, 1, \dots, n$ and $n = 0, 1, \dots$ See, for instance, Charalambides (2002).

6.2.2 Hypergeometric partitions tree distributions

Let $\mathbb{X} = (0, 1]$, and let the law of $(\xi_1, \dots, \xi_N)$ be the distribution $\mathscr{H}(\mathcal{N})$, introduced in Section 4.5.1. Our aim is to evaluate $\int_{\mathbb{X}} f(x) \mathrm{d}\tilde{e}(x)$. For simplicity, assume that f is a monotone function.

Combining (4.23) and (4.25), one obtains a closed form for the predictive expectation:

$$\mathbb{E}(\xi_{n+1} \mid \xi(n)) = \sum_{m=1}^{+\infty} 2^{-m} \sum_{\substack{(\varepsilon_1 \dots \varepsilon_m) \in E^m: \\ \varepsilon_m = 1}} \prod_{k=1}^{m} \frac{\alpha_{\varepsilon_1 \dots \varepsilon_k} - \tilde{n}_{\varepsilon_1 \dots \varepsilon_k}}{\alpha_{\varepsilon_1 \dots \varepsilon_{k-1} 0} + \alpha_{\varepsilon_1 \dots \varepsilon_{k-1} 1} - \tilde{n}_{\varepsilon_1 \dots \varepsilon_{k-1}}}, \tag{6.3}$$

where $\tilde{n}_\varepsilon$ denotes $\sum_{i=1}^{n} \delta_{\{\xi_i \in I_\varepsilon\}}$ for every ε in E^*. This formula can be plugged in (6.1) to obtain the Bayes rule for $t(\tilde{e}) = \int_{\mathbb{X}} f(x)\ \mathrm{d}\,\tilde{e}(x)$, being f the identity function. The problem is that the series in (6.3) is too cumbersome to be evaluated by a partial sum. Other procedures are better to find a good approximation of (6.3).

Denote

$$\mathcal{S}_m := \{\varepsilon \in E^m : \xi_i \notin I_\varepsilon, \quad i = 1, \dots, n\}$$

for each $m \geq 1$, and write:

$$\mathbb{E}(f(\xi_{n+1}) \mid \xi(n)) = \sum_{\varepsilon \in \mathcal{S}_m} \int_{I_\varepsilon} f(x)\ \mathrm{d}\,\mathcal{L}_{\xi_{n+1}|\xi(n)}(x) + \sum_{\varepsilon \in E^m \cap \mathcal{S}_m^c} \int_{I_\varepsilon} f(x)\ \mathrm{d}\,\mathcal{L}_{\xi_{n+1}|\xi(n)}(x). \tag{6.4}$$

One can find the exact value of the first sum in (6.4) and approximate the second one. In fact, if no observations fall in I_ε (for $\varepsilon = (\varepsilon_1 \dots \varepsilon_m)$ in E^m), then, by (4.25) and then by (4.22), for any $\varepsilon' = (\varepsilon_1 \dots \varepsilon_m \, \varepsilon_{m+1} \dots \varepsilon_h)$ in E^h with $h > m$,

$$\begin{aligned}
P(\xi_{n+1} \in I_{\varepsilon'} \mid \xi(n)) &= \frac{\alpha_{\varepsilon_1} - \tilde{n}_{\varepsilon_1}}{\alpha_0 + \alpha_1 - n} \cdot \dots \cdot \frac{\alpha_{\varepsilon_1 \dots \varepsilon_m}}{\alpha_{\varepsilon_1 \dots \varepsilon_{m-1} 0} + \alpha_{\varepsilon_1 \dots \varepsilon_{m-1} 1} - \tilde{n}_{\varepsilon_1 \dots \varepsilon_{m-1}}} \\
&\quad \cdot \frac{\alpha_{\varepsilon_1 \dots \varepsilon_{m+1}}}{\alpha_{\varepsilon_1 \dots \varepsilon_m 0} + \alpha_{\varepsilon_1 \dots \varepsilon_m 1}} \cdots \frac{\alpha_{\varepsilon_1 \dots \varepsilon_h}}{\alpha_{\varepsilon_1 \dots \varepsilon_{h-1} 0} + \alpha_{\varepsilon_1 \dots \varepsilon_{h-1} 1}} \\
&= P(\xi_{n+1} \in I_\varepsilon \mid \xi(n)) \; P(\xi_{n+1} \in I_{\varepsilon'} \mid \xi_{n+1} \in I_\varepsilon) \\
&= \frac{P(\xi_{n+1} \in I_\varepsilon \mid \xi(n))}{P(\xi_{n+1} \in I_\varepsilon)} \; P(\xi_{n+1} \in I_{\varepsilon'}) \\
&= \frac{P(\xi_{n+1} \in I_\varepsilon \mid \xi(n))}{P(\xi_1 \in I_\varepsilon)} \; P(\xi_1 \in I_{\varepsilon'}).
\end{aligned}$$

Hence, for each dyadic interval I_ε containing no observations,

$$\mathcal{L}_{\xi_{n+1} | \xi(n)}|_{I_\varepsilon} = \frac{P(\xi_{n+1} \in I_\varepsilon \mid \xi(n))}{P(\xi_1 \in I_\varepsilon)} \; \mathcal{L}_{\xi_1},$$

denoting by $\mu|_B$ the restriction of a probability measure μ defined on $\mathscr{X}$ to the σ-field $\mathscr{X}_B := \{A \cap B : A \in \mathscr{X}\}$, for B in $\mathscr{X}$. Therefore, the first sum in (6.4) becomes:

$$\sum_{\varepsilon \in \mathcal{S}_m} \frac{P(\xi_{n+1} \in I_\varepsilon \mid \xi(n))}{P(\xi_1 \in I_\varepsilon)} \; \mathbb{E}(f(\xi_1) \, \mathbb{I}_{\{\xi_1 \in I_\varepsilon\}}).$$

The second sum is contained by the interval

$$\left(\sum_{\varepsilon \in E^m \cap \mathcal{S}_m^c} P(\xi_{n+1} \in I_\varepsilon \mid \xi(n)) \, f(l_\varepsilon), \sum_{\varepsilon \in E^m \cap \mathcal{S}_m^c} P(\xi_{n+1} \in I_\varepsilon \mid \xi(n)) \, f(l_\varepsilon + 2^{-m}) \right], \tag{6.5}$$

where l_ε denotes the lower bound of I_ε, i.e. $l_{\varepsilon_1 \dots \varepsilon_m} = \sum_{j=1}^m \varepsilon_j 2^{-j}$. So, an approximation for $\mathbb{E}(f(\xi_{n+1}) \mid \xi(n))$ is obtained taking the middle point of the interval (6.5), i.e.

$$\begin{aligned}
&\mathbb{E}(f(\xi_{n+1}) \mid \xi(n)) \\
&\qquad \approx \sum_{\varepsilon \in \mathcal{S}_m} \frac{P(\xi_{n+1} \in I_\varepsilon \mid \xi(n))}{P(\xi_1 \in I_\varepsilon)} \; \mathbb{E}(f(\xi_1) \, \mathbb{I}_{\{\xi_1 \in I_\varepsilon\}}) \\
&\qquad\quad + \sum_{\varepsilon \in E^m \cap \mathcal{S}_m^c} P(\xi_{n+1} \in I_\varepsilon \mid \xi(n)) \cdot \frac{f(l_\varepsilon) + f(l_\varepsilon + 2^{-m})}{2},
\end{aligned} \tag{6.6}$$

so that the error is bounded above by half length of the interval in (6.5), that is

$$\sum_{\varepsilon \in E^m \cap \mathcal{S}_m^c} P(\xi_{n+1} \in I_\varepsilon \mid \xi(n)) \cdot \frac{f(l_\varepsilon + 2^{-m}) - f(l_\varepsilon)}{2}.$$

If, for instance, f is the identity function, $\alpha_\varepsilon = K$ for all ε in E^*(so that $\mathcal{L}_{\xi_1}$ is uniform over $(0, 1]$), (6.6) becomes

$$\mathbb{E}(\xi_{n+1} \mid \xi(n)) \approx \sum_{\varepsilon \in E^m} P(\xi_{n+1} \in I_\varepsilon \mid \xi(n)) \cdot (l_\varepsilon + 2^{-m-1}), \tag{6.7}$$

and the error is not greater than

$$\sum_{\varepsilon \in E^m \cap \mathcal{S}_m^c} P(\xi_{n+1} \in I_\varepsilon \mid \xi(n)) \cdot 2^{-m-1}.$$

A fictitious population was created, generating $N = 2000$ random variables $\xi_1, \ldots, \xi_N$, taking value in the unit interval, and having joint distribution $\mathscr{H}(\mathcal{N})$, with α_ε constantly equal to $K = 2500$. A sample of size one hundred was taken, the sample mean was calculated, and the finitary Bayes rule was evaluated by means of (6.7), with $m = 10$. Then, other hundred units were added to the sample, and the two estimates were calculated again. This was done for eight times more. Figure 6.1 compares the finitary Bayes rules to the classical sample means obtained. One can see that the finitary Bayesian estimate is closer to the population mean

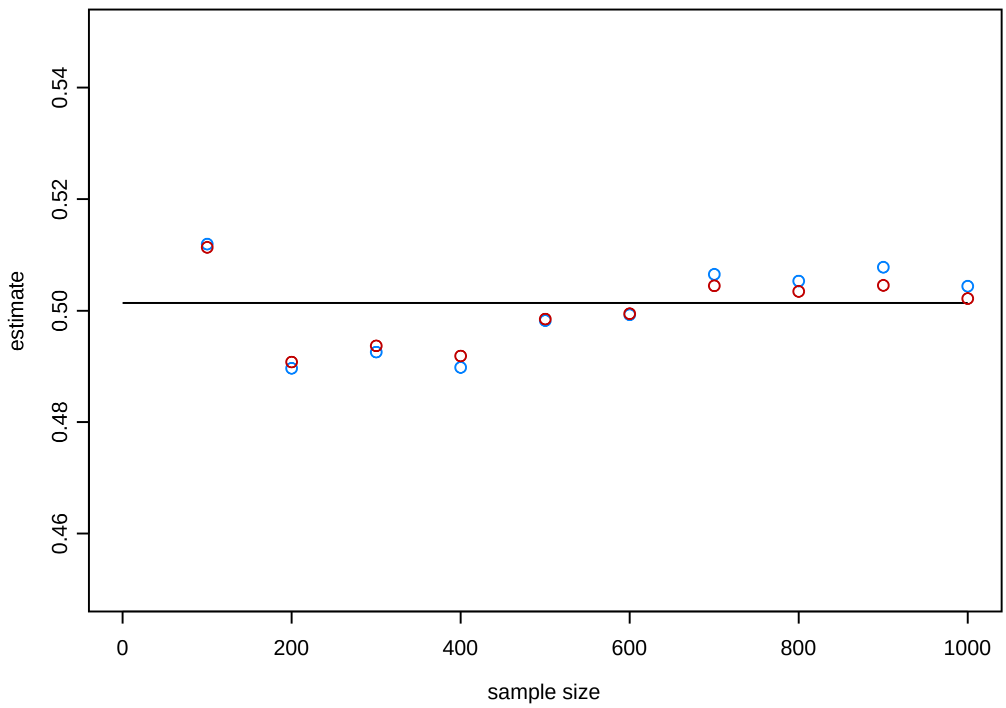


Figure 6.1: Population mean estimates, increasing the sample size. Sample mean (blue circles), finitary Bayes estimate (red circles), true population mean (black line).

than the traditional empirical estimate.

6.3 Regression problems

Assume that two different (real-valued) phenomena are observed on each statistical unit. Formally, that means that each observation ζ_i is a pair (ξ_i, η_i) of real numbers. Hence, the range of each observation $\mathbb{X} \times \mathbb{Y}$ is a subset of $\mathbb{R}^2$. Let us denote by $\xi := (\xi_1, \ldots, \xi_N)$ and by $\eta := (\eta_1, \ldots, \eta_N)$ the sequence of observations related respectively to the first and the second phenomenon. Assume that the statistician has to investigate about a possible relationship between the two phenomena. Hence, some functional space $\mathscr{F}$ of maps from $\mathbb{X}$ into $\mathbb{Y}$ is considered, and the purpose will be to choose a map from $\mathscr{F}$ to express the dependence of the second phenomenon on the first one.

So $\mathscr{F}$ coincides with the space of all possible actions $\mathbb{A}$. The elements f_τ of $\mathscr{F}$ can be indexed - as usual in regression procedures - by a parameter τ belonging to some space T. In what follows, we shall take as $\mathscr{F}$ the space of all affine functions from $\mathbb{R}$ into itself, that is $\tau = (\alpha, \beta)$ and $f_\tau(x) = \alpha + \beta x$. Other possible choices for f_τ are known to be polynomials or $f_\tau(x) = \alpha e^{\beta x}$.

At this point two different approaches are possible. One may be concerned only with the relationship between the two components ξ_{n+1}, η_{n+1} of the outcoming observation, and we can just take the squared loss error. In this way, the quantity to be estimated is not a function of $\tilde{e}$, but only of the next observation, and, therefore, the past observations $\zeta(n)$ does not enter into the loss function, but only into the predictive distribution.

Instead, one may be interested in "approximating" the whole sequence of $(\eta_1, \ldots, \eta_N)$ by $(f_\tau(\xi_1), \ldots, f_\tau(\xi_N))$. In this other case, a natural choice for the loss function can be taken to be $\int_{\mathbb{X}^2} (x-y)^2 \tilde{e}(\mathrm{d}x, \mathrm{d}y)$, i.e. $\sum_{i=1}^N (\eta_i - f_\theta(\xi_i))^2/N$. Minimizing this loss function gives rise to the estimate, under square loss function, of the conditional expectation $\mathbb{E}(Z_2 \mid Z_1)$ (where the law of (Z_1, Z_2) is $\tilde{e}$) taking as $\mathbb{A}$ the affine functions.

6.3.1 First approach

Assume we have observed $\zeta(n)$ and we are concerned with finding the formal Bayes rule for (ξ_{n+1}, η_{n+1}) when the action space is the set of all affine functions from $\mathbb{X}$ into $\mathbb{Y}$ and the loss function is $L(x_1, x_2; \tau) = (x_2 - f_\tau(x_1))^2$ where $f_\tau(x) = \alpha + \beta\, x$. In formula we want to calculate:

$$\operatorname{argmin}_{(\alpha,\beta)} \mathbb{E}((\eta_{n+1} - (\alpha + \beta\xi_{n+1}))^2 \mid \zeta(n)), \tag{6.8}$$

which yields:

$$\begin{aligned} \hat{\alpha}_n &= \mathbb{E}(\eta_{n+1} \mid \zeta(n)) - \hat{\beta}\,\mathbb{E}(\xi_{n+1} \mid \zeta(n)) \\ \hat{\beta}_n &= \frac{\operatorname{Cov}(\xi_{n+1}, \eta_{n+1} \mid \zeta(n))}{\operatorname{Var}(\xi_{n+1} \mid \zeta(n))}. \end{aligned} \tag{6.9}$$

In fact, $(\hat{\alpha}_n, \hat{\beta}_n)$ is the only zero of the gradient of the function we want to minimize in (6.8), which is a convex function since its hessian matrix

$$H = 2 \cdot \begin{pmatrix} 1 & \mathbb{E}(\xi_{n+1}|\zeta(n)) \\ \mathbb{E}(\xi_{n+1}|\zeta(n)) & \mathbb{E}(\xi_{n+1}^2|\zeta(n)) \end{pmatrix} \tag{6.10}$$

is definitive positive $[\det(H) = 2\operatorname{Var}(\xi_{n+1} \mid \zeta(n))]$.

6.3.2 Second approach

For each $p \in \mathbb{P}$, let p_1 and $p_{2|1}$ denote respectively the marginal law of Z_1 and the conditional law of Z_2 given Z_1 when p is the law of (Z_1, Z_2). Suppose we are concerned with finding a Bayes rule for $t(\tilde{e}; x)$, where $t(p; x) = \int y p_{2|1}(\mathrm{d}y|x)$, choosing as action space $\mathbb{A}$ the set of the affine functions. We can take as loss function:

$$L_a(p, \tau) = \int |t(p; x) - f_\tau(x)|^2 \mu(\mathrm{d}x; p). \tag{6.11}$$

If we put $\mu(\mathrm{d}x; p) = p_1(\mathrm{d}x)$ and (Z_1, Z_2) has law $\tilde{e}$, (6.11) yields:

$$\begin{aligned} L_a(\tilde{e}, \tau) &= \mathbb{E}[(\mathbb{E}[Z_2|Z_1] - f_\tau(Z_1))^2] \\ &= \mathbb{E}[(\mathbb{E}[Z_2|Z_1])^2] + \mathbb{E}[f_\tau(Z_1)^2] - 2\,\mathbb{E}[Z_2 f_\tau(Z_1)] \\ &= \mathbb{E}[(\mathbb{E}[Z_2|Z_1])^2] + \sum_{i=1}^{N} f_\tau(\xi_i)(f_\tau(\xi_i) - 2\eta_i). \end{aligned} \tag{6.12}$$

At this point, note that it is equivalent to minimize either $\tau \mapsto L_a(\tilde{e},\tau)$ or

$$\tau \mapsto L_b(\tilde{e},\tau) := \int_{\mathbb{R}} (y - f_\tau(x))^2 \tilde{e}(\mathrm{d}x\,\mathrm{d}y) = \frac{1}{N}\sum_{i=1}^N (\eta_i - f_\tau(\xi_i))^2. \tag{6.13}$$

If $f_\tau(x) = \alpha + \beta x$ ($\tau = (\alpha,\beta)$), the minimizer of

$$\tau \mapsto \mathbb{E}[L_b(\tilde{e},\tau)|\zeta(n)] \tag{6.14}$$

is given by

$$\begin{aligned} \alpha_n^* &= \mathbb{E}(Y) + \beta_n^*\,\mathbb{E}(X) \\ \beta_n^* &= \frac{Cov(X,Y)}{Var(X)}, \\ \text{where} \quad \mathcal{L}_{(X,Y)} &= \frac{n}{N}\tilde{e}_n + \left(1 - \frac{n}{N}\right)\mathcal{L}_{\zeta_{n+1}|\zeta(n)}. \end{aligned}$$

In fact,

$$\mathbb{E}(L_a(\tilde{e},\tau) \mid \zeta(n)) = \mathbb{E}(L(X,Y)),$$

and, therefore, proceeding as in Section 6.3.1, one can see that the minimizer of (6.14) can be obtained from (6.9) replacing $\mathcal{L}_{(\xi_{n+1},\eta_{n+1})|\zeta(n)}$ with $\mathcal{L}_{(X,Y)}$.

Note that when $n = N$, (α_n^*, β_n^*) is the least square estimate of τ. Moreover, if n is fixed and N diverges, (α_n^*, β_n^*) converges to $(\hat{\alpha}_n, \hat{\beta}_n)$.

6.3.3 A numerical example

As a matter of example, assume that an exchangeable law for $(\zeta_1,\ldots,\zeta_N)$ has been assessed such that: $\mathcal{L}_{(\xi_1,\ldots,\xi_N)}$ is $\mathscr{H}(\mathcal{N})$, with α_ε equal to a constant K for each ε in E^*, the conditional expectation $\mathbb{E}(\eta_i \mid \xi_1,\ldots,\xi_N)$ is an affine function of ξ_i (as $i = 1,\ldots,N$), and each $\mathcal{L}_{\eta_i|\xi_1,\ldots,\xi_N}$ depends on $(\xi_1,\ldots,\xi_N)$ only through $\mathbb{E}(\eta_i \mid \xi_1,\ldots,\xi_N)$. So, setting $\gamma_i := \eta_i - \mathbb{E}(\eta_i \mid \xi_i)$ for $i = 1,\ldots,N$, the random vectors $(\gamma_1,\ldots,\gamma_N)$ and $(\xi_1,\ldots,\xi_N)$ are stochastically independent.

This assessment is tantamount to assuming the existence of N exchangeable random elements $(\gamma_1,\ldots,\gamma_N)$ of $\mathbb{Y}$ and of two real number a and b, such that

$$\eta_i = a + b\xi_i + \gamma_i \qquad \text{and}$$

$$\mathbb{E}(\gamma_i) = 0$$

as $i = 1, \ldots, N$. Assume that $(\gamma_1, \ldots, \gamma_N)$ is distributed as the initial segment of an (infinite) exchangeable sequence directed by a Dirichlet process with parameter α equal to a Gaussian distribution with mean zero and variance σ^2. In this way, the total mass parameter a is one.

A fictitious population of one thousand units ($N = 1000$) was created, generating N random vectors $\zeta_1, \ldots, \zeta_N$ with the distribution described above and $a = 0$, $b = 1$, $K = 500$, $\sigma^2 = 1/4$. Then, a sample of size sixteen was taken, and $(\alpha^*_{16}, \beta^*_{16})$ was calculated.

The results are shown in Figure 6.2: the Bayes rule line is closer to the whole population minimum squares line than the sample minimum squares line.

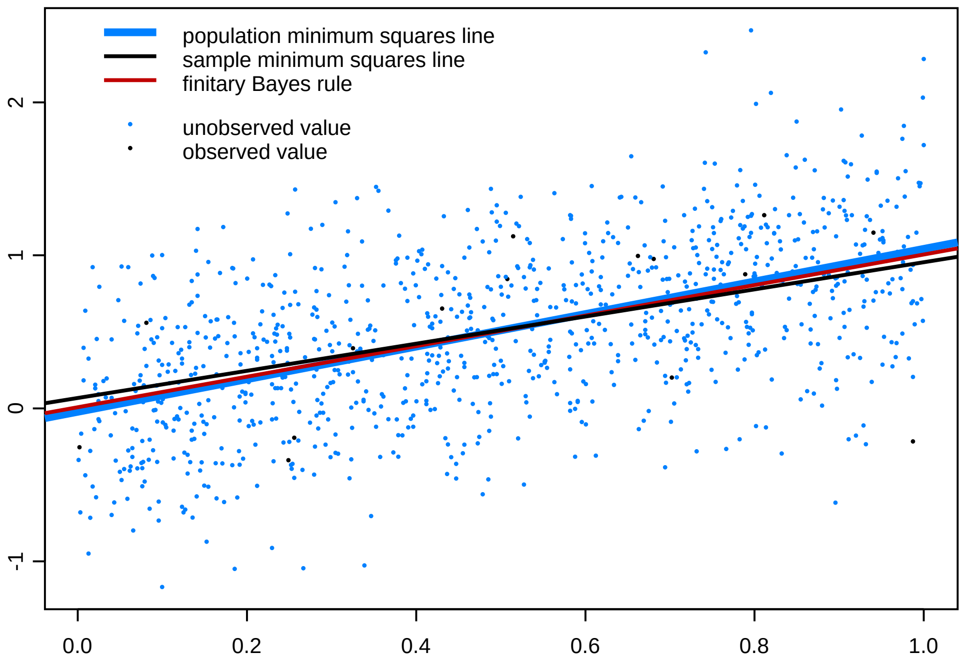


Figure 6.2: Linear regression.

Then, another sample of ten units was taken from the same population, and the estimate (α^*_n, β^*_n) was calculated again, but on the basis of a "bad guess" about the parameters a and b: $a_0 = -0.3$, $b_0 = 1.5$. The sample size was progressively increased adding other units to the sample, and the same estimate was calculated each time.

As shown in Figure 6.3, the wrong choice of a_0, b_0 affect the result, but, as the sample size increases, the finitary Bayes rule line (the red line) gets closer to the population minimum squares (blue) line.

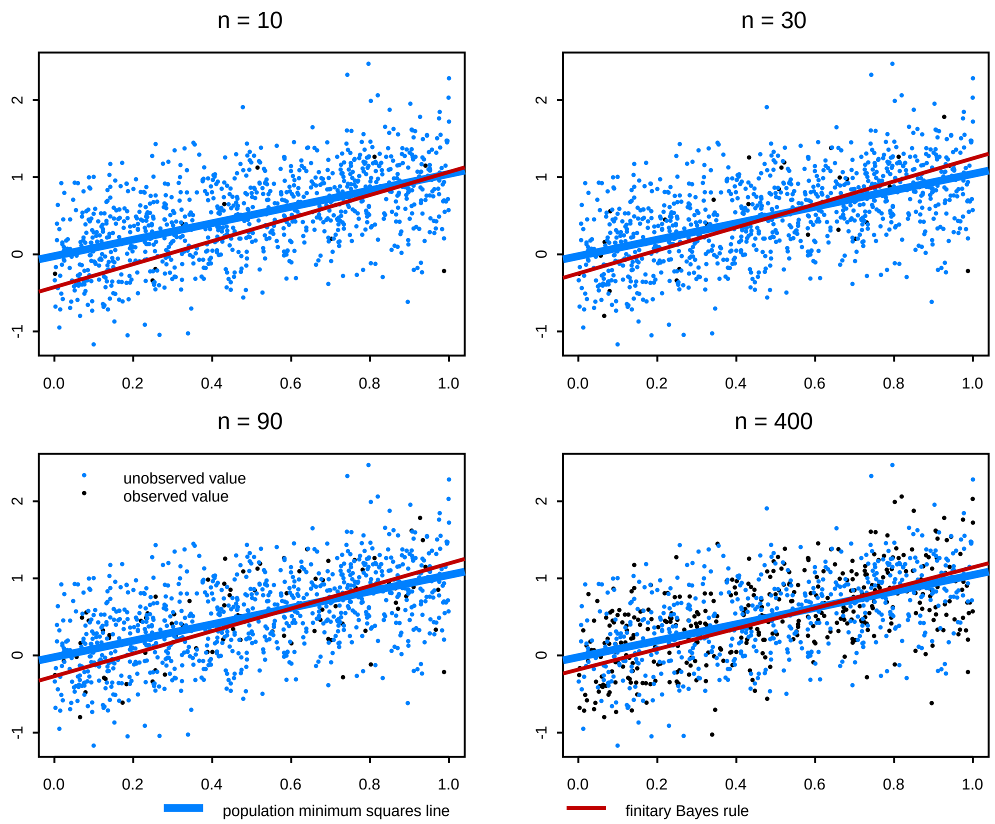


Figure 6.3: Linear regression. Population minimum squares (blue), finitary Bayes rule (red) line, for different sample sizes n.

Appendix A

Proofs

A.1 Proof of Theorem 3.4

Let $\tilde{\mathscr{S}}_N$ be the class of measurable rectangles of $\mathbb{X}^N$ of the form $B_{\beta_1} \times \cdots \times B_{\beta_N}$, where $\{B_1, \ldots, B_m\}$ is any (not fixed) finite partition of $\mathbb{X}$, B_j belongs to $\mathscr{G}$ for each j, and $\beta' = (\beta_1, \ldots, \beta_N)$ is a vector in $\{1, \ldots, m\}^N$. Now, let $\mathscr{S}_N = \tilde{\mathscr{S}}_N \cup \{\emptyset\}$. The first step of the proof of Theorem 3.4 consists of the following lemma:

Lemma A.1. *$\mathscr{S}_N$ is a semialgebra that generates the algebra $\mathscr{A}^N$ and the σ-algebra $\mathscr{X}^N$.*

Proof. In order to prove that $\mathscr{S}_N$ is a semialgebra, consider two sets A and B in $\mathscr{S}_N$, and assume that $A = A_{\alpha_1} \times \cdots \times A_{\alpha_N}$ and $B = B_{\beta_1} \times \cdots \times B_{\beta_N}$, where $\{A_1, \ldots, A_k\}$ and $\{B_1, \ldots, B_m\}$ are two partitions of sets in $\mathscr{G}$ and $\alpha' = (\alpha_1, \ldots, \alpha_N)$, $\beta' = (\beta_1, \ldots, \beta_N)$ are two vectors respectively in$\{1, \ldots, k\}^N$ and $\{1, \ldots, m\}^N$. Then

$$A \cap B = (A_{\alpha_1} \times \cdots \times A_{\alpha_N}) \cap (B_{\beta_1} \times \cdots \times B_{\beta_N}) = C_{\gamma_1} \times \cdots \times C_{\gamma_N},$$

where $C_{\gamma_i} := A_{\alpha_i} \cap B_{\beta_i}$. If $A \cap B \neq \emptyset$, then, for each i, C_{γ_i} belongs to the partition generated by the sets $A_1, \ldots, A_k, B_1, \ldots, B_m$. In any case, $A \cap B \in \mathscr{S}_N$. Hence $\mathscr{S}_N$ is closed under intersection. Moreover, if T belongs to $\mathscr{S}_N$ and $T = B_{\beta_1} \times \cdots \times B_{\beta_N}$, then

$$T^c = \cup_{\eta' \neq \beta'} B_{\eta_1} \times \cdots \times B_{\eta_N}$$

and therefore $\mathscr{S}_N$ is a semialgebra.

Let us prove that $\mathscr{A}^N$ is the algebra generated by $\mathscr{S}_N$. Denote by $\mathscr{A}(\mathscr{G}^N)$ the algebra generated by the class $\mathscr{G}^N$ of cartesian products of sets in $\mathscr{G}$. First, note that $\mathscr{A}(\mathscr{G}^N)$ coincides with the algebra $\mathscr{A}(\mathscr{S}_N)$ generated by $\mathscr{S}_N$. In fact, $\mathscr{A}(\mathscr{G}^N) \supset \mathscr{A}(\mathscr{S}_N)$ since $\mathscr{G}^N \supset \mathscr{S}_N$. On the other hand, if $D_1, \ldots, D_N$ are sets in $\mathscr{G}$, then the product $D_1 \times \cdots \times D_N$ is a finite union of sets in $\mathscr{S}_N$ (take the partition generated by $D_1, \ldots, D_N$); this means that $\mathscr{G}^N \subset \mathscr{A}(\mathscr{S}_N)$ and therefore $\mathscr{A}(\mathscr{G}^N) \subset \mathscr{A}(\mathscr{S}_N)$. But $\mathscr{A}(\mathscr{G}^N)$ coincides with $\mathscr{A}^N$. In fact, on one side $\mathscr{A}(\mathscr{G}^N) \subset \mathscr{A}^N$ since $\mathscr{G}^N$ is a subset of the class of cartesian products of sets in $\mathscr{A}$. On the other side, recall that $\mathscr{A}$ is the class of finite disjoint unions of sets in $\mathscr{G}$ [i.e. $\mathscr{A}$ is the algebra generated by $\mathscr{G}$] and therefore if $D_1, \ldots, D_N$ are sets in $\mathscr{A}$, $D_1 \times \cdots \times D_N$ is a finite disjoint union of sets in $\mathscr{G}^N$, in other words it belongs to $\mathscr{A}(\mathscr{G}^N)$ and therefore $\mathscr{A}^N \subset \mathscr{A}(\mathscr{G}^N)$.

Let us prove that $\mathscr{X}^N$ is the σ-algebra generated by $\mathscr{S}_N$. $\mathscr{G}^N$ (i.e. the class of cartesian products of sets in $\mathscr{G}$) and $\mathscr{S}_N$ generate the same σ-algebra since they generate the same algebra as we just proved. Hence, it is sufficient to prove that $\mathscr{G}^N$ generates $\mathscr{X}^N$. Since $\mathscr{G}^N$ is a subset of the class of measurable rectangles, the σ-algebra $\sigma(\mathscr{G}^N)$ generated by $\mathscr{G}^N$ is a subset of $\mathscr{X}^N$.

We shall prove that $\sigma(\mathscr{G}^N) \supset \mathscr{X}^N$ by induction about N.

Since $\mathscr{G}^1$ coincides with $\mathscr{G}$, the case $N = 1$ is trivial. Suppose that the thesis is true for $N - 1$. If C belongs to $\mathscr{G}^{N-1}$, let

$$\mathscr{F}_C := \big\{D : C \times D \in \sigma(\mathscr{G}^N)\big\}.$$

Note that $\mathscr{F}_C$ is a λ-system[1] containing the π-system[2] $\mathscr{G}$ and therefore, by Dynkin's $\pi - \lambda$ theorem it contains the σ-algebra generated by $\mathscr{G}$, that is $\mathscr{X}$. Hence, if $D \in \mathscr{X}$ and $C \in \mathscr{G}^{N-1}$, $C \times D \in \sigma(\mathscr{G}^N)$, i.e. $\mathscr{G}^{N-1} \subset \mathscr{F}_D$, where

$$\mathscr{F}_D := \big\{C : C \times D \in \sigma(\mathscr{G}^N)\big\}.$$

[1] A class Λ of sets is said to be a λ-system of subset of $\mathbb{X}$ if and only if:

1. $\mathbb{X} \in \Lambda$
2. $S, T \in \Lambda,\ S \subset T \Longrightarrow T \setminus S \in \Lambda$
3. $S_n \in \Lambda,\ S_n \uparrow S \Longrightarrow S \in \Lambda$.

[2] A π-system is a class of sets closed under intersections.

By Dynkin's theorem, $\mathscr{F}_D$ contains the sigma-algebra generated by $\mathscr{G}^{N-1}$, which is $\mathscr{X}^{N-1}$ by induction hypothesis. Therefore when $D_i \in \mathscr{X}$, $D_1 \times \cdots \times D_{N-1} \times D$ belongs to $\sigma(\mathscr{G}^N)$. This implies that $\mathscr{X}^N \subset \sigma(\mathscr{G}^N)$ as desired. $\square$

Proof of Theorem 3.4. Define $\tilde{\rho}_N$ on $\mathscr{S}_N$ by setting $\tilde{\rho}_N(\emptyset) = 0$ and

$$\tilde{\rho}_N(B_{\beta_1} \times \cdots \times B_{\beta_N}) := \frac{\psi_{B_1,\ldots,B_m}(N_1,\ldots,N_m)}{\binom{N}{N_1,\ldots,N_m}} \tag{A.1}$$

where $\{B_1,\ldots,B_m\}$ is a finite partition of $\mathbb{X}$, $B_j \in \mathscr{G}$ for each j, $(\beta_1,\ldots,\beta_N) \in \{1,\ldots,m\}^N$ and $N_j = |\{i = 1,\ldots,N : \beta_i = j\}|$. Recall that $\psi_{B_1,\ldots,B_m}$ is the p.m.f. of $(\tilde{N}(B_1),\ldots,\tilde{N}(B_m))$. When $m = 2$ we shall write for simplicity $\psi_{B_1}(N_1)$ instead of $\psi_{B_1,B_2}(N_1, N - N_1)$. We shall prove that $\tilde{\rho}_N$ is finitely additive, i.e. for any $B \in \mathscr{S}_N$ that is a finite disjoint union of sets $A_{\alpha'} = A_{\alpha_1} \times \cdots \times A_{\alpha_N} \in \mathscr{S}_N$, with $\alpha' = (\alpha_1,\ldots,\alpha_N) \in \mathcal{A}$ and $\mathcal{A} \subset \mathbb{N}^N$ such that $\tilde{\rho}_N(B) = \sum_{\alpha' \in \mathcal{A}} \tilde{\rho}_N(A_{\alpha'})$. Since $B \in \mathscr{S}_N$, B can be written as $B_{\beta_1} \times \cdots \times B_{\beta_N}$, where $\{B_1,\ldots,B_m\}$ is a partition whose elements belong to $\mathscr{G}$. Suppose that $\{B_{1,1},\ldots,B_{m,k_m}\}$ is a partition not coarser than $\{B_1,\ldots,B_m\}$ such that $B_{j,l} \in \mathscr{G}$ and $B_{j,l}$ is a subset of B_j for $l = 1,\ldots,k_j$ and $j = 1,\ldots,m$. Let

$$\mathcal{S} := \left\{\gamma' : \gamma' = (\gamma_1,\ldots,\gamma_N) \in \mathbb{N}^N, \gamma_i = 1,\ldots,k_{\beta_i} \text{ for } i = 1,\ldots,N\right\}.$$

Hence,

$$B = \bigcup_{\gamma' \in \mathcal{S}} (B_{\beta_1,\gamma_1} \times \cdots \times B_{\beta_N,\gamma_N}),$$

where the sets in the union are pairwise disjoint and the union is finite. Note that

$$\tilde{\rho}_N(B) = \sum_{\gamma' \in \mathcal{S}} \tilde{\rho}_N(B_{\beta_1,\gamma_1} \times \cdots \times B_{\beta_N,\gamma_N}). \tag{A.2}$$

In fact if we let

$$\mathcal{T} := \left\{N' = (N_{1,1},\ldots,N_{m,k_m}) \in \mathbb{N}^{\sum k_j} : \sum_{l=1}^{k_j} N_{j,l} = N_l \text{ for } j = 1,\ldots,m\right\}$$

and

$$\mathcal{S}(N') := \Big\{\gamma' : \gamma' = (\gamma_1,\ldots,\gamma_N) \in \mathbb{N}^N,$$
$$|\{i = 1,\ldots,N : \gamma_i = l,\, \beta_i = j\}| = N_{j,l} \text{ for } l = 1,\ldots,k_j \text{ and } j = 1,\ldots,m\Big\}$$

for each $N' \in \mathcal{T}$, then by (A.1),

$$\begin{aligned}
&\sum_{\gamma' \in \mathcal{S}} \tilde{\rho}_N(B_{\beta_1,\gamma_1} \times \cdots \times B_{\beta_N,\gamma_N}) \\
&= \sum_{N' \in \mathcal{T}} \sum_{\gamma' \in \mathcal{S}(N')} \tilde{\rho}_N(B_{\beta_1,\gamma_1} \times \cdots \times B_{\beta_N,\gamma_N}) \\
&= \sum_{N' \in \mathcal{T}} \sum_{\gamma' \in \mathcal{S}(N')} \frac{\psi_{B_{1,1},\ldots,B_{m,k_m}}(N_{1,1},\ldots,N_{m,k_m})}{\binom{N}{N_{1,1},\ldots,N_{m,k_m}}}.
\end{aligned} \tag{A.3}$$

Since in the last term of (A.3) each addend depends only on N', we obtain

$$\begin{aligned}
&\sum_{\gamma' \in \mathcal{S}} \tilde{\rho}_N(B_{\beta_1,\gamma_1} \times \cdots \times B_{\beta_N,\gamma_N}) \\
&= \sum_{N' \in \mathcal{T}} \binom{N_1}{N_{1,1},\ldots,N_{1,k_1}} \cdots \binom{N_m}{N_{m,1},\ldots,N_{m,k_m}} \frac{\psi_{B_{1,1},\ldots,B_{m,k_m}}(N_{1,1},\ldots,N_{m,k_m})}{\binom{N}{N_{1,1},\ldots,N_{m,k_m}}} \\
&= \frac{\binom{N_1}{N_{1,1},\ldots,N_{1,k_1}} \cdots \binom{N_m}{N_{m,1},\ldots,N_{m,k_m}}}{\binom{N}{N_{1,1},\ldots,N_{m,k_m}}} \sum_{N' \in \mathcal{T}} \psi_{B_{1,1},\ldots,B_{m,k_m}}(N_{1,1},\ldots,N_{m,k_m}) \\
&= \frac{1}{\binom{N}{N_1,\ldots,N_m}} \sum_{N' \in \mathcal{T}} \psi_{B_{1,1},\ldots,B_{m,k_m}}(N_{1,1},\ldots,N_{m,k_m}) \\
&= \frac{\psi_{B_1,\ldots,B_m}(N_1,\ldots,N_m)}{\binom{N}{N_1,\ldots,N_m}} = \tilde{\rho}_N(B).
\end{aligned} \tag{A.4}$$

In fact note that by hypothesis (2),

$$\psi_{B_1,\ldots,B_m}(N_1,\ldots,N_m) = \sum_{N' \in \mathcal{T}} \psi_{B_{1,1},\ldots,B_{m,k_m}}(N_{1,1},\ldots,N_{m,k_m}). \tag{A.5}$$

We are now in position to prove that $\tilde{\rho}_N$ is finitely-additive on $\mathscr{S}_N$. Let $(A_{\alpha'})_{\alpha' \in \mathcal{A}}$ be a class of pairwise disjoint sets belonging to $\mathscr{S}_N$, where $\mathcal{A}$ is a finite subset of $\mathbb{N}^N$, $A_{\alpha'} := A_{(\alpha_1,\ldots,\alpha_N)} := A_{1,\alpha_1} \times \cdots \times A_{N,\alpha_N}$, and suppose that $B = A_1 \times \cdots \times A_N := \cup_{\alpha' \in \mathcal{A}} A_{\alpha'}$ also belongs to $\mathscr{S}_N$. Notice that $A_i = \cup_{\alpha' \in \mathcal{A}} A_{i,\alpha_i}$ for each i. Denote by $\{C_1, \ldots, C_k\}$ the partition generated by the class of sets $\{A_{1,\alpha_1}, \ldots, A_{N,\alpha_N} : \alpha' \in \mathcal{A}\}$. The elements of such partition belong to $\mathscr{G}$ since $\mathscr{G}$ is closed under intersection. Then applying (A.2) twice, we obtain that

$$\begin{aligned}
\tilde{\rho}_N(B) &= \sum_{\substack{C_{l_i} \subset A_i \\ i=1,\ldots,N}} \tilde{\rho}_N(C_{l_1} \times \cdots \times C_{l_N}) \\
&= \sum_{\alpha' \in \mathcal{A}} \sum_{\substack{C_{l_i} \subset A_{i,\alpha_i} \\ i=1,\ldots,N}} \tilde{\rho}_N(C_{l_1} \times \cdots \times C_{l_N}) \\
&= \sum_{\alpha' \in \mathcal{A}} \tilde{\rho}_N(A_{1,\alpha_1} \times \cdots \times A_{N,\alpha_N}).
\end{aligned} \tag{A.6}$$

Therefore $\tilde{\rho}_N$ is finitely-additive on $\mathscr{S}_N$. Hence, it is known that $\tilde{\rho}_N$ has a unique finitely-additive extension ρ_N on the algebra $\mathscr{A}^N$ generated by $\mathscr{S}_N$[3].

Let us demonstrate the second part of the thesis. We already proved that hypothesis (3) is necessary for σ-additivity of ρ_N, but we shall prove that it is also sufficient.

We resort to a V.V. Sazonov's result about perfect measures. Sazonov (1965) calls *perfect* any measure μ on some measurable space $(\Omega, \mathscr{F})$ if, for each real-valued $\mathscr{F}$-measurable function f and for each subset E of the line such that $f^{-1}(E) \in \mathscr{F}$, there exists a Borel set B such that $B \subset E$ and $\mu(f^{-1}(E)) = \mu(f^{-1}(B))$. Any measure on $(\mathbb{X}, \mathscr{X})$ is perfect since $\mathbb{X}$ is a Polish space endowed with its Borel σ-algebra $\mathscr{X}$. He shows that any finitely-additive measure μ on the algebra of rectangles such that each marginal is a perfect measure is countably additive. It is known that any measure on the Borel σ-field of a Polish (i.e. complete separable metric) space is perfect, since any tight measure μ on a metric space (i.e. such that $\forall \epsilon > 0$, $\exists$ a compact set K_ϵ such that $\mu(K_\epsilon^c) < \epsilon$) is perfect and any measure on a Polish space is tight [see Parthasarathy (1967), pages 28–32]. So any measure on $(\mathbb{X}, \mathscr{X})$ is perfect. Moreover recall that for a finitely additive measure on an algebra, continuity from above is sufficient for countable additivity. Consequently, we only need to prove that if hypothesis (3) holds, than $\rho_{(i)}(A) := \rho_N(\mathbb{X}^{i-1} \times A \times \mathbb{X}^{N-i})$ is continuous from above on $\mathscr{A}$.

Let $(C_n)_n$ be a sequence of events in $\mathscr{A}$ that converges from above to the empty set. Since $\tilde{\rho}_N$ is finitely additive and the coordinate functions $\xi_1, \ldots, \xi_N$ are identically distributed under ρ_N, then

$$\begin{aligned} \rho_N(\mathbb{X}^{i-1} \times C_n \times \mathbb{X}^{N-i}) &= \rho_N(\{\xi_i \in C_n\}) \\ &= \mathbb{E}(\tilde{e}(C_n)), \end{aligned} \tag{A.7}$$

which converges to zero since $(\tilde{e}(C_n))_n$ is a sequence of r.v.'s having – by hypothesis (1) – the same finite support $\{0, 1, \ldots, N\}$ when hypothesis (3) is satisfied.

In conclusion, when hypothesis (3) holds, ρ_N is σ-additive on $\mathscr{A}^N$, and therefore, by Carathéodary's Extension Theorem, ρ_N has an unique extension[3] that is a measure on the σ-field $\mathscr{X}^N$ generated by $\mathscr{A}^N$. $\square$

[3]See for instance the appendix on measure theory of Durrett (1996),pages 440–464.

A.2 Properties of partitions tree distributions

This section contains the proofs of some properties of partitions tree distributions. To begin, let us recall some notation already introduced in Chapter 4. As usual, let $\xi_1, \dots, \xi_N$ be the coordinate functions on $(\mathbb{X}^N, \mathscr{X}^N, P)$, and let $\Pi = (\pi_m)_{m=0}^{\infty}$ be a separating binary tree of partitions of $\mathbb{X}$, such that $\mathscr{G} = \cup_{m \geq 0} \pi_m$ generates $\mathscr{X}$. Given B in π_m ($m \geq 1$), define $\mathrm{ge}(B) = C$, where $C \in \pi_{m-1}$ and $B \subset C$. Moreover, let

$$\mathrm{ge}^{(0)}(\cdot) := \cdot \ , \qquad \mathrm{ge}^{(1)}(\cdot) := \mathrm{ge}(\cdot) \ ,$$
$$\mathrm{ge}^{(2)}(\cdot) := \mathrm{ge}(\mathrm{ge}(\cdot)) \ , \qquad \mathrm{ge}^{(3)}(\cdot) := \mathrm{ge}(\mathrm{ge}^{(2)}(\cdot)) \ , \quad \dots$$

Remark A.2. Condition 4.1.1 requires that for each B in $\mathscr{G}$ there exists a constant c_B such that

$$\mathbb{E}(\tilde{N}(B) \mid \tilde{N}(\mathrm{ge}(B))) = c_B \tilde{N}(\mathrm{ge}(B)). \tag{A.8}$$

Taking expectation on both sides, one obtains:

$$\mathbb{E}(\tilde{N}(B)) = c_B \, \mathbb{E}(\tilde{N}(\mathrm{ge}(B))),$$

and therefore

$$c_B = \frac{\mathbb{E}(\tilde{N}(B))}{\mathbb{E}(\tilde{N}(\mathrm{ge}(B)))}$$

whenever $P(\tilde{N}(\mathrm{ge}(B)) = 0)) < 1$, while in the other case (A.8) is satisfied for any value of c_B. In other words, Condition 4.1.3 holds if and only if

$$\mathbb{E}\left(\frac{\tilde{N}(B)}{\tilde{N}(\mathrm{ge}(B))} \,\middle|\, \tilde{N}(\mathrm{ge}(B)) = M \right) = \frac{\mathbb{E}(\tilde{N}(B))}{\mathbb{E}(\tilde{N}(\mathrm{ge}(B)))} \tag{A.9}$$

for each B in $\mathscr{G}$ such that $\tilde{N}(B)$ is not degenerate at zero, and for each $M \geq 1$ such that $P(\tilde{N}(\mathrm{ge}(B))) = M)$ is positive.

We observe an important consequence of Condition 4.1.2:

Proposition A.3. *Let $(B_{(m)})_{m\in\mathbb{N}}$ be a sequence of sets such that $B_{(m)}$ belongs to π_m, and $B_{(m+1)} \subset B_{(m)}$ for each $m \in \mathbb{N}$.*
If Condition 4.1.2 holds, then the sequence $(\tilde{N}(B_{(m)}))_{m\in\mathbb{N}}$ of r.v.'s is a Markov Chain.

Proof. By Condition 4.1.2,

$$\begin{aligned} &P(\tilde{N}(B_{(m)}) = M_m \mid \tilde{N}(B_{(k)}) = M_k,\, k = 1, \dots, m-1) \\ &\quad = P(\tilde{N}(B_{(m)}) = M_m \mid \tilde{N}(B_{(m-1)}) = M_{m-1}, \sum_{\substack{C\in\pi_{m-1}: \\ C\subset B_{(k-1)}\setminus B_{(k)}}} \tilde{N}(C) = M_{k-1} - M_k,\ k < m) \\ &\quad = P(\tilde{N}(B_{(m)}) = M_m \mid \tilde{N}(B_{(m-1)}) = M_{m-1}). \end{aligned}$$

□

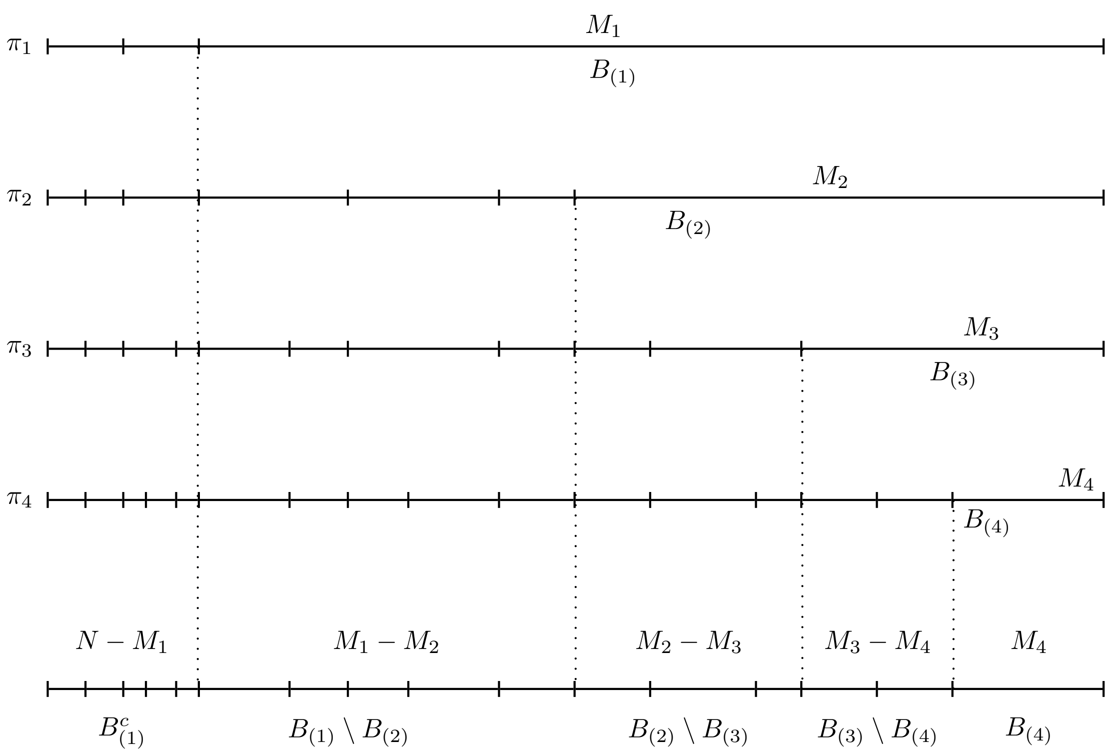


Figure A.1: Scheme for the proof of Proposition A.3.

Proposition A.4. *If Condition 4.1.3.a holds, then, for each B in $\mathscr{G}$ such that $\tilde{N}(B)$ is not degenerate at zero,*

$$P\{\xi_1 \in B\} = \mathbb{E}\left(\frac{\tilde{N}(B_1)}{N}\right) \prod_{j=2}^{m} \mathbb{E}\left(\frac{\tilde{N}(B_j)}{M_{j-1}} \mid \tilde{N}(B_{j-1}) = M_{j-1}\right) \qquad (B \in \pi_m), \tag{A.10}$$

where $B_m = B$ and, for $j < m$, B_j denotes the set in π_j containing B_m, and M_j is any positive value such that $P(\tilde{N}(B_j) = M_j)$ is positive.
Equivalently,

$$P(\xi_1 \in B) = P(\xi_1 \in \mathrm{ge}(B)) \cdot \mathbb{E}\left(\frac{\tilde{N}(B)}{M} \mid \tilde{N}(\mathrm{ge}(B)) = M\right),$$

for each B in $\mathscr{G}$ such that $\tilde{N}(B)$ is not degenerate at zero and for each positive M such that $P(\tilde{N}(\mathrm{ge}(B)) = M)$ is positive.

Proof. Write the left hand side in (A.10) as

$$\begin{aligned} P\{\xi_1 \in B\} &= \frac{\mathbb{E}(\tilde{N}(B_m))}{N} \\ &= \frac{\mathbb{E}(\tilde{N}(B_1))}{N} \prod_{j=2}^{m} \frac{\mathbb{E}(\tilde{N}(B_j))}{\mathbb{E}(\tilde{N}(B_{j-1}))}, \end{aligned}$$

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

Remark A.5 (Some properties of the law of the empirical measure). We now point out some properties of the law of the empirical measure, which are just a consequence of exchangeability of P. Let us denote $\tilde{N}(\cdot) := N\,\tilde{e}(\cdot) = \sum_{i=1}^{N} \delta_{\xi_i}(\cdot)$. If B belongs to $\mathscr{X}$ and $\{B_1, \dots, B_m\}$ is a measurable partition of $\mathbb{X}$, then:

$$P(\xi_1 \in B) = \mathbb{E}(\tilde{e}(B)) = \mathbb{E}\left(\frac{\tilde{N}(B)}{N}\right) \tag{A.11a}$$

$$\begin{aligned} &P\Big\{\xi_1 \in B, \tilde{N}(B_1) = N_1, \dots, \tilde{N}(B_m) = N_m\Big\} \\ &\qquad = \mathbb{E}\left(\frac{\tilde{N}(B)}{N}\, \mathbb{I}_{\{\tilde{N}(B_1)=N_1,\dots,\tilde{N}(B_m)=N_m\}}\right). \end{aligned} \tag{A.11b}$$

Moreover if $B = B_j$ for some j, then

$$\begin{aligned} &P\Big\{\xi_1 \in B, \tilde{N}(B_1) = N_1, \dots, \tilde{N}(B_m) = N_m\Big\} \\ &\qquad = \frac{N_j}{N}\, P\Big\{\tilde{N}(B_1) = N_1, \dots, \tilde{N}(B_m) = N_m\Big\}. \end{aligned} \tag{A.11c}$$

A.2.1 Proof of Proposition 4.1 and Proposition 4.2

In order to demonstrate Propositions 4.1 and 4.2, we need some lemmas.

Lemma A.6. *Equation*

$$\mathcal{L}_{\tilde{N}_{m+1}|\tilde{N}_m} = \bigtimes_{C\in\pi_m} \mathcal{L}_{\tilde{N}(B):\,\mathrm{ge}(B)=C|\tilde{N}(C)} \tag{A.12}$$

holds true for each m in $\mathbb{N}$ if and only if

$$\mathcal{L}_{\tilde{N}_{m+h}|\tilde{N}_m} = \bigtimes_{C\in\pi_m} \mathcal{L}_{\tilde{N}(B):\,\mathrm{ge}^{(h)}(B)=C|\tilde{N}(C)}, \tag{A.13}$$

for any $h \in \mathbb{N}^+$ and for any $m \in \mathbb{N}$.

Proof. Of course, (A.13) implies (A.12) (take $h = 1$). We shall prove by induction on h that (A.12) implies (A.13). Therefore, let us suppose that (A.12) holds and that

$$\mathcal{L}_{\tilde{N}_{m+h-1}|\tilde{N}_m} = \bigtimes_{C\in\pi_m} \mathcal{L}_{\tilde{N}(B):\,\mathrm{ge}^{(h-1)}(B)=C|\tilde{N}(C)} \qquad \text{for each } m \geq 1. \tag{A.14}$$

Hence, for any $m \geq 1$ and for any vector $N(k_m) = (N_{m,1}, \ldots, N_{m,k_m})$ in $\mathbb{N}^{k_m}$ whose components sum up to N,

$$\begin{aligned} P(\tilde{N}_{m+h} = N(k_{m+h}) \mid \tilde{N}_m = N(k_m)) \\ &= P(\tilde{N}_{m+h} = N(k_{m+h}) \mid \tilde{N}_{m+1} = N(k_{m+1})) \\ &\quad \cdot P(\tilde{N}_{m+1} = N(k_{m+1}) \mid \tilde{N}_m = N(k_m)), \end{aligned} \tag{A.15}$$

where for any $r \in \mathbb{N}^+$, $N(k_{m+r}) := (N_{m+r,1}, \ldots, N_{m+r,k_{m+r}}) \in \mathbb{N}^{k_{m+r}}$ is such that

$$\sum_{l:\, B_{m+r,\,l} \subset B_{m,\,j}} N_{m+r,\,l} = N_{m,\,j} \qquad \text{and} \qquad P(\tilde{N}_{m+1} = N(k_{m+1})) > 0.$$

By (A.12) and by induction hypothesis (A.14), (A.15) becomes:

$$
\begin{aligned}
&P(\tilde{N}_{m+h} = N(k_{m+h}) \mid \tilde{N}_m = N(k_m)) \\
&\quad = \prod_{l=1}^{k_m+1} P\Big(\bigcap_{s:B_{m+h,s}\subset B_{m+1,l}} \{\tilde{N}(B_{m+h,s}) = N_{m+h,s}\} \mid \tilde{N}(B_{m+1,l}) = N_{m+1,l}\Big) \\
&\quad\quad \cdot \prod_{j=1}^{k_m} P\Big(\bigcap_{w:B_{m+1,w}\subset B_{m,j}} \{\tilde{N}(B_{m+1,w}) = N_{m+1,w}\} \mid \tilde{N}(B_{m,j}) = N_{m,j}\Big) \\
&\quad = \prod_{j=1}^{k_m} \Bigg[P\Big(\bigcap_{w:B_{m+1,w}\subset B_{m,j}} \{\tilde{N}(B_{m+1,w}) = N_{m+1,w}\} \mid \tilde{N}(B_{m,j}) = N_{m,j}\Big) \\
&\quad\quad \cdot \prod_{\substack{l: \\ B_{m+1,l}\subset B_{m,j}}} P\Big(\bigcap_{s:B_{m+h,s}\subset B_{m+1,l}} \{\tilde{N}(B_{m+h,s}) = N_{m+h,s}\} \mid \tilde{N}(B_{m+1,l}) = N_{m+1,l}\Big)\Bigg].
\end{aligned}
\tag{A.16}
$$

Again by induction hypothesis (A.14) from the last expression we obtain:

$$
\begin{aligned}
&P(\tilde{N}_{m+h} = N(k_{m+h}) \mid \tilde{N}_m = N(k_m)) \\
&\quad = \prod_{j=1}^{k_m} \Bigg[P\Big(\bigcap_{w:B_{m+1,w}\subset B_{m,j}} \{\tilde{N}(B_{m+1,w}) = N_{m+1,w}\} \mid \tilde{N}(B_{m,j}) = N_{m,j}\Big) \\
&\quad\quad \cdot P\Big(\bigcap_{s:B_{m+h,s}\subset B_{m,j}} \{\tilde{N}(B_{m+h,s}) = N_{m+h,s}\} \mid \tilde{N}(B_{m+1,l}) = N_{m+1,l}\Big)\Bigg] \\
&\quad = \prod_{j=1}^{k_m} \Bigg[P\Big(\bigcap_{w:B_{m+1,w}\subset B_{m,j}} \{\tilde{N}(B_{m+1,w}) = N_{m+1,w}\} \mid \tilde{N}(B_{m,j}) = N_{m,j}\Big) \\
&\quad\quad \cdot P\Big(\bigcap_{s:B_{m+h,s}\subset B_{m,j}} \{\tilde{N}(B_{m+h,s}) = N_{m+h,s}\} \mid \bigcap_{w:B_{m+1,w}\subset B_{m,j}} \{\tilde{N}(B_{m+1,w}) = N_{m+1,w}\}\Big)\Bigg] \\
&\quad = \prod_{j=1}^{k_m} \Bigg[P\Big(\bigcap_{s:B_{m+h,s}\subset B_{m,j}} \{\tilde{N}(B_{m+h,s}) = N_{m+h,s}\} \mid \tilde{N}(B_{m,j}) = N_{m,j}\Big)\Bigg]
\end{aligned}
\tag{A.17}
$$

as desired. □

Lemma A.7. *If Condition 4.1.2 holds, then for any m and l such that $m > l$ and for any $B \in \pi_m$, $\tilde{N}(B)$ and $(\tilde{N}(C) : C \in \pi_{m-l} \setminus \{\mathrm{ge}^{(l)}(B)\})$ are conditionally independent given $\tilde{N}(\mathrm{ge}^{(l)}(B))$, i.e.*

$$
\mathcal{L}_{(\tilde{N}(B))|\tilde{N}_{m-l}} = \mathcal{L}_{(\tilde{N}(B))|\tilde{N}(\mathrm{ge}^{(l)}(B))}. \tag{A.18}
$$

Proof. We shall give the proof by induction on l. For $l = 1$, the hypothesis trivially implies the thesis. Hence suppose that:

$$\mathcal{L}_{(\tilde{N}(B))|\tilde{N}_{m-l+1}} = \mathcal{L}_{(\tilde{N}(B))|\tilde{N}(\mathrm{ge}^{(l-1)}(B))}. \tag{A.19}$$

If we let $N(k_{m-l}) \in \{0, \ldots, N\}^{k_{m-l}}$ such that $P(\tilde{N}_{m-l} = N(k_{m-l})) > 0$, then:

$$\begin{aligned} &P(\tilde{N}(B) = M \mid \tilde{N}_{m-l} = N(k_{m-l})) \\ &\quad = \sum_{N(k_{m-1}) \in \mathcal{C}} P(\tilde{N}(B) = M \mid \tilde{N}_{m-1} = N(k_{m-1})) \\ &\qquad\qquad \cdot P(\tilde{N}_{m-1} = N(k_{m-1}) \mid \tilde{N}_{m-l} = N(k_{m-l})), \end{aligned} \tag{A.20}$$

where $\mathcal{C}$ is the set of all vectors $N(k_{m-1}) = (N_{m-1,1}, \ldots, N_{m-1,k_{m-1}})$ such that $\sum\limits_{B_{m-1,h} \subset B_{m-l,j}} N_{m-1,h} = N_{m-l,j}$ for each j.

For simplicity let us assume that $\mathrm{ge}(B) = B_{m-1,1}$ and that $\mathrm{ge}^{(l)}(B) = B_{m-l,1}$. Moreover, denote:

$$\mathcal{C}_{x_1} := \{(x_2, \ldots, x_{k_{m-1}}) : (x_1, \ldots, x_{k_{m-1}}) \in \mathcal{C}\}.$$

Hence, by hypothesis and by (A.19), (A.20) becomes:

$$\begin{aligned} &P(\tilde{N}(B) = M \mid \tilde{N}_{m-l} = N(k_{m-l})) \\ &\quad = \sum_{N_{m-1,1}=0}^{N_{m-l,1}} P(\tilde{N}(B) = M \mid \tilde{N}(B_{m-1,1}) = N_{m-1,1}) \\ &\qquad \cdot \sum_{(N_{m-1,2},\ldots,N_{m-1,k_{m-1}}) \in \mathcal{C}_{N_{m-1,1}}} P(\tilde{N}_{m-1} = N(k_{m-1}) \mid \tilde{N}_{m-l} = N(k_{m-l})) \\ &\quad = \sum_{N_{m-1,1}=0}^{N_{m-l,1}} P(\tilde{N}(B) = M \mid \tilde{N}(B_{m-1,1}) = N_{m-1,1}) \\ &\qquad \cdot P(\tilde{N}(B_{m-1,1}) = N_{m-1,1} \mid \tilde{N}_{m-l} = N(k_{m-l})) \\ &\quad = \sum_{N_{m-1,1}=0}^{N_{m-l,1}} P(\tilde{N}(B) = M \mid \tilde{N}(B_{m-1,1}) = N_{m-1,1}) \\ &\qquad \cdot P(\tilde{N}(B_{m-1,1}) = N_{m-1,1} \mid \tilde{N}(B_{m-l,1}) = N_{m-l,1}) \\ &\quad = P(\tilde{N}(B) = M \mid \tilde{N}(B_{m-l,1}) = N_{m-l,1}). \end{aligned}$$

□

Define, for each C in $\mathscr{G}$ such that $\tilde{N}(C)$ is not degenerate at zero,

$$\mathcal{S}(C) = \Big\{M = 1, \dots, N : P(\tilde{N}(C) = M) > 0\Big\}.$$

Lemma A.8. *If Condition 4.1.2 and Condition 4.1.3 hold, then for any m and l such that $m > l$ and for any $B \in \pi_m$ such that $\tilde{N}(\mathrm{ge}^{(l)}(B))$ is not degenerate at zero, the conditional expectation $\mathbb{E}\left(\frac{\tilde{N}(B)}{M} \mid \tilde{N}(\mathrm{ge}^{(l)}(B)) = M\right)$ does not depend on M, as M varies in $\mathcal{S}(\mathrm{ge}^{(l)}(B))$, that is*

$$\mathbb{E}\left(\frac{\tilde{N}(B)}{\tilde{N}(\mathrm{ge}^{(l)}(B))} \mid \tilde{N}(\mathrm{ge}^{(l)}(B)) = M\right) = \frac{\mathbb{E}(\tilde{N}(B))}{\mathbb{E}(\tilde{N}(\mathrm{ge}^{(l)}(B)))}$$

for each M in $\mathcal{S}(\mathrm{ge}^{(l)}(B))$.

Proof. We can prove this by induction on l. If $l = 1$, the thesis coincides with Condition 4.1.3.

Now assume that, for each $B \in \mathscr{G}$ such that $P(\tilde{N}(\mathrm{ge}^{(l-1)}(B)) = 0) < 1$, and for each M in $\mathcal{S}(\mathrm{ge}^{(l-1)}(B))$,

$$\mathbb{E}\left(\frac{\tilde{N}(B)}{\tilde{N}(\mathrm{ge}^{(l-1)}(B))} \mid \tilde{N}(\mathrm{ge}^{(l-1)}(B)) = M\right) = \frac{\mathbb{E}(\tilde{N}(B))}{\mathbb{E}(\tilde{N}(\mathrm{ge}^{(l-1)}(B)))}. \tag{A.21}$$

Equation (A.21) is tantamount to saying that

$$\mathbb{E}(\tilde{N}(B) \mid \tilde{N}(\mathrm{ge}^{(l-1)}(B))) = \frac{\mathbb{E}(\tilde{N}(B))}{\mathbb{E}(\tilde{N}(\mathrm{ge}^{(l-1)}(B)))}\tilde{N}(\mathrm{ge}^{(l-1)}(B)). \tag{A.22}$$

By Condition 4.1.2 we can apply Proposition A.3 and we obtain:

$$\begin{aligned}\mathbb{E}(\tilde{N}(B) \mid \tilde{N}(\mathrm{ge}^{(l)}(B))) &= \mathbb{E}(\mathbb{E}(\tilde{N}(B) \mid \tilde{N}(\mathrm{ge}^{(l)}(B)), \tilde{N}(\mathrm{ge}(B))) \mid \tilde{N}(\mathrm{ge}^{(l)}(B))) \\ &= \mathbb{E}(\mathbb{E}(\tilde{N}(B) \mid \tilde{N}(\mathrm{ge}(B))) \mid \tilde{N}(\mathrm{ge}^{(l)}(B))),\end{aligned} \tag{A.23}$$

which by Condition 4.1.3 becomes:

$$\mathbb{E}\left(\tilde{N}(\mathrm{ge}(B))\frac{\mathbb{E}(\tilde{N}(B))}{\mathbb{E}\,\tilde{N}(\mathrm{ge}(B))} \mid \tilde{N}(\mathrm{ge}^{(l)}(B))\right) = \frac{\mathbb{E}(\tilde{N}(B))}{\mathbb{E}\,\tilde{N}(\mathrm{ge}(B))}\,\mathbb{E}(\tilde{N}(\mathrm{ge}(B)) \mid \tilde{N}(\mathrm{ge}^{(l)}(B))). \tag{A.24}$$

By induction hypothesis (A.21), the right hand side of (A.24) is equal to:

$$\frac{\mathbb{E}(\tilde{N}(B))}{\mathbb{E}\,\tilde{N}(\mathrm{ge}(B))}\frac{\mathbb{E}(\tilde{N}(\mathrm{ge}(B)))}{\mathbb{E}\,\tilde{N}(\mathrm{ge}^{(l)}(B))}\tilde{N}(\mathrm{ge}^{(l)}(B)) = \frac{\mathbb{E}(\tilde{N}(B))}{\mathbb{E}\,\tilde{N}(\mathrm{ge}^{(l)}(B))}\tilde{N}(\mathrm{ge}^{(l)}(B)).$$

So, for each $M \in \mathcal{S}(\mathrm{ge}^{(l)}(B))$,

$$\mathbb{E}(\tilde{N}(B) \mid \tilde{N}(\mathrm{ge}^{(l)}(B)) = M) = \frac{\mathbb{E}(\tilde{N}(B))}{\mathbb{E}\,\tilde{N}(\mathrm{ge}^{(l)}(B))}M,$$

which yields the thesis. $\square$

Recall that, for each C in $\mathscr{G}$, $\mathfrak{T}(C)$ denotes the vector obtained ordering the elements of the set $\{j = 1, \ldots, k_m : \mathrm{ge}(B_{m,j}) = C\}$.

Lemma A.9. *Assume that Conditions 4.1.1 – 4.1.2 are satisfied.*
Let $A_1, \ldots, A_N$ be N sets in π_{m+1}, and denote

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

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 $n \leq N$,*

$$\begin{aligned} &P(\tilde{N}(B_{m+1,1} = N_1, \ldots, \tilde{N}(B_{m+1,k_{m+1}}) = N_{k_{m+1}} \mid \xi_1 \in A_1, \ldots, \xi_n \in A_n, \tilde{N}(C) = N_C : C \in \pi_m) \\ &= \prod_{C \in \pi_m} P(\tilde{N}(B_{m+1,j}) = N_j : j \in \mathfrak{T}(C) \mid \tilde{N}(C) = N_C,\, \xi_i \in A_i : i \leq n, \mathrm{ge}(A_i) = C), \end{aligned} \tag{A.25}$$

where $N_C := \sum_{j \in \mathfrak{T}(C)} N_j$ for each $C \in \pi_m$.

Proof. Notice that

$$\begin{aligned} &P(\xi_1 \in A_1, \ldots, \xi_N \in A_N \mid \tilde{N}(C) = N_C : C \in \pi_m) \\ &\quad = P(\tilde{N}(B_{m+1,1} = N_1, \ldots, \tilde{N}(B_{m+1,k_{m+1}}) = N_{k_{m+1}} \mid \tilde{N}(C) = N_C : C \in \pi_m)) \\ &\qquad\qquad / \prod_{C \in \pi_m} \binom{N_C}{N_j : j \in \mathfrak{T}(C)}, \end{aligned} \tag{A.26}$$

which, by hypothesis, becomes

$$\begin{aligned} &\prod_{C \in \pi_m} \left(P(\tilde{N}(B_{m+1,j}) = N_j : j \in \mathfrak{T}(C) \mid \tilde{N}(C) = N_C) / \binom{N_C}{N_j : j \in \mathfrak{T}(C)} \right) \\ &\quad = \prod_{C \in \pi_m} P(\xi_i \in A_i : i \leq N, \mathrm{ge}(A_i) = C \mid \tilde{N}(C) = N_C). \end{aligned} \tag{A.27}$$

Combining equations (A.26) and (A.27), one obtains

$$\begin{aligned} &P(\xi_1 \in A_1, \ldots, \xi_N \in A_N \mid \tilde{N}(C) = N_C : C \in \pi_m) = \\ &\qquad \prod_{C \in \pi_m} P(\xi_i \in A_i : i \leq N, \mathrm{ge}(A_i) = C \mid \tilde{N}(C) = N_C). \end{aligned} \tag{A.28}$$

At this stage, denote

$$\begin{aligned} n_j &:= |\{i = 1, \ldots, n : A_i = B_{m+1,j}\}| && (j = 1, \ldots, k_{m+1}) \\ n_C &:= \sum_{j \in \mathfrak{T}(C)} n_j && (\text{for each } C \text{ in } \pi_m) \end{aligned}$$

Hence, the left hand side of (A.25) is equal to

$$\begin{aligned}&P(\xi_{n+1}\in A_{n+1},\dots,\xi_N\in A_N\mid \xi_1\in A_1,\dots,\xi_n\in A_n,\,\tilde N(C)=N_C:C\in\pi_m)\\&\qquad\Big/\prod_{C\in\pi_m}\binom{N_C-n_C}{N_j-n_j:\ j\in\mathcal{T}(C)}.\end{aligned}\tag{A.29}$$

The conditional probability in (A.29) can be rewritten in this way:

$$\frac{P(\xi_1\in A_1,\dots,\xi_N\in A_N\mid \tilde N(C)=N_C:C\in\pi_m)}{P(\xi_1\in A_1,\dots,\xi_n\in A_n\mid \tilde N(C)=N_C:C\in\pi_m)},$$

and, by (A.28), it becomes

$$\begin{aligned}&\prod_{C\in\pi_m}\frac{P(\xi_i\in A_i:i\le N,\mathrm{ge}(A_i)=C\mid \tilde N(C)=N_C)}{P(\xi_i\in A_i:i\le n,\mathrm{ge}(A_i)=C,\mid \tilde N(C)=N_C)}\\&\quad=\prod_{C\in\pi_m}P(\xi_i\in A_i:\ i\le N,A_i\subset C\mid \tilde N(C)=N_C,\,\xi_i\in A_i:\ i\le n,A_i\subset C).\end{aligned}\tag{A.30}$$

If one substitutes the conditional probability in (A.29) with the right hand side of (A.30), then the right hand side of (A.25) is obtained as desired. □

At this stage, Proposition 4.2 and in part Proposition 4.1 will be proved just for the case $n=1$.

Lemma A.10. *Assume that:*

(i) Condition 4.1.2 holds,

(ii) For each $m\ge 0$, and each B in π_{m+1} such that $\tilde N(\mathrm{ge}(B))$ is not degenerate at zero, $\mathbb{E}(\frac{\tilde N(B)}{\tilde N(\mathrm{ge}(B))}\mid \tilde N(\mathrm{ge}(B))=M)$ does not depend on M (M in $\mathcal{S}(\mathrm{ge}(B))$),

Then:

(iii) Denoting by B_m^x the set of π_m which x belongs to,

$$\begin{aligned}&P(\tilde N(B_{m,1})=N_1,\dots,\tilde N(B_{m,k_m})=N_{k_m}\mid \xi_1=x_1)\\&\qquad\overset{P\text{-}a.s.}{=}P(\tilde N(B_{m,1})=N_1,\dots,\tilde N(B_{m,k_m})=N_{k_m}\mid \xi_1\in B_m^{x_1});\end{aligned}\tag{A.31}$$

(iv) For each $m\ge 0$ and for each $B\in\pi_{m+1}$,

$$\mathcal{L}_{\tilde N(B)\mid \tilde N_m,\,\xi_1}=\mathcal{L}_{\tilde N(B)\mid \tilde N(\mathrm{ge}(B)),\,\xi_1}.\tag{A.32}$$

Proof. First we shall show that (i) and (ii) imply (iii). Let

$$\mathcal{N}_h = \{x \in \mathbb{X} : P(\xi_1 \in B_h^x) = 0\}, \quad \mathcal{N} = \bigcup_{h \in \mathbb{N}} \mathcal{N}_h.$$

Note that $\mathcal{N}$ has P-probability zero.

In order to prove (iii), we shall equivalently show that for any l and any $B \in \pi_l$

$$\int_{\{\xi_1 \in B \cap \mathcal{N}^c\}} \varphi(\xi_1)\, \mathrm{d}P \quad = \quad P(\tilde{N}(B_{m,\,1}) = N_1, \ldots, \tilde{N}(B_{m,\,k_m}) = N_{k_m}, \xi_1 \in B), \tag{A.33}$$

where

$$\varphi(x) := P(\tilde{N}(B_{m,\,1}) = N_1, \ldots, \tilde{N}(B_{m,\,k_m}) = N_{k_m} \mid \xi_1 \in B_m^x),$$

for $x \in \mathcal{N}^c$. Let us suppose that

$$P(\tilde{N}(B_{m,\,1}) = N_1, \ldots, \tilde{N}(B_{m,\,k_m}) = N_{k_m}) > 0$$

(the opposite case is trivial).

We shall first consider the case in which $l > m$. If $l > m$, (A.33) becomes

$$\begin{aligned} &P(\tilde{N}(B_{m,\,1}) = N_1, \ldots, \tilde{N}(B_{m,\,k_m}) = N_{k_m} \mid \xi_1 \in B_{m,\tilde{\jmath}}) \cdot P(\xi_1 \in B) \\ &\qquad = P(\tilde{N}(B_{m,\,1}) = N_1, \ldots, \tilde{N}(B_{m,\,k_m}) = N_{k_m}, \xi_1 \in B), \end{aligned} \tag{A.34}$$

where $\tilde{\jmath}$ denotes the index such that $B_{m,\tilde{\jmath}} = \mathrm{ge}^{(l-m)}(B)$. Equation (A.34) can be re-written in the form:

$$\begin{aligned} &P(\tilde{N}(B_{m,\,1}) = N_1, \ldots, \tilde{N}(B_{m,\,k_m}) = N_{k_m}, \xi_1 \in B_{m,\tilde{\jmath}}) \cdot P(\xi_1 \in B) \\ &\qquad = P(\tilde{N}(B_{m,\,1}) = N_1, \ldots, \tilde{N}(B_{m,\,k_m}) = N_{k_m}, \xi_1 \in B) \cdot P(\xi_1 \in B_{m,\tilde{\jmath}}). \end{aligned} \tag{A.35}$$

In order to verify (A.35), notice that, combining Lemma A.7 with Lemma A.8, we obtain that

$$\begin{aligned} &N_{m,\tilde{\jmath}} \;\cdot\; \mathbb{E}(\tilde{N}(B)) \\ &\qquad = \mathbb{E}(\tilde{N}(B_{m,\tilde{\jmath}})) \;\cdot\; \mathbb{E}(\tilde{N}(B)) \mid \tilde{N}(B_{m,\,1}) = N_1, \ldots, \tilde{N}(B_{m,\,k_m}) = N_{k_m}). \end{aligned} \tag{A.36}$$

If we apply (A.11a) three times in equation (A.36) – for the second factor of the right hand side, considering $P(\cdot \mid \tilde{N}(B_{m,\,1}) = N_1, \ldots, \tilde{N}(B_{m,\,k_m}) = N_{k_m})$ instead of $P(\cdot)$ –, we obtain

$$\frac{N_{m,\tilde{\jmath}}}{N} \cdot P(\xi_1 \in B)$$
$$= P(\xi_1 \in B_{m,\tilde{\jmath}}) \cdot P(\xi_1 \in B \mid \tilde{N}(B_{m,\,1}) = N_1, \ldots, \tilde{N}(B_{m,\,k_m}) = N_{k_m}). \quad \text{(A.37)}$$

that is

$$\frac{N_{m,\tilde{\jmath}}}{N} \cdot P(\xi_1 \in B) \cdot P(\tilde{N}(B_{m,\,1}) = N_1, \ldots, \tilde{N}(B_{m,\,k_m}) = N_{k_m})$$
$$= P(\xi_1 \in B_{m,\tilde{\jmath}}) \cdot P(\xi_1 \in B, \tilde{N}(B_{m,\,1}) = N_1, \ldots, \tilde{N}(B_{m,\,k_m}) = N_{k_m}), \quad \text{(A.38)}$$

which by (A.11b) yields (A.35).

Now suppose that $l \leq m$. Hence the left hand side of (A.33) is equal to

$$\sum_{B_{m,j} \subset B} \int_{\{\xi_1 \in B_{m,j}\}} \varphi(\xi_1) \, \mathrm{d}P$$
$$= \sum_{B_{m,j} \subset B} P(\tilde{N}(B_{m,\,1}) = N_1, \ldots, \tilde{N}(B_{m,\,k_m}) = N_{k_m} \mid \xi_1 \in B_{m,j}) \cdot P(\xi_1 \in B_{m,j})$$
$$= P(\tilde{N}(B_{m,\,1}) = N_1, \ldots, \tilde{N}(B_{m,\,k_m}) = N_{k_m}, \xi_1 \in B),$$

and therefore (A.33) is satisfied in this case, too.

Apply Lemma A.9 - with $n = 1$ - to show that (i)–(iii) imply (iv).

□

Proposition A.11. *Assume that:*

(*i*) *Condition 4.1.2 holds;*

(*ii*) *For any $m \geq 0$, and every B in π_{m+1} such that $\tilde{N}(\mathrm{ge}(B))$ is not degenerate at zero,*

(*a*) $\mathbb{E}(\frac{\tilde{N}(B)}{\tilde{N}(\mathrm{ge}(B))} \mid \tilde{N}(\mathrm{ge}(B)) = M)$ *does not depend on M ($M \in \mathcal{S}(\mathrm{ge}(B))$),*

(*b*) *for any $n < N$ and $A_1, \ldots, A_n$ in π_{m+1},*

$$\mathbb{E}\left(\frac{\sum_{i=n+1}^N \delta_{\xi_i}(B)}{\sum_{i=n+1}^N \delta_{\xi_i}(\mathrm{ge}(B))} \mid \sum_{i=n+1}^N \delta_{\xi_i}(\mathrm{ge}(B)) = M, \xi_1 \in A_1, \ldots, \xi_n \in A_n \right)$$

does not depend on M, as M varies in the following set:

$$\Big\{ j = 1, \ldots, N : P\big(\xi_1 \in A_1, \ldots, \xi_n \in A_n, \textstyle\sum_{i=n+1}^N \delta_{\xi_i}(\mathrm{ge}(B)) = j\big) > 0 \Big\}.$$

Then:

(*iii*) *For any $n \leq N$,*

$$P(\tilde{N}(B_{m,1}) = N_1, \ldots, \tilde{N}(B_{m,k_m}) = N_{k_m} \mid \xi_1 = x_1, \ldots, \xi_n = x_n)$$
$$\overset{P\text{-a.s.}}{=} P(\tilde{N}(B_{m,1}) = N_1, \ldots, \tilde{N}(B_{m,k_m}) = N_{k_m} \mid \xi_1 \in B_m^{x_1}, \ldots, \xi_n \in B_m^{x_n}) \quad (A.39)$$

where B_m^x denotes the set of π_m which x belongs to;

(*iv*) *For any $n \leq N$, any $m \geq 0$ and any $B \in \pi_{m+1}$,*

$$\mathcal{L}_{\tilde{N}(B)|\tilde{N}_m,\, \xi(n)} = \mathcal{L}_{\tilde{N}(B)|\tilde{N}(\mathrm{ge}(B)),\, \xi(n)}.$$

Proof. The proof will be done by induction. The thesis was already proved for the case $n = 1$ in Lemma A.10.

Hence, we just need to prove that the thesis holds when

$$P(\tilde{N}(B_{m,1}) = N_1, \ldots, \tilde{N}(B_{m,k_m}) = N_{k_m} \mid \xi_1 = x_1, \ldots, \xi_{n-1} = x_{n-1})$$
$$\overset{P\text{-a.s.}}{=} P(\tilde{N}(B_{m,1}) = N_1, \ldots, \tilde{N}(B_{m,k_m}) = N_{k_m} \mid \xi_1 \in B_m^{x_1}, \ldots, \xi_{n-1} \in B_m^{x_{n-1}}) \quad (A.40a)$$

and for any $m \geq 0$ and any $B \in \pi_{m+1}$,

$$\mathcal{L}_{\tilde{N}(B)|\tilde{N}_m,\, \xi(n-1)} = \mathcal{L}_{\tilde{N}(B)|\tilde{N}(\mathrm{ge}(B)),\, \xi(n-1)}. \quad (A.40b)$$

Let $(\eta_n, \ldots, \eta_N)$ be a random vector, whose distribution is $\mathcal{L}_{(\xi_n,\, \ldots,\, \xi_N)|(\xi_1,\, \ldots,\, \xi_{n-1})}$. By (ii.b), the conditional expectation

$$\mathbb{E}\left(\frac{\sum_{i=n}^N \delta_{\eta_i}(B)}{\sum_{i=n}^N \delta_{\eta_i}(\mathrm{ge}(B))} \mid \sum_{i=n}^N \delta_{\eta_i}(\mathrm{ge}(B)) = M \right) \quad (A.41)$$

is constant w.r.t. $M \in \mathcal{S}(\mathrm{ge}(B))$, i.e. $P(\tilde{N}(\mathrm{ge}(B)) = M) > 0$. In order to show that, first note that (A.41) is equal to

$$\mathbb{E}\left(\frac{\sum_{i=n}^N \delta_{\xi_i}(B)}{\sum_{i=n}^N \delta_{\xi_i}(\mathrm{ge}(B))} \mid \sum_{i=n}^N \delta_{\xi_i}(\mathrm{ge}(B)) = M,\, \xi(n-1) = x(n-1) \right). \quad (A.42)$$

Observe now that if X, Y, Z are three r.v.'s such that X and Y are discrete, then we can write:

$$\mathbb{E}(X \mid Y = y, Z = z) = \frac{\mathbb{E}(X \mathbb{I}_{\{Y=y\}} \mid Z = z)}{P(Y = y \mid Z = z)}. \quad (A.43)$$

Applying first (A.43) and then (A.40a), (A.42) becomes

$$\frac{\mathbb{E}\left(\frac{1}{M}\sum_{i=n}^{N}\delta_{\xi_i}(B)\mathbb{I}_{\{\sum_{i=n}^{N}\delta_{\xi_i}(\mathrm{ge}(B))=M\}} \mid \xi(n-1)=x(n-1)\right)}{P\left(\sum_{i=n}^{N}\delta_{\xi_i}(\mathrm{ge}(B))=M \mid \xi(n-1)=x(n-1)\right)}$$
$$=\frac{\mathbb{E}\left(\frac{1}{M}\sum_{i=n}^{N}\delta_{\xi_i}(B)\mathbb{I}_{\{\sum_{i=n}^{N}\delta_{\xi_i}(\mathrm{ge}(B))=M\}} \mid \xi(n-1)\in B_m^{x_1}\times\cdots\times B_m^{x_{n-1}}\right)}{P\left(\sum_{i=n}^{N}\delta_{\xi_i}(\mathrm{ge}(B))=M \mid \xi(n-1)\in B_m^{x_1}\times\cdots\times B_m^{x_{n-1}}\right)},$$

which is equal – again by (A.43) – to the expectation in (ii.b) when A_i coincides with $B_m^{x_i}$ for each i.

Moreover, by (A.40b), Condition 4.1.2 still holds when we substitute the random vector $(\xi_1,\dots,\xi_N)$ with $(\eta_n,\dots,\eta_N)$. Hence, we can apply Lemma A.10 to $(\eta_n,\dots,\eta_N)$, obtaining that

$$\begin{aligned}&P(\textstyle\sum_{i=n}^{N}\delta_{\eta_i}(B_{m,1})=M_1,\dots,\sum_{i=n}^{N}\delta_{\eta_i}(B_{m,k_m})=M_{k_m} \mid \eta_n=x_n)\\&\quad=P(\textstyle\sum_{i=n}^{N}\delta_{\eta_i}(B_{m,1})=M_1,\dots,\sum_{i=n}^{N}\delta_{\eta_i}(B_{m,k_m})=M_{k_m} \mid \eta_n\in B_m^{x_n})\end{aligned}\tag{A.44}$$

and

$$\mathcal{L}_{\tilde N(B)\mid\tilde N_m,\,\eta_n}=\mathcal{L}_{\tilde N(B)\mid\tilde N(\mathrm{ge}(B)),\,\eta_n}.\tag{A.45}$$

In fact, consider that $\tilde N=\sum_{i=n}^{N}\delta_{\eta_i}+\sum_{i=1}^{n-1}\delta_{x_i}$ if $\xi(n-1)=x(n-1)$. If one combines (A.40a) and (A.44), (iii) is obtained; while (A.45) yields (iv). □

Now we are in position to conclude:

Proof of Propositions 4.1 and 4.2. Proposition 4.2 is a consequence of Proposition A.11, while Proposition 4.1 is proved combining Proposition 4.2 and Lemma A.9.

□

A.2.2 Some further properties of a partitions tree law

The following proposition says that a partition tree distribution w.r.t. some separating binary tree Π does not depend "too much" on Π.

Proposition A.12. *If P is a partitions tree distribution w.r.t. $\Pi=\{\pi_m\}_m$, then P is a partitions tree distribution w.r.t. any subsequence of* Π.

Proof. Our aim is to prove that if the random vectors $\tilde{N}_m$ and $\tilde{N}_{m+1}$ satisfy, for each m, Condition 4.1, then, for each $h > 1$, the random vectors $\tilde{N}_m$ and $\tilde{N}_{m+h}$ satisfy the analogous conditions, too, that is:

(i) The collections of random variables $\{\tilde{N}(B) : \mathrm{ge}^{(h)}(B) = C\}$, as C varies in π_m, are conditionally independent given $\tilde{N}_m$.

(ii) For each C in π_m, the collections $\{\tilde{N}(B) : \mathrm{ge}^{(h)}(B) = C\}$ and $\{\tilde{N}(B) : B \in \pi_m \setminus \{C\}\}$ are conditionally independent given $\tilde{N}(C)$.

(iii) For each $m \geq 0$ and every B in π_{m+h} such that $\tilde{N}(\mathrm{ge}^{(h)}(B))$ is not degenerate at zero,

   a. $\mathbb{E}(\frac{\tilde{N}(B)}{\tilde{N}(\mathrm{ge}^{(h)}(B))} \mid \tilde{N}(\mathrm{ge}^{(h)}(B)) = M)$ does not depend on M $(M \in \mathcal{S}(\mathrm{ge}(B)))$,

   b. for any $n < N$ and $A_1, \ldots, A_n$ in π_{m+1},

$$\mathbb{E}\left(\frac{\sum_{i=n+1}^{N} \delta_{\xi_i}(B)}{\sum_{i=n+1}^{N} \delta_{\xi_i}(\mathrm{ge}^{(h)}(B))} \mid \sum_{i=n+1}^{N} \delta_{\xi_i}(\mathrm{ge}^{(h)}(B)) = M, \xi_1 \in A_1, \ldots, \xi_n \in A_n\right)$$

   does not depend on M, as M varies in

$$\Big\{j = 1, \ldots, N : P\big(\xi(n) \in A_1 \times \cdots \times A_n, \textstyle\sum_{i=n+1}^{N} \delta_{\xi_i}(\mathrm{ge}^{(h)}(B)) = j\big) > 0\Big\}.$$

Conditions (i) and (ii) hold by Lemma A.6. In order to prove (iii), fix n and take $(\eta_{n+1}, \ldots, \eta_N)$ to be a random sequence with distribution $\mathcal{L}_{\xi_{n+1},\ldots,\xi_N|(\xi_1,\ldots,\xi_n)}$. Since it was proved that partitions tree laws are conjugate (Proposition 4.1), Lemma A.8 for $n = 1, \ldots, N$ can be applied to get the desired result. □

The following proposition shows how Condition 4.1.3 can be reformulated in terms of the distribution of $(\xi_1, \ldots, \xi_N)$.

For each $C \in \mathscr{G}$ and each $1 \leq n \leq N$, denote

$$\mathcal{S}_n(C) := \Big\{M = 0, \ldots, N-n : P(\tilde{N}(C) = M + n) > 0\Big\}.$$

Proposition A.13. *Assume that Conditions 4.1.1–4.1.2 are satisfied.*

The following facts are equivalent:

(i) Condition 4.1.3 holds;

(ii) For any C in $\mathscr{G}$, any N-tuple $(A_1, \ldots, A_N)$ of sets such that $\mathrm{ge}(A_i) = C$, and any n such that $1 \leq n \leq N$, 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 $\mathcal{S}_n(C)$;

(iii) Under the same assumptions of (ii),

$$\begin{aligned} &P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in C, \ldots, \xi_{n+M} \in C, \xi_{n+M+1} \in C^c, \ldots, \xi_N \in C^c) \\ &\quad = P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in C, \ldots, \xi_n \in C) \end{aligned}$$

for any M in $\mathcal{S}_n(C)$.

Proof. Denote $F_0 := \mathbb{X}^N$ and, for each n such that $1 \leq n \leq N$,

$$\begin{aligned} F_n &:= \{\xi_1 \in A_1, \ldots, \xi_n \in A_n\} \\ E_n^{(M)} &:= \{\xi_1 \in C, \ldots, \xi_{n+M} \in C, \xi_{n+M+1} \in C^c, \ldots, \xi_N \in C^c\} \\ D_n^{(M)} &:= F_n \cap E_n^{(M)} \end{aligned}$$

Thus:

$$D_n^{(M)} := \{\xi_1 \in A_1, \ldots, \xi_n \in A_n, \xi_{n+1} \in C, \ldots, \xi_{n+M} \in C, \xi_{n+M+1} \in C^c, \ldots, \xi_N \in C^c\}.$$

We shall show that (i) is tantamount to (ii) and (ii) entails (iii). It is trivial to prove that (iii) implies (ii).

In order to show that (i) and (ii) are equivalent, we shall assume, without loss of generality, that $A_1, \ldots, A_N$ are such that $P(F_n)$ is positive for each $n \geq 1$. We want to prove that Condition 4.1.3 holds if and only if $P(F_n \mid E_n^{(M)})$ does not depend on M, as M varies in the following set:

$$\mathcal{U}_n := \Big\{M = 0, \ldots, N-n : P(D_n^{(M)}) > 0\Big\}.$$

First, we show that, by Lemma A.9, Condition 4.1.3 is satisfied if and only if

$$\mathbb{E}\left(\frac{\sum_{i=n}^N \delta_{\xi_i}(A_n)}{\sum_{i=n}^N \delta_{\xi_i}(\mathrm{ge}(A_n))} \mid \sum_{i=n}^N \delta_{\xi_i}(\mathrm{ge}(A_n)) = M, \xi_1 \in A_1, \ldots, \xi_{n-1} \in A_{n-1}\right) \tag{A.46}$$

does not depend on M ($M \in \mathfrak{U}_n$), just when $A_1, \ldots, A_{n-1}$ are such that, for each $i \leq (n-1)$, A_i is either a descendant or the complementary of $\mathrm{ge}(A_n)$. To prove this, let $A_1, \ldots, A_n$ all belong to the same π_m (for some m) and denote $\tilde{N}_{(n)} = \sum_{i=n}^{N} \delta_{\xi_i}$, and

$$A_{i,j} = \begin{cases} A_i & \text{if } \mathrm{ge}(A_i) = B_{m,j} \\ B^c_{m,j} & \text{otherwise,} \end{cases}$$

as $j = 1, \ldots, k_m$ and $i = 1, \ldots, n-1$. In virtue of exchangeability, by Lemma A.9, for any $m \geq 1$ and any vector $(M_{m+1,1}, \ldots, M_{m+1,k_{m+1}})$ summing up to $(N - n + 1)$, letting

$$M_{m,\,l} := \sum_{j: B_{m+1,\,j} \subset B_{m,\,l}} M_{m+1,\,j} \qquad (l = 1, \ldots, k_m),$$

one can write

$$\begin{aligned} &P(\tilde{N}_{(n)}(B_{m+1,\,j}) = M_{m+1,\,j}, j = 1, \ldots, k_{m+1} \mid \tilde{N}_{(n)}(B_{m,\,j}) = M_{m,\,j}, j = 1, \ldots, k_m,\, \xi_1 \in A_1, \ldots, \xi_n \in A_{n-1}) \\ &= \prod_{j=1}^{k_m} P(\tilde{N}_{(n)}(B_{m+1,\,l}) = M_{m+1,\,l} : \mathrm{ge}(B_{m+1,\,l}) = B_{m,\,j} \mid \tilde{N}_{(n)}(B_{m,\,j}) = M_{m,\,j},\, \xi_1 \in A_{1,j}, \ldots, \xi_n \in A_{n,j}), \end{aligned}$$

which implies:

$$\begin{aligned} &\mathcal{L}_{(\tilde{N}_{(n)}(B_{m+1,\,l}) :\, \mathrm{ge}(B_{m+1,\,l}) = B_{m,\,j}) | (\tilde{N}_{(n)}(B_{m,\,r}), r=1,\ldots,k_m,\, \xi_1 \in A_1, \ldots, \xi_{n-1} \in A_{n-1})} = \\ &\qquad \mathcal{L}_{(\tilde{N}_{(n)}(B_{m+1,\,l}) :\, \mathrm{ge}(B_{m+1,\,l}) = B_{m,\,j}) | (\tilde{N}_{(n)}(B_{m,\,j}),\, \xi_1 \in A_{1,\,j}, \ldots, \xi_{n-1} \in A_{n-1,\,j})}. \end{aligned} \tag{A.47}$$

Define

$$V_{j,i} := \begin{cases} l & \text{if} \quad \xi_i \in B_{m+1,\,l} \subset B_{m,j} \\ 0 & \text{if} \quad \xi_i \in B^c_{m,j}, \end{cases}$$

for $j = 1, \ldots, k_m$, $i = 1, \ldots, N$. Hence, (A.47) becomes

$$\begin{aligned} &\mathcal{L}_{(\tilde{N}_{(n)}(B_{m+1,\,l}) :\, \mathrm{ge}(B_{m+1,\,l}) = B_{m,\,j}) | (\tilde{N}_{(n)}(B_{m,\,r}),\, V_{r,i} : r=1,\ldots,k_m, i=1,\ldots,n-1)} = \\ &\qquad \mathcal{L}_{(\tilde{N}_{(n)}(B_{m+1,\,l}) :\, \mathrm{ge}(B_{m+1,\,l}) = B_{m,\,j}) | (\tilde{N}_{(n)}(B_{m,\,j}),\, V_{j,i} : i=1,\ldots,n-1)}. \end{aligned} \tag{A.48}$$

Therefore, we can say that, for each $1 \leq j \leq k_m$ and each A such that $\mathrm{ge}(A) = B_{m,j}$, $\tilde{N}_{(n)}(A)$ and $\{V_{r,i} : r \neq j, i \leq n-1\}$ are conditionally independent given $(\tilde{N}_{(n)}(B_{m,j}), V_{j,i} : i = 1, \ldots, n-1)$.

For this reason, let us assume that, for each $i \leq n-1$, A_i is either a descendant or the complementary of $C = \mathrm{ge}(A_n)$. In virtue of exchangeability, for each $M \in \mathcal{U}_n$, the conditional expectation (A.46) is equal to

$$\begin{aligned}
&\mathbb{E}(\textstyle\sum_{i=n}^{N}\delta_{\xi_i}(A_n)/M \mid \sum_{i=n}^{N}\delta_{\xi_i}(C) = M, \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1}) \\
&\quad = \frac{\mathbb{E}(\sum_{i=n}^{N}\delta_{\xi_i}(A_n)\mathbb{I}_{\{\sum_{j=n}^{N}\delta_{\xi_j}(C)=M\}} \mid \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1})}{M\, P(\sum_{i=n}^{N}\delta_{\xi_i}(C) = M) \mid \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1})} \\
&\quad = \frac{\binom{N-n}{M}\,\mathbb{E}(\sum_{i=n}^{n+M}\delta_{\xi_i}(A_n)\mathbb{I}_{D_{n-1}^{(M)}} \mid \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1})}{\binom{N-n}{M}\, M\, P(D_{n-1}^{(M)} \mid \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1})} \\
&\quad = \frac{P(\xi_i \in A_n, D_{n-1}^{(M)} \mid \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1})}{P(D_{n-1}^{(M)} \mid \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1})} \\
&\quad = P(\xi_n \in A_n \mid D_{n-1}^{(M)}).
\end{aligned} \tag{A.49}$$

Denote by h the number of A_i, with $i \leq n-1$, that are descendants of C. If $h = 0$, by exchangeability, the last term (A.49) is equal to $P(\xi_1 \in A_n \mid D_0^{(M+1)})$. If $h > 0$, let $(l_1, \dots, l_h)$ be such that $l_1 < \cdots < l_h$ and $A_{l_i} \subset C$ for each $i \leq n-1$. Then, by exchangeability, $P(\xi_n \in A_n \mid D_{n-1}^{(M)})$ is equal to

$$P(\xi_n \in A_n \mid \xi_1 \in A_{l_1}, \dots, \xi_h \in A_{l_h}, \xi_{h+1} \in C, \dots, \xi_{M+h} \in C, \xi_{M+h+1} \in C^c, \dots, \xi_N \in C^c).$$

Therefore Condition 4.1.3 is tantamount to require that, whenever $A_1, \dots, A_N$ are descendants of C, for each $n \geq 1$ (fixed), $P(\xi_n \in A_n \mid D_{n-1}^{(M)})$ does not depend on M, as M varies in $\mathcal{S}_n(C)$. At this point, one can see that Condition 4.1.3 is satisfied if (ii) holds, i.e. $P(F_n \mid E_n^{(M)})$ is constant w.r.t. $M \in \mathcal{S}_n(C)$. In fact, for any $n > 1$ and any M in $\mathcal{S}_{n-1}(C)$, $E_{n-1}^{(M)} = E_n^{(M-1)}$, and therefore

$$P(\xi_n \in A_n \mid D_{n-1}^{(M)}) = \frac{P(F_n \cap E_{n-1}^{(M)})}{P(F_{n-1} \cap E_{n-1}^{(M)})} = \frac{P(F_n \mid E_{n-1}^{(M)})}{P(F_{n-1} \mid E_{n-1}^{(M)})} = \frac{P(F_n \mid E_n^{(M-1)})}{P(F_{n-1} \mid E_{n-1}^{(M)})}.$$

If $n = 1$, then, for any $M \in \mathcal{S}_0(C)$, $P(\xi_1 \in A_1 \mid D_0^{(M)}) = P(F_1 \mid E_0^{(M)}) = P(F_1 \mid E_1^{(M-1)})$.

On the other hand, if $P(\xi_n \in A_n \mid D_{n-1}^{(M)})$ is a constant function of $M \in \mathcal{S}_n(C)$ (for each $n \geq 1$), so is $P(F_n \mid E_n^{(M)})$ too, since

$$P(F_n \mid E_n^{(M)}) = P(\xi_1 \in A_1 \mid D_0^{(M+n)})P(\xi_2 \in A_2 \mid D_1^{(M+n-1)}) \cdots P(\xi_n \in A_n \mid D_{n-1}^{(M)}).$$

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:

$$
\begin{aligned}
&P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in C, \ldots, \xi_n \in C) \\
&\quad = \sum_{t_1,\ldots,\, t_{N-n}} P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in C, \ldots, \xi_n \in C, \xi_{n+1} \in C^{t_1}, \ldots, \xi_N \in C^{t_{N-n}}) \\
&\qquad\quad \cdot P(\xi_{n+1} \in C^{t_1}, \ldots, \xi_N \in C^{t_{N-n}} \mid \xi_1 \in C, \ldots, \xi_n \in C) \\
&\quad = \sum_{M \in \mathcal{S}_n(C)} \sum_{\sum_{i=1}^{N-n} t_i = M} P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid E_n^{(M)}) \\
&\qquad\quad \cdot P(\xi_{n+1} \in C^{t_1}, \ldots, \xi_N \in C^{t_{N-n}} \mid \xi_1 \in C, \ldots, \xi_n \in C),
\end{aligned}
\tag{A.50}
$$

where the sum runs over all vectors $(t_1, \ldots, t_{N-n})$ in $\{0,1\}^{N-n}$. By (ii), (A.50) becomes:

$$
\begin{aligned}
&P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid \xi_1 \in C, \ldots, \xi_n \in C) \\
&\quad = P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid E_n^{(M)}) \sum_{t_1,\ldots,\, t_{N-n}} P(\xi_{(n,N)} \in C^{t_1} \times \cdots \times C^{t_{N-n}} \mid \xi_1 \in C, \ldots, \xi_n \in C) \\
&\quad = P(\xi_1 \in A_1, \ldots, \xi_n \in A_n \mid E_n^{(M)}),
\end{aligned}
$$

and the proof is done. □

Proposition A.14. *Assume that Conditions 4.1.1—4.1.3 hold. If $A_1, \ldots, A_n$ belong to π_m for some m and $P(\xi_1 \in A_1, \ldots, \xi_{n-1} \in A_{n-1}, \xi_n \in \mathrm{ge}(A_n)) > 0$, then*

$$
\begin{aligned}
&P(\xi_n \in A_n \mid \xi_1 \in A_1, \ldots, \xi_{n-1} \in A_{n-1}, \xi_n \in \mathrm{ge}(A_n)) \\
&\quad = P(\xi_n \in A_n \mid \xi_n \in \mathrm{ge}(A_n), \xi_i \in A_i :\ \mathrm{ge}(A_i) = \mathrm{ge}(A_n),\ i = 1, \ldots, n-1)
\end{aligned}
\tag{A.51}
$$

for each $2 \le n \le N$.

Proof. Let $\mathrm{ge}(A_n) = C$ and notice that any random sequence $(\eta_1, \ldots, \eta_{N+1-n})$ such that

$$
\mathcal{L}_{(\eta_1,\ldots,\eta_{N+1-n})}(\,\cdot\,) = P((\xi_n, \ldots, \xi_N) \in \cdot\ \mid \xi_1 \in A_1, \ldots, \xi_{n-1} \in A_{n-1})
$$

trivially satisfies Condition 4.1.3. Hence, by Proposition A.4, one can write

$$
\begin{aligned}
&P(\xi_n \in A_n \mid \xi_1 \in A_1, \ldots, \xi_{n-1} \in A_{n-1}) \\
&\quad = P(\xi_n \in C \mid \xi_1 \in A_1, \ldots, \xi_{n-1} \in A_{n-1}) \\
&\qquad \cdot \tfrac{1}{M}\, \mathbb{E}(\textstyle\sum_{i=n}^N \delta_{\xi_i}(A_n) \mid \sum_{i=n}^N \delta_{\xi_i}(C) = M,\ \xi_1 \in A_1, \ldots, \xi_{n-1} \in A_{n-1})
\end{aligned}
$$

for each $M \in \mathcal{S}_{n-1}(C)$.

Therefore,

$$\begin{aligned}
&P(\xi_n \in A_n \mid \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1}, \xi_n \in C) \\
&\qquad = \frac{P(\xi_n \in A_n \mid \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1})}{P(\xi_n \in C \mid \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1})} \\
&\qquad = \frac{1}{M}\, \mathbb{E}(\textstyle\sum_{i=n}^N \delta_{\xi_i}(A_n) \mid \sum_{i=n}^N \delta_{\xi_i}(C) = M,\, \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1}),
\end{aligned} \tag{A.52}$$

for M in $\mathcal{S}_{n-1}(C)$.

The last term in (A.52), by (A.49), is equal to

$$P(\xi_n \in A_n \mid \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1}, \xi_n \in C, \dots, \xi_{n+M-1} \in C, \xi_{n+M} \in C^c \dots, \xi_N \in C^c).$$

Hence, for each M in $\mathcal{S}_{n-1}(C)$,

$$\begin{aligned}
&P(\xi_n \in A_n \mid \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1}, \xi_n \in C, \dots, \xi_{n+M-1} \in C, \xi_{n+M} \in C^c, \dots, \xi_N \in C^c) \\
&\qquad = P(\xi_n \in A_n \mid \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1}, \xi_n \in C).
\end{aligned} \tag{A.53}$$

By (A.52), we need to prove that

$$\begin{aligned}
&\frac{1}{M}\, \mathbb{E}(\textstyle\sum_{i=n}^N \delta_{\xi_i}(A_n) \mid \sum_{i=n}^N \delta_{\xi_i}(C) = M,\, \xi_1 \in A_1, \dots, \xi_{n-1} \in A_{n-1}) \\
&\qquad = P(\xi_n \in A_n \mid \xi_n \in C, \xi_i \in A_i, i = 1, \dots, n-1 : \mathrm{ge}(A_i) = C),
\end{aligned} \tag{A.54}$$

for each M in $\mathcal{S}_{n-1}(C)$.

Denote

$$A_{(i)} = \begin{cases} A_i & \text{if } A_i \subset C \\ C^c & \text{otherwise,} \end{cases} \qquad i = 1, \dots, n$$

$$A' := \{\xi_1 \in A_{(1)}, \dots, \xi_1 \in A_{(n-1)}\}.$$

By (A.48), the left-hand side of (A.54) is the same as

$$\frac{1}{M}\, \mathbb{E}(\textstyle\sum_{i=n}^N \delta_{\xi_i}(A_n) \mid \sum_{i=n}^N \delta_{\xi_i}(C) = M, A'). \tag{A.55}$$

Arguing as in (A.49), one realizes that (A.55) is equal to

$$P(\xi_n \in A_n \mid A', \xi_n \in C, \dots, \xi_{n+M-1} \in C, \xi_{n+M} \in C^c, \dots, \xi_N \in C^c),$$

which, by exchangeability, becomes:

$$P(\xi_{h+1} \in A_n \mid \xi_1 \in A_{l_1}, \dots, \xi_h \in A_{l_h}, \xi_{h+1} \in C, \dots, \xi_{h+M} \in C, \xi_{h+M+1} \in C^c, \dots, \xi_N \in C^c), \tag{A.56}$$

where $h = |\{i \le n-1 : \mathrm{ge}(A_i) = C\}|, l_1 < \cdots < l_h$ and $\mathrm{ge}(A_{l_i}) = C$.

Applying (A.53) (with $h+1$ in place of n and $(A_{l_1}, \dots, A_{l_h}, A_n)$ in place of $(A_1, \dots, A_n)$), (A.56) is the same as

$$P(\xi_{h+1} \in A_n \mid \xi_{h+1} \in C, \xi_1 \in A_{l_1}, \dots, \xi_h \in A_{l_h})$$

that, by exchangeability, is equal to the right-hand side of (A.54), as desired. □

A.3 Random partitions

Lemma A.15. *For $i = 1, \dots, N$, let α_i^* be an exchangeable probability measure on $(Z^i, \mathcal{Z}^i)$ and ρ be a probability measure on $\mathcal{P}_N$ for which (5.7) holds true. If $(\zeta_1, \dots, \zeta_N)$ is a random vector such that*

$$P\{\zeta_1 \in C_1, \dots, \zeta_N \in C_N\} = \sum_{\pi \in \mathcal{P}_N} \rho(\pi) \prod_{i=1}^{|\pi|} \alpha^*_{|\pi_i|}(\times_{j \in \pi_i} C_j)$$

is satisfied for every $C_1, \dots, C_N$ in $\mathcal{Z}$, then it is exchangeable.

Proof. For every permutation σ in S_N and for every subset $I = \{i_1, \dots, i_k\}$ of $\{1, \dots, N\}$ define $g_\sigma(I) = \{\sigma(i_1), \dots, \sigma(i_k)\}$, and for every partition π in $\mathcal{P}_N$ let $f_\sigma(\pi)$ be the partition whose blocks are given by

$$g_\sigma(\pi_i) \qquad i = 1, \dots, |\pi|.$$

Observe that if $F : \mathcal{P}_N \times S_N \to \mathbb{R}$ is a function such that $F(\pi, \sigma) = F(f_\sigma(\pi), id)$ for every π in $\mathcal{P}_N$, then

$$\sum_{\pi \in \mathcal{P}_N} F(\pi, \sigma) = \sum_{\pi \in \mathcal{P}_N} F(\pi, id). \tag{A.57}$$

Given $(C_1, \dots, C_N)$ in $\mathscr{X}^N$, for every σ in S_N and every π in $\mathcal{P}_N$ set

$$F(\pi, \sigma) = P(\zeta_1 \in C_{\sigma(1)}, \dots, \zeta_N \in C_{\sigma(N)}, \Pi = \pi) = \rho(\pi) \prod_{i=1}^{|\pi|} \alpha^*_{|\pi_i|}(\times_{j \in \pi_i} C_{\sigma(j)}).$$

Now, note that $\rho(\pi) = \rho(f_\sigma(\pi)))$ and, moreover, that

$$\prod_{i=1}^{|\pi|} \alpha^*_{|\pi_i|}(\times_{j\in\pi_1} C_{\sigma(j)}) = \prod_{i=1}^{|f_\sigma(\pi)|} \alpha^*_{|f_\sigma(\pi)_i|}(\times_{j\in f_\sigma(\pi)_i} C_j).$$

Hence, by (A.57) exchangeability follows. □

Proof of Proposition 5.1. For every $k = 1, \dots, N$ let α^*_k be defined by

$$\alpha^*_k(A_1 \times B_1 \times \dots A_k \times B_k) = \int_{\cap_{j=1}^k A_j} \prod_{i=1}^k \kappa_x(\times_{j=1}^k B_j)\ \alpha_k(\mathrm{d}x) \tag{A.58}$$

for every $A_1, \dots, A_k$ in $\mathscr{X}$ and every $B_1, \dots, B_k$ in $\mathcal{Y}$. Since κ_x is an exchangeable kernel, it follows that α^* is exchangeable for every k. Hence Lemma A.15 yields the thesis. □

Proof of Proposition 5.2. Expressions (5.9) and (5.10) follow from (5.6), (5.4), and the exchangeability of the (ξ, η)s, while (5.11) can be deduced from (5.9). As to (5.12), note first that

$$\mathbb{E}[f(\xi_1)]\ \mathbb{E}[g(\xi_1)] = \sum_{l,k=1}^N w_l\, w_k\, M_{1,l}\, M_{1,k}. \tag{A.59}$$

Moreover, it is easy to check that

$$\mathbb{E}(f(\xi_1)g(\xi_2)) = \sum_{i=1}^N \sum_{\pi\in\mathcal{P}_N:1,2\in\pi_i} \rho(\pi)\, M_{1,|\pi_i|}(fg) + \sum_{i\neq j} \sum_{\pi\in\mathcal{P}_N:1\in\pi_i\ 2\in\pi_j} \rho(\pi)\, M_{1,|\pi_i|}(f)\, M_{1,|\pi_j|}(g).$$

At this stage, rewrite w_i as $t_i + \sum_{j=1}^N r_{ij}$ to get (5.12). □

In order to prove Propositions 5.3 and 5.4, we need two simple preliminary results.

Lemma A.16. *The conditional law of $\tilde\pi$ given $(\xi(n), \eta(n))$ turns out to be the same of the conditional law of $\tilde\pi$ given $\xi(n)$ and coincides with* (5.15).

Proof. Let $A := A_1 \times \dots \times A_n$ and $B := B_1 \times \dots \times B_n$, with A_i in $\mathscr{X}$ and B_i in $\mathcal{Y}$ for every

$i = 1, \ldots, n$. Given π in $\mathcal{P}_N$, let $\check{\pi} := \pi|_n$ and $\nu := |\check{\pi}|$. Observe that

$$\begin{aligned}
&\mathbb{E}[\tau_{\xi(n)}(\pi)\ \mathbb{I}_{\{\xi(n)\in A,\,\eta(n)\in B\}}] \\
&\quad = \rho(\pi) \sum_{\Pi\in\mathcal{P}_N} \rho(\Pi)\ \mathbb{E}\Big[\frac{\mathbb{I}_{\{\check{\pi}=\Pi|_n\}}\ \prod_{i=1}^{\nu} a_{|\check{\pi}_i|}(\xi_i^*)\ \mathbb{I}_{\{\xi_i^*\in\cap_{j\in\check{\pi}_i}A_j\}}\ \kappa_{\xi_i^*}\{\times_{j\in\check{\pi}_i}B_j\}}{\sum_{q\in\mathcal{P}_N}\rho(q)\ \mathbb{I}_{\{q|_n=\Pi|_n\}}\prod_{i=1}^{|q|_n|} a_{|q_i|}(\xi_i^*)}\ \Big|\ \Pi(\xi(N)) = \Pi\Big] \\
&\quad = \rho(\pi) \sum_{\Pi\in\mathcal{P}_N} \rho(\Pi)\ \mathbb{E}\Big[\frac{\mathbb{I}_{\{\check{\pi}=\Pi|_n\}}\ \prod_{i=1}^{\nu} a_{|\check{\pi}_i|}(\xi_i^*)\ \mathbb{I}_{\{\xi_i^*\in\cap_{j\in\check{\pi}_i}A_j\}}\ \kappa_{\xi_i^*}\{\times_{j\in\check{\pi}_i}B_j\}}{\sum_{q\in\mathcal{P}_N}\rho(q)\ \mathbb{I}_{\{\check{\pi}=q|_n\}}\ \prod_{i=1}^{\nu} a_{|q_i|}(\xi_i^*)}\ \Big|\ \Pi(\xi(N)) = \Pi\Big] \\
&\quad = \rho(\pi)\int_{X^\nu}\prod_{i=1}^{\nu} a_{|\check{\pi}_i|}(x_i)\ \mathbb{I}_{\{x_i\in\cap_{j\in\check{\pi}_i}A_j\}}\ \kappa_{\xi_i^*}\{\times_{j\in\check{\pi}_i}B_j\} \\
&\qquad\quad \cdot \frac{\sum_{\Pi\in\mathcal{P}_N}\rho(\Pi)\ \mathbb{I}_{\{\check{\pi}=\Pi|_n\}}\prod_{i=1}^{\nu} a_{\Pi_i}(x_i)}{\sum_{q\in\mathcal{P}_N}\rho(q)\ \mathbb{I}_{\{q|_n=\check{\pi}\}}\prod_{i=1}^{\nu} a_{|q_i|}(x_i)}\ \mu(\mathrm{d}x_1)\ldots\mu(\mathrm{d}x_\nu) \\
&\quad = \rho(\pi)\ \mathbb{E}[\mathbb{I}_{\{\xi(n)\in A,\,\eta(n)\in B\}}\mid \Pi(\xi(N)) = \pi] \\
&\quad = P\{\tilde{\pi} = \pi,\ \xi(n)\in A,\ \eta(n)\in B\},
\end{aligned}$$

which yields the thesis. □

Lemma A.17. *The conditional law of* $\xi(n,N)$ *given* $(\xi(n),\,\eta(n),\,\tilde{\pi})$ *turns out to be equal to the conditional law of* $\xi(n,N)$ *given* $(\xi(n),\,\tilde{\pi})$ *and it is given by* (5.14).

Proof. Take any A in $\mathscr{X}^{N-n}$, B in $\mathcal{Y}^n$, C in $\mathscr{X}^n$, and π in $\mathcal{P}_N$, and observe that

$$\begin{aligned}
&P(\{\xi(n,N)\in A\}\mid \eta(n)\in B, \xi(n)\in C, \tilde{\pi}=\pi) \\
&\qquad = \frac{P\{\eta(n)\in B\mid \xi(n,N)\in A, \xi(n)\in C, \tilde{\pi}=\pi\}\ P\{\xi(n,N)\in A\mid \xi(n)\in C, \tilde{\pi}=\pi\}}{P\{\eta(n)\in B\mid \xi(n)\in C, \tilde{\pi}=\pi\}}
\end{aligned}$$

whenever $P(\xi(n,N)\in A) > 0$ and $P(\eta(n)\in B,\ \xi(n)\in C,\ \tilde{\pi}=\pi) > 0$. Now, it is immediate to check that $P\{\eta(n)\in B\mid \xi(n)\in C,\ \xi(n,N)\in A,\ \tilde{\pi}=\pi\} = P\{\eta(n)\in B\mid \xi(n)\in C,\ \tilde{\pi}=\pi\}$, and then the thesis follows. □

Proof of Propositions 5.3 and 5.4. First of all, note that

$$\begin{aligned}
&\mathcal{L}_{\xi(n,N),\,\eta(n,N)|\xi(n),\eta(n),\tilde{\pi}}(\mathrm{d}x_{n+1}\ldots\mathrm{d}x_N\,\mathrm{d}y_{n+1}\ldots\mathrm{d}y_N) \\
&\quad = g_{n,N}(\mathrm{d}x_{n+1}\ldots\mathrm{d}x_N\mid\xi(n),\,\tilde{\pi})\ h_{n,N}(\mathrm{d}y_{n+1}\ldots\mathrm{d}y_N\mid(\xi_1,\ldots,\xi_n,\,x_{n+1},\ldots,x_N),\ \eta(n),\,\tilde{\pi}).
\end{aligned}$$

Hence, by means of Lemmas A.16 and A.17, one gets

$$\begin{aligned}
&\mathcal{L}_{\xi(n,N),\eta(n,N)|\xi(n),\eta(n)}(\mathrm{d}x_{n+1}\dots\mathrm{d}x_N\,\mathrm{d}y_{n+1}\dots\mathrm{d}y_N)\\
&=\sum_{\pi}\mathcal{L}_{\tilde{\pi}|\xi(n),\eta(n)}(\pi)\;\mathcal{L}_{\xi(n,N),\eta(n,N)|\xi(n),\eta(n),\tilde{\pi}}(\mathrm{d}x_{n+1}\dots\mathrm{d}x_N\mathrm{d}y_{n+1}\dots\mathrm{d}y_N)\\
&=\sum_{\pi}\tau_{\xi(n)}(\pi)\;g_n(\mathrm{d}x_{n+1}\dots\mathrm{d}x_N\mid\xi(n),\,\pi)\cdot\\
&\qquad\qquad\cdot\,h_{n,N}(dy_{n+1}\dots dy_N\mid(\xi_1,\dots,\xi_n,\,x_{n+1},\dots,x_N),\;\eta(n),\,\pi).
\end{aligned}$$

At this stage both propositions follow easily. $\square$

Bibliography

Aldous, D. J. (1985), "Exchangeability and related topics," in *École d'été de probabilités de Saint-Flour, XIII—1983*, Berlin: Springer, vol. 1117 of *Lecture Notes in Math.*, pp. 1–198.

Berliner, L. M. and Hill, B. M. (1988), "Bayesian nonparametric survival analysis," *J. Amer. Statist. Assoc.*, 83, 772–779.

Billingsley, P. (1995), *Probability and measure*, Wiley Series in Probability and Mathematical Statistics, New York: John Wiley & Sons Inc., 3rd ed., a Wiley-Interscience Publication.

Blackwell, D. and MacQueen, J. B. (1973), "Ferguson distributions via Pólya urn schemes," *Ann. Statist.*, 1, 353–355.

Charalambides, C. A. (2002), *Enumerative combinatorics*, CRC Press Series on Discrete Mathematics and its Applications, Chapman & Hall/CRC, Boca Raton, FL.

de Finetti, B. (1930), "Funzione caratteristica di un fenomeno aleatorio," in *Memorie della Reale Accademia dei Lincei*, vol. V of *IV*, pp. 86–133.

— (1937a), "La prévision: ses lois logiques, ses sources subjectives," *Ann. Instit. H. Poincaré*, 7.

— (1937b), "A proposito di 'correlazione'," *Supplemento Statistico ai Nuovi problemi di Politica, Storia ed Economia*, 3.

— (1972), *Probability, induction and statistics. The art of guessing*, John Wiley & Sons, London-New York-Sydney, Wiley Series in Probability and Mathematical Statistics.

— (1974), "Bayesianism: its unifying role for both the foundations and the applications of statistics," *Int. Stat. Rev.*, 42, 117–130.

— (1975), *Theory of Probability*, London: John Wiley & Sons.

Dellacherie, C. and Meyer, P.-A. (1980), *Probabilités et potentiel. Tomes I et II*, Paris: Hermann.

Diaconis, P. (1988), "Recent progress on de Finetti's notions of exchangeability," in *Bayesian statistics, 3 (Valencia, 1987)*, New York: Oxford Univ. Press, Oxford Sci. Publ., pp. 111–125.

Diaconis, P. and Freedman, D. (1980), "Finite exchangeable sequences," *Ann. Probab.*, 8, 745–764.

Doksum, K. (1974), "Tailfree and neutral random probabilities and their posterior distributions," *Ann. Probability*, 2.

Drăghici, L. and Ramamoorthi, R. V. (2000), "A note on the absolute continuity and singularity of Polya tree priors and posteriors," *Scand. J. Statist.*, 27, 299–303.

Dubins, L. and Freedman, D. (1967), "Random distribution functions," in *Proceedings, Fifth Berkeley Simp. on Math. Statistics and Probability*, eds. LeCam, L. and Neyman, J., Berkeley / Los Angeles: Univ. of California Press, pp. 183–214.

Durrett, R. (1996), *Probability: theory and examples*, Belmont, CA: Duxbury Press, 2nd ed.

Ericson, W. A. (1969), "Subjective Bayesian Models in Sampling Finite Populations," *Journal of Royal Statistical Society*, Ser. B, 31, 195–233.

Ferguson, T. S. (1973), "A Bayesian analysis of some nonparametric problems," *Ann. Statist.*, 1, 209–230.

— (1974), "Prior distributions on spaces of probability measures," *Ann. Statist.*, 2, 615–629.

Fortini, S., Ladelli, L., and Regazzini, E. (2000), "Exchangeability, predictive distributions and parametric models," *Sankhyā Ser. A*, 62, 86–109.

Freedman, D. A. (1963), "On the asymptotic behavior of Bayes' estimates in the discrete case," *Ann. Math. Statist.*, 34, 1386–1403.

Ghosh, J. K. and Ramamoorthi, R. V. (2003), *Bayesian nonparametrics*, Springer Series in Statistics, New York: Springer-Verlag.

Ghosh, M. and Meeden, G. (1997), *Bayesian Methods for Finite Population Sampling*, Monographs on Statistics and Applied Probability, London, UK: Chapman & Hall.

Gnedin, A. and Pitman, J. (2005), "Exchangeable Gibbs partitions and Stirling triangles," *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 325, 83–102, 244–245.

Hansen, B. and Pitman, J. (2000), "Prediction rules for exchangeable sequences related to species sampling," *Statist. Probab. Lett.*, 46, 251–256.

Hanson, T. and Johnson, W. O. (2002), "Modeling regression error with a mixture of Polya trees," *J. Amer. Statist. Assoc.*, 97, 1020–1033.

Hill, B. M. (1968), "Posterior distribution of percentiles: Bayes' theorem for sampling from a population," *J. Amer. Statist. Assoc.*, 63, 677–691.

— (1979), "Posterior moments of the number of species in a finite population and the posterior probability of finding a new species," *J. Amer. Statist. Assoc.*, 74, 668–673.

— (1980), "Invariance and robustness of the posterior distribution of characteristics of a finite population, with reference to contingency tables and the sampling of species," in *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*, ed. Zellner, A., Amsterdam: North-Holland, pp. 383–395.

— (1993), "Parametric models for A_n: splitting processes and mixtures," *J. Roy. Statist. Soc. Ser. B.*, 55, 423–433.

Hsuan, F. (1979), "A stepwise Bayes procedure," *Ann. Statist.*, 7, 860–868.

Hume, D. (1748), *An enquiry concerning human understanding*, London.

James, L. F. (2005), "Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages," *Ann. Statist.*, 33, 1771–1799.

Johnson, N. L., Kemp, A. W., and Kotz, S. (2005), *Univariate discrete distributions*, Wiley Series in Probability and Statistics, Hoboken, NJ: Wiley-Interscience [John Wiley & Sons], 3rd ed.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997), *Discrete multivariate distributions*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, New York: John Wiley & Sons Inc., a Wiley-Interscience Publication.

Kemp, C. D. and Kemp, A. W. (1956), "Generalized hypergeometric distributions," *J. Roy. Statist. Soc. Ser. B.*, 18, 202–211.

Kerns and Szekely (2005), "de Finetti theorem for finite exhangeable sequences," To appear in J.Theoret. Probab.

Kingman, J. (1978a), "Random partitions in population genetics," *Proc. Roy. Soc. A.*, 361.

Kingman, J. F. C. (1967), "Completely random measures," *Pacific J. Math.*, 21, 59–78.

— (1978b), "The representation of partition structures," *J. London Math. Soc. (2)*, 18, 374–380.

— (1978c), "Uses of exchangeability," *Ann. Probability*, 6, 183–197.

Lane, D. A. and Sudderth, W. D. (1978), "Diffuse Models for Sampling and Predictive Inference," *Ann. Statist.*, 6, 1318–1336.

Lavine, M. (1992), "Some aspects of Pólya tree distributions for statistical modelling," *Ann. Statist.*, 20, 1222–1235.

— (1994), "More aspects of Pólya tree distributions for statistical modelling," *Ann. Statist.*, 22, 1161–1176.

Lo, A. Y. (1991), "A characterisation of the Dirichlet process," *Statist. Probab. Lett.*, 12, 185–187.

Mauldin, R. D., Sudderth, W. D., and Williams, S. C. (1992), "Pólya trees and random distributions," *Ann. Statist.*, 20, 1203–1221.

Mauldin, R. D. and Williams, S. C. (1990), "Reinforced random walks and random distributions," *Proc. Amer. Math. Soc.*, 110, 251–258.

Meeden, G. and Ghosh, M. (1983), "Choosing Between Experiments: application to Finite Population Sampling," *Ann. Statist.*, 11, 296–305.

Meeden, G. and Vardeman, S. (1991), "A Noninformative Bayesian Approach to Interval Estimation in Finite Population Sampling," *J. Amer. Statist. Assoc.*, 86, 972–098.

Muliere, P. and Walker, S. (1997), "A Bayesian non-parametric approach to survival analysis using Polya trees," *Scand. J. Statist.*, 24, 331–340.

Neath, A. A. (2003), "Polya tree distributions for statistical modeling of censored data," *J. Appl. Math. Decis. Sci.*, 7, 175–186.

Nieto-Barajas, L. E., Prünster, I., and Walker, S. G. (2004), "Normalized random measures driven by increasing additive processes," *Ann. Statist.*, 32, 2343–2360.

Paddock, S. M. (2002), "Bayesian nonparametric multiple imputation of partially observed data with ignorable nonresponse," *Biometrika*, 89, 529–538.

Parthasarathy, K. R. (1967), *Probability measures on metric spaces*, Probability and Mathematical Statistics, No. 3, New York: Academic Press Inc.

Pitman, J. (1995), "Exchangeable and partially exchangeable random partitions," *Proc. Roy. Soc. A.*, 102.

— (1996), "Some developments of the Blackwell-MacQueen urn scheme," in *Statistics, probability and game theory*, Hayward, CA: Inst. Math. Statist., vol. 30 of *IMS Lecture Notes Monogr. Ser.*, pp. 245–267.

— (2003), "Poisson-Kingman partitions," in *Statistics and science: a Festschrift for Terry Speed*, Beachwood, OH: Inst. Math. Statist., vol. 40 of *IMS Lecture Notes Monogr. Ser.*, pp. 1–34.

— (2006), *Combinatorial Stochastic Processes*, vol. 1875/2006 of *Lecture Notes in Math.*, Berlin/Heidelberg: Springer.

Prünster, I. (2002), "Random probability measures derived from increasing additive processes and their applications to Bayesian statistics," Ph.D. thesis, Università degli Studi di Pavia, Pavia (Italy).

Regazzini, E. (1978), "Intorno ad alcune questioni relative alla definizione del premio secondo la teoria della credibilità," *Giorn. Ist. Ital. Attuari*, 41, 77–89.

— (1991), "Coherence, exchangeability and statistical models," in *Atti del convegno – Sviluppi metodologici nei diversi approcci all'inferenza statistica*, Società Italiana di Statistica, Cagliari: Pitagora Editrice Bologna.

— (2004), "Bayesian statistics - Notes by Eugenio Regazzini for a a SMI Course held in Cortona (7/26 – 8/14/2004)," .

Regazzini, E., Lijoi, A., and Prünster, I. (2003), "Distributional results for means of normalized random measures with independent increments," *Ann. Statist.*, 31, 560–585, dedicated to the memory of Viatcheslav V. Sazonov.

Regazzini, E. and Petris, G. (1992), "Some critical aspects of the use of exchangeability in statistics," *J. It. Statist. Soc.*, 1, 103–130.

Roberts, H. V. (1965), "Probabilistic prediction," *J. Amer. Statist. Assoc.*, 60, 50–62.

Roussanov, B. V. (1999), "Stepwise Bayes distributional inference for a finite population," *Statist. Neerlandica*, 53, 361–374.

Sangalli, L. (2006), "Some developments of normalized random measures with independent increments," To appear in *Sankhyā*.

Sarno, E. (1998), "Dependence structures of Polya tree autoregressive models," *Statistica (Bologna)*, 58, 363–373 (1999).

Sazonov, V. V. (1965), "On perfect measures," *Izv. Akad. Nauk SSSR Ser. Mat. (Russian), Moscow, 1962, pp. 391–414 ; Amer. Math. Soc. Transl. (2)*, 48, 229–254.

Schervish, M. J. (1995), *Theory of statistics*, Springer Series in Statistics, New York: Springer-Verlag.

Sethuraman, J. (1994), "A constructive definition of Dirichlet priors," *Statist. Sinica*, 4.

Spizzichino, F. (1982), "Extendibility of symmetric probability distributions and related bounds," in *Exchangeability in probability and statistics (Rome, 1981)*, Amsterdam: North-Holland, pp. 313–320.

Walker, S. and Muliere, P. (1997), "A characterisation of Polya tree distributions," *Statist. Probab. Lett.*, 31, 163–168.

Wood, G. R. (1992), "Binomial mixtures and finite exchangeability," *Ann. Probab.*, 20, 1167–1173.