Prior processes for Bayesian nonparametrics

Donato Michele Cifarelli e Pietro Muliere
Istituto di Metodi Quantitativi
Università "Luigi Bocconi" - Milano

Piercesare Secchi
Dipartimento di Matematica "Francesco Brioschi"
Politecnico di Milano

September 21, 1999

Abstract

Following an idea of Regazzini and Petris (1992) we present a method for constructing random probability measures for Bayesian nonparametrics based on de Finetti's Representation Theorem.

This work is part of a book on 'Bayesian theory in a completely predictive perspective' that the authors are writing under contract with John Wiley & Sons, Ltd.

1 Introduction

In Bayesian nonparametrics the rôle of the parameter appearing in a statistical model is taken by a probability distribution; therefore, the parameter space becomes a class of probability distributions defined on a given sample space. A Bayesian usually considers the class of all probability distributions on the sample space and defines a probability on this class, the so called prior distribution.

A precise description of a Bayesian nonparametric problem, the first to our knowledge, appears in de Finetti (1935) in a paper about the problem of fitting a smooth curve onto an empirical distribution. At that time workable prior distributions on spaces whose elements are probabilities were not known. For these we have to wait until the 60' when a few schemes were put forward. D.A. Freedman (1963) and J. Fabius (1964) introduced a class of random probability distributions named tailfree. Unfortunately, these papers were not directly aimed at Bayesian nonparametric analysis so that their importance with respect to this topic was not immediately recognized. J.E. Rolph (1968) suggested a scheme for constructing random probability distributions based on moments evaluation. A definite impetus to nonparametric inference whithin the Bayesian approach to statistics came eventually with the papers by Ferguson (1973, 1974) and Doksum (1974), based on the cited works of Fabius and Freedman, as well as on the papers by Dubins and Freedman (1963, 1966) and Freedman (1965). In these papers a particular tailfree prior distribution is presented, namely the Dirichlet process, that has the double advantage of having a large support, with respect to a suitable topology on the space of probability distributions on the sample space, and of being analytically manageable for Bayesian posterior computations. Since then the literature on Bayesian nonparametrics has grown enormously and the need for solutions of new problems has caused the introduction of new prior distributions.

In the next pages, following an idea of Regazzini and Petris (1992), we present a method for constructing prior distributions for Bayesian nonparametrics based on de Finetti's Representation Theorem. After having applied the method for a direct introduction of the Dirichlet process, we will consider the more general class of tailfree processes and a special subclass of them named Polya tree processes.

2 The nonparametric model

Let $\mathbf{X}$ be a metric space, separable and complete, endowed with its Borel σ-field $\mathcal{X}$. We write $\mathbf{P}$ for the class of all probability measures defined on $(\mathbf{X}, \mathcal{X})$ and we endow $\mathbf{P}$ with the σ-field $\mathcal{P}$ generated by the topology of weak convergence, that is the smallest σ-field according to which are measurable all mappings $\wp_A : \mathbf{P} \to [0,1]$, with $A \in \mathcal{X}$, defined for all $p \in \mathbf{P}$ by

$$\wp_A(p) = p(A).$$

In Bayesian statistics one often encounters the following situation: given a sample of n values in $\mathbf{X}$, these are assumed to be realizations of n random variables belonging to an infinite exchangeable sequence $\{X_n\}$ of random variables with values in $\mathbf{X}$. Exchangeability and de Finetti's Representation Theorem imply the existence of a random element $P \in \mathbf{P}$ such that, conditionally on $P = p$, the random variables X_n's are independent and identically distributed with probability distribution p. Initial opinions about the sequence $\{X_n\}$ are expressed through the probability distribution ν of P and updated by means of Bayes Theorem; all inferential questions related to the sequence $\{X_n\}$ are then answered on the basis of the conditional probability distribution of P, given the observed sample from $\{X_n\}$. The aim of this section is to construct a formal framework where this problem can be embedded.

Let us first consider the space $(\mathbf{X}^\infty, \mathcal{X}^\infty)$ where $\mathbf{X}^\infty$ indicates the set of all infinite sequences $x = (x_1, x_2, \ldots)$ of elements of $\mathbf{X}$ whereas $\mathcal{X}^\infty$ is the σ-field generated by the subsets of $\mathbf{X}^\infty$ of the type

$$A_1 \times \cdots \times A_n \times \mathbf{X}^\infty = \{x \in \mathbf{X}^\infty : x_1 \in A_1, \ldots, x_n \in A_n\}$$

with $n \geq 1$ and $A_1, \ldots, A_n \in \mathcal{X}$. It is a standard result of measure theory (Ash 1972, Corollary 2.7.3) that, for all $p \in \mathbf{P}$, there exist a unique probability p^∞ on $(\mathbf{X}^\infty, \mathcal{X}^\infty)$, called the product probability measure, such that

$$p^\infty(A_1 \times \cdots \times A_n \times \mathbf{X}^\infty) = \prod_{i=1}^{n} p(A_i)$$

for all $n \geq 1$ and $A_1, \ldots, A_n \in \mathcal{X}$.

Since we are concerned with an inferential problem whose basic ingredients are an infinite exchangeable sequence $\{X_n\}$ of random variables with values in $\mathbf{X}$ and a random element $P \in \mathbf{P}$, the natural space where to embed it is the product space $(\mathbf{X}^\infty \times \mathbf{P}, \mathcal{X}^\infty \times \mathcal{P})$ where

$$\mathbf{X}^\infty \times \mathbf{P} = \{(x, p) : x \in \mathbf{X}^\infty, p \in \mathbf{P}\}$$

while $\mathcal{X}^\infty \times \mathcal{P}$ is the smallest σ-field of subsets of $\mathbf{X}^\infty \times \mathbf{P}$ containing the family of measurable rectangles $\{A \times B : A \in \mathcal{X}^\infty, B \in \mathcal{P}\}$.

The probability distribution ν of P, often called *the prior probability distribution*, plays a pivotal role for the definition of a probability measure on $(\mathbf{X}^\infty \times \mathbf{P}, \mathcal{X}^\infty \times \mathcal{P})$. In fact, given a probability measure ν defined on the space $(\mathbf{P}, \mathcal{P})$, we set

$$\pi(C) = \int_{\mathbf{P}} p^\infty(\{x \in \mathbf{X}^\infty : (x,p) \in C\})\nu(dp) \tag{2.1}$$

for all $C \in \mathcal{X}^\infty \times \mathcal{P}$. Note that, when $C = (A_1 \times \cdots \times A_n \times \mathbf{X}^\infty) \times B$, with $n \geq 1, A_i \in \mathcal{X}$ for $i = 1, \ldots, n$ and $B \in \mathcal{P}$, then

$$\pi(C) = \int_B \prod_{i=1}^n p(A_i)\nu(dp).$$

2.2 **Lemma.** *π is a probability measure on* $(\mathbf{X}^\infty \times \mathbf{P}, \mathcal{X}^\infty \times \mathcal{P})$.

Proof. First observe that

$$\pi(\mathbf{X}^\infty \times \mathbf{P}) = \int_{\mathbf{P}} p^\infty(\mathbf{X}^\infty)\nu(dp) = 1.$$

Now let C_1, C_2 be two disjoint sets in $\mathcal{X}^\infty \times \mathcal{P}$; then

$$\begin{aligned}
&\pi(C_1 \bigcup C_2) \\
&\quad = \int_{\mathbf{P}} p^\infty(\{x \in \mathbf{X}^\infty : (x,p) \in C_1 \bigcup C_2\})\nu(dp) \\
&\quad = \int_{\mathbf{P}} p^\infty(\{x \in \mathbf{X}^\infty : (x,p) \in C_1\} \bigcup \{x \in \mathbf{X}^\infty : (x,p) \in C_2\})\nu(dp) \\
&\quad = \int_{\mathbf{P}} p^\infty(\{x \in \mathbf{X}^\infty : (x,p) \in C_1\})\nu(dp) + \int_{\mathbf{P}} p^\infty(\{x \in \mathbf{X}^\infty : (x,p) \in C_2\})\nu(dp) \\
&\quad = \pi(C_1) + \pi(C_2)
\end{aligned}$$

where the third equality follows from the fact that the sets

$$\{x \in \mathbf{X}^\infty : (x,p) \in C_1\} \text{ and } \{x \in \mathbf{X}^\infty : (x,p) \in C_2\}$$

are disjoint for all $p \in \mathbf{P}$. This proves that π is finitely additive.

In order to prove that π is countably additive, let $C_1, C_2, \ldots, C_n, \ldots \in \mathcal{X}^\infty \times \mathcal{P}$ be an infinite sequence of sets decreasing to the empty set; note that, given any $p \in \mathbf{P}$, the sets

$$\{x \in \mathbf{X}^\infty : (x,p) \in C_n\} \in \mathcal{X}^\infty, \quad n = 1, 2, \ldots,$$

decrease to the empty set when n grows to infinity. Then

$$\begin{aligned}
&\lim_{n\to\infty} \pi(C_n) \\
&\quad = \lim_{n\to\infty} \int_{\mathbf{P}} p^\infty(\{x \in \mathbf{X}^\infty : (x,p) \in C_n\})\nu(dp) \\
&\quad = \int_{\mathbf{P}} \lim_{n\to\infty} p^\infty(\{x \in \mathbf{X}^\infty : (x,p) \in C_n\})\nu(dp) = 0
\end{aligned}$$

where the next to the last equality holds because of Dominated Convergence Theorem.

Hence π is a probability on $(\mathbf{X}^\infty \times \mathbf{P}, \mathcal{X}^\infty \times \mathcal{P})$. ◇

2.3 **Definition**. *Given a probability measure ν on $(\mathbf{P}, \mathcal{P})$, we call statistical model the triple*

$$(\mathbf{X}^\infty \times \mathbf{P}, \mathcal{X}^\infty \times \mathcal{P}, \pi),$$

where the probability π is defined as in (2.1).

Since π is a probability measure on the product space $(\mathbf{X}^\infty \times \mathbf{P}, \mathcal{X}^\infty \times \mathcal{P})$, it induces two probability measures on the spaces $(\mathbf{X}^\infty, \mathcal{X}^\infty)$ and $(\mathbf{P}, \mathcal{P})$ respectively. In fact, for all $B \in \mathcal{P}$,

$$\pi(\mathbf{X}^\infty \times B) = \nu(B),$$

the prior probability distribution. Analogously, for all $A \in \mathcal{X}^\infty$, set

$$\tau(A) = \pi(A \times \mathbf{P});$$

τ is a probability measure on the space $(\mathbf{X}^\infty, \mathcal{X}^\infty)$ commonly called the law of the process $\{X_n\}$.

2.4 **Lemma**. *τ is exchangeable and its de Finetti measure is ν.*

Proof. Let $n \geq 1, A_1, \ldots, A_n \in \mathcal{X}$. For each permutation $\sigma = (\sigma(1), \ldots, \sigma(n))$ of $(1, \ldots, n)$,

$$\begin{aligned}
&\tau(A_{\sigma(1)} \times \cdots \times A_{\sigma(n)} \times \mathbf{X}^\infty) \\
&\quad = \int_{\mathbf{P}} \prod_{i=1}^{n} p(A_i)\nu(dp) = \tau(A_1 \times \cdots \times A_n \times \mathbf{X}^\infty);
\end{aligned} \tag{2.5}$$

this proves that τ is exchangeable.

The fact that ν is the de Finetti measure of τ follows from (2.5) and de Finetti Representation Theorem. ◇

Any random variable belonging to an infinite exchangeable sequence $\{X_n\}$ and with values in $\mathbf{X}$ can be regarded as a projection of $\mathbf{X}^\infty \times \mathbf{P}$ into $\mathbf{X}$. In fact in what follows we will assume that, for all $n \geq 1$, $X_n : \mathbf{X}^\infty \times \mathbf{P} \to \mathbf{X}$ is defined by setting $X_n(x,p) = x_n$ for all $x = (x_1, x_2, \ldots) \in \mathbf{X}^\infty$ and $p \in \mathbf{P}$. Given $n \geq 1$, let τ_n be the probability distribution induced on $(\mathbf{X}^n, \mathcal{X}^n)$ by the random vector $(X_1, \ldots, X_n)$: that is, for all $A \in \mathcal{X}^n$, set

$$\tau_n(A) = \pi((X_1, \ldots, X_n) \in A) = \tau(\{x \in \mathbf{X}^\infty : (x_1, \ldots, x_n) \in A\}). \tag{2.6}$$

Therefore, for all $n \geq 1$ and $A_1, \ldots, A_n \in \mathcal{X}$,

$$\pi(X_1 \in A_1, \ldots, X_n \in A_n) = \tau(A_1 \times \cdots \times A_n \times \mathbf{X}^\infty);$$

which shows that the sequence $\{X_n\}$ has law equal to τ and therefore is exchangeable with de Finetti measure ν.

Analogously, if $P : \mathbf{X}^\infty \times \mathbf{P} \to \mathbf{P}$ is defined by setting $P(x,p) = p$ for all $x = (x_1, x_2, \ldots) \in \mathbf{X}^\infty$ and $p \in \mathbf{P}$, then it is easy to check that the random element $P \in \mathbf{P}$ has probability distribution ν. It is well to remark that, since for all $n \geq 1$, $A_1, \ldots, A_n \in \mathcal{X}$ and $B \in \mathcal{P}$,

$$\pi(X_1 \in A_1, \ldots, X_n \in A_n, P \in B) = \pi((A_1 \times \ldots \times A_n \times \mathbf{X}^\infty) \times B) = \int_B \prod_{i=1}^n p(A_i)\nu(dp)$$

we may conclude that, conditionally on $P = p$, the random variables of the sequence $\{X_n\}$ are independent and identically distributed with probability distribution p.

Furthermore, since $\mathbf{X}^\infty \times \mathbf{P}$ is separable and complete, for every $n \geq 1$ and $(x_1, \ldots, x_n) \in \mathbf{X}^n$, there exists a regular conditional probability q on $\mathcal{X}^\infty \times \mathcal{P}$ given $X_1 = x_1, \ldots, X_n = x_n$ (Ash 1972, Theorem 6.6.5). This means that, for $n \geq 1$, there exist a function $q : \mathbf{X}^n \times (\mathcal{X}^\infty \times \mathcal{P}) \to [0,1]$ such that:

(a) For all $(x_1, \ldots, x_n) \in \mathbf{X}^n$, $q((x_1, \ldots, x_n), \cdot)$ is a probability on $\mathcal{X}^\infty \times \mathcal{P}$.

(b) For all $C \in \mathcal{X}^\infty \times \mathcal{P}$, $q(\cdot, C)$ is a measurable map from $(\mathbf{X}^n, \mathcal{X}^n)$ to $([0,1], \mathcal{B}[0,1])$.

(c) For all $A \in \mathcal{X}^n, C \in \mathcal{X}^\infty \times \mathcal{P}$,

$$\pi(C \bigcap (X_1, \ldots, X_n)^{-1}(A)) = \int_A q((x_1, \ldots, x_n), C)\tau_n(dx_1 \cdots dx_n).$$

We will often write $\pi(C|X_1 = x_1, \ldots, X_n = x_n)$ for $q((x_1, \ldots, x_n), C)$.

Given $n \geq 1$ and $(x_1, \ldots, x_n) \in \mathbf{X}^n$ the probability $\nu(\cdot|X_1 = x_1, \ldots, X_n = x_n)$ on $(\mathbf{P}, \mathcal{P})$, defined by setting

$$\nu(B|X_1 = x_1, \ldots, X_n = x_n) = q((x_1, \ldots, x_n), \mathbf{X}^\infty \times B) = \pi(\mathbf{X}^\infty \times B|X_1 = x_1, \ldots, X_n = x_n)$$

for all $B \in \mathcal{P}$, is called the *posterior probability distribution* of P: this is the conditional distribution of P given $X_1 = x_1, \ldots, X_n = x_n$. Analogously, given $m, n \geq 1$ and $(x_1, \ldots, x_n) \in \mathbf{X}^n$, the probability $\tau_m(\cdot|X_1 = x_1, \ldots, X_n = x_n)$ on $(\mathbf{X}^m, \mathcal{X}^m)$ defined by setting

$$\tau_m(A|X_1 = x_1, \ldots, X_n = x_n) = q((x_1, \ldots, x_n), (A \times \mathbf{X}^\infty) \times \mathbf{P}),$$

for all $A \in \mathcal{X}^m$, is called *predictive probability distribution* of $X_{n+1}, \ldots, X_{n+m}$: this is in fact the conditional probability distribution of $X_{n+1}, \ldots, X_{n+m}$ given $X_1 = x_1, \ldots, X_n = x_n$.

3 Constructing random probability measures

The aim of this section is to discuss a method for constructing probability measures on the space $(\mathbf{P}, \mathcal{P})$. The method looks appealing for determining prior probability distributions for nonparametric statistical models; as a first illustration, we will use it in the next section for constructing the celebrated Dirichlet process introduced by Ferguson (1973). The same method will then be used for constructing tailfree processes, very general priors of which Dirichlet processes, as well as the Polya tree processes of Mauldin, Sudderth and Williams (1992), are a subclass.

If ν is a probability defined on $(\mathbf{P}, \mathcal{P})$, for all $n \geq 1$ and $A_1, \ldots, A_n$ distinct elements of $\mathcal{X}$ let us indicate with $q_{A_1,\ldots,A_n}$ the probability distribution induced on $([0,1]^n, \mathcal{B}^n[0,1])$ by the random vector $(\wp_{A_1}, \ldots, \wp_{A_n})$; that is

$$q_{A_1,\ldots,A_n}(C) = \nu(\{p \in \mathbf{P} : (\wp_{A_1}(p), \ldots, \wp_{A_n}(p)) \in C\}) = \nu(\{p \in \mathbf{P} : (p(A_1), \ldots, p(A_n)) \in C\}) \tag{3.1}$$

for every Borel subset C of $\mathcal{B}^n[0,1]$.

Set

$$\mathbf{Q} = \{q_{A_1,\ldots,A_n} : n \geq 1 \text{ and } A_1, \ldots, A_n \text{ distinct elements of } \mathcal{X}\}.$$

We begin by enumerating some natural consistency properties satisfied by the elements of $\mathbf{Q}$:

(C_1) For all $n \geq 1$ and $A_1, \ldots, A_n$ distinct elements of $\mathcal{X}$, $q_{A_1,\ldots,A_n}$ is a probability measure on $([0,1]^n, \mathcal{B}^n[0,1])$ such that, for every Borel subset C of $\mathcal{B}[0,1]^n$,

$$q_{A_1,\ldots,A_n}(C) = q_{A_{\sigma(1)},\ldots,A_{\sigma(n)}}(\sigma C)$$

where $\sigma = (\sigma(1), \ldots, \sigma(n))$ is any permutation of $(1, \ldots, n)$ and

$$\sigma C = \{(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \in [0,1]^n : (x_1, \ldots, x_n) \in C\}.$$

This is obvious since, for every Borel subset C of $[0,1]^n$, the two sets

$$\{p \in \mathbf{P} : (\wp_{A_1}(p), \ldots, \wp_{A_n}(p)) \in C\} \text{ and } \{p \in \mathbf{P} : (\wp_{A_{\sigma(1)}}(p), \ldots, \wp_{A_{\sigma(n)}}(p)) \in \sigma C\}$$

describe the same event.

(C_2)

$$q_{\mathbf{X}} = \delta_1$$

where δ_x indicates the point mass at x, for all $x \in \Re$.

This property is true since

$$\wp_{\mathbf{X}}(p) = p(\mathbf{X}) = 1$$

for all $p \in \mathbf{P}$.

(C_3) Let $n \geq 1$ and $A_1, \ldots, A_n$ distinct elements of $\mathcal{X}$. Let $B_1, \ldots, B_m \in \mathcal{X}$ be a finite partition of $\mathbf{X}$ such that

$$A_1 = \bigcup_{(1)} B_j, \ldots, A_n = \bigcup_{(n)} B_j$$

where, for $i = 1, \ldots, n$, (i) indicates the set $\{j \in \{1, \ldots, n\} : B_j \subseteq A_i\}$ which we assume to be nonempty. Then, for every Borel subset C of $[0,1]^n$,

$$q_{A_1,\ldots,A_n}(C) = q_{B_1,\ldots,B_m}(\{(x_1, \ldots, x_m) \in [0,1]^m : (\sum_{(1)} x_j, \ldots, \sum_{(n)} x_j) \in C\}).$$

This property follows after noticing that, for all $p \in \mathbf{P}$ and $i = 1, \ldots, n$,

$$\wp_{A_i}(p) = p(\bigcup_{(i)} B_j) = \sum_{(i)} p(B_j) = \sum_{(i)} \wp_{B_j}(p).$$

(C_4) Let $\{A_n\}$ be a sequence of elements of $\mathcal{X}$ decreasing to the empty set. Then q_{A_n} weakly converges to δ_0, when n goes to infinity.

This follows from the fact that, for all $p \in \mathbf{P}$,

$$\lim_{n\to\infty} \wp_{A_n}(p) = \lim_{n\to\infty} p(A_n) = 0 = \wp_{\emptyset}(p).$$

Therefore a probability measure ν on $(\mathbf{P}, \mathcal{P})$ generates a family $\mathbf{Q}$ of probability distributions whose elements are indexed by finite sequences of distinct elements of $\mathcal{X}$ and satisfy the consistency conditions $(C_1) - (C_4)$. The main question we want to address in this section is whether, given a family

$$\mathbf{Q} = \{q_{A_1,\ldots,A_n} : n \geq 1 \text{ and } A_1, \ldots, A_n \text{ distinct elements of } \mathcal{X}\}$$

whose elements are probability measures satisfying $(C_1)-(C_4)$, there exists a unique probability measure ν on $(\mathbf{P},\mathcal{P})$ such that, for all $n\geq 1$ and distinct $A_1,\ldots,A_n\in\mathcal{X}$, $q_{A_1,\ldots,A_n}$ is the probability distribution of the random vector $(\wp_{A_1},\ldots,\wp_{A_n})$. The next theorem answers positively to this question; the argument for its proof, based on an idea of Regazzini and Petris (1992) and Regazzini (1996), considers an exchangeable probability measure μ generated by $\mathbf{Q}$ on the space $(\mathbf{X}^\infty,\mathcal{X}^\infty)$ and shows that its unique de Finetti measure is the right ν on $(\mathbf{P},\mathcal{P})$.

3.2 **Theorem**. *If the elements of* $\mathbf{Q}$ *satisfy the consistency conditions* $(C_1)-(C_4)$, *there exists a unique probability measure* ν *on* $(\mathbf{P},\mathcal{P})$ *such that, for all* $n\geq 1$ *and* $A_1,\ldots,A_n$ *distinct elements of* $\mathcal{F}$,

$$\nu(\{p\in\mathbf{P}:(\wp_{A_1}(p),\ldots,\wp_{A_n}(p))\in C\})=q_{A_1,\ldots,A_n}(C)$$

for every Borel subset C *of* $[0,1]^n$.

The theorem will be proved by means of the following lemmas that are stated with the assumption that the elements of $\mathbf{Q}$ satisfy $(C_1)-(C_4)$.

For all $n\geq 1$, let $\mathcal{F}^n$ be the field of finite unions of disjoint rectangles of the type

$$R=D_1\times\cdots\times D_n$$

with $D_1,\ldots,D_n\in\mathcal{X}$. Given $n\geq 1$, for every rectangle $R=D_1\times\cdots\times D_n\in\mathcal{F}^n$ we define

$$\mu_n(R)=\mu_n(D_1\times\cdots\times D_n)=\int_{[0,1]^m}y_1^{r_1}\cdots y_m^{r_m}q_{A_1,\ldots,A_m}(dy_1\cdots dy_m) \tag{3.3}$$

where $A_1,\ldots,A_m$ are the distinct elements of the sequence $D_1,\ldots,D_n$ and, for $i=1,\ldots,m$, r_i indicates the number of times the set A_i appears in the sequence. If $F=\bigcup_{i=1}^k R_i$ is a finite disjoint union of rectangles of $\mathcal{F}^n$ we define

$$\mu_n(F)=\sum_{i=1}^k\mu_n(R_i). \tag{3.4}$$

3.5 **Lemma**. *For all* $n\geq 1$, *the set function* μ_n *is a well defined, finitely additive exchangeable probability on* $\mathcal{F}^n$.

Proof. Let $n\geq 1$. We begin by showing that μ_n is well defined on $\mathcal{F}^n$.

First notice that, if $R=D_1\times\cdots\times D_n$ is a rectangle, the definition of $\mu_n(R)$ does not depend on the order of the sets of the sequence $D_1,\ldots,D_n$ nor on the order of the distinct elements $A_1, \ldots, A_n$ of the sequence since the elements of $\mathbf{Q}$ satisfy (C_1). This also implies that, if μ_n is a well defined probability on $\mathcal{F}^n$, then it is exchangeable.

Next assume that the rectangle R is a finite union of disjoint rectangles of $\mathcal{F}^n$; that is assume that $R = \bigcup_{i=1}^k R_i$ where, for $i = 1, \ldots, k$,

$$R_i = D_{i,1} \times \ldots \times D_{i,n}$$

with $D_{i,1}, \ldots, D_{i,n} \in \mathcal{X}$, and the R_i are disjoint. Then

$$R = (\bigcup_{i=1}^k D_{i,1}) \times \ldots \times (\bigcup_{i=1}^k D_{i,n})$$

and according to (3.3),

$$\mu_n(R) = \int_{[0,1]^m} y_1^{r_1} \cdots y_m^{r_m} q_{A_1,\ldots,A_m}(dy_1 \cdots dy_m) \tag{3.6}$$

where $A_1, \ldots, A_m$ are the distinct sets appearing in the sequence $\bigcup_{i=1}^k D_{i,1}, \ldots, \bigcup_{i=1}^k D_{i,n}$. We need to show that

$$\mu_n(R) = \sum_{i=1}^k \mu_n(R_i). \tag{3.7}$$

Let $\{B_1, \ldots, B_t\}$ be a partition of $\mathbf{X}$ generated by the sets $D_{i,j}$ with $i = 1, \ldots, k$ and $j = 1, \ldots, n$; that is the class of sets in $\mathcal{X}$ obtained by taking all the possible intersections of the sets $D_{i,j}$ and their complements. For $i = 1, \ldots, k$ and $j = 1, \ldots, n$ let $(i, j) = \{s \in \{1, \ldots, t\} : B_s \subseteq D_{i,j}\}$ and $(j) = \bigcup_{i=1}^k (i, j)$. Then (3.6) and property (C_3) imply that

$$\mu_n(R) = \int_{[0,1]^t} \prod_{j=1}^n [\sum_{r \in (j)} y_r] q_{B_1,\ldots,B_t}(dy_1 \cdots dy_t).$$

However (3.3) and property (C_3) imply also that, for $i = 1, \ldots, k$,

$$\mu_n(R_i) = \int_{[0,1]^t} \prod_{j=1}^n [\sum_{r \in (i,j)} y_r] q_{B_1,\ldots,B_t}(dy_1 \cdots dy_t).$$

Therefore (3.7) is proved if we show that

$$\prod_{j=1}^n [\sum_{r \in (j)} y_r] = \sum_{i=1}^k \prod_{j=1}^n [\sum_{r \in (i,j)} y_r] \tag{3.8}$$

for all $(y_1, \ldots, y_t) \in [0, 1]^t$. For any given $(y_1, \ldots, y_t) \in [0, 1]^t$, imagine that the sets $B_1, \ldots, B_t$ have measures $y_1, \ldots, y_t$ respectively; then both sides of (3.8) would represent the product measure of R. Therefore they must be the same.

Now let $F = \bigcup_{i=1}^{k} R_i$ where the R_i are disjoint rectangles of $\mathcal{F}^n$ and assume that F can also be represented as $\bigcup_{s=1}^{k'} R'_s$ where the R'_i are also disjoint rectangles of $\mathcal{F}^n$. Then we need to show that

$$\sum_{i=1}^{k} \mu_n(R_i) = \sum_{s=1}^{k'} \mu_n(R'_s). \tag{3.9}$$

However, for $s = 1, \ldots, k'$,

$$R'_s = R'_s \bigcap F = \bigcup_{i=1}^{k} (R'_s \bigcap R_i)$$

so that, by (3.7),

$$\mu_n(R'_s) = \sum_{i=1}^{k} \mu_n(R'_s \bigcap R_i);$$

hence

$$\sum_{s=1}^{k'} \mu_n(R'_s) = \sum_{s,i} \mu_n(R'_s \bigcap R_i).$$

Analogously, one shows that

$$\sum_{i=1}^{k} \mu_n(R_i) = \sum_{s,i} \mu_n(R'_s \bigcap R_i)$$

and this proves (3.9).

Therefore μ_n is well defined on $\mathcal{F}^n$. Definitions (3.3) and (3.4) imply that μ_n is nonnegative and additive. Finally note that

$$\mu_n(\mathbf{X}^n) = \int_{[0,1]} y^n q_{\mathbf{X}}(dy) = 1$$

because of (C_2). Hence μ_n is a finitely additive probability on $\mathcal{F}^n$. ◇

3.10 **Lemma**. *For all $n \geq 1$ and every sequence $\{E_k\}$ of elements of $\mathcal{F}^n$ such that $E_1 \supseteq E_2 \supseteq \ldots$ and $\bigcap_{k=1}^{\infty} E_k = \emptyset$:*

$$\lim_{k \to \infty} \mu_n(E_k) = 0.$$

Proof. Let us first prove that the lemma is true for μ_1. In fact, $\mathcal{F}^1 = \mathcal{X}$ and if $\{E_k\}$ is a sequence of elements of $\mathcal{X}$ decreasing to the empty set, then

$$\lim_{k \to \infty} \mu_1(E_k) = \lim_{k \to \infty} \int_{[0,1]} y q_{E_k}(dy) = 0$$

because of (C_4). This fact and Lemma 3.5 prove that μ_1 is a probability measure defined on $\mathcal{X}$.

Now notice that, for all $A \in \mathcal{X}$ and $n \geq 2$,

$$\mu_n(A \times \mathbf{X}^{n-1}) = \mu_n(\mathbf{X} \times A \times \mathbf{X}^{n-2}) = \ldots = \mu_n(\mathbf{X}^{n-1} \times A) = \mu_1(A).$$

In fact this is trivially true when $A = \mathbf{X}$; when $A \subset \mathbf{X}$ we have, for instance,

$$\mu_n(A \times \mathbf{X}^{n-1}) = \int_{[0,1]^2} y_1 y_2^{n-1} q_{A,\mathbf{X}}(dy_1 dy_2) = \int_{[0,1]^2} y_1 (y_1 + y_2)^{n-1} q_{A,A^c}(dy_1 dy_2)$$

where the first equality is true by definition whereas the next one follows from (C_3). However, for all $D \in \mathcal{B}[0,1]$,

$$q_{A,A^c}(\{(y_1,y_2) \in [0,1]^2 : y_1 + y_2 \in D\}) = q_{\mathbf{X}}(D) = \delta_1(D)$$

where the first equality follows from (C_3) while the last one is true because of (C_2). Therefore, for all $D_1, D_2 \in \mathcal{B}[0,1]$,

$$\begin{aligned} & q_{A,A^c}(\{(y_1,y_2) \in [0,1]^2 : y_1 \in D_1, y_1 + y_2 \in D_2\}) \\ = \; & \delta_1(D_2) q_{A,A^c}(\{(y_1,y_2) \in [0,1]^2 : y_1 \in D_1\}) \\ = \; & \delta_1(D_2) q_A(D_1) \end{aligned}$$

where the last equality follows once again from (C_3). This implies that

$$\mu_n(A \times \mathbf{X}^{n-1}) = \int_{[0,1]} y_1 q_A(dy_1) = \mu_1(A_1).$$

Since $\mathbf{X}$ is a separable and complete metric space endowed with its Borel σ-field $\mathcal{X}$, for all $A \in \mathcal{X}$ and $\epsilon > 0$ there is a compact set $K \in \mathcal{X}$, $K \subseteq A$, such that $\mu_1(A \backslash K) < \epsilon$ (Ash, Theorem 4.3.8). This implies that, for all $n \geq 1$, $\epsilon > 0$ and $F \in \mathcal{F}^n$, there is a set $K \in \mathcal{F}^n$ compact with respect to the product topology of $\mathbf{X}^n$ and such that $K \subseteq F$ and $\mu_n(F \backslash K) < \epsilon$. In fact let $F = \bigcup_{i=1}^m R_i$ where $R_1, \ldots, R_m$ are disjoint rectangles of $\mathcal{F}^n$; for $i = 1, \ldots, m$, assume that

$$R_i = A_{i,1} \times \cdots \times A_{i,n}$$

with $A_{i,j} \in \mathcal{X}$ for $j = 1, \ldots, n$. For $i = 1, \ldots, m$ and $j = 1, \ldots, n$ let $K_{i,j}$ be a compact subset of $A_{i,j}$ such that

$$\mu_1(A_{i,j} \backslash K_{i,j}) < \frac{\epsilon}{nm}$$

and set $K_i = K_{i,1} \times \cdots \times K_{i,n}$; then $K_i \subseteq R_i$. Let $K = \bigcup_{i=1}^m K_i \in \mathcal{F}^n$. K is compact with respect to the product topology of $\mathbf{X}^n$ and

$$\mu_n(F \backslash K) = \mu_n(\bigcup_{i=1}^m (R_i \backslash K_i))$$

$$\begin{aligned}
&= \sum_{i=1}^{m} \mu_n(R_i \backslash K_i) \\
&\leq \sum_{i=1}^{m} \mu_n\Big(\bigcup_{j=1}^{n} \mathbf{X} \times \cdots \times \mathbf{X} \times A_{i,j} \backslash K_{i,j} \times \mathbf{X} \times \cdots \times \mathbf{X}\Big) \\
&\leq \sum_{i=1}^{m}\sum_{j=1}^{n} \mu_1(A_{i,j} \backslash K_{i,j}) < \epsilon
\end{aligned}$$

where the first equality follows from the fact that the rectangles R_i are disjoint.

Now consider a sequence $\{E_k\}$ of elements of $\mathcal{F}^n$ such that $E_1 \supseteq E_2 \supseteq \ldots$ and $\bigcap_{k=1}^{\infty} E_k = \emptyset$. By way of contradiction assume that $\{\mu_n(E_k)\}$ does not converge to 0. Then, there is a $\epsilon > 0$ such that

$$\mu_n(E_k) \geq \epsilon \tag{3.11}$$

for all k. For $s = 1, 2, \ldots$, let K_s be a compact subset of E_s such that $\mu_n(E_s \backslash K_s) < \epsilon 2^{-s}$. Then, for all $k = 1, 2, \ldots$,

$$\begin{aligned}
\mu_n\Big(E_k \backslash \bigcap_{s=1}^{k} K_s\Big) &\leq \sum_{s=1}^{k} \mu_n(E_k \backslash K_s) \\
&\leq \sum_{s=1}^{k} \mu_n(E_s \backslash K_s) \\
&\leq \frac{\epsilon}{2}
\end{aligned}$$

where the second inequality follows from the fact that $E_k \subseteq E_{k-1} \subseteq \ldots \subseteq E_1$. Therefore, for all $k \geq 1$,

$$\mu_n\Big(\bigcap_{s=1}^{k} K_s\Big) \geq \frac{\epsilon}{2};$$

hence, for all $k \geq 1$, $\bigcap_{s=1}^{k} K_s$ is non empty, and since the K_s are compact, this implies that $\bigcap_{s=1}^{k} K_s \neq \emptyset$. However this contradicts the assumption that $\bigcap_{s=1}^{\infty} E_s = \emptyset$ since $K_s \subseteq E_s$, for all s. Therefore (3.11) is false and

$$\lim_{k \to \infty} \mu_n(E_k) = 0.$$

◇

The previous two lemmas show that, for all $n \geq 1$, μ_n is a probability measure defined on the field of rectangles $\mathcal{F}^n$. By means of Caratheodory's Theorem we may then extend uniquely μ_n to a probability measure on the σ-field generated by $\mathcal{F}^n$ which coincides with $\mathcal{X}^n$, the product σ-field in $\mathbf{X}^n$; we will continue to write μ_n for the probability measure on $(\mathbf{X}^n, \mathcal{X}^n)$ thus constructed.

3.12 **Lemma**. *There is a unique probability measure μ on $(\mathbf{X}^\infty, \mathcal{X}^\infty)$ such that, for all $n \geq 1$ and $B \in \mathcal{X}^n$,*

$$\mu(\{(x_1, x_2, \ldots) \in \mathbf{X}^\infty : (x_1, \ldots, x_n) \in B\}) = \mu_n(B).$$

Proof. We show that the sequence of probability measures

$$\mu_1, \mu_2, \ldots \text{ on } (\mathbf{X}^1, \mathcal{X}^1), (\mathbf{X}^2, \mathcal{X}^2), \ldots$$

possess Kolmogorov's consistency property (see Shiryayev (1984), Theorem 2.3.3). For this it is enough to verify that, for all $n \geq 1$ and every rectangle $R = D_1 \times \cdots D_n$, with $D_i \in \mathcal{X}$,

$$\mu_{n+1}(R \times \mathbf{X}) = \mu_n(R). \tag{3.13}$$

Before proving (3.13), assume that $A_1, \ldots, A_m$ are $m \geq 1$ distinct elements of $\mathcal{X}$ such that $A_i \subset \mathbf{X}$ for all $i = 1, \ldots, m$, and consider the probability measure $q_{X, A_1, \ldots, A_m}$ defined on $\mathcal{B}^{m+1}[0,1]$. If $B_1, \ldots, B_k$ is a finite partition of $\mathbf{X}$ with elements in $\mathcal{X}$ and such that

$$A_1 = \bigcup_{(1)} B_j, \ldots, A_n = \bigcup_{(m)} B_j$$

where the set $(i) = \{j \in \{1, \ldots, k\} : B_j \subseteq A_i\}$ is assumed to be nonempty for $i = 1, \ldots, m$, property (C_3) implies that,

$$\begin{aligned} &q_{X, A_1, \ldots, A_m}(C_0 \times \cdots \times C_m) \\ &\quad = q_{B_1, \ldots, B_k}(\{(x_1, \ldots, x_k) \in [0,1]^k : \sum_{i=1}^k x_j \in C_0, \sum_{(1)} x_j \in C_1, \ldots, \sum_{(m)} x_j \in C_m\}), \end{aligned}$$

for all $C_0, \ldots, C_m \in \mathcal{B}[0,1]$. However

$$q_{B_1, \ldots, B_k}(\{(x_1, \ldots, x_k) \in [0,1]^k : \sum_{i=1}^k x_j \in C_0\}) = q_{\mathbf{X}}(C_0) = \delta_1(C_0)$$

where the first equality follows from (C_3) whereas the next one is true because of (C_2). Therefore

$$\begin{aligned} &q_{X, A_1, \ldots, A_m}(C_0 \times \cdots \times C_m) \\ &\quad = \delta_1(C_0) q_{B_1, \ldots, B_k}(\{(x_1, \ldots, x_k) \in [0,1]^k : \sum_{(1)} x_j \in C_1, \ldots, \sum_{(m)} x_j \in C_m\}) \\ &\quad = \delta_1(C_0) q_{A_1, \ldots, A_m}(C_1 \times \cdots \times C_m) \end{aligned} \tag{3.14}$$

for all $C_0, C_1, \ldots, C_m \in \mathcal{B}[0,1]$.

Now let $n \geq 1$ and $R = D_1 \times \cdots \times D_n$ with $D_i \in \mathcal{X}$ for $i = 1, \ldots, n$. If $D_i = \mathbf{X}$ for all $i = 1, \ldots, n$, (3.13) is evident. If at least one $D_i \subset \mathbf{X}$, let $A_1, \ldots, A_m$ be the distinct elements different from $\mathbf{X}$ of the sequence $D_1, \ldots, D_n$; set r_0 to be the number of times $\mathbf{X}$ appears in the sequence $D_1, \ldots, D_n$ and, for $i = 1, \ldots, m$, let r_i be the number of times the set A_i appears in the same sequence. Then

$$\begin{aligned}
&\mu_{n+1}(R \times \mathbf{X}) \\
&\quad = \int_{[0,1]^{m+1}} y_0^{r_0+1} \cdots y_m^{r_m} q_{\mathbf{X}, A_1, \ldots, A_m}(dy_0 \cdots dy_m) \\
&\quad = \int_{[0,1]^m} y_1^{r_1} \cdots y_m^{r_m} q_{A_1, \ldots, A_m}(dy_1 \cdots dy_m) \\
&\quad = \mu_n(R)
\end{aligned}$$

where the last two equalities follow from (3.14) and the definition of μ_n. This shows that (3.13) holds and concludes the proof of the lemma. ◇

3.15 **Lemma**. *The probability measure μ on $(\mathbf{X}^\infty, \mathcal{B}^\infty)$ is exchangeable.*

Proof. Because of the previous lemma, exchangeability for μ follows if we show that, for all $n \geq 1$ and $D_1, \ldots, D_n \in \mathcal{X}$,

$$\mu_n(D_1 \times \cdots \times D_n) = \mu_n(D_{\sigma(1)} \times \cdots \times D_{\sigma(n)})$$

for every permutation $\sigma = (\sigma(1), \ldots, \sigma(n))$ of $(1, \ldots, n)$. But this is true since μ_n is exchangeable. ◇

3.16 **Remark**. It is now easy to see that a measure μ on $(\mathbf{X}^\infty, \mathcal{X}^\infty)$ is exchangeable if and only if there is a class $\mathbf{Q}$ of probability measures indexed by finite sequences of distinct elements of $\mathcal{X}$ and satisfying the consistency conditions $(C_1) - (C_4)$ such that, for all $n \geq 1$ and $D_1, \ldots, D_n \in \mathcal{X}$,

$$\mu(D_1 \times \cdots \times D_n \times \mathbf{X}^\infty)$$

is equal to the right member of (3.3). The previous lemmas prove the sufficiency part. To prove that the condition is also necessary let μ be exchangeable and define the elements of $\mathbf{Q}$ by means of (3.1) where ν is the unique de Finetti measure of μ. ◇

Since μ is an exchangeable probability measure on $(\mathbf{X}^\infty, \mathcal{B}^\infty)$, it follows from de Finetti's Representation Theorem that there exists a unique probability measure ν on $(\mathbf{P}, \mathcal{P})$ such that, for all $n \geq 1$ and all $C_1, \ldots, C_n \in \mathcal{X}$,

$$\mu(C_1 \times \cdots \times C_n \times \mathbf{X} \cdots) = \int_{\mathbf{P}} \prod_{i=1}^{n} p(C_i)\nu(dp).$$

3.17 **Lemma**. *For all $n \geq 1$ and $A_1, \ldots, A_n$ distinct elements of $\mathcal{X}$,*

$$\nu(\{p \in \mathbf{P} : (\wp_{A_1}(p), \ldots, \wp_{A_n}(p)) \in C\}) = q_{A_1,\ldots,A_n}(C) \tag{3.18}$$

for every $C \in \mathcal{B}^n[0,1]$.

Proof. Fix $n \geq 1$ and $A_1, \ldots, A_n$ distinct elements of $\mathcal{X}$. For all $k_1, \ldots, k_n$ integers,

$$\begin{aligned} &E_\nu[\wp_{A_1}^{k_1} \cdots \wp_{A_n}^{k_n}] \\ &\quad = \int_{\mathbf{P}} p^{k_1}(A_1) \cdots p^{k_n}(A_n)\nu(dp) \\ &\quad = \mu(A_1 \times \cdots \times A_1(k_1 \text{ times}) \times \cdots\cdots \times A_n \times \cdots \times A_n(k_n \text{ times})) \\ &\quad = \int_{[0,1]^n} y_1^{k_1} \cdots y_n^{k_n} q_{A_1 \ldots A_n}(dy_1 \cdots dy_n). \end{aligned}$$

Since the probability distribution of $(\wp_{A_1}, \ldots, \wp_{A_n})$ is completely determined by its moments, this shows that it must be $q_{A_1 \ldots A_n}$ and proves the lemma. ◇

3.19 **Lemma**. *There is a unique probability measure ν on $(\mathbf{P}, \mathcal{P})$ for which (3.18) holds.*

Proof. Let $\tilde{\nu}$ be another probability measure on $(\mathbf{P}, \mathcal{P})$ for which (3.18) holds: let $\tilde{\mu}$ be the exchangeable probability measure on $(\mathbf{X}^\infty, \mathcal{X}^\infty)$ defined by setting, for all $A \in \mathcal{X}^\infty$,

$$\tilde{\mu}(A) = \int_{\mathbf{P}} p^\infty(A)\tilde{\nu}(dp).$$

Then, for all $n \geq 1$ and $D_1, \ldots, D_n \in \mathcal{X}$,

$$\begin{aligned} &\tilde{\mu}(D_1 \times \cdots D_n \times \mathbf{X}^\infty) \\ &\quad = \int_{\mathbf{P}} \prod_{i=1}^{n} p(D_i)\tilde{\nu}(dp) \\ &\quad = \int_{[0,1]^m} y_1^{r_1} \cdots y_m^{r_m} q_{A_1,\ldots,A_m}(dy_1 \cdots dy_m) \\ &\quad = \int_{\mathbf{P}} \prod_{i=1}^{n} p(D_i)\nu(dp) \\ &\quad = \mu(D_1 \times \cdots D_n \times \mathbf{X}^\infty) \end{aligned}$$

where $A_1, \ldots, A_m$ are the distinct elements of the sequence $D_1, \ldots, D_n$ and r_i counts the number of times the set A_i appears in the sequence for $i = 1, \ldots, n$.

This implies that $\tilde{\mu} = \mu$. Since the de Finetti measure of μ is unique, $\tilde{\nu} = \nu$. $\diamond$

The last lemma concludes the proof of Theorem 3.2 and settles affirmatively the main question of the section: given a family $\mathbf{Q}$ of probability distributions indexed by finite sequences of distinct elements of $\mathcal{X}$ and satisfying consistency conditions $(C_1) - (C_4)$ there is a unique probability measure ν on $(\mathbf{P}, \mathcal{P})$ for which (3.18) holds.

When applying Theorem 3.2 it is often handy to define the elements of $\mathbf{Q}$ by first defining $q_{A_1,\ldots,A_n}$ when $A_1, \ldots, A_n$ is a finite partition of $\mathbf{X}$ with elements in $\mathcal{X}$, that is a finite measurable partition of $\mathbf{X}$, and then using the hoped for consistency conditions $(C_1) - (C_4)$ for defining the remaining elements of $\mathbf{Q}$

In fact let

$$\mathbf{Q}' = \{q_{B_1,\ldots,B_n} : n \geq 1 \text{ and } B_1, \ldots, B_n \text{ measurable partition of } \mathbf{X}\}$$

and assume that

(C_1') For all $n \geq 1$ and $B_1, \ldots, B_n$ measurable partition of $\mathbf{X}$, $q_{B_1,\ldots,B_n}$ is a probability measure on $([0,1]^n, \mathcal{B}^n[0,1])$ such that, for every $C \in \mathcal{B}^n[0,1]$,

$$q_{B_1,\ldots,B_n}(C) = q_{B_{\sigma(1)},\ldots,B_{\sigma(n)}}(\sigma C)$$

where $\sigma = (\sigma(1), \ldots, \sigma(n))$ is any permutation of $(1, \ldots, n)$ and

$$\sigma C = \{(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \in [0,1]^n : (x_1, \ldots, x_n) \in C\}.$$

(C_2')

$$q_{\mathbf{X}} = \delta_1.$$

(C_3') Let $n \geq 1$ and $B_1, \ldots, B_n$ be a finite measurable partition of $\mathbf{X}$. if $B_1', \ldots, B_m'$ is a finite measurable partition of $\mathbf{X}$ finer than $B_1, \ldots, B_n$ let

$$(i) = \{j \in \{1, \ldots, m\} : B_j' \subseteq B_i\}$$

for $i = 1, \ldots, n$. Then, for every $C \in \mathcal{B}^n[0,1]$,

$$q_{B_1,\ldots,B_n}(C) = q_{B_1',\ldots,B_m'}(\{(x_1, \ldots, x_m) \in [0,1]^m : (\sum_{(1)} x_j, \ldots, \sum_{(n)} x_j) \in C\}).$$

(C_4') Let $\{A_n\}$ be a sequence of elements of $\mathcal{X}$ decreasing to the empty set. Then, for all $a \neq 0$,

$$\lim_{n\to\infty} q_{A_n,A_n^c}(\{(y_1,y_2)\in[0,1]^2 : y_1 \le a\}) = \delta_0((-\infty,a]).$$

Now, given $n \ge 1$ and any finite collection $\mathcal{A} = \{A_1,\dots,A_n\}$ of elements of $\mathcal{X}$, suppose that $\nabla = \{B_1,\dots,B_m\}, m \ge n$, is a finite measurable partition of $\mathbf{X}$ such that:

$$A_1 = \bigcup_{(1)} B_j, \dots, A_n = \bigcup_{(n)} B_j$$

where $(i) = \{j \in \{1,\dots,m\} : B_j \subseteq A_i\}$ for $i = 1,\dots,n$. We introduce the measurable map $\phi_{\mathcal{A};\nabla} : [0,1]^m \to [0,1]^n$ defined by setting, for all $(y_1,\dots,y_m) \in [0,1]^m$,

$$\phi_{\mathcal{A};\nabla}((y_1,\dots,y_m)) = (\sum_{(1)} y_j, \dots, \sum_{(n)} y_j)$$

with the convention that $\sum_{(i)} y_j = 0$ if $\{j \in \{1,\dots,m\} : B_j \subseteq A_i\} = \emptyset$ for some $i = 1,\dots,n$.

If $n \ge 1$ and $A_1,\dots,A_n$ are distinct elements of $\mathcal{X}$, let $B_1,\dots,B_N$ be a finite measurable partition of $\mathbf{X}$ generated by $A_1,\dots,A_n$; that is a class of sets obtained by taking all the possible intersections of the sets A_i and their complements, $i = 1,\dots,n$. By considering $\phi_{A_1,\dots,A_n;B_1,\dots,B_N}$ as a random vector which maps $[0,1]^N$ into $[0,1]^n$, the probability measure $q_{A_1,\dots,A_n}$ is now defined as the probability distribution of $\phi_{A_1,\dots,A_n;B_1,\dots,B_N}$ when $([0,1]^N, \mathcal{B}^N[0,1])$ is endowed with the probability measure $q_{B_1,\dots,B_N}$. In other words, for all $C \in \mathcal{B}^n[0,1]$, we define

$$q_{A_1,\dots,A_n}(C) = q_{B_1,\dots,B_N}(\phi^{-1}_{A_1,\dots,A_n;B_1,\dots,B_N}(C)). \tag{3.20}$$

Hence, by means of the class $\mathbf{Q}'$ whose elements satisfy $(C_1')-(C_4')$ we generated a class $\tilde{\mathbf{Q}}$ whose elements, defined by (3.20), are probability distributions indexed by finite sequences of distinct elements of $\mathcal{X}$.

3.21 **Lemma**. *The elements of the class* $\tilde{\mathbf{Q}}$ *satisfy consistency conditions* $(C_1)-(C_4)$.

Proof. In order to prove that (C_1) is satisfied consider, for example, a measurable partition B_1, B_2, B_3, B_4 of $\mathbf{X}$ and let $A_1 = B_1 \bigcup B_2$ and $A_2 = B_1 \bigcup B_3$; note that $B_1,\dots,B_4$ is a partition generated by A_1, A_2. Then, for any rectangle $C_1 \times C_2$, with $C_i \in \mathcal{B}[0,1]$ for $i = 1,2$,

$$\begin{aligned} q_{A_1,A_2}(C_1 \times C_2) &= q_{B_1,\dots,B_4}(\phi^{-1}_{A_1,A_2;B_1,\dots,B_4}(C_1 \times C_2)) \\ &= q_{B_1,\dots,B_4}(\phi^{-1}_{A_2,A_1;B_1,\dots,B_4}(C_2 \times C_1)) = q_{A_2,A_1}(C_2 \times C_1) \end{aligned}$$

since $\phi^{-1}_{A_1,A_2;B_1,\ldots,B_4}(C_1 \times C_2) = \phi^{-1}_{A_2,A_1;B_1,\ldots,B_4}(C_2 \times C_1)$. Following a natural extension of the previous argument, one proves in complete generality that the elements of $\tilde{\mathbf{Q}}$ satisfy (C_1).

Property (C_2) is the same as (C_2')

For proving that (C_3) is satisfied, let $n \geq 1$ and $A_1, \ldots, A_n$ be distinct elements of $\mathcal{X}$. If $D_1, \ldots, D_k$ is a measurable partition of $\mathbf{X}$ such that $A_i = \bigcup_{(i)} D_j$, for $i = 1, \ldots, n$, then a measurable partition $B_1, \ldots, B_m$ generated by $A_1, \ldots, A_n$ is not finer than the partition $D_1, \ldots, D_k$. We need to prove that, for all $C \in \mathcal{B}^n[0,1]$,

$$\begin{aligned} &q_{A_1,\ldots,A_n}(C) \\ &\quad = q_{D_1,\ldots,D_k}(\{(x_1,\ldots,x_k) \in [0,1]^k : (\sum_{(1)} x_j, \ldots, \sum_{(n)} x_j) \in C\}) \\ &\quad = q_{D_1,\ldots,D_k}(\phi^{-1}_{A_1,\ldots,A_n;D_1,\ldots,D_k}(C)). \end{aligned} \tag{3.22}$$

However, by definition,

$$q_{A_1,\ldots,A_n}(C) = q_{B_1,\ldots,B_m}(\phi^{-1}_{A_1,\ldots,A_n;B_1,\ldots,B_m}(C))$$

and from property (C_3') it follows that

$$q_{B_1,\ldots,B_m}(E) = q_{D_1,\ldots,D_k}(\phi^{-1}_{B_1,\ldots,B_m;D_1,\ldots,D_k}(E))$$

for all $E \in \mathcal{B}^m[0,1]$. Hence, (3.22) follows after checking that

$$\phi^{-1}_{A_1,\ldots,A_n;D_1,\ldots,D_k}(C) = \phi^{-1}_{B_1,\ldots,B_m;D_1,\ldots,D_k}(\phi^{-1}_{A_1,\ldots,A_n;B_1,\ldots,B_{m+1}}(C)).$$

Finally, (C_4) is satisfied if we show that, for every sequence $\{A_n\}$ of elements of $\mathcal{X}$ decreasing to the empty set,

$$\lim_{n\to\infty} q_{A_n}([0,a]) = \delta_0([0,a]) = 1$$

for all $a \neq 0$. However, by (3.20),

$$q_{A_n}([0,a]) = q_{A_n,A_n^c}(\{(y_1,y_2) \in [0,1]^2 : y_1 \leq a\})$$

so that (C_4) follows from (C_4'). ◇

The previous Lemma and Theorem 3.2 show that the class $\mathbf{Q}'$ of probability distributions indexed by the finite measurable partitions of $\mathbf{X}$ define a probability measure ν on $(\mathbf{P}, \mathcal{P})$ when $(C_1') - (C_4')$ are satisfied; it is a simple exercise to show along the lines of Lemma 3.19 that such a ν is uniquely defined by the elements of $\mathbf{Q}'$.

4 The Dirichlet process

As a first example of a probability measure on $(\mathbf{P}, \mathcal{P})$ constructed through the approach described in the previous section, we consider the celebrated Dirichlet process introduced by Ferguson (1973) as a flexible prior for Bayesian nonparametrics .

We begin with the definition of the Dirichlet distribution. For $n \geq 2$ and $i = 1, \ldots, n$, let $\alpha_i \geq 0$ and such that $\sum_{i=1}^n \alpha_i > 0$, and set $Z_1, \ldots, Z_n$ to be independent real random variables with distribution Gamma$(\alpha_1, 1)$,...,Gamma$(\alpha_n, 1)$, respectively, where we assume by convention that Gamma$(0, 1)$ is the point mass at 0.

4.1 **Definition**. *The probability distribution on* $([0,1]^n, \mathcal{B}^n[0,1])$ *of the random vector*

$$(\frac{Z_1}{\sum_{i=1}^n Z_n}, \ldots, \frac{Z_n}{\sum_{i=1}^n Z_n})$$

is said to be Dirichlet with parameters $(\alpha_1, \ldots, \alpha_n)$ *and denoted by* $\mathcal{D}[\cdot; \alpha_1, \ldots, \alpha_n]$.

Note that $\mathcal{D}[\cdot; \alpha_1, \ldots, \alpha_n]$ is singular with respect to the Lebesgue measure on $([0,1]^n, \mathcal{B}^n[0,1])$; however, if $\alpha_i > 0$ for all $i = 1, \ldots, n$, then the probability distribution induced by $\mathcal{D}[\cdot|\alpha_1, \ldots, \alpha_n]$ on $([0,1]^{n-1}, \mathcal{B}^{n-1}[0,1])$ via the orthogonal projection which drops the kth coordinate, with $k \in \{1, \ldots, n\}$, is absolutely continuous with respect to the Lebesgue measure and has density

$$\begin{aligned} &d_{\alpha_1,\ldots,\alpha_n}(y_1, \ldots, y_{k-1}, y_{k+1}, \ldots, y_n) \\ &= \frac{\Gamma(\sum_{i=1}^n \alpha_i)}{\prod_{i=1}^n \Gamma(\alpha_i)}[\prod_{\{i \in \{1,\ldots,n\}: i \neq k\}} y_i^{\alpha_i - 1}][1 - \sum_{\{i \in \{1,\ldots,n\}: i \neq k\}} y_i]^{\alpha_k - 1} I_{S_{n-1}}((y_1, \ldots, y_{k-1}, y_{k+1}, \ldots, y_n)) \end{aligned}$$

where

$$S_{n-1} = \{(t_1, \ldots, t_{n-1}) \in [0,1]^{n-1} : t_i > 0 \text{ for } i = 1, \ldots, n-1 \text{ and } \sum_{i=1}^{n-1} t_i < 1\}$$

is the $n-1$-dimensional simplex. For $n = 2$ and $k = 2$ this is the density of a Beta distribution with parameters (α_1, α_2).

Let now α be a finite measure defined on $(\mathbf{X}, \mathcal{X})$. For all finite partitions $B_1, \ldots, B_n$ of $\mathbf{X}$ such that $n \geq 2$ and $B_i \in \mathcal{X}$, for $i = 1, \ldots, n$, that is for all finite measurable partitions of $\mathbf{X}$, define a probability measure $q_{B_1,\ldots,B_n}$ on the Borel σ-field $\mathcal{B}^n[0,1]$ of $[0,1]^n$ by setting

$$q_{B_1,\ldots,B_n}(C) = \mathcal{D}[C; \alpha(B_1), \ldots, \alpha(B_n)]. \tag{4.2}$$

for all $C \in \mathcal{B}^n[0,1]$. Moreover set,

$$q_{\mathbf{X}}(C) = \delta_1(C) \tag{4.3}$$

for all $C \in \mathcal{B}^n([0,1])$. Let $\mathbf{Q}'_{\mathcal{D}}$ be the class of probability distributions defined by (4.2) and (4.3).

4.4 **Theorem**. *The elements of $\mathbf{Q}'_{\mathcal{D}}$ satisfy $(C'_1) - (C'_4)$.*

Proof. (C'_1) and (C'_2) are obvious. (C'_3) follows from the well known additive property of the Dirichlet distribution. Finally, (C'_4) is satisfied if we show that, for every sequence $\{A_n\}$ of elements of $\mathcal{X}$ decreasing to the empty set,

$$\lim_{n\to\infty} q_{A_n,A_n^c}(\{(y_1,y_2) \in [0,1]^2 : y_1 \leq a) = \delta_0([0,a]) = 1 \tag{4.5}$$

for all $a \neq 0$. However, if $\alpha(A_N) = 0$ for a given $N \geq 1$, then $q_{A_n,A_n^c}([0,a] \times [0,1]) = 1 = \delta_0([0,a])$ for all $n \geq N$, so that (4.5) follows trivially. When $\alpha(A_n) > 0$ for all n, notice that the distribution induced by q_{A_n,A_n^c} on $([0,1], \mathcal{B}[0,1])$ via the projection which drops the last coordinate is the Beta distribution with parameters $(\alpha(A_n), \alpha(A_n^c))$: therefore it follows from Markov inequality that,

$$q_{A_n,A_n^c}([0,a] \times [0,1]) \geq 1 - \frac{\alpha(A_n)}{x\alpha(\mathbf{X})}$$

for all $a \neq 0$ and all n. This implies (4.5) since $\lim_{n\to\infty} \alpha(A_n) = 0$. ◇

4.6 **Definition**. *Given a finite measure α on $(\mathbf{X}, \mathcal{X})$, the unique probability measure determined by the class $\mathbf{Q}'_{\mathcal{D}}$ on the space $(\mathbf{P}, \mathcal{P})$ is called Dirichlet process with parameter α and denoted by $\mathcal{D}_\alpha$.*

4.7 **Example**. Let us consider the case where $\mathbf{X} = \Re$. If $-\infty < x_1 < \ldots < x_n < \infty$ are n real numbers and

$$A_1 = (-\infty, x_1], A_2 = (x_1, x_2], \ldots, A_n = (x_{n-1}, x_n],$$

set

$$F(x_1) = \wp_{A_1}, F(x_2) = \wp_{A_1} + \wp_{A_2}, \ldots, F(x_n) = \sum_{i=1}^{n} \wp_{A_i}.$$

Then, for $0 \leq y_1 \leq \ldots \leq y_n \leq 1$,

$$\begin{aligned}
&\mathcal{D}_\alpha(F(x_1) \leq y_1, \ldots, F(x_n) \leq y_n) \\
&\quad = \mathcal{D}_\alpha(\{p \in \mathbf{P} : p(A_1) \leq y_1, \ldots, \sum_{i=1}^{n} p(A_i) \leq y_n\}) \\
&\quad = \frac{\Gamma(\alpha(\Re))}{\Gamma(\alpha((-\infty, x_1]))\Gamma(\alpha((x_1, x_2]))\cdots\Gamma(\alpha((x_n, +\infty)))} \cdot \\
&\quad\quad \cdot \int_{\{(z_1,\ldots,z_n)\in[0,1]^n : z_1 \leq \ldots \leq z_n, z_1 \leq y_1, \ldots, z_n \leq y_n\}} z_1^{\alpha((-\infty,x_1])-1} (z_1 - z_2)^{\alpha((x_1,x_2])-1} \cdots \\
&\quad\quad\quad \cdots (z_{n-1} - z_n)^{\alpha((x_{n-1},x_n])-1} (1 - z_n)^{\alpha((x_n,\infty))-1} dz_1 \cdots dz_n.
\end{aligned}$$

In particular, for $x, y \in [0,1]$,

$$\mathcal{D}_\alpha(F(x) \le y) = \frac{\Gamma(\alpha(\mathbf{X}))}{\Gamma(\alpha((-\infty,x])\Gamma(\alpha((x,+\infty)))} \int_0^y z^{\alpha((-\infty,x])-1}(1-z)^{\alpha((x,+\infty))-1} dz,$$

that is $F(x)$ has Beta distribution with parameters $(\alpha((-\infty,x]), \alpha((x,+\infty))$. Therefore, for instance,

$$E(F(x)) = \frac{\alpha((-\infty,x])}{\alpha(\mathbf{X})}.$$

$\diamond$

In the rest of the section we want to present a few basic properties of the Dirichlet process which make it an appealing prior distribution for Bayesian nonparametric statistics. We therefore consider an infinite exchangeable sequence of random variables $\{X_n\}$ with values in $\mathbf{X}$ and we assume that, given $P = p \in \mathbf{P}$, they are independent and identically distributed with probability distribution p, where P is a random element of $(\mathbf{P}, \mathcal{P})$ with probability distribution $\mathcal{D}_\alpha$. It is well to point out that, according to our definition, P is the only random probability measure for which the distribution of $(P(B_1), \ldots, P(B_n))$ is Dirichlet with parameters $(\alpha(B_1), \ldots, \alpha(B_n))$ for every measurable partition $B_1, \ldots, B_n$ of $\mathbf{X}$. In particular, if $B \in \mathcal{X}$ and $\alpha(B) > 0$, $P(B)$ has Beta distribution with parameters $(\alpha(B), \alpha(\mathbf{X}) - \alpha(B))$.

As we showed in Section 6.2, the right framework where to examine inferential questions regarding the sequence $\{X_n\}$ and the random probability distribution P is the statistical model $(\mathbf{X}^\infty \times \mathbf{P}, \mathcal{X}^\infty \times \mathcal{P}, \pi)$ where π is defined as in (2.1) with $\nu = \mathcal{D}_\alpha$.

Let us begin by proving that the marginal probability distribution of X_1 is the measure α normalized.

4.8 **Theorem**. *For all* $A \in \mathcal{X}$,

$$\tau_1(A) = \frac{\alpha(A)}{\alpha(\mathbf{X})}.$$

Proof. Fix $A \in \mathcal{X}$. Then,

$$\tau_1(A) = \pi(X_1 \in A) = \int_{\mathbf{P}} p(A)\mathcal{D}_\alpha(dp) = E[P(A)] = \frac{\alpha(A)}{\alpha(\mathbf{X})}.$$

$\diamond$

For all $n \ge 1$ and $(x_1, \ldots, x_n) \in \mathbf{X}^n$ we can easily compute the posterior distribution $\mathcal{D}_\alpha(\cdot | X_1 = x_1, \ldots, X_n = x_n)$ whose existence was established in Section 6.2. Recall that, for $x \in \mathbf{X}$, δ_x indicates the point mass at x.

4.9 **Theorem**. *For all $n \geq 1$ and $(x_1, \ldots, x_n) \in \mathbf{X}^n$,*

$$\mathcal{D}_\alpha(\cdot|X_1 = x_1, \ldots, X_n = x_n) = \mathcal{D}_{\alpha + \sum_{i=1}^n \delta_{x_i}}.$$

Proof. We consider only the case $n = 1$; the general case then follows by induction on n.

Let $x_1 \in \mathbf{X}$. In order to prove that $\mathcal{D}_\alpha(\cdot|X_1 = x_1)$ is $\mathcal{D}_{\alpha+\delta_{x_1}}$, it is enough to show that, for all $k \geq 2$ and all finite measurable partitions $B_1, \ldots, B_k$, the probability distribution of the random vector$(\wp_{B_1}, \ldots, \wp_{B_k})$ induced by $\mathcal{D}_\alpha(\cdot|X_1 = x_1)$ on $([0,1]^k, \mathcal{B}^k[0,1])$ is Dirichlet with parameters $(\alpha(B_1) + \delta_{x_1}(B_1), \ldots, \alpha(B_k) + \delta_{x_1}(B_k))$. This is equivalent to show that the conditional probability distribution of $(\wp_{B_1}(P), \ldots, \wp_{B_k}(P))$ given $X_1 = x_1$, is Dirichlet with parameters $(\alpha(B_1) + \delta_{x_1}(B_1), \ldots, \alpha(B_k) + \delta_{x_1}(B_k))$. In fact, for all $C \in \mathcal{B}^k[0,1]$,

$$\begin{aligned}
&\mathcal{D}_\alpha[(\wp_{B_1}, \ldots, \wp_{B_k}) \in C|X_1 = x_1] \\
&\quad = \pi[\mathbf{X}^\infty \times \{p \in \mathbf{P} : (\wp_{B_1}(p), \ldots, \wp_{B_k}(p)) \in C\}|X_1 = x_1] \\
&\quad = \pi[(\wp_{B_1}(P), \ldots, \wp_{B_k}(P)) \in C|X_1 = x_1]
\end{aligned}$$

where the first equality follows from the definition of $\mathcal{D}_\alpha(\cdot|X_1 = x_1)$ whereas the next equality follows from definition of P.

Let $k \geq 2$, $B_1, \ldots, B_k$ be a finite measurable partition of $\mathbf{X}$ and $A \in \mathcal{X}$. Set $B_s^0 = A \bigcap B_s$ and $B_s^1 = A^c \bigcap B_s$, for $s = 1, \ldots, k$. Then, for all $(t_1, \ldots, t_k) \in [0,1]^k$,

$$\pi[\wp_{B_1}(P) \leq t_1, \ldots, \wp_{B_k}(P) \leq t_k, X_1 \in A] = \sum_{s=1}^k \pi[\wp_{B_1}(P) \leq t_1, \ldots, \wp_{B_k}(P) \leq t_k, X_1 \in B_s^0] \tag{4.10}$$

For $s = 1, \ldots, k$, consider

$$\begin{aligned}
&\pi[\wp_{B_1}(P) \leq t_1, \ldots, \wp_{B_k}(P) \leq t_k, X_1 \in B_s^0] \qquad (4.11)\\
&\quad = \pi[\wp_{B_1^0}(P) + \wp_{B_1^1}(P) \leq t_1, \ldots, \wp_{B_k^0}(P) + \wp_{B_k^1}(P) \leq t_k, X_1 \in B_s^0] \\
&\quad = \int_{\mathbf{X}^\infty \times \{p \in \mathbf{P} : \wp_{B_1^0}(p) + \wp_{B_1^1}(p) \leq t_1, \ldots, \wp_{B_k^0}(p) + \wp_{B_k^1}(p) \leq t_k\}} \\
&\quad \pi[X_1 \in B_s^0|\wp_{B_1^0}(P), \ldots, \wp_{B_k^0}(P), \wp_{B_1^1}(P), \ldots, \wp_{B_k^1}(P)]d\pi \\
&\quad = \int_{\{p \in \mathbf{P} : \wp_{B_1^0}(p) + \wp_{B_1^1}(p) \leq t_1, \ldots, \wp_{B_k^0}(p) + \wp_{B_k^1}(p) \leq t_k\}} p(B_s^0)\mathcal{D}_\alpha(dp)
\end{aligned}$$

where the first equality follows from the fact that $\wp_{B_1} = \wp_{B_1^0} + \wp_{B_1^1}, \ldots, \wp_{B_k} = \wp_{B_k^0} + \wp_{B_k^1}$, whereas the last one is true because, for all $p \in \mathbf{P}$ the conditional probability distribution of

X_1 given $P = p$ is p; therefore

$$\begin{aligned}
&\pi[X_1 \in B_s^0 | \wp_{B_1^0}(P), \ldots, \wp_{B_k^0}(P), \wp_{B_1^1}(P), \ldots, \wp_{B_k^1}(P)] \\
&\quad = E[\pi[X_1 \in B_s^0 | P] | \wp_{B_1^0}(P), \ldots, \wp_{B_k^0}(P), \wp_{B_1^1}(P), \ldots, \wp_{B_k^1}(P)] \\
&\quad = E[\wp_{B_s^0}(P) | \wp_{B_1^0}(P), \ldots, \wp_{B_k^0}(P), \wp_{B_1^1}(P), \ldots, \wp_{B_k^1}(P)] \\
&\quad = \wp_{B_s^0}(P)
\end{aligned}$$

almost surely with respect to π.

If $\alpha(B_s^0) = 0$, $\mathcal{D}_\alpha[\{p \in \mathbf{P} : p(B_s^0) = 0\}] = 1$ so that the last quantity in (4.11) is equal to zero.

If $\alpha(B_s^0) = \alpha(\mathbf{X})$, $\mathcal{D}_\alpha[\{p \in \mathbf{P} : p(B_s^0) = 1\}] = 1$ and in this case the last quantity in (4.11) is equal to $\delta_1([0, t_s])$.

If $0 < \alpha(B_s^0) < \alpha(\mathbf{X})$, set $\beta_i = \alpha(B_i^0)$, if $i = 1, \ldots, k$, $\beta_i = \alpha(B_{i-k}^1)$, if $i = k+1, \ldots, 2k$, and $C = \{i \in \{1, \ldots, 2k\} : \beta_i \neq 0\}$. Since $\alpha(B_s^0) < \alpha(\mathbf{X})$ there is a $j \neq s$ such that $\beta_j \neq 0 :$ let $C^* = \{i \in C : i \neq j\}$. Finally set $\beta_i^s = \beta_s + 1$ if $i = s$ and $\beta_i^s = \beta_i$ if $i \in \{1, \ldots, 2k\}, i \neq s$. Then,

$$\begin{aligned}
&\int_{\{p \in \mathbf{P} : \wp_{B_1^0}(p) + \wp_{B_1^1}(p) \leq t_1, \ldots, \wp_{B_k^0}(p) + \wp_{B_k^1}(p) \leq t_k\}} p(B_s^0) \mathcal{D}_\alpha(dp) \\
&\quad = \int_E z_s \frac{\Gamma(\alpha(\mathbf{X}))}{\prod_{i \in C} \Gamma(\beta_i)} [\prod_{i \in C^*} z_i^{\beta_i - 1}] (1 - \sum_{i \in C^*} z_i)^{\beta_j - 1} \prod_{i \in C^*} dz_i \\
&\quad = \frac{\beta_s}{\alpha(\mathbf{X})} \int_E \frac{\Gamma(\alpha(\mathbf{X}) + 1)}{\prod_{i \in C} \Gamma(\beta_i^s)} [\prod_{i \in C^*} z_i^{\beta_i^s - 1}] (1 - \sum_{i \in C^*} z_i)^{\beta_j - 1} \prod_{i \in C^*} dz_i \\
&\quad = \frac{\beta_s}{\alpha(\mathbf{X})} D[\{(z_1, \ldots, z_{2k}) \in [0,1]^{2k} : z_i + z_{i+k} \leq t_i, \text{ for } i \in \{1, \ldots, k\}\}; \beta_1^s, \ldots, \beta_{2k}^s] \\
&\quad = \frac{\alpha(B_s^0)}{\alpha(\mathbf{X})} D[\{(z_1, \ldots, z_k) \in [0,1]^k : z_1 \leq t_1, \ldots, z_k \leq t_k\}; \alpha(B_1), \ldots, \alpha(B_s) + 1, \ldots, \alpha(B_k)]
\end{aligned}$$

where

$$\begin{aligned}
E = \{&(z_1, \ldots, z_{2k}) \in [0,1]^{2k} : \\
&z_l = 0 \text{ if } \beta_l = 0, \text{ for } l \in \{1, \ldots, 2k\}, \text{ and } z_i + z_{i+k} \leq t_i, \text{ for } i \in \{1, \ldots, k\}\}.
\end{aligned}$$

Note that the last equality follows from the well known additive property of the Dirichlet distribution.

Therefore, whatever the value of $\alpha(B_s^0)$, for $s = 1, \ldots, k$,

$$\begin{aligned}
&\pi[\wp_{B_1}(P) \leq t_1, \ldots, \wp_{B_k} \leq t_k, X_{1 \in B_s^0}] \\
&\quad = \frac{\alpha(B_s^0)}{\alpha(\mathbf{X})} D[\{(z_1, \ldots, z_k) \in [0,1]^k : z_1 \leq t_1, \ldots, z_k \leq t_k\}; \alpha(B_1), \ldots, \alpha(B_s) + 1, \ldots, \alpha(B_k)];
\end{aligned}$$

this and (4.10) imply that

$$\pi[\wp_{B_1} \le t_1, \dots, \wp_{B_k} \le t_k, X_1 \in A]$$
$$= \sum_{s=1}^{k} \frac{\alpha(B_s^0)}{\alpha(\mathbf{X})} D[\{(z_1, \dots, z_k) \in [0,1]^k : z_1 \le t_1, \dots, z_k \le t_k\}; \alpha^s(B_1), \dots, \alpha^s(B_k)]$$

where $\alpha^s(B_i) = \alpha(B_s) + 1$ if $s = i$ whereas $\alpha^s(B_i) = \alpha(B_i)$ if $s \ne i$.

However, for all $s = 1, \dots, k$,

$$\frac{\alpha(B_s^0)}{\alpha(\mathbf{X})} = \frac{1}{\alpha(\mathbf{X})} \int_{A \bigcap B_s} \alpha(dx)$$

so that

$$\pi[\wp_{B_1}(P) \le t_1, \dots, \wp_{B_k}(P) \le t_k, X_1 \in A] \tag{4.12}$$
$$= \sum_{s=1}^{k} D[\{[0,t_1] \times \cdots \times [0,t_k]; \alpha^s(B_1), \dots, \alpha^s(B_k)] \frac{1}{\alpha(\mathbf{X})} \int_{A \bigcap B_s} \alpha(dx)$$
$$= \sum_{s=1}^{k} \frac{1}{\alpha(\mathbf{X})} \int_{A \bigcap B_s} D[[0,t_1] \times \cdots \times [0,t_k]; \alpha(B_1) + \delta_x(B_1), \dots, \alpha(B_k) + \delta_x(B_k)] \alpha(dx)$$
$$= \frac{1}{\alpha(\mathbf{X})} \int_A D[[0,t_1] \times \cdots \times [0,t_k]; \alpha(B_1) + \delta_x(B_1), \dots, \alpha(B_k) + \delta_x(B_k)] \alpha(dx).$$

Being (4.12) true for all $A \in \mathcal{X}$ and all $(t_1, \dots, t_k) \in [0,1]^k$ this shows that the conditional distribution of $(\wp_{B_1}(P), \dots, \wp_{B_k}(P))$ given $X_1 = x$ is Dirichlet with parameters $(\alpha(B_1) + \delta_x(B_1), \dots, \alpha(B_k) + \delta_x(B_k))$ since the probability distribution of X_1 is the measure α normalized, as follows from Theorem 4.8. Therefore $\mathcal{D}_\alpha(\cdot | X_1 = x)$ is $\mathcal{D}_{\alpha+\delta_x}$. ◇

The next result shows that if the random probability P has probability distribution $\mathcal{D}_\alpha$, then P is discrete almost surely. Before stating the theorem, note that for any given $p \in \mathbf{P}$ the set $\{x \in \mathbf{X} : p(\{x\}) > 0\} \in \mathcal{X}$ since $\mathbf{X}$ is separable; moreover p is discrete if and only if $p(\{x \in \mathbf{X} : p(\{x\}) > 0\}) = 1$. One can in fact prove that the set $\{p \in \mathbf{P} : p(\{x \in \mathbf{X} : p(\{x\}) > 0\}) = 1\}$ of all discrete probability distributions on $(\mathbf{X}, \mathcal{X})$ is an element of $\mathcal{P}$.

4.13 **Theorem**.

$$\mathcal{D}_\alpha(\{p \in \mathbf{P} : p(\{x \in \mathbf{X} : p(\{x\}) > 0\}) = 1\}) = 1. \tag{4.14}$$

Proof. Let us consider the function $P(\{X_1\}) : \mathbf{X}^\infty \times \mathbf{P} \to [0,1]$ which maps $((x_1, x_2, \dots), p) \in \mathbf{X}^\infty \times \mathbf{P}$ in $p(\{x_1\})$. It is a matter of technicalities to show that $P(\{X_1\})$ is measurable; therefore we may compute $\pi(P(\{X_1\}) > 0)$.

In fact,

$$\pi(P(\{X_1\}) > 0) = \frac{1}{\alpha(\mathbf{X})} \int_{\mathbf{X}} \pi(P(\{x_1\}) > 0 | X_1 = x_1)\alpha(dx_1) \tag{4.15}$$

since the marginal probability distribution of X_1 is the measure α normalized. Given $X_1 = x_1$, the conditional probability distribution of P is $D_{\alpha+\delta_{x_1}}$ and therefore $P(\{x_1\})$ has distribution Beta with parameters $(\alpha(\{x_1\}) + 1, \alpha(\mathbf{X}) - \alpha(\{x_1\})$; hence $P(\{x_1\})$ is strictly positive with probability one. This shows that $\pi(P(\{X_1\}) > 0) = 1$.

However,

$$\pi(P(\{X_1\}) > 0) = \int_{\mathbf{P}} \pi(p(\{X_1\}) > 0 | P = p)\mathcal{D}_\alpha(dp). \tag{4.16}$$

But, given $P = p$, the probability distribution of X_1 is p so that

$$\pi(p(\{X_1\}) > 0 | P = p) = p(\{x \in \mathbf{X} : p(\{x\}) > 0\}). \tag{4.17}$$

Equations (4.15), (4.16) and (4.17) imply (4.14.) ◇

The fact that P is discrete with probability one when P has probability distribution $\mathcal{D}_\alpha$ is sometimes considered as a drawback to the use of the Dirichlet process for applications. As such, this is an ill-conceived criticism. Notice, for instance, that any continuous probability distribution on the real line gives probability one to the set of irrational numbers; but this does not stop applied statisticians from using continuous distributions for the analysis of rational data. In point of fact Ferguson (1973) remarks that the support of $\mathcal{D}_\alpha$ is the set of all probabilities on $(\mathbf{X}, \mathcal{X})$ whose support is contained in the support of α; a lot of probabilities that are not discrete may well belong to this set. In any case, in the next section we will introduce a probability distribution on $(\mathbf{P}, \mathcal{P})$ which may select with probability one continuous probabilities on $(\mathbf{X}, \mathcal{X})$.

5 Tailfree processes

The Dirichlet process considered in the previous section belongs to a very general class of probability distributions defined on $(\mathbf{P}, \mathcal{P})$ which were named *tailfree processes* by Fabius (1964); they were extensively studied by Freedman (1963), Ferguson (1973) and Doksum (1974) among others. The aim of this section is to construct tailfree processes as probabilities on $(\mathbf{P}, \mathcal{P})$ using the method described in Section 3. In addition to the Dirichlet process, as a special instance of a tailfree process we will also get the *Polya tree process* introduced by Mauldin, Sudderth and Williams (1992) and Lavine (1992,1994) and studied in the next section.

Let $\{\nabla_m\}$ be a sequence of finite measurable partitions of $\mathbf{X}$, that is finite partitions whose elements belong to $\mathcal{X}$. In the rest of the section, for all $m \geq 1$, we will denote with $\{B_{m,1}, \ldots, B_{m,k_m}\}$ the partition ∇_m; however the order of appearance of the sets $B_{m,j}$ is not going to matter.

5.1 **Definition**. *A sequence $\{\nabla_m\}$ of finite measurable partitions of $\mathbf{X}$ is called a tree of partitions of $(\mathbf{X}, \mathcal{X})$ if:*

(i) $\nabla_0 = \{\mathbf{X}\}$;

(ii) for all $m \geq 1$, ∇_m is finer than ∇_{m-1};

(iii) the σ-field generated by $\bigcup_m \nabla_m$ is $\mathcal{X}$.

Before proceeding any further we need a few useful notations.

If $B \in \nabla_m$ and $n \leq m$, we call *predecessor of B at level n* the unique set $ps(B, n) \in \nabla_n$ such that $B \subseteq ps(B, n)$. Note that $ps(B, m) = B$; moreover if $C \in \nabla_n, D \in \nabla_k$ with $k \leq n \leq m$, and $ps(B, n) = C$ while $ps(C, k) = D$, then $ps(B, k) = D$. That is:

$$ps(B, k) = ps(ps(B, n), k). \tag{5.2}$$

For $A \in \mathcal{X}$ and $m \geq 1$, we set

$$\mathcal{E}_m(A) = \{B \in \nabla_m : B \bigcap A \neq \emptyset\}.$$

Finally, for all $x \in \mathbf{X}$ and $m \geq 0$ let $\xi_m(x)$ be the unique set in ∇_m which contains x; clearly $ps(\xi_{m+1}(x), m) = \xi_m(x)$.

Tree of partitions exist if and only if $\mathcal{X}$ is countably generated. Moreover, the following lemma describes how to represent the elements of $\mathcal{X}$ when $\{\nabla_n\}$ is a tree of partitions of $(\mathbf{X}, \mathcal{X})$.

5.3 **Lemma**. *Let $\{\nabla_n\}$ be a tree of partitions of $(\mathbf{X}, \mathcal{X})$. Then, for each $A \in \mathcal{X}$,*

$$A = \bigcap_n \bigcup_{B \in \mathcal{E}_n(A)} B.$$

Proof. Let

$$\mathcal{F} = \{A \in \mathcal{X} : A = \bigcap_n \bigcup_{B \in \mathcal{E}_n(A)} B\}.$$

We show that $\mathcal{F}$ is a monotone class containing the field $\mathcal{G}$ whose elements are the empty set and finite unions of sets of $\bigcup_n \nabla_n$; since $\mathcal{G}$ generates $\mathcal{X}$, the Monotone Class Theorem (Ash (1972), Theorem 1.3.9) guarantees that $\mathcal{F} = \mathcal{X}$.

First notice that, for all $A \in \mathcal{X}$, the sequence of sets $\{\bigcup_{B \in \mathcal{E}_n(A)} B\}$ is decreasing with n.

Now, let A be a finite union of elements of $\bigcup_n \nabla_n$; then there is an $m \geq 1$ such that A is a finite union of elements of ∇_m. Therefore, for all $n \geq m$,

$$A = \bigcup_{B \in \mathcal{E}_n(A)} B$$

and this implies that

$$A = \bigcap_n \bigcup_{B \in \mathcal{E}_n(A)} B.$$

Therefore $\mathcal{F}$ contains $\mathcal{G}$.

Now let $\{A_m\}$ be a sequence of elements of $\mathcal{F}$ monotonically increasing to the set A. Then

$$A = \bigcup_m \bigcap_n [\bigcup_{B \in \mathcal{E}_n(A_m)} B] = \bigcap_n [\bigcup_{B \in \mathcal{E}_n(\bigcup_m A_m)} B]$$

and the last equality proves that $A \in \mathcal{F}$. Analogously, if $\{A_m\}$ is a sequence of elements of $\mathcal{F}$ monotonically decreasing to the set A, then

$$A = \bigcap_m \bigcap_n [\bigcup_{B \in \mathcal{E}_n(A_m)} B] = \bigcap_n [\bigcup_{B \in \mathcal{E}_n(\bigcap_m A_m)} B]$$

and the last equality proves that $A \in \mathcal{F}$. Therefore $\mathcal{F}$ is a monotone class. $\diamond$

Given a tree of partitions we may define a probability on $(\mathbf{X}, \mathcal{X})$ by specifying the values of the conditional probability of each element of $\bigcup_m \nabla_m$ given its predecessor in the tree. We give a detailed description of the procedure since along the same idea one can also construct random probability measures on $(\mathbf{X}, \mathcal{X})$, that is probability distributions on $(\mathbf{P}, \mathcal{P})$.

Let $\{v_{m,B} : m \geq 0, B \in \nabla_m\}$ be a double sequence of numbers such that:

(t_1) $v_{m,B} \in [0, 1]$ for all $m \geq 0$ and $B \in \nabla_m$.

(t_2) $v_{0,\mathbf{X}} = 1$.

(t_3) For all $m \geq 0$ and $B \in \nabla_m$,

$$\sum_{C \in \mathcal{E}_{m+1}(B)} v_{m+1,C} = 1.$$

(t_4) For all sequences $\{B_n\}$ decreasing to the empty set and such that, for each $n \geq 0$, B_n is union of sets in ∇_n,

$$\lim_{n\to\infty} \sum_{\{B\in\mathcal{E}_n(B_n)\}} \prod_{i=1}^{n} v_{i,ps(B,i)} = 0$$

For all sets A belonging to the field $\mathcal{G}$ whose elements are the empty set and finite unions of sets of $\bigcup_n \nabla_n$ define

$$p(A) = \inf_n \sum_{B\in\mathcal{E}_n(A)} \prod_{i=1}^{n} v_{i,ps(B,i)}. \tag{5.4}$$

Note that if $A \in \mathcal{G}$ is not empty there is an $m \geq 0$ such that A is a finite union of sets in ∇_m; using (t_3) it thus follows that

$$\sum_{B\in\mathcal{E}_n(A)} \prod_{i=1}^{n} v_{i,ps(B,i)} = \sum_{B\in\mathcal{E}_m(A)} \prod_{i=1}^{m} v_{i,ps(B,i)}$$

for all $n \geq m$. Therefore the infimum appearing on the right side of (5.4) is attained at $n = m$ since the sequence $\{\sum_{B\in\mathcal{E}_n(A)} \prod_{i=1}^{n} v_{i,ps(B,i)}\}$ is decreasing with n.

It is now easy to check that the set function $p : \mathcal{G} \to [0,1]$ defined in (5.4) is a probability on $\mathcal{G}$ which is countably additive because of (t_4). Therefore p can be extended to $\mathcal{X}$ in a unique way by means of Charatheodory Theorem (Ash (1972), Theorem 1.3.10). In particular, because of Lemma 5.3,

$$p(A) = \lim_n \sum_{B\in\mathcal{E}_n(A)} p(B) = \lim_n \sum_{B\in\mathcal{E}_n(A)} \prod_{i=1}^{n} v_{i,ps(B,i)}$$

for all $A \in \mathcal{X}$.

5.5 **Definition**. *A probability distribution $\mathcal{T}$ on $(\mathbf{P}, \mathcal{P})$ is said to be tailfree with respect to a tree of partitions $\{\nabla_n\}$ if there is a sequence of random variables $\{V_{n,B} : n \geq 0, B \in \nabla_n\}$ defined on a probability space $(\Omega, \sigma, \lambda)$ with values in $([0,1], \mathcal{B}[0,1])$ such that the families of random variables $\{V_{1,B} : B \in \nabla_1\}, \{V_{2,B} : B \in \nabla_2\}, ...$ are mutually independent and, for all $m \geq 1$ and $C \in \mathcal{B}[0,1]^{k_m}$,*

$$\mathcal{T}((\wp_{B_{m,1}}, ..., \wp_{B_{m,k_m}}) \in C) = \lambda((\prod_{i=1}^{m} V_{i,ps(B_{m,1},i)}, \ldots, \prod_{i=1}^{m} V_{i,ps(B_{m,k_m},i)}) \in C).$$

If P is a random element of $\mathbf{P}$ with distribution $\mathcal{T}$ and $\mathcal{T}$ is tailfree, this essentially means that the families of random variables

$$\{P(B|ps(B,0)) : B \in \nabla_1\}, \{P(B|ps(B,1)) : B \in \nabla_2\}, ... \tag{5.6}$$

are mutually independent. With the words of Doksum (1974), "when the sets of one partition, say ∇_m, are divided into new sets for the next partition ∇_{m+1}, the relative random probabilities assigned to this new sets are independent of the corresponding relative random probabilities assigned to the sets in other partitions." In fact, if P is such that the families (5.6) are independent, then its distribution $\mathcal{T}$ is tailfree since, for $m \geq 1$ and $B \in \nabla_m$ one can define $V_{B,m} = P(B|ps(B, m-1))$.

5.7 **Example**. A Dirichlet process $\mathcal{D}_\alpha$ on $(\mathbf{P}, \mathcal{P})$ is tailfree with respect to any tree of partitions $\{\nabla_m\}$.

Given a tree of partitions $\{\nabla_m\}$ of $(\mathbf{X}, \mathcal{X})$ we may construct a probability distribution on $(\mathbf{P}, \mathcal{P})$ tailfree with respect to $\{\nabla_m\}$ via Theorem 3.2. In fact, let $\{V_{m,B} : m \geq 0, B \in \nabla_m\}$ be a sequence of random variables defined on a probability space $(\Omega, \sigma, \lambda)$. Assume that, with probability one with respect to the probability λ :

(T_1) $V_{m,B} \in [0,1]$ for all $m \geq 0$ and $B \in \nabla_m$.

(T_2) $V_{0,\mathbf{X}} = 1$.

(T_3) For all $m \geq 0$ and $B \in \nabla_m$,

$$\sum_{C \in \mathcal{E}_{m+1}(B)} V_{m+1,C} = 1.$$

Moreover assume that:

(T_4) For all sequences $\{B_n\}$ decreasing to the empty set and such that, for each $n \geq 0$, B_n is union of sets in ∇_n,

$$\lim_{n\to\infty} \sum_{B \in \mathcal{E}_n(B_n)} \prod_{i=1}^{n} E_\lambda[V_{i,ps(B,i)}] = 0$$

where E_λ denotes the expected value computed according to λ.

(T_5) The collections of random variables

$$\{V_{1,B} : B \in \nabla_1\}, \{V_{2,B} : B \in \nabla_2\}, ...$$

are mutually independent.

For all $m \geq 0$, we define on $([0,1]^{k_m}, \mathcal{B}^{k_m}[0,1])$ the probability distribution

$$q_{B_{m,1},\ldots,B_{m,k_m}}(C) = \lambda[(\prod_{i=1}^{m} V_{i,ps(B_{m,1},i)}, \ldots, \prod_{i=1}^{m} V_{i,ps(B_{m,k_m},i)}) \in C] \tag{5.8}$$

for $C \in \mathcal{B}^{k_m}[0,1]$.

The next lemma shows that an additive property holds for the probability distributions defined in (5.8).

5.9 **Lemma**. *Let $r > m \geq 1$. Then, for all $C \in \mathcal{B}^{k_m}[0,1]$,*

$$q_{B_{m,1},\ldots,B_{m,k_m}}(C) = q_{B_{r,1},\ldots,B_{r,k_r}}(\phi^{-1}_{B_{m,1},\ldots,B_{m,k_m};B_{r,1},\ldots,B_{r,k_r}}(C)). \tag{5.10}$$

Proof. Fix $C \in \mathcal{B}^{k_m}[0,1]$.

Let us first consider the case where $r = m+1$. Note that

$$\begin{aligned}
&q_{B_{m+1,1},\ldots,B_{m+1,k_{m+1}}}(\phi^{-1}_{B_{m,1},\ldots,B_{m,k_m};B_{m+1,1},\ldots,B_{m+1,k_{m+1}}}(C)) \\
&\quad = q_{B_{m+1,1},\ldots,B_{m+1,k_{m+1}}}(\{(x_1,\ldots,x_{k_{m+1}}) \in [0,1]^{k_{m+1}} : (\sum_{(1)} x_j, \ldots, \sum_{(k_m)} x_j) \in C\}) \\
&\quad = \lambda[(\sum_{(1)} \prod_{i=1}^{m+1} V_{i,ps(B_{m+1,j},i)}, \ldots, \sum_{(k_m)} \prod_{i=1}^{m+1} V_{i,ps(B_{m+1,j},i)}) \in C]
\end{aligned}$$

where $(s) = \{j \in \{1,\ldots,k_{m+1}\} : ps(B_{m+1,j}, m) = B_{m,s}\}$ for $s = 1,\ldots,k_m$. The first equality follows from the definition of the map $\phi_{B_{m,1},\ldots,B_{m,k_m};B_{m+1,1},\ldots,B_{m+1,k_{m+1}}}$ whereas the second one holds because of the definition of $q_{B_{m+1,1},\ldots,B_{m+1,k_{m+1}}}$. Now notice that, for $s = 1,\ldots,k_m$,

$$\begin{aligned}
\sum_{(s)} \prod_{i=1}^{m+1} V_{i,ps(B_{m+1,j},i)} &= \sum_{(s)} V_{m+1,ps(B_{m+1,j},m+1)} \prod_{i=1}^{m} V_{i,ps(B_{m+1,j},i)} \\
&= \sum_{(s)} V_{m+1,B_{m+1,j}} \prod_{i=1}^{m} V_{i,ps(B_{m+1,j},i)} \\
&= [\prod_{i=1}^{m} V_{i,ps(B_{m,s},i)}] \sum_{(s)} V_{m+1,B_{m+1,j}} \\
&= \prod_{i=1}^{m} V_{i,ps(B_{m,s},i)}
\end{aligned}$$

where the second equality holds because $ps(B_{m+1,j}, m+1) = B_{m+1,j}$, the next one is true since $ps(B_{m+1,j}, i) = ps(B_{m,s}, i)$ for $i \leq m$ as follows from (5.2) and, finally, the last equality is implied by (T_4). Therefore

$$\begin{aligned}
&q_{B_{m+1,1},\ldots,B_{m+1,k_{m+1}}}(\phi^{-1}_{B_{m,1},\ldots,B_{m,k_m};B_{m+1,1},\ldots,B_{m+1,k_{m+1}}}(C)) \\
&\quad = \lambda[(\prod_{i=1}^{m} V_{i,ps(B_{m,1},i)}, \ldots, \prod_{i=1}^{m} V_{i,ps(B_{m,k_m},i)}) \in C] \\
&\quad = q_{B_{m,1},\ldots,B_{m,k_m}}(C).
\end{aligned}$$

This proves the lemma when $r = m+1$; for $r > m+1$, the lemma follows by induction on $r-m$ after checking that

$$\phi^{-1}_{\nabla_m;\nabla_r}(C) = \phi^{-1}_{\nabla_{r-1};\nabla_r}(\phi^{-1}_{\nabla_m;\nabla_{r-1}}(C)).$$

◇

We are now ready to define the elements of the class

$$\mathbf{Q}_{\mathcal{T}} = \{q_{A_1,\ldots,A_n} : n \geq 1 \text{ and } A_1,\ldots,A_n \text{ distinct elements of } \mathcal{X}\}$$

which will determine, via Theorem 3.2, a tailfree probability distribution on $(\mathbf{P}, \mathcal{P})$.

Notice that, for all $A \in \mathcal{X}$, the sequence $\{\sum_{B\in\mathcal{E}_n(A)} \prod_{i=1}^n V_{i,ps(B,i)}\}$ is bounded between 0 and 1 and monotonically decreasing with n. In fact for all $n \geq 0$

$$\sum_{B\in\mathcal{E}_{n+1}(A)} \prod_{i=1}^{n+1} V_{i,ps(B,i)} \leq \sum_{B\in\nabla_{n+1}:ps(B,n)\in\mathcal{E}_n(A)} \prod_{i=1}^{n+1} V_{i,ps(B,i)} = \sum_{B\in\mathcal{E}_n(A)} \prod_{i=1}^{n} V_{i,ps(B,i)}$$

where the last equality follows from (T_3). Therefore the limit of the sequence

$$\{\sum_{B\in\mathcal{E}_n(A)} \prod_{i=1}^{n} V_{i,ps(B,i)}\}$$

exists on a set with λ probability one.

For all $n \geq 1$ and $A_1,\ldots,A_n$ distinct elements of $\mathcal{X}$, we define

$$q_{A_1,\ldots,A_n}(C) = \lambda((\lim_m \sum_{B\in\mathcal{E}_m(A_1)} \prod_{i=1}^{m} V_{i,ps(B,i)}, \ldots, \lim_m \sum_{B\in\mathcal{E}_m(A_n)} \prod_{i=1}^{m} V_{i,ps(B,i)}) \in C) \tag{5.11}$$

for $C \in \mathcal{B}^n[0,1]$. It is again a consequence of (T_3) that, when $A_1,\ldots,A_n$ belong to a partition of the tree $\{\nabla_m\}$, the previous definition is consistent with (5.8).

5.12 **Theorem**. *The elements of* $\mathbf{Q}_{\mathcal{T}}$ *satisfy* $(C_1)-(C_4)$.

Proof. Property (C_1) is trivial whereas (C_2) follows immediately from (T_1).

Let now $A_1,\ldots,A_n, n \geq 1$, be distinct elements of $\mathcal{X}$ and suppose that $D_1,\ldots,D_k$ is a finite partition of $\mathbf{X}$ with elements in $\mathcal{X}$ and such that

$$A_1 = \bigcup_{(1)} D_j, \ldots, A_n = \bigcup_{(n)} D_j$$

where, as usual, $(s) = \{j \in \{1,\ldots,k\} : D_j \subseteq A_i\}$ for $s = 1,\ldots,n$. Property (C_3) is proved if we show that, for all $A_s \in \{A_1,\ldots,A_n\}$

$$\lim_m \sum_{B\in\mathcal{E}_m(A_s)} \prod_{i=1}^{m} V_{i,ps(B,i)} = \sum_{j\in(s)} [\lim_m \sum_{B\in\mathcal{E}_m(D_j)} \prod_{i=1}^{m} V_{i,ps(B,i)}] \tag{5.13}$$

on a set of λ probability one. It is in fact enough if we show that (5.13) holds for $A_s = D_1 \bigcup D_2$ with $D_1, D_2 \in \mathcal{X}$ disjoint. Note that in this case, for all $m \geq 1$,

$$\sum_{B\in\mathcal{E}_m(A_s)} \prod_{i=1}^{m} V_{i,ps(B,i)}$$
$$= \sum_{B\in\mathcal{E}_m(D_1)} \prod_{i=1}^{m} V_{i,ps(B,i)} + \sum_{B\in\mathcal{E}_m(D_2)} \prod_{i=1}^{m} V_{i,ps(B,i)} - \sum_{B\in\mathcal{E}_m(D_1)\bigcap\mathcal{E}_m(D_2)} \prod_{i=1}^{m} V_{i,ps(B,i)}$$

therefore, in order to prove (5.13), we need to show that

$$\lim_m \sum_{B\in\mathcal{E}_m(D_1)\bigcap\mathcal{E}_m(D_2)} \prod_{i=1}^{m} V_{i,ps(B,i)} = 0$$

with λ probability one.

Notice that,

$$\begin{aligned}\bigcap_m [\bigcup_{B\in\mathcal{E}_m(D_1)\bigcap\mathcal{E}_m(D_2)} B] &\subseteq \{\bigcap_m [\bigcup_{B\in\mathcal{E}_m(D_1)} B]\}\bigcap\{\bigcap_m [\bigcup_{B\in\mathcal{E}_m(D_2)} B]\} \\ &= D_1 \bigcap D_2 \\ &= \emptyset\end{aligned}$$

where the next to the last equality follows from Lemma 5.3. Therefore, if for all $m \geq 0$ we set $C_m = \bigcup_{B\in\mathcal{E}_m(D_1)\bigcap\mathcal{E}_m(D_2} B$, we get a sequence of sets $\{C_m\}$ decreasing to the empty set and such that $C_m \in \nabla_m$ for all $m \geq 0$. Let us now consider the sequence of random variables $\{\sum_{B\in\mathcal{E}_m(C_m)} \prod_{i=1}^{m} V_{i,ps(B,i)}\}$ which converges to 0 in mean because of (T_4) and (T_5). Since $C_{m+1} \subseteq C_m$ for all $m \geq 0$, $ps(B,m) \in \mathcal{E}_m(C_m)$ if $B \in \mathcal{E}_{m+1}(C_{m+1})$; therefore

$$\sum_{B\in\mathcal{E}_{m+1}(C_{m+1})} \prod_{i=1}^{m+1} V_{i,ps(B,i)} \leq \sum_{B\in\nabla_{m+1}:ps(B,m)\in\mathcal{E}_m(C_m)} \prod_{i=1}^{m+1} V_{i,ps(B,i)} = \sum_{B\in\mathcal{E}_m(C_m)} \prod_{i=1}^{m} V_{i,ps(B,i)}$$

where the last equality follows from (T_3). This shows that the sequence

$$\{\sum_{B\in\mathcal{E}_m(C_m)} \prod_{i=1}^{m} V_{i,ps(B,i)}\}$$

is monotonically decreasing with m and it thus converges with λ probability one; the limit must be 0 since the sequence is bounded between 0 and 1 and converges in mean to 0. Hence

$$\lambda[\lim_m \sum_{B\in\mathcal{E}_m(D_1)\bigcap\mathcal{E}_m(D_2)} \prod_{i=1}^{m} V_{i,ps(B,i)} = 0] = 1$$

and thus (5.13) follows when $A_s = D_1 \bigcup D_2$.

Finally let $\{A_n\}$ be a sequence of elements of $\mathcal{X}$ decreasing to the empty set. Fix $x \in (0,1]$; then Markov's inequality implies that, for each n,

$$q_{A_n}([0,x]) \geq 1 - \frac{1}{x}\int_0^1 t q_{A_n}(dt). \tag{5.14}$$

However

$$\begin{aligned}\int_0^1 t q_{A_n}(dt) &= E_\lambda[\lim_m \sum_{B\in\mathcal{E}_m(A_n)} \prod_{i=1}^m V_{i,ps(B,i)}] \\ &\leq E_\lambda[\sum_{B\in\mathcal{E}_n(A_n)} \prod_{i=1}^m V_{i,ps(B,i)}] \\ &= \sum_{B\in\mathcal{E}_n(A_n)} \prod_{i=1}^m E_\lambda[V_{i,ps(B,i)}]\end{aligned} \tag{5.15}$$

where the last equality is true because of (T_5).

For all $n \geq 0$ set $C_n = \bigcup_{B\in\mathcal{E}_n(A_n)} B$; then $C_n \in \nabla_n$. Moreover $\{C_n\}$ is a sequence of sets monotonically decreasing to the empty set. In fact, suppose that $y \in C_{n+1}$; then $\xi_{n+1}(y) \bigcap A_{n+1} \neq \emptyset$ and this implies that $\xi_n(y) \bigcap A_n \neq \emptyset$ since $ps(\xi_{n+1}(y), n) = \xi_n(y)$ and $A_{n+1} \subseteq A_n$. Therefore $\xi_n(y) \in \mathcal{E}_n(A_n)$ and $y \in C_n$ which proves that $C_{n+1} \subseteq C_n$. In order to verify that $\bigcap_n C_n = \emptyset$, suppose that this is not true and let $y \in \bigcap_n C_n$. Then, $y \in C_n$ for all n. Fix $k \geq 0$ and consider A_k; since $y \in C_k$, $\xi_k(y) \bigcap A_k \neq \emptyset$ and this implies that $\xi_m(y) \bigcap A_k \neq \emptyset$ for all $m \leq k$ because $\xi_k(y) \subseteq \xi_m(y)$. However if $m > k$, $\xi_m(y) \bigcap A_m \neq \emptyset$ since $y \in C_m$, but this implies that $\xi_m(y) \bigcap A_k \neq \emptyset$ because $A_m \subseteq A_k$. This shows that $y \in \bigcap_m [\bigcup_{B\in\mathcal{E}_m(A_k)} B] = A_k$ for all k; hence $y \in \bigcap_k A_k$ which is a contradiction because this set is empty.

Therefore the sequence $\{C_n\}$ satisfies the assumptions of (T_4) and this implies that

$$\lim_{n\to\infty} \sum_{B\in\mathcal{E}_n(C_n)} \prod_{i=1}^n E_\lambda[V_{i,ps(B,i)}] = \lim_{n\to\infty} \sum_{B\in\mathcal{E}_n(A_n)} \prod_{i=1}^n E_\lambda[V_{i,ps(B,i)}] = 0. \tag{5.16}$$

Equations (5.14), (5.15) and (5.16) show that $\lim_n q_{A_n}([0,x]) = 1$ for all $x \in (0,1]$. Therefore (C_4) is satisfied. $\diamond$

We now consider a statistical model $(\mathbf{X}^\infty \times \mathbf{P}, \mathcal{X}^\infty \times \mathcal{P}, \pi)$ where the probability π is defined as in (2.1) with $\nu = \mathcal{T}$ being tailfree with respect to a tree of partitions $\{\nabla_n\}$. The next result shows how to compute the marginal distribution of an element of an exchangeable sequence $\{X_n\}$ whose de Finetti measure is a tailfree process $\mathcal{T}$.

5.17 **Theorem**. *For all $A \in \mathcal{X}$,*

$$\tau_1(A) = \lim_n \sum_{B\in\mathcal{E}_m(A)} \prod_{i=1}^n E_\lambda[V_{i,ps(B,i)}].$$

Proof. For all $A \in \mathcal{X}$,

$$\tau_1(A) = \pi(X_1 \in A) = E_{\mathcal{T}}(P(A)) = E_{\mathcal{T}}(\lim_n \sum_{B \in \mathcal{E}_n(A)} P(B))$$

since $A = \bigcap_n [\bigcup_{B \in \mathcal{E}_n(A)} B]$ because of Lemma 5.3. However, since $\mathcal{T}$ is tailfree, the law of $P(B)$ is the same as the law of $\prod_{i=1}^n V_{i,ps(B,i)}$ for all $n \geq 1$ and all $B \in \nabla_n$. Therefore

$$E_{\mathcal{T}}(\lim_n \sum_{B \in \mathcal{E}_n(A)} P(B)) = E_\lambda[\lim_n \sum_{B \in \mathcal{E}_m(A)} \prod_{i=1}^n V_{i,ps(B,i)}] = \lim_n \sum_{B \in \mathcal{E}_m(A)} \prod_{i=1}^n E_\lambda[V_{i,ps(B,i)}]$$

where the last equality follows from Monotone Convergence Theorem and (T_5). ◇

Tailfree processes are conjugate; if P is a random element of $(\mathbf{P}, \mathcal{P})$ whose distribution is tailfree with respect to a tree of partitions of $(\mathbf{X}, \mathcal{X})$, then the posterior probability distribution of P is again tailfree with respect to the same tree of partitions. The next results prove this fact which make tailfree processes useful for Bayesian nonparametrics.

Assume that the distribution $\mathcal{T}$ of P is tailfree with respect to a tree of partitions $\{\nabla_n\}$ of $(\mathbf{X}, \mathcal{X})$: we indicate with $\{V_{n,B}\}$ the double sequence of random variables defined on a space $(\Omega, \sigma, \lambda)$ whose existence is stated in Definition 5.5. For all $n \geq 0$, let $E_n = \{x \in \mathbf{X} : \prod_{i=1}^n E_\lambda[V_{i,\xi_i(x)}] = 0\}$; set $E = \bigcup_n E_n$.

5.18 **Lemma**. *For all $n \geq 0$, $E_n \in \mathcal{X}$ and $\tau_1(E_n) = 0$. Therefore $\tau_1(E) = 0$.*

Proof. Let $n \geq 1$. If $x \in E_n$

$$E_\lambda[\prod_{i=1}^n V_{i,ps(\xi_n(x),i)}] = \prod_{i=1}^n E_\lambda[V_{i,ps(\xi_n(x),i)}] = \prod_{i=1}^n E_\lambda[V_{i,\xi_i(x)}] = 0$$

where the first equality holds because the sequences $\{V_{1,B} : B \in \nabla_1\}$, $\{V_{2,B} : B \in \nabla_2\}, \ldots$ are mutually independent whereas the next to the last equality follows from the fact that $ps(\xi_{n+1}(x), n) = \xi_n(x)$ for all $n \geq 0$ and $x \in \mathbf{X}$.

Now suppose $y \in \xi_n(x)$: then $\xi_n(y) = \xi_n(x)$. Therefore $y \in E_n$ if and only if $x \in E_n$. This shows that

$$E_n = \bigcup_{B \in \nabla_n : E_\lambda[\prod_{i=1}^n V_{i,ps(B,i)}]=0} B.$$

Being a finite union of elements of ∇_n, $E_n \in \mathcal{X}$. Moreover, since the sets of ∇_n are disjoint,

$$\tau_1(E_n) = \sum_{B \in \nabla_n : E_\lambda[\prod_{i=1}^n V_{i,ps(B,i)}]=0} \tau_1(B) = \sum_{B \in \nabla_n : E_\lambda[\prod_{i=1}^n V_{i,ps(B,i)}]=0} E_\lambda[\prod_{i=1}^n V_{i,ps(B,i)}] = 0$$

where the next to the last equality follows from Theorem 5.17.

The fact that $\tau_1(E) = 0$ now follows easily. $\diamond$

5.19 **Lemma**. *Let $n \geq 1$. For all $x \in E^c$ and $C \in \mathcal{B}^{k_n}[0,1]$,*

$$\mathcal{T}[(\wp_{B_{n,1}}, ..., \wp_{B_{n,k_n}}) \in C | X_1 = x] = \frac{E_\lambda[I_C[(\prod_{i=1}^n V_{i,ps(B_{n,1},i)}, ..., \prod_{i=1}^n V_{i,ps(B_{n,k_n},i)})] \prod_{i=1}^n V_{i,\xi_i(x)}]}{\prod_{i=1}^n E_\lambda[V_{i,\xi_i(x)}]}. \tag{5.20}$$

Proof. Fix $C \in \mathcal{B}^{k_n}[0,1]$; we need to verify that, for all $A \in \mathcal{X}$,

$$\pi((P(B_{n,1}), ..., P(B_{n,k_n})) \in C, X_1 \in A) = \int_{A \bigcap E^c} h(x)\tau_1(dx) \tag{5.21}$$

where, for all $x \in E^c$, the expression of $h(x)$ is that of the right member of (5.20). Because of Lemma 5.3 it is in fact enough to verify that (5.21) holds for $A \in \bigcup_m \nabla_m$; therefore assume that $A \in \nabla_m$ with $m \geq 0$.

Case I: $n \leq m$. For each $x \in A$ we have that $\xi_i(x) = ps(A, i)$ if $i \leq n$. Therefore $h(x)$ is constant for $x \in A$ and

$$\begin{aligned}
&\int_{A \bigcap E^c} h(x)\tau_1(dx) \\
&= \frac{E_\lambda[I_C[(\prod_{i=1}^n V_{i,ps(B_{n,1},i)}, ..., \prod_{i=1}^n V_{i,ps(B_{n,k_n},i)})] \prod_{i=1}^n V_{i,ps(A,i)}]}{\prod_{i=1}^n E_\lambda[V_{i,ps(A,i)}]} \tau_1(A) \\
&= \frac{E_\lambda[I_C[(\prod_{i=1}^n V_{i,ps(B_{n,1},i)}, ..., \prod_{i=1}^n V_{i,ps(B_{n,k_n},i)})] \prod_{i=1}^n V_{i,ps(A,i)}]}{\prod_{i=1}^n E_\lambda[V_{i,ps(A,i)}]} \prod_{i=1}^m E_\lambda[V_{i,ps(A,i)}] \\
&= E_\lambda[I_C[(\prod_{i=1}^n V_{i,ps(B_{n,1},i)}, ..., \prod_{i=1}^n V_{i,ps(B_{n,k_n},i)})] \prod_{i=1}^m V_{i,ps(A,i)}] \\
&= E_\pi[I_C[(P(B_{n,1}), ..., P(B_{n,k_n}))]P(A)] \\
&= \pi((P(B_{n,1}), ..., P(B_{n,k_n})) \in C, X_1 \in A)
\end{aligned}$$

where the third equality holds because the sequences $\{V_{1,B} : B \in \nabla_1\}, \{V_{2,B} : B \in \nabla_2\}, ...$ are mutually independent.

Case II: $m < n$. Then:

$$\begin{aligned}
&\pi((P(B_{n,1}), ..., P(B_{n,k_n})) \in C, X_1 \in A) \\
&= \sum_{B \in \mathcal{E}_n(A)} \pi((P(B_{n,1}), ..., P(B_{n,k_n})) \in C, X_1 \in B) \\
&= \sum_{B \in \mathcal{E}_n(A)} \int_{B \bigcap E^c} h(x)\tau_1(dx) \\
&= \int_{A \bigcap E^c} h(x)\tau_1(dx)
\end{aligned}$$

where the next to the last equality holds because of Case I. ◇

5.22 **Theorem**. *Let $\mathcal{T}$ be a tailfree process with respect to a tree of partitions $\{\nabla_n\}$ of $(\mathbf{X}, \mathcal{X})$. Then, for all $n \geq 1$ and $(x_1, ...x_n) \in \mathbf{X}^n$,*

$$\mathcal{T}(\cdot|X_1 = x_1, ..., X_n = x_n)$$

is again a tailfree process with respect to $\{\nabla_n\}$.

Proof. We consider the case $n = 1$; the general case then follows by induction on n.

Let $x \in E^c$ be fixed. For all $n \geq 1$ define

$$Q_n(C) = \frac{E_\lambda(I_C((V_{n,B_{n,1}}, ..., V_{n,B_{n,k_n}}))V_{n,\xi_n(x)})}{E_\lambda(V_{n,\xi_n(x)})}$$

for $C \in \mathcal{B}^{k_n}[0,1]$. It's easy to check that Q_n is a probability measure on $([0,1]^{k_n}, \mathcal{B}^{k_n}[0,1])$. On the product space

$$([0,1]^{k_1} \times [0,1]^{k_2} \times ..., \mathcal{B}^{k_1}[0,1] \times \mathcal{B}^{k_2}[0,1] \times ...)$$

we define the product probability measure $Q = Q_1 \times Q_2 \times ....$ Finally, for $n \geq 1$ and $B \in \nabla_n$ we set $W_{n,B}$ to be the (n, B)-projection of $[0,1]^{k_1} \times [0,1]^{k_2} \times ...$ into $[0,1]$. Therefore, for all $n \geq 1$, Q_n is the probability distribution of $(W_{n,B_{n,1}}, ..., W_{n,B_{n,k_n}})$ and the sequences of random variables

$$\{W_{1,B} : B \in \nabla_1\}, \{W_{2,B} : B \in \nabla_2\}, ...$$

are mutually independent. Moreover, for $n \geq 1$ and $C \in \mathcal{B}^{k_n}[0,1]$,

$$\begin{aligned} &Q((\prod_{i=1}^n W_{i,ps(B_{n,1},i)}, ..., \prod_{i=1}^n W_{i,ps(B_{n,k_n},i)}) \in C) \\ &= \frac{E_\lambda[I_C[(\prod_{i=1}^n V_{i,ps(B_{n,1},i)}, ..., \prod_{i=1}^n V_{i,ps(B_{n,k_n},i)})] \prod_{i=1}^n V_{i,\xi_i(x)}]}{\prod_{i=1}^n E_\lambda[V_{i,\xi_i(x)}]} \\ &= \mathcal{T}[(P(B_{n,1}), ..., P(B_{n,k_n})) \in C|X_1 = x] \end{aligned} \tag{5.23}$$

where the first equality holds because of the definition of $Q_1, ..., Q_n$ whereas the next one is true because of Lemma 5.19.

Since $\tau_1(E) = 0$, equations (5.23) show that $\mathcal{T}(\cdot|X_1 = x)$ is again tailfree with respect to $\{\nabla_n\}$. ◇

We have seen that Dirichlet processes are tailfree; in fact they are tailfree with respect to any tree of partitions of $\mathbf{X}$. However there are tailfree processes with properties not shared by any Dirichlet process; for instance, a tailfree process may assign probability one to a set of probability measures on $(\mathbf{P}, \mathcal{P})$ which are absolutely continuous with respect to τ_1. In order to prove this result, let $\mathcal{T}$ be tailfree with respect to a tree of partitions $\{\nabla_n\}$ of $(\mathbf{X}, \mathcal{X})$; we assume in the following that the probability measure τ_1 is such that

$$\tau_1(B) = \int_{\mathbf{P}} p(B)\mathcal{T}(dp) > 0,$$

for all $n \geq 0$ and for all $B \in \nabla_n$.

On the product space $(\mathbf{X} \times \mathbf{P}, \mathcal{X} \times \mathcal{P})$ define the product probability measure $\tau_1 \times \mathcal{T}$ and, for all $n \geq 0$ and $(x, p) \in \mathbf{X} \times \mathbf{P}$, set

$$f_n(x, p) = \frac{p(\xi_n(x))}{\tau_1(\xi_n(x))}.$$

For all $n \geq 0$, let $\mathcal{F}_n \subseteq \mathcal{X} \times \mathcal{P}$ be the sigma-field generated by the rectangles $D \times A$ where D is the union of sets in ∇_n whereas $A \in \mathcal{P}$. Since, for all $n \geq 1$, ∇_{n+1} is a finer partition of ∇_n, $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$; therefore the sequence of sigma-fields $\{\mathcal{F}_n\}$ is a filtration of $(\mathbf{X} \times \mathbf{P}, \mathcal{X} \times \mathcal{P})$.

5.24 **Lemma**. *The sequence $\{f_n\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n\}$ and the product measure $\tau_1 \times \mathcal{T}$.*

Proof. We leave it as an exercise to prove that, for all $n \geq 0$, f_n is $\mathcal{F}_n$ measurable. Next, notice that, for all $n \geq 0$, $f_n \geq 0$ and, using Fubini,

$$\begin{aligned}
&\int_{\mathbf{X}\times\mathbf{P}} f_n(x,p)\tau_1(dx)\mathcal{T}(dp) \\
&\quad = \int_{\mathbf{P}} (\int_{\mathbf{X}} f_n(x,p)\tau_1(dx))\mathcal{T}(dp) \\
&\quad = \int_{\mathbf{P}} (\sum_{B\in\nabla_n} \int_B f_n(x,p)\tau_1(dx))\mathcal{T}(dp).
\end{aligned}$$

However, if $x \in B \in \nabla_n$, $\xi_n(x) = B$; therefore, for all $p \in \mathbf{P}$,

$$\int_B f_n(x,p)\tau_1(dx) = \frac{p(B)}{\tau_1(B)}\tau_1(B) = p(B).$$

Hence

$$\int_{\mathbf{P}} (\sum_{B\in\nabla_n} \int_B f_n(x,p)\tau_1(dx))\mathcal{T}(dp) = \int_{\mathbf{P}} \sum_{B\in\nabla_n} p(B)\mathcal{T}(dp) = \int_{\mathbf{P}} 1\mathcal{T}(dp) = 1.$$

Therefore, for all $n \geq 0$, the expected value of f_n is finite and equal to one.

Now let $n \geq 0$, D be the finite union of elements of ∇_n and $G \in \mathcal{P}$. Then

$$\begin{aligned}
&\int_{D\times G} f_{n+1}(x,p)\tau_1(dx)\mathcal{T}(dp)\\
&\quad = \int_G (\int_D f_{n+1}(x,p)\tau_1(dx))\mathcal{T}(dp)\\
&\quad = \int_G (\sum_{B\in\mathcal{E}_{n+1}(D)} \int_B f_{n+1}(x,p)\tau_1(dx))\mathcal{T}(dp)\\
&\quad = \int_G (\sum_{B\in\mathcal{E}_{n+1}(D)} p(B))\mathcal{T}(dp),
\end{aligned}$$

where the last equality holds because $f_n(x,p) = \frac{p(B)}{\tau_1(B)}$ for all $x \in B \in \nabla_{n+1}$. But ∇_{n+1} is a partition of $\mathbf{X}$ finer than ∇_n : hence

$$\sum_{B\in\mathcal{E}_{n+1}(D)} p(B) = p(D) = \sum_{B\in\mathcal{E}_n(D)} p(B).$$

Therefore

$$\begin{aligned}
&\int_G (\sum_{B\in\mathcal{E}_{n+1}(D)} p(B))\mathcal{T}(dp)\\
&\quad = \int_G (\sum_{B\in\mathcal{E}_n(D)} p(B))\mathcal{T}(dp)\\
&\quad = \int_G (\int_D f_n(x,p)\tau_1(dx))\mathcal{T}(dp)\\
&\quad = \int_{D\times G} f_n(x,p)\tau_1(dx)\mathcal{T}(dp).
\end{aligned}$$

This shows that

$$\int_{D\times G} f_{n+1}(x,p)\tau_1(dx)\mathcal{T}(dp) = \int_{D\times G} f_n(x,p)\tau_1(dx)\mathcal{T}(dp);$$

since the algebra of rectangles of type $D \times G$ generates $\mathcal{F}_n$ the previous equality also proves that

$$E(f_{n+1}|\mathcal{F}_n) = f_n$$

where the conditional expectation is taken with respect to the product measure $\tau_1 \times \mathcal{T}$. Hence $\{f_n\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n\}$ and the product measure $\tau_1 \times \mathcal{T}$. $\diamond$

Since for all $n \geq 0$, $f_n \geq 0$, by Doob's Martingale Convergence Theorem there is a function $f : \mathbf{X} \times \mathbf{P} \to [0,1]$ such that $\lim_{n\to\infty} f_n = f$ on a set of probability $\tau_1 \times \mathcal{T}$ one. When the sequence $\{f_n\}$ is uniformly integrable, the convergence is in $L_1(\mathbf{X} \times \mathbf{P}, \mathcal{X} \times \mathcal{P}, \tau_1 \times \mathcal{T})$ and the following theorem holds: it states that $\mathcal{T}$ assigns probability one to a set of probability measures absolutely continuous with respect to τ_1.

5.25 **Theorem**. *Assume that $\{f_n\}$ is uniformly integrable. Then*

$$\mathcal{T}(\{p \in \mathbf{P} : p(A) = \int_A f(x,p)\tau_1(dx), \text{ for all } A \in \mathcal{X}\}) = 1. \tag{5.26}$$

Proof. Fix $k \geq 0$ and $D \in \nabla_k$. Using Fubini, for all $n \geq k$ and $G \in \mathcal{P}$,

$$\begin{aligned}
&\int_{D\times G} f_n(x,p)\tau_1(dx)\mathcal{T}(dp) \\
&\quad = \int_G (\sum_{B\in\mathcal{E}_n(D)} \int_B f_n(x,p)\tau_1(dx))\mathcal{T}(dp) \\
&\quad = \int_G p(D)\mathcal{T}(dp).
\end{aligned}$$

However, since $\{f_n\}$ is uniformly integrable and $\lim_{n\to\infty} f_n = f$ in $L_1(\mathbf{X}\times\mathbf{P}, \mathcal{X}\times\mathcal{P}, \tau_1\times\mathcal{T})$,

$$\lim_{n\to\infty} \int_{D\times G} f_n(x,p)\tau_1(dx)\mathcal{T}(dp) = \int_{D\times G} f(x,p)\tau_1(dx)\mathcal{T}(dp).$$

Therefore, for all $G \in \mathcal{P}$,

$$\int_G p(D)\mathcal{T}(dp) = \int_{D\times G} f(x,p)\tau_1(dx)\mathcal{T}(dp) = \int_G (\int_D f(x,p)\tau_1(dx))\mathcal{T}(dp).$$

From this it follows that

$$\mathcal{T}(\{p \in \mathbf{P} : p(D) = \int_D f(x,p)\tau_1(dx)\}) = 1 \tag{5.27}$$

Now let $\mathcal{G} = \bigcap_{B\in\bigcup_{n=0}^\infty \nabla_n}\{p \in \mathbf{P} : p(B) = \int_B f(x,p)\tau_1(dx)\}$. Since each partition ∇_n has a finite number of elements, equation (5.27) implies that $\mathcal{T}(\mathcal{G}) = 1$. Let $p \in \mathcal{G}$; then, for all $A \in \mathcal{X}$,

$$\begin{aligned}
p(A) &= \lim_{n\to\infty} \sum_{B\in\mathcal{E}_n(A)} p(B) \\
&= \lim_{n\to\infty} \sum_{B\in\mathcal{E}_n(A)} \int_B f(x,p)\tau_1(dx) \\
&= \int_A f(x,p)\tau_1(dx)
\end{aligned}$$

where the first and the last equality are a consequence of Lemma 5.3. Hence

$$\mathcal{T}(\{p \in \mathbf{P} : p(A) = \int_A f(x,p)\tau_1(dx), \text{ for all } A \in \mathcal{X}\}) = 1.$$

◇

A sufficient condition for uniform integrability of the sequence $\{f_n\}$ (see, for instance, Lemma II.6.3 in Shiryayev (1984)) is the existence of a $k > 1$ such that

$$\sup_n \int_{\mathbf{X}\times\mathbf{P}} f_n^k(x,p)\tau_1(dx)\mathcal{T}(dp) < \infty.$$

This condition implies the next useful Lemma and the following Corollary, a first version of which appeared already in Kraft (1964); recall that, for a process $\mathcal{T}$ tailfree with respect to a tree of partitions $\{\nabla_n\}$, there is a sequence $\{V_{n,B}\}$ of random variables defined on a space $(\Omega, \sigma, \lambda)$ for which the conditions stated in Definition 5.5 are satisfied.

5.28 **Lemma**. *If*

$$\sum_{n=0}^{\infty} \sup_{B\in\nabla_n} \frac{Var_\lambda(V_{n,B})}{(E_\lambda^2(V_{n,B}))} < \infty, \tag{5.29}$$

then the martingale $\{f_n\}$ is uniformly integrable.

Proof. We will verify that, when (5.29) is satisfied,

$$\sup_n \int_{\mathbf{X}\times\mathbf{P}} f_n^2(x,p)\tau_1(dx)\mathcal{T}(dp) < \infty;$$

this implies uniform integrability of $\{f_n\}$.

Fix $n \geq 0$, and note that

$$\int_{\mathbf{X}\times\mathbf{P}} f_n^2(x,p)\tau_1(dx)\mathcal{T}(dp) = \int_{\mathbf{X}} (\int_{\mathbf{P}} (\frac{p(\xi_n(x))}{\tau_1(\xi_n(x))})^2 \mathcal{T}(dp))\tau_1(dx).$$

However, for all $x \in \mathbf{X}$, it follows from Lemma 5.17 and (T_3) that

$$\tau_1(\xi_n(x)) = \prod_{i=1}^{n} E_\lambda(V_{i,ps(\xi_n,i)(x)}) = \prod_{i=1}^{n} E_\lambda(V_{i,\xi_i(x)}).$$

Moreover,

$$\int_{\mathbf{P}} p^2(\xi_n(x))\mathcal{T}(dp) = E_\lambda(\prod_{i=1}^{n} V_{i,ps(\xi_n(x),i)}^2) = \prod_{i=1}^{n} E_\lambda(V_{i,\xi_i(x)}^2)$$

where the last equality follows from the independence property stated in (T_5). Therefore

$$\int_{\mathbf{X}\times\mathbf{P}} f_n^2(x,p)\tau_1(dx)\mathcal{T}(dp) = \int_{\mathbf{X}} \prod_{i=1}^{n} \frac{E_\lambda(V_{i,\xi_i(x)}^2)}{E_\lambda^2(V_{i,\xi_i(x)})}\tau_1(dx). \tag{5.30}$$

Let us indicate with

$$M = \sum_{n=0}^{\infty} \sup_{B\in\nabla_n} \frac{\mathrm{Var}_\lambda(V_{n,B})}{(E_\lambda^2(V_{n,B}))} < \infty$$

which does not depend on n. Then

$$\begin{aligned}
&\log(\sup_{x\in\mathbf{X}}\prod_{i=1}^{n}\frac{E_\lambda(V^2_{i,\xi_i(x)})}{E^2_\lambda(V_{i,\xi_i(x)})}) \\
&\le \log(\prod_{i=1}^{n}\sup_{x\in\mathbf{X}}\frac{E_\lambda(V^2_{i,\xi_i(x)})}{E^2_\lambda(V_{i,\xi_i(x)})}) \\
&= \sum_{i=1}^{n}\log(\sup_{x\in\mathbf{X}}\frac{E_\lambda(V^2_{i,\xi_i(x)})}{E^2_\lambda(V_{i,\xi_i(x)})}) \\
&\le \sum_{i=1}^{n}\sup_{x\in\mathbf{X}}[\frac{E_\lambda(V^2_{i,\xi_i(x)})}{E^2_\lambda(V_{i,\xi_i(x)})}-1] \\
&= \sum_{i=1}^{n}\sup_{x\in\mathbf{X}}\frac{\mathrm{Var}_\lambda(V_{i,\xi_i(x)})}{E^2_\lambda(V_{i,\xi_i(x)})} \\
&\le \sum_{i=1}^{\infty}\sup_{x\in\mathbf{X}}\frac{\mathrm{Var}_\lambda(V_{i,\xi_i(x)})}{E^2_\lambda(V_{i,\xi_i(x)})} \\
&= \sum_{i=1}^{\infty}\sup_{B\in\nabla_i}\frac{\mathrm{Var}_\lambda(V_{i,B})}{E^2_\lambda(V_{i,B})} = M;
\end{aligned}$$

the second inequality follows from the fact that $\log z \le z-1$, for all $z \ge 0$.

Hence

$$\sup_{x\in\mathbf{X}}\prod_{i=1}^{n}\frac{E_\lambda(V^2_{i,\xi_i(x)})}{E^2_\lambda(V_{i,\xi_i(x)})}) \le e^M;$$

this and (5.30) imply that

$$\int_{\mathbf{X}\times\mathbf{P}} f_n^2(x,p)\tau_1(dx)\mathcal{T}(dp) \le e^M.$$

Therefore,

$$\sup_n \int_{\mathbf{X}\times\mathbf{P}} f_n^2(x,p)\tau_1(dx)\mathcal{T}(dp) < \infty.$$

◇

5.31 **Corollary**. *If for all $n \ge 0$ and $B \in \nabla_n$, $E_\lambda(V_{n,B}) > 0$ and*

$$\sum_{n=0}^{\infty}\sup_{B\in\nabla_n}\frac{Var_\lambda(V_{n,B})}{(E^2_\lambda(V_{n,B}))} < \infty$$

then there exists a measurable $f : \mathbf{X}\times\mathbf{P} \to [0,\infty)$ such that

$$\mathcal{T}(\{p\in\mathbf{P} : p(A) = \int_{\mathbf{X}} f(x,p)\tau_1(dx) \textit{ for all } A\in\mathcal{X}\}).$$

Proof. For all $n \geq 0$ and $B \in \nabla_n$, $\tau_1(B) > 0$ since each random variable of the sequence $\{V_{n,B}\}$ has strictly positive expected value. Then the result follows from the previous Lemma and Theorem 5.25. ◇

Examples of tailfree processes which select continuous probability measures with probability one are deferred to the next section where we will consider a special class of tailfree processes.

6 Polya trees

Polya trees are a special class of tailfree processes; they have been considered by Ferguson (1974), Mauldin, Sudderth and Williams (1992) and by Lavine (1992, 1994). They generalize the Dirichlet process with the advantage that they can be constructed to give probability one to the set of continuous or absolutely continuous probability distributions on $(\mathbf{X}, \mathcal{X})$.

6.1 **Definition**. *Let $\{\nabla_n\}$ be a tree of partitions of $(\mathbf{X}, \mathcal{X})$ such that, for a given $r > 0$ and for all $n \geq 0$ and $B \in \nabla_n$ there are exactly r sets $C_1, ..., C_r$ in ∇_{n+1} such that $ps(C_i, n) = B$ for $i = 1, ..., r$. Then, a probability distribution $\mathcal{T}$ on $(\mathbf{P}, \mathcal{P})$ which is tailfree with respect to $\{\nabla_n\}$ is said to be a Polya tree if the sequence of random variables $\{V_{n,B} : n \geq 0, B \in \nabla_n\}$ appearing in Definition 5.5 is such that for all $n \geq 0$ and $B \in \nabla_n$ the joint distribution of the random vector $(V_{n+1,C_1}, ..., V_{n+1,C_r})$ is Dirichlet and these vectors are independent for different $B \in \nabla_n$.*

In the rest of the section we consider a few examples of Polya trees with the aim to show their greater tractability over more general tailfree processes. All of them will be *binary Polya trees*, meaning that they will be defined by means of trees of binary partitions.

Let $\nabla_0 = \{\mathbf{X}\}$ and set $\nabla_1 = \{B_0, B_1\}$ to be a measurable partition for $\mathbf{X}$. Then let (B_{00}, B_{01}) be a measurable partition of B_0 and (B_{10}, B_{11}) a measurable partition of B_1 and set $\nabla_2 = \{B_{00}, B_{01}, B_{10}, B_{11}\}$. Continue in this fashion ad infinitum; the tree of partitions so constructed is called a tree of binary partitions of $\mathbf{X}$. For $m \geq 1$, let $\epsilon = \epsilon_1 ... \epsilon_m$ with $\epsilon_i \in \{0, 1\}$ for $i = 1, ..., m$; then ϵ defines in a unique way a set $B_\epsilon \in \nabla_m$. We use the notation Y_ϵ for the random variable V_{m,B_ϵ}.

Set $Y = 1$ and assume that (Y_0, Y_1) has Dirichlet distribution with parameters (α_0, α_1). For all $m \geq 1$ and $\epsilon = \epsilon_1 ... \epsilon_m$, let $(Y_{\epsilon 0}, Y_{\epsilon 1})$ have Dirichlet distribution with parameters $(\alpha_{\epsilon 0}, \alpha_{\epsilon 1})$: hence with probability one $Y_{\epsilon 0} = 1 - Y_{\epsilon 1}$ and it is distributed according to a Beta with parameters $(\alpha_{\epsilon 0}, \alpha_{\epsilon 1})$. Moreover, assume that the random variables of the collection $\{Y_{\epsilon 0}\}$ are independent. With these assumptions conditions $(T_1)-(T_3)$ and (T_5) are satisfied; when condition (T_4) is also satisfied, the tree of partitions $\{\nabla_n\}$ and the sequence of random variables $\{Y_\epsilon\}$ define a particular Polya tree distribution on $(\mathbf{P}, \mathcal{P})$ called binary Polya tree with parameters $(\{\nabla_n\}, \{\alpha_\epsilon\})$.

6.2 **Example**. Let $\mathbf{X} = (0,1]$ endowed with its Borel sigma-field $\mathcal{B}(0,1]$ and assume that, for all $n \geq 1$,

$$\nabla_n = \{(\frac{i-1}{2^n}, \frac{i}{2^n}] : i = 1, ..., 2^n\}.$$

Let $\{\alpha_n\}$ be a sequence of nonnegative real numbers. Define Y_0 to be a random variable with Beta distribution with parameters (α_1, α_1) and set $Y_1 = 1 - Y_0$. For all $n \geq 0$ and $\epsilon = \epsilon_1 ... \epsilon_n$ with $\epsilon_i \in \{0,1\}$, let $Y_{\epsilon 0}$ be a random variable independent from Y_ϵ and from $Y_{\epsilon'}$ for all $\epsilon' = \epsilon'_1 ... \epsilon'_n, \epsilon'_i \in \{0,1\}$, different from ϵ. Assume that the distribution of $Y_{\epsilon 0}$ is Beta with parameters $(\alpha_{n+1}, \alpha_{n+1})$; set $Y_{\epsilon 1} = 1 - Y_{\epsilon 0}$.

The tree of partitions $\{\nabla_n\}$ and the set of random variables Y's define a binary Polya tree; in fact it's easy to check that conditions $(T_1)-(T_5)$ of the previous section hold. In order to verify that (T_4) is satisfied notice that, if $\{B_n\}$ is a sequence of sets decreasing to the empty set and such that, for each $n \geq 0$, B_n is union of sets in ∇_n, then

$$\begin{aligned}
\lim_{n\to\infty} \sum_{\epsilon=\epsilon_1...\epsilon_n : B_\epsilon \in \mathcal{E}_n(B_n)} \prod_{i=1}^{n} E[Y_{\epsilon_1 ... \epsilon_n}] &= \lim_{n\to\infty} \sum_{\epsilon=\epsilon_1...\epsilon_n : B_\epsilon \in \mathcal{E}_n(B_n)} \frac{1}{2^n} \\
&= \lim_{n\to\infty} l(B_n) = 0
\end{aligned}$$

where l is used for the Lebesgue measure on $(0,1]$, since , for all $n \geq 0$ and $\epsilon = \epsilon_1 ... \epsilon_n$, $E[Y_\epsilon] = \frac{1}{2}$.

We note that, for all $n \geq 1$ and $i = 1, ..., 2^n$,

$$\tau_1((\frac{i-1}{2^n}, \frac{i}{2^n}]) = \frac{1}{2^n};$$

therefore $\tau_1 = l$, the Lebesgue measure on $(0,1]$.

Moreover, since for all $n \geq 0$,

$$Var[Y_\epsilon] = \frac{1}{4(2\alpha_n + 1)}$$

we have that

$$\sum_{n=0}^{\infty} \sup_{B_\epsilon \in \nabla_n} \frac{Var(Y_\epsilon)}{E^2(Y_\epsilon)} = \sum_{n=0}^{\infty} \frac{1}{2\alpha_n + 1}.$$

Setting, for instance, $\alpha_n = \alpha n^2$ with $\alpha > 0$ for $n = 1, 2, ...$ yields a probability distribution $\mathcal{T}$ on the set of probability measures defined on $((0,1], \mathcal{B}(0,1])$ which almost surely selects probabilities absolutely continuous with respect to the Lebesgue measure, as follows from Corollary 5.31.

Dubins and Freedman (1963) showed that setting $\alpha_n = \alpha > 0$ for all $n \geq 1$, yields instead a binary Polya tree which almost surely selects continuous probabilities on $((0,1], \mathcal{B}(0,1])$ singular with respect to the Lebesgue measure.

Finally if $\alpha_n = 2^n$ for $n = 1, 2...$, we obtain a Dirichlet process as follows from the next theorem.

6.3 **Theorem**. *Let $\mathcal{T}$ be a binary Polya tree with parameters $(\{\nabla_n\}, \{\alpha_\epsilon\})$. If, for all $n \geq 1$ and $\epsilon = \epsilon_1 ... \epsilon_n, \epsilon_i \in \{0,1\}$,*

$$\alpha_\epsilon = \alpha_{\epsilon 0} + \alpha_{\epsilon 1}$$

then $\mathcal{T}$ is a Dirichlet process.

Proof. Let P be a random element of $\mathbf{P}$ with probability distribution $\mathcal{T}$. In order to prove that $\mathcal{T}$ is a Dirichlet process we need to show that there is a finite measure α on $(\mathbf{X}, \mathcal{X})$ such that, for any finite partition $D_1, ..., D_k$ of $\mathbf{X}$ with $D_i \in \mathcal{X}, i = 1, ..., k$, the random vector $(P(D_1), ..., P(D_k))$ has Dirichlet distribution with parameters $(\alpha(D_1), ..., \alpha(D_k))$.

By means of Charatheodory's Theorem, it is not difficult to show that there is a unique finite measure α on $(\mathbf{X}, \mathcal{X})$ such that, for all $n \geq 1$ and $\epsilon = \epsilon_1 ... \epsilon_n$,

$$\alpha(B_\epsilon) = \alpha_\epsilon.$$

This is were we use the assumption that, for all $n \geq 1$ and $\epsilon = \epsilon_1 ... \epsilon_n$,

$$\alpha_\epsilon = \alpha_{\epsilon 0} + \alpha_{\epsilon 1}.$$

Fix $n \geq 1$ and consider the partition $\nabla_n = \{B_{0...00}, B_{0...01}, ..., B_{1...10}, B_{1...1}\}$; by computing its mixed moments, one may show that the random vector

$$(P(B_{0...00}), P(B_{0...01}), ..., P(B_{1...10}), P(B_{1...1}))$$

has Dirichlet distribution with parameters $(\alpha(B_{0...00}), \alpha(B_{0...01}), ..., \alpha(B_{1...10}), \alpha(B_{1...1})))$.

Now let $D_1, ..., D_k$ be a measurable partition of $\mathbf{X}$ and consider the distribution of the random vector $(P(D_1), ..., P(D_k))$; because of (5.11) its distribution is the weak limit, as

$n \to \infty$, of a Dirichlet with parameters

$$(\sum_{B \in \mathcal{E}_n(D_1)} \alpha(B), ..., \sum_{B \in \mathcal{E}_n(D_k)} \alpha(B)).$$

Using Lemma 5.3 one proves that this is a Dirichlet with parameters $(\alpha(D_1), ..., \alpha(D_k))$. $\diamond$

6.4 **Remark**. If $\mathcal{T}$ is a Dirichlet process with parameter α and $\{\nabla_n\}$ is a binary partition of $\mathbf{X}$, for all $n \geq 1$ and $B_\epsilon \in \nabla_n$ define $\alpha_\epsilon = \alpha(B_\epsilon)$. Then $\mathcal{T}$ is a binary Polya tree with parameters $(\{\nabla_n\}, \{\alpha_\epsilon\})$. Moreover, since α is a measure,

$$\alpha_\epsilon = \alpha_{\epsilon 0} + \alpha_{\epsilon 1}$$

for all ϵ. $\diamond$

We conclude the section with the determination of the posterior distribution of a binary Polya tree; therefore let $(\mathbf{X}^\infty \times \mathbf{P}, \mathcal{X}^\infty \times \mathcal{P}, \pi)$ be a statistical model where the probability π is defined as in (2.1) with $\nu = \mathcal{T}$ a binary Polya tree.

6.5 **Theorem**. *Let $\mathcal{T}$ be a binary Polya tree with parameters $(\{\nabla_n\}, \{\alpha_\epsilon\})$. Then, for all $n \geq 1$ and $(x_1, ..., x_n) \in \mathbf{X}^n$,*

$$\mathcal{T}(\cdot | X_1 = x_1, ..., X_n = x_n)$$

is again a binary Polya tree with parameters $(\{\nabla_n\}, \{\tilde{\alpha}_\epsilon\})$ where, for all $m \geq 1$ and $\epsilon = \epsilon_1 ... \epsilon_m, \epsilon_i \in \{0, 1\}$,

$$\tilde{\alpha}_\epsilon = \alpha_\epsilon + \sum_{i=1}^n I_{B_\epsilon}(x_i).$$

Proof. We will consider only the case $n = 1$; the general case will then follow by induction on n.

Let $x \in \mathbf{X}$. Define a family $\{\tilde{Y}_\epsilon\}$ of random variables such that, for all $m \geq 1$ and $\epsilon = \epsilon_1 ... \epsilon_m$ with $\epsilon_i \in \{0, 1\}$, the vector $(\tilde{Y}_{\epsilon 0}, \tilde{Y}_{\epsilon 1})$ has Dirichlet distribution with parameters $(\tilde{\alpha}_{\epsilon 0}, \tilde{\alpha}_{\epsilon 1})$. Moreover assume that the vectors $(\tilde{Y}_{\epsilon 0}, \tilde{Y}_{\epsilon 1})$ and $(\tilde{Y}_{\epsilon' 0}, \tilde{Y}_{\epsilon' 1})$ are independent for different ϵ and ϵ'. By using Lemma 5.19, one may show that, for all $m \geq 1$, if

$$\nabla_m = \{B_{0...00}, B_{0...01}, ..., B_{1...10}, B_{1...11}\}$$

then

$$\mathcal{T}((\wp_{B_{0\ldots00}}, \wp_{B_{0\ldots01}}, \ldots, \wp_{B_{1\ldots10}}, \wp_{B_{1\ldots11}}) \in C | X_1 = x)$$
$$= E[I_C[(\tilde{Y}_0 \cdots \tilde{Y}_{0\ldots00}, \tilde{Y}_0 \cdots \tilde{Y}_{0\ldots01}, \ldots, \tilde{Y}_1 \cdots \tilde{Y}_{1\ldots10}, \tilde{Y}_1 \cdots \tilde{Y}_{1\ldots11})]]$$

for all $C \in \mathcal{B}[0,1]^{2^m}$. For illustration, we show this only for the case $m = 2$.

Fix $C \in \mathcal{B}[0,1]^4$. Then, since $\mathcal{T}$ is a binary Polya tree with parameters $(\{\nabla_m\}, \{\alpha_\epsilon\})$, it follows from Lemma 5.19 that

$$\mathcal{T}((\wp_{B_{00}}, \wp_{B_{01}}, \wp_{B_{10}}, \wp_{B_{11}}) \in C | X_1 = x)$$
$$= \frac{E(I_C[(Y_0 Y_{00}, Y_0 Y_{01}, Y_1 Y_{10}, Y_1 Y_{11}) \in C] Y_{\epsilon_1(x)} Y_{\epsilon_1(x)\epsilon_2(x)})}{E(Y_{\epsilon_1(x)} Y_{\epsilon_1(x)\epsilon_2(x)})} \tag{6.6}$$

where $B_{\epsilon_1(x)}$ and $B_{\epsilon_1(x)\epsilon_2(x)}$ are the sets in ∇_1 and ∇_2 respectively to which the point x belongs. Now, since $Y_{\epsilon_1(x)}$ and $Y_{\epsilon_1(x)\epsilon_2(x)}$ are independent and Beta distributed, one can check that

$$E(Y_{\epsilon_1(x)} Y_{\epsilon_1(x)\epsilon_2(x)}) = \frac{\alpha_{\epsilon_1(x)}}{\alpha_0 + \alpha_1} \frac{\alpha_{\epsilon_1(x)\epsilon_2(x)}}{\alpha_{\epsilon_1(x)0} + \alpha_{\epsilon_1(x)1}}.$$

Set K to be the quantity described by the previous equation. Analogously, using the fact that the random vectors (Y_0, Y_1), (Y_{00}, Y_{01}) and (Y_{10}, Y_{11}) are independent and Dirichlet distributed with parameters (α_0, α_1), $(\alpha_{00}, \alpha_{01})$ and $(\alpha_{10}, \alpha_{11})$ respectively, one shows that the numerator of the right member of (6.6) is equal to

$$KE(I_C((\tilde{Y}_0 \tilde{Y}_{00}, \tilde{Y}_0 \tilde{Y}_{01}, \tilde{Y}_1 \tilde{Y}_{10}, \tilde{Y}_1 \tilde{Y}_{11})).$$

◇

References

[1] Ash, R.B. (1972). *Real Analysis and Probability*, Academic Press, New York.

[2] de Finetti, B. (1935), Il problema della perequazione. *Atti della Società Italiana per il Progresso delle Scienze, (XII Riunione)*, Napoli.

[3] Doksum, K.A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.*,**2**, 183-201.

[4] Dubins, L. and Freedman, D. (1963). Random distribution functions. *Bull. Amer. Math. Soc.*, **69**, 548-551.

[5] Dubins, L. and Freedman, D. (1966). Random distribution functions. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.*, 2, 183–214. University of California Press.

[6] Fabius, J. (1964). Asymtotic Behavior of Bayes estimate. *Ann. Math. Statist.* **35**, 846-856.

[7] Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209-230.

[8] Ferguson, T.S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.*, **2**, 615-629.

[9] Freedman, D.A. (1963). On the asymptotic behavior of Bayes' estimate in the discrete case. *Ann. Math. Statist.* **34**, 1386-1403.

[10] Freedman, D.A. (1965). On the asymptotic behavior of Bayes' estimate in the discrete case. *Ann. Math. Statist.* **36**, 454–456.

[11] Kraft, C.H. (1964). A class of distribution function processes which have derivatives. *J. Appl. Prob.*, 1, 385–388.

[12] Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modeling. *ann. Statist.*, **20**, 1203-1221.

[13] Lavine, M. (1994). More aspects of Plya trees for statistical modelling. *Ann. Statist.*, **22**, 1164-1176.

[14] Mauldin, R.D., Sudderth, W.D. and Williams, S.C. (1992). Polya trees and random distributions. *Ann. Statist.*, **20**, 1203-1221.

[15] Rolph, J.E. (1968). Bayesian Estimation of Mixing distributions, *Annal. Math. Statist.*, **39**, 1289- 1302.

[16] Regazzini, E. and Petris, G. (1992). Some critical aspects on the use of exchangeability in statistics. *Journal of the Italian Statistical Society*, 1(1), 103–130.

[17] Regazzini, E. (1996). *Impostazione non parametrica di problemi d'inferenza statistica bayesiana*. Dottorato di Ricerca in Statistica Metodologica – Materiali Didattici N.2, Dipartimento di Statistica, Probabilità e Statistica Applicata, Università di Roma "La Sapienza".

[18] Shiryayev, A.N. (1984). *Probability*, Springer-Verlag, New York.