

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Bayesian nonparametric inference on a Frechet class

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Dreassi, E., Pratelli, L., Rigo, P. (2025). Bayesian nonparametric inference on a Frechet class. THEORY OF PROBABILITY AND MATHEMATICAL STATISTICS, 113, 51-64 [10.1090/tpms/1241].

Availability:

This version is available at: <https://hdl.handle.net/11585/1005809> since: 2025-11-19

Published:

DOI: <http://doi.org/10.1090/tpms/1241>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

BAYESIAN NONPARAMETRIC INFERENCE ON A FRÉCHET CLASS

EMANUELA DREASSI, LUCA PRATELLI, AND PIETRO RIGO

ABSTRACT. Let $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{G}, \nu)$ be probability spaces and (Z_n) be a sequence of random variables with values in $(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$. Let $\Gamma(\mu, \nu)$ be the collection of all probability measures p on $\mathcal{F} \otimes \mathcal{G}$ such that

$$p(A \times \mathcal{Y}) = \mu(A) \quad \text{and} \quad p(\mathcal{X} \times B) = \nu(B) \quad \text{for all } A \in \mathcal{F} \text{ and } B \in \mathcal{G}.$$

In this paper, we build some probability measures Π on $\Gamma(\mu, \nu)$. In addition, for each such Π , we assume that (Z_n) is exchangeable with de Finetti's measure Π and we evaluate the conditional distribution $\Pi(\cdot \mid Z_1, \dots, Z_n)$. In Bayesian nonparametrics, if $(Z_1, \dots, Z_n)$ are the available data, Π and $\Pi(\cdot \mid Z_1, \dots, Z_n)$ can be regarded as the prior and the posterior, respectively. To support this interpretation, it suffices to think of a problem where the unknown probability distribution of some bivariate phenomenon is constrained to have marginals μ and ν . Finally, analogous results are obtained for the set $\Gamma(\mu)$ of those probability measures on $\mathcal{F} \otimes \mathcal{G}$ with marginal μ on $\mathcal{F}$ (but arbitrary marginal on $\mathcal{G}$). That is, we introduce some priors on $\Gamma(\mu)$ and we evaluate the corresponding posteriors.

1. INTRODUCTION

An intriguing problem in Bayesian nonparametrics is to build a (suitable) prior distribution supported by a given subset $\mathcal{P}_0$ of the set of all probability measures on the sample space. Depending on $\mathcal{P}_0$, such a problem may be quite hard. This paper focuses on this problem when $\mathcal{P}_0$ is a Fréchet class (to be meant as in Section 1.2).

Throughout, $(\mathcal{Z}, \mathcal{H})$ is a measurable space and $(Z_n : n \geq 1)$ is a sequence of random variables with values in $(\mathcal{Z}, \mathcal{H})$. All random elements appearing in this paper are defined on the same probability space, say $(\Omega, \mathcal{A}, \mathbb{P})$. Moreover, $\mathcal{P}$ is the collection of all probability measures on $\mathcal{H}$ and Ψ is the σ -field over $\mathcal{P}$ generated by the evaluation maps $p \mapsto p(H)$ for all $H \in \mathcal{H}$.

1.1. Bayesian nonparametric inference with exchangeable data. In a nutshell, the standard framework of Bayesian nonparametrics can be summarized as follows.

Let Z be a random element, taking values in $(\mathcal{Z}, \mathcal{H})$, and let $\mathcal{L}(Z)$ denote its probability distribution. The goal is to make inference on $\mathcal{L}(Z)$, or on some functional of $\mathcal{L}(Z)$, based on the available observations $Z_1, Z_2, \dots$ on Z . To this end, some assumptions on the data sequence (Z_n) are to be made. It is quite usual, even if not mandatory, to assume (Z_n) exchangeable or partially exchangeable; see e.g. [3], [5], [6], [8], [13], [14]. In the sequel, (Z_n) is supposed to be exchangeable.

2020 *Mathematics Subject Classification.* 60G09, 60G57, 62F15, 62G99.

Key words and phrases. Bayesian nonparametrics, copula, exchangeability, Fréchet class, mass transportation, random probability measure.

We refer to [1] for an exhaustive treatment of exchangeability. Here, we just recall that (Z_n) is exchangeable if

$$(Z_{j_1}, \dots, Z_{j_n}) \sim (Z_1, \dots, Z_n)$$

for all $n \geq 1$ and all permutations $(j_1, \dots, j_n)$ of $(1, \dots, n)$. An i.i.d. sequence is obviously exchangeable while the converse is not true (just take $Z_n = V$ for all $n \geq 1$, where V is a non-degenerate random variable). However, under mild conditions on $(\mathcal{Z}, \mathcal{H})$, if (Z_n) is exchangeable there is a unique probability measure Π on Ψ such that

$$\mathbb{P}(Z_1 \in H_1, \dots, Z_k \in H_k) = \int p(H_1) \dots p(H_k) \Pi(dp)$$

for all $k \geq 1$ and $H_1, \dots, H_k \in \mathcal{H}$. Such Π is called the *de Finetti's measure* or the *prior distribution*.

If (Z_n) is exchangeable, then, for each $n \geq 1$, the sequence $(Z_{n+k} : k \geq 1)$ is still exchangeable conditionally on $(Z_1, \dots, Z_n)$. Hence, there is an a.s. unique probability measure $\Pi(\cdot \mid Z_1, \dots, Z_n)$ on Ψ such that

$$\mathbb{P}(Z_{n+1} \in H_1, \dots, Z_{n+k} \in H_k \mid Z_1, \dots, Z_n) = \int p(H_1) \dots p(H_k) \Pi(dp \mid Z_1, \dots, Z_n)$$

a.s. for all $k \geq 1$ and $H_1, \dots, H_k \in \mathcal{H}$. If $(Z_1, \dots, Z_n)$ are the available observations, $\Pi(\cdot \mid Z_1, \dots, Z_n)$ is called the *posterior distribution*.

In real problems, to make inference on $\mathcal{L}(Z)$, one first selects a prior Π and then evaluates the posterior $\Pi(\cdot \mid Z_1, \dots, Z_n)$. Clearly, evaluating the posterior may be very difficult, and one is often forced to use numerical approximations. Anyway, if and when $\Pi(\cdot \mid Z_1, \dots, Z_n)$ is available, any inference regarding $\mathcal{L}(Z)$ should be based on it. Among other things, given $\Pi(\cdot \mid Z_1, \dots, Z_n)$, one can evaluate the *predictive distribution*

$$\mathbb{P}(Z_{n+1} \in H \mid Z_1, \dots, Z_n) = \int p(H) \Pi(dp \mid Z_1, \dots, Z_n) \quad \text{for all } H \in \mathcal{H}.$$

Suppose now that $\mathcal{L}(Z)$ is known to belong to some class $\mathcal{P}_0 \in \Psi$. This actually happens in a number of situations. Some obvious examples are:

- $\mathcal{P}_0 = \{p \in \mathcal{P} : p \text{ is absolutely continuous with respect to } \lambda\}$
where λ is a reference measure on $\mathcal{H}$; see [5];
- $\mathcal{P}_0 = \{p \in \mathcal{P} : p \text{ is invariant under } G\}$
where G is a group of transformations on $\mathcal{Z}$; see [7, Example 17], [11], [15];
- $\mathcal{P}_0 = \{p \in \mathcal{P} : p = p^* \text{ on } \mathcal{H}^*\}$
where $\mathcal{H}^* \subset \mathcal{H}$ and p^* is a given element of $\mathcal{P}$;
- $\mathcal{Z} = \mathbb{R}$ and $\mathcal{P}_0 = \{p \in \mathcal{P} : p \text{ has given moments up to a certain order}\}$.

Obviously, if it is known that $\mathcal{L}(Z) \in \mathcal{P}_0$, the prior Π should satisfy

$$\Pi(\mathcal{P}_0) = 1.$$

However, to build a reasonable prior, with large support and satisfying $\Pi(\mathcal{P}_0) = 1$, is usually hard; see e.g. [11]. Similarly, once a prior is selected, to obtain the posterior may be difficult as well.

1.2. **Fréchet classes.** In the sequel,

$$(\mathcal{X}, \mathcal{F}, \mu) \quad \text{and} \quad (\mathcal{Y}, \mathcal{G}, \nu)$$

are probability spaces. To avoid annoying complications, $(\mathcal{X}, \mathcal{F})$ and $(\mathcal{Y}, \mathcal{G})$ are assumed to be *standard Borel spaces*.

Let

$$\mathcal{Z} = \mathcal{X} \times \mathcal{Y} \quad \text{and} \quad \mathcal{H} = \mathcal{F} \otimes \mathcal{G}.$$

For each $p \in \mathcal{P}$, the marginals of p are

$$p_1(A) = p(A \times \mathcal{Y}) \quad \text{and} \quad p_2(B) = p(\mathcal{X} \times B) \quad \text{for all } A \in \mathcal{F} \text{ and } B \in \mathcal{G}.$$

Obviously, p_1 and p_2 are probability measures on $\mathcal{F}$ and $\mathcal{G}$, respectively. The *Fréchet class* of the pair (μ, ν) is

$$\Gamma(\mu, \nu) = \{p \in \mathcal{P} : p \text{ has marginal } \mu \text{ on } \mathcal{F} \text{ and } \nu \text{ on } \mathcal{G}\}.$$

Similarly, in this paper,

$$\Gamma(\mu) = \{p \in \mathcal{P} : p \text{ has marginal } \mu \text{ on } \mathcal{F}\}$$

is said to be the Fréchet class of μ .

Our goal is to build some priors Π such that $\Pi(\mathcal{P}_0) = 1$ where

$$\mathcal{P}_0 = \Gamma(\mu, \nu) \quad \text{or} \quad \mathcal{P}_0 = \Gamma(\mu).$$

In addition, assuming to have an exchangeable sequence (Z_n) with one of such priors, we evaluate the corresponding posterior.

The first choice of $\mathcal{P}_0$ does not require a lot of comments. Indeed, $\Gamma(\mu, \nu)$ plays a role in a number of settings, including contingency tables, mass transportation, Wasserstein distances, statistical physics, gradient flows and their many ramifications; see e.g. [2], [17] and references therein. We just note that $\Gamma(\mu, \nu)$ reduces to the collection of bivariate copulas if $\mathcal{X} = \mathcal{Y} = [0, 1]$ and μ and ν are uniform. Hence, in this case, our problem is to build a (reasonable) prior on the set of bivariate copulas. The second choice of $\mathcal{P}_0$, while less intriguing than the first, makes some interest as well. In general, $\Gamma(\mu)$ comes into play if one marginal of $\mathcal{L}(Z)$ is known but the other is not. As an example, suppose $Z = (X, Y)$ where Y is a response variable and X is a vector of covariates. If $\mathcal{L}(X)$ is known, say $\mathcal{L}(X) = \mu$, making inference on $\mathcal{L}(Z)$ amounts to making inference on the conditional distribution $\mathcal{L}(Y | X)$. Hence, for instance, $\Gamma(\mu)$ is meaningful when $\mathcal{L}(X)$ is known and the object of inference is some functional of $\mathcal{L}(Y | X)$, such as the regression function $g(x) = E(Y | X = x)$. Note also that there are situations where $\mathcal{L}(X)$ is given while $\mathcal{L}(Y | X)$ is arbitrary. A well known example, concerning variable selection, is the knockoff procedure by Barber and Candès; see e.g. [4] and [10].

1.3. **Content of this paper.** As already noted, we aim to introduce some priors, supported by $\Gamma(\mu, \nu)$ or $\Gamma(\mu)$, and to evaluate the corresponding posteriors.

To fix ideas, let us consider $\Gamma(\mu, \nu)$ (the case $\Gamma(\mu)$ is quite analogous). A prior on $\Gamma(\mu, \nu)$ can be regarded as the probability distribution of a random probability measure (see Section 2) taking values in $\Gamma(\mu, \nu)$. Accordingly, each of our priors is the probability distribution of a measurable map $P : \Omega \rightarrow \Gamma(\mu, \nu)$, and our main task is to build such P . In Section 3, in addition to $P \in \Gamma(\mu, \nu)$, we require P to admit a density with respect to the product measure $\mu \times \nu$. This has two advantages. First, at least for us, it is easier to argue in terms of densities than in terms of probability measures. Second, if P is absolutely continuous with respect $\mu \times \nu$, we are actually dealing with a dominated statistical model. Hence, the posterior can be evaluated via Bayes theorem.

Once a random probability measure P has been built, we look for explicit formulas for its probability distribution. Our results are more or less informative depending on the different situations. In some cases, we completely obtain the finite dimensional distributions of P , that is, we evaluate the distribution of the random vector $(P(H_1), \dots, P(H_m))$ where $H_1, \dots, H_m$ is a finite measurable partition of $\mathcal{X} \times \mathcal{Y}$. In other cases, we only find the distribution of $P(H)$ for fixed $H \in \mathcal{H}$. Finally, as regards the posterior, we obtain the conditional distribution of $(P(H_1), \dots, P(H_m))$ or $P(H)$ given $(Z_1, \dots, Z_n)$.

2. RANDOM PROBABILITY MEASURES

In this section, we just report some well known definitions. Recall that $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space and $(\mathcal{Z}, \mathcal{H})$ a measurable space. Moreover, $\mathcal{P}$ is the set of all probability measures on $\mathcal{H}$ and Ψ the σ -field over $\mathcal{P}$ generated by the evaluation maps $p \mapsto p(H)$ for all $H \in \mathcal{H}$.

A *random probability measure* (r.p.m.) on $(\mathcal{Z}, \mathcal{H})$ is a measurable map

$$P : (\Omega, \mathcal{A}) \rightarrow (\mathcal{P}, \Psi).$$

Equivalently, an r.p.m. is a map $P : \Omega \rightarrow \mathcal{P}$ such that $\omega \mapsto P(\omega, H)$ is a real random variable for fixed $H \in \mathcal{H}$. Obviously, the distribution of an r.p.m. P is the probability measure Π on Ψ defined as

$$\Pi(L) = \mathbb{P}(P \in L) \quad \text{for } L \in \Psi.$$

In the sequel, $P(H)$ denotes the real random variable $\omega \mapsto P(\omega, H)$.

Let $\alpha_1, \dots, \alpha_m$ be non-negative constants such that $\sum_{i=1}^m \alpha_i > 0$ and let $U_1, \dots, U_m$ be independent Gamma random variables with scale parameter 1 and shape parameter α_i (with $U_i = 0$ a.s. if $\alpha_i = 0$). The Dirichlet law of parameters $\alpha_1, \dots, \alpha_m$, denoted

$$\text{Dir}(\alpha_1, \dots, \alpha_m),$$

is the probability distribution of the vector

$$\left(\frac{U_1}{\sum_{i=1}^m U_i}, \dots, \frac{U_m}{\sum_{i=1}^m U_i} \right).$$

Let $c > 0$ be a constant and $\lambda \in \mathcal{P}$. A r.p.m. P is said to be Dirichlet with parameters c and λ if

$$(P(H_1), \dots, P(H_m)) \sim \text{Dir}[c\lambda(H_1), \dots, c\lambda(H_m)]$$

whenever $H_1, \dots, H_m$ is a finite measurable partition of $\mathcal{Z}$. A Dirichlet prior is the probability distribution of a Dirichlet r.p.m. Dirichlet priors are a basic ingredient of Bayesian nonparametrics. They also play a role in many other frameworks, including population genetics and species sampling. Here, we just recall their conjugacy property. Suppose (Z_n) is exchangeable and the prior Π is Dirichlet with parameters c and λ . Then, the posterior $\Pi(\cdot \mid Z_1, \dots, Z_n)$ is still Dirichlet with parameters $c + n$ and $\frac{c\lambda + \sum_{i=1}^n \delta_{Z_i}}{c+n}$, where δ_z denotes the unit mass at z for every $z \in \mathcal{Z}$. See e.g. [9], [12], [18].

Despite their various merits, however, Dirichlet priors do not work for our purposes. Suppose in fact $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, $\mathcal{H} = \mathcal{F} \otimes \mathcal{G}$, and Π is a Dirichlet prior. The parameter λ of Π is a probability measure on $\mathcal{F} \otimes \mathcal{G}$. If λ is not trivial on $\mathcal{F}$, then

$$\Pi(\Gamma(\mu, \nu)) = \Pi(\Gamma(\mu)) = 0.$$

To see this, take a Dirichlet r.p.m. P with distribution Π . Since λ is not trivial on $\mathcal{F}$, there is $A \in \mathcal{F}$ such that $\lambda(A \times \mathcal{Y}) \in (0, 1)$. Hence, $P(A \times \mathcal{Y})$ has a beta distribution, which in turn implies

$$\Pi(\Gamma(\mu, \nu)) \leq \Pi(\Gamma(\mu)) = \mathbb{P}(P \in \Gamma(\mu)) \leq \mathbb{P}[P(A \times \mathcal{Y}) = \mu(A)] = 0.$$

Finally, a *copula* is a distribution function on $\mathbb{R}^n$ whose marginals are uniform on $[0, 1]$. If C is a copula and $F_1, \dots, F_n$ are distribution functions on $\mathbb{R}$, then

$$(2.1) \quad F(x_1, \dots, x_n) = C[F_1(x_1), \dots, F_n(x_n)], \quad (x_1, \dots, x_n) \in \mathbb{R}^n,$$

is a distribution function on $\mathbb{R}^n$ with marginals $F_1, \dots, F_n$. Conversely, for each distribution function F on $\mathbb{R}^n$ (with marginals $F_1, \dots, F_n$) there is a copula C satisfying equation (2.1). Moreover, such C is unique whenever $F_1, \dots, F_n$ are continuous. In the sequel, we are concerned with bivariate copulas (i.e., $n = 2$).

3. SOME PRIORS ON $\Gamma(\mu, \nu)$

Let again $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ and $\mathcal{H} = \mathcal{F} \otimes \mathcal{G}$. In this section, we denote by λ the product measure

$$\lambda = \mu \times \nu$$

and we focus on

$$\mathcal{P}_0 = \{p \in \Gamma(\mu, \nu) : p \text{ is absolutely continuous with respect to } \lambda\}.$$

Our goal is to introduce some priors Π such that $\Pi(\mathcal{P}_0) = 1$ and to obtain the corresponding posteriors.

Some more notation is needed. Let

$$\mathcal{D} = \{f : f \text{ is a probability density with respect to } \lambda \text{ and } P_f \in \Gamma(\mu, \nu)\}$$

where P_f is the probability measure

$$P_f(H) = \int_H f d\lambda \quad \text{for all } H \in \mathcal{H}.$$

The set $\mathcal{D}$ is equipped with the σ -field Ψ^* generated by the map $f \mapsto P_f$. Precisely, Ψ^* is the collection of events of the form $\{f \in \mathcal{D} : P_f \in L\}$ for some $L \in \Psi$. A *random density* is a random element of $(\mathcal{D}, \Psi^*)$. Moreover, we denote by

$$Z_n = (X_n, Y_n), \quad n \geq 1,$$

a sequence of random variables taking values in $(\mathcal{X} \times \mathcal{Y}, \mathcal{H})$.

One advantage of $\mathcal{P}_0$ is that it is a *dominated statistical model*, in the sense that each element of $\mathcal{P}_0$ is absolutely continuous with respect to a fixed probability measure on $\mathcal{H}$. Hence, if the prior is concentrated on $\mathcal{P}_0$, the posterior can be obtained via Bayes theorem. This is formalized in the next result.

Lemma 3.1. *(Z_n) is exchangeable and the prior is concentrated on $\mathcal{P}_0$ if and only if there is a random density f such that*

$$(3.1) \quad (Z_n) \mid f \text{ is i.i.d. with } Z_1 \sim f.$$

Moreover, if Π denotes the probability distribution of f , one obtains

$$\Pi(df \mid Z_1, \dots, Z_n) = \frac{\prod_{i=1}^n f(Z_i)}{\int \prod_{i=1}^n \phi(Z_i) \Pi(d\phi)} \Pi(df) \quad \text{a.s.}$$

whenever $\int \prod_{i=1}^n \phi(Z_i) \Pi(d\phi) \in (0, \infty)$. (Here, Π and $\Pi(\cdot \mid Z_1, \dots, Z_n)$ are probability measures on $(\mathcal{D}, \Psi^*)$).

Proof. The “if” part is trivial. Conversely, since $(\mathcal{X}, \mathcal{F})$ and $(\mathcal{Y}, \mathcal{G})$ are standard Borel spaces, if (Z_n) is exchangeable there is an r.p.m. P on $(\mathcal{X} \times \mathcal{Y}, \mathcal{H})$ such that

$$(Z_n) \mid P \text{ is i.i.d. with } Z_1 \sim P.$$

Since the prior is concentrated on $\mathcal{P}_0$, one also obtains $P \in \mathcal{P}_0$ a.s. Hence, to get (3.1), it suffices to take f as the density of P with respect to λ . Finally, as noted above, since $\mathcal{P}_0$ is a dominated statistical model, the posterior can be evaluated via Bayes theorem. Precisely, assume condition (3.1), call Π the probability distribution of the random density f , and define

$$\Pi^*(df) = \frac{\prod_{i=1}^n f(Z_i)}{\int \prod_{i=1}^n \phi(Z_i) \Pi(d\phi)} \Pi(df) \quad \text{whenever} \quad \int \prod_{i=1}^n \phi(Z_i) \Pi(d\phi) \in (0, \infty).$$

Such Π^* is an r.p.m. on $(\mathcal{D}, \Psi^*)$. Moreover, using Fubini’s theorem, it is straightforward to show that

$$\int P_f(H_1) \dots P_f(H_k) \Pi^*(df) = \mathbb{P}(Z_{n+1} \in H_1, \dots, Z_{n+k} \in H_k \mid Z_1, \dots, Z_n)$$

a.s. for all $k \geq 1$ and $H_1, \dots, H_k \in \mathcal{H}$. Therefore, $\Pi(\cdot \mid Z_1, \dots, Z_n) = \Pi^*$ a.s. $\square$

An r.p.m. P on $(\mathcal{X} \times \mathcal{Y}, \mathcal{H})$ such that $P \in \mathcal{P}_0$ a.s. can be identified with a random density. Similarly, a prior on $\mathcal{P}_0$ can be identified with the probability distribution of a random density. Accordingly, in the rest of this section, we will speak of random densities rather than priors on $\mathcal{P}_0$. Importantly, by Lemma 3.1, once a random density has been defined, the posterior can be written in closed form. Hence, from a Bayesian perspective, it remains only to introduce some (suitable) random densities. We will do it following two different approaches.

3.1. First approach. Let

$$f(x, y) = 1 + g(x) h(y) \quad \text{for all } (x, y) \in \mathcal{X} \times \mathcal{Y},$$

where

$$(3.2) \quad g : \mathcal{X} \rightarrow \mathbb{R} \text{ and } h : \mathcal{Y} \rightarrow \mathbb{R} \text{ are Borel functions such that} \\ \sup |g| \leq 1, \quad \sup |h| \leq 1 \quad \text{and} \quad E_\mu(g) = E_\nu(h) = 0.$$

Then, f is a non-negative Borel function on $\mathcal{X} \times \mathcal{Y}$. In addition, for all $A \in \mathcal{F}$ and $B \in \mathcal{G}$, Fubini’s theorem yields

$$\begin{aligned} \int_{A \times \mathcal{Y}} f d\lambda &= \mu(A) + E_\mu(g \mathbb{1}_A) E_\nu(h) = \mu(A), \\ \int_{\mathcal{X} \times B} f d\lambda &= \nu(B) + E_\mu(g) E_\nu(h \mathbb{1}_B) = \nu(B). \end{aligned}$$

Therefore, f is a probability density with respect to λ and $P_f \in \Gamma(\mu, \nu)$, that is, $f \in \mathcal{D}$. Similarly, one still obtains $f \in \mathcal{D}$ if f is defined as

$$f(x, y) = 1 + \sum_{n=1}^{\infty} u_n g_n(x) h_n(y)$$

where u_n is a constant, $\sum_n |u_n| \leq 1$, and the pair (g_n, h_n) satisfies condition (3.2).

Based on the previous remarks, to build a random density, it suffices to select g_n , h_n and u_n randomly. A straightforward way to do it is to let

$$(3.3) \quad f(x, y) = 1 + \sum_{n=1}^{\infty} U_n g_n(x) h_n(y)$$

where g_n and h_n are fixed deterministic functions satisfying condition (3.2) while the U_n are real random variables such that $\sum_n |U_n| \leq 1$ a.s. Then, for each $H \in \mathcal{H}$,

$$P_f(H) = \lambda(H) + \sum_{n=1}^{\infty} a_n(H) U_n \quad \text{where } a_n(H) = \int_H g_n(x) h_n(y) \lambda(dx, dy).$$

Hence, the distribution of $P_f(H)$ is available at least when the random variables U_n are independent. In this case, in fact,

$$E\left\{\exp(it P_f(H))\right\} = \exp(it \lambda(H)) \prod_{n=1}^{\infty} \phi_n(a_n(H) t) \quad \text{for all } t \in \mathbb{R}$$

where ϕ_n is the characteristic function of U_n .

In passing, we also note that, if $(\mathcal{X}, \mathcal{F}) = (\mathcal{Y}, \mathcal{G})$ and $\mu = \nu$, it is possible to let $g_n = h_n$ for all n . In this case,

$$P_f(A \times B) = P_f(B \times A) \quad \text{for all } A, B \in \mathcal{F}$$

so that the r.p.m. P_f is exchangeable. Hence, if $\mu = \nu$ and $g_n = h_n$ for all n , one a.s. selects exchangeable elements of $\Gamma(\mu, \mu)$.

The random density (3.3) is very simple but, just for this reason, potentially useful in real problems. We now turn to a more involved random density.

Say that F is a *random distribution function* on $\mathbb{R}^n$ if

$$F(x_1, \dots, x_n) = P\left((-\infty, x_1] \times \dots \times (-\infty, x_n]\right), \quad (x_1, \dots, x_n) \in \mathbb{R}^n,$$

for some r.p.m. P on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Write P_F to denote such P and m to indicate the Lebesgue measure on $[0, 1]$. A *random bivariate copula* is a random distribution function C on $\mathbb{R}^2$ such that $P_C \in \Gamma(m, m)$ a.s.

We now assume

$$\mathcal{X} = \mathcal{Y} = [0, 1] \quad \text{and} \quad \mu = \nu = m.$$

However, the results obtained in this case can be easily extended to $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. In fact, if μ and ν are Borel probabilities on $\mathbb{R}$ and C is a random bivariate copula, then

$$F(x, y) = C[F_\mu(x), F_\nu(y)], \quad (x, y) \in \mathbb{R}^2,$$

is a random distribution function on $\mathbb{R}^2$ with marginals F_μ and F_ν , where

$$F_\mu(x) = \mu((-\infty, x]) \quad \text{and} \quad F_\nu(y) = \nu((-\infty, y]).$$

Hence, any random bivariate copula C induces a random distribution function F on $\mathbb{R}^2$ such that $P_F \in \Gamma(\mu, \nu)$. Equivalently, any prior on bivariate copulas induces a prior on $\Gamma(\mu, \nu)$.

Let $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ be a bounded Borel function and $W_i = \{W_i(t) : t \in [0, 1]\}$ be a real stochastic process with Borel paths, where $i = 1, 2$. Suppose also that

$$(3.4) \quad \sup_{t \in [0, 1]} \left| \varphi(W_i(t)) - \int_0^1 \varphi(W_i(s)) ds \right| \leq 1 \quad \text{for } i = 1, 2.$$

Under these conditions, define

$$g(t) = \varphi(W_1(t)) - \int_0^1 \varphi(W_1(s)) ds \quad \text{and} \quad h(t) = \varphi(W_2(t)) - \int_0^1 \varphi(W_2(s)) ds.$$

Our next result deals with the random density $f(x, y) = 1 + g(x)h(y)$ when W_1 and W_2 are standard Brownian motions. We denote by

$$L_i = \{L_i(t, x) : t \in [0, 1], x \in \mathbb{R}\}$$

the local time process corresponding to W_i . Here, L_i is meant according to [16, p. 201]. Among other things, L_i has non-negative continuous paths and the map $x \mapsto 2L_i(t, x)$ is a density of the occupation measure of W_i on $[0, t]$. The latter measure is defined as

$$\mathcal{O}_{i,t}(A) = \int_0^t \mathbb{1}_A(W_i(s)) ds \quad \text{for all } A \in \mathcal{B}(\mathbb{R}).$$

Theorem 3.2. *Let W_1 and W_2 be standard Brownian motions and $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ be a Borel function such that $-1/2 \leq \varphi \leq 1/2$ or $0 \leq \varphi \leq 1$. Define*

$$f(x, y) = 1 + \left\{ \varphi(W_1(x)) - \int_0^1 \varphi(W_1(s)) ds \right\} \left\{ \varphi(W_2(y)) - \int_0^1 \varphi(W_2(s)) ds \right\}$$

for all $(x, y) \in [0, 1]^2$. Then, f is a random density and

$$P_f([0, a] \times [0, b]) = ab + 4U_1(a)U_2(b)$$

for all $0 \leq a, b \leq 1$, where

$$U_i(t) = \int_{-\infty}^{\infty} \{L_i(t, x) - tL_i(1, x)\} \varphi(x) dx \quad \text{for } i = 1, 2 \text{ and } t \in [0, 1].$$

Proof. Condition (3.4) holds since $-1/2 \leq \varphi \leq 1/2$ or $0 \leq \varphi \leq 1$. Hence, f is a random density. Let $\mathcal{O}_{i,t}$ be the occupation measure of W_i on $[0, t]$. Since $2L_i(t, \cdot)$ is a density of $\mathcal{O}_{i,t}$, one obtains

$$\int_0^t \varphi(W_i(s)) ds = \int_{-\infty}^{\infty} \varphi(x) \mathcal{O}_{i,t}(dx) = 2 \int_{-\infty}^{\infty} L_i(t, x) \varphi(x) dx.$$

Therefore,

$$\begin{aligned} P_f([0, a] \times [0, b]) &= \int_0^a \int_0^b f(x, y) dx dy \\ &= ab + \left\{ \int_0^a \varphi(W_1(s)) ds - a \int_0^1 \varphi(W_1(s)) ds \right\} \left\{ \int_0^b \varphi(W_2(s)) ds - b \int_0^1 \varphi(W_2(s)) ds \right\} \\ &= ab + 4 \int_{-\infty}^{\infty} \{L_1(a, x) - aL_1(1, x)\} \varphi(x) dx \int_{-\infty}^{\infty} \{L_2(b, x) - bL_2(1, x)\} \varphi(x) dx. \end{aligned}$$

□

The random variables $U_1(a)$ and $U_2(b)$ are independent whenever W_1 and W_2 are independent. In this case, to evaluate the distribution of $P_f([0, a] \times [0, b])$, it suffices to know the distribution of $U_i(t)$. Possibly, the latter distribution is known, at least if φ has a simple form, but we are not aware of this fact.

3.2. Second approach. We now develop a different idea to build a random density. We still assume $\mathcal{X} = \mathcal{Y} = [0, 1]$ and $\mu = \nu = m$ where m is the Lebesgue measure on $[0, 1]$. As explained in Section 3.1, however, any result obtained in this case extends easily to $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. Note also that $\lambda = \mu \times \nu = m \times m$ is Lebesgue measure on $[0, 1]^2$.

Fix an integer $k \geq 1$ and define the intervals

$$I_j = \left(\frac{j-1}{k}, \frac{j}{k} \right] \quad \text{for } j = 1, \dots, k.$$

Call Σ the set of all permutations $\sigma = (\sigma(1), \dots, \sigma(k))$ of $(1, \dots, k)$. For each $\sigma \in \Sigma$, denote by f_σ the uniform density (with respect to λ) on the set

$$S_\sigma = \bigcup_{j=1}^k I_j \times I_{\sigma(j)}.$$

Hence,

$$f_\sigma(x, y) = k \mathbb{1}_{S_\sigma}(x, y) = k \sum_{j=1}^k \mathbb{1}_{I_j}(x) \mathbb{1}_{I_{\sigma(j)}}(y).$$

For any Borel set $A \subset [0, 1]$,

$$\begin{aligned} \int_{A \times [0, 1]} f_\sigma d\lambda &= k \lambda(S_\sigma \cap (A \times [0, 1])) = \sum_{j=1}^k m(A \cap I_j) = m(A), \\ \int_{[0, 1] \times A} f_\sigma d\lambda &= k \lambda(S_\sigma \cap ([0, 1] \times A)) = \sum_{j=1}^k m(A \cap I_{\sigma(j)}) = m(A). \end{aligned}$$

Therefore, $P_{f_\sigma} \in \Gamma(m, m)$ or, equivalently, $f_\sigma \in \mathcal{D}$. Since $\mathcal{D}$ is a convex set, one also obtains

$$\sum_{\sigma \in \Sigma} u_\sigma f_\sigma \in \mathcal{D}$$

provided each u_σ is a non-negative constant and $\sum_{\sigma \in \Sigma} u_\sigma = 1$.

Based on the previous remarks, to get a random density, it suffices to select the weights u_σ in a random way. To do this, we write Σ as

$$\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_{k!}\}$$

and we fix a random vector $U = (U_1, \dots, U_{k!})$ satisfying $U_i \geq 0$ for all i and $\sum_{i=1}^{k!} U_i = 1$ a.s. In this notation, we focus on the random density

$$(3.5) \quad f = \sum_{i=1}^{k!} U_i f_{\sigma_i}.$$

Theorem 3.3. *If f is defined by (3.5), then*

$$P_f([0, a] \times [0, b]) = \sum_{i=1}^{k!} U_i \left\{ \sum_{j=1}^{[ka]} m(I_{\sigma_i(j)} \cap [0, b]) + (ka - [ka]) m(I_{\sigma_i([ka]+1)} \cap [0, b]) \right\}$$

for all $a, b \in [0, 1]$. Moreover, f can be written as

$$f(x, y) = k \sum_{j,h=1}^k D_{j,h} \mathbb{1}_{I_j}(x) \mathbb{1}_{I_h}(y) \quad \text{where} \quad D_{j,h} = \sum_{i: \sigma_i(j)=h} U_i.$$

Proof. Just note that

$$\begin{aligned}
P_f([0, a] \times [0, b]) &= \sum_{i=1}^{k!} U_i P_{f_{\sigma_i}}([0, a] \times [0, b]) = k \sum_{i=1}^{k!} U_i \lambda(S_{\sigma_i} \cap ([0, a] \times [0, b])) \\
&= k \sum_{i=1}^{k!} U_i \sum_{j=1}^k m(I_j \cap [0, a]) m(I_{\sigma_i(j)} \cap [0, b]) \\
&= \sum_{i=1}^{k!} U_i \left\{ \sum_{j=1}^{[ka]} m(I_{\sigma_i(j)} \cap [0, b]) + (ka - [ka]) m(I_{\sigma_i([ka]+1)} \cap [0, b]) \right\}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{1}{k} f(x, y) &= \sum_{i=1}^{k!} U_i \sum_{j=1}^k \mathbb{1}_{I_j}(x) \mathbb{1}_{I_{\sigma_i(j)}}(y) = \sum_{j=1}^k \mathbb{1}_{I_j}(x) \sum_{i=1}^{k!} U_i \mathbb{1}_{I_{\sigma_i(j)}}(y) \\
&= \sum_{j=1}^k \mathbb{1}_{I_j}(x) \sum_{h=1}^k \mathbb{1}_{I_h}(y) \sum_{i: \sigma_i(j)=h} U_i = \sum_{j,h=1}^k D_{j,h} \mathbb{1}_{I_j}(x) \mathbb{1}_{I_h}(y).
\end{aligned}$$

□

By Theorem 3.3, for simple choices of the vector U , the probability distribution of $P_f([0, a] \times [0, b])$ can be written in closed form. For instance, one could select a proper subset of permutations, say $\{\sigma_1, \dots, \sigma_m\}$, and define

$$U_{m+1} = \dots = U_{k!} = 0 \quad \text{and} \quad (U_1, \dots, U_m) \sim \text{Dir}(\alpha_1, \dots, \alpha_m)$$

for some $\alpha_1, \dots, \alpha_m > 0$. Note also that, if ka and kb are both integers, Theorem 3.3 yields

$$P_f([0, a] \times [0, b]) = \frac{1}{k} \sum_{i=1}^{k!} \text{card}\{j : 1 \leq j \leq ka \text{ and } \sigma_i(j) \leq kb\} U_i.$$

Finally, the random density f could be generalized by replacing k with an integer-valued random variable K independent of U . We don't develop this hint but we note that, if $\mathbb{P}(K = k) > 0$ for each $k \geq 1$, the probability distribution of P_f has a very large support. To motivate the latter (vague) claim, we prove that any element of $\Gamma(m, m)$ can be weakly approximated by P_g for some density $g \in \mathcal{D}$ of the form

$$(3.6) \quad g(x, y) = \sum_{j,h=1}^k d_{j,h} \mathbb{1}_{I_j}(x) \mathbb{1}_{I_h}(y)$$

where the $d_{j,h}$ are suitable constants. In the next result, d_{BL} denotes the *bounded Lipschitz metric* between Borel probabilities on $[0, 1]$, that is

$$d_{BL}(p, q) = \sup_{\phi} \left| \int \phi dp - \int \phi dq \right|$$

where p and q are Borel probabilities on $[0, 1]$ and $\sup$ is over the 1-Lipschitz functions $\phi : [0, 1] \rightarrow [-1, 1]$.

Theorem 3.4. *Given $k \geq 1$ and $p \in \Gamma(m, m)$, define g by (3.6) where*

$$d_{j,h} = k^2 p(I_j \times I_h).$$

Then,

$$g \in \mathcal{D} \quad \text{and} \quad d_{BL}(p, P_g) \leq \frac{2\sqrt{2}}{k}.$$

Proof. Since $p \in \Gamma(m, m)$, it is straightforward to check that

$$P_g(A \times [0, 1]) = P_g([0, 1] \times A) = m(A)$$

for each Borel set $A \subset [0, 1]$. Hence, $g \in \mathcal{D}$. Next, fix a 1-Lipschitz function $\phi : [0, 1] \rightarrow [-1, 1]$ and define

$$\alpha = \sum_{j,h=1}^k \phi(j/k, h/k) p(I_j \times I_h).$$

Since ϕ is 1-Lipschitz,

$$|\phi(x, y) - \phi(j/k, h/k)| \leq \|(x, y) - (j/k, h/k)\| \leq \frac{\sqrt{2}}{k} \quad \text{for all } (x, y) \in I_j \times I_h$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^2$. Therefore,

$$\begin{aligned} \left| \int \phi dp - \alpha \right| &= \left| \sum_{j,h=1}^k \int_{I_j \times I_h} \{\phi(x, y) - \phi(j/k, h/k)\} p(dx, dy) \right| \\ &\leq \sum_{j,h=1}^k \sup_{(x,y) \in I_j \times I_h} |\phi(x, y) - \phi(j/k, h/k)| p(I_j \times I_h) \leq \frac{\sqrt{2}}{k}. \end{aligned}$$

Since $P_g(I_j \times I_h) = p(I_j \times I_h)$ for all j and h , one similarly obtains

$$\left| \int \phi dP_g - \alpha \right| \leq \frac{\sqrt{2}}{k}.$$

Hence,

$$\left| \int \phi dp - \int \phi dP_g \right| \leq \left| \int \phi dp - \alpha \right| + \left| \alpha - \int \phi dP_g \right| \leq \frac{2\sqrt{2}}{k}.$$

□

4. DIRICHLET TYPE PRIORS ON $\Gamma(\mu)$

To build a prior on $\Gamma(\mu)$, a naive idea is to fix an r.p.m. Q on $(\mathcal{Y}, \mathcal{G})$ and let

$$P = \mu \times Q.$$

Then, P is an r.p.m. on $(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$ and $P(A \times \mathcal{Y}) = \mu(A)$ for all $A \in \mathcal{F}$, so that the probability distribution Π of P satisfies $\Pi(\Gamma(\mu)) = 1$. It is tempting to take Q as a Dirichlet r.p.m. In this case, P is basically a mixture of Dirichlet r.p.m.'s as defined in [3]. The main features of P are collected in the next result. To state it, for all $x \in \mathcal{X}$ and $H \subset \mathcal{X} \times \mathcal{Y}$, we denote by

$$H^x = \{y \in \mathcal{Y} : (x, y) \in H\}$$

the section of H with respect to x .

Theorem 4.1. *Let $P = \mu \times Q$, where Q is a Dirichlet r.p.m. on $(\mathcal{Y}, \mathcal{G})$ with parameters c and ν (as defined in Section 2). Moreover, let $Z_n = (X_n, Y_n)$, $n \geq 1$, be an exchangeable sequence such that*

$$(Z_n) \mid P \text{ is i.i.d. with } Z_1 \sim P.$$

Then:

(a) For each finite measurable partition $H_1, \dots, H_m$ of $\mathcal{X} \times \mathcal{Y}$,

$$(P(H_1), \dots, P(H_m)) \sim \int_{\mathcal{X}} \text{Dir}[c\nu(H_1^x), \dots, c\nu(H_m^x)] \mu(dx);$$

(b) For each $n \geq 1$, conditionally on $(Z_1, \dots, Z_n)$, one obtains

$$(P(H_1), \dots, P(H_m)) \sim \int_{\mathcal{X}} \text{Dir}[\nu_n(H_1^x), \dots, \nu_n(H_m^x)] \mu(dx)$$

where $H_1, \dots, H_m$ is a finite measurable partition of $\mathcal{X} \times \mathcal{Y}$ and ν_n is the random measure

$$\nu_n = c\nu + \sum_{i=1}^n \delta_{Y_i};$$

(c) For all $A \in \mathcal{F}$ and $B \in \mathcal{G}$,

$$\mathbb{P}(X_{n+1} \in A, Y_{n+1} \in B \mid Z_1, \dots, Z_n) = \mu(A) \frac{\nu_n(B)}{c+n} \quad a.s.$$

Proof. First note that $H_1^x, \dots, H_m^x$ is a finite measurable partition of $\mathcal{Y}$ for each $x \in \mathcal{X}$. Since Q is Dirichlet with parameters c and ν , it follows that

$$(Q(H_1^x), \dots, Q(H_m^x)) \sim \text{Dir}[c\nu(H_1^x), \dots, c\nu(H_m^x)].$$

Moreover, $P(H) = \int_{\mathcal{X}} Q(H^x) \mu(dx)$ for each $H \in \mathcal{F} \otimes \mathcal{G}$. Hence,

$$\begin{aligned} (P(H_1), \dots, P(H_m)) &= \left(\int_{\mathcal{X}} Q(H_1^x) \mu(dx), \dots, \int_{\mathcal{X}} Q(H_m^x) \mu(dx) \right) \\ &= \int_{\mathcal{X}} (Q(H_1^x), \dots, Q(H_m^x)) \mu(dx) \sim \int_{\mathcal{X}} \text{Dir}[c\nu(H_1^x), \dots, c\nu(H_m^x)] \mu(dx). \end{aligned}$$

This proves point (a). As to (b), it suffices noting that:

– Conditionally on $(Y_1, \dots, Y_n)$, the r.p.m. Q is Dirichlet with parameters $(c+n)$ and $\frac{\nu_n}{c+n}$; Hence, conditionally on $(Y_1, \dots, Y_n)$, one obtains

$$(Q(H_1^x), \dots, Q(H_m^x)) \sim \text{Dir}[\nu_n(H_1^x), \dots, \nu_n(H_m^x)];$$

– P is conditionally independent of $(X_1, \dots, X_n)$ given $(Y_1, \dots, Y_n)$.

Taking these two facts into account, point (b) can be proved exactly as (a). Finally, to get (c), consider a sequence

$$Z_n^* = (X_n^*, Y_n^*)$$

of random variables with values in $\mathcal{X} \times \mathcal{Y}$ and predictive distributions

$$\mathbb{P}(Z_{n+1}^* \in \cdot \mid Z_1^*, \dots, Z_n^*) = \frac{c(\mu \times \nu) + \sum_{i=1}^n (\mu \times \delta_{Y_i^*})}{c+n}.$$

By the results of [7] (see Theorems 7 and 9), (Z_n^*) is exchangeable and has the same prior as (Z_n) . (That is, the prior of both (Z_n^*) and (Z_n) is the probability distribution of P). Therefore, $(Z_n) \sim (Z_n^*)$, which in turn implies

$$\mathbb{P}(X_{n+1} \in A, Y_{n+1} \in B \mid Z_1, \dots, Z_n) = \frac{c\mu(A)\nu(B) + \mu(A) \sum_{i=1}^n \delta_{Y_i}(B)}{c+n} = \mu(A) \frac{\nu_n(B)}{c+n}.$$

□

The idea underlying Theorem 4.1 can be developed further. Suppose

$$\mathcal{X} = \mathcal{Y} = \mathbb{R}$$

and denote again by $F_\mu(t) = \mu((-\infty, t])$ and $F_\nu(t) = \nu((-\infty, t])$ the distribution functions corresponding to μ and ν . Moreover, fix a bivariate copula C , a random distribution function G on $\mathbb{R}$, and define

$$(4.1) \quad F(x, y) = C[F_\mu(x), G(y)] \quad \text{for all } (x, y) \in \mathbb{R}^2.$$

Then, F is a random distribution function on $\mathbb{R}^2$ and $P_F \in \Gamma(\mu)$ by construction. (Recall that P_F denotes the r.p.m. induced by F). If C is the product copula, equation (4.1) yields $F(x, y) = F_\mu(x)G(y)$ so that $P_F = \mu \times P_G$. Hence, in the special case where C is the product copula and P_G is Dirichlet, P_F reduces to the r.p.m. involved in Theorem 4.1.

Theorem 4.2. *Define F by (4.1) where P_G is Dirichlet with parameters c and ν . Fix $(x, y) \in \mathbb{R}^2$ such that $0 < F_\mu(x), F_\nu(y) < 1$ and denote by $\mathbb{B}_y$ the beta distribution function with parameters $cF_\nu(y)$ and $c(1 - F_\nu(y))$. Then,*

$$\mathbb{P}(F(x, y) \leq a) = \mathbb{B}_y[r(x, a)] \quad \text{for all } a \in [0, F_\mu(x)]$$

where

$$r(x, a) = \sup\{v \in [0, 1] : C[F_\mu(x), v] = a\}.$$

Moreover, if $Z_n = (X_n, Y_n)$, $n \geq 1$, is an exchangeable sequence such that

$$(Z_n) \mid F \text{ is i.i.d. with } Z_1 \sim F,$$

then

$$\mathbb{P}(F(x, y) \leq a \mid Y_1, \dots, Y_n) = \mathbb{B}_{n,y}[r(x, a)] \quad \text{a.s. for all } n \geq 1,$$

where $\mathbb{B}_{n,y}$ is a beta random distribution function with parameters

$$cF_\nu(y) + \sum_{i=1}^n \mathbb{1}_{\{Y_i \leq y\}} \quad \text{and} \quad c(1 - F_\nu(y)) + \sum_{i=1}^n \mathbb{1}_{\{Y_i > y\}}.$$

Proof. For each $v \in [0, 1]$, define

$$\phi_x(v) = C[F_\mu(x), v].$$

Then, ϕ_x is a non-decreasing continuous map from $[0, 1]$ onto $[0, F_\mu(x)]$. Moreover,

$$\{F(x, y) \leq a\} = \{\phi_x(G(y)) \leq a\} = \{G(y) \leq r(x, a)\} \quad \text{for each } a \in [0, F_\mu(x)].$$

Since P_G is Dirichlet with parameters c and ν , then $G(y)$ has a beta distribution with parameters $cF_\nu(y)$ and $c(1 - F_\nu(y))$. Therefore,

$$\mathbb{P}(F(x, y) \leq a) = \mathbb{P}(G(y) \leq r(x, a)) = \mathbb{B}_y[r(x, a)] \quad \text{for each } a \in [0, F_\mu(x)].$$

Next, conditionally on $(Y_1, \dots, Y_n)$, P_G is Dirichlet with parameters $(n + c)$ and $\frac{\nu_n}{n+c}$ where $\nu_n = c\nu + \sum_{i=1}^n \delta_{Y_i}$. Hence, $G(y)$ has a beta distribution with parameters

$$\nu_n((-\infty, y]) = cF_\nu(y) + \sum_{i=1}^n \mathbb{1}_{\{Y_i \leq y\}} \quad \text{and} \quad \nu_n((y, \infty)) = c(1 - F_\nu(y)) + \sum_{i=1}^n \mathbb{1}_{\{Y_i > y\}}.$$

Therefore,

$$\mathbb{P}(F(x, y) \leq a \mid Y_1, \dots, Y_n) = \mathbb{P}(G(y) \leq r(x, a) \mid Y_1, \dots, Y_n) = \mathbb{B}_{n,y}[r(x, a)] \quad \text{a.s.}$$

□

It is worth noting that Theorem 4.2 provides the conditional distribution of F given $(Y_1, \dots, Y_n)$ but not the conditional distribution of F given $(Z_1, \dots, Z_n)$. In fact, we did not find any (useful) general formula for the latter.

In Theorem 4.2, C is a *fixed* bivariate copula. An obvious development is to allow C to be random. For instance, C could be distributed according to Section 3. Anyway, if C is random, $r(x, a)$ is random as well (as it depends on C). To make this fact explicit, in our last result, we write $r_C(x, a)$ instead of $r(x, a)$.

Corollary 4.3. *Define F by (4.1) where P_G is Dirichlet with parameters c and ν and C is a random bivariate copula. Suppose C is independent of G and fix $(x, y) \in \mathbb{R}^2$ such that $0 < F_\mu(x)$, $F_\nu(y) < 1$. Then,*

$$\mathbb{P}(F(x, y) \leq a) = \int \mathbb{B}_y[r_C(x, a)] \Pi(dC) \quad \text{for all } a \in [0, F_\mu(x)]$$

where Π denotes the probability distribution of C .

Proof. Since C and G are independent, it suffices to apply Theorem 4.2. $\square$

A last remark is that the results of this section (suitably adapted) are still true if the Dirichlet is replaced by any r.p.m. with a known probability distribution.

BIBLIOGRAPHY

1. Aldous D.J. (1985) Exchangeability and related topics, *Ecole d'été de Probabilités de Saint-Flour XIII, Lecture Notes in Math.* 1117, Springer, Berlin.
2. Ambrosio L., Gigli N., Savaré G. (2008) *Gradient flows*, 2nd ed., Birkhäuser, Basel.
3. Antoniak C.E. (1974) Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems, *Ann. Statist.*, 2, 1152–1174.
4. Barber R.F., Candes E.J. (2015) Controlling the false discovery rate via knockoffs, *Ann. Statist.*, 43, 2055–2085.
5. Berti P., Pratelli L., Rigo P. (2013) Exchangeable sequences driven by an absolutely continuous random measure, *Ann. Probab.*, 41, 2090–2102.
6. Berti P., Dreassi E., Pratelli L., Rigo P. (2021) A class of models for Bayesian predictive inference, *Bernoulli*, 27, 702–726.
7. Berti P., Dreassi E., Leisen F., Pratelli L., Rigo P. (2023) Kernel based Dirichlet sequences, *Bernoulli*, 29, 1321–1342.
8. Berti P., Dreassi E., Leisen F., Pratelli L., Rigo P. (2025) A probabilistic view on predictive constructions for Bayesian learning, *Statist. Science*, 40, 25–39.
9. Blackwell D., MacQueen J.B. (1973) Ferguson distributions via Pólya urn schemes, *Ann. Statist.*, 1, 353–355.
10. Candes E.J., Fan Y., Janson L., Lv J. (2018) Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection, *J. R. Statist. Soc. B*, 80, 551–577.
11. Dalal S.R. (1979) Dirichlet invariant processes and applications to nonparametric estimation of symmetric distribution functions, *Stoch. Process. Appl.*, 9, 99–107.
12. Ferguson T.S. (1973) A Bayesian analysis of some nonparametric problems, *Ann. Statist.*, 1, 209–230.
13. Fortini S., Petrone S. (2025) Exchangeability, prediction and predictive modeling in Bayesian statistics, *Statist. Science*, 40, 40–67.
14. Ghosal S., van der Vaart A.W. (2017) *Fundamentals of nonparametric Bayesian inference*, Cambridge Univ. Press.
15. Hosseini R., Zarepour M. (2021) Bayesian bootstrapping for symmetric distributions, *Statistics*, 55, 711–732.
16. Karatzas I., Shreve S.E. (1991) *Brownian motion and stochastic calculus*, Second edition, Springer, New York.
17. Koehl P., Delarue M., Orland H. (2019) Optimal transport at finite temperature, *Physical Review*, 100, 013310.
18. Sethuraman J. (1994) A constructive definition of Dirichlet priors, *Statist. Sinica*, 4, 639–650.

EMANUELA DREASSI, DIPARTIMENTO DI STATISTICA, INFORMATICA, APPLICAZIONI “G. PARENTI”, UNIVERSITÀ DI FIRENZE, VIALE MORGAGNI 59, 50134 FIRENZE, ITALY

Email address: `emanuela.dreassi@unifi.it`

LUCA PRATELLI, ACCADEMIA NAVALE, VIALE ITALIA 72, 57100 LIVORNO, ITALY

Email address: `luca_pratelli@marina.difesa.it`

PIETRO RIGO (CORRESPONDING AUTHOR), DIPARTIMENTO DI SCIENZE STATISTICHE “P. FORTUNATI”, UNIVERSITÀ DI BOLOGNA, VIA DELLE BELLE ARTI 41, 40126 BOLOGNA, ITALY

Email address: `pietro.rigo@unibo.it`

Received 14/FEB/2025