Some Useful Properties of the Dirichlet Process

Jean-Marie Rolin*

January 1992

Abstract

The aim of this paper is firstly to recall the close connection between the Dirichlet process and the Gamma process in very general spaces and to show how it can be used to derive interesting properties of the Dirichlet process. Secondly, it is shown that the characterization of the Dirichlet process through independence relations between associated σ-fields entails a nice description of the posterior distribution of the Dirichlet process at points where there is either no or at least one observation.

Keywords: Dirichlet process, Gamma measure, Nonparametric Bayesan Statistic.

AMS-MOS Primary: 62 F15
Secondary: 62G99 60G57

This work was financed, in part, by contrat No 26 of the programme "Pôle d'attraction interuniversitaire" of the Belgian government.

* PROB and CORE, Université Catholique de Louvain, Belgium.

ISSN-0771 3894

1. Introduction

This paper has two main objectives. The first one is to show how the close connection between the Dirichlet process and the Gamma process may be used to simplify some proofs of known properties of the Dirichlet process and to provide new such properties. The second one is to show that the Dirichlet process, in the nondegenerate case, is entirely characterized by independence relations between σ-fields associated to the process. Corollaries of these independence relations show how exactly the Dirichlet process is revised by the observations providing a new insight to the transformation prior to posterior distribution.

2. The Dirichlet Process and the Gamma Process

Let us consider an abstract probability space $(\Omega, \mathcal{M}, P)$ and a standard Borel Space $(S, \mathcal{S})$. Recall that a measurable space is standard Borel if there exists a bijective bimeasurable function on $(S, \mathcal{S})$ into a Borel subset of the real line and that a Polish space is standard Borel (see, e.g., Breiman (1968) or Dellacherie–Meyer (1975)). A random σ-finite (resp. finite, resp. probability) measure on $(S, \mathcal{S})$ is an application $A : \Omega \times \mathcal{S} \to (\mathbb{R}^+, \mathcal{B}^+)$, i.e., into the nonnegative real line, denoted $A(B)(\omega)$, such that for any $B \in \mathcal{S}$, $A(B)$ is a positive random variable and for any $\omega \in \Omega$, $A(\cdot)(\omega)$ is a σ-finite (resp. finite, resp. probability) measure on $(S, \mathcal{S})$. For $x \in S$, we use the simplified notation $A(x)$ instead of $A(\{x\})$. It is known that the existence of such a random measure is guaranteed as soon as, for any measurable nontrivial partition of S, $\mathcal{P} = \{B_\ell : 1 \leq \ell \leq k\}$, the distribution of the vector $\{A(B_\ell) : 1 \leq \ell \leq k\}$ is consistently specified. This means that $\lambda(B) = E[A(B)]$ must be a σ-finite (resp. finite, resp. probability along with $A(B) \in [0, 1]$ $\forall$ $B \in \mathcal{S}$ and $A(S) = 1$ a.s.) measure on $(S, \mathcal{S})$ and that the distribution of $\{A(B_1) + A(B_2),\ A(B_\ell) : 3 \leq \ell \leq k\}$ computed from the distribution of $\{A(B_\ell) : 1 \leq \ell \leq k\}$ must be the same as the provided distribution of $\{A(B_1 \cup B_2),\ A(B_\ell) : 3 \leq \ell \leq k\}$ (see Ferguson (1973) or Daley–Vere Jones (1988)). Along these lines, Ferguson (1973) defined the Dirichlet process as follows.

Definition 2.1. *A random probability measure F is a Dirichlet process of parameter $\alpha\mu$ where $\alpha \in \mathbb{R}_0^+$ and μ is a probability measure on $(S, \mathcal{S})$, to be written as $F \sim \mathcal{D}i(\alpha\mu)$, if for any measurable nontrivial partition $\mathcal{P} = \{B_\ell : 1 \leq \ell \leq k\}$ of S, $\{F(B_\ell : 1 \leq \ell \leq k\}$ has the Dirichlet distribution* $\mathrm{Di}[\alpha\mu(B_\ell) : 1 \leq \ell < k]$ *(see*

Wilks (1962)) which may be characterized as follows,

$$\begin{aligned} &\forall\ a_\ell > -\alpha\mu(B_\ell), \quad 1 \le \ell \le k \\ &E\left[\prod_{1\le\ell\le k} F(B_\ell)^{a_\ell}\right] = \frac{\Gamma(\alpha)}{\Gamma(\alpha+a)} \prod_{1\le\ell\le k} \frac{\Gamma[\alpha\mu(B_\ell)+a_\ell]}{\Gamma[\alpha\mu(B_\ell)]} \mathbf{1}_{\{\mu(B_\ell)>0\}} \end{aligned} \tag{2.1}$$

where $a = \sum_{1\le\ell\le k} a_\ell$. ∎

Clearly $F(S) = 1$ a.s. since $E[F(S)^a] = 1\ \forall\ a > -\alpha$ and $\forall\ B \in \mathcal{S}$, $F(B) \sim \text{Be}[\alpha\mu(B), \alpha\mu(B^c)]$, i.e., has a beta distribution. Therefore, $E[F(B)] = \mu(B)$. For more details and properties, see Ferguson (1973).

Next, we define the Gamma process.

Definition 2.2. *A random σ-finite measure A is a Gamma process of parameters $\beta \in \mathbb{R}_0^+$ and ρ, a σ-finite measure on $(S, \mathcal{S})$, to be written as $A \sim \mathcal{G}a(\beta, \rho)$, if for any measurable nontrivial partition of S, $\mathcal{P} = \{B_\ell : 1 \le \ell \le k\}$, the random variables $\{A(B_\ell),\ 1 \le \ell \le k\}$ are independent, to be written as $\perp\!\!\!\perp_{1\le\ell\le k} A(B_\ell)$, and for any $B \in \mathcal{S}$, $A(B) \sim \Gamma(\beta, \rho(B))$, i.e., has a gamma distribution of shape parameter $\rho(B)$ and scale parameter β which may be characterized as follows:*

$$\begin{aligned} &\forall\ a > -\rho(B) \\ &E[A(B)^a] = \frac{\Gamma[\rho(B)+a]}{\Gamma[\rho(B)]} \quad \frac{1}{\beta^a} \mathbf{1}_{\{\rho(B)>0\}}. \end{aligned} \tag{2.2}$$

∎

The following theorems closely relate these two processes.

Theorem 2.3. *Let $\alpha, \beta \in \mathbb{R}_0^+$ and μ be a probability measure on $(S, \mathcal{S})$. If $A \sim \mathcal{G}a(\beta, \alpha\mu)$ and $F = \frac{1}{A(S)} A$, then $F \perp\!\!\!\perp A(S)$, $A(S) \sim \Gamma(\beta, \alpha)$ and $F \sim \mathcal{D}i(\alpha\mu)$.* ∎

Proof: Clearly F is a random probability measure and looking at a measurable nontrivial partition of S, $\mathcal{P} = \{B_\ell : 1 \le \ell \le k\}$, it is a standard result that $\left\{\frac{A(B_\ell)}{A(S)} : 1 \le \ell \le k\right\} \perp\!\!\!\perp A(S)$ and $F(B_\ell) = \frac{A(B_\ell)}{A(S)}$ shows that $\{F(B_\ell) : 1 \le \ell \le k\} \sim \text{Di}[\{\alpha\mu(B_\ell) : 1 \le \ell \le k\}]$. This is an alternative definition of the Dirichlet distribution (see Ferguson (1973) p. 211).

The next theorem is the dual of Theorem 2.3.

Theorem 2.4. *Let $\alpha, \beta \in \mathbb{R}_0^+$ and μ be a probability measure on $(S, \mathcal{S})$. If $F \sim \mathcal{D}i(\alpha\mu)$ and $U \sim \Gamma(\beta, \alpha)$ with $U \perp\!\!\!\perp F$, then if $A = UF$, $A \sim \mathcal{G}a(\beta, \alpha\mu)$.*

Proof: Clearly A is a random finite measure and for any measurable nontrivial partition of S, $\mathcal{P} = \{B_\ell : 1 \leq \ell \leq k\}$, if $a_\ell > -\alpha\mu(B_\ell)$ and $a = \sum_{1\leq\ell\leq k} a_\ell$, by (2.1) and (2.2):

$$\begin{aligned} E\left[\prod_{1\leq\ell\leq k} A(B_\ell)^{a_\ell}\right] &= E\left[U^a \prod_{1\leq\ell\leq k} F(B_\ell)^{a_\ell}\right] \\ &= E[U^a] E\left[\prod_{1\leq\ell\leq k} F(B_\ell)^{a_\ell}\right] \\ &= \frac{\Gamma(\alpha+a)}{\Gamma(\alpha)\beta^a} \frac{\Gamma(\alpha)}{\Gamma(\alpha+a)} \prod_{1\leq\ell\leq k} \frac{\Gamma[\alpha\mu(B_\ell)+a_\ell]}{\Gamma[\alpha\mu(B_\ell)]} \mathbb{1}_{\{\mu(B_\ell)>0\}} \\ &= \prod_{1\leq\ell\leq k} \frac{\Gamma[\alpha\mu(B_\ell)+a_\ell]}{\Gamma[\alpha\mu(B_\ell)]\beta^{a_\ell}} \mathbb{1}_{\{\alpha\mu(B_\ell)>0\}} \end{aligned}$$

.

This shows that $\perp\!\!\!\perp_{1\leq\ell\leq k} A(B_\ell)$ and $A(B_\ell) \sim \Gamma[\beta, \alpha\mu(B_\ell)]$. ∎

Theorem 2.3 appears in Ferguson (1973), Section 4, as an alternative definition of the Dirichlet process. However, instead of using the general definition of a Gamma process (Definition 2.2), he used a representation of a process with independent and gamma increments on the real line obtained by Ferguson and Klass (1972). But this representation was obtained under the supplementary assumption that this process has no fixed points of discontinuity, i.e., μ diffuse measure on $(S, \mathcal{S})$. This is too restrictive since incidentally the posterior distribution of a Dirichlet process is a Dirichlet process of nondiffuse parameter $(\alpha\mu + \sum_{1\leq i\leq n} \varepsilon_{X_i})$. He also obtained a representation of $F \sim \mathcal{D}i(\alpha\mu)$ as

$$F = \sum_{1\leq k<\infty} P_k \varepsilon_{V_k} \tag{2.3}$$

where $P_1^\infty \perp\!\!\!\perp V_1^\infty$, $\perp\!\!\!\perp_{1\leq k<\infty} V_k$, $V_k \sim \mu$ and $P_k > P_{k+1}\ \forall\ 1 \leq k < \infty$. He also produced the distribution of P_k given $P_1^{k-1} = (P_1, P_2, \ldots, P_{k-1})$ which only depends on α. We will show later on that this representation heavily relies on the fact that μ is diffuse and that the use of this representation may furnish false results when μ is not diffuse.

The usefulness of Theorems 2.3 and 2.4 relies in the fact that simple to prove properties of the Gamma process may be used to obtain not easy to prove properties of the Dirichlet process.

Let us first recall the main properties of a Gamma process as defined in (2.2).

3. Main Properties of the Gamma Process

We restrict our attention to the case where ρ is a finite measure on $(S, \mathcal{S})$. Extensions to the σ-finite case are then easily obtained.

For a nonnegative measurable function f on $(S, \mathcal{S})$, if

$$A(f) = \int_S f(x)A(dx) = \int_S f dA \tag{3.1}$$

the distribution of a random finite measure A is entirely characterized by the following functional

$$\psi_A(f) = -\ln E[e^{-A(f)}]. \tag{3.2}$$

Since

$$\begin{aligned} \psi(p, B) &= -\ln E[e^{-pA(B)}] \\ &= \ln\left(1 + \frac{p}{\beta}\right)\rho(B) \\ &= \chi(p)\lambda(B) \end{aligned} \tag{3.3}$$

where

$$\chi(p) = \beta \ln\left(1 + \frac{p}{\beta}\right) \tag{3.4}$$

$$\lambda(B) = E[A(B)] = \frac{1}{\beta}\rho(B). \tag{3.5}$$

Using simple measurable functions and the independence properties, it is easily seen that

$$\psi_A(f) = \int_S \ln\left(1 + \frac{1}{\beta}f(x)\right)\rho(dx). \tag{3.6}$$

We first decompose A by separating fixed and random atoms. Let

$$I = \{x \in S : \rho(x) > 0\}. \tag{3.7}$$

This is a countable set and we define

$$A_d(B) = A(B \cap I) \tag{3.8}$$
$$A_c(B) = A(B \cap I^c)$$

and

$$\rho_c(B) = \rho(B \cap I^c).$$

Now

$$A = A_c + A_d, \tag{3.9}$$
$$A_c \perp\!\!\!\perp A_d$$

and

$$A_c \sim \mathcal{G}a(\beta, \rho_c).$$

Furthermore,

$$A_d = \sum_{x \in I} A(x)\varepsilon_x, \tag{3.10}$$
$$\perp\!\!\!\perp_{x \in I} A(x) \quad \text{and} \quad A(x) \sim \Gamma[\beta, \rho(x)].$$

Next, we define on $(S \times I\!R_0^+, \mathcal{S} \otimes I\!B_0^+)$, the Levy counting process and the Levy measure as

$$N(D) = |\{x \in S : (x, A_c(x)) \in D\}| \tag{3.11}$$
$$\nu(D) = E[N(D)].$$

Since A_c has no fixed atoms, it follows from the theory of purely random measures (see e.g., Daley–Vere Jones (1988)) that N is a Poisson process of parameter ν, i.e.,

$$E[e^{-pN(D)}] = e^{-(1-e^{-p})\nu(D)}. \tag{3.12}$$

Therefore if f is a nonnegative measurable function on $(S \times I\!R_0^+, \mathcal{S} \otimes I\!B_0^+)$ such that $f(x, 0) = 0 \; \forall \; x \in S$,

$$\sum_{x \in S} f[x, A_c(x)] = \int_{S \times I\!R_0^+} f(x, u) N(dx, du) \tag{3.13}$$

and

$$E\left[e^{-\int_{S\times\mathbb{R}_0^+} f(x,u)N(dx,du)}\right] = e^{-\int_{S\times\mathbb{R}_0^+} 1-e^{-f(x,u)}\nu(dx,du)} \tag{3.14}$$

$$E\left[\sum_{x\in S} f[x, A_c(x)]\right] = \int_{S\times\mathbb{R}_0^+} f(x,u)\nu(dx,du). \tag{3.15}$$

Now since A_c is a discrete measure a.s., $A_c(B) = \sum_{x\in S} \mathbf{1}_B(x)A_c(x)$ and therefore, using (3.3), (3.9), and (3.14), we obtain,

$$\ln\left(1+\frac{p}{\beta}\right)\rho_c(B) = \int_{B\times\mathbb{R}_0^+}(1-e^{-pu})\nu(dx,du).$$

We deduce from this that

$$\int_{B\times\mathbb{R}_0^+} ue^{-pu}\nu(dx,du) = \frac{1}{\beta+p}\rho_c(B),$$

i.e.,

$$\nu(dx,du) = \frac{e^{-\beta u}}{u}du\rho_c(dx). \tag{3.16}$$

4. Some Examples of Easily Obtained Properties

A first example is the characterization of the distribution of a functional of the Dirichlet process. For f, a nonnegative measurable function on $(S,\mathcal{S})$, we define

$$F(f) = \int_S fdF = \int_S f(x)F(dx). \tag{4.1}$$

Proposition 4.1. *If $F\sim\mathcal{D}i(\alpha\mu)$, then $\forall\ p\in\mathbb{R}_0^+$*

$$E\left[(1+pF(f))^{-\alpha}\right] = e^{-\alpha\int_S \ln(1+pf)d\mu}. \tag{4.2}$$

Proof: By (3.6) if $A\sim\mathcal{G}a(1,\alpha\mu)$

$$E\left[e^{-pA(f)}\right] = e^{-\alpha\int_S \ln[1+pf]d\mu}.$$

By Theorem (2.4),

$$\begin{aligned} E\left[e^{-pA(f)}\right] &= E\left[e^{-pUF(f)}\right] \\ &= E\left[E\left[e^{-pF(f)U}\mid F\right]\right] \\ &= E\left[\{1+pF(f)\}^{-\alpha}\right] \end{aligned}$$

since $F \perp\!\!\!\perp U$ and $U \sim \Gamma(1, \alpha)$. ∎

This result appears in Cifarelli and Regazzini (1990) (Lemma 1). It was obtained under several restrictive assumptions by a quite difficult proof. Let us recall that another characterization of the distribution of a functional of the Dirichlet process was obtained in the same spirit as ours by Hannum, Hollander and Langberg (1981). Their result goes like this: since $F(f) = \frac{1}{A(S)}A(f)$,

$$P[F(f) \leq t] = P\left[\int_S [f(x) - t]A(dx) \leq 0\right] \tag{4.3}$$
$$= P[T_t \leq 0],$$

and from (3.6)

$$E[e^{-pT_t}] = e^{-\alpha \int_S \ln[1+p[f(x)-t]]\mu(dx)}. \tag{4.4}$$

This relation was obtained in their Theorem 2.3 by a quite long proof.

A second example is concerned with the atoms of a Dirichlet process.

Proposition 4.2. *Let $F \sim \mathcal{D}i(\alpha\mu)$ where μ is a diffuse probability measure on $(S, \mathcal{S})$, i.e. $\mu(x) = 0 \quad \forall\ x \in S$. Then for any $B \in \mathcal{S}$ and any nonnegative measurable function g on $([0,1], \mathcal{B}\cap[0,1])$ with $g(0) = 0$,*

$$E\left[\sum_{x\in B} g[F(x)]\right] = \mu(B)\int_0^1 g(x)\frac{\alpha(1-x)^{\alpha-1}}{x}dx. \tag{4.5}$$

In particular, $\forall\ B \in \mathcal{S}$ and $a \in \mathbb{R}_0^+$,

$$E\left[\sum_{x\in B} F(x)^a\right] = \mu(B)\frac{\Gamma(a)\Gamma(\alpha+1)}{\Gamma(\alpha+a)}. \tag{4.6}$$

Proof: From (3.15), $\forall\ B \in \mathcal{S}$ and $a \in \mathbb{R}_0^+$

$$E\left[\sum_{x\in B} A(x)^a\right] = \int_{B\times\mathbb{R}_0^+} u^a \frac{e^{-u}}{u}du\ \alpha\mu(dx)$$
$$= \alpha\mu(B)\Gamma(a).$$

But by Theorem (2.4),

$$E\left[\sum_{x\in B} A(x)^a\right] = E\left[U^a \sum_{x\in B} F(x)^a\right]$$
$$= E[U^a]E\left[\sum_{x\in B} F(x)^a\right]$$
$$= \frac{\Gamma(\alpha+a)}{\Gamma(\alpha)} E\left[\sum_{x\in B} F(x)^a\right]$$

Therefore,

$$E\left[\sum_{x\in B} F(x)^a\right] = \alpha\mu(B)\frac{\Gamma(a)\Gamma(\alpha)}{\Gamma(\alpha+a)}$$
$$= \mu(B)\int_0^1 x^{a-1}\alpha(1-x)^{\alpha-1}dx.$$

Relation (4.5) being verified for all polynomials without constant term is therefore verified for any nonnegative measurable function vanishing at 0 on $([0,1], I\!B\cap[0,1])$ by the Stone–Weierstrass and monotone class theorems. ∎

Since μ is a diffuse probability measure, the representation of Ferguson (Formula (2.3)) i.e.,

$$F = \sum_{1\le k<\infty} P_k \varepsilon_{V_k} \tag{4.7}$$

holds and as a corollary of Proposition 4.2, we have the following proposition.

Proposition 4.3. *Let $F \sim \mathcal{D}i(\alpha\mu)$ where μ is diffuse and $F = \sum_{1\le k<\infty} P_k\varepsilon_{V_k}$. Then for any nonnegative measurable function f on $(S\times[0,1], \mathcal{S}\otimes(I\!B\cap[0,1]))$ with $f(x,0) = 0 \ \ \forall\ x \in S$*

$$E\left[\sum_{1\le k<\infty} f(V_k, P_k)\right] = \int_{S\times[0,1]} f(x,u)\frac{\alpha(1-u)^{\alpha-1}}{u}\mu(dx)\ du. \tag{4.8}$$

Proof: Indeed

$$\sum_{1\le k<\infty} f(V_k, P_k) = \sum_{x\in S} f[x, F(x)].$$

By Proposition 4.2, (4.8) is true for all f of the form $f(x,u) = \mathbf{1}_B(x)g(u)$. It is then true for all f such that $f(x,0) = 0\ \forall\ x \in S$. ∎

To our knowledge, Propositions 4.2 and 4.3 are new results. A last example is the result obtained by Ferguson (1973) (Theorem 4).

Proposition 4.4. *Let $F \sim \mathcal{D}i(\alpha\mu)$ and let f_1 and f_2 be nonnegative measurable functions on $(S, \mathcal{S})$. Then*

$$E[F(f_1)F(f_2)] = \mu(f_1)\mu(f_2) + \frac{1}{\alpha+1}[\mu(f_1 f_2) - \mu(f_1)\mu(f_2)]. \tag{4.9}$$

Proof: Let B and $C \in \mathcal{S}$ and $A \sim \mathcal{G}a(1, \rho)$. Then, by independence,

$$\begin{aligned} E[A(B)A(C)] &= E[A(B\cap C)^2] + \rho(B\cap C)\rho(B^c\cap C) \\ &\quad + \rho(B\cap C^c)\rho(B\cap C) + \rho(B\cap C^c)\rho(B^c\cap C) \\ &= \{\rho(B\cap C)+1\}\rho(B\cap C) \\ &\quad + \rho(B\cap C^c)\rho(B\cap C) + \rho(B)\rho(B^c\cap C) \\ &= \rho(B\cap C) + \rho(B)\rho(C). \end{aligned}$$

But

$$\begin{aligned} E[A(B)A(C)] &= E[U^2]E[F(B)F(C)] \\ &= \alpha(\alpha+1)E[F(B)F(C)] \end{aligned}$$

Therefore,

$$E[F(B)F(C)] = \frac{1}{\alpha+1}\mu(B\cap C) + \frac{\alpha}{\alpha+1}\mu(B)\mu(C)$$

and by a monotone class argument

$$E[F(f_1)F(f_2)] = \frac{1}{\alpha+1}\mu(f_1 f_2) + \frac{\alpha}{\alpha+1}\mu(f_1)\mu(f_2).$$

■

In his proof, Ferguson uses the representation (4.7) (only valid whan μ is diffuse) and shows in a clever tricky way that

$$E\left[\sum_{1\le k<\infty} P_k^2\right] = \frac{1}{\alpha+1}. \tag{4.10}$$

This is easily obtained from (4.6) or (4.8) when μ is diffuse. However (4.10) is false for arbitrary μ. This may be seen as follows: if $X_1 \perp\!\!\!\perp X_2 \mid F$, $(X_1 \mid F) \sim F$ and $(X_2 \mid F) \sim F$,

$$P[X_1 = X_2 \mid F] = \sum_{x\in S} F(x)^2 = \sum_{1\le k<\infty} P_k^2.$$

But, since $(F \mid X_1) \sim \mathcal{D}i(\alpha\mu + \varepsilon_{X_1})$,

$$(X_2 \mid X_1) \sim \frac{\alpha}{\alpha+1}\mu + \frac{1}{\alpha+1}\varepsilon_{X_1}.$$

Therefore,

$$\begin{aligned} E\left[\sum_{x\in S} F(x)^2\right] &= P[X_1 = X_2] \\ &= E\left[P[X_2 = X_1 \mid X_1]\right] \\ &= \frac{1}{\alpha+1} + \frac{\alpha}{\alpha+1}E[\mu(X_1)] \\ &= \frac{1}{\alpha+1} + \frac{\alpha}{\alpha+1}\sum_{x\in S}\mu(x)^2. \end{aligned}$$

Incidentally, Yamato (1984) also used the representation (4.7) without assuming μ diffuse in computing characteristic functions of functionals of a Dirichlet process.

5. Characterization of the Dirichlet Process by Independence Relations

The following lemma is well known (see, e.g. Fabius (1973)), but we will again give a simple proof using the Gamma process.

Lemma 5.1. *If $F \sim \mathcal{D}i(\alpha\mu)$, then for any nontrivial measurable partition of S, $\mathcal{P} = \{B_\ell : 1 \le \ell \le k\}$, if $\mu(B_1 \cup B_2) > 0$, then*

$$(i) \qquad \frac{F(B_1)}{F(B_1 \cup B_2)} \perp\!\!\!\perp \{F(B_\ell) : 3 \le \ell \le k\}$$

$$(ii) \qquad \frac{F(B_1)}{F(B_1 \cup B_2)} \sim \mathrm{Be}[\alpha\mu(B_1), \alpha\mu B_2)].$$

Proof: If $A \sim \mathcal{G}a(1, \alpha\mu)$, then $A(B_1) \perp\!\!\!\perp A(B_2) \perp\!\!\!\perp \{A(B_\ell) : 3 \le \ell \le k\}$. But, as in the proof of Theorem 2.3,

$$\frac{A(B_1)}{A(B_1) + A(B_2)} \perp\!\!\!\perp A(B_1) + A(B_2)$$

and

$$\frac{A(B_1)}{A(B_1) + A(B_2)} = \frac{F(B_1)}{F(B_1 \cup B_2)} \sim \mathrm{Be}[\alpha\mu(B_1), \alpha\mu(B_2)].$$

This shows (ii) and implies that

$$\frac{F(B_1)}{F(B_1 \cup B_2)} \perp\!\!\!\perp \{A(B_1) + A(B_2), A(B_\ell) : 3 \le \ell \le k\}.$$

But $A(S) = A(B_1) + A(B_2) + \sum_{3\le\ell\le k} A(B_\ell)$. Therefore

$$\frac{F(B_1)}{F(B_1 \cup B_2)} \perp\!\!\!\perp \{A(S), F(B_\ell) : 3 \le \ell \le k\}.$$

■

We now define random conditional probability measures and σ-fields associated to it. For elementary properties of (conditional) independence between σ-fields and related notation, the reader may refer to Florens, Mouchart and Rolin (1990). If F is a random probability measure and $A \in \mathcal{S}$, we define the random probability conditional on A, F^A as

$$F^A(B) = F(B \mid A) = \frac{F(B \cap A)}{F(A)} \mathbf{1}_{\{F(A)>0\}} + \mathbf{1}_{\{F(A)=0\}}. \tag{5.1}$$

For $A, B \in \mathcal{S}$, we define

$$\mathcal{F}_B^A = \sigma\{F(B \cap C \mid A) : C \in \mathcal{S}\}, \tag{5.2}$$

i.e., the knowledge of $\mathcal{F}_B^A$ corresponds to the knowledge of all probabilities of subsets of B conditional on A. In particular, we write

$$\mathcal{F}_A = \mathcal{F}_A^S \qquad \mathcal{F}^A = \mathcal{F}_S^A. \tag{5.3}$$

The following results are generalizations of results obtained by Fabius (1972), (1973), for the Dirichlet distribution and by Docksum (1973) for the Dirichlet process on the real line when they considered several concepts of neutrality. See also Ferguson (1974), Antoniak (1971), Rolin (1982), James and Mosiman (1980).

Theorem 5.2. *Let F be a random probability measure on $(S, \mathcal{S})$. Then $F \sim \mathcal{D}i(\alpha\mu)$ if and only if $\forall\ A, B \in \mathcal{S}$ with $0 < \mu(A) < 1$,*

$$(i) \qquad F(B \mid A) \perp\!\!\!\perp \mathcal{F}_{A^c}$$

$$(ii) \qquad F(B) \sim \mathrm{Be}[\alpha\mu(B), \alpha\mu(B^c)].$$

Proof: To prove the "only if" part note that (ii) is clear and that for any nontrivial measurable partition of A^c, $\mathcal{P} = \{B_\ell : 1 \le \ell \le k\}$ we obtain in completing this partition with $A \cap B$ and $A \cap B^c$ by Lemma 5.1, that $F(B \mid A) = \frac{F(B \cap A)}{F(B \cap A) + F(B^c \cap A)}$ a.s. is independent of $\{F(B_\ell) : 1 \le \ell \le k\}$ and this proves (i).

To prove the "if" part, we first note that (i) and (ii) imply that

$$\begin{aligned} &\forall\ A, B \in \mathcal{S} \quad \text{with} \quad \mu(A) > 0 \\ &F(B \mid A) \sim \text{Be}[\alpha\mu(B \cap A), \alpha\mu(B^c \cap A)]. \end{aligned} \tag{5.4}$$

Indeed since $F(B \cap A) = F(A)F(B \mid A)$ a.s. and by (i), $F(A) \perp\!\!\!\perp F(B \mid A)$, $\forall\ a > 0$,

$$E[F(B \cap A)^a] = E[F(A)^a]E[F(B \mid A)^a].$$

Therefore,

$$E[F(B \mid A)^a] = \frac{\Gamma[\alpha\mu(B \cap A) + a]\Gamma(\alpha)}{\Gamma(\alpha + a)\Gamma[\alpha\mu(B \cap A)]} \Big/ \frac{\Gamma[\alpha\mu(A) + a]\Gamma(\alpha)}{\Gamma[\alpha + a)\Gamma[\alpha\mu(A)]}$$

$$= \frac{\Gamma[\alpha\mu(B \cap A) + a]\Gamma[\alpha\mu(A)]}{\Gamma[\alpha\mu(A) + a]\Gamma[\alpha\mu(B \cap A)]}.$$

Hence the desired result since $\mu(A) = \mu(B \cap A) + \mu(B^c \cap A)$.

Now if we consider a nontrivial measurable partition of S, $\mathcal{P} = \{B_\ell : 1 \le \ell \le k\}$ with $\mu(B_\ell) > 0\ \forall\ 1 \le \ell \le k$, we define $A_\ell = \underset{1 \le r \le \ell}{\cup} B_r$ and $Z_\ell = F(A_\ell \mid A_{\ell+1})$ $1 \le \ell \le k-1$. Now by (i) $Z_\ell \perp\!\!\!\perp \mathcal{F}_{A^c_{\ell+1}}$ and by (5.4) $Z_\ell \sim \text{Be}[\alpha\mu(A_\ell), \alpha\mu(A_{\ell+1} - A_\ell)]$. Note also that Z_ℓ is $\mathcal{F}_{A^c_\ell}$-measurable since $Z_\ell = \frac{F(A_\ell)}{F(A_\ell)+F(A_{\ell+1} \cap A^c_\ell)}$. Therefore $\underset{1 \le \ell \le k-1}{\perp\!\!\!\perp} Z_\ell$. Now

$$F(B_\ell) = F(A_\ell) - F(A_{\ell-1}) = (1 - Z_{\ell-1}) \prod_{\ell \le r \le k-1} Z_r.$$

Hence $\forall\ a_\ell > -\alpha\mu(B_\ell)\ \ 1 \le \ell \le k$

$$\prod_{1 \le \ell \le k} F(B_\ell)^{a_\ell} = \prod_{1 \le \ell \le k-1} (1 - Z_\ell)^{a_{\ell+1}} Z_\ell^{b_\ell}$$

where $b_\ell = \sum_{1 \le r \le \ell} a_r$. Therefore

$$\begin{aligned} E\left[\prod_{1 \le \ell \le k} F(B_\ell)^{a_\ell}\right] &= \prod_{1 \le \ell \le k-1} E\left[(1 - Z_\ell)^{a_{\ell+1}} Z_\ell^{b_\ell}\right] \\ &= \prod_{1 \le \ell \le k-1} \frac{\Gamma[\alpha\mu(A_\ell) + b_\ell]\Gamma[\alpha\mu(A_{\ell+1} - A_\ell) + a_{\ell+1}]\Gamma[\alpha\mu(A_{\ell+1})]}{\Gamma[\alpha\mu(A_\ell)]\Gamma[\alpha\mu(A_{\ell+1} - A_\ell)]\Gamma[\alpha\mu(A_{\ell+1}) + b_{\ell+1}]} \end{aligned}$$

$$= \frac{\Gamma(\alpha)}{\Gamma(\alpha + b_k)} \prod_{1 \le \ell \le k} \frac{\Gamma[\alpha\mu(B_\ell) + a_\ell]}{\Gamma[\alpha\mu(B_\ell)]}.$$

Hence, by (2.1), this shows that $F \sim \mathcal{D}i(\alpha\mu)$. ∎

The next theorem shows that in fact under some mild nontriviality conditions (i) implies (ii) and consequently the Dirichlet process is almost entirely characterized by (i) a condition of independence stating that any random probability conditional on A is independent of all random probabilities of subsets of A^c for any $A \in \mathcal{S}$.

Theorem 5.3. *Let F be a random probability measure on $(S, \mathcal{S})$ and let $\mu(B) = E[F(B)]$. If*

(i) $\quad \forall\ A, B \in \mathcal{S}, \quad 0 < \mu(A) < 1$
$\quad F(B \mid A) \perp\!\!\!\perp \mathcal{F}_{A^c}$

(ii) $\quad \exists\ B_0 \in \mathcal{S}, \quad 0 < \mu(B_0) < 1 \quad$ *such that*
$\quad P[F(B_0) \neq \mu(B_0)] > 0$

(iii) $\quad \exists \{B_i : 1 \leq i \leq 3\} \subset \mathcal{S} \quad$ *partition of* S *such that*
$\quad \mu(B_i) > 0 \quad \forall\ 1 \leq i \leq 3,$

then $\exists\ \alpha \in]0, \infty[$ such that $F \sim \mathcal{D}i(\alpha\mu)$.

Proof: We will only sketch the proof. It suffices to show that $\exists\ \alpha \in]0, \infty[$ such $\forall\ B \in \mathcal{S}$, $F(B) \sim \text{Be}[\alpha\mu(B), \alpha\mu(B^c)]$. This is clearly true if $\mu(B) = 0$ or $\mu(B) = 1$. Next from (i) and (ii), it may be shown that $\forall\ B \in \mathcal{S}$ such that $0 < \mu(B) < 1$, $P[F(B) \neq \mu(B)] > 0$. Afterwards, $\forall\ B \in \mathcal{S}$ such that $0 < \mu(B) < 1$, under (iii), it is possible to find a measurable partition $\{C_1, C_2, C_3\}$ with $\mu(C_i) > 0\ \forall\ 1 \leq i \leq 3$, $C_1 = B$ or $C_1 = B^c$ and $C_2, C_3 \in \{B \cap B_i, B \cap B_i^c,\ B^c \cap B_i,\ B^c \cap B_i^c,\ 1 \leq i \leq 3\}$. Finally, if $X = F(C_1^c)$ and $Y = F(C_2 \mid C_1^c)$, X and Y are nondegenerate random variables which by (i) are independent. But $Z = 1 - XY = F(C_2^c)$ and $U = \frac{1-X}{1-XY} = F(C_1 \mid C_2^c)$ are also nondegenerate independent random variables. In this situation, it has been shown by James and Mosiman (1980) and Fabius (1973) (see also Rolin (1982)), that $F(C_1^c) \sim B[\alpha\mu(C_1^c), \alpha\mu(C_1)]$ where $\alpha = \frac{E[F(C_1)F(C_1^c)]}{V[F(C_1)]}$. ∎

This theorems shows the natural character of the Dirichlet process since besides degenerate random probability measures, the Dirichlet process is the only random probability measure satisfying the independence relation (i).

Theorem 5.2 entails some interesting corollaries. If F is a random probability measure on $(S, \mathcal{S})$ and if $\mathcal{P} = \{B_\ell : 1 \le \ell \le k\}$ is a nontrivial measurable partition of S with $\mu(B_\ell) > 0 \ \ \forall\ 1 \le \ell \le k$, then F may be decomposed as follows:

$$F = \sum_{1 \le \ell \le k} F(B_\ell) F^{B_\ell} \tag{5.5}$$

and Theorem 5.2 permits to completely describe the distributions of the terms involved in the decomposition.

Corollary 5.4. *Let $F \sim \mathcal{D}i(\alpha\mu)$. Then $\forall\ A \in \mathcal{S}$ such that $0 < \mu(A) < 1$,*

(i) $\qquad F(A) \perp\!\!\!\perp F^A \perp\!\!\!\perp F^{A^c}$

(ii) $\qquad F^A \sim \mathcal{D}i[\alpha\mu(A)\mu^A]$

(iii) $\qquad F^{A^c} \sim \mathcal{D}i[\alpha\mu(A^c)\mu^{A^c}]$

(iv) $\qquad F(A) \sim \mathrm{Be}[\alpha\mu(A), \alpha\mu(A^c)]$

Proof: Let $\mathcal{P} = \{B_\ell : 1 \le \ell \le k\}$ be any nontrivial measurable partition of S and let $A_\ell = \bigcup_{1 \le r \le \ell} B_r$. By repeated uses of (i), we have, as in the proof of Theorem 5.2, $\perp\!\!\!\perp_{1 \le \ell \le k-1} F(A_\ell \mid A_{\ell+1} \cap A) \perp\!\!\!\perp \mathcal{F}_{A^c}$ and therefore $\{F(B_\ell \mid A) : 1 \le \ell \le k\} \perp\!\!\!\perp \mathcal{F}_{A^c}$. This shows that $F^A \perp\!\!\!\perp \mathcal{F}_{A^c}$. This implies that $F^A \perp\!\!\!\perp F(A)$ and since $\mathcal{F}_{A^c} = \mathcal{F}^{A^c} \vee \sigma\{F(A)\}$ this shows that $F^A \perp\!\!\!\perp F^{A^c} \perp\!\!\!\perp F(A)$. To show that F^A is a Dirichlet process it suffices to prove (i) and (ii) of Theorem 5.2 for F^A. But (ii) for F^A is equivalent to (5.4). Now for $B, C \in \mathcal{S}$ such that $0 < \mu^A(C) < 1$, i.e., $0 < \mu(A \cap C) < 1$, $F^A(B \mid C) = F(B \mid A \cap C)$. Therefore $F^A(B \mid C) \perp\!\!\!\perp \mathcal{F}_{A^c \cup C^c}$. But it is easily seen that $\mathcal{F}_{A^c \cup C^c} = \mathcal{F}_{A^c} \vee \mathcal{F}^A_{C^c}$. Indeed, $\forall\ B \in \mathcal{S}$

$$F[B \cap (A^c \cup C^c)] = F(B \cap A^c) + F(A) F(B \cap C^c \mid A).$$

Therefore $F^A(B \mid C) \perp\!\!\!\perp \mathcal{F}^A_{C^c} \perp\!\!\!\perp \mathcal{F}_{A^c}$, i.e., (i) is true for F^A. ∎

This corollary may be generalized as follows.

Corollary 5.5. *Let $F \sim \mathcal{D}i(\alpha\mu)$. Then for any nontrivial measurable partition of S, $\mathcal{P} = \{B_\ell : 1 \le \ell \le k\}$ with $\mu(B_\ell) > 0 \ \ \forall\ 0 \le \ell \le k$,*

(i) $\qquad \perp\!\!\!\perp_{1 \le \ell \le k} F^{B_\ell} \perp\!\!\!\perp \{F(B_\ell) : 1 \le \ell \le k\}$

(ii) $\qquad \{F(B_\ell) : 1 \le \ell \le k\} \sim \mathrm{Di}[\alpha\mu(B_\ell) : 1 \le \ell \le k]$

(iii) $\qquad F^{B_\ell} \sim \mathcal{D}i[\alpha\mu(B_\ell)\mu^{B_\ell}] \qquad \forall\ 1 \le \ell \le k.$

Proof: Let $\mathcal{G} = \sigma\{F(B_\ell) : 1 \leq \ell \leq k\}$. By Corollary 5.4, $\mathcal{F}^{B_1} \perp\!\!\!\perp \mathcal{F}_{B_1^c}$. But $\mathcal{F}_{B_1^c} = \mathcal{F}_{\bigcup_{2\leq\ell\leq k} B_\ell} = \mathcal{G} \vee \bigvee_{2\leq\ell\leq k} \mathcal{F}^{B_\ell}$. Once again by Corollary 5.4,

$$\mathcal{F}^{B_2} \perp\!\!\!\perp (\mathcal{G} \vee \bigvee_{3\leq\ell\leq k} \mathcal{F}^{B_\ell}).$$

Therefore

$$\mathcal{F}^{B_1} \perp\!\!\!\perp \mathcal{F}^{B_2} \perp\!\!\!\perp \left(\bigvee_{3\leq\ell\leq k} \mathcal{F}^{B_\ell} \vee \mathcal{G} \right)$$

and $\perp\!\!\!\perp_{1\leq\ell\leq k} \mathcal{F}^{B_\ell} \perp\!\!\!\perp \mathcal{G}$. This finishes the proof. ∎

Particular interesting applications of Corollary 5.4 is to look at the decomposition of μ into its diffuse and purely atomic parts and to characterize the posterior distribution of the Dirichlet process with diffuse parameter.

For the first one, we define for $F \sim \mathcal{D}i(\alpha\mu)$

$$\begin{aligned} &I = \{x \in S : \mu(x) > 0\} \\ &\alpha_d = \alpha\mu(I) \qquad \alpha_c = \alpha\mu(I^c) \\ &\mu_d = \mu^I \qquad\quad \mu_c = \mu^{I^c} \\ &F_d = F^I \qquad\quad F_c = F^{I^c} \\ &P = F(I). \end{aligned} \tag{5.6}$$

Note that

$$\begin{aligned} \alpha\mu &= \alpha_c\mu_c + \alpha_d\mu_d \\ F &= PF_d + (1-P)F_c \end{aligned} \tag{5.7}$$

and that

$$F_d = \sum_{x\in I} F_d(x)\varepsilon_x \tag{5.8}$$

where

$$F_d(x) = \frac{F(x)}{F(I)}. \tag{5.9}$$

By Corollary 5.4, we obtain the following results:

$$P \perp\!\!\!\perp F_c \perp\!\!\!\perp F_d \tag{5.10}$$

$$P \sim \mathrm{Be}(\alpha_d, \alpha_c) \tag{5.11}$$

$$F_c \sim \mathcal{D}i(\alpha_c \mu_c) \tag{5.12}$$

$$F_d \sim \mathcal{D}i(\alpha_d \mu_d) \tag{5.13}$$

and in particular

$$F_d(x) \sim \mathrm{Be}[\alpha\mu(x), \alpha - \alpha\mu(x)]. \tag{5.14}$$

Let us now consider random variables

$$X_i : (\Omega, \mathcal{M}) \to (S, \mathcal{S}), \quad 1 \le i \le n, \tag{5.15}$$

which form an i.i.d. sample distributed according to F where $F \sim \mathcal{D}i(\alpha\mu)$ with μ a diffuse probability measure on $(S, \mathcal{S})$, i.e.,

$$\perp\!\!\!\perp_{1 \le i \le n} X_i \mid F, \tag{5.16}$$

$$(X_i \mid F) \sim F \qquad \forall\ 1 \le i \le n \tag{5.17}$$

i.e.,

$$P[X_i \in B \mid F] = F(B) \qquad \forall\ B \in \mathcal{S}, \quad \forall\ 1 \le i \le n.$$

Ferguson (1973) showed that the posterior distribution of F is given by

$$(F \mid X_1^n) \sim \mathcal{D}i(\alpha\mu + \sum_{1 \le i \le n} \varepsilon_{X_i}). \tag{5.18}$$

Since F is discrete, equalities are likely to happen between the X_i's. So let M_n be the number of distinct values, i.e.,

$$M_n = \sum_{1 \le i \le n} \mathbf{1}_{\bigcap_{1 \le j \le i-1} \{X_i \ne X_j\}}. \tag{5.19}$$

Let Z_k $(1 \le k \le M_n)$ be the k-th distinct value appearing in the sequence $(X_1, X_2, \ldots, X_n)$ and let $D_{n,k}$ be the number of X_i's equal to Z_k, i.e.,

$$D_{n,k} = \sum_{1 \le i \le n} \mathbf{1}_{\{X_i = Z_k\}}. \tag{5.20}$$

With these definitions we have, if $\widehat{F}_n$ is the empirical distribution of the sample,

$$n\widehat{F}_n = \sum_{1 \le i \le n} \varepsilon_{X_i} = \sum_{1 \le k \le M_n} D_{n,k} \varepsilon_{Z_k}. \tag{5.21}$$

Now, if μ is a diffuse probability measure, the atoms of $\alpha\mu + n\widehat{F}_n$ are $\{Z_k : 1 \leq k \leq M_n\}$. Therefore if

$$P_n = \sum_{1 \leq k \leq M_n} F(Z_k) \tag{5.22}$$

$$F_d(Z_k) = P_n^{-1} F(Z_k) \tag{5.23}$$

and

$$F_d = \sum_{1 \leq k \leq M_n} F_d(Z_k) \varepsilon_{Z_k}. \tag{5.24}$$

We see that a posteriori, we can decompose F as follows

$$F = (1 - P_n)F_c + P_n F_d \tag{5.25}$$

with the following properties

$$P_n \perp\!\!\!\perp F_c \perp\!\!\!\perp F_d \mid X_1^n \tag{5.26}$$

$$P_n \mid X_1^n \sim \text{Be}(n, \alpha) \tag{5.27}$$

$$F_c \mid X_1^n \sim \mathcal{D}i(\alpha\mu) \tag{5.28}$$

$$F_d \mid X_1^n \sim \mathcal{D}i(n\widehat{F}_n). \tag{5.29}$$

This shows that P_n and F_c are independent of the sample and that F_c has the same distribution as the prior distribution. At last (5.29) may be rewritten as

$$\begin{aligned} &\{F_d(Z_k) : 1 \leq k \leq M_n\} \mid X_1^n \\ &\sim \text{Di}[\{D_{n,k} : 1 \leq k \leq M_n\}] \end{aligned} \tag{5.30}$$

and in particular,

$$F_d(Z_k) \mid X_1^n \sim \text{Be}[D_{n,k}, n - D_{n,k}]. \tag{5.31}$$

Now from (5.25),

$$F(Z_k) = P_n F_d(Z_k), \tag{5.32}$$

and from (5.18)

$$F(Z_k) \mid X_1^n \sim \text{Be}[D_{n,k}, \alpha + n - D_{n,k}]. \tag{5.33}$$

Recalling (5.27) and (5.26) this sheds a new light on how is revised the Dirichlet process at points where there are some observations.

REFERENCES

Antoniak, C.E. (1974) Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics*, 2, 1152–1174.

Breiman (1968) *Probability*. Reading Menlo Park and London: Addison-Wesley.

Cifarelli, D.M. and E. Reggazzini (1990) Distribution functions of means of a Dirichlet process. *Annals of Statistics*, 18, 429–442.

Daley, D.J. and D. Vere-Jones (1988) *An Introduction to the Theory of Point Processes*. New York: Srpinger-Verlag.

Dellacherie, C. and P.A. Meyer (1975) *Probabilité et Potentiel*. Paris: Hermann.

Doksum, K. (1974) Tailfree and neutral random probabilities and their posterior distributions. *Annals of Probability*, 2, 183–201.

Fabius, J. (1972) *Neutrality and Dirichlet Distributions*. Trans. Sixth Prague Conf. Inform. Theor., Statist. Dec. Functions, Random Processes.

Fabius, J. (1973) Two characterizations of the Dirichlet distribution. *Annals of Statistics*, 1, 583–587.

Ferguson, T.S. (1973) A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1, 209–230.

Ferguson, T.S. (1974) Prior distributions on spaces of probability measures. *Annals of Statistics*, 2, 615–629.

Ferguson, T.S. and M.J. Klass (1972). A representation of independent increment processes without Gaussian components. *Annals of Mathematical Statistics*, 43, 1634–1643.

Florens, J.P., Mouchart, M. and J.M. Rolin (1990). *Elements of Bayesian Statistics*. New York: Marcel Dekker.

Hannum, R.C., Hollander, M. and N.A. Langberg (1981). Distributional results for random functionals of a Dirichlet process. *Annals of Probability*, 9, 665–670.

James, I.R. and J.E. Mosimann (1980). A new characterization of the Dirichlet

distribution through neutrality. *Annals of Statistics*, 8, 183–189.

Rolin, J.M. (1983). Nonparametric Bayesian statistics: A stochastic process approach. *Lecture Notes in Statistics*, 16, 108–133.

Wilks, S.S. (1962). *Mathematical Statistics*. New York: Wiley.

Yamato, H. (1984). Characteristic functions of means of distributions chosen from a Dirichlet process. *Annals of Probability*, 12, 262–267.