Characterisations of Random Dirichlet Means

Nils Lid Hjort and Andrea Ongaro

University of Oslo and Università di Milano-Bicocca

– *October 2003* –

Abstract. Two characterisations of a random mean from a Dirichlet process, as a limit of finite sums of a simple symmetric form and as a solution of certain stochastic equations, are investigated. These are used to reach results on and new insight into such random means. In particular, identities involving functional transforms and recursive moment formulae are established. Furthermore, characterisations for several choices of the Dirichlet process parameter (leading to symmetric, unimodal, stable and finite mixture distributions) are provided. The theory also extends to the case of several random Dirichlet means simultaneously, a situation not covered earlier in the literature.

Key words: *Dirichlet process, Hilbert transform, random means, random probability measure, stochastic equation*

1. Introduction and summary

The Dirichlet process was brought to the attention of the statistical community as the first genuine nonparametric prior for use in Bayesian statistics by Ferguson (1973, 1974). It is still a cornerstone in nonparametric Bayesian methodology, where it is often used separately or as an ingredient in more complicated priors. It is also a favourite special case of larger classes of nonparametric priors, like neutral to the right and tailfree processes, Beta processes and Pólya trees; see Walker, Damien, Laud and Smith (1998) and Hjort (2003) for recent reviews and discussion. This paper is concerned with distributional aspects for important functionals of the Dirichlet process. In Bayesian contexts one is more interested in the posterior distributions of such functionals, i.e. given a set of observations, but since the random distribution underlying the data continues to be a Dirichlet process given the data, only with updated parameters, the distributional results we reach here continue to be relevant.

Let P be such a Dirichlet process on some sample space Ω, with parameter aP_0, in terms of a probability distribution P_0 and a positive strength parameter a. It is characterised by the property that for any measurable partition $(A_1,\ldots,A_k)$, the vector $(P(A_1),\ldots,P(A_k))$ is Dirichlet distributed with parameter vector $(aP_0(A_1),\ldots,aP_0(A_k))$, and we write $P\sim\mathrm{Dir}(aP_0)$. We shall be interested in the random mean $\theta=\theta(P)=\int g\,\mathrm{d}P$. Here g is in principle any measurable function making the integral finite a.s.. Note that the mean can be represented as

$$\theta=\int_\Omega g(x)\,\mathrm{d}P(x)=\int_{-\infty}^{\infty} y\,\mathrm{d}Q(y), \tag{1.1}$$

where $Q = Pg^{-1}$ is the transformed Dirichlet process with parameter $aQ_0 = aP_0g^{-1}$ on the real line.

There is a growing literature on exact formulae for and numerical approximations to distributions of random Dirichlet means. Cifarelli and Regazzini (1990, 1994) complemented earlier results of Hannum, Hollander and Langberg (1981) and Yamato (1984), and gave rather complicated formulae in terms of limits of integrals in regions of the complex plane, following inversion of certain transforms. Diaconis and Kemperman (1996) provided new tools and surprising connections to other areas of probability and mathematics, and also gave some explicit formulae for the special case of $a = 1$. The recent paper of Regazzini, Guglielmo and Di Nunno (2002) sums up earlier work and also discusses successes and difficulties with attempts at the quite non-trivial computer implementation of the mathematical results.

Random Dirichlet means can be used in several statistical contexts in addition to the most natural one, which is to make inference on such mean parameters in a framework of nonparametric Bayesian statistics. Diaconis and Kemperman made various connections, also to topics of current interest in mathematics. Ferguson (1983) and Lo (1984) were early papers working on Bayesian density estimation where the prior involves a kernel smoothed integral of a Dirichlet process; see also Hjort (1996). In a Bayesian regression framework one would like to model noise with mean zero around the main structure, and this could be accomplished with a Dirichlet process having its mean subtracted. This calls for somewhat complicated calculations involving simultaneous aspects of the process and its mean, and where results and moment formulae of this paper are relevant. Finally, in Hjort (2003) a model for random shapes is being discussed, involving a smoothed normalised gamma process to represent a process of random radii. This is the same as a process of integrals with respect to a Dirichlet process.

We shall investigate here two representations of a random Dirichlet mean with the aim of demonstrating their potential both to reach new results and to give independent and simpler proof of known ones. More specifically, the contents of the paper are as follows. In Section 2 and 3 the representations are introduced and applied to prove general results on random Dirichlet means. Specifically, in Section 2 Dirichlet means are constructed as limits of finite sums of simple symmetric random variables. This enables us to obtain necessary and sufficient conditions for finiteness of a Dirichlet mean and a new characterisation in terms of characteristic functions. As a consequence of the latter, a strategy for simulating from the exact distribution of a Dirichlet mean is developed, provided only that the strength parameter a is an integer; furthermore, the asymptotic behaviour of θ for a going to infinity is established.

In Section 3 Dirichlet means are characterised as solutions of certain stochastic equations. For any fixed value of the parameter a, such stochastic equations are shown to establish a one-to-one correspondence between the law of a Dirichlet mean and the distribution $Q_0 = P_0g^{-1}$, leading to an expression of the Hilbert transform of Q_0 in terms of

Hilbert transforms of the corresponding random mean. As a further direct consequence of such equations, simple recursive formulae for direct and centralised moments together with necessary and sufficient conditions for their existence are derived.

The two above representations are then applied in Section 4 to obtain useful information on Dirichlet means associated with distributions Q_0 belonging to some general classes. In particular, we consider symmetric, unimodal, stable and finite mixture distributions. A number of examples is also provided. A number of examples, including applications to the general setting of mixtures of Dirichlet processes, are also provided. Finally, in Section 5 some further topics are briefly discussed, where the theory and methods developed in this article can be applied.

2. Construction as limit of symmetric distributions

One of the main tools used in the following to derive results on the random mean $\theta = \int g \, \mathrm{d}P$ is based on approximations via finite sums of certain symmetric distributions. This stems from the following simple approximation of a Dirichlet process. Let $\xi_1, \cdots, \xi_m$ be independent from P_0 and independent of $(\beta_1, \ldots, \beta_m)$, which we give a symmetric Dirichlet distribution with parameter $(a/m, \ldots, a/m)$. Then the random probability measure

$$P_m = \sum_{j=1}^{m} \beta_j \delta(\xi_j) \tag{2.1}$$

converges in distribution to a Dirichlet process with parameter aP_0 as $m \to \infty$; for a proof, see Hjort and Ongaro (2003). Here $\delta(\xi)$ denotes unit point mass at position ξ. As a direct consequence of this one obtains (see Theorem 4.2 in Kallenberg, 1986) convergence in distribution of

$$\theta_m = \int g \, \mathrm{d}P_m = \sum_{j=1}^{m} g(\xi_j)\beta_j \tag{2.2}$$

to θ when g is a continuous function with compact support. It is the symmetry of the Dirichlet distribution used in (2.1) and (2.2) that for some applications gives a simpler treatment than if working with other characterisations of the Dirichlet process.

The random probability (2.1) has been considered by several authors, independently and in quite unrelated contexts; an extensive list of references is given in Ishwaran and Zarepour (2002, Section 4). In particular, Ishwaran and Zarepour proved convergence of θ_m under the assumption that g is P_0-integrable. In the following theorem we shall prove convergence of θ_m under completely general conditions, i.e. whenever θ exists finite. As a consequence of this approximation, we shall also derive necessary and sufficient conditions for finiteness of θ and an identity in terms of characteristic functions which fully determines its distribution. The latter identity produces, as a special case, the so-called Hilbert transform of order a of θ. Hereafter we shall denote by ξ a random variable with distribution P_0.

THEOREM 1. *Let $P \sim \mathrm{Dir}(aP_0)$ and consider $\theta = \int g\,\mathrm{d}P$ for a measurable g for which*

$$\mathrm{E}\log\{1+|g(\xi)|\} = \int \log(1+|g|)\,\mathrm{d}P_0 = \int_{-\infty}^{\infty} \log(1+|y|)\,\mathrm{d}Q_0(y) \quad \textit{is finite.} \tag{2.3}$$

Let furthermore $G_a \sim \mathrm{Gam}(a,1)$ be independent of θ. Then θ is a.s. finite and its distribution is fully characterised by the characteristic function of $(G_a\theta, G_a)$:

$$\mathrm{E}\exp[i\,(tG_a\theta + sG_a)] = \exp[-a\mathrm{E}\log\{1-i(tg(\xi)+s)\}] \quad \text{for } t,s \in \mathcal{R}. \tag{2.4}$$

As a special case one obtains:

$$\mathrm{E}\,\exp(itG_a) = \mathrm{E}\Big(\frac{1}{1-it\theta}\Big)^a = \exp[-a\,\mathrm{E}\log\{1-itg(\xi)\}] \quad \textit{for } t \in \mathcal{R}. \tag{2.5}$$

Furthermore, θ_m converges in distribution to θ. If on the other hand $\mathrm{E}\log(1+|g(\xi)|)$ is infinite, then $\int |g|\,\mathrm{d}P$ is a.s. infinite, so that θ does not exist finite.

PROOF. Write $\beta_j = G_j/S_m$ in terms of independent and identically distributed $G_j \sim \mathrm{Gam}(a/m,1)$, with sum $S_m = \sum_{j=1}^m G_j$. This leads to $\theta_m = R_m/S_m$, with $R_m = \sum_{j=1}^m G_j g(\xi_j)$ being a random mixture of independent small gammas. We shall first compute the characteristic function of (R_m, S_m) and its limit under condition (2.3).

Use $\mathrm{E}\exp(-itG_j) = (1-it)^{-a/m}$ to find

$$\begin{aligned}\mathrm{E}\big[\exp(-i(tR_m+sS_m))\,|\,\xi_1,\ldots,\xi_m\big] &= \prod_{j=1}^m \{1-i(tg(\xi_j)+s)\}^{-a/m}\\ &= \exp\Big[-a\frac{1}{m}\sum_{j=1}^m \log\{1-i(tg(\xi_j)+s)\}\Big].\end{aligned}$$

Note that if (2.3) holds, then automatically also $\mathrm{E}\log\{1-i(tg(\xi)+s)\}$ is finite for all $t,s\in\mathcal{R}$. Under this assumption,

$$\mathrm{E}\exp(-i(tR_m+sS_m)) \to \exp[-a\mathrm{E}\log\{1-i(tg(\xi)+s)\}] \quad \text{as } m\to\infty \tag{2.6}$$

by the law of large numbers.

Suppose now that g is bounded and write $\theta = \int y\,\mathrm{d}Q(y)$ with $Q \sim \mathrm{Dir}(aP_0g^{-1})$. Convergence of θ_m to θ is then a consequence, by Theorem 4.2 in Kallenberg (1986), of convergence in distribution of P_m in (2.1). This in turn implies $(R_m, S_m) \sim (G_a\theta_m, G_a) \to_d (G_a\theta, G_a)$, where $G_a \sim \mathrm{Gam}(a,1)$ is independent of θ_m and θ. It follows then by (2.6) that relation (2.4) must hold when g is bounded.

Consider now a general measurable g and let us see when θ exists finite, i.e. when $\theta^+ = \int |g|\,\mathrm{d}P < \infty$ a.s.. Let $h_k(y) = yI\{|y|\le k\} + kI\{y>k\} - kI\{y<-k\}$. Then as $|h_k(y)| \uparrow |y|$, by the monotone convergence theorem we have $\theta_k^+ = \int |h_k(g)|\,\mathrm{d}P \uparrow \theta^+$

a.s., where θ^+ is an extended non-negative random variable. This implies $G_a\theta_k^+ \uparrow G_a\theta^+$, which implies convergence of the Laplace transform $\mathrm{E}\exp(-tG_a\theta_k^+)$ to $\mathrm{E}\exp(-tG_a\theta^+)$ for $t > 0$. Note that this convergence takes place even if $G_a\theta^+$ is not finite with probability one. Noticing that $|h_k(g)| \leq k$, the same argument used to prove (2.6) gives

$$\mathrm{E}\exp(-tG_a\theta_k^+) = \exp\{-a\mathrm{E}\log(1 + t|h_k(g(\xi))|\}.$$

The latter, again by the monotone convergence theorem, tends to $\exp\{-a\mathrm{E}\log(1+t|g(\xi)|\}$, which is then the Laplace transform of $G_a\theta^+$. Under condition (2.3) $\mathrm{E}\log(1 + t|g(\xi)|\}$ is finite for any positive t, so that such a Laplace transform tends to 1 as $t \uparrow 0^+$. This implies that $G_a\theta^+$ and therefore θ^+ are finite a.s..

If, on the other hand, $\mathrm{E}\log\{1 + |g(\xi)|\} = \infty$ so that $\mathrm{E}\log(1 + t|g(\xi)|\} = \infty$ for any positive t, then

$$\Pr\{G_a\theta^+ < \infty\} = \lim_{t\to 0^+} \mathrm{E}\exp(-tG_a\theta^+) = 0.$$

As a consequence $\theta^+ = \infty$ a.s..

Let us prove that (2.4) holds under condition (2.3). Consider $\theta_k = \int h_k(g)\,\mathrm{d}P$. As $h_k(g) \to g$ and $|h_k(g)| \leq |g|$ with $|g|$ integrable, we have by dominated convergence theorem $\theta_k \to \theta$ a.s.. This implies convergence in distribution of $(G_a\theta_k, G_a)$ to $(G_a\theta, G_a)$ and therefore convergence of the corresponding characteristic functions. As h_k is bounded, the characteristic function $\mathrm{E}\exp\{i\,(tG_a\theta_k + sG_a)\}$ is equal to $\exp[-a\mathrm{E}\log\{1 - i(th_k(g(\xi)) + s)\}]$. Furthermore, under condition (2.3), $\mathrm{E}\log\{1 - i(tg(\xi) + s)\}$ is finite, and it is possible to apply dominated convergence theorem to show that such characteristic function converges to $\exp[-a\mathrm{E}\log\{1 - i(tg(\xi) + s)\}]$. This proves relation (2.4).

Furthermore since by (2.6) the characteristic function of $(G_a\theta_m, G_a)$ converges to the characteristic function of $(G_a\theta, G_a)$, this implies $(G_a\theta_m, G_a) \to_d (G_a\theta, G_a)$ and by continuous mapping theorem $\theta_m \to_d \theta$. Finally, relation (2.5) is obtained by setting $s = 0$ in (2.4) and then computing the characteristic function of $G_a\theta$ conditionally on θ. ■

Condition (2.3) for finiteness of θ and relation (2.5) are known (see Regazzini, Guglielmo and Di Nunno, 2002, and references therein). Their proof is based on properties of the Gamma process and its relation with the Dirichlet process. The above proof only uses elementary tools and well-known properties of the Dirichlet distribution.

Attempts at inverting the new derived expression (2.4) to exhibit the underlying density $f_a(r, s)$ for $(R, S) \equiv (G_a\theta, G_a)$ lead to complexities resembling those encountered in Cifarelli and Regazzini (1990) and Regazzini, Guglielmi and Di Nunno (2002). A relative simple formula emerges only for $a = 1$, where the density d_1 of θ is equal to

$$d_1(t) = \pi^{-1}\sin(\pi F_0(t))\exp\Big\{-\int_{-\infty}^{\infty}\log|y - t|\,\mathrm{d}Q_0(y)\Big\}, \tag{2.7}$$

where $F_0(t)$ denotes the distribution function associated with $Q_0 = P_0 g^{-1}$. For this case (θ, G_a) has density $d_1(t)\exp(-s)$, implying a density for (R, S) of the form

$$f_1(r, s) = d_1(r/s)\exp(-s)s^{-1}. \tag{2.8}$$

This also leads to a simulation recipe for a an arbitrary integer as follows. First notice that (R, S) is an infinitely divisible random vector, as the right hand side of (2.4) is a characteristic function for any positive a. Let then $(\bar{R}_i, \bar{S}_i)$ for $i = 1, \ldots, a$ be independently drawn from $f_1(r, s)$. This is accomplished by drawing a unit exponential $\bar{S}_i$ and independently a θ_i from $f_1(t)$ using a rejection-acceptance routine, and then setting $\bar{R}_i = \theta_i \bar{S}_i$. Then (R, S), with $R = \sum_{i=1}^k \bar{R}_i$ and $S = \sum_{i=1}^k \bar{S}_i$, comes from the density $f_a(r, s)$ of (R, S), and $\theta = R/S$ is a simulated outcome from the exact density $d_a(t)$ of θ. This, in particular, allows exact simulation from the posterior distribution of θ whenever the total mass parameter of the prior is an integer number. It is also clear that $f_a(r, s)$ can be expressed as the convolution $f_1 \star \cdots \star f_1$, although this generally produces analytically complicated expressions.

As a further application of relation (2.4) and, in particular, of the infinitely divisible property of (R, S), we may for example rather easily determine the behaviour of θ for a large.

PROPOSITION 1. *Let θ_0 and σ_0 be mean and standard deviation of θ, assumed finite. Then $(a+1)^{1/2}(\theta - \theta_0)$ tends to a normal $(0, \sigma_0^2)$ as a grows.*

PROOF. Take a as an integer, for simplicity, and write θ as $\bar{R}(a)/\bar{S}(a)$, where $\bar{R}(a) = \sum_{i=1}^a \bar{R}_i$ and $\bar{S}(a) = \sum_{i=1}^a \bar{S}_i$, with $(\bar{R}_i, \bar{S}_i)$ being independent from the same distribution, namely the one of (R, S) for the case of $a = 1$. These have mean vector and covariance matrix equal to

$$\begin{pmatrix} \theta_0 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \sigma_0^2 + \theta_0^2 & \theta_0 \\ \theta_0 & 1 \end{pmatrix},$$

respectively, as found by a little analysis. The conclusion follows from the central limit theorem and the delta method. ∎

We note that the result of course must hold with scaling $a^{1/2}$ too; we use $(a+1)^{1/2}$ here to match the exact variance. It may also be pointed out that the representation of θ as a ratio between i.i.d. averages may be used to give suitable modifications to the first order asymptotic result above, for example via saddlepoint approximations. Some such computational schemes would rely on some knowledge of the distribution of (R, S) for the special case of $a = 1$: a formula for the density of such distribution is given in (2.8).

Notice also that this result implies approximate normality of the posterior distribution of the random mean when the number of observations n is large, as the posterior distribution of a Dirichlet process is still a Dirichlet process, with strength parameter $a + n$.

3. Stochastic equations characterisations

The second basic tool used to reach results on the random mean θ is given by a representation of θ as solutions of certain stochastic equations. Such stochastic equations can be derived from the following representation of the Dirichlet process (see Sethuraman and Tiwari, 1982, Sethuraman, 1994). Let

$$P = \sum_{j=1}^{\infty} \gamma_j \delta(\xi_j), \tag{3.1}$$

where the ξ_js are independent from P_0 and independent of the weights, which are constructed in terms of a sequence $\{B_j\}$ of independent Beta$(1, a)$ variables as follows: $\gamma_1 = B_1$, $\gamma_2 = (1 - B_1)B_2$, $\gamma_3 = (1 - B_1)(1 - B_2)B_3$, and so on. Then P is a.s. a Dir(aP_0) process.

As a consequence we have the following distributional equation. In the following, '$=_d$' means equality in distribution.

PROPOSITION 2. *If $Y = g(\xi)$ is such that* $\mathrm{E}\log(1 + |Y|)$ *is finite, then $\theta = \int g\,\mathrm{d}P$ satisfies the stochastic equation*

$$\theta =_d BY + (1 - B)\theta, \quad \textit{where } B \sim \mathrm{Beta}(1, a), \tag{3.2}$$

and where on the right hand side B, Y and θ are independent.

PROOF. The infinite series representation (3.1) allows θ a $\sum_{j=1}^{\infty} \gamma_j Y_j$ representation, the random sum being a.s. convergent exactly under condition (2.3). This leads to

$$\begin{aligned}
\theta &= B_1 Y_1 + (1 - B_1)B_2 Y_2 + (1 - B_1)(1 - B_2)B_3 Y_3 + \cdots \\
&= B_1 Y_1 + (1 - B_1)\{B_2 Y_2 + (1 - B_2)B_3 Y_3 + (1 - B_2)(1 - B_3)B_4 Y_4 + \cdots\} \\
&= B_1 Y_1 + (1 - B_1)\theta',
\end{aligned}$$

where θ' has the very same construction as has θ. ∎

Such 'stochastic equations' imply identities involving characteristic functions or Laplace transforms. Specifically, for g nonnegative, write L_0 and L for the Laplace transforms of respectively Y and θ. Then conditioning first on (B, Y) in (1), and then taking the mean value w.r.t. Y, leads to

$$L(u) = \int_0^1 L_0(ub)L(u(1 - b))\beta_a(b)\,\mathrm{d}b, \tag{3.3}$$

where $\beta_a(b)$ is the Beta$(1, a)$ density. Similarly an identity might be put up using convolution, involving the density f for θ in terms of the density f_0 for Y.

These identities determine in principle L for given L_0 and f for given f_0, although the exact solution might be hard to come by.

Stochastic equation (3.2) is used in the bounded case by Guglielmi (1998) to prove an expression for the generalised Stieltjes transform of θ. Furthermore, related work constructs simulations from θ by exploiting a Markov chain of the form $\theta_i = B_i Y_i + (1 - B_i)\theta_{i-1}$

for $i \geq 2$, with B_i and Y_i independent from respectively the Beta$(1,a)$ and $Q_0 = P_0 g^{-1}$. The Markov chain has the required equilibrium distribution. Guglielmi and Tweedie (2001) and Guglielmi, Holmes and Walker (2001) work with such a chain, following up earlier work by Feigin and Tweedie (1989).

An equivalent stochastic equation, which will be often employed in the following, can be obtained for θ by multiplying both sides of equation (3.2) by an independent Gamma random variable with shape parameter $a+1$. By exploiting independence properties of the Beta distribution one has:

$$\theta G_{a+1} =_d G_1 Y + G_a \theta, \tag{3.4}$$

where θ is independent of G_{a+1} on the left hand side and G_1, G_a, θ and Y are independent on the right hand side, with G_s denoting a Gamma random variable with parameter $(s,1)$.

There is some literature on similarly-structured stochastic equations, but in contexts very different from the present. See Gjessing and Paulsen (1997), Dufresne (1998) and references therein for an integro-differential equation approach and for a list of similar equations with exact solutions.

As a simple example, consider the case $Y \sim$ Beta$(c, 1-c)$, with $0 < c < 1$. Then (3.4) becomes $\theta G_{a+1} =_d G_c + G_a \theta$, where all the random variables on the left and the right hand side are independent. As a direct consequence of Theorem 2 in Dufresne (1998), this equation admits the simple solution $\theta \sim$ Beta$(a+\frac{1}{2}, a+\frac{1}{2})$ when $c = \frac{1}{2}$. This case has also been derived in Cifarelli and Melilli (2000) on the basis of the generalised Stieltjes transform of θ.

Stochastic equation (3.2) can be shown to establish, for any given a, a one-to-one correspondence between the law of a random Dirichlet mean and the distribution Q_0: for any 'permissible' choice of Q_0 the corresponding random mean θ is uniquely determined by (3.2) and, conversely, any random variable distributed as a Dirichlet mean uniquely determines through (3.2) the corresponding distribution $Q_0 = P_0 g^{-1}$. Furthermore stochastic equation (3.4) yields a characterisation, via Hilbert transforms, of Q_0 in terms of θ.

Proposition 3. *Let $P \sim \mathrm{Dir}(aP_0)$ and $\theta = \int g \, \mathrm{d}P$ where g is a real measurable function. For any fixed $a > 0$, there is a one-to-one correspondence between the class $\mathcal{Q} = \{Q_0 = P_0 g^{-1} \colon \int \log(1+|y|) \, \mathrm{d}Q_0(y) < \infty\}$ and the class $\mathcal{L} = \{\mathcal{L}(\theta) \colon \theta \text{ exists finite a.s.}\}$ where $\mathcal{L}(\theta)$ denotes the law of θ. Such a correspondence is given by the distributional equation (3.2) and by the transforms (2.5) and*

$$\mathrm{E}\left(\frac{1}{1-itY}\right) = \mathrm{E}\left(\frac{1}{1-it\theta}\right)^{a+1} \Big/ \mathrm{E}\left(\frac{1}{1-it\theta}\right)^{a} \quad \text{for } t \in \mathcal{R}, \tag{3.5}$$

where $Y = g(\xi) \sim Q_0$.

Proof. If we fix Q_0, then equation (3.2) uniquely defines the law of θ by Lemma 3.3 in Sethuraman (1994), whenever θ exists finite, i.e. whenever $Q_0 \in \mathcal{Q}$. Conversely, let us fix the law of θ. Suppose that both $Y \sim Q_0 \in \mathcal{Q}$ and $Y' \sim Q_0' \in \mathcal{Q}$ satisfy equation (3.2) and therefore equation (3.4). It follows that

$$G_1 Y + G_a \theta =_d G_1 Y' + G_a \theta. \tag{3.6}$$

By computing the characteristic functions of both sides of equation (3.6) and noticing that by (2.5), $\mathrm{E}\exp(itG_a\theta) \neq 0$, one can show that equation (3.6) implies $G_1 Y =_d G_1 Y'$, which in turn implies $Y =_d Y'$.

Relation (3.5) can be derived by computing the characteristic function of the left and the right hand side of equation (3.4). ∎

The existence of a one-to-one correspondence is proved in Lijoi and Regazzini (2001) via a completely different route, by making use of complex theory arguments. Relation (3.5) appears to be new as well as the fact that the stochastic equation (3.2) determines a bijection between random Dirichlet means and the corresponding distribution Q_0.

Another application of stochastic equation (3.2) gives a description of θ in terms of its moments. Expressions for moments and sufficient conditions for their existence are obtained in Regazzini (1998) and Epifani (1999) in terms of complete Bell exponential polynomials. The proofs of such elaborated expressions is rather technical: it is based on the theory of special functions (Regazzini, 1998) and on approximating the Dirichlet process by Bernshteĭn polynomials (Epifani, 1999).

Here we shall derive, as a straightforward consequence of (3.2), a simple recursive formula for direct moments $\mathrm{E}\theta^p$ and centralised moments $\mathrm{E}(\theta-\theta_0)^p$. As a further consequence of the stochastic equations (3.2) and (3.4), we shall also give new necessary and sufficient conditions for existence of such moments.

PROPOSITION 4. *Let $P \sim \mathrm{Dir}(aP_0)$ and $\theta = \int g\,\mathrm{d}P$ where g is a measurable function such that $\mathrm{E}\log(1+|g(\xi)|)$ is finite. Then, for any positive integer p, the following three conditions are equivalent: (1) $\mathrm{E}|\theta|^p < \infty$, (2) $\mathrm{E}|g(\xi)|^p < \infty$, and (3) $\mathrm{E}(\int |g|\,\mathrm{d}P)^p < \infty$. Furthermore, under any of the above conditions, we have*

$$\mathrm{E}(\theta - x)^p = a(p-1)! \sum_{j=0}^{p-1} \frac{1}{j!(a+j)^{[p-j]}} \mathrm{E}(Y-x)^{p-j}\,\mathrm{E}(\theta-x)^j, \tag{3.7}$$

where x is an arbitrary real number and $y^{[p]} = y(y+1)\cdots(y+p-1)$.

PROOF. Let us first prove expression (3.7) assuming that $\mathrm{E}|\theta|^p < \infty$ and $\mathrm{E}|Y|^p < \infty$, where $Y = g(\xi)$. From (3.2),

$$(\theta - x)^p =_d \sum_{j=0}^{p} \binom{p}{j} \{B(Y-x)\}^{p-j}(1-B)^j(\theta-x)^j$$

for any x, implying

$$\mathrm{E}(\theta-x)^p = \frac{1}{1-\mathrm{E}(1-B)^p} \sum_{j=0}^{p-1} \binom{p}{j} \mathrm{E}B^{p-j}(1-B)^j\,\mathrm{E}(Y-x)^{p-j}\,\mathrm{E}(\theta-x)^j.$$

Formula (3.7) then follows by recalling that $\mathrm{E}B^{p-j}(1-B)^j = a(p-j)!\,\Gamma(a+j)/\Gamma(1+a+p)$.

Let us consider now the equivalence of the three conditions. Clearly, (3) implies (1). Let us show that (1) implies (2).

As a consequence of stochastic equation (3.4), condition (1) implies that $\mathrm{E}|\theta G_{a+1}|^p$ and therefore $\mathrm{E}|G_1 Y + G_a \theta|^p$ are finite. By independence of $G_1 Y$ and $G_a \theta$, this implies (see, for example, Lemma 3 of section V.6 in Feller, 1996) that $\mathrm{E}|G_1 Y|^p < \infty$, which in turn implies $\mathrm{E}|Y|^p < \infty$.

We conclude the proof by showing that (3) is a consequence of (2). We prove it by induction. It is easy to check, using the Sethuraman representation, that the results holds for $n = 1$. Suppose now that it is true for an arbitrary n, that is suppose that $\mathrm{E}|Y|^n < \infty$ implies $\mathrm{E}(\theta^+)^n < \infty$, where $\theta^+ = \int |g| \,\mathrm{d}P$, and let us prove it for $n+1$. If $\mathrm{E}|Y|^{n+1} < \infty$ then $\mathrm{E}|Y|^n < \infty$ which then implies, by hypothesis, $\mathrm{E}(\theta^+)^n < \infty$. Consider now, for a positive integer k, the function $t_k(x) = |x| I\{|x| \leq k\} + k I\{|x| > k\}$. Clearly, $0 \leq t_k(x) \leq k$ and $t_k(x) \uparrow |x|$. Moreover, let $\theta_k = \int t_k(g) \,\mathrm{d}P$ and notice that, by the monotone convergence theorem, $\theta_k \uparrow \theta^+$ a.s.. As θ_k and $Y_k = t_k(Y)$ are bounded random variables, they admit moments of any order and we can therefore apply the recursive formula (3.7) obtaining

$$\mathrm{E}(\theta_k)^{n+1} = a\, n! \sum_{j=0}^{n} \frac{1}{j!(a+j)^{[n+1-j]}} \mathrm{E}(Y_k)^{n+1-j}\, \mathrm{E}(\theta_k)^j. \tag{3.8}$$

By the monotone convergence theorem, one then has

$$\mathrm{E}(\theta^+)^{n+1} = a\, n! \sum_{j=0}^{n} \frac{1}{j!(a+j)^{[n+1-j]}} \mathrm{E}(|Y|)^{n+1-j}\, \mathrm{E}(\theta^+)^j,$$

the expression on the right being finite, as $\mathrm{E}|Y|^{n+1}$ and $\mathrm{E}(\theta^+)^n$ are finite. ■

Formula (3.7) gives an easily implementable recursive computational scheme: high order centralised moments can be readily obtained and then used, for example, to obtain accurate approximation of the density of θ. For one such idea, see Section 5.2.

It is also worth noticing the equivalence of condition (1) with the apparently stronger condition (3): such equivalence is by no means true for a general random probability measure.

4. Results for some general classes of distributions

We shall apply here the two representations above introduced to derive results on the random Dirichlet mean when Q_0 belongs to some general classes of distributions.

4.1 Symmetric and unimodal distributions. As a direct consequence of Proposition 3 and of stochastic equation (3.2) we may obtain a simple proof of the following result which characterises the class of symmetric random means. A different proof, based on a contour integral expression of the characteristic function of θ derived through multiple hypergeometric functions, can be found in Lijoi and Regazzini (2001).

PROPOSITION 5. *Suppose θ exists finite. Then θ is symmetric if and only if Q_0 is symmetric.*

PROOF. Suppose that the law of θ is symmetric, i.e. $\theta =_d -\theta$. One has $\theta =_d \beta Y + (1-\beta)\theta$ and $-\theta =_d \beta(-Y)+(1-\beta)(-\theta)$ and by the symmetry of θ follows $\theta =_d \beta(-Y)+(1-\beta)\theta$ which then implies, by Proposition 3, that $Y =_d (-Y)$. An analogous argument shows the converse. ■

A partially analogous result can be obtained using representation (2.2) if one consider symmetric and unimodal distribution. We shall say that a distribution function F is unimodal with vertex x_0 if there exists a real number x_0 such that $F(x)$ is convex for $x < x_0$ and concave for $x > x_0$. This means that F is absolutely continuous, except possibly at x_0, and that its density is monotone in the intervals $\{x < x_0\}$ and $\{x > x_0\}$.

PROPOSITION 6. *Suppose that the distribution of $Y = g(\xi)$ is symmetric and unimodal with vertex c and such that* $\mathrm{E}\log(1+|Y|) < \infty$. *Then the distribution of $\theta = \int g \,\mathrm{d}P$ is symmetric and unimodal with the same vertex.*

PROOF. We can assume without loss of generality $c = 0$, as $\int (g-c)\,\mathrm{d}P = \theta - c$. Consider the approximation (2.2) $\theta_m = \sum_{i=1}^m Y_i\beta_i$ to θ and let us first prove that θ_m is symmetric and unimodal with vertex 0 (denoted by S-U). Clearly, $Y_i\beta_i \,|\, \beta_i$ is S-U. The same is true for $\sum_{i=1}^m Y_i\beta_i \,|\, (\beta_1,\ldots,\beta_m)$, as sum of independent S-U random variables is S-U: see, for example, Theorem 4.5.5 in Lukacs (1970). But this holds also unconditionally, i.e. for θ_m, since it is easy to check that if the conditional distribution of $T \,|\, V$ is a.s. S-U then T is also S-U. The limit θ is symmetric by the Proposition 5 or alternatively by noticing that, since θ_m and $-\theta_m$ converge respectively to θ and $-\theta$ and $\theta_m =_d -\theta_m$, it follows that $\theta =_d -\theta$. Finally, θ is also unimodal being limit of unimodal random variables; see for a proof Theorem 4.5.4 in Lukacs (1970). ■

The converse to Proposition 6 is not true: a counterexample is given by the case $Y \sim \mathrm{Beta}(\frac{1}{2},\frac{1}{2})$ and $\theta \sim \mathrm{Beta}(a+\frac{1}{2}, a+\frac{1}{2})$ considered below equation (3.4).

4.2 STABLE DISTRIBUTIONS. Next proposition gives a characterisation for the family of random means obtained starting from stable distributions: it essentially says that if $Y = g(\xi) \sim Q_0$ has a stable law, then $\theta = \int g\,\mathrm{d}P$ is a scale mixture of Y. Several characterisations of the mixing distribution are also given.

Recall that a general nondegenerate stable law $\mathrm{Stab}(\mu,\sigma,p,\gamma)$ has characteristic function given by

$$\phi(t,\mu,\sigma,p,\gamma) = \exp\{i\mu t - |\sigma t|^p(1+i\gamma K(p,t))\} \tag{4.1}$$

where

$$K(p,t) = (\operatorname{sign} t)\,\tan(\tfrac{1}{2}p\pi) \quad \text{if } p \neq 1,\ 0 < p \leq 2$$

while being equal to $(2/\pi)(\operatorname{sign} t)\log|t|$ if $p = 1$, and $\sigma > 0$, $\mu \in R$, and $-1 \leq \gamma \leq 1$. We shall denote by Φ the class of all stable laws $\mathrm{Stab}(\mu,\sigma,p,\gamma)$ and by Φ^- the subclass of Φ obtained by excluding the case $\{p = 1,\ \gamma \neq 0\}$.

PROPOSITION 7. *Let $Q_0 \in \Phi^-$. This is equivalent to assuming that if $Y, Y_1, \ldots, Y_n$ are i.i.d. random variables from Q_0, then there exist $0 < p \le 2$ and $\mu \in R$ such that one of the following two equivalent conditions holds:*
(a) $\sum_{i=1}^{n} Y_i \sim n^{1/p} Y + \mu(n - n^{1/p}) \quad \text{for } n \ge 1;$
(b) $a_1(Y_1 - \mu) + a_2(Y_2 - \mu) \sim (a_1^p + a_2^p)^{1/p}(Y - \mu) \quad \text{for } a_1,\, a_2 \ge 0.$
Then $\theta = \int g \, \mathrm{d}P$, where $P \sim \mathrm{Dir}(aP_0)$, has distribution equal to

$$\theta - \mu \sim (Y - \mu)\, T,$$

where $Y \sim Q_0$ is independent from the nonnegative random variable T. Furthermore T is the limit in distribution of $T_m = (\sum_{j=1}^{m} \beta_j^p)^{1/p}$ where $(\beta_1, \ldots, \beta_m)$ has a Dirichlet distribution with symmetric parameter $(a/m, \ldots, a/m)$. It can also be expressed as $T^p =_d \sum_{j=1}^{\infty} \gamma_j^p$, where the γ_js are the random weights in representation (3.1). Finally, T is uniquely determined by the following stochastic equation:

$$T^p \sim B^p + (1 - B)^p \, T^p, \tag{4.2}$$

where $B \sim \mathrm{Beta}(1, a)$ is independent of T.

PROOF. It is immediate to see that if $Q_0 \in \Phi^-$ then (b) is satisfied. Furthermore (b) implies (a). The fact that (a) implies $Q_0 \in \Phi^-$ follows from general theory of stable laws (see e.g. Hoffmann-Jorgensen, 1994).

Let us now derive the characteristic function of $\theta_m = \sum_{j=1}^{m} \beta_j Y_j$. By conditioning on the β_js, we have

$$\mathrm{E}\exp(it\theta_m) = \mathrm{E}\phi\Big(t, \mu, \Big(\sum_{j=1}^{m} \beta_j^p\Big)^{1/p} \sigma, p, \gamma\Big),$$

which implies $(\theta_m - \mu) \sim (Y - \mu)\, T_m$, where $Y \sim P_0$ is independent of $T_m = (\sum_{j=1}^{m} \beta_j^p)^{1/p}$. Convergence of T_m^p to $\sum_{j=1}^{\infty} \gamma_j^p$ and relation (4.2) are shown in Hjort and Ongaro (2003). The result then follows from convergence of θ_m to θ. ■

Another characterisation of the distribution of T^p via Laplace transform is given in Hjort and Ongaro (2003), where the asymptotic behaviour of T^p as a tends to infinity is also established.

Special examples of the above result are the Cauchy and the normal distributions corresponding, respectively, to the $p = 1$ and $p = 2$. In particular, for the Cauchy distribution one has $T_m = T = 1$ which gives the known correspondence (Yamato, 1984) $\theta =_d Y$.

A further representation of θ can be given for mixture of Cauchy distributions: if Y is a scaled mixture of Cauchy, then so is θ and the mixing distribution for θ can be written as a random mean from a Dirichlet process with parameter proportional to the distribution mixing the Cauchy.

PROPOSITION 8. *Let $X \sim R$ be a nonnegative random variable and let $Q_0 = P_0 g^{-1}$ be a R-mixture of scaled Cauchys, i.e. Q_0 is the law of XY where $Y \sim Cauchy$ is independent of X. Then $\gamma = \int g \, \mathrm{d}P$, with $P \sim \mathrm{Dir}(aP_0)$, is distributed as*

$$\gamma =_d Y\theta$$

where Y is independent of the nonnegative random variable θ. Furthermore $\theta =_d \int x \, \mathrm{d}P$ where $P \sim \mathrm{Dir}(aR)$ and is therefore uniquely determined by the stochastic equation

$$\theta =_d BX + (1-B)\theta, \quad \text{where } B \sim \mathrm{Beta}(1, a),$$

with X, B and θ independent.

PROOF. Consider $\gamma_m = \sum_{j=1}^m \beta_j \varepsilon_j$, where ε_j is from the Cauchy mixture Q_0, that is, $\varepsilon_j \mid \sigma_j$ is Cauchy (σ_j) where σ_j comes from R. Hence $\mathrm{E}\exp(iu\varepsilon_j) = \mathrm{E}\exp(-|u|\sigma_j)$ and

$$\mathrm{E}[\exp(iu\beta_j\varepsilon_j) \mid \beta] = \mathrm{E}[\exp(-|u|\beta_j\sigma_j) \mid \beta].$$

It follows that $\mathrm{E}\exp(iu\gamma_m) = \mathrm{E}\exp(-|u|\theta_m)$, where θ_m is our usual approximation to θ. This implies that $\mathrm{E}\exp(-|u|\theta_m)$ converges to $\mathrm{E}\exp(-|u|\theta)$, which can be proven to be the characteristic function of $Y\theta$. ■

4.3 MIXTURE DISTRIBUTIONS. Here we derive results for another broad class of interest, giving characterisations of the distribution of θ when the base measure P_0 is a finite mixture distribution. A constructive proof of such a characterisation may be given based on the approximation θ_m to θ. Here we shall give a simpler proof, obtained by checking that given decomposition of θ satisfies relation (2.5).

PROPOSITION 9. *Suppose P_0 has a mixture representation of the form $\sum_{i=1}^k p_i P_{0,i}$ and let $P \sim \mathrm{Dir}(aP_0)$. If g is a measurable function such that $\mathrm{E}\log(1 + |g(\xi)|)$ is finite, then $\theta = \int g \, \mathrm{d}P$ admits the decomposition*

$$\theta = \sum_{i=1}^k D_i \theta_i, \tag{4.3}$$

in which $D = (D_1, \ldots, D_k)$ is Dirichlet $(ap_1, \ldots, ap_k)$ and independent of $\theta_1, \ldots, \theta_k$. These are independent among themselves and $\theta_i = \int g \, \mathrm{d}P_i$, where $P_i \sim \mathrm{Dir}(ap_i P_{0,i})$.

PROOF. It is enough to check that the above given decomposition of θ possesses the right Hilbert transform, i.e. satisfies relation (2.5). Notice that existence of the θ_is follows from the finiteness assumption on $\mathrm{E}\log(1 + |g(\xi)|)$. By conditioning on $\sum_{i=1}^k D_i\theta_i$ we have

$$\mathrm{E}\exp\Big\{it\Big(G_a \sum_{i=1}^k D_i\theta_i\Big)\Big\} = \mathrm{E}\Big(\frac{1}{1 - it\sum_{i=1}^k D_i\theta_i}\Big)^a.$$

From well known independence properties of the Dirichlet distribution follows the distributional equation $G_a(\sum_{i=1}^k D_i\theta_i) =_d \sum_{i=1}^k G_{ap_i}\theta_i$, where on the left hand side G_a is independent of the sum and all the random variables variables on the right hand side are independent. As a consequence we have that the left hand side of the above equation is equal to

$$\mathrm{E}\exp\Big(it\sum_{i=1}^k G_{ap_i}\theta_i\Big) = \exp\Big\{-a\sum_{i=1}^k p_i\int \log(1-itg(x))\,\mathrm{d}P_{0,i}(x)\Big\},$$

completing the proof. ■

Proposition 9 can be applied in several different ways. First one may consider the simple case where $P_{0,i} = P_0$ and $p_i = 1/k$, obtaining a representation of θ as symmetric Dirichlet mixture of i.i.d. copies of a random mean from a Dirichlet process with total mass a/k. This allows to extend knowledge on the distribution of θ from a Dirichlet(cP_0) to the distribution of θ from Dirichlet(kcP_0) for integer k.

For example, when a is an integer one has $\theta = \sum_{i=1}^a D_i\theta_i$ where D is from the flat Dirichlet$(1,\ldots,1)$ and the θ_is are from a Dirichlet process with total mass 1. As the density of such random means is known (see 2.7), this permits to perform exact simulation from θ. Such simulation method is equivalent to the one described below expression (2.8).

As a special case, consider $g(\xi) =_d e^C/(1+e^C)$, where C is a Cauchy distributed random variable, and let $a=1$. Then θ has a uniform distribution on $(0,1)$ (Diaconis and Kemperman, 1996). By applying (4.3) we can derive the density of θ for $a=2$:

$$d_2(t) = -2[(1-t)\log(1-t) + t\log t)].$$

Notice also that if we take the limit for $k\to\infty$ in (4.3) with $P_{0,i} = P_0$ and $p_i = 1/k$, we obtain the Sethuraman sum representation of a random mean; this is because as $k\to\infty$, θ_i converges in distribution to $g(\xi)$ and the symmetric Dirichlet distributed D, after a size-biased sampling reordering, converges in distribution to Sethuraman random weights γ_is (see Section 4 in Billingsley, 1999).

As a second application of Proposition 9 consider partitioning the sample space in separate regions $A_1,\ldots,A_k$ such that $P(A_i) > 0$ and let $P_{0,i}(A) = P_0(A\cap A_i)/P_0(A_i)$ be the normalised restriction of P_0 to A_i. We thus obtain a decomposition of θ in terms of independent random means from Dirichlet processes defined on different regions. For example, by taking $\Omega^+ = \{x\colon g(x)\geq 0\}$ and Ω^- its complement, one has

$$\theta = B\theta^+ - (1-B)\theta^-$$

where $B = \mathrm{Beta}(aP_0(\Omega^+), a(1-P_0(\Omega^+)))$, $\theta^+ = \int_{\Omega^+} g^+\,\mathrm{d}P^+$ and $\theta^- = \int_{\Omega^-} g^-\,\mathrm{d}P^-$. Here g^+ and g^- denote the positive and negative part of g, so that $g = g^+ - g^-$, and P^+ and P^- are Dirichlet processes with parameters $aP/P(\Omega^+)$ and $aP/P(\Omega^-)$ on Ω^+ and Ω^-. This allows studying properties of a general random mean from properties of random means of positive functions, extending in particular the known examples of distributions of random means.

One can then consider the general case where the $P_{0,i}$ are neither identical nor defined on different regions. Let us first take the following simple case: $P_0 = pP_0' + (1-p)\delta(x)$, where x is a fixed point belonging to the sample space. Then θ admits the simple representation

$$\theta =_d g(x) + B\{\theta' - g(x)\},$$

where $B \sim \text{Beta}(ap, a(1-p))$ is independent of $\theta' = \int g\, \mathrm{d}P'$, with $P' \sim \text{Dir}(apP_0')$. For example, if $P_0' = \text{Beta}(\frac{1}{2}, \frac{1}{2})$, $g(x) = 0$ and $p = 1 - \frac{1}{2a}$, with $a > \frac{1}{2}$, we obtain $\theta =_d \text{Beta}(a - \frac{1}{2}, a + \frac{1}{2})$, with an analogous result holding for $g(x) = 1$.

More substantially, the above mixture model can be used to construct nonparametric priors centred at a multimodal distribution. For example, consider the bimodal distribution $P_0 = pP_{0,1} + (1-p)P_{0,2}$, where $P_{0,i} = \text{Cauchy}(\mu_i, \sigma^2)$. Then a direct application of Propositions 7 and 9 gives

$$\theta =_d \mu_2 + B(\mu_1 - \mu_2) + Y,$$

where $Y =_d \text{Cauchy}(0, \sigma^2)$ is independent of B. Alternatively, one can consider mixture of normals, i.e. $P_{0,i} = \text{Normal}(\mu_i, \sigma^2)$, obtaining the representation

$$\theta =_d \mu_2 + B(\mu_1 - \mu_2) + Y\{B^2 T_{ap}^2 + (1-B)^2 T_{a(1-p)}^2\}^{1/2},$$

where B, Y, T_{ap}, $T_{a(1-p)}$ are independent. Here $Y \sim \text{Normal}(0, \sigma^2)$ and $T_c^2 =_d \sum_{j=1}^{\infty} \gamma_j^2$ where the γ_is are the random weights defined in (3.1) in terms of i.i.d. Beta(1,c) variables.

As a further example, consider the posterior distribution of θ given a random sample of observations $X_1, \ldots, X_n$. This is a random mean from a Dirichlet process with strength parameter $a + n$ and probability measure given by the mixture $w_n P_0 + (1 - w_n) F_n$, where $F_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ is the empirical distribution and $w_n = 1/(a+n)$. The random mean from a Dirichlet process with parameter nF_n is $\theta_{F_n} = \sum_{i=1}^n D_i g(X_i)$ where the D_is come from a symmetric Dirichlet $(1, \ldots, 1)$; the density of such random mean can be given an explicit and relatively simple expression (see Cifarelli and Melilli, 2000, example 6). By applying the above proposition it follows that the posterior random mean θ^n can be decomposed as

$$\theta^n =_d B\theta + (1-B)\theta_{F_n} \tag{4.4}$$

where $B \sim \text{Beta}(a, n)$, θ and θ_{F_n} are independent. This gives a useful description of θ^n whenever knowledge is available on the prior random mean θ.

4.4. Mixtures of Dirichlet priors. Finally, Proposition 9 can also be fruitfully applied to make inference in nonparametric hierarchical models, where the Dirichlet process is used as a second stage prior. Such models received considerable attention in the literature (see, for example, West, Müller and Escobar, 1994, Escobar and West, 1998, MacEachern, 1998). They can be described as follows.

A sample $\mathbf{X_n} = (X_1, \dots, X_n)$ is observed. Conditionally on $P \sim \mathrm{Dir}(aP_0)$ and the unobserved random vector $\mathbf{Y_n} = (Y_1, \dots, Y_n)$, the X_is are independent with distribution $F(\cdot\,|\,Y_i)$. Furthermore, $\mathbf{Y_n} = (Y_1, \dots, Y_n)$ is a sample from P, i.e., conditionally on P, the Y_is are i.i.d. with distribution P. Slightly more general version including a vector of parameters can also be accommodated into this framework.

For such hierarchical models, the posterior distribution of P is a mixture of Dirichlet processes (Antoniak, 1974). More specifically, it can be represented by mixing the posterior Dirichlet distribution of P given the sample $\mathbf{Y_n}$ with the distribution $H_{X_n}(\cdot)$ of $\mathbf{Y_n}\,|\,\mathbf{X_n}$. In symbols:

$$P^n = P\,|\,\mathbf{X_n} \sim \int \mathrm{Dir}\Big(aP_0 + \sum_{i=1}^n \delta_{y_i}\Big)\,\mathrm{d}H_{X_n}(y_1, \dots, y_n). \tag{4.5}$$

If the conditional distribution $F(\cdot\,|\,Y_i)$ of $X_i\,|\,Y_i$ admits a density $f(\cdot\,|\,Y_i)$ then H_{X_n} has density proportional to $\prod_{i=1}^n f(X_i\,|\,y_i)$ with respect to the marginal distribution of $\mathbf{Y_n}$. Furthermore, efficient algorithms to simulate from H_{X_n} have been developed (Escobar, 1994, MacEachern, 1994, Escobar and West, 1995 and Ishwaran and James, 2001).

Consider now the posterior random mean $\theta^n = \int g\,\mathrm{d}P^n$. Then, as $\theta^n\,|\,\mathbf{Y_n} =_d \theta\,|\,\mathbf{Y_n}$, the discussion leading to (4.4) gives

$$\theta^n\,|\,\mathbf{Y_n} =_d \theta B + (1-B)\sum_{i=1}^n D_i g(Y_i).$$

If we then average with respect to the distribution of $\mathbf{Y_n}\,|\,\mathbf{X_n}$, we obtain the representation

$$\theta^n =_d \theta B + (1-B)\sum_{i=1}^n D_i g(Z_i),$$

where $(Z_1, \dots, Z_n) \sim H_{X_n}$ is independent of all the other random quantities. It follows that also in this more general hierarchical model we are able to simulate from the posterior mean θ^n whenever simulation procedures are available for the prior random mean θ.

5. Further extensions

Here we briefly outline some further themes that can be worked with using the machinery we have developed in this paper.

5.1. The multidimensional case. Diaconis and Kemperman (1996) mention specifically that there seems to be no results in the literature on the simultaneous distribution of several random means. This would be needed for determining the distribution of a random variance, for example. The methods developed in this paper lend themselves nicely also to the multidimensional framework. The following can be proved along the lines of results arrived at above.

Let $g_1, \dots, g_k$ be nonnegative functions on Ω with finite integrals $\int \log(1+g_i)\,\mathrm{d}P_0$. Then the simultaneous distribution of the p Dirichlet means $\theta_i = \int g_i\,\mathrm{d}P$ is determined by the following transform, which we may term 'the simultaneous Hilbert transform of order a', as follows:

$$\mathrm{E}\exp\Big\{-a\log\Big(1+\sum_{i=1}^k u_i\theta_i\Big)\Big\} = \exp\Big[-a\int\log\Big\{1+\sum_{i=1}^k u_i g_i(\xi)\Big\}\,\mathrm{d}P_0(\xi)\Big].$$

We point out that the two quantities appearing here are identical to the simultaneous Laplace transform $\mathrm{E}\exp(-\sum_{i=1}^k u_i R_i)$, where $R_i = G_a\theta_i$ with G_a a Gamma variable with parameters $(a,1)$, independent of the θ_is, cf. (2.4). We may also view these R_is as integrals of functions w.r.t. a Gamma process. The simultaneous distribution of $(\theta_1,\ldots,\theta_k)$ is determined by the above transform, and can also be characterised via the simultaneous Laplace transform of $(R_1,\ldots,R_k,G_a)$, which becomes

$$\mathrm{E}\exp\Big(-\sum_{i=1}^k u_iR_i - vG_a\Big) = \exp\Big[-a\int\log\Big\{1+\sum_{i=1}^k u_i g_i(\xi)+v\Big\}\,\mathrm{d}P_0(\xi)\Big].$$

5.2. Approximating the density of a random Dirichlet mean. To explain the following method, assume for simplicity of presentation that P_0 is confined to the unit interval. Our task is to approximate the density f of $\theta = \int_0^1 x\,\mathrm{d}P$. To this end, consider

$$f_m(t) = f_0(t)c_m(b)^{-1}\exp(b_1t+b_2t^2+\cdots+b_mt^m) \quad \text{on } [0,1],$$

where f_0 is a suitably chosen start approximation (which could be $f_0(t)=1$) and $c_m(b) = \int_0^1 f_0(t)\exp(b_1t+\cdots+b_mt^m)\,\mathrm{d}t$ the necessary integration constant. The idea is to select the coefficients $b_1,\ldots,b_m$ so as to give optimal approximation quality of f_m to the real f. We do this by minimising the Kullback–Leibler distance $\int f\log(f/f_m)\,\mathrm{d}t$, which is seen to be the same as maximising the function $A_m(b) = \sum_{j=1}^m b_j\xi_j - \log c_m(b)$ with respect to $b=(b_1,\ldots,b_m)$, where $\xi_j = \int t^j f(t)\,\mathrm{d}t$ is the real jth moment of θ. These moments can easily be found numerically via (3.7), say for j up to $m=100$. The maximisation task is also rather easily done using optimisation algorithms available in software packages like `Splus`, helped here by the fact that $A_m(b)$ is concave in b.

This method will often work very well in practice, but we lack precise results about the quality of the resulting approximation. Results reached in e.g. Barron and Sheu (1991) have some relevance, but are typically derived under too restrictive conditions.

References

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

Barron, A.R. and Sheu, C. (1991). Approximations of density functions by sequences of exponential families. *Annals of Statistics* **19**, 1347–1369.

Billingsley, P. (1999). *Convergence of Probability Measures* (2nd ed.). Wiley, New York.

Cifarelli, D.M. and Regazzini, E. (1990). Distribution functions of means of a Dirichlet process. *Annals of Statistics* **18**, 429–442; corrigendum, ibid. (1994) **22**, 1633–1634.

Cifarelli, D.M., Melilli, E. (2000). Some new results for Dirichlet priors. *Annals of Statistics* **28**, 1390–1413.

Diaconis, P. and Kemperman, J. (1996). Some new tools for Dirichlet priors. In *Bayesian Statistics 5*, J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.), 97–106. Oxford University Press, Oxford.

Dufresne, D. (1998). Algebraic properties of Beta and Gamma distributions, and applications. *Advances in Applied Mathematics* **20**, 285–299.

Epifani, I. (1999). Some results on the moments of vector-valued functionals of the Ferguson-Dirichlet process. Quaderno 380/P, Dipartimento di Matematica, Politecnico di Milano.

Escobar, M.D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* **89**, 268–277.

Escobar, M.D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588.

Escobar, M.D. and West, M. (1998). Computing nonparametric hierarchical models. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (eds. D. Day, P. Müller and D. Sinha), Lecture Notes in Statistics **133**, Springer-Verlag, New York, pp. 1–22.

Feigin, P.B. and Tweedie, R.L. (1989). Linear functionals and Markov chains associated with Dirichlet processes. *Mathematics Proceedings of the Cambridge Philosophical Society* **105**, 579–585.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications,* Vol. II. Wiley, New York.

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. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics*, Chernoff Festschrift Volume (H. Rizvi, J.S. Rustagi and D.O. Siegmund, eds.), 287–302. Academic Press, New York.

Gjessing, H. and Paulsen, J. (1997). Present value distributions with applications to ruin theory and stochastic equations. *Stochastic Processes and their Applications* **71**, 123–144.

Guglielmi, A. (1998). A simple procedure calculating the generalized Stieltjies transform of the mean of a Dirichlet process. *Statistics & Probability Letters* **38**, 299–303.

Guglielmi, A., Holmes, C.C. and Walker, S.G. (2001). Perfect simulation involving functionals of a Dirichlet process. To appear.

Guglielmi, A. and Tweedie, R. L. (2001). Markov chain Monte Carlo estimation of the law of the mean of a Dirichlet process. *Bernoulli* **7**, 593–604.

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

Hjort, N.L. (1996). Bayesian approaches to non- and semiparametric density estimation (with discussion contributions). In *Bayesian Statistics 5*, proceedings of the Fifth International València Meeting on Bayesian Statistics (eds. J. Berger, J. Bernardo, A.P. Dawid, A.F.M. Smith), 223–253. Oxford University Press.

Hjort, N.L. (2003). Topics in nonparametric Bayesian statistics [with discussion]. In *Highly Structured Stochastic Systems* (eds. P.J. Green, N.L. Hjort and S. Richardson), Oxford University Press, 455–478.

Hjort, N.L. and Ongaro, A. (2003). On the distribution of random Dirichlet jumps. Submitted for publication.

Hoffmann-Jorgensen, J. (1994). *Probability With a View Toward Statistics,* Chapman & Hall, New York.

Ishwaran, H. and James, L.F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96**, 161–173.

Ishwaran, H. and Zarepour, M. (2002). Exact and approximate sum-representations for the Dirichlet process. *The Canadian Journal of Statistics* **30**, 1–16.

Kallenberg, O. (1986). *Random Measures.* Akademie-Verlag, Berlin and Academic Press, London.

Lijoi, A. and Regazzini, E. (2001). Means of a Dirichlet process and multiple hypergeometric functions. Quaderno IAMI 01.09, CNR-IAMI, Milano.

Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Annals of Statistics* **12**, 351–357.

Lukacs, E. (1970). *Characteristic functions* (2nd ed.). Griffin, London.

MacEachern, S.N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Communications in Statistics – Simulation* **23**, 727–741.

MacEachern, S.N. (1998). Computational methods for mixture of Dirichlet process models. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (eds. D. Day, P. Müller and D. Sinha), Lecture Notes in Statistics **133**, Springer-Verlag, New York, pp. 23–43.

Regazzini, E. (1998). An example of the interplay between statistics and special functions, in *Tricomi's ideas and contemporary applied mathematics, Atti dei Convegni Lincei* **147**, 303–320. Accademia Nazionale dei Lincei e Accademia delle Scienze di Torino.

Regazzini, E., Guglielmi, A. and Di Nunno, G. (2002). Theory and numerical analysis for exact distributions of functionals of a Dirichlet process. *Annals of Statistics* **30**, 1376–1411.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.

Sethuraman, J. and Tiwari, R. (1982). Convergence of Dirichlet measures and the interpretation of their parameter. In *Proceedings of the Third Purdue Symposium on Statistical Decision Theory and Related Topics*, S.S. Gupta and J. Berger (eds.), 305–315. Academic Press, New York.

Walker, S.G., Damien, P., Laud, P.W., and Smith, A.F.M. (1998). Bayesian nonparametric inference for random distributions and related functions (with discussion). *Journal of the Royal Statistical Society* **B 61**, 485–527.

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