On the Distribution of Random Dirichlet Jumps

Nils Lid Hjort and Andrea Ongaro

University of Oslo and Università di Milano-Bicocca

– *September 2003* –

ABSTRACT. The Dirichlet process has been extensively studied over the last thirty years, along with various generalisations, and remains a fundamental tool for nonparametric Bayesian statistics. The probabilistic structure of its jumps has not drawn so much attention in those contexts, however, but has been examined in somewhat unrelated literature, ranging from probabilistic number theory, population genetics, mathematical ecology, and size-biased sampling theory. This paper connects some of these theories and results together, using a new limit type representation of the Dirichlet process. This in particular allows simpler derivation of some of the previous results in the literature. Some new results are also reached.

KEY WORDS: *Poisson-Dirichlet distribution, Dirichlet process, jump sizes, Laplace transform, Lévy process, size-biased sampling, point processes.*

1. Introduction and summary

The Dirichlet process, introduced by Ferguson (1973, 1974), set off a veritable wave of research in nonparametric Bayesian directions. It remains a cornerstone and master tool in that area, also in connection with various more general constructions, as witnessed also by recent overviews over the field, see e.g. Walker, Damien, Laud and Smith (1998) and Hjort (2003).

The growing literature on Dirichlet processes, in various probabilistical and statistical contexts, appears not to have directly studied distributional aspects of the sizes of the jumps $\gamma_1, \gamma_2, \ldots$ that make up these processes, however. Relevant results have been obtained, interestingly, in quite unrelated literature, in contexts ranging from probabilistic number theory (long cycles of random permutations, large prime divisors of random numbers and other combinatorial structures) to population genetics (largest surviving groups with random mutation rates, allele frequencies in a diffusion model), ecology (species abundance models) and size-biased sampling. The distribution of the sizes of the jumps arranged in a descending order is known in such contexts as Poisson-Dirichlet distribution and was introduced by Kingman (1975) (see also Kingman, 1993). Some of the most important results on such distribution are exposited in Billingsley (1999, Section 4; see also his bibliographical notes, p. 267), which however do not touch on statistical or Bayesian connections. These deductions have demanded rather technical and difficult analysis, and Billingsley (p. 54) admits to 'considerable uphill calculations' being necessary; see also comments made by Holst (2002). We are able to obtain these and related results with a rather more pedestrian hill angle being required, exploiting a special construction of the Dirichlet process exhibited in Section 2 below, and a further connection to Gamma processes. Insight into aspects of the Dirichlet jumps also leads to new perspectives for several other related variables. We shall among other results characterise the random variables $W_q = \sum_{j=1}^{\infty} \gamma_j^q$, which have applications in genetics and in Bayesian nonparametrics (namely to random means and configuration of ties in a sample from the Dirichlet process). We shall also investigate the random numbers of jumps sizes fallen in any given region, leading to point processes characterisations of the γ_js; more generally results will be given on variables of the type $\sum_{j=1}^{\infty} f(\gamma_j)$.

The contents of the paper are as follows. We start in Section 2 with several representations of the process; these constructions supplement each other and may be utilised for different purposes. We in particular express the Dirichlet process as a limit of random probabilities with symmetrically defined weights, and relate this representation to normalised Gamma processes. In Section 3 we use this connection to provide an alternative derivation of the distribution of the i biggest Dirichlet jumps.

We go on in Section 4 to results regarding the distribution of the W_q variables mentioned above. These have probability interpretations, and are related to distribution of alleles in genetics models. Our methods allow us a quite straightforward derivation of the limit distribution of W_qs as the prior strength parameter, say a, tends to infinity. Such results have been reached in Joyce, Krone and Kurtz (2002), but our derivation appears to be much simpler. Then in Section 5 and 6 we make connections to the theory of random relabelling and size-biased sampling, leading to precise results about means, variances and higher moments for variables of the general type $\sum_{j=1}^{\infty} f(\gamma_j)$. These connections are also exploited to obtain characterisations of the γ_js interpreted as a point process on the interval $[0, 1]$. As a direct consequence an independent simple proof of an explicit expression of the density of the biggest Dirichlet jumps is exhibited. Our paper ends with a generous list of complementary comments and results.

2. Constructions and representations of the Dirichlet process

Let P_0 be a fixed probability distribution on some sample space and let a be positive. The original definition, from Ferguson (1973), is that P is a Dirichlet process with parameter aP_0 if for any measurable partition $A_1, \ldots, A_k$ the random vector $(P(A_1), \ldots, P(A_k))$ has the Dirichlet distribution with parameters $(aP_0(A_1), \ldots, aP_0(A_k))$. This section briefly reviews various representations of the Dirichlet process and includes a constructive version via a limit of symmetrically defined finite-dimensional Dirichlets.

2.1. The infinite sum representation. An intriguing probabilistic construction is due to Sethuraman and Tiwari (1982). Let

$$P = \sum_{j=1}^{\infty} \gamma_j \delta(\xi_j), \tag{2.1}$$

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; see also Sethuraman (1994).

2.2. Construction via a limit of symmetric distributions. The following construction has some resemblance to the (2.1) representation, but is different in that the finite-dimensional Dirichlet distribution involved is symmetric. This will allow some Dirichlet consequences to be derived in a simpler fashion than with (2.1). Let $\xi_1, \ldots, \xi_m$ be independent from P_0, let independently $(\beta_1, \ldots, \beta_m)$ be symmetric Dirichlet with parameter $(a/m, \ldots, a/m)$, and consider the random distribution

$$P_m = \sum_{j=1}^{m} \beta_j \delta(\xi_j). \tag{2.2}$$

We then have the following.

THEOREM 2.1. *The random probability measure P_m of (2.2) converges in distribution to the Dirichlet process with parameter aP_0, as $m \to \infty$, in the sense of convergence of all finite-dimensional distributions.*

PROOF. It suffices to consider a measurable partition $A_1, \ldots, A_k$, since the distribution of $(P(B_1), \ldots, P(B_l))$ with potentially overlapping subsets can be expressed via convolutions of random probabilities of disjoint sets. Conditional on $\xi_1, \ldots, \xi_m$ it is clear that

$$\big(P_m(A_1), \ldots, P_m(A_k)\big) \sim \mathrm{Dir}\big(a\widehat{P}_m(A_1), \ldots, a\widehat{P}_m(A_k)\big),$$

where $\widehat{P}_m$ is the empirical distribution of points $\xi_1, \ldots, \xi_m$. But $\widehat{P}_m(A)$ tends almost surely to $P_0(A)$, uniformly over all sets A, and the Dirichlet distribution is continuous in its parameters. Hence the finite-dimensional distributions of P_m tend to those of F. ■

The random probability (2.2) has been considered by several authors, independently and in quite unrelated contexts; see references given in Ishwaran and Zarepour (2002), Section 4, and in Hjort (2003, Sections 2 and 3).

When P_m and P are confined to a bounded real interval $[a, b]$, the finite-dimensional convergence suffices for full convergence $F_m \to_d F$ in the Skorokhod topology, writing F_m and F for the random cumulative distributions of P_m and P. The reason is that the necessary tightness of the P_m sequence, see e.g. Billingsley (1999), here follows from the fact that F_m is monotone for each m. This secures $g(P_m) \to_d g(P)$ for each continuous functional g.

It is also worth noting that if the sample space Ω is a locally second countable Hausdorff topological space (for example a Euclidean space), then convergence of finite-dimensional distributions implies, by Kallenberg (1983, theorem 4.2), convergence in distribution of the linear functional $\int f \, \mathrm{d}P_m$ to $\int f \, \mathrm{d}P$ for any real continuous function f with compact support. In fact, it can be proved (Hjort and Ongaro, 2003) that convergence of the above linear functional takes place for any measurable f such that $\int f \, \mathrm{d}P$ exist finite almost surely.

2.3. Relation to Gamma processes. It is well known that the Dirichlet process can be represented as a normalised Gamma process: if H is a gamma process with parameter aP_0 and total value $S = H(\Omega) \sim \mathrm{Gam}(a, 1)$, then $P =_d H/S$ where $P \sim \mathrm{Dir}(aP_0)$ and '$=_d$' means equality in distribution. An analogous construction can be considered for P_m, leading to the following approximation of the Gamma process. 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 the representation

$$P_m = H_m/S_m \quad \text{where} \quad H_m = \sum_{j=1}^{m} G_j \delta(\xi_j). \tag{2.3}$$

Here H_m is not a gamma process, but rather a mixture of such over the possible configurations of the ξ_js. Conditionally on these one has $H_m(A) \sim \mathrm{Gam}(a\widehat{P}_m(A), 1)$, with independence over disjoint sets. Again, the empirical distribution $\widehat{P}_m$ goes a.s. to P_0, which by continuity of the gamma distribution in its parameters implies that $H_m \to_d H$, a gamma process with parameter aP_0. It is also clear that this convergence is simultaneous with $S_m \to_d S$, from which follows $H_m/S_m \to_d H/S =_d P$ once more.

3. The sizes of Dirichlet process jumps

In this section we use the constructions associated with (2.2) and (2.3) above to derive first the simultaneous distribution of the biggest random Gamma process jumps, and then of the biggest random Dirichlet process jumps. This amounts to a new derivation that in several respects appears to be simpler than previous chains of arguments; see again Billingsley (1999, Section 4) for such 'uphill calculations'.

3.1. Weak convergence of H_m and P_m jumps. In Section 2 the random distribution P_m was represented as $H_m/H_m(\Omega)$. The limiting processes P and H are discrete a.s., as is made particularly clear via the representation (2.1), featuring probability weights $\gamma_1, \gamma_2, \ldots$. These do decrease in expected value, but one might nevertheless have $\gamma_3 > \gamma_2$, and so on, making it a complicated task to find the distribution of the biggest jump, for example.

In the following, take P_0 to be free of atoms, so that all ξ_js of (2.1) are distinct a.s. and each γ_j really represents a single jump. In the $P_m = \sum_{j=1}^m \beta_j \delta(\xi_j)$ representation, sort the weights as $\beta_{m,(1)} > \cdots > \beta_{m,(m)}$. We shall identify aspects of the limiting distribution of the largest jumps, and do this via the $\beta_j = G_j/S_m$ representation. Let $G_{m,(1)} > \cdots > G_{m,(m)}$ be the ranked gamma weights. Then $\beta_{m,(i)} = G_{m,(i)}/S_m$, and these are again independent of S_m. In particular,

$$\mathrm{E}G^p_{m,(i)} = \mathrm{E}\beta^p_{m,(i)}\, a(a+1)\cdots(a+p-1) \quad \text{for all } p. \tag{3.1}$$

The plan is to use this for finding all moments for $\beta_{(i)}$, the ith biggest jump in the Dirichlet process.

First we need to show that the $G_{m,(i)}$s and the $\beta_{m,(i)}$s do actually converge to the random ordered jumps of, respectively, the Gamma and the Dirichlet process, as suggested by the convergence of P_m and H_m. A proof of this can be given borrowing arguments from Section 5 of Kingman (1975), as follows. As the distribution of the random jumps of the Gamma and Dirichlet process is independent of the parameter P_0, we shall take, without loss of generality, P_0 uniform on $[0,1]$, so that $H_t \equiv H([0,t]) \sim \mathrm{Gam}(at,1)$ for $0 \le t \le 1$. Define

$$C_{i,m} = H_{i/m} - H_{(i-1)/m} \quad \text{for } i = 1,\ldots,m.$$

Then the $C_{i,m}$s have the same distribution of the $G_{i,m}$s and therefore the ordered values $C_{m,(i)}$ are distributed as the $G_{m,(i)}$s. For any given realization of the Gamma process H, it is not difficult to check that $C_{m,(i)}$ converges, as $m \to \infty$, to the i-th biggest jump $G_{(i)}$ of the process. This implies convergence in distribution of $(G_{m,(1)},\ldots,G_{m,(m)})$ to $(G_{(1)},\ldots,G_{(m)})$ and therefore of the sequences $(G_{m,(1)},\ldots,G_{m,(m)},0,\ldots)$ to the sequence of ordered Gamma process jumps $(G_{(1)},G_{(2)},\ldots)$.

Convergence of $J_m = (\beta_{m,(1)},\cdots,\beta_{m,(m)},0,\ldots)$ to the ordered Dirichlet process jumps $J = (\beta_{(1)},\beta_{(2)},\ldots)$ follows immediately from the above once we notice the following simple facts. The ordered Dirichlet jumps can be represented as $\beta_{(i)} = G_{(i)}/S$ as a consequence of the representation of the Dirichlet process as normalised Gamma process. Analogously, the $\beta_{m,(i)}$s can be represented as $C_{m,(i)}/\sum_{i=1}^m C_{m,(i)}$; furthermore we have $S = H_1 = \sum_{i=1}^m C_{m,(i)} = \sum_{i=1}^\infty G_{(i)}$.

In the following, keep i fixed while m grows, and let $Z_{m,i} = G_{m,(i+1)} + \cdots + G_{m,(m)}$ and $Z_i = \sum_{j=i+1}^\infty G_{(j)}$. Previous arguments imply

$$(G_{m,(1)},\ldots,G_{m,(i)},Z_{m,i}) \to_d (G_{(1)},\ldots,G_{(i)},Z_i).$$

As the $\beta_{(i)}$s can be represented as $G_{(i)}/S$, where $S = G_{(1)}+\cdots+G_{(i)}+Z_i$, their distribution can be obtained by deriving the asymptotic distribution of the sequence of random vectors $(G_{m,(1)},\ldots,G_{m,(i)},Z_{m,i})$, which is the subject of the next lemma.

To this end let us introduce the so-called exponential integral function $E_1(t) = \int_t^\infty x^{-1}\exp(-x)\,\mathrm{d}x$ and the special density $g_a(x)$ on $(0,\infty)$ having Laplace transform

$$\int_0^\infty \exp(-ux)g_a(x)\,\mathrm{d}x = \exp\{-aT(u)\}, \quad \text{where } T(u) = \int_0^u \{1-\exp(-v)\}v^{-1}\,\mathrm{d}v. \tag{3.2}$$

Formula (4.13) in Billingsley (1999) gives $g_a(x)$ in explicit form; terming this explicit form cumbersome is an understatement, however.

LEMMA 3.1. *The simultaneous density of* $(G_{m,(1)},\dots,G_{m,(i)})$ *converges to*

$$k_i(t_1,\dots,t_i) = a^i t_1^{-1}\cdots t_i^{-1}\exp\{-(t_1+\cdots+t_i) - aE_1(t_i)\} \quad \text{on } t_1>\cdots>t_i. \tag{3.3}$$

Furthermore, $(G_{m,(1)},\dots,G_{m,(i)},Z_{m,i})$ *converges in distribution to a random vector with density* $k_i(t_1,\dots,t_i)h(z\,|\,t_i)$*, where*

$$h(z\,|\,t_i) = \exp(-z)g_a(t_i^{-1}z)t_i^{-1}/\exp\{-aT(t_i)\} \quad \text{for } z>0. \tag{3.4}$$

PROOF. The density of $(G_{m,(1)},\dots,G_{m,(i)})$ can be written

$$k_{m,i}(t_1,\dots,t_i) = \frac{m!}{1!\cdots 1!(m-i)!}\gamma_{a/m}(t_1)\cdots\gamma_{a/m}(t_i)\Gamma_{a/m}(t_i)^{m-i} \quad \text{for } t_1>\cdots>t_i>0,$$

where Γ_ε and γ_ε are the cumulative and density for a gamma $(\varepsilon,1)$ variable. An investigation reveals that $\Gamma_\varepsilon(t) = 1-\varepsilon E_1(t)+O(\varepsilon^2)$, and result (3.3) follows.

Proving the additional result on the conditional density of the remaining sum $Z_{m,i}$ given the i largest gamma jumps takes some more efforts. Starting with the density $m!\,\gamma_{a/m}(t_1)\cdots\gamma_{a/m}(t_m)$ for the m ordered G_is, and the density used above for the i largest, one finds the conditional density

$$h_m(t_{i+1},\dots,t_m\,|\,t_1,\dots,t_i) = (m-i)!\,\gamma_{a/m}(t_{i+1})\cdots\gamma_{a/m}(t_m)/\Gamma_{a/m}(t_i)^{m-i}$$

on $t_1>\cdots>t_m$. But this shows that the $m-i$ smallest ordered contributions behave as the order statistics vector of a sample of size $m-i$ from the density $\gamma_{a/m}(t)/\Gamma_{a/m}(t_i)$ on $[0,t_i]$. In other words, $Z_{m,i}$ can be represented as as sum of $m-i$ independent variables from a density with Laplace transform

$$\int_0^{t_i}\frac{\exp(-ux)}{\Gamma_{a/m}(t_i)}\frac{1}{\Gamma(a/m)}x^{a/m-1}\exp(-x)\,\mathrm{d}x = \frac{\Gamma_{a/m}((u+1)t_i)}{\Gamma_{a/m}(t_i)}\frac{1}{(u+1)^{a/m}}.$$

The conditional Laplace transform of $Z_{m,i}$ is therefore

$$\mathrm{E}\{\exp(-uZ_{m,i})\,|\,t_1,\dots,t_i\} = \Big(\frac{\Gamma_{a/m}((u+1)t_i)}{\Gamma_{a/m}(t_i)}\Big)^{m-i}\frac{1}{(u+1)^{a(1-i/m)}},$$

which is seen to converge to $\exp\{-aM(u\,|\,t_i)\}$, where

$$\begin{aligned} M(u\,|\,t_i) &= -E_1(t_i)+E_1((u+1)t_i)+\log(u+1) \\ &= \int_{t_i}^{(u+1)t_i}\{y^{-1}-y^{-1}\exp(-y)\}\,\mathrm{d}y = T((u+1)t_i)-T(t_i). \end{aligned}$$

But this is demonstrably also the Laplace transform of the density (3.4). ■

We note that result (3.3) also may be shown following general Lévy process theory as presented in Ferguson (1973, 1974), using the Lévy measure of the Gamma process; see also Section 7.3 where other results connecting (3.3) to the point process representation of a Gamma process are presented.

The arguments used here are however more direct and use only elementary tools. Ferguson's methods do not apply to the study of jumps of the Dirichlet process itself, which is our next concern.

3.2. Distribution of largest jumps. Results above make us ready for studying the largest Dirichlet jumps.

THEOREM 3.1. *The i biggest Dirichlet process jumps have simultaneous density*

$$\kappa_i(b_1,\ldots,b_i) = a^i\Gamma(a)\exp(a\gamma)b_1^{-1}\cdots b_{i-1}^{-1}b_i^{a-2}g_a((1-b_1-\cdots-b_i)/b_i) \tag{3.5}$$

for $1 > b_1 > \cdots > b_i$, $b_1+\cdots+b_i < 1$, *where* g_a *is the special density determined by (3.2).*

PROOF. One passes from the density of the $i+1$ variables $G_{(1)},\ldots,G_{(i)},Z_i$ given by Lemma 3.1 to that of the $i+1$ transformed variables $\beta_{(1)} = G_{(1)}/S,\ldots,\beta_{(i)} = G_{(i)}/S$ and $S = G_{(1)}+\cdots+G_{(i)}+Z_i$ via elementary calculations, involving a Jacobian determinant, giving

$$k_i(sb_1,\ldots,sb_i)h(s(1-b_1-\ldots-b_i)\,|\,sb_i)s^i.$$

Using (3.3) and (3.4) in conjunction with the independently proved relation $T(u)-E_1(u) = \log u+\gamma$, one arrives at $\kappa_i(b_1,\ldots,b_i)$ of (3.5) times the $\mathrm{Gam}(a,1)$ density of S. This proves the required formula and in addition independence between jump sizes and S. ■

Integrating out the other variables in (3.5) to reach a separate formula for the density $d_i(b)$ of $\beta_{(i)}$ involves special properties of the intricate density g_a, and is not easy. Formula (4.16) in Billingsley (1999) provides one still cumbersome expression (again, exposited in a very different context). We may more easily derive the full moment sequences, however, exploiting (3.1), the point being that is easier to derive an expression for the ith largest jump for the gamma process. With arguments used to prove (3.3) and (3.4) one finds that $G_{(i)}$ has density

$$\ell_i(t) = \{a^{i-1}/(i-1)!\}\,at^{-1}E_1(t)^{i-1}\exp\{-t-aE_1(t)\}. \tag{3.6}$$

Combining this with the limit version of (3.1) one establishes

$$\mathrm{E}(\beta_{(i)})^p = \frac{1}{a^{[p]}}\frac{a^{i-1}}{(i-1)!}\int_0^\infty t^{p-1}aE_1(t)^{i-1}\exp\{-t-aE_1(t)\}\,\mathrm{d}t \quad \text{for } p=1,2,3,\ldots. \tag{3.7}$$

Figures 1 and 2 display respectively the means and the standard deviations of the four largest jumps, as a function of parameter a.

– *Figures 1 and 2 approximately here* –

REMARK 3.1. A distribution on a bounded interval is identified by its sequence of moments, so in principle (3.7) suffices to determine the density of $\beta_{(i)}$. But our moment sequence (3.6) agrees fully with the corresponding one exhibited in Billingsley's equation (4.17). This gives another proof of his complicated formula (4.16) for $d_i(b)$. This independent proof uses only the simpler (3.3) part of Lemma 3.1, and not the more laborious (3.4). Similarly one may with more efforts arrive at formulae for all product moments for $\beta_{(i)}$s, using the $G_{(i)} = \beta_{(i)} S$ connection as in (3.1), and identify them as those found by Griffiths (1979, p. 145) for a density which turns out to be equivalent to our (3.5) above, thus giving another proof of this formula. ■

We reiterate that result (3.7) has been proven elsewhere, notably in Arratia, Barbour and Tavaré (1997) and Billingsley (1999), but in very different contexts. Our contribution regarding this is to give simpler proofs via the Gamma process and the symmetric approximations of sections 2.2 and 2.3.

4. The distribution of W_qs

This section is concerned with aspects of the random quantities

$$W_q = \sum_{j=1}^{\infty} \gamma_j^q \quad \text{for } q = 1, 2, 3, \ldots, \tag{4.1}$$

defined in terms of the weights γ_j appearing in the (2.1) representation. It is clear that $W_1 = 1$ and that $1 \geq W_2 \geq W_3 \geq \cdots$. These quantities are important for several reasons.

First, they have probability interpretations; W_q is the random probability that all observations $X_1, \ldots, X_q$ in a sample of size q from the Dirichlet process prior P are equal. They also lead to different interpretations in the different contexts where the Dirichlet process jumps are used. This is relevant, in particular, in hierarchical clusters models where the structure of ties in a sample from P determines the structure of groups. A further example comes from genetics where the Dirichlet process jumps give the equilibrium distribution for certain alleles diffusion models. Here the W_q is the probability that q genes randomly sampled according to the alleles frequencies specified by the equilibrium distribution are all of the same type. This gives a measure of the qth order population homozygosity. From a totally different perspective, the W_qs are also connected to the distributions of random Dirichlet means (cf. Section 7.1).

4.1. On the exact distribution of W_qs. An approach towards understanding and analysing the W_qs is via the construction of Section 2.2, which in view of weak convergence $J_m \to_d J$, cf. the previous section, leads to

$$W_{m,q} = \sum_{j=1}^{m} \beta_j^q = \sum_{j=1}^{m} \beta_{m,(j)}^q \to_d W_q = \sum_{j=1}^{\infty} \beta_{(j)}^q = \sum_{j=1}^{\infty} \gamma_j^q. \tag{4.2}$$

Also, $W_q = U_q/S^q = U_q/U_1^q$ in terms of

$$U_{m,q} = \sum_{j=1}^{m} G_{m,(j)}^q \to_d U_q = \sum_{j=1}^{\infty} G_{(j)}^q \quad \text{for} \quad q = 1, 2, \ldots. \tag{4.3}$$

Consequently, the distribution of the W_qs is fully characterised by the distribution of the U_qs, which is the the subject of the next proposition.

PROPOSITION 4.1. *The Laplace transform of the random vector* $U = (U_1, \ldots, U_q)$ *is given by*

$$\mathrm{E}\exp(-v_1 U_1 - v_2 U_2 - \cdots - v_q U_q) = \exp\{-aF(v_1, v_2, \ldots, v_q)\},$$

where

$$F(v_1, v_2, \ldots, v_q) = -\gamma - \int_0^\infty \log x \exp\{-((v_1+1)x + v_2 x^2 + \cdots + v_q x^q\} \\ (1 + v_1 + 2v_2 x + \cdots + q v_q x^{q-1})\,\mathrm{d}x,$$

with $\gamma = -\Gamma'(1) = 0.5772...$ *being the Euler constant.*

PROOF. From convergence of $(G_{m,(1)}, \ldots, G_{m,(m)}, 0, \ldots)$ to $(G_{(1)}, G_{(2)}, \ldots)$, we have joint convergence in distribution of the quantities $\sum_{i=1}^m G_{m,(i)}^q = \sum_{i=1}^m G_i^q$ to U_q, where G_i are i.i.d. Gamma random variables with parameter a/m. We shall therefore obtain the Laplace transform of U as limit of the Laplace transform of the quantities $\sum_{i=1}^m G_i^q$. To this end we have

$$\begin{aligned}\mathrm{E}\exp\Big(-\sum_{l=1}^q v_l G_i^l\Big) &= \int_0^\infty \frac{x^{a/m-1}}{\Gamma(a/m)} \exp\{-(1+v_1)x - v_2 x^2 - \cdots - v_q x^q\}\,\mathrm{d}x \\ &= \int_0^\infty \frac{x^{a/m}}{\Gamma(a/m+1)} \exp\{-(1+v_1)x - v_2x^2 - \cdots - v_q x^q\} \\ &\qquad (1 + v_1 + 2v_2 x + \cdots + q v_q x^{q-1})\,\mathrm{d}x\end{aligned}$$

via partial integration. Now use $x^{a/m} \doteq 1 + (a/m)\log x$ and $\Gamma(1 + a/m) \doteq 1 - (a/m)\gamma$ to reach

$$\mathrm{E}\exp\Big\{-\sum_{l=1}^q v_l \Big(\sum_{i=1}^m G_i^l\Big)\Big\} = \Big\{\mathrm{E}\exp\Big(-\sum_{l=1}^q v_i G_i^l\Big)\Big\}^m \to \exp\{-aF(v_1, v_2, \ldots, v_q)\}.$$

This proves our claim. ■

Note from the proposition that U is infinitely divisible in a. If e.g. $(\widetilde{U}_1, \ldots, \widetilde{U}_q)$ stems from the above construction, for $a = 2$, then

$$(\widetilde{U}_1, \ldots, \widetilde{U}_q) =_d (U_1 + U_1', \ldots, U_q + U_q'),$$

where U and U' are independent and come from the distribution with $a = 1$.

4.2. Stochastic equations and moments of the W_qs. Using the representation of γ_js inherent in (2.1), we may write

$$\begin{aligned} W_q &= B_1^q + (1-B_1)^q B_2^q + (1-B_1)^q (1-B_2)^q B_3^q + \cdots \\ &= B_1^q + (1-B_1)^q \{B_2^q + (1-B_2)^q B_3^q + (1-B_2)^q (1-B_3)^q B_4^q + \cdots\}, \end{aligned}$$

which is seen to imply the stochastic equation

$$W_q =_d B^q + (1-B)^q W_q. \tag{4.4}$$

Here B is a Beta $(1,a)$ variable, and for the right hand side B and W_q are independent. It is worthwhile to note here that the stochastic equations (4.4) hold jointly for $W_1, \ldots, W_q$. Notice also that, by Lemma 3.3 of Sethuraman (1994), (4.4) uniquely determines the distribution of W_q.

This representation has a number of consequences. Among them is the possibility of simulating realisations of W_q, using a Markov chain Monte Carlo scheme of the type

$$W_{q,n+1} = B_{n+1}^q + (1-B_{n+1})^q W_{q,n} \quad \text{for } n = 1,2,3,\ldots,$$

where the B_ns are i.i.d. Beta $(1,a)$ variables. One may also use (4.4) to compute the full moment sequence, recursively. Specifically, let

$$M_a(b,c) = \mathrm{E}B^b(1-B)^c = a\frac{\Gamma(1+b)\Gamma(a+c)}{\Gamma(1+a+b+c)},$$

where B is a Beta$(1,a)$ variable. Then we derive

$$\mathrm{E}W_q = M_a(q,0)/\{1-M_a(0,q)\}, \tag{4.5}$$

and

$$\mathrm{E}W_q^r = \sum_{j=1}^{q} \binom{r}{j} M_a(qj,0) M_a(0,qrj)\, \mathrm{E}W_q^{r-j} \Big/ \{1-M_a(0,qr)\} \quad \text{for } r \geq 2. \tag{4.6}$$

4.3. Asymptotics for growing a. In genetics, the distribution of the Dirichlet process jumps is used as stationary distribution of the neutral infinite alleles models. Here the parameter a is interpreted as the mutation rate. Motivated by questions arising in problems of genetic diffusion Joyce, Krone and Kurtz (2002) study the asymptotic behaviour of the Dirichlet jumps when a goes to infinity. In particular, they give results on the (random) probability of any given configuration of groups of alleles for a sample drawn from a population where the alleles frequencies are distributed as Dirichlet process jumps. The asymptotic behaviour of such a probability is determined by the asymptotics of the W_qs of (4.1). The main result of their paper is based on the proof of asymptotic normality of the W_qs, which is given in Theorem 1. Their proof is long and rather technical, requiring several preliminary results based on martingale and stochastic integral theory.

As a consequence of results in Section 4.1, we are able to derive the asymptotic normality of the W_qs in a straightforward fashion, making use only of elementary tools, as follows. Proposition 4.1 implies that U is an infinitely divisible random vector. In particular, for any integer a, $U = U(a)$ can be written as a sum of a i.i.d. copies of $U(1)$. A direct application of the central limit theorem and of the delta method then gives the asymptotic distribution of the $W_q = W_q(a)$s.

THEOREM 4.1. *Let $Z_q(a) = \sqrt{a}\{a^{q-1}W_q(a) - (q-1)!\}$. Then as $a \to \infty$,*

$$Z(a) = (Z_2(a), Z_3(a), \ldots) \to_d Z,$$

where $Z = (Z_2, Z_3, \ldots)$ is a random sequence characterised by the following multivariate normal marginal distributions: $(Z_2, \ldots, Z_q) \sim \mathrm{N}(0, B)$, with $B_{i,j} = (i+j-1)! - i!j!$.

PROOF. It is enough to prove convergence of the first q components of Z, for an arbitrary integer q. The multivariate central limit theorem gives

$$\sqrt{a}\Big(\frac{U(a)}{a} - \mathrm{E}U(1)\Big) \to_d \mathrm{N}(0, \Sigma),$$

where Σ is the covariance matrix of $U(1)$. Noticing that $\mathrm{E}U_q = \mathrm{E}W_q\,\mathrm{E}U_1^q$ and using (4.5), one has $\mathrm{E}U(1) = (1, 1, 2!, \ldots, (q-1)!)$. Similarly, using $\mathrm{E}U_qU_l = \mathrm{E}W_qW_l\,\mathrm{E}U_1^{q+l}$ and a bi-dimensional version of (4.4) it easy to check that $\Sigma_{i,j} = (i+j-1)!$.

Let now $V(a) = U(a)/a$ and consider the following transformation g of $V(a)$:

$$g(V(a)) = \Big(\frac{V_2}{V_1^2}, \frac{V_3}{V_1^3}, \ldots, \frac{V_q}{V_1^{q-1}}\Big) = (aW_2, a^2W_3, \ldots, a^{q-1}W_q).$$

Via the delta method one obtains

$$\sqrt{a}\{g(V(a)) - g(\mathrm{E}U_1)\} \to_d \mathrm{N}(0, \nabla g^T \Sigma \nabla g),$$

where ∇g indicates the matrix of first derivatives of g evaluated at $\mathrm{E}U_1$. Simple algebra leads to the result. ∎

5. Size-biased sampling and the distribution of ordered jumps

Size-biased sampling is a random relabelling scheme which has applications in many different fields: population genetics (Ewens, 1979, 1990), where the interest lies in alleles frequencies; mathematical ecology (Engen, 1979), where the interest lies in species abundances; Bayesian nonparametrics (Ongaro, 2003), where it is used for the derivation of posterior distributions; and various probabilistic contexts such as heap processes (Donnelly, 1991), recursive splitting of intervals (Lloyd and Williams, 1988) and cycle structure of random permutations (Billingsley, 1999). It can be informally understood as follows.

5.1. Size-biased sampling. The size-biased sampling $(\beta_1^s, \beta_2^s, \ldots)$ of a sequence of random weights $(\beta_1, \beta_2, \ldots)$ is given by the following random relabelling scheme: β_1^s is equal to β_i with probability β_i; having chosen $\beta_1^s = \beta_i$ as above, β_2^s is equal to β_j, for $j \neq i$, with conditional probability $\beta_j/(1-\beta_i)$; given β_1^s and β_2^s, β_3^s is equal to β_l, where $l \notin \{i, j\}$, with conditional probability $\beta_l/(1-\beta_i-\beta_j)$, and so on.

If β_i represents the random frequency of the ith group of a certain population, in the new labelling, the ith group will become the first one with probability β_i; conditionally on this first choice, the j-th group, for $j \neq i$, will become the second one with probability $\beta_j/(1-\beta_i)$, and so on.

It can be shown (see for example Donnelly and Joyce, 1989) that the distribution of the random weights γ_i appearing in the Sethuraman representation (2.1) of the Dirichlet process is invariant under size-biased sampling. By taking the size-biased version of Sethuraman weights and using invariance we obtain the following result.

PROPOSITION 5.1. *Let $\mathcal{S}^k$ be the k-dimensional simplex $\mathcal{S}^k = \{\mathbf{x}_k \in \mathbf{R}^k : \sum_{i=1}^k x_i < 1, x_i > 0 \text{ for } i = 1, \ldots, k\}$ and let f be a positive measurable function defined on $\mathcal{S}^k$. Then*

$$\begin{aligned}\sum_{\substack{i_1,\ldots,i_k\\ \text{distinct}}} \mathrm{E}[f(\beta_{(i_1)}, \ldots, \beta_{(i_k)})] &= \sum_{\substack{i_1,\ldots,i_k\\ \text{distinct}}} \mathrm{E}[f(\gamma_{i_1}, \ldots, \gamma_{i_k})] \\ &= \mathrm{E}\Big[f(\gamma_1, \ldots, \gamma_k)\frac{(1-\gamma_1)\cdots(1-\gamma_1-\cdots-\gamma_{k-1})}{\gamma_1\cdots\gamma_k}\Big].\end{aligned} \tag{5.1}$$

Furthermore, the latter expression is equal to

$$\int_{\mathcal{S}^k} f(x_1, \ldots, x_k)\, a^k \frac{(1-x_1-\cdots-x_k)^{a-1}}{x_1\cdots x_k}\, \mathrm{d}x_1\cdots \mathrm{d}x_k. \tag{5.2}$$

PROOF. The conditional distribution of the size-biased sampling $(\gamma_1^s, \ldots, \gamma_k^s)$ of $(\gamma_1, \gamma_2, \ldots)$ is defined as follows (see e.g. Ongaro, 2003):

$$\begin{aligned}&\Pr\{(\gamma_1^s, \ldots, \gamma_k^s) \in A_k \mid \gamma_1, \gamma_2, \ldots\} \\ &\qquad = \sum_{\substack{i_1,\ldots,i_k\\ \text{distinct}}} I_{A_k}(\gamma_{i_1}, \ldots, \gamma_{i_k})\, \gamma_{i_1} \frac{\gamma_{i_2}}{1-\gamma_{i_1}} \cdots \frac{\gamma_{i_k}}{1-\gamma_{i_1}-\cdots-\gamma_{i_{k-1}}},\end{aligned}$$

where $k \geq 1$, A_k is a suitable measurable set and $I_{A_k}(\cdot)$ denotes the indicator function of the set A_k. By taking expectation and using invariance of the distribution of the γ_is we have

$$\begin{aligned}\Pr\{(\gamma_1, \ldots, \gamma_k) \in A_k\} &= \Pr\{(\gamma_1^s, \ldots, \gamma_k^s) \in A_k\} \\ &= \sum_{\substack{i_1,\ldots,i_k\\ \text{distinct}}} \mathrm{E}\Big[I_{A_k}(\gamma_{i_1}, \ldots, \gamma_{i_k})\, \gamma_{i_1} \frac{\gamma_{i_2}}{1-\gamma_{i_1}} \cdots \frac{\gamma_{i_k}}{1-\gamma_{i_1}-\cdots-\gamma_{i_{k-1}}}\Big].\end{aligned}$$

This implies that, for any positive measurable function h, it holds

$$\mathrm{E}[h(\gamma_1, \ldots, \gamma_k)] = \sum_{\substack{i_1,\ldots,i_k\\ \text{distinct}}} \mathrm{E}\Big[h(\gamma_{i_1}, \ldots, \gamma_{i_k})\, \gamma_{i_1} \frac{\gamma_{i_2}}{1-\gamma_{i_1}} \cdots \frac{\gamma_{i_k}}{1-\gamma_{i_1}-\cdots-\gamma_{i_{k-1}}}\Big];$$

by choosing

$$h(\gamma_1,\ldots,\gamma_k)=f(\gamma_1,\ldots,\gamma_k)\frac{(1-\gamma_1)\cdots(1-\gamma_1-\cdots-\gamma_{k-1})}{\gamma_1\cdots\gamma_k}$$

one obtains (5.1). The density q of $(\gamma_1,\ldots,\gamma_k)$ can be derived by standard calculations from the density of the Beta independent variables B_j (see expression 2.1):

$$q(\gamma_1,\ldots,\gamma_k)=a^k\frac{(1-\gamma_1-\cdots-\gamma_k)^{a-1}}{(1-\gamma_1)\cdots(1-\gamma_1-\cdots-\gamma_{k-1})}.$$

This proves formula (5.2). ■

Notice that the quantities in (5.2) fully identify the distribution of the ordered jumps $\beta_{(i)}$s. This is because from (5.2) one can recover the distribution of the γ_is; the distribution of the $\beta_{(i)}$s is then obtained by ordering the γ_is. The potential of Proposition 5.1 will be demonstrated through some simple examples in the following section and, in a more systematic way, in Section 6.

5.2. Illustrations. By setting $k=$ in Proposition 5.1 we obtain:

$$\mathrm{E}\sum_{j=1}^{\infty}f(\gamma_j)=\mathrm{E}\sum_{j=1}^{\infty}f(\beta_{(j)})=\mathrm{E}\frac{f(\gamma_1)}{\gamma_1}=\int_0^1\frac{f(x)}{x}\,a(1-x)^{a-1}\,\mathrm{d}x.$$

We will illustrate this for a few examples.

First, setting $f(x)=I_{(y,1]}(x)$ corresponds to the variable $N_y=\sum_{j=1}^{\infty}I\{\gamma_j>y\}$, the number of jumps bigger than a threshold y. For this variable we find

$$\begin{aligned}\mathrm{E}N_y&=\sum_{j=1}^{\infty}\Pr\{\gamma_j>y\}=\sum_{j=1}^{\infty}\Pr\{\beta_{(j)}>y\}\\&=\sum_{1\le j<1/y}\Pr\{\beta_{(j)}>y\}=\int_y^1\frac{1}{x}\,a(1-x)^{a-1}\mathrm{d}x.\end{aligned}$$

As a consequence, we have the following relation on the marginal densities $d_i(y)$ of the $\beta_{(i)}$s:

$$\sum_{1\le i<1/y}d_i(y)=\frac{1}{y}\,a(1-y)^{a-1},$$

which can also be expressed as

$$\sum_{i=1}^{n}d_i(y)=\frac{1}{y}\,a(1-y)^{a-1}\quad\text{on}\quad\frac{1}{n+1}\le y<\frac{1}{n}.$$

For example, for $y\ge\frac{1}{2}$ one has $d_1(y)=ay^{-1}(1-y)^{a-1}$.

A second illustration takes $f(x)=xI\{x\ge y\}$, for which we find

$$\mathrm{E}\sum_{j=1}^\infty \gamma_j I\{\gamma_j\ge y\}=\int_y^1 a(1-x)^{a-1}\,\mathrm{d}x=(1-y)^a.$$

When simulating Dirichlet processes using the (2.1) representation it is in practice necessary to truncate the sum, and the above result may be used to establish a satisfactory truncation level, securing a small enough value of the tail of the series with high enough probability.

Yet another illustration is for $f(x)=x^q$ with $q>0$. Then

$$\mathrm{E}W_q=\mathrm{E}\sum_{j=1}^\infty \gamma_j^q=a\frac{\Gamma(a)\Gamma(q)}{\Gamma(a+q)},$$

agreeing with results of Section 4. More generally, let $f(x_1,\ldots,x_k)=x_1^{c_1}\cdots x_k^{c_k}$ where the c_is are positive constants. Then

$$\mathrm{E}\sum_{\substack{i_1,\ldots,i_k\\ \text{distinct}}}\gamma_{i_1}^{c_1}\cdots\gamma_{i_k}^{c_k}=a^k\frac{\Gamma(c_1)\ldots\Gamma(c_k)\Gamma(a)}{\Gamma(c_1+\cdots+c_k+a)}.$$

When $k=2$ the proposition gives

$$\mathrm{E}\sum_{i\ne j}f(\gamma_i,\gamma_j)=\int\int_{x_1+x_2\le 1}f(x_1,x_2)\,a^2\frac{(1-x_1-x_2)^{1-a}}{x_1x_2}\,\mathrm{d}x_1\,\mathrm{d}x_2.$$

This allows us to find also the variance of N_y above, as

$$\mathrm{E}N_y^2=\mathrm{E}N_y+\int\int_{\substack{x_1+x_2\le 1\\ x_1\ge y,x_2\ge y}}a^2\frac{(1-x_1-x_2)^{1-a}}{x_1x_2}\,\mathrm{d}x_1\,\mathrm{d}x_2. \tag{5.3}$$

6. Higher moments and the multivariate frequency spectrum

We found in the previous section that the random relabelling theory and size-biased theory could be utilised to find formulae for moments of $\sum_{j=1}^\infty f(\beta_{(j)})$ type variables. Here we take this theory further, going on to product and higher order moments, tying in our results with the so-called multivariate frequency spectrum

$$t_k(x_1,\ldots,x_k)=a^k\frac{(1-x_1-\cdots-x_k)^{a-1}}{x_1\cdots x_k}. \tag{6.1}$$

This term is used by Watterson (1976), where it is derived as a limit from certain finite neutral alleles models. It is given the following interpretation: $t_k(\mathbf{x}_k)\,\mathrm{d}x_1\cdots\mathrm{d}x_k$ represents the expected product of the numbers of alleles whose frequencies are in the small intervals

$(x_1, x_1+\mathrm{d}x_1), \ldots, (x_k, x_k+\mathrm{d}x_k)$, which in the limit becomes the probability that there are frequencies lying in those intervals. Griffith (1988) also uses the frequency spectrum and more precisely integrals of the type displayed in (5.3) to express the distribution function of $\beta_{(i)}$ and the probability distribution of the variable N_y. Series expansion of such integrals useful for numerical evaluation are also given.

Our work below leads to new results and insights for the multivariate frequency spectrum, showing in particular its relevance for describing various aspects of the Dirichlet process jumps interpreted as a point process on the interval $[0,1]$.

6.1. Higher moments and the jump counting process. Using Proposition 5.1, one can obtain all moments of random variables of the type $Y_f = \sum_{j=1}^{\infty} f(\gamma_j)$. For example, one has

$$\mathrm{E}Y_f^3 = \mathrm{E}\Big\{\sum_i f(\gamma_i)^3 + 3 \sum_{\substack{i,j \\ \text{distinct}}} f(\gamma_i)^2 f(\gamma_j) + \sum_{\substack{i,j,l \\ \text{distinct}}} f(\gamma_i) f(\gamma_j) f(\gamma_l)\Big\}.$$

This leads to

$$\mathrm{E}Y_f^3 = \int_0^1 f^3(x) t_1(x)\,\mathrm{d}x + 3\int_{\mathcal{S}^2} f(x_1)^2 f(x_2) t_2(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 + \int_{\mathcal{S}^3} f(x_1)f(x_2)f(x_3)t_3(x_1,x_2,x_3)\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3.$$

A generalisation of this argument produces the following formula for the generic n-th moment:

$$\mathrm{E}Y_f^n = \sum_{k=1}^n \sum_{(n_1,\ldots,n_k)\in D_n} \binom{n}{n_1 \cdots n_k} \frac{1}{m_1! \cdots m_n!} \int_{\mathcal{S}^k} f(x_1)^{n_1} \cdots f(x_k)^{n_k} t_k(x_1,\ldots,x_k)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k,$$

where D_n is the set of all positive integers $i_1 \geq \cdots \geq i_k \geq 1$, such that $i_1 + \ldots + i_k = n$ and m_i is the number of n_js equal to i. The above given moments formula fully characterises the distribution of Y_f when this is bounded.

In particular, consider $f(x) = q_1 I\{x \in A_1\} + \cdots + q_k I\{x \in A_k\}$ where the q_is are positive constants and the A_is are measurable subsets of $[0,1]$. Then the above formula determines the distribution of Y_f which in turn determines the distribution of the random vector $(N(A_1), \ldots, N(A_k))$ representing the number of jumps fallen in the subsets A_i. The latter distribution completely identifies the distribution of the γ_is or of the $\beta_{(i)}$s viewed as a point process on $[0,1]$. This gives another characterisation of the distribution of the Dirichlet jumps in terms of integrals of type (5.2).

Similar reasoning leads to expressions for joint moments of the type $\mathrm{E}(Y_f^{n_1} Y_g^{n_2})$ as a function of the integrals (5.2). For example, one has

$$\mathrm{E}(Y_f Y_g) = \int_0^1 f(x)g(x)t_1(x)\,\mathrm{d}x + \int_{\mathcal{S}^2} f(x_1)g(x_2)t_2(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2.$$

The study of variables of the form $N(A) = \sum_{j=1}^{\infty} I(\gamma_j \in A)$, which are a particular instance of Y_f type variables, turns out to be particularly fruitful for reaching new insight into the Dirichlet jumps viewed as a point process. It also gives a better understanding of the role of the frequency spectrum t_k.

If A contains an open interval $(0, \epsilon)$ around zero then $N(A) = \infty$ a.s.. Suppose on the other hand that $N(A)$ is bounded away from zero, i.e. $\inf(A) = c > 0$. Then $N(A)$ is a bounded integer valued variable such that $N(A) \leq 1/c$ a.s.. It therefore admits moments of any order, which can be computed using the general formula for $\mathrm{E}Y_f^n$ given above. It can more easily be described through its factorial moments, as follows.

Let $m_{[n]}(A) = \mathrm{E}N(A)^{[n]}$ be the n-th factorial moment of $N(A)$. It is easy to check that

$$N(A)^{[n]} = \sum_{\substack{i_1, \ldots, i_n \\ \text{distinct}}} I\{\gamma_{i_1} \in A\} \cdots I\{\gamma_{i_n} \in A\}.$$

By taking expectation and using Proposition 5.1 one then has

$$m_{[n]}(A) = \int_{\mathcal{S}^n \cap A^{(n)}} t_n(x_1, \ldots, x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n,$$

where $A^{(n)}$ is the product set $A \times \cdots \times A$. Clearly, $m_{[n]}(A) = 0$ if $n > 1/c$.

It is then possible to use the factorial moments to obtain other characterisations of $N(A)$, such as the probability generating function and the probability distribution. In particular,

$$p_n(A) \equiv \Pr\{N(A) = n\} = \frac{1}{n!} \sum_{k=0}^{\infty} \frac{1}{k!} (-1)^k \int_{\mathcal{S}^{n+k} \cap A^{(n+k)}} t_{n+k}(x_1, \ldots, x_{n+k}) \, \mathrm{d}x_1 \cdots \mathrm{d}x_{n+k}. \tag{6.2}$$

Notice that the series appearing in the above expressions are in fact finite sums as the $(n + k)$dimensional integral appearing in such expressions vanishes for $n + k > 1/c$. The latter formula generalises Theorem 4 in Griffiths (1988), which only holds for $A = (y, 1]$.

The case $n = 0$ in (6.2) is of special interest because it corresponds to the so-called avoidance function $\Pr\{N(A) = 0\}$, which gives a further full characterisation of the jumps γ_is viewed as a point process on $[0, 1]$.

By choosing A equal to $A_y = (y, 1]$ in (6.2), one can obtain information on the distribution of the ordered jumps $\beta_{(j)}$s. Clearly $\{N(A_y) = n\} = \{\beta_{(n+1)} \leq y < \beta_{(n)}\}$ and $\{N(A_y) \leq n\} = \{\beta_{(n+1)} \leq y\}$; it follows in particular that we can easily obtain expressions for the distribution function of $\beta_{(j)}$ for $j = 1, \ldots,$ in analogy to those shown in Griffiths (1988), Theorem 2. For example, by setting $n = 0$ we have

$$p_0(A_y) = \Pr\{\beta_{(1)} \leq y\} = 1 + \sum_{1 \leq k < 1/y} \frac{1}{k!} (-1)^k \int_{\mathcal{S}^k \cap A_y^{(k)}} t_k(x_1, \ldots, x_k) \, \mathrm{d}x_1 \cdots \mathrm{d}x_k. \tag{6.3}$$

The above results can be generalised to the vector $(N(A_1), \ldots, N(A_k))$, where the A_is are disjoint measurable subsets of $[0,1]$, leading to a new interpretation of the function t_n. An extension of the argument used to derive $\mathrm{E}N(A)^{[n]}$ shows that

$$\mathrm{E}\Big\{N(A_1)^{[n_1]} \cdots N(A_k)^{[n_k]}\Big\} = \int_{\mathcal{S}^n \cap A^{(n_1)} \times \cdots \times A^{(n_k)}} t_n(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n,$$

where $n = n_1 + \cdots + n_k$. It follows that if we set $t_n = 0$ outside the n-dimensional simplex $\mathcal{S}^n$, then t_n is the n-factorial moment density of the point process formed by the jumps γ_is; this means that it is the density with respect to the Lebesgue measure on $[0,1]^{[n]}$ of the factorial moment measure $M_{[n]}$ defined on rectangles by

$$M_{[n]}(A^{(n_1)} \times \cdots \times A^{(n_k)}) = \mathrm{E}\Big\{N(A_1)^{[n_1]} \cdots N(A_k)^{[n_k]}\Big\}.$$

Thus the function t_n has the following meaning (see Daley and Vere-Jones, 1988): $t_n(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_k$ is the probability that one jump is located in each of the infinitesimal intervals $\mathrm{d}x_1, \ldots, \mathrm{d}x_k$, in agreement with the above given interpretation by Watterson (1976).

In analogy with the univariate case, from the factorial moment measure one can derive the probability generating functional of the jumps point process and a (new) description of its distribution. The probability generating functional $G(h)$ of the jumps point process can be defined as $G(h) = \mathrm{E}\{\Pi_i h(\gamma_i)\}$, for suitable complex-valued functions h such that $|h(x)| \leq 1$. An expansion of $G(h)$ in terms of the factorial moment densities along the lines of Daley and Vere-Jones (1988) gives

$$G(1+\eta) = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{\mathcal{S}^k} \eta(x_1) \cdots \eta(x_k) t_k(x_1, \ldots, x_k)\, \mathrm{d}x_1 \cdots \mathrm{d}x_k,$$

where $\eta(x) = 1 - h(x)$, provided the series on the right is convergent when $\eta(x)$ is replaced with its absolute value. This has applications to the study of random means $\theta = \int g\, \mathrm{d}P = \sum_{j=1}^{\infty} g(\xi_j)\gamma_j$ from a Dirichlet process P. Let $k(t) = \mathrm{E}\exp\{itg(\xi_i)\}$ be the characteristic function of the random variable $g(\xi_i)$. Then, by conditioning on the γ_is, the characteristic function of θ can be written as

$$\mathrm{E}\exp(it\theta) = \mathrm{E}\{\Pi_{j=1}^{\infty} k(t\gamma_j)\}.$$

That is, the characteristic function of θ is equal to probability generating functional $G(1+q)$, with $q(x) = k(tx) - 1$, and can be computed, in principle, by the above formula.

A description of the distribution of the point process formed by the jumps can be obtained by giving the distribution of $N(A)$ for any set A with finitely many points and then, conditionally on $N(A) = n$, the symmetric distribution $\Pi_n(A)$ of the n random points lying in A. A formula for the probability distribution $p_n(A)$ of $N(A)$ has been given in

(6.2), whereas the density π_n with respect to Lebesgue measure of the distribution $\Pi_n(A)$ can be written as follows (see again Daley and Vere-Jones, 1988):

$$\pi_n(x_1,\ldots,x_n) = \frac{1}{n!\,p_n(A)}\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\int_{A^{(k)}} t_{n+k}(x_1,\ldots,x_n,y_1,\ldots,y_k)\,\mathrm{d}y_1\cdots\mathrm{d}y_k.$$

In particular, if we choose A equal to A_y, a change of variables in the integrals together with expression (6.3) yield the formula

$$\begin{aligned}\pi_n(x_1,\ldots,x_n) &= \frac{t_n(x_1,\ldots,x_n)}{n!\,p_n(A_y)}\Big\{1+\sum_{1\le k<d}\frac{(-1)^k}{k!}\int_{A^{(k)}_{1/d}} t_k(y_1,\ldots,y_k)\,\mathrm{d}y_1\cdots\mathrm{d}y_k\Big)\\ &= \frac{1}{n!\,p_n(A_y)}t_n(x_1,\ldots,x_n)\ \Pr\Big\{\beta_{(1)}\le\frac{y}{1-x_1\ldots-x_n}\Big\},\end{aligned}$$

where $d=(1-x_1\ldots-x_n)/y$; this gives the density of the (unordered) n biggest jumps given that there are exactly n jumps bigger than y.

6.2. Back to the ordered jumps. From the above expression one can derive the density of the n ordered biggest jumps through the following operations. First one can obtain the unconditional density of the biggest unordered jumps by multiplying π_n by $p_n(A)$ and setting y equal to the minimum of the x_is. The density of the ordered biggest jumps is then obtained by multiplying it by $n!$ and further restricting the domain of the density to the set $\{x_1>\ldots>x_n\}$. This leads to the following expression:

$$\begin{aligned}\kappa_n(x_1,\ldots,x_n) &= t_n(x_1,\ldots,x_n)\Big\{1+\sum_{1\le k<b}\frac{(-1)^k}{k!}\int_{A^{(k)}_{1/b}} t_k(y_1,\ldots,y_k)\,\mathrm{d}y_1\cdots\mathrm{d}y_k\Big)\\ &= t_n(x_1,\ldots,x_n)\ \Pr\Big\{\beta_{(1)}\le\frac{x_n}{1-x_1\ldots-x_n}\Big\},\end{aligned}$$

where $b=(1-x_1\ldots-x_n)/x_n$. This gives an independent simple derivation for the density of the biggest jumps; one can check that such expression is equivalent to the one given in formulas (4.13) and (4.15) in Billingsley (1999).

6.3. Back to the symmetrically distributed Dirichlet jumps. The above results have also consequences for the symmetric random weights β_is in the limiting representation of Section 2.2, proving the following strong form of convergence.

PROPOSITION 6.1. *Let $\beta_{1,m},\ldots,\beta_{m,m}$ have a symmetric Dirichlet distribution with parameter $(a/m,\ldots,a/m)$ and let the γ_i be the random weights in Sethuraman representation (2.1). Then as m tends to ∞,*

$$\sum_{\substack{j_1,\ldots,j_k\\ \text{distinct}}}\mathrm{E}[f(\beta_{j_1,m},\ldots,\beta_{j_k,m})]\to\sum_{\substack{i_1,\ldots,i_k\\ \text{distinct}}}\mathrm{E}[f(\gamma_{i_1},\ldots,\gamma_{i_k})],\qquad(6.4)$$

where the indexes j_is run from 1 to m and f is a positive measurable function defined on $\mathcal{S}^k$ such that $\int_{\mathcal{S}^k} f(x_1,\ldots,x_k)\,t_k(x_1,\ldots,x_k)\,\mathrm{d}x_1\cdots\mathrm{d}x_k$ is finite.

PROOF. By Proposition 5.1 we have to prove that the left hand side converges to $\int_{\mathcal{S}^k} f(x_1,\ldots,x_k) t_k(x_1\cdots x_k)\,\mathrm{d}x_1\cdots\mathrm{d}x_k$. By symmetry of the $\beta_{i,m}$s, for $m>k$, one has

$$\sum_{\substack{j_1,\ldots,j_k\\ \text{distinct}}} \mathrm{E}[f(\beta_{j_1,m},\ldots,\beta_{j_k,m})] = m^{[k]} \frac{\Gamma(a)}{\Gamma(a/m)^k \Gamma(a-ak/m)} \int_{\mathcal{S}^k} f c_m \,\mathrm{d}x_1\cdots\mathrm{d}x_k$$

where $c_m(x_1\cdots x_k) = (x_1\cdots x_k)^{a/m-1}(1-x_1-\cdots-x_k)^{a-ak/m-1}$. The factor in front of the integral is easily seen to converge to a^k whereas the function c_m goes to $t_k\, a^{-k}$. Furthermore it easy to check that c_m increases to its limit in the region $\{x_1\cdots x_k < (1-x_1-\cdots-x_k)^k\}$ and decreases to the same limit on the complementary region. The result then follows by splitting the integral in two pieces corresponding to the two regions and then applying separately dominated convergence theorem (for example in the version given in Hoffmann-Jorgensen, 1994). ∎

That this form of convergence is stronger than convergence in distribution of the corresponding point processes is made clear in the following corollary.

COROLLARY 6.1. *Let N_m be the finite point process on $[0,1]$ formed by the random points $\beta_{1,m},\ldots,\beta_{m,m}$ which have a symmetric Dirichlet distribution with parameter $(a/m,\ldots,a/m)$ and let N be the point process formed by the Dirichlet process jumps γ_is. Then relation (6.4) implies convergence in distribution of N_m to N.*

PROOF. Consider the point process M_m on $[1,\infty)$ obtained by taking the reciprocals of the points of N_m, i.e. M_m is formed by the random points $1/\beta_{i,m}, i=1,\ldots,m$ and analogously define M as the point process formed by taking the reciprocals of the points of N. It is enough to prove convergence in distribution of M_m to M. Notice that M is a boundedly finite point process (i.e. it has a finite number of points in any bounded set), so that it suffices to prove convergence in distribution of $(M_m(A_1),\ldots,M_m(A_k))$ to $(M(A_1),\ldots,M(A_k))$ for bounded measurable sets A_i (see for example Theorem 9.1.VI in Daley and Vere-Jones, 1988). Moreover the A_is can be taken to be disjoint, as the general case can be proved by writing the possibly overlapping sets as a union of disjoint sets and then exploiting additivity of M_m and M. As $M_m(A)$ and $M(A)$ are bounded random variables for any bounded set A, convergence in distribution of the $M_m(A_i)$s is implied by convergence of joint moments. The latter convergence follows from convergence of factorial moments $\mathrm{E}\big[M_m(A_1)^{[n_1]}\cdots M_m(A_k)^{[n_k]}\big]$, as ordinary moments can be expressed as finite linear combinations of factorial moments. But this is a consequence of Proposition 6.1 if we choose

$$f(x_1,\ldots,x_n) = \prod_{i=1}^{n_1} I\{1/x_i\in A_1\} \prod_{i=n_1+1}^{n_1+n_2} I\{1/x_i\in A_2\} \quad\cdots\quad \prod_{i=n-n_k+1}^{n} I\{1/x_i\in A_k\}$$

where $n=n_1+\cdots+n_k$. To see this one can simply check that, for such a choice of f and for disjoint sets A_is, it holds

$$M_m(A_1)^{[n_1]}\cdots M_m(A_k)^{[n_k]} = \sum_{\substack{j_1,\ldots,j_n\\ \text{distinct}}} f(\beta_{j_1,m},\ldots,\beta_{j_n,m}),$$

and analogously for M. ■

7. Concluding remarks

In this section we provide various concluding comments and additional results.

7.1. Application to random means. Consider the random Dirichlet mean $\theta = \int g\, dP$ where P is a Dirichlet process with parameter aP_0 and g is a measurable function such that θ exist finite a.s.. There are many papers in the literature studying aspects of the distribution of θ, see e.g. Cifarelli and Regazzini (1990), Diaconis and Kemperman (1996), Cifarelli and Melilli (2000), Guglielmi and Tweedie (2001), Regazzini, Guglielmi and Di Nunno (2002) and Hjort (2003).

Note that θ may be represented as $\sum_{j=1}^{\infty} g(\xi_j)\gamma_j$, from (2.1) and also as the distributional limit of $\theta_m = \int g\, dP_m = \sum_{j=1}^{m} g(\xi_j)\beta_j$, as mentioned in Section 2.2; there are therefore relevant connections to results about jumps and ordered jumps, from the present article.

To make one such connection, consider the problem of characterising the distribution of $\theta = \int x\, \mathrm{d}P$ when $P \sim \mathrm{Dir}(aP_0)$ and P_0 is standard normal. Then $\theta_m = \sum_{j=1}^{m} \xi_j\beta_j$ with ξ_js being independent and standard normal, which means that θ_m given the β vector is $\mathrm{N}(0, W_{2,m})$, where $W_{2,m} = \sum_{j=1}^{m} \beta_j^2$. In Section 4.1 we showed that $W_{m,2}$ converges in distribution to W_2. This implies that θ can be represented as the following random scale mixture of normals:

$$\theta =_d X\sqrt{W}_2,$$

where X is a standard normal independent of W_2. It follows that θ has density $\int_0^1 \phi(u/\sigma)$ $h(\sigma)\,\mathrm{d}\sigma$, where $h(\sigma)$ is the density of $W_2^{1/2}$. Knowledge of the Laplace transform given in Proposition 4.1 and of the moment sequence (4.6) can then be used to derive exact or approximate expressions for the density of W_2 and therefore of θ.

More generally, suppose that $X \sim P_0$ is distributed as a stable random variable of order q, $0 < q \leq 2$, and position parameter μ. Then it can be shown (Hjort and Ongaro, 2003) that

$$\theta - \mu =_d (X - \mu)W_q^{1/q},$$

with X independent of W_q. The theory developed in Section 4 can readily be extended to cover the case of q real, thus providing useful information for the stable distribution case.

7.2. Relation to Bernshteĭn polynomials. For a single set A, consider $P_m(A)$ of Section 2.2, which conditional on the ξ_js is a Beta $(a\widehat{P}_m(A), a\widehat{P}_m(A^c))$. It is in other words a mixture of Betas with binomial weights;

$$\Pr\{P_m(A) \in C\} = \sum_{y=0}^{m} B(C; ay/m, a(1-y/m))\mathrm{bin}_m(y,p) = \mathrm{E}\, B(C; aY/m, a(1-Y/m)),$$

in terms of a Y which is binomial (m,p) with $p = P_0(A)$, and where $B(C;\alpha,\beta)$ indicates the probability for a $\mathrm{Beta}(\alpha,\beta)$ variable to belong to a set C. Thus our approximation $\Pr\{P_m(A)\in C\}$ is nothing but the Bernshteĭn polynomial approximation to the continuous function $h(p) = B(C; ap, a(1-p))$ at the point $p = P_0(A)$.

To assess the quality of the approximation involved, assume that $|h(q) - h(p)| \le M|q-p|$ for all p and q; one may show that the h under study, for given C, has a bounded derivative, and M may be taken as the maximum value of $|h'(p)|$. One finds

$$\begin{aligned}|\mathrm{E}_p h(Y/m) - h(p)| &\le \mathrm{E}_p|h(Y/m) - h(p)|\, I\{|Y/m - p| < \delta\}\\ &\qquad + \mathrm{E}_p|h(Y/m) - h(p)|\, I\{|Y/m-p| \ge \delta\}\\ &\le M\delta + p(1-p)/(m\delta^2) \le M\delta + \tfrac14/(m\delta^2).\end{aligned}$$

This is minimised for $\delta = (2Mm)^{-1/3}$, giving

$$\max_A |\Pr\{P_m(A)\in C\} - \Pr\{F(A)\in C\}| \le (3/2)2^{-1/3}M^{2/3}m^{-1/3}.$$

More generally, let $\mathrm{Dir}(C;\alpha_1,\ldots,\alpha_k)$ indicate the probability that a Dirichlet random vector with parameters $(\alpha_1,\ldots,\alpha_k)$ lands in the set C. Then

$$\begin{aligned}\Pr\{(P_m(A_1),\ldots,P_m(A_k))\in C\} &= \sum_{y_1,\ldots,y_k} \mathrm{Dir}(C; ay_1/m,\ldots,ay_k/m)\\ &\qquad\qquad \times \mathrm{mnom}_m(y_1,\ldots,y_k;p_1,\ldots,p_k)\\ &= \mathrm{E}\,\mathrm{Dir}(C; aY_1/m,\ldots,aY_k/m),\end{aligned}$$

in terms of a multinomial vector $(Y_1,\ldots,Y_k)$ with parameters m and $p_j = P_0(A_j)$ for $j = 1,\ldots,k$. This would be the generalised Bernshteĭn mth degree polynomial of $k-1$ variables, used to approximate the continuous function $h(p_1,\ldots,p_{k-1}) = \mathrm{Dir}(C; ap_1,\ldots,ap_k)$.

Assume $|h(p) - h(q)| \le M\|p-q\|$ for all vectors p and q; this holds when M is the maximum over all $\|\partial h(p)/\partial p\|$, which can be shown to be finite for given set C. Then

$$\max_{A_1,\ldots,A_k} |\Pr\{(P_m(A_1),\ldots,P_m(A_k))\in C\} - \mathrm{Dir}(C;ap_1,\ldots,ap_k)| \le M\delta + 1/(m\delta^2),$$

and the right hand side is minimised for $\delta = (2/M)^{1/3}m^{-1/3}$ with minimal value $2^{1/3}(3/2)$ $M^{2/3}m^{-1/3}$. One may sharpen the error bound here by turning to a higher order Markov inequality replacing the ingredient in the above argument which employs the Chebyshov inequality. With $\Pr\{\|Y/m - p\| \ge \delta\} \le \mathrm{E}\|Y/m-p\|^4/\delta^4$, for example, some additional work leads to an error bound of size $O(\delta + (m^2\delta^4)^{-1})$, which leads to a sharper error bound of size $O(m^{-2/5})$.

7.3. On the distribution of $(G_{(1)},\ldots,G_{(i)},Z_i)$. The distribution of $(G_{(1)},\ldots,G_{(i)},Z_i)$ given by (3.3) and (3.4) can be understood as follows. Conditionally on $G_{(i)}$, the random vector $(G_{(1)},\dots,G_{(i-1)})$ and Z_i are independent. In particular, $(G_{(1)},\dots,G_{(i-1)})\,|\,\{G_{(i)}=t_i\}$ is distributed as the order statistics from a random sample of size $i-1$ from the density

$$\frac{1}{E_1(t_i)}\frac{1}{x}\exp(-x) \quad \text{on} \quad x>t_i,$$

and $Z_i\,|\,G_{(i)}=t_i$ has density given by (3.4). Furthermore, the density of $G_{(i)}$ is given in (3.6). If we make the transforms $Y_j=E_1(G_{(j)})$, so that $Y_1<\cdots<Y_i$, it is straightforward to check from (3.3) that $(Y_1,\dots,Y_{i-1})\,|\,Y_i$ is distributed as the order statistics from a random sample of size $i-1$ from a uniform random variable on $(0,Y_i)$ and Y_i is a $\mathrm{Gam}(a,1)$ random variable. The Y_js can also be represented as follows: $(Y_1,\dots,Y_i)\sim(T_1,\dots,T_i)$ where $T_j=W_1+\ldots+W_j$ and the W_js are i.i.d. random variables with exponential distribution of parameter a. This implies that

$$(G_{(1)},\dots,G_{(i)})\sim(E_1^{-1}(T_1),\dots,E_1^{-1}(T_i)).$$

In principle, these results can be employed to derive the distribution or to simulate from the distribution of any random quantity based on the ordered weights $\beta_{(j)}$s. In particular, notice that the random variable $Z_i\,|\,G_{(i)}$ is infinitely divisible and can be simulated using techniques developed, for example, by Damien, Laud and Smith (1995) and Wolpert and Ickstadt (1998).

The above results also imply that the $G_{(i)}$s form a Poisson point process with intensity $\frac{a}{y}\,e^{-y}$, proving once more that they are distributed as the random jumps of a Gamma process.

7.4. Yet more on the ordered Dirichlet jumps. From expression (3.5) one has that the random vector

$$\beta_{(1)},\dots,\beta_{(i-1)}\,|\,\beta_{(i)}=b_i,\beta_{(1)}+\cdots+\beta_{(i-1)}=b_{i-1}^{+}$$

has density proportional to $b_1^{-1}\cdots b_{i-1}^{-1}$, where $b_{i-1}=b_{i-1}^{+}-b_1-\cdots-b_{i-2}$, on the set $b_1>\cdots>b_i$. Such a distribution is independent of a so that $(\beta_{(i)},\beta_{(1)}+\cdots+\beta_{(i-1)})$ can be thought of as a sufficient statistics for a.

Furthermore, the distribution of $(\beta_{(i)},\beta_{(1)}+\cdots+\beta_{(i-1)})$ can be recovered by considering the transform $(1/(1+T+W),W(1+T+W))$, where

$$(W,T)=\Big(\frac{\beta_{(1)}+\cdots+\beta_{(i-1)}}{\beta_{(i)}},\frac{1-\beta_{(1)}-\cdots-\beta_{(i)}}{\beta_{(i)}}\Big)=\Big(\frac{G_{(1)}+\cdots+G_{(i-1)}}{G_{(i)}},\frac{Z}{G_{(i)}}\Big).$$

The density and Laplace transform of (W,T) can be easily derived by conditioning on $G_{(i)}$ and using results of Section 5.1.

Starting from the distribution of the random vector $\beta_{(1)},\dots,\beta_{(i-1)}\,|\,\beta_{(i)}=b_i,\beta_{(1)}+\cdots+\beta_{(i-1)}=b_{i-1}^{+}$, one can also derive the distribution of the normalised ordered weights $Q_1,\dots,Q_{i-1}$, with $Q_j=\beta_{(j)}/(\beta_{(1)}+\cdots+\beta_{(i-1)})$. In particular, one can show that

$Q_1, \ldots, Q_{i-1} \mid Q_i = q_i$ has density proportional to $q_1^{-1} \cdots q_{i-1}^{-1}$, where $q_{i-1} = 1 - q_1 - \cdots - q_{i-2}$, on the set $q_1 > \cdots > q_i$. It follows that, conditionally on Q_i, the distribution of the relative sizes of the ordered jumps is not influenced by the parameter a. The distribution of $Q_i = 1/W$ can then be derived by conditioning on $G_{(i)}$ as mentioned above.

References

Arratia, R., Barbour, A.D. and Tavaré, S. (1997). Random combinatorial structures. *Notices of the American Mathematical Society* **44**, 903–910.

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.

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

Damien, P., Laud, P.W. and Smith, A.F.M. (1995). Approximate random variate generation from infinitely divisible distributions with applications to Bayesian inference. *Journal of the Royal Statistical Society Series B* **57**, 547–563.

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.

Donnelly, P. (1991). The heaps process, libraries and size-biased permutations. *Journal of Applied Probability*, **28**, 322–335.

Donnelly, P. and and Joyce, P. (1989). Continuity and weak convergence of ranked and size-biased permutations on the infinite simplex. *Stochastic Processes and their Applications* **31**, 89–103.

Engen, S. (1979). *Stochastic Abundance Models with Emphasis on Biological Communities and Species Diversity.* Chapman and Hall, London.

Ewens, W.J. (1979). *Mathematical Population Genetics.* Springer-Verlag, New York.

Ewens, W.J. (1990). Population genetics theory – the past and the future. In *Mathematical and Statistical Developments of Evolutionary Theory*, ed. S. Lessard. Kluwer, Dordrecht.

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.

Griffiths, R.C. (1979). On the distribution of allele frequencies in a diffusion model. *Theoretical Population Biology* **15**, 140–158.

Griffiths, R.C. (1988). On the distribution of points in a Poisson Dirichlet process. *Journal of Applied Probability* **25**, 336–345.

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.

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). Characterisations of random Dirichlet means. Manuscript.

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

Holst, L. (2002). The Poisson–Dirichlet distribution and its relatives revisited. Technical report, Department of Mathematics, University of Stockholm.

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

Joyce, P., Krone, S.M. and Kurtz, T.G. (2002). Gaussian limits associated with the Poisson–Dirichlet distribution and the Ewens sampling formula. *The Annals of Applied Probability* **12**, 101–124.

Kallenberg, O. (1983). *Random Measures* [3rd ed.]. Akademie-Verlag, Berlin.

Kingman, J.F.C. (1975). Random discrete distributions [with discussion]. *Journal of the Royal Statistical Society Series B* **37**, 1–22.

Kingman, J.F.C. (1993). *Poisson Processes.* Oxford University Press.

Lloyd, C.J. and Williams, E.J. (1988). Recursive splitting of an interval when the proportions are identical and independent random variables. *Stochastic Processes and their Applications*, **28**, 111–122.

Ongaro, A. (2003). Size-biased sampling and discrete nonparametric Bayesian inference. Technical report, Dipartimento di Statistica, Università di Milano Bicocca. Submitted for publication.

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.

Watterson, G.A. (1976). The stationary distribution of the infinitely-many neutral alleles diffusion model. *Journal of Applied Probability* **13**, 639–651.

Wolpert, R.L. and Ickstadt, K. (1988). Simulation of Lévy random fields. In *Practical Nonparametric and Semiparametric Bayesian Statistics*, Lecture Notes in Statistics 133, Springer, New York, 227–242.