Statistica Sinica 4(1994), 639-650

A CONSTRUCTIVE DEFINITION OF DIRICHLET PRIORS

Jayaram Sethuraman

Florida State University

Abstract: In this paper we give a simple and new constructive definition of Dirichlet measures removing the restriction that the basic space should be $\mathcal{R}_k$. We also give complete, self contained proofs of the three basic results for Dirichlet measures:

1. The Dirichlet measure is a probability measure on the space of all probability measures.
2. It gives probability one to the subset of discrete probability measures.
3. The posterior distribution is also a Dirichlet measure.

Key words and phrases: Bayesian nonparametrics, random probability measures, Dirichlet measures.

1. Introduction

Dirichlet measures form a class of distributions of a random probability measure P on a measurable space $(\mathcal{X}, \mathcal{B})$ and are useful in Bayesian nonparametrics. The purpose of this paper is to give a constructive definition of Dirichlet measures for arbitrary measurable spaces $(\mathcal{X}, \mathcal{B})$, and to give a self contained proof showing that it satisfies the main properties **P1**, **P2** and **P3**, which are defined later in this section. This is done in Sections 2, 3 and 4. This definition simplifies the proofs of earlier results of Ferguson (1973) and Blackwell (1973) and is also useful to prove new results. This is illustrated by the examples in Section 5.

The following notations are required to give a rigorous definition of a Dirichlet measure.

Let X be a random variable, representing "data", taking values in a measurable space $(\mathcal{X}, \mathcal{B})$ and let its unknown probability measure be P. The unknown probability distribution P, is the "parameter" in the nonparametric problem, and it takes values in $\mathcal{P}$, the collection of all probability measures on $(\mathcal{X}, \mathcal{B})$. Let $\mathcal{C}$ be the smallest σ-field generated by sets of the form $\{P : P(A) < r\}$ where $A \in \mathcal{B}$ and $r \in [0, 1]$. Let ν be a probability measure on $(\mathcal{P}, \mathcal{C})$. Such a probability measure ν can be used as a prior distribution for P. The Bayesian solution is to compute the posterior distribution, ν^X, of P given X, and use it for decision making.

If $\mathcal{X}$ is a finite set $\{1, 2, \ldots, k\}$, say, then every probability measure P on $\mathcal{X}$

is given by a vector $(p_1 = P(\{1\}), \dots, p_k = P(\{k\}))$ taking values in the simplex $\Delta_k = \{(p_1, \dots, p_k) : 0 \le p_1 \le 1, \dots, 0 \le p_k \le 1, \sum p_j = 1\}$, which is a subset of $\mathcal{R}^k$. It is therefore easy to define probability measures on Δ_k. In particular, we can use the Dirichlet measures on finite dimensional spaces, which are defined in the next paragraph.

Let $(\gamma_1, \gamma_2, \dots, \gamma_k)$ be a vector such that $\gamma_j \ge 0$, $j = 1, 2, \dots, k$, and such that $\sum \gamma_j > 0$. Let $z_{\gamma_j}, j = 1, 2, \dots, k$, be independent Gamma random variables with scale parameter 1 and shape parameters $\gamma_j, j = 1, 2, \dots, k$, respectively. Let $z = \sum z_{\gamma_j}$ and $y_j = (z_{\gamma_j}/z), j = 1, 2, \dots, k$. The joint distribution of the random variable $(y_1, y_2, \dots, y_k)$, taking values in Δ_k, is defined to be k-dimensional Dirichlet measure $\mathcal{D}_{(\gamma_1, \gamma_2, \dots, \gamma_k)}$.

Let $\mathbf{e}_j$ denote the k-dimensional vector consisting of 0's, except for the jth co-ordinate, which is equal to 1. Notice that the Dirichlet measure $\mathcal{D}_{\mathbf{e}_j}$ puts all its probability mass at the point $\mathbf{e}_j$. Furthermore, it is interesting to note that $\mathcal{D}_{2\mathbf{e}_j} = \mathcal{D}_{\mathbf{e}_j}$. This fact will use used later in the proof of Theorem 4.3.

When $\mathcal{X}$ is not a finite space, P takes values in an infinite dimensional space, and hence the definition of a prior distribution for P has always required a more careful description of the attendant measure theoretic problems. Bayesian nonparametrics with general data spaces $\mathcal{X}$, becomes feasible only if one can define a large class of prior distributions ν on $(\mathcal{P}, \mathcal{C})$ and also obtain the corresponding posterior distributions ν^X. The class of Dirichlet measures, which are probability distributions on $(\mathcal{P}, \mathcal{C})$, forms one such family of prior distributions.

The intuitive definition of a Dirichlet measure in the general case is easy to give. Let $\mathcal{M}$ be the class of non-zero finite measures on $(\mathcal{X}, \mathcal{B})$ and let $\alpha \in \mathcal{M}$. A probability distribution ν on $(\mathcal{P}, \mathcal{C})$ is said to be a Dirichlet measure with parameter α if for every measurable partition $\{B_1, B_2, \dots, B_k\}$ of $\mathcal{X}$, the distribution of $(P(B_1), P(B_2), \dots, P(B_k))$ under ν is the finite dimensional Dirichlet distribution $\mathcal{D}_{(\alpha(B_1), \alpha(B_2), \dots, \alpha(B_k))}$. When such a probability measure ν on $(\mathcal{P}, \mathcal{C})$ can be demonstrated to exist, it will be denoted by $\mathcal{D}_\alpha$.

There are three main properties of Dirichlet measures that make them useful in Bayesian nonparametrics. Apart from their marginals having finite dimensional Dirichlet distributions, they possess the following three properties:

P1 $\mathcal{D}_\alpha$ is a probability measure on $(\mathcal{P}, \mathcal{C})$,

P2 $\mathcal{D}_\alpha$ gives probability one to the subset of all discrete probability measures on $(\mathcal{X}, \mathcal{B})$, and

P3 the posterior distribution $\mathcal{D}_\alpha^X$ is the Dirichlet measure $\mathcal{D}_{\alpha+\delta_X}$ where δ_X is the probability measure degenerate at X.

We now give a brief review of earlier work on the definition of a Dirichlet measure and indicate some of the difficulties. Dirichlet measures were introduced in Ferguson (1973) and in Blackwell and McQueen (1973). We will now describe the work in these two papers.

It is easy to see that the distributions of $(P(B_1), P(B_2), \ldots, P(B_k))$ give rise to a consistent family of measures over the class of all partitions $(B_1, B_2, \ldots, B_k)$. Ferguson (1973) argued on the basis of the Kolmogorov consistency theorem, that this fact leads to a unique probability measure on $[0,1]^{\mathcal{B}}$ with its associated Kolmogorov σ-field. Furthermore, for any given sequence of disjoint measurable sets $B_1, B_2, \ldots$, the probability is one that

$$P(\cup B_j) = \sum P(B_j), \tag{1.1}$$

where $P(\cdot)$ is the canonical representation of a point in $[0,1]^{\mathcal{B}}$. This set of probability one may depend on the sequence $B_1, B_2, \ldots$. Such a P is a member of $\mathcal{P}$ if and only if (1.1) were true for all disjoint sequences $B_1, B_2, \ldots$. The collection of such disjoint sequences is uncountable. This presents a problem in making this definition rigorous and establishing property **P1**. For the special case where $\mathcal{X}$ is the real line, or more generally a separable complete metric space, one can use a result of Harris ((1968), Lemma 6.1). This result states that a verification of (1.1) for a select countable number of cases of disjoint sequences of sets is sufficient to ensure that (1.1) holds for all disjoint countable sets and that the set function P is a probability measure. An appeal to this result is one way to show that there is a probability measure on $(\mathcal{P}, \mathcal{C})$ with the required properties and this defines the Dirichlet measure $\mathcal{D}_\alpha$.

In a later section, Ferguson ((1973), Section 4) gives an alternative constructive definition of the Dirichlet measure which shows that it gives probability one to the subset of discrete probability measures. However, it takes some effort to see that the two definitions are equivalent.

Ferguson (1973) also establishes the posterior distribution property **P3** by using a very peculiar definition (see his Definition 2) for the joint distribution of (P, X), rather than the straightforward definition that the distribution of X given P is P.

Blackwell and McQueen (1973) appeal to the famous theorem of de Finetti to show that there is a one-to-one correspondence between sequences of exchangeable random variables and probability measures on $(\mathcal{P}, \mathcal{C})$. A particular case of exchangeable random variables, namely the generalized Pòlya urn scheme, corresponds to the Dirichlet measure. In this paper and in Blackwell (1973), they establish the three properties **P1**, **P2** and **P3**. Their proof is elegant but quite indirect and also requires the space $\mathcal{X}$ to be a separable complete metric space.

Freedman (1963) and Fabius (1964) contain early work on tail-free priors, which include Dirichlet priors, for the case when $\mathcal{X}$ is the set of integers or $[0,1]$.

We briefly summarize the rest of the paper. Let $(\mathcal{X}, \mathcal{B})$ be an *arbitrary* measurable space. Let $\mathcal{E}$ be the usual Borel σ-field restricted to $[0,1]$. In Section 2, we define a function P based on a sequence of i.i.d. random variables $(\theta_n, Y_n), n = 1, 2, \ldots$, taking values in $([0,1] \times \mathcal{X}, \mathcal{E} \times \mathcal{B})$. See (2.1). By its very definition, P is a random measure taking values in $(\mathcal{P}, \mathcal{S})$ and giving probability one to the subset of discrete probability measures on $(\mathcal{X}, \mathcal{B})$. This establishes properties **P1** and **P2**. We give a direct proof, in Theorem 3.4 of Section 3, that the finite dimensional marginal distributions of P are Dirichlet distributions. This establishes that the distribution of P is a Dirichlet measure. In Theorem 4.3 of Section 4 we prove property **P3** thus establishing that the posterior distribution is also a Dirichlet measure. The definition and proofs are all given in some detail to make this paper self contained.

This constructive definition of a Dirichlet measure was presented at an invited paper of an IMS meeting in 1980 and also announced in Sethuraman and Tiwari (1982) which dealt with the convergence of Dirichlet measures. This definition has since been used by several authors to simplify previous calculations and to obtain new results involving Dirichlet measures. For instance see Doss (1991), Ferguson (1983), Ferguson, Phadia and Tiwari (1992), Kumar and Tiwari (1989).

2. Constructive Definition of the Dirichlet Measure

Let α be a non-zero finite measure on $(\mathcal{X}, \mathcal{B})$. Let $\beta(B) = \alpha(B)/\alpha(\mathcal{X})$ be the normalized probability measure arising from α. Let $B(\gamma, \delta)$ stand for the Beta distribution on $[0,1]$ with parameters γ and δ. This Beta distribution is the marginal distribution of the first co-ordinate of the Dirichlet measure $\mathcal{D}(\gamma, \delta)$ on the two-dimensional simplex Δ_2 defined earlier. Let $\mathcal{N} = \{1, 2, \ldots\}$ be the set of positive integers and let $\mathcal{F}$ be the σ-field of all subsets of $\mathcal{N}$. Let $\{\Omega, \mathcal{S}, Q\}$ be a probability space supporting a collection of random variables $(\boldsymbol{\theta}, \mathbf{Y}, I) = ((\theta_j, Y_j), j = 1, 2, \ldots, I)$ taking values in $(([0,1] \times \mathcal{X})^\infty \times \mathcal{N}, (\mathcal{E} \times \mathcal{B})^\infty \times \mathcal{F})$, with a joint distribution defined as follows. The random variables $(\theta_1, \theta_2, \ldots)$ are independently and identically distributed (i.i.d.) with a common Beta distribution $B(1, \alpha(\mathcal{X}))$. The random variables $(Y_1, Y_2, \ldots)$ are independent of the $(\theta_1, \theta_2, \ldots)$ and i.i.d. among themselves with common distribution β. Let $p_1 = \theta_1$ and let $p_n = \theta_n \prod_{1 \le m \le n-1}(1 - \theta_m)$ for $n = 2, 3, \ldots$. Notice that $\sum_{1 \le m \le n} p_m = 1 - \prod_{1 \le m \le n}(1 - \theta_m) \to 1$ with Q-probability one. Let $Q(I = n | (\boldsymbol{\theta}, \mathbf{Y})) = p_n, n = 1, 2, \ldots$. The existence of a probability space $(\Omega, \mathcal{S}, Q)$ and such a sequence of random variables $(\boldsymbol{\theta}, \mathbf{Y}, I)$ follows from the usual construction of a product measure, and does not require any restrictions on $(\mathcal{X}, \mathcal{B})$, such as its being a separable complete metric space.

Define

$$P(\boldsymbol{\theta}, \mathbf{Y}; B) = P(B) = \sum_{n=1}^{\infty} p_n \delta_{Y_n}(B), \tag{2.1}$$

where $\delta_x(\cdot)$ stands for the probability measure degenerate at x.

This is the new constructive definition of a Dirichlet measure. As convenience dictates, we drop all or part of the arguments $(\boldsymbol{\theta}, \mathbf{Y}, B)$ and denote the random measure in (2.1) by P, for simplicity of notation. Since P is clearly a measurable map from $(\Omega, \mathcal{S})$ into $(\mathcal{P}, \mathcal{C})$ and takes values in the subset of discrete probability measures, properties **P1** and **P2** are self evident.

Notice that the random variable I introduced above has not been used in the definition of P. It will be used later, in Section 4, to prove the posterior distribution property **P3**.

A more direct way to describe the constructive definition in (2.1) is as follows. Let $Y_1, Y_2, \ldots$ be i.i.d. with common distribution β. Let $\{p_1, p_2, \ldots\}$ be the probabilities from a discrete distribution on the integers with discrete failure rate $\{\theta_1, \theta_2, \ldots\}$ which are i.i.d. with a Beta distribution $B(1, \alpha(\mathcal{X}))$. Let P be the random probability measure that puts weights p_n at the degenerate measures δ_{Y_n}, $n = 1, 2, \ldots$. This is the random probability measure P described in (2.1). The alternative definition given in Ferguson ((1973), Section 4) uses a different set of random weights which are arranged in decreasing order. The use of unordered weights in this paper simplifies all our calculations. It is interesting to note that the weights used by Ferguson (1973) are equivalent to our weights rearranged in decreasing order. However, it is not clear that there is an easy way to unorder the weights of Ferguson (1973) to obtain weights with the simple structure of (2.1).

3. The Distribution of the Random Measure P is $\mathcal{D}_\alpha$

We will digress a little before establishing that the distribution of P is the Dirichlet measure $\mathcal{D}_\alpha$.

Let $\theta_n^* = \theta_{n+1}, Y_n^* = Y_{n+1}, n = 1, 2, \ldots$, and let $J = I - 1$. Using the definition $(\boldsymbol{\theta}^*, \mathbf{Y}^*, J) = ((\theta_1^*, \theta_2^*, \ldots), (Y_1^*, Y_2^*, \ldots), J)$, we see that the random probability measure P in (2.1) satisfies

$$P(\boldsymbol{\theta}, \mathbf{Y}; B) = \theta_1 \delta_{Y_1}(B) + (1 - \theta_1) P(\boldsymbol{\theta}^*, \mathbf{Y}^*; B). \tag{3.1}$$

Notice that $(\boldsymbol{\theta}^*, \mathbf{Y}^*)$ has the same distribution as $(\boldsymbol{\theta}, \mathbf{Y})$ and is independent of (θ_1, Y_1). Thus we can re-write (3.1) as the following distributional equation for P:

$$P \stackrel{\text{st}}{=} \theta_1 \delta_{Y_1} + (1 - \theta_1) P, \tag{3.2}$$

where on the right hand side P is independent of (θ_1, Y_1).

Theorem 3.4 below uses the distributional equation (3.2) to show that the distribution of P is the Dirichlet measure $\mathcal{D}_\alpha$. The proof of this theorem uses well known facts about finite dimensional Dirichlet measures and a result on the uniqueness of solutions to distributional equations, which are given below as Lemmas 3.1, 3.2 and 3.3.

Lemma 3.1. *Let $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_k)$ and $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_k)$ be k-dimensional vectors. Let U, V be independent k-dimensional random vectors with Dirichlet distributions $\mathcal{D}_{\boldsymbol{\gamma}}$ and $\mathcal{D}_{\boldsymbol{\delta}}$, respectively. Let W be independent of (U, V) and have a Beta distribution $B(\gamma, \delta)$, where $\gamma = \sum \gamma_j$ and $\delta = \sum \delta_j$. Then the distribution of $WU + (1 - W)V$ is the Dirichlet distribution $\mathcal{D}_{\boldsymbol{\gamma}+\boldsymbol{\delta}}$.*

Lemma 3.2. *Let $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)$, $\gamma = \sum \gamma_j$ and let $\beta_j = \gamma_j / \gamma, j = 1, 2, \ldots, k$. Then*

$$\sum \beta_j \mathcal{D}_{\boldsymbol{\gamma}+\mathbf{e}_j} = \mathcal{D}_{\boldsymbol{\gamma}}.$$

This conclusion can also be written as $E(\mathcal{D}_{\boldsymbol{\gamma}+\mathbf{Z}}) = \mathcal{D}_{\boldsymbol{\gamma}}$, where $\mathbf{Z}$ is a random vector that takes the values $\mathbf{e}_j$ with probability $\gamma_j/\gamma, j = 1, \ldots, k$.

The proofs of these two lemmas are found in many standard text books, for instance in Wilks ((1962), Section 7).

Lemma 3.3 stated and proved below shows that certain distributional equations have unique solutions. Such results appear in several areas of statistics, notably in renewal theory. For a recent work which gives more general results see Goldie (1991). The following lemma is sufficient for our purposes. Its proof, which is not new, is given here to make this paper self contained.

Lemma 3.3. *Let W, U be a pair of random variables where W takes values in $[-1, 1]$ and U takes values in a linear space. Suppose that V is a random variable taking values in the same linear space as U and which is independent of (W, U) and satisfies the distributional equation*

$$V \stackrel{\text{st}}{=} U + WV. \tag{3.3}$$

Suppose that $P(|W| = 1) \neq 1$. Then there is only one distribution for V that satisfies (3.3).

Proof. Let V and V' be two random variables whose distributions are not equal but satisfy Equation (3.3). Let (W_n, U_n) be independent copies of (W, U) which are independent of V, V'. Let $V_1 = V$, $V_1' = V'$ and define, recursively, $V_{n+1} = U_n + W_n V_n$ and $V_{n+1}' = U_n + W_n V_n'$ for $n = 1, 2, \ldots$. From the distributional equation (3.3), the V_n's have the same distribution as V and the V_n''s have the same distribution as V'. However,

$$|V_{n+1} - V_{n+1}'| = |W_n|\, |V_n - V_n'| = \prod_{1 \le m \le n} |W_m|\, |V_1 - V_1'| \to 0$$

with probability 1, since the W_n's are i.i.d., $P(|W_1| \leq 1) = 1$, and $P(|W_1| = 1) < 1$. This contradicts the supposition that the distributions of V and V' are unequal and proves that if the distribution of V satisfies (3.3), then it is unique.

Theorem 3.4. *Let $\{B_1, B_2, \ldots, B_k\}$ be a measurable partition of $\mathcal{X}$ and let $\mathbf{P} = (P(B_1), P(B_2), \ldots, P(B_k))$. Then the distribution of $\mathbf{P}$ is the k-dimensional Dirichlet measure $\mathcal{D}_{(\alpha(B_1),\alpha(B_2),\ldots,\alpha(B_k))}$.*

Proof. Let $\mathbf{D} = (\delta_{Y_1}(B_1), \delta_{Y_1}(B_2), \ldots, \delta_{Y_1}(B_k))$. Notice that $P(\mathbf{D} = \mathbf{e}_j) = P(Y_1 \in B_j) = \beta(B_j), j = 1, 2, \ldots, k$. From (3.2) we see that $\mathbf{P}$ satisfies the distributional equation

$$\mathbf{P} \stackrel{\text{st}}{=} \theta_1 \mathbf{D} + (1 - \theta_1)\mathbf{P}, \tag{3.4}$$

where, on the right, θ_1 has a Beta distribution $B(1, \alpha(\mathcal{X}))$, $\mathbf{D}$ is independent of θ_1 and takes the value $\mathbf{e}_j$ with probability $\beta(B_j), j = 1, 2, \ldots, k$, and the k-dimensional random vector $\mathbf{P}$ is independent of $(\theta_1, \mathbf{D})$.

We first verify that the k-dimensional Dirichlet measure for $\mathbf{P}$ satisfies the distributional equation (3.4) and then show that this solution is the unique solution.

Let the distribution of $\mathbf{P}$ on the right of (3.4) be the k-dimensional Dirichlet measure $\mathcal{D}_{(\alpha(B_1),\alpha(B_2),\ldots,\alpha(B_k))}$. The k-dimensional Dirichlet measure $\mathcal{D}_{\mathbf{e}_j}$ gives probability 1 to $\mathbf{e}_j$. Given that $\mathbf{D} = \mathbf{e}_j$, the distribution of $\theta_1 \mathbf{D} + (1 - \theta_1)\mathbf{P}$ is the distribution of $\theta_1 \mathcal{D}_{\mathbf{e}_j} + (1 - \theta_1)\mathcal{D}_{(\alpha(B_1),\alpha(B_2),\ldots,\alpha(B_k))}$ and this, by Lemma 3.1, is $\mathcal{D}_{(\alpha(B_1),\alpha(B_2),\ldots,\alpha(B_k))+\mathbf{e}_j}$. Summing over the distribution of $\mathbf{D}$ is equivalent to taking a mixture of these Dirichlet measures with weights $\beta(B_j) = \alpha(B_j)/\alpha(\mathcal{X})$, which by Lemma 3.2, is equal to $\mathcal{D}_{(\alpha(B_1),\alpha(B_2),\ldots,\alpha(B_k))}$. This verifies that the k-dimensional Dirichlet measure satisfies the distributional equation (3.4). Lemma 3.3 shows that this solution is unique. This completes the proof of Theorem 3.4.

4. The Posterior Distribution of P is $\mathcal{D}_{\alpha+\delta_X}$

Let $X = Y_I$. Then X is a random variable from $(\Omega, \mathcal{S})$ into $\mathcal{X}$ defined explicitly as a function of $(\boldsymbol{\theta}, \mathbf{Y}, I)$. The next lemma shows that the distribution of X given P is P and hence the joint distribution of (P, X) is that of the "parameter" and "data" in a Bayesian nonparametric problem.

Lemma 4.1. *The distribution of X given P is P.*

Proof. Let $B \in \mathcal{B}$. By direct calculation, we get

$$\begin{aligned} Q(X \in B|(\boldsymbol{\theta}, \mathbf{Y})) &= \sum_n Q(X \in B|I = n, (\boldsymbol{\theta}, \mathbf{Y}))Q(I = n|(\boldsymbol{\theta}, \mathbf{Y})) \\ &= \sum_n \delta_{Y_n}(B) p_n = P(B). \end{aligned}$$

Since this conditional probability is a function of P, it immediately follows that $Q(\cdot|P)$ exists as a regular conditional probability and $Q(X \in B|P) = P(B)$ with Q-probability 1.

We now come to the posterior distribution of P, i.e. the distribution of P given X. We do this by separately obtaining the conditional distribution of $(\boldsymbol{\theta}, \mathbf{Y})$ given $I = 1$ and given $I > 1$. When f and g are functions of $(\boldsymbol{\theta}, \mathbf{Y}, I)$, we will use the notations $\mathcal{L}(f)$ and $\mathcal{L}(f|g)$ to denote the distribution of f and the conditional distribution of f given g, under Q, respectively.

Lemma 4.2. *The following are the conditional distributions of* $(\boldsymbol{\theta}, \mathbf{Y}, I)$ *given* $I = 1$ *and given* $I > 1$*:*

$$\mathcal{L}((\theta_1, Y_1), (\boldsymbol{\theta}^*, \mathbf{Y}^*)|I = 1) = B(2, \alpha(\mathcal{X})) \times \mathcal{L}(\boldsymbol{\theta}, \mathbf{Y}) \tag{4.1}$$

and

$$\mathcal{L}((\theta_1, Y_1), (\boldsymbol{\theta}^*, \mathbf{Y}^*, J)|I > 1) = B(1, \alpha(\mathcal{X}) + 1) \times \mathcal{L}(\boldsymbol{\theta}, \mathbf{Y}, I). \tag{4.2}$$

Proof. Note that $Q(I = 1|(\boldsymbol{\theta}, \mathbf{Y})) = \theta_1$. Thus, if $A_i \in \mathcal{E}, B_i \in \mathcal{B}, i = 1, 2, \ldots, n$, we have the relation

$$\begin{aligned} &Q\{\theta_i \in A_i, Y_i \in B_i, i = 1, 2, \ldots, n, I = 1\} \\ &\quad \propto \int I(x_i \in A_i, y_i \in B_i, i = 1, 2, \ldots, n)\, x_1 \prod_{1 \le i \le n} [(1 - x_i)^{\alpha(\mathcal{X})-1} dx_i \beta(dy_i)]. \end{aligned}$$

This implies that, conditional on $I = 1$, θ_1 has distribution $B(2, \alpha(\mathcal{X}))$, the distributions of $\theta_i, i = 2, 3, \ldots, n$, and $Y_i, i = 1, 2, \ldots, n$, are all unchanged, and all these are independent. This gives all the finite dimensional conditional distributions and proves (4.1). The proof of (4.2) follows along the same lines since $Q(I > 1|(\boldsymbol{\theta}, \mathbf{Y})) = 1 - \theta_1$.

Theorem 4.3. *The posterior distribution of* P *given* X *is the Dirichlet measure* $\mathcal{D}_{\alpha+\delta_X}$.

Proof. Let $P^* = P(\boldsymbol{\theta}^*, \mathbf{Y}^*)$. We can rewrite (3.1) as $P = \theta_1 \delta_{Y_1} + (1 - \theta_1)P^*$. When $I = 1$, we use (4.1) and obtain

$$\begin{aligned} \mathcal{L}(P|X, I = 1) &= \mathcal{L}(\theta_1 \delta_{Y_1} + (1 - \theta_1)P^*|X, I = 1) \\ &\stackrel{\text{st}}{=} \theta_1' \delta_X + (1 - \theta_1')P^{**}, \end{aligned} \tag{4.3}$$

where θ_1' has distribution $B(2, \alpha(\mathcal{X}))$, and P^{**} is a random probability measure, independent of θ_1', whose distribution is the Dirichlet measure $\mathcal{D}_\alpha$. The random probability measure putting all its mass on the degenerate measure δ_X is the Dirichlet measure $\mathcal{D}_{\delta_X}$ which is also equal to $\mathcal{D}_{2\delta_X}$. Since θ_1' has a Beta distribution $B(2, \alpha(\mathcal{X}))$, this latter choice allows us to use Lemma 3.1 to obtain

$$\mathcal{L}(P|X, I=1) \overset{\text{st}}{=} \mathcal{D}_{\alpha+2\delta_X}. \tag{4.4}$$

When $I > 1$, we use (4.2) and first obtain

$$\mathcal{L}(\boldsymbol{\theta}^*, \mathbf{Y}^*, X|I>1) = \mathcal{L}(\boldsymbol{\theta}, \mathbf{Y}, X) \tag{4.5}$$

since $X = Y_I = Y_J^*$ on $I > 1$. Thus

$$\begin{aligned}\mathcal{L}(P|X, I>1) &= \mathcal{L}(\theta_1\delta_{Y_1} + (1-\theta_1)P^*|X, I>1)\\ &\overset{\text{st}}{=} \theta_1''\delta_{Y_1} + (1-\theta_1'')P^{***},\end{aligned} \tag{4.6}$$

where Y_1 has distribution β, θ_1'' is independent of Y_1 and has distribution $B(1, \alpha(\mathcal{X})+1)$, and P^{***} is a random probability measure, independent of (Y_1, θ_1''), whose distribution is $\mathcal{L}(P|X)$, in view of (4.5). We can combine (4.3) and (4.6) to obtain a distributional equation for $\mathcal{L}(P|X)$ as follows.

$$\mathcal{L}(P|X) \overset{\text{st}}{=} A(\theta_1'\delta_X + (1-\theta_1')P^{**}) + (1-A)(\theta_1''\delta_{Y_1} + (1-\theta_1'')P^{***}), \tag{4.7}$$

where all the random variables on the right are independent and have the distributions previously specified, and the random variable A takes values 1 and 0 with probabilities $\frac{1}{(\alpha(\mathcal{X})+1)}$ and $\frac{\alpha(\mathcal{X})}{(\alpha(\mathcal{X})+1)}$, respectively. Notice that the distribution of P^{***} is $\mathcal{L}(P|X)$ which makes (4.7) a distributional equation.

From Lemma 3.3 we conclude that if there is a solution to (4.7), it will be a unique solution. We now verify that $\mathcal{L}(P|X) = \mathcal{D}_{\alpha+\delta_X}$ satisfies the distributional equation (4.7). Relation (4.4) can be rewritten as

$$\theta_1'\delta_X + (1-\theta_1')P^{**} \overset{\text{st}}{=} \mathcal{D}_{\alpha+2\delta_X}. \tag{4.8}$$

By conditioning on Y_1 and using Lemma 3.1, and then taking expectations with respect to Y_1, we find that

$$\theta_1''\delta_{Y_1} + (1-\theta_1'')P^{***} \overset{\text{st}}{=} E(\mathcal{D}_{\alpha+\delta_X+\delta_{Y_1}}), \tag{4.9}$$

where Y_1 has distribution β. Let Z be a random variable in $(\mathcal{X}, \mathcal{B})$ with distribution $\frac{1}{(\alpha(\mathcal{X})+1)}\delta_X + \frac{\alpha(\mathcal{X})}{(\alpha(\mathcal{X})+1)}\beta = \frac{\alpha+\delta_X}{(\alpha(\mathcal{X})+1)}$. Combining (4.8) and (4.9), and using Lemma 3.2 on mixtures of Dirichlet measures, we conclude the distribution of the random measure in the right hand side of (4.7) is equal to

$$\begin{aligned}\frac{1}{(\alpha(\mathcal{X})+1)}\mathcal{D}_{\alpha+2\delta_X} + \frac{\alpha(\mathcal{X})}{(\alpha(\mathcal{X})+1)}E(\mathcal{D}_{\alpha+\delta_X+\delta_{Y_1}}) &\overset{\text{st}}{=} E(\mathcal{D}_{\alpha+\delta_X+\delta_Z})\\ &\overset{\text{st}}{=} \mathcal{D}_{\alpha+\delta_X}.\end{aligned}$$

This proves that $\mathcal{D}_{\alpha+\delta_X}$ is the posterior distribution of P given X.

Suppose that the random variables $X_1, X_2, \ldots, X_n$ given P are i.i.d. with common distribution P, where P is distributed according to the Dirichlet measure $\mathcal{D}_\alpha$. This is the situation when a sample of size n is drawn. The posterior distribution of P given $(X_1, \ldots, X_n)$ follows from Theorem 4.3 and the following general fact described in the next paragraph.

Let the random variables $X_1, X_2, \ldots, X_n$ given P be i.i.d. with common distribution P, where P has an arbitrary prior distribution ν on $(\mathcal{P}, \mathcal{C})$. Let ν^{X_1} be the posterior distribution of P given X_1. Then the joint distribution of $(X_2, X_3, \ldots, X_n)$ given X_1 is the same as the joint distribution of the random variables $(Y_2, Y_3, \ldots, Y_n)$ which can be described after introducing a new random probability measure P' as follows: given P', $Y_2, Y_3, \ldots, Y_n$ are i.i.d. with common distribution P', and P' has distribution ν^{X_1}.

If the prior distribution of P is $\mathcal{D}_\alpha$, then from this remark and Theorem 4.3, it follows that the posterior distribution of P given $(X_1, X_2, \ldots, X_n)$ is $\mathcal{D}_{\alpha+\Sigma\delta_{X_i}}$.

5. Examples

The first application of our construction of Dirichlet measures was made in Sethuraman and Tiwari (1982) wherein the weak convergence of a sequence of Dirichlet measures was established. There is no other method available to prove such weak convergence. See that paper for the details.

More conventional applications have appeared in several papers, wherein old and new results have been proved using our construction of Dirichlet measure. See, for instance, Ferguson (1983), Ferguson, Phadia and Tiwari (1992), Kumar and Tiwari (1989).

We now give another application wherein one needs to simulate a random measure P with a Dirichlet measure and a random variable X with distribution P. It is impossible to generate such a random measure P, even with our construction, since it will mean generating an infinite number of random variables. However, if our interest is in X, then this can be accomplished because our construction of a Dirichlet measure shows that one needs to know only a finite number of the infinite number of the variables in the definition of P. This has been done in a recent paper of Doss (1991). We now sketch some of the details.

Let α be a non-zero measure in $\mathcal{M}$ and $\beta(B) = \alpha(B)/\alpha(\mathcal{X})$ be the normalized measure of α. Suppose that the prior distribution of P is the Dirichlet measure $\mathcal{D}_\alpha$. Let the distribution of X given P be P. This is the standard setup that we have discussed up to now. Fix a non-empty set A in $\mathcal{B}$, which is not a singleton set. Let $Y = I(x \in A)$, where $I(\cdot)$ is the indicator function. Suppose that it is observed that $X \in A$, that is $Y = 1$, but the value of X is not observed. This happens in the standard models of censoring. What is $\mathcal{L}(P|Y=1)$, the posterior distribution of P given $Y=1$? This and other generalizations are addressed in Doss (1991). For any probability measure Q on $(\mathcal{X},\mathcal{B})$, let Q_A be the version of Q truncated to A, defined by

$$Q_A(B)=Q(B\cap A)/Q(A).$$

The following two conditional distributions are straightforward: $\mathcal{L}(P|X,Y=1)=\mathcal{D}_{\alpha+\delta_X}$, and $\mathcal{L}(X|P,Y=1)=P_A$.

This suggests that one can use the Markov chain successive substitution method described in Gelfand and Smith (1990) to simulate an observation from $\mathcal{L}(P|Y=1)$, as follows. Let (P_0,X_0) be arbitrary point in $\mathcal{P}\times\mathcal{X}$. For $n=1,2,\ldots$, let P_n have distribution $\mathcal{L}(P|X_{n-1},Y=1)$ and X_n have distribution $\mathcal{L}(X|P_n,Y=1)$. Then from the results in Athreya, Doss and Sethuraman (1992) as shown in Doss (1991), it follows that (P_n,X_n) and also averages based on $(P_0,X_0),\ldots,(P_n,X_n)$, converge in distribution to $\mathcal{L}((P,X)|Y=1)$, as $n\to\infty$. By retaining P alone we obtain an approximation to $\mathcal{L}(P|Y=1)$.

This simulation requires generating observations from both distributions $\mathcal{L}(P|X,Y=1)$ and $\mathcal{L}(X|P,Y=1)$, but we see below that we can generate from the latter distribution and bypass the former distribution. In other words, we will show below that, we can generate X_n without the full knowledge of P_n, which will be difficult to generate since it has a Dirichlet distribution. From our constructive definition, P_n is of the form $\sum_{j=1}^{\infty}p_j\delta_{Y_j}$, where p_j and Y_j are random variables depending on the parameter of the Dirichlet distribution of P_n. Also X_n is just Y_J, conditioned to lie in A by the simple rejection method, where J is an integer valued random variable taking the value j with probability p_j. This random index J can be generated on the basis of a uniform random variable U by putting $J=\min\{j:\sum_{1\le r\le j}p_r\ge U\}$, which requires the evaluation of only a finite number of p_r's. If $Y_J\in A$, we let $X_n=Y_J$, and this means that we need to generate only $Y_1,\ldots,Y_J$. If $Y_J\notin A$, we repeat the procedure by using another uniform random variable U to generate a J until $Y_J\in A$. Thus one can ignore the problem of generating P_n and go straight to generating X_n. For some large n, we declare that $\mathcal{L}(P|Y=1)$ is approximated by $\mathcal{L}(P_{n+1}|X_n,Y=1)$ which is $\mathcal{D}_{\alpha+\delta_{X_n}}$, and compute approximations to functionals of $\mathcal{L}(P|Y=1)$. This example is an illustration of the power of our constructive definition. More details and other more useful generalizations are given in Doss (1991).

Acknowledgement

Professor David Blackwell was visiting Florida State University in Spring 1979 when I was giving a seminar on Bayesian nonparametrics. It was then that I discovered this new construction of a Dirichlet measure. I wish to thank Professor Blackwell for his inspiration and encouragement.

This research was supported by U. S. Army Research Office Grant DAAL03-90-G-0103 and DAAH04-93-G-0201.

References

Athreya, K. B., Doss, H. and Sethuraman, J. (1992). A proof of convergence of the Markov chain simulation method. Florida State University Technical Report.

Blackwell, D. (1973). Discreteness of Ferguson selections. *Ann. Statist.* **1**, 356-358.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Pòlya schemes. *Ann. Statist.* **1**, 353-355.

Doss, H. (1991). Bayesian nonparametric estimation for incomplete data via successive substitution sampling. Florida State University Technical Report No. 850.

Fabius, J. (1964). Asymptotic behavior of Bayes' estimates. *Ann. Math. Statist.* **35**, 846-856.

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

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on his Sixtieth Birthday* (Edited by M. H. Rizvi, J. Rustagi and D. Siegmund), 287-302, Academic Press.

Ferguson, T. S., Phadia, E. G. and Tiwari, R. C. (1992). Bayesian nonparametric inference. *Current Issues in Statistical Inference; Essays in Honor of D. Basu* (Edited by M. Ghosh and P. K. Pathak). IMS Lecture Notes - Monograph Series, **17**, 127-150.

Freedman, D. A. (1963). Invariants under mixing which generalize de Finetti's theorem: Continuous time parameter. *Ann. Math. Statist.* **34**, 1194-1216.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398-409.

Goldie, C. M. (1991). Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1**, 126-166.

Harris, T. E. (1968). Counting measures, monotone random set functions. *Z. Wahrsch. verw. Gebiete* **10**, 102-119.

Kumar, S. and Tiwari, R. C. (1989). Bayes estimation of reliability under a random environment governed by a Dirichlet prior. *IEEE Trans. Reliab.* **38**, 218-223.

Sethuraman, J. and Tiwari, R. C. (1982). Convergence of Dirichlet measures and the interpretation of their parameter. *Statistical Decision Theory and Related Topics III* **2**, 305-315.

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

Department of Statistics, Florida State University, Tallahassee, FL 32306, U.S.A.

(Received November 1992; accepted February 1994)