A NOTE ON DIRICHLET PROCESS

by

D. Basu¹ and R. C. Tiwari²

February, 1980
FSU Statistics Report M536

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

¹Research partially supported by NSF Grant No. 79-04693.

²Research supported by Government of India, Ministry of Education.
A Note on Dirichlet Process

by

D. Basu and R. C. Tiwari
(The Florida State University)

ABSTRACT

This expository article is concerned primarily with the question of existence of a Dirichlet process. A random probability measure on a measurable space \((X, A)\) is a stochastic process \(\{P(A): A \in A\}\) — a collection of random variables indexed by the measurable sets in \(X\) — such that almost every realization of the process is a probability measure on \((X, A)\). Given a finite measure \(\alpha\) on \((X, A)\), a Dirichlet process \(D^\alpha\) is a random probability measure on \((X, A)\) such that, for every partition \((A_1, A_2, \ldots, A_k)\) of \(X\) into a finite number of measurable sets, the joint distribution of the random variables \((P(A_1), P(A_2), \ldots, P(A_k))\) is singular Dirichlet with parameters \((\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_k))\).

Part one of this article deals with the familiar case when \(X\) is a finite set. Properties of the \(k\)-dimensional Dirichlet distribution are so expositied as to motivate Blackwell's (1973) constructive definition of the Dirichlet process. In part two, the case where \((X, A)\) is a Borel space is discussed in some details.
Introduction

This report is concerned primarily with the question of existence of a Dirichlet process on a Borel space \((X, A)\). Part one of this report deals with the familiar case when \(X\) is a finite set. Properties of the \(k\)-dimensional Dirichlet distribution are so exposited as to motivate Blackwell's (1973) constructive definition of the Dirichlet process. In part two, the case where \((X, A)\) is a Borel space is discussed in some details.

Section 2 deals with the Bayesian (parametric) inference and the family of natural conjugate priors. Section 3 is devoted to some characterizations of the Dirichlet distribution and elucidations of its useful properties. In Sections 4 and 5, the results of Section 3 are extended and it is shown that there exists a Dirichlet process on \((X, A)\), when \(X\) is a finite or a countable infinite set.

In Section 6, some preliminary material on the Dirichlet process is presented. Some useful results on a Borel space are given in Section 7. Blackwell's (1973) construction of a Dirichlet process on a Borel space \((X, A)\) is discussed in Section 8. Section 9 is devoted to the study of some properties of a Dirichlet process.
PART ONE: The Dirichlet distribution

2. Bayesian Inference and Natural Conjugate Priors.

Let \( X \) be an observable random variable (r.v.) with a statistical model that is characterized by a probability density function (p.d.f.) \( f(\cdot | \theta) \), where \( \theta \in \Theta \) is the unknown parameter of interest. Before any data are collected, a Bayesian represents his prior opinion about \( \theta \) by a distribution on \( \Theta \), called prior distribution. If he observes \( X \) \( n \)-times and denotes by \( D_1^n = (x_1, x_2, \ldots, x_n) \) the data so obtained, then his opinion about \( \theta \) is represented by the distribution of \( \theta \) given \( D_1^n \), called posterior distribution.

Let \( q(\cdot) \) be the prior p.d.f. of \( \theta \). The posterior p.d.f. of \( \theta \) given \( D_1^n \), namely \( q(\cdot | D_1^n) \), is expressed by the relation

\[
q(\theta | D_1^n) \propto q(\theta) L(\theta | D_1^n).
\]

Here, \( L(\theta | D_1^n) \) is the likelihood function of \( \theta \) at point \( D_1^n \), and the proportionality symbol \( \propto \) is used to indicate that the posterior p.d.f. of \( \theta \) given \( D_1^n \) is equal to the right side of (2.1) divided by the factor \( \int_{\Theta} L(\theta | D_1^n)q(\theta) d\theta \), which does not involve \( \theta \).

This is how new knowledge, obtained through data, may be combined with prior knowledge. The Bayesian continually updates his knowledge as more observations are taken. Clearly,

\[
q(\theta | D_1^{n+m}) \propto q(\theta) L(\theta | D_1^{n+m})
\]

\[
\propto q(\theta) L(\theta | D_1^n) L(\theta | D_{n+1}^{n+m})
\]

\[
\propto q(\theta | D_1^n) L(\theta | D_{n+1}^{n+m}).
\]
Thus the opinion $q(\theta | D_1^{n+m})$ based on data $D_1^{n+m}$ may be regarded as the posterior based on data $D_1^{n+1}$ and prior $q(\theta | D_1^n)$. This process of updating opinion may go through many stages.

It is clear from the relation (2.1) that the change of opinion about $\theta$ after the data are obtained is effected through the likelihood function. In context of a chosen statistical model, a Bayesian will regard the likelihood function as the sole reservoir of all the relevant information about the parameter that is contained in the data. This is usually stated as:

**The Likelihood Principle:** Two sets of data generating equivalent likelihood functions contain the same relevant information about the parameter.

Two likelihood functions are said to be equivalent if one of them is a constant multiple of the other, where the constant may depend on the data. (See Basu (1975) for more on the likelihood principle.)

In many situations, it is convenient to access the prior within a family $\mathcal{C}$ of distributions. The class $\mathcal{C}$ should be large enough to accommodate various shades of opinion about the parameter. Further, if $q(\cdot) \in \mathcal{C}$ is a prior p.d.f. of $\theta$, then the posterior p.d.f. $q(\theta | D_1^n)$ of $\theta$ given the data $D_1^n$ ought to be in a simple computable form. If $q(\theta | D_1^n) \in \mathcal{C}$ for all $q(\cdot) \in \mathcal{C}$ and data $D_1^n$, then $\mathcal{C}$ is called a **conjugate family of priors**.

It frequently happens that a conjugate family of priors naturally co-exists with a given statistical model for the observable r.v. $X$. Suppose the model is such that there exists an $n_0 > 0$ with the property that for all data $D_1^n$ with $n \geq n_0$ the induced likelihood function $L(\theta | D_1^n)$ is integrable (with respect to some integrating measure $\mu$) over the parameter space $\Theta$.

Consider then the family $\mathcal{C}_0$ of p.d.f.'s $q(\theta)$ of the form:
\[ q(\theta) = \frac{L(\theta | D_1^n)}{\int_{\theta} L(\theta | D_1^n) d\mu(\theta)} \]

If the prior p.d.f. \( q(\theta) \) corresponds to a so called prior data \( D_1^n = (y_1, y_2, \ldots, y_n) \), then with the current data \( D_1^m = (x_1, x_2, \ldots, x_m) \) the posterior p.d.f. \( \theta | D_1^{n+m} \) will correspond to the likelihood function \( L(\theta | D_1^{n+m}) \), where \( D_1^{n+m} \) is the extended data \( (y_1, y_2, \ldots, y_n, x_1, x_2, \ldots, x_m) \). Thus for each prior \( q(\cdot) \in C_0 \), the posterior \( q(\cdot | D_1^m) \) belongs to \( C_0 \) for all possible current data \( D_1^m \).

The natural conjugate family \( C_0 \) of prior distributions take on a simple form when, irrespective of the sample size \( n \), there exists a sufficient statistic \( T = (T_1, T_2, \ldots, T_k) \) of fixed and small dimension \( k(k \geq 1) \). Then

\[ L(\theta | D_1^n) = H_n(\theta, T_1, T_2, \ldots, T_k), \]

where \( T_1, T_2, \ldots, T_k \) are functions of \( D_1^n \). In this case, the natural conjugate family \( C_0 \) of prior distributions is characterized by \( k + 1 \) super-parameters, namely, particular values of \( T_1, T_2, \ldots, T_k \) and \( n \).

For example, suppose each observation on an observable r.v. \( X \) belongs to one of the \( k + 1 \) mutually exclusive and collectively exhaustive categories. Let \( p_i (0 < p_i < 1) \) be the probability that an observation belongs to the \( i^{\text{th}} \) category, \( i = 1, 2, \ldots, k + 1 \), where \( \sum_{i=1}^{k+1} p_i = 1 \), then we may regard \((p_1, p_2, \ldots, p_k)\) as the model parameters. Suppose \( X \) is observed \( n \)-times and let \( D_1^n \) be the data \((x_1, x_2, \ldots, x_n)\) collected. Furthermore, let \( n_i \) denote the number of \( x \)'s that belong to the \( i^{\text{th}} \) category, \( i = 1, 2, \ldots, k + 1 \). Then each \( n_i \) is a non-negative integer and \( \sum_{i=1}^{k+1} n_i = n \).

Also, since \( \sum_{i=1}^{k+1} n_i = n \), we may regard \( T = (n_1, n_2, \ldots, n_k) \) as the \( k \)-dimensional
sufficient statistic. Before the data are collected, \((n_1, n_2, \ldots, n_k)\) are r.v.'s having a multinomial distribution with parameters \(n\) and \((p_1, p_2, \ldots, p_k)\).

The likelihood function \(L(p_1, p_2, \ldots, p_k | D^n) = \prod_{i=1}^{k} p_i^{n_i} (1 - \sum_{i=1}^{k} p_i)^{n_{k+1}}\). Consider the family \(C_0\) of distributions having p.d.f.

\[
q(p_1, p_2, \ldots, p_k) = \prod_{i=1}^{k} p_i^{a_i-1} \left(1 - \sum_{i=1}^{k} p_i\right)^{a_{k+1}-1}, \quad p_i > 0, \quad i = 1, 2, \ldots, k, \quad \sum_{i=1}^{k} p_i < 1,
\]

where each \(a_i\) is a positive integer. Then for any prior p.d.f. \(q(p_1, p_2, \ldots, p_k)\) in \(C_0\) and any data \(D^n\), the posterior p.d.f. \(q(p_1, p_2, \ldots, p_k | D^n) \propto \prod_{i=1}^{k} p_i^{a_i+n_i-1} \left(1 - \sum_{i=1}^{k} p_i\right)^{n_{k+1}+a_{k+1}-1}\), so \(q(p_1, p_2, \ldots, p_k | D^n)\) is also in \(C_0\).

Thus the family \(C_0\) is the natural conjugate family of prior distributions for the parameters \((p_1, p_2, \ldots, p_k)\). This is a sub-family of the family of the Dirichlet distributions, defined in the next section.

3. The Dirichlet distribution

This section is devoted to the study of the family of the Dirichlet distributions, as the natural conjugate family for the parameters of a multinomial distribution, and its characterizations. The Dirichlet distribution is defined as follows:

**Definition 3.1.** Let \(a_i > 0, \ i = 1, 2, \ldots, k+1\). The r.v.'s \((Y_1, Y_2, \ldots, Y_k)\) are said to have a Dirichlet distribution with parameters \((a_1, a_2, \ldots, a_{k+1})\),
denoted by \((Y_1, Y_2, \ldots, Y_k) \sim D(\alpha_1, \alpha_2, \ldots, \alpha_{k+1})\), if the joint distribution of \((Y_1, Y_2, \ldots, Y_k)\) has the p.d.f.

\[
f(y_1, y_2, \ldots, y_k) = \text{const.} \cdot y_1^{\alpha_1-1} y_2^{\alpha_2-1} \cdots y_k^{\alpha_k-1} (1 - y_1 - \cdots - y_k)^{\alpha_{k+1}-1},
\]

over the \(k\)-dimensional simplex \(S_k\) defined by the inequalities \(y_i > 0, i = 1, 2, \ldots, k, \sum_{i=1}^{k} y_i < 1\).

More generally, in the above definition we can take \(\alpha_i \geq 0\) for each \(i\), and \(\sum_{i=1}^{k+1} \alpha_i > 0\) and if \(\alpha_i = 0\) for some \(i\) then the corresponding \(Y_i = 0\) with probability one.

For \(k = 1\), the Dirichlet distribution \(D(\alpha_1, \alpha_2)\) for \(Y_1\) is the familiar Beta distribution with parameters \(\alpha_1\) and \(\alpha_2\), \(\text{Beta}(\alpha_1, \alpha_2)\).

The proof of the following basic proposition has already been outlined in the previous section.

**Proposition 3.1.** Let \(D(\alpha_1, \alpha_2, \ldots, \alpha_{k+1})\) be the prior probability model for the parameters \((p_1, p_2, \ldots, p_k)\) in the statistical model of a \(k+1\) valued r.v. \(X\). Then, with \(n\) independent observations on \(X\) giving rise to the sample frequencies \(n_1, n_2, \ldots, n_{k+1}\) for the \(k+1\) values of the r.v. \(X\), the posterior distribution of \((p_1, p_2, \ldots, p_k)\) will be \(D(\alpha_1 + n_1, \alpha_2 + n_2, \ldots, \alpha_{k+1} + n_{k+1})\).

The rest of this section is devoted to some characterizations of the Dirichlet distribution and elucidations of some of its more useful properties.

First of all, note that if we define \(Y_{k+1}^* = 1 - \sum_{i=1}^{k} y_i\) then the joint distribution of \((Y_1, Y_2, \ldots, Y_{k+1})\) is singular with respect to the \(k+1\) dimensional Lebesgue measure \(\lambda_{k+1}^*\) on \(R_{k+1}\). The support of this singular distribution is the \(k\)-dimensional simplex \(S_{k+1}\) defined by the inequalities
\[ y_i > 0, \ i = 1, 2, \ldots, k + 1, \ \sum_{i=1}^{k+1} y_i = 1. \] The joint p.d.f. (with respect to the \( k \)-dimensional Lebesque measure on \( E_{k+1} \)) of the \( k + 1 \) variables may be neatly represented as const. \( \prod_{i=1}^{k+1} y_i \). We shall use the notation \( \prod_{i=1}^{k+1} y_i \).

\( (Y_1, Y_2, \ldots, Y_{k+1}) \sim D_s(\alpha_1, \alpha_2, \ldots, \alpha_{k+1}) \) to denote that the joint distribution of the r.v.'s \( (Y_1, Y_2, \ldots, Y_{k+1}) \) is a singular Dirichlet with parameters \( (\alpha_1, \alpha_2, \ldots, \alpha_{k+1}) \).

The following result follows immediately:

**Proposition 3.2.** If \( i_1, i_2, \ldots, i_k \) is any sequence of distinct integers from the set \( \chi = \{1, 2, \ldots, k + 1\} \) then \( (Y_{i_1}, y_{i_2}, \ldots, y_{i_k}) \sim D(\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_k}) \).

A characterization of the Dirichlet distribution in terms of mutually independent Beta r.v.'s is given by

**Proposition 3.3.** Let \( (Y_1, Y_2, \ldots, Y_k) \sim D(\alpha_1, \alpha_2, \ldots, \alpha_{k+1}) \). Let \( U_1 = Y_1 \), and \( U_i = \frac{Y_i}{1 - \sum_{j=1}^{i-1} y_j} \), \( i = 2, 3, \ldots, k \). Then \( U_i \sim \text{Beta}(\alpha_i, \alpha_{i+1} + \alpha_{i+2} + \ldots + \alpha_{k+1}) \), \( i = 1, 2, \ldots, k \), and \( U_1, U_2, \ldots, U_k \) are mutually independent.

**Proof.** The joint p.d.f. of \( (Y_1, Y_2, \ldots, Y_k) \) is

\[
f(y_1, y_2, \ldots, y_k) = \text{const.} \prod_{i=1}^{k} y_i^{\alpha_i - 1} (1 - \sum_{i=1}^{k} y_i)^{\alpha_{k+1} - 1}, (y_1, y_2, \ldots, y_k) \in S_k.
\]

Consider the one-one transformation of \( S_k \) onto the \( k \)-dimensional cube \((0, 1)^k\) given by the relation

\[
y_1 = u_1, \ y_i = u_i \prod_{j=1}^{i-1} (1 - u_j), i = 2, 3, \ldots, k.
\]

The Jacobian

\[
\left| \frac{\partial(y_1, y_2, \ldots, y_k)}{\partial(u_1, u_2, \ldots, u_k)} \right| = \prod_{i=1}^{k} (1 - u_i)^{k-i}.
\]
It follows then that the joint p.d.f. of \((U_1, U_2, \ldots, U_k)\) is
\[
g(u_1, u_2, \ldots, u_k) = \text{const.} \prod_{i=1}^{k} \frac{\alpha_i^{u_i - 1} (1 - u_i)^{\alpha_i + \alpha_{i+1} + \cdots + \alpha_{k+1} - 1}}{u_i}.
\]

We note that the converse of Prop. 3.3 is clearly true.

As a byproduct of the above proposition we immediately have that the

r.v. \(Y_i \sim \text{Beta}(\alpha_i, \alpha - \alpha_i)\), where \(\alpha = \sum_{i=1}^{k+1} \alpha_i\). This fact together with Prop. 3.2 then gives

**Corollary 3.1.** If \((Y_1, Y_2, \ldots, Y_k) \sim D(\alpha_1, \alpha_2, \ldots, \alpha_{k+1})\), then

\(Y_i \sim \text{Beta}(\alpha_i, \alpha - \alpha_i)\), \(i = 1, 2, \ldots, k\).

This corollary may be generalized by using the converse of Prop. 3.3 to

**Corollary 3.2.** If \((Y_1, Y_2, \ldots, Y_k) \sim D(\alpha_1, \alpha_2, \ldots, \alpha_{k+1})\), then

\((Y_1, Y_2, \ldots, Y_r) \sim D(\alpha_1, \alpha_2, \ldots, \alpha_r, \alpha - \sum_{j=1}^{r} \alpha_i)\).

A more general extension is then given by Prop. 3.2 and the converse of Prop. 3.3 as:

**Corollary 3.3.** For any subset \(\{i_1, i_2, \ldots, i_r\}\) of \(X\), \((Y_{i_1}, Y_{i_2}, \ldots, Y_{i_r}) \sim D(\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_r}, \alpha - \sum_{j=1}^{r} \alpha_{i_j})\).

The converse of prop. 3.3 is again useful in establishing the following:

**Proposition 3.4.** Let \((Y_1, Y_2, \ldots, Y_k) \sim D(\alpha_1, \alpha_2, \ldots, \alpha_{k+1})\). Then, for any integer \(r\) such that \(2 \leq r \leq k\),

\[
\frac{Y_1}{1-Y_1}, \ldots, \frac{Y_r}{1-Y_r}, \ldots, \frac{Y_k}{1-Y_k}
\]

is independent of \((Y_1, Y_2, \ldots, Y_{r-1})\).

Also, \((\frac{Y_1}{1-Y_1}, \frac{Y_{r+1}}{1-Y_{r+1}}, \ldots, \frac{Y_k}{1-Y_k}) \sim D(\alpha_1, \alpha_{r+1}, \ldots, \alpha_{k+1})\).
Proof. Consider the mutually independent r.v.'s \((U_1, U_2, \ldots, U_k)\) defined as in Prop. 3.3. Note that \((U_1, U_2, \ldots, U_{r-1})\) is independent of \((u_r, U_{r+1}, \ldots, U_k)\).

It follows then that \((U_1, U_2(1 - U_1), \ldots, U_{r-1} \Pi_{i=1}^{r-2} (1 - U_i))\) is independent of \((U_r, U_{r+1}(1 - U_r), \ldots, U_k \Pi_{i=r}^{k-1} (1 - U_i))\). That is, \((Y_1, Y_2, \ldots, Y_{r-1})\) independent of \(Y_r\), \(Y_{r+1}\), \ldots, \(Y_k\) \(i = r\). Now, from the converse of Prop. 3.3 it follows that \((U_r, U_{r+1}(1 - U_r), \ldots, U_k \Pi_{i=r}^{k-1} (1 - U_i)) \sim D(\alpha_r, \alpha_{r+1}, \ldots, \alpha_{k+1})\).

Thus, \(\frac{Y_r}{1 - Y_1 \cdots Y_{r-1}}, \frac{Y_{r+1}}{1 - Y_1 \cdots Y_{r-1}}, \ldots, \frac{Y_k}{1 - Y_1 \cdots Y_{r-1}} \sim D(\alpha_r, \alpha_{r+1}, \ldots, \alpha_{k+1})\).

Thus, the conditional r.v.'s \((Y_r, Y_{r+1}, \ldots, Y_k) \mid Y_1, Y_2, \ldots, Y_{r-1}\) have the property that the joint distribution of \(\frac{Y_r}{1 - Y_1 \cdots Y_{r-1}}, \frac{Y_{r+1}}{1 - Y_1 \cdots Y_{r-1}}, \ldots, \frac{Y_k}{1 - Y_1 \cdots Y_{r-1}}\) given \((Y_1, Y_2, \ldots, Y_{r-1})\) is \(D(\alpha_r, \alpha_{r+1}, \ldots, \alpha_{k+1})\), independent of \((Y_1, Y_2, \ldots, Y_{r-1})\).

The Dirichlet distribution can also be characterized in terms of mutually independent Gamma r.v.'s. This is given by the following:

Proposition 3.5. Let \(Z_1, Z_2, \ldots, Z_{k+1}\) be mutually independent Gamma r.v.'s with the common scale parameter \(\beta > 0\) and possibly different shape parameters \(\alpha_i > 0, i = 1, 2, \ldots, k + 1\). Let \(Z = \sum_{i=1}^{k+1} Z_i^\beta\) and \(Y_i = \frac{Z_i}{Z}, i = 1, 2, \ldots, k + 1\).
Then \((Y_1, Y_2, \ldots, Y_k) \sim D(\alpha_1, \alpha_2, \ldots, \alpha_{k+1})\). Also, \((Y_1, Y_2, \ldots, Y_k)\) is independent of \(Z\).

**Proof.** The joint p.d.f. of the \(Z_i\)'s is

\[
f(z_1, z_2, \ldots, z_{k+1}) = \text{Const. } e^{-\sum_{i=1}^{k+1} z_i} \prod_{i=1}^{k} z_i^{\alpha_i-1}, z_i > 0, i = 1, 2, \ldots, k+1.
\]

Consider the transformation

\[
z = \sum_{i=1}^{k+1} z_i, y_i = \frac{z_i}{z}, i = 1, 2, \ldots, k,
\]

the reverse transformation being

\[
z_i = z y_i, i = 1, 2, \ldots, k, \text{ and } z_{k+1} = z(1 - \sum_{i=1}^{k} y_i).
\]

The Jacobian

\[
\left|\frac{\partial(z_1, z_2, \ldots, z_{k+1})}{\partial(z, y_1, \ldots, y_k)}\right| = z^k.
\]

It follows then that the joint p.d.f. of \((Z, Y_1, \ldots, Y_k)\) is

\[
g(z, y_1, \ldots, y_k) = \text{Const. } e^{-\beta z} \prod_{i=1}^{k} y_i^{\alpha_i-1} (1 - \sum_{i=1}^{k} y_i)^{\alpha_{k+1}-1}
\]

The converse of Prop. 3.5 is clearly true. That is, if \(Z\) is a r.v. having a

Gamma distribution with shape parameter \(\alpha = \sum_{i=1}^{k+1} \alpha_i > 0\) and scale parameter \(\beta > 0\), denoted by \(Z \sim G(\alpha, \beta)\), and if \(Z\) is independent of \((Y_1, Y_2, \ldots, Y_k)\),

where \((Y_1, Y_2, \ldots, Y_k) \sim D(\alpha_1, \alpha_2, \ldots, \alpha_{k+1})\). Then the r.v.'s
\[ Z_i = Z Y_i, \ i = 1, 2, \ldots, k, \text{ and } Z_{k+1} = Z(1 - \sum_{i=1}^{k} Y_i) \] are mutually independent and \( Z_i \sim G(\alpha_i, \beta), \ i = 1, 2, \ldots, k + 1. \)

That \((Y_1, Y_2, \ldots, Y_k)\) is independent of \(Z\) in Prop. 3.5 is a consequence of the Basu Theorem (Basu (1955)): Let \((X, A, \{P_\theta: \theta \in \Theta\})\) be a statistical model. If \(T\) is a boundedly complete sufficient statistic then any ancillary statistic is independent of \(T\).

To see that \((Y_1, Y_2, \ldots, Y_k)\) is independent of \(Z\), proceed as follows:

Regard the parameters \((\alpha_1, \alpha_2, \ldots, \alpha_{k+1})\), where each \(\alpha_i > 0\), as known constants and \(\beta > 0\) as the unknown parameter. With \((Z_1, Z_2, \ldots, Z_{k+1})\) as the sample, \(Z = Z_1 + Z_2 + \ldots + Z_{k+1}\) is a complete sufficient statistic.

The vector valued statistic \(\left(\frac{Z_1}{Z}, \frac{Z_2}{Z}, \ldots, \frac{Z_{k+1}}{Z}\right)\) is scale invariant. Since \(\beta > 0\) is a scale parameter, it follows that \(\left(\frac{Z_1}{Z}, \frac{Z_2}{Z}, \ldots, \frac{Z_{k+1}}{Z}\right)\) is an ancillary statistic. From the Basu theorem, then, it follows that

\[ \left(\frac{Z_1}{Z}, \frac{Z_2}{Z}, \ldots, \frac{Z_{k+1}}{Z}\right) = (Y_1, Y_2, \ldots, Y_{k+1}) \] is independent of \(Z\). However, since

\[ Y_{k+1} = 1 - \sum_{i=1}^{k} Y_i, \] then \((Y_1, Y_2, \ldots, Y_k)\) is independent of \(Z\).

Notice that for \(k = 2\), the above proposition is the familiar result:

Remark 3.1. Let \(Z_1 \sim G(\alpha_1, \beta)\), \(i = 1, 2\), and let \(Z_1\) and \(Z_2\) be independent.

Then \(\frac{Z_1}{Z_1 + Z_2}\) is independent of \(Z_1 + Z_2\). Also, \(\frac{Z_1}{Z_1 + Z_2} \sim \text{Beta}(\alpha_1, \alpha_2)\) and \(Z_1 + Z_2 \sim G(\alpha_1 + \alpha_2, \beta)\).
The following is an alternative proof of Prop. 3.4 using the Gamma characterization of the Dirichlet distribution.

**Alternative Proof of Prop. 3.4.** Note that
\[
\begin{align*}
\frac{Z_r}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}, & \quad \frac{Z_{r+1}}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}, & \quad \ldots, & \quad \frac{Z_{k+1}}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}
\end{align*}
\]
is independent of
\[
(Z_r + Z_{r+1} + \ldots + Z_{k+1}). \quad \text{Also,} \quad (Z_1, Z_2, \ldots, Z_{r-1}) \quad \text{is independent of}
\]
\[
\begin{align*}
\frac{Z_r}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}, & \quad \frac{Z_{r+1}}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}, & \quad \ldots, & \quad \frac{Z_{k+1}}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}
\end{align*}
\]
and
\[
(Z_r + Z_{r+1} + \ldots + Z_{k+1}). \quad \text{Thus, the r.v.'s} \quad (Z_1, Z_2, \ldots, Z_{r-1}),
\]
\[
(Z_r + Z_{r+1} + \ldots + Z_{k+1}), \quad \text{and} \quad (\frac{Z_r}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}, \frac{Z_{r+1}}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}, \ldots,
\]
\[
\frac{Z_{k+1}}{Z_r + Z_{r+1} + \ldots + Z_{k+1}})
\]
are mutually independent and so \((Z_1, Z_2, \ldots, Z_{r-1})
\]
\[
(Z_r + Z_{r+1} + \ldots + Z_{k+1}) \quad \text{is independent of}
\]
\[
\frac{Z_r}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}, \quad \frac{Z_{r+1}}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}, \quad \ldots,
\]
\[
\frac{Z_{k+1}}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}.
\]
Thus, \((Z_1 + Z_2 + \ldots + Z_{k+1}, Z_1 + Z_2 + \ldots + Z_{k+1}, \ldots,
\]
\[
\frac{Z_{r-1}}{Z_1 + Z_2 + \ldots + Z_{k+1}}, \quad \frac{Z_{r-1}}{Z_1 + Z_2 + \ldots + Z_{k+1}})
\]
is independent of \((\frac{Z_r}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}, \frac{Z_{r+1}}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}, \ldots,
\]
\[
\frac{Z_{k+1}}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}).
\]
That is, \((Y_1, Y_2, \ldots, Y_{r-1}) \quad \text{is independent of}
\]
\[
\left(\frac{Y_r}{1-Y_1 - \ldots - Y_{r-1}}, \frac{Y_{r+1}}{1-Y_1 - \ldots - Y_{r-1}}, \ldots, \frac{Y_k}{1-Y_1 - \ldots - Y_{r-1}}\right).
\]
Now following along the lines of Prop. 3.5, \((\frac{Z_r}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}, \frac{Z_{r+1}}{Z_r + Z_{r+1} + \ldots + Z_{k+1}}, \ldots,
\]
\[
\frac{Z_{k+1}}{Z_r + Z_{r+1} + \ldots + Z_{k+1}})
\]
\sim D(\alpha_r, \alpha_{r+1}, \ldots, \alpha_{k+1}). \quad \text{Thus,} \quad \left(\frac{Y_r}{1-Y_1 - \ldots - Y_{r-1}}, \frac{Y_{r+1}}{1-Y_1 - \ldots - Y_{r-1}}, \ldots,
\]
\[
\frac{Y_k}{1-Y_1 - \ldots - Y_{r-1}}\right) \sim D(\alpha_r, \alpha_{r+1}, \ldots, \alpha_{k+1}). \quad \Box
4. Extension of the results of Section 3

Suppose $A$ is any subset of the set $X = \{1, 2, \ldots, k+1\}$ and

$(y_1, y_2, \ldots, y_{k+1})$ is any given point in the simplex $E_{k+1}$. Define

$$P(A) = \sum_{i \in A} y_i.$$ 

Then, $P$ is a probability measure on $X$, and $P$ can be identified by the point $(y_1, y_2, \ldots, y_{k+1})$ in $E_{k+1}$. In this way, $E_{k+1}$ represents the class of all probability measures on $X$. If $(Y_1, Y_2, \ldots, Y_{k+1})$ is a random point in $E_{k+1}$ then $P(A) = \sum_{i \in A} Y_i$ is random probability measure of the set $A$, and $P$ is a random probability measure on $X$. Thus, a random probability measure on $X$ can be viewed as a probability measure on $E_{k+1}$. In particular, we shall now consider $D(\alpha_1, \alpha_2, \ldots, \alpha_{k+1})$ as a distribution on $E_{k+1}$. So let $(Y_1, Y_2, \ldots, Y_k) \sim D(\alpha_1, \alpha_2, \ldots, \alpha_{k+1})$, and let $Z_1, Z_2, \ldots, Z_{k+1}$ be mutually independent Gamma r.v.'s with $Z_i \sim G(\alpha_i, \beta)$, $i = 1, 2, \ldots, k+1$, defined as in Prop. 3.5.

Put $Z(A) = \sum_{i \in A} Z_i$ and $\alpha(A) = \sum_{i \in A} \alpha_i$. Then

$$P(A) = \sum_{i \in A} Y_i = \frac{Z(A)}{Z(X)}.$$

For any subsets $A$ and $B$ of $X$, let $P(A|B)$ be the random conditional probability of $A$ given $B$ defined as:

$$P(A|B) = \begin{cases} \frac{P(AB)}{P(B)}, & \text{if } P(B) > 0 \\ 0, & \text{if } P(B) = 0. \end{cases}$$

Note that for any collection $(A_1, A_2, \ldots, A_n)$ of disjoint subsets of $X$, the r.v.'s $Z(A_1), Z(A_2), \ldots, Z(A_n)$ are mutually independent and $Z(A_1) \sim G(\alpha(A_1), \beta)$, $i = 1, 2, \ldots, n$.

The following is a general property of the Dirichlet distribution.
Proposition 4.1. Let $Y_i$, $i = 1, 2, \ldots, k + 1$, be defined as in Prop. 3.5. Let $X$ be partitioned into non-empty subsets $A_1, A_2, \ldots, A_{m+1}$, $1 \leq m \leq k$. Then, $(P(A_1), P(A_2), \ldots, P(A_m)) \sim D(\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_{m+1}))$. Also, $(P(A_1), P(A_2), \ldots, P(A_m))$ is independent of $Z(X)$.

Proof. Note that $P(A_i) = \frac{Z(A_i)}{Z(X)}$, $i = 1, 2, \ldots, m + 1$, and that $Z(A_1), Z(A_2), \ldots, Z(A_{m+1})$ are mutually independent Gamma r.v.'s with the common scale parameter $\beta > 0$ and different shape parameters $\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_{m+1})$, respectively. Now, invoke Prop. 3.5.

The following result is an immediate corollary.

Corollary 4.1. For any two disjoint subsets $A_1$ and $A_2$ of $X$, $\frac{Z(A_1)}{Z(A_1 \cup A_2)}$ is independent of $Z(A_1 \cup A_2)$. Also, $\frac{Z(A_1)}{Z(A_1 \cup A_2)} \sim Beta(\alpha(A_1), \alpha(A_2))$, and $Z(A_1 \cup A_2) \sim G(\alpha(A_1 \cup A_2), \beta)$.

As a direct consequence of the above corollary we have the following:

Remark 4.1. The marginal distribution of the sum of any $r$, $1 \leq r \leq k$, r.v.'s $Y_{i_1}, Y_{i_2}, \ldots, Y_{i_r}$ is a Beta $(\alpha_{i_1} + \alpha_{i_2} + \ldots + \alpha_{i_r} - \sum_{j=1}^{r} \alpha_{i_j})$, where $i_1, i_2, \ldots, i_r$ is any sequence of $r$ distinct integers from $X = \{1, 2, \ldots, k + 1\}$. The following Propositions 4.2 and 4.3 are, respectively, extensions of Propositions 3.3 and 3.4.

Proposition 4.2. Let $A_i, i = 2, \ldots, m + 1, 1 \leq m \leq k$, be defined as in Prop. 4.1. Then, the r.v.'s $P(A_1), P(A_2 | A_1^C), P(A_3 | A_1^C A_2^C), \ldots,$

$P(A_m | A_1^C A_2^C \cdots A_{m-1}^C)$ are mutually independent with $P(A_1) \sim Beta(\alpha(A_1), \alpha(A_1^C))$ and $P(A_r | A_1^C A_2^C \cdots A_{r-1}^C) \sim Beta(\alpha(A_r), \alpha(A_r^C A_r A_1^C \cdots A_{r-1}^C)), r = 2, 3, \ldots, m$, where for any subset $A$ of $X$, $A^C$ denotes the complement of $A$. 

Proposition 4.3. Let $A_i, i = 1, 2, \ldots, m + 1, 1 \leq m \leq k$, be defined as in Prop. 4.1. Then $(P(A_r | A_1^c A_2^c \ldots A_{r-1}^c), P(A_{r+1} | A_1^c A_2^c \ldots A_{r-1}^c), \ldots, P(A_m | A_1^c A_2^c \ldots A_{r-1}^c))$ is independent of $(P(A_1), P(A_2), \ldots, P(A_{r-1}))$, $2 \leq r \leq m$. Also, $(P(A_r | A_1^c A_2^c \ldots A_{r-1}^c), P(A_{r+1} | A_1^c \ldots A_{r-1}^c), \ldots, P(A_m | A_1^c A_2^c \ldots A_{r-1}^c)) \sim D(\alpha(A_r), \alpha(A_{r+1}), \ldots, \alpha(A_{m+1}))$. The proofs of Propositions 4.2 and 4.3 follow along the lines of Propositions 3.3 and 3.4, respectively.

The next result is a preliminary to Prop. 4.4.

Lemma 4.1. Let $B_1$ and $B_2$ be any two subsets of $X$. Then, the r.v.'s $P(B_1), P(B_2 | B_1), P(B_2 | B_1^c)$ are mutually independent.

Proof. It suffices to show that the r.v.'s $Z(B_1) Z(B_1 B_2) Z(B_1 B_2^c) Z(B_1^c B_2)$ are mutually independent. Since the r.v.'s $Z(B_1 B_2), Z(B_1 B_2^c), Z(B_1^c B_2), Z(B_1^c B_2^c)$ are mutually independent, the pairs $(Z(B_1 B_2), Z(B_1 B_2^c))$ and $(Z(B_1^c B_2), Z(B_1^c B_2^c))$ of r.v.'s are independent both "within and between (w.b.)". Applying Corollary 4.1 to each of the two pairs we have then that the pairs $(Z(B_1 B_2) Z(B_1), Z(B_1))$ and $(Z(B_1 B_2^c) Z(B_1^c), Z(B_1^c))$ of r.v.'s are independent both "w.b.". Thus, the r.v.'s $Z(B_1 B_2) Z(B_1 B_2^c) Z(B_1^c B_2) Z(B_1^c B_2^c)$ are mutually independent. Applying Corollary 4.1 to the last two r.v.'s we finally conclude that the r.v.'s $Z(B_1 B_2) Z(B_1 B_2^c) Z(B_1^c B_2) Z(B_1^c B_2^c)$ are mutually independent.

For any subset $B$ of $X$, define $B^t$ as $B$ when $t = 1$ and as $B^c$ when $t = 0$.

We shall now state Prop. 4.4.
Proposition 4.4. For any collection $B_1, B_2, \ldots, B_n$ of subsets of $X$, the r.v.'s $P(B_i), \{P(B_2 | B_1^c); t_1 = 0 \text{ or } 1\}, \ldots, \{P(B_n | B_1 B_2^c \cdots B_{n-1}^c); t_1 = 0 \text{ or } 1, i = 1, 2, \ldots, n-1\}$ are mutually independent with $P(B_1)$ \\

Beta($\alpha(B_1^c), \alpha(B_1^0)$) and $P(B_r | B_1 B_2^c \cdots B_{r-1}^c B_r^c) \sim Beta(\alpha(B_1^c) B_2^c \cdots B_{r-1}^c B_r^c, \alpha(B_1^c B_2^c \cdots B_{r-1}^c B_r^c)), t_i = 0 \text{ or } 1, i = 1, 2, \ldots, r-1, r = 2, 3, \ldots, n$.

Proof. We proceed by induction. For $n = 1$ there is nothing to prove.

For $n = 2$, the proof is given in Lemma 4.1. Now, suppose the result holds for $n - 1$. Then to prove that it also holds for $n$, we must show that for $t_1 = 0$ or $1, i = 1, 2, \ldots, n - 1$ are mutually independent. It suffices to show that the two sets $
abla$ and $\{Z(B_1 B_2^c \cdots B_{n-1}^c B_n^c); t_1 = 0 \text{ or } 1, i = 1, 2, \ldots, n - 1\}$ of $2^{n-1}$ r.v.'s in each are independent both "w.b." Since the $2^n$ r.v.'s $\{Z(B_1 B_2^c \cdots B_n^c); t_1 = 0 \text{ or } 1, i = 1, 2, \ldots, n\}$ are mutually independent, the $2^{n-1}$ pairs $\{(Z(B_1 B_2^c \cdots B_{n-1}^c B_n^c), Z(B_1 B_2^c \cdots B_{n-1}^c B_n^c)); t_1 = 0 \text{ or } 1, i = 1, 2, \ldots, n - 1\}$ of r.v.'s are independent both "w.b." Applying Corollary 4.1 to each of the $2^{n-1}$ pairs we have then that the
2^{n-1} \text{ pairs } \left\{ \frac{Z(B_1 B_2 \ldots B_{n-1} \cdot n)}{t_i^{1} t_2^{2} t_{n-1}^{n-1}}, \frac{Z(B_1 B_2 \ldots B_{n-1} \cdot n)}{t_i^{1} t_2^{2} t_{n-1}^{n-1}} \right\}; t_i = 0 \text{ or } 1, \ i = 1, 2, \ldots, n - 1 \right\}, \ \ \ text{of } 2^{n-1} \text{ r.v.'s are independent both } \text{"w.b.". Thus, the two sets } \\
\left\{ \frac{Z(B_1 B_2 \ldots B_{n-1} \cdot n)}{t_i^{1} t_2^{2} t_{n-1}^{n-1}}, t_i = 0 \text{ or } 1, \ i = 1, 2, \ldots, n - 1 \right\} \text{ and } \left\{ Z(B_1 B_2 \ldots B_{n-1} \cdot n) \right\}, \ \ \ t_i = 0 \text{ or } 1, \ i = 1, 2, \ldots, n - 1 \right\} \text{ of } 2^{n-1} \text{ r.v.'s in each are independent both } \text{"w.b.". Also, observe that for each } r, 1 \leq r \leq n, \\
\frac{Z(B_1 B_2 \ldots B_{r-1} \cdot r)}{Z(B_1 B_2 \ldots B_{r-1} \cdot r)}; t_i = 0 \text{ or } 1, \ i = 1, 2, \ldots, r - 1. \right\} \\
\text{The following result is an immediate corollary.} \\
\textbf{Corollary 4.2.} \ \text{For any subsets } B_1 \supset B_2 \supset \ldots \supset B_n \text{ of } \chi, \text{ the r.v.'s } P(B_1), \\
P(B_2 | B_1), \ldots, P(B_n | B_{n-1}) \text{ are mutually independent.} \\
The following result is a consequence of Corollary 4.2 with } B_1 = \{1, 2, \ldots, k - i + 1\}, i = 1, 2, \ldots, k. \\
\textbf{Remark 4.2.} \ \text{Let } Y_1, Y_2, \ldots, Y_k \text{ be defined as in Prop. 3.5. Define} \\
V_i = \frac{Y_1 + Y_2 + \ldots + Y_i}{Y_1 + Y_2 + \ldots + Y_i}, \ i = 1, 2, \ldots, k - 1, \text{ and } V_k = Y_1 + Y_2 + \ldots + Y_k. \\
Then, the r.v.'s } V_1, V_2, \ldots, V_k \text{ are mutually independent, and } V_i \sim \\
\text{Beta}(a_1 + a_2 + \ldots + a_i, a_{i+1}), i = 1, 2, \ldots, k. \\
\text{A collection } B_1, B_2, \ldots, B_n \text{ of subsets of } \chi \text{ forms a separating sequence if } \\
\text{for any two distinct points } x_1 \text{ and } x_2 \text{ in } \chi, \text{ there exists a set } B_i \text{ in the collection} \\
\text{which contains either } x_1 \text{ or } x_2 \text{ but not both. Equivalently } \bigcap_{i=1}^{n} B_i \text{ is either a} \\
single point set or the empty set, where each } t_i = 0 \text{ or } 1.
Remark 4.3. If \{B_1, B_2, \ldots, B_n\} is a separating sequence and the r.v.'s \(P(B_1), \ldots, P(B_n)\); \(t_1 = 0 \text{ or } 1\), \(t_2 = 0 \text{ or } 1\), \(t_{n-1} = 0 \text{ or } 1\), \(i = 1, 2, \ldots, n-1\) are mutually independent with \(P(B_1) \sim \text{Beta}(\alpha(B_1), \alpha(B_1))\) and \(P(B_r | B_1 \ldots B_{r-2} B_{r-1}) \sim \text{Beta}(\alpha(B_1 \ldots B_{r-2} B_{r-1}), \alpha(B_1 \ldots B_{r-2} B_{r-1} B_r))\), \(t_1 = 1, 2, \ldots, r-1, r = 2, 3, \ldots, n\), then, as a consequence of our discussion in this section for every partition \((A_1, A_2, \ldots, A_n)\) of \(X\), \((P(A_1), P(A_2), \ldots, P(A_n)) \sim D_\text{\(A\)}(\alpha(A_1), \alpha(A_2), \ldots, (A_n))\). In this case, the random probability measure \(P\) said to be a Dirichlet process on \((X, A)\) (see Definition 6.1), where \(A\) is the set of all subsets of \(X\).

Up to this stage, only finite dimensional Dirichlet distributions were considered. A more general Dirichlet distribution may be defined as follows:

Let \(\alpha_1, \alpha_2, \ldots\) be a sequence of numbers satisfying \(\alpha_i > 0\) for each \(i\), and \(\sum_{i=1}^{\infty} \alpha_i < \infty\). A sequence \((Y_1, Y_2, \ldots)\) of r.v.'s such that \(0 \leq Y_i \leq 1\) for each \(i\), and \(\sum_{i=1}^{\infty} Y_i = 1\) is said to have a Dirichlet distribution with parameters \((\alpha_1, \alpha_2, \ldots)\), if for each \(k\), \((Y_1, Y_2, \ldots, Y_k) \sim D(\alpha_1, \alpha_2, \ldots, \alpha_k, \sum_{i=k+1}^{\infty} \alpha_i)\).

In the above definition, we can take \(\alpha_i > 0\) for each \(i\), and \(0 < \sum_{i=1}^{\infty} \alpha_i < \infty\).

and if \(\alpha_i = 0\) for some \(i\) then the corresponding \(Y_i = 0\) with probability one.

The results of Sections 3 and 4 may be generalized for this Dirichlet distribution. In particular, remark 4.3 and its natural generalization shows the existence of a Dirichlet process on a finite or a countably infinite space.

In part II we shall show how this approach leads us to the Blackwell's (1973) constructive definition of a Dirichlet process on a Borel space.
PART TWO: The Dirichlet Process

6. Dirichlet Process Preliminaries

In the Bayesian analysis of nonparametric problem there is a sequence \( \{X_1, X_2, \ldots\} \) of independent identically distributed (i.i.d.) random variables with a common unknown distribution \( P \); that is, given \( P = P \) the \( X_n \)'s are i.i.d. \( P \). Here \( P \) is regarded as the parameter and belongs to \( P \), the class of all probability measures on a given space \((\mathbb{X}, \mathcal{A})\). A prior for \( P \) is a probability measure on \((P, \sigma(P))\), where \( \sigma(P) \) is the smallest \( \sigma \)-field of subsets of \( P \) such that the map \( P \times \mathcal{P}(A) \) from \( P \) into \([0, 1]\) is \( \sigma(P) \)-measurable \( \forall A \in \mathcal{A} \). This prior may be viewed as a stochastic process \( \{P(A) : A \in \mathcal{A}\} \) whose sample functions are probability measures on \((\mathbb{X}, \mathcal{A})\). As in the parametric case, a class of processes satisfying the following properties is desired:

(I) it is wide enough to accommodate various shades of opinion about \( P \),

(II) if a prior is selected from this class, then the posterior distribution given a sample of observations from the true (unknown) distribution is manageable analytically, and it belongs to the class, i.e., the class is closed under "the Bayesian operation."

The class of Dirichlet process introduced by Ferguson (1973) is especially convenient since it satisfies the properties (I) and (II).

Let us look back at the definition of a random probability measure as given in the abstract of the paper. A random probability measure \( P \) on an arbitrary measurable space \((\mathbb{X}, \mathcal{A})\) may be viewed as a measurable map from a probability space \((\Omega, \sigma(\Omega), \mu)\) to the space \((P, \sigma(P))\). It may also be regarded as a transition function from \((\Omega, \sigma(\Omega), \mu)\) into \((P, \sigma(P)) \rightarrow P(\cdot, \cdot)\) is a measurable map from \( \Omega \times \mathcal{A} \) into \([0, 1]\) such that (i) for every \( \omega \) in \( \Omega \), \( P(\omega, \cdot) \) is a
probability measure on \((X, A)\), and (ii) for every set \(A\) in, \(P(\cdot, A)\) is 
\(\omega(\mathcal{P})\) - measurable, i.e., \(P(\cdot, A)\) is a random variable with values in 
\([0, 1]\). The distribution of \(P\), namely \(\omega^{-1}\), is the prior probability mea-
sure on \((\mathcal{P}, \omega(\mathcal{P}))\). Therefore, this paper can be thought of as dealing with 
a class of random probability measures, with a class of stochastic processes, 
or with a class of prior probabilities.

The Dirichlet process is defined as follows:

**Definition 6.1.** Let \(\alpha\) be a finite measure on \((X, A)\). A Dirichlet process 
\(\widetilde{D}^\alpha\) is a random probability measure on \((X, A)\) such that, for every partition 
\((A_1, A_2, \ldots, A_k)\) of \(X\) into a finite number of measurable sets, the joint 
distribution of the random variables \(P(A_1), P(A_2), \ldots, P(A_k)\) is a singular Dirichlet with parameters \((\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_k))\).

Ferguson (1973) shows through the Kolmogorov extension theorem that 
there exists a probability measure on \(([0, 1]^A, \mathcal{O}([0, 1]^A))\) yielding the 
above finite dimensional Dirichlet distributions. Here \([0, 1]^A\) is the product 
space having for each of its factors the closed unit interval \([0, 1]\), 
there being as many factors as elements of \(A\). In most of the applications, \(A\) 
has uncountably many elements. Equivalently, \([0, 1]^A\) may be viewed as the 
class of all set functins defined on \(A\) with values in \([0, 1]\). Also, \(\sigma([0, 1]^A)\) 
is the product \(\sigma\)-field for \([0, 1]^A\), the \(\sigma\)-field generated by the measurable 
cylinders having a finite base, i.e., a cylinder whose base is determined 
by measurable restrictions on a finite number of coordinates of \([0, 1]\).

Note that the set \(F\) of measurable cylinders having a countable base is a 
\(\sigma\)-field, and such measurable cylinders are in \(([0, 1]^A)\). Thus, \(F\) is equal 
to \(\sigma([0, 1]^A)\). Viewing \([0, 1]\) as a class of set functions, each set in
as a class of set functions, each set in \([0, 1]^A\) may be defined by restrictions on a countable collection \(\{p(A_n); n = 1, 2, \ldots\}\), where \(\{A_1, A_2, \ldots\}\) is a given countable subset of \(A\) and \(p\) denotes an element of \([0, 1]^A\). Observe that with \(A\) uncountable the single point sets in \([0, 1]^A\) are not in \(\sigma([0, 1]^A)\). Also, the class \(P\) of all probability measures on \((X, A)\) does not belong to \(\sigma([0, 1]^A)\); it is not determined by a countable number of restrictions when \(A\) is uncountable. Thus a statement like "a Dirichlet process gives probability one to the class \(P\)" is not meaningful.

Berk and Savage (1975) discuss other technical problems relating to measurability in addition to some fundamental difficulties with Ferguson's definition of a Dirichlet process. However, as proved by Blackwell (1973), none of these difficulties arise when \((X, A)\) is a Borel space. The basic idea of Blackwell's paper is the construction of a mapping from some probability space to \((P, \sigma(P))\) that produces a random probability measure \(P\) on \((X, A)\) which is a Dirichlet process.

Section 7 is devoted to some basic results. A discussion on Blackwell's construction is presented in Section 8. In addition, in Section 9 we present simple proofs for the following assertions:

1. The Dirichlet process \(D^\alpha\) has its support on the class of all discrete probability measures.

2. Under some conditions on the parameter \(\alpha\), for almost all realization \(P\) of \(P\) the set \(\text{Ep}\) of discrete mass points of \(P\) is dense in \(X\).
7. Some useful results on a Borel space

Let $(X, A)$ be a Borel space—a complete separable metric space $X$ with $A$ being the $\sigma$-field generated by the open subsets of $X$. Since $X$ is a separable metric space, $A$ is countably generated. So without any loss of generality we can assume that there exists a countable field $\mathcal{B} = \{B_1, B_2, \ldots\}$ such that its Borel extension is $A$. The family $\mathcal{B}$ forms a separating sequence, i.e., for any two distinct points $x_1$ and $x_2$ in $X$ there exists a set $B_n \in \mathcal{B}$ which contains either $x_1$ or $x_2$ but not both. Equivalently

$$\bigcap_{i=1}^{\infty} B_{t_i}$$

is either a single point set or the empty set, where each $t_i = 0$ or $1$.

Consider all sequences $t = (t_1, t_2, \ldots)$ such that each $t_i = 0$ or $1$. Let $T = \{t\}$ be the class of all such sequences and $T$ be the $\sigma$-field for $T$—the $\sigma$-field generated by the cylinders having a finite base, the so called Kolmogorov's sets. Then $(T, T)$ is a Borel space.

Consider the map $\xi: X \times T$ defined by $\xi(x) = (\xi_1(x), \xi_2(x), \ldots)$, where $\xi_i$ is the indicator of $B_{t_i}$, $i = 1, 2, \ldots$. Notice that $\xi$ is a measurable map since each coordinate is measurable. Also, $\xi$ is one-one since any two $x$'s that agree on all $\xi_i$'s are the same. Then we have the following:

**Lemma 7.1.** $\xi(X)$ is a Borel subset of $T$.

The proof of the above lemma is a direct consequence of

**The Kuratowski Theorem** (Parthasarathy [1967], Theorem 3.9): If $\rho$ is a one-one measurable map from a Borel subset $E_1$ of a complete separable metric space into another complete separable metric space with $\rho(E_1) = E_2$, then $E_2$ is a Borel set. Also, the map $\rho$ from $E_1$ onto $E_2$ is one-one and bimeasurable.
Let $[0, 1]^\infty$ be the set of all sequences $(w_1, w_2, \ldots)$ with $0 \leq w_n \leq 1$ for each $n$. Note that $([0, 1]^\infty, \sigma([0, 1]^\infty))$ is a Borel space. Consider the map $\eta: P \to [0, 1]^\infty$ defined as $\eta(P) = (P(B_1), P(B_2), \ldots)$. The map is one-to-one since $B$ is a field. And it is measurable since each of its coordinates is a measurable map of $P$ into $[0, 1]$. Let $S$ be the range of the map $\eta$.

From Kuratowski's theorem it then follows:

**Lemma 7.2.** $S$ is a Borel subset of $[0, 1]^\infty$, and the map $\eta$ from $P$ onto $S$ is one-to-one and bimeasurable.

For the remainder of this section we need only to assume $A$ is countably generated and contains the single point sets. Let $B$ be a countable field that generates $A$.

The following result is a preliminary to Prop. 7.1.

**Lemma 7.3.** Let $P$ be a probability measure on $(X, A)$. Then, for every $x \in X$ we have

$$\inf_{n: x \in B_n} P(B_n) = P\{x\}.$$

**Proof.** Let $C_1, C_2, \ldots$ be an enumeration of sets in $B$ that contain $x$. Then, $n \to C_n = \{x\}$. Defining $D_n = C_1 C_2 \ldots C_n$, we have $P(D_n) = P(\{x\})$. \hfill $\Box$

Consider the set of all pairs $(P, x)$ such that $P \in P$ and $x \in X$; that is, the product space $P \times X$. Equip this with product $\sigma$-field $\sigma(P) \times A$.

Let $E = \{(P, x): P(\{x\}) > 0\}$. The following result is useful.

**Proposition 7.1.** $E \in \sigma(P) \times A$. 

Proof. It suffices to show that the map \((P, x) \mapsto P(\{x\})\) from \(P \times X\) into \([0, 1]\) is \(\sigma(P) \times \mathcal{A}\)-measurable. Consider the map \(H : \mathbb{R}_2 \to \mathbb{R}_1\) defined as
\[
H(a, b) = \begin{cases} 
a, & \text{if } b \neq 0 \\
1, & \text{if } b = 0
\end{cases}
\]
Now, the map \(P \mapsto P(B_n)\) is \(\sigma(P)\)-measurable \(\forall n\), and the map \(x \mapsto I_{B_n}(x)\) is \(\mathcal{A}\)-measurable \(\forall n\). Also, observe that \(H\) is a measurable map from \(\mathbb{R}_2\) into \(\mathbb{R}_1\). Therefore, the map \((P, x) \mapsto H(P(B_n), I_{B_n}(x))\) is \(\sigma(P) \times \mathcal{A}\)-measurable \(\forall n\), and so is \(\inf_n H(P(B_n), I_{B_n}(x))\). Note that \(\inf_n H(P(B_n), I_{B_n}(x)) = \inf_{\mathbb{R}_2 : x \in B_n} P(B_n) = P(\{x\})\), where the last equality follows from Lemma 7.3. Thus the map \((P, x) \mapsto P(\{x\})\) is \(\sigma(P) \times \mathcal{A}\)-measurable. \(\|\)

For \(P \in \mathcal{P}\), let \(E_P = \{x : P(\{x\}) > 0\}\) be the \(P\)-section of \(E\). \(E_P\) is the discrete mass points of \(P\). Also, if \(P\) is a discrete probability measure, then \(E_P\) is the support of \(P\).

We have the following:

**Position 7.2.** The map \(\psi : \mathcal{P} \to [0, 1]\) defined as \(\psi(P) = P(E_P)\), the discrete mass of \(P\), is \(\sigma(P)\)-measurable.

**Proof.** Observe that \(\forall P \in \mathcal{P}\), the set \(E_P \in \mathcal{A}\) since it is countable and the single point set \(\{x\} \in \mathcal{A} \forall x \in X\). Thus the map \(P \mapsto P(E_P)\) is \(\sigma(P)\)-measurable. \(\|\)

**Corollary 7.1.** The class \(\mathcal{P}_0\) of all discrete probability measures on \((X, \mathcal{A})\) is \(\sigma(P)\)-measurable.

**Proof.** Observe that \(\mathcal{P}_0 = \{P \in \mathcal{P} : \psi(P) = 1\}\). \(\|\)
8. Existence of a Dirichlet Process

We proceed to prove the existence of a dirichlet process $D^\alpha$ on a Borel space $(X, \mathcal{A})$ corresponding to any finite measure $\alpha$ on $A$. Choose and fix a countable field $\mathcal{B} = \{B_1, B_2, \ldots\}$ of sets in $X$ such that $\mathcal{B}$ is a generator of the $\sigma$-field $\mathcal{A}$. The map $x \rightarrow \xi(x) = \{\xi_1(x), \xi_2(x), \ldots\}$, where $\xi_n$ is the indicator of $B_n$, is then a one-one bimeasurable map of $(X, \mathcal{A})$ into $(T, T)$. A probability measure $Q$ on $(T, T)$ defines a probability measure $P = Q^\xi$ on $(X, \mathcal{A})$ provided $Q[\xi(X)] = 1$.

To simplify our notations we denote a typical point $(t_1, t_2, \ldots, t_n)$ of the product space $T_n = \{0, 1\}^n$ by $s_n$. By $s_n^o$ we denote the point in $T_{n+1}$ that is obtained by augmenting $s_n$ by $o$; that is, $s_n^o = (t_1, t_2, \ldots, t_n, o)$, and similarly for $s_n^1$. Finally, we denote by $[s_n]$ the cylinder set of all points in $T$ whose first $n$ coordinates form the vector $s_n$. For example, $[o]$ is the set of all $t \in T$ such that $t_1 = 0$.

It is easily seen that a probability measure $Q$ on $(T, T)$ is uniquely defined by a double sequence $\omega$ of numbers in the closed unit interval $[0, 1]$:

$$(8.1) \quad \omega = \{u, (u_0, u_1), (u_{oo}, u_{ol}, u_{lo}, u_{ll}), \ldots\},$$

where $u = Q([1])$, $u_0 = Q([o,1]|[o])$, $u_1 = Q([1,1]|[1])$, $u_{oo} = Q([o,o,1]|[o,o])$ and so on, a typical term of the $(n + 1)^{th}$ block of the double sequence being $u_n = Q([s_n,1]|[s_n])$, $s_n \in T_n$.

Let $\Omega$ denote the space of all double sequences $\omega$ with its coordinates lying in $[0, 1]$. The probability measure on $(T, T)$ that coexists with each $\omega \in \Omega$ is denoted by $Q_\omega$. If $\mathcal{C}$ is equipped with the product $\sigma$-field $\sigma(\Omega)$, then the
map \( \omega \mapsto Q_{\omega} \) defines a transition function from \((\Omega, \sigma(\Omega))\) to \((T, T)\). If \((\Omega, \sigma(\Omega))\) is equipped with a probability measure \(\mu\), then we have a random probability measure on \((T, T)\) which we denote by \(Q_{\mu}\). How do we choose \(\mu\) so that \(P_{\mu} = Q_{\mu} \xi\) is a Dirichlet process on \((X, \mathcal{A})\) with parameter \(\alpha\)?

For an arbitrary but fixed \(n\), consider the partition \(\{B_n^s : s \in T^n\}\) of \(X\), where by \(B_n^s\) we denote the set \(B_1^{s_1}B_2^{s_2} \cdots B_n^{s_n}\). If \(P_{\mu}\) is \(D^n_\alpha\) on \((X, \mathcal{A})\), then the joint distribution of the \(2^n\) r.v.'s \(\{P_{\mu}(B_n^s) : s \in T^n\}\) is singular Dirichlet with parameters \(\{\alpha(B_n^s) : s \in T^n\}\). Invoking Proposition 4.4 we then have:

\[
(8.2) \quad P_{\mu}(B_1), \Gamma_{\mu}(B_2 | B_1^O), P_{\mu}(B_2 | B_1), P_{\mu}(B_3 | B_1^O B_2^O), \ldots
\]

are mutually independent random variables with

\[
(8.3) \quad P_{\mu}(B_1) \sim \text{Beta}(\alpha(B_1), \alpha(B_1^O))\text{ and } P_{\mu}(B_m | B_{m-1}^O) \sim \text{Beta}(\alpha(B_m), \alpha(B_m^O)), s_m \in T_m, m = 1, 2, \ldots, n.
\]

Observe that the map \(\xi : X \mapsto T\) transforms \(B_1\) to \([1]\), \(B_m\) to \([s_m]\), \(P_{\mu}(B_1)\) to \(\cap_{\mu}(1)\), and so on. It is, therefore, clear that if under \(\mu\) the coordinates of \(\omega\) are mutually independent and are distributed as

\[
(8.4) \quad u_1 \sim \text{Beta}(\alpha(B_1), \alpha(B_1^O))
\]

\[
(8.5) \quad u_{s_n} \sim \text{Beta}(\alpha(B_n), \alpha(B_n^O)), n = 1, 2, \ldots,
\]

then (8.2) and (8.3) hold true for all \(n\).
Theorem 8.1. (Blackwell (1973)): If under $\mu$ coordinates of $\omega$ are mutually independent and (8.4) holds, then $P_\mu$ is $D^\alpha$ on $(X, A)$.

Proof: Since (8.2) and (8.3) hold, it follows (Remark 4.3) that $P_\mu(B_n) \sim Beta(\alpha(B_n), \alpha(\emptyset_n))$ for all $n$. Since $B$ is a field, $X = B_n$ for some $n$. Therefore, $P_\mu(X) \sim Beta(\alpha(X), \alpha)$, that is, $P_\mu(X) = 1$ a.s.[$\mu$]. This proves that $P_\mu$ is a random probability measure on $(X, A)$.

To prove that $\tilde{i}_\mu(A) \sim Beta(\alpha(A), \alpha(A^C))$ for each $A \in A$ we proceed as follows. The map $P \rightarrow (P(B_1), P(B_2), \ldots)$ from $P$ to $S$ is one-one and bimeasurable (Lemma 7.2). For any $A \in A$, the map $P \rightarrow P(A)$ from $P$ to $[0, 1]$ is measurable. Hence there exists a measurable map $h_A: S \rightarrow [0, 1]$ such that $h_A(P(B_1), P(B_2), \ldots) = P(A)$ for all $P \in P$. For each $n$, the joint distribution of $\tilde{i}_\mu(B_1), \tilde{i}_\mu(B_2), \ldots, \tilde{i}_\mu(B_n)$ is well defined in terms $\mu$. And for different $n$ these joint distributions are mutually consistent. The Kolmogrov Extension Theorem, therefore, guarantees that the joint distribution of the whole sequence $\tilde{i}_\mu(B_1), \tilde{i}_\mu(B_2), \ldots$ is well defined. If we denote this joint distribution by $\Pi_\mu$, then $\tilde{i}_\mu(A) \sim \Pi_\mu h_A^{-1}$.

Consider now the hypothetical situation where we might have started with $\tilde{\beta} = \{A, B_1, B_2, \ldots\}$ as the generator of $A$. Proceeding as before, we would then have defined a random probability measure $P_\mu$ on $(X, A)$. For this random probability measure, it is clear that $P_\mu(A) \sim Beta(\alpha(A), \alpha(A^C))$ and $(P_\mu(B_1), P_\mu(B_2), \ldots) \sim \Pi_\mu$. Therefore, $\Pi_\mu h_A^{-1}$ is $Beta(\alpha(A), \alpha(A^C))$. This proves that $i_\mu(A) \sim Beta(\alpha(A), \alpha(A^C))$ for all $A \in A$.

The above argument goes through, word for word, for an arbitrary measurable partition $(A_1, A_2, \ldots, A_k)$ of $X$ leading to the conclusion

$$(P_\mu(A_1), P_\mu(A_2), \ldots, P_\mu(A_k)) \sim D_S(\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_k)).$$

This proves that $P_\mu$ is $D^\alpha$ on $(X, A)$.

The existence theorem of the previous section may be restated as:

Theorem 9.1: If \((X, \mathcal{A})\) is a Borel space, then, for each finite measure \(\alpha\) on \((X, \mathcal{A})\), there exists a probability measure \(D^{\alpha}\) on \((P, \sigma(P))\) such that,

\[
P \sim D^{\alpha}, (\mathcal{P}(A_1), \mathcal{P}(A_2), \ldots, \mathcal{P}(A_k)) \sim D_{\mathcal{S}}(\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_k))
\]

for any measurable partition \((A_1, A_2, \ldots, A_k)\) of \(X\).

Let \(P_o\) be the family of discrete probability measure on \((X, \mathcal{A})\). That \(P_o\) belongs to \(\sigma(P)\) has been noted in Corollary 7.1.

Theorem 9.2: If \(P \sim D^{\alpha}\), then almost every realization of \(P\) is a discrete probability measure on \((X, \mathcal{A})\), that is,

\[
D^{\alpha}(P_o) = 1
\]

Historical Note: Ferguson (1973) gave a rather involved argument to prove this result. Blackwell (1973), and Blackwell and MacQueen (1973) gave alternative arguments for the same result. The proof given here is a streamlined and mathematically rigorous version of an heuristic argument given by Berk and Savage (1975).

Consider the pair \((P, X)\) of random entities such that (i) \(P \sim D^{\alpha}\) and

(ii) \(X|P \sim P\), that is, conditional on \(P = p\), the probability distribution of \(X\) on \((X, \mathcal{A})\) is \(P\). Let \(\Delta^{\alpha}\) denote the joint distribution of \((P, X)\) on the product space \((P \times X, \sigma(P) \times \mathcal{A})\). The marginal distribution of \(X\) is then easily verified to be the normalized measure \(\overline{\alpha} = \alpha/\alpha(X)\). It is well known (see Ferguson (1973)) that \(P|X = x \sim D^{\alpha} x\), where \(\delta\) denotes the degenerate probability measure with its whole mass concentrated at \(x\).

We have noted earlier (Proposition 7.1) that \(E = \{(P, x): P(\{x\}) > 0\}\) belongs to \(\sigma(P) \times \mathcal{A}\). The following proposition is a preliminary to the proof of Theorem 9.2.

Proposition 9.1: \(\Delta^{\alpha}(E) = 1\).
Proof: Writing $E^x$ for the $x$-section of $E$, we have

$$\Delta^\alpha(E) = \int \hat{\alpha}(E^x) d\tilde{\alpha}(x)$$

Now, $E^x$ is the set of all $P \in \mathcal{F}$ such that $P\{\{x\}\} > 0$. Under the distribution $D^{\alpha+\delta}_x$, the random variable $P(\{x\})$ is positive with probability one – this is because the $\alpha + \delta$ measure of the set $\{x\}$ is positive. Therefore,

$$\int_D x(E^x) = \int_D \delta x \{ P : P\{x\} > 0 \} = 1$$

for all $x$. In other words $\Delta^\alpha(E) = 1$. \[\|\]

**Proof of Theorem 9.1.** Consider now the $P$-section $E_P = \{x : P(\{x\}) > 0\}$ of the set $E$.

Since $X$, given $P = P$ is distributed as $P$, we have

$$\Delta^\alpha(E) = \int \psi(P) d\psi^\alpha(P)$$

where $\psi(P) = P(E_P)$ is the discrete mass of $P$.

Since $\Delta^\alpha(E) = 1$, we at once have $\psi(P) = 1$ a.s. $[\mathcal{D}^\alpha]$. But $\{P : \psi(P) = 1\} = \mathcal{P}_0$.

This completes the proof of Theorem 9.2. \[\|\]

Let $\mathcal{V} = \{V\}$ be the collection of all open sets in $X$. Since $X$ is a separable metric space, there exists a countable subcollection $\{V_1, V_2, \ldots\}$ of open sets such that every $V$ contains some $V_n$. Let $P'$ be the collection of all $P \in \mathcal{F}$ such that $P(V) > 0$ for all $V \in \mathcal{V}$. Similarly, let $P_n = \{P : P(V_n) > 0\}$.

It is then clear that $P' = \bigcap_{n=1}^{\infty} P_n$.

**Theorem 9.3:** If $\alpha(V) > 0$ for all $V \in \mathcal{V}$ then $\mathcal{D}^\alpha(P') = 1$.

**Proof:** Since $P(V_n) \sim \text{Beta}(\alpha(V_n), \alpha(V_n))$ and $\alpha(V_n) > 0$ it follows that $P(V_n) > 0$ a.s. $[\mathcal{D}^\alpha]$, that is, $\mathcal{D}^\alpha(P_n) = 1$. Therefore, $\mathcal{D}^\alpha(P') = 1$. \[\|\]

The set $P_0 \cap P'$ is the collection of all discrete probability measure $P$ on $(X, \mathcal{A})$ such that the mass points of $P$ are everywhere dense in $X$. Putting Theorems (9.2) and (9.3) together we finally have:
Theorem 9.4: If $P \sim \delta^f$ and the $\alpha$-measure of every open subset of $X$ is positive, then for almost every realization $P$ of $P$ it is true that $P$ is discrete with its mass points everywhere dense in $X$.

Further properties of the Dirichlet process will be discussed in a forthcoming note.

Acknowledgement: The authors are greatly indebted to Professor David Blackwell for the benefit of some prolonged discussions and consultations during the Fall and Winter Quarters of 1978-79 when he was visiting the Florida State University.

References


