CONTRIBUTIONS TO THE THEORY
OF DIRICHLET PROCESSES

by

Ramesh M. Korwar and Myles Hollander

FSU Statistics Report M221
ONR Technical Report No. 63
AFOSR Technical Report No. 3

October, 1971
The Florida State University
Department of Statistics
Tallahassee, Florida

1Research sponsored by the Office of Naval Research under Contract NONR-988(08)
NR 042-004 and the Air Force Office of Scientific Research, AFSC, USAF, under
Grant No. AFOSR-71-2058. The United States Government is authorized to
reproduce and distribute reprints for governmental purposes notwithstanding
any copyright notation hereon.
ABSTRACT

Contributions to the Theory of Dirichlet Processes

by

Ramesh M. Korwar and Myles Hollander

The Florida State University

We derive some basic properties of a sample $X_1,\ldots,X_n$ from a Dirichlet process. Let $r_i = 0$ if $X_i = X_k$ for some $k = 1, \ldots, i-1, \text{ and } 1$ otherwise. We first establish i) the distribution of $\sum_{i=1}^{n} r_i$, the number of distinct observations in the sample, and ii) certain conditional and unconditional joint distributions of the $X_i$'s and $r_i$'s. These results are used to prove a weak law of large numbers for $Z_n = \{ \sum_{i=1}^{n} r_i X_i \} / \sum_{i=1}^{n} r_i$. The weak law is then applied to obtain the consistency of a Bayes estimator of the index of the transition measure of a mixture of Dirichlet processes.
CONTRIBUTIONS TO THE THEORY OF DIRICHLET PROCESSES

by Ramesh M. Korwar and Myles Hollander

The Florida State University

0. Introduction and summary. Ferguson (1969) has introduced the Dirichlet process (Definition 1.2) for generating random distribution functions. He uses the process as a prior on a set of probability measures in order to consider certain nonparametric problems from a Bayesian approach. Antoniak (1969) extended Ferguson's work by introducing mixtures of Dirichlet processes (Definition 1.5). These mixtures enabled Antoniak to consider certain problems that could not be treated by the Dirichlet process. The main contributions of this paper are:

(1) In Section 2 we derive some basic properties of a sample $X_1, \ldots, X_n$ (Definition 1.3) from a Dirichlet process. Since the distribution chosen by a Dirichlet process is discrete with probability one (cf. Ferguson (1969), Blackwell (1969)), the sample values need not be distinct.

Let

$$r_i = \begin{cases} 0 & \text{if } X_i = X_k \text{ for some } k = 1, \ldots, i - 1 \\ 1 & \text{otherwise.} \end{cases}$$

(0.1)

In Proposition 2.1 we give the distribution of $r = \sum_{i=1}^{n} r_i$, the number of distinct observations in the sample. Propositions 2.3 and 2.4 give results

---

1Research supported by the Office of Naval Research Contract No. NONR-988(08) Task Order NR 042-004 and by Air Force of Scientific Research Grant AFOSR-71-2058.
concerning certain conditional and unconditional distributions of the $r_i$'s and $X_i$'s. In Theorem 2.7 we establish the joint unconditional distribution of $X_1, \ldots, X_n$, generalizing a result (see Theorem 1.1) of Ferguson.

(2) Theorem 2.8 gives a weak law of large numbers for samples from a Dirichlet process. Let $Z_n = \{ \frac{\sum_{i=1}^{n} r_i X_i}{\sum_{i=1}^{n} r_i} \}$, the average of the distinct observations. Under certain restrictions on the parameter $\alpha$ and the space $(X, \mathcal{A})$ of the Dirichlet process, we derive an "in-probability" result for the sequence $\{Z_n\}$.

(3) Antoniak has derived a Bayes estimator for the index (see Definitions 1.4 and 1.5) of the transition measure of a mixture of Dirichlet processes. In Section 3 we generalize his univariate result to $k$ dimensions. We then apply the weak law to obtain the consistency of this Bayes estimator, and in particular, the consistency of Antoniak's estimator. (Antoniak did not discuss consistency.)

1. Preliminaries. In this section we list some basic definitions and results that will be used in the sequel.

**Definition 1.1** (Ferguson) Let $Z_1, \ldots, Z_k$ be independent random variables with $Z_j$ having a gamma distribution with shape parameter $\alpha_j \geq 0$ and scale parameter $1$, $j = 1, \ldots, k$. Let $\alpha_j > 0$ for some $j$. The Dirichlet distribution with parameter $(\alpha_1, \ldots, \alpha_k)$, denoted by $\mathcal{D}(\alpha_1, \ldots, \alpha_k)$, is defined as the distribution of $(Y_1, \ldots, Y_k)$, where $Y_j = Z_j / \sum_{i=1}^{k} Z_i$, $j = 1, \ldots, k$. 

\[ C_1, \ldots, C_n. \]

\[ Q\{X_1 \in C_1, \ldots, X_n \in C_n | P(A_1), \ldots, P(A_m), P(C_1), \ldots, P(C_n) \} = \prod_{i=1}^n P(C_i) \text{ a.s.}, \]

where \( Q \) denotes probability.

We may view a sample of size \( n \) from a Dirichlet process as follows. The process chooses a random distribution \( F \) say, and then given \( F \), \( X_1, \ldots, X_n \) is a random sample from \( F \).

Definitions 1.4-1.5 define a stochastic process \( \{P(A), A \in A\} \) which may be viewed as a Dirichlet process where the parameter \( \alpha \) is random.

**DEFINITION 1.4 (Antoniak).** Let \( (X, A) \) and \( (V, B) \) be two measurable spaces. A transition measure \( \alpha \) is a mapping of \( V \times A \) into \([0, \infty)\) such that

(a) for every \( y \in V \), \( \alpha(y, \cdot) \) is a finite, non-negative, non-null measure on \( (X, A) \); and

(b) for every \( A \in A \), \( \alpha(\cdot, A) \) is measurable on \( (V, B) \).

**DEFINITION 1.5 (Antoniak).** Let \( (X, A) \) be a measurable space, let \( (V, B, H) \) be a probability space, called the index space, and let \( \alpha \) be a transition measure on \( V \times A \). We say \( F \) is a mixture of Dirichlet processes on \( (X, A) \) with mixing distribution \( H \) on the index space \( (V, B) \), and transition measure \( \alpha \) on \( V \times A \) if, for all \( k = 1, 2, \ldots \) and any measurable partition \( A_1, \ldots, A_k \) of \( X \), we have

\[ Q\{P(A_1) \leq z_1, \ldots, P(A_k) \leq z_k \} = \int \left( D(z_1, \ldots, z_k | \alpha(y, A_1), \ldots, \alpha(y, A_k)) \right) dH(y), \]

where \( D(\cdot | \beta_1, \ldots, \beta_k) \) denotes the distribution function of a Dirichlet...
distribution with parameter \((\beta_1, \ldots, \beta_k)\).

The next definition is that of a sample from a mixture of Dirichlet processes. (Compare Definition 1.6 with Definition 1.3.)

**Definition 1.6** (Antoniak). Let \(P\) be a mixture of Dirichlet processes on \((X, \mathcal{A})\) with mixing distribution \(\mathbb{H}\) on the index space \((Y, \mathcal{B})\) and transition measure \(\alpha\) on \(Y \times \mathcal{A}\). We say that the \(X\)-valued random variables \(X_1, \ldots, X_n\) constitute a sample of size \(n\) from the mixture \(P\) if for any \(m = 1, 2, \ldots\) and measurable sets \(A_1, \ldots, A_m, C_1, \ldots, C_n\) we have

\[
Q(X_1 \in C_1, \ldots, X_n \in C_n \mid Y, P(A_1), \ldots, P(A_m), P(C_1), \ldots, P(C_n)) = \prod_{i=1}^{n} P(C_i) \quad \text{a.s.}
\]

Theorem 1.1 gives the distribution of a sample of size 1 from a Dirichlet process. We generalize this theorem in Section 2 (see Theorem 2.7).

**Theorem 1.1** (Ferguson). Let \(P\) be a Dirichlet process on \((X, \mathcal{A})\) with parameter \(\alpha\), and let \(X\) be a sample of size 1 from \(P\). Then for \(A \in \mathcal{A}\),

\[
Q(X \in A) = \frac{\alpha(A)}{\alpha(X)}.
\]

Theorem 1.2 presents the conditional distribution of a Dirichlet process, given a sample from it. The conditional distribution is also Dirichlet but with a different parameter.

**Theorem 1.2** (Ferguson). Let \(P\) be a Dirichlet process on \((X, \mathcal{A})\) with parameter \(\alpha\), and let \(X_1, \ldots, X_n\) be a sample of size \(n\) from \(P\). Then the conditional distribution of \(P\) given \(X_1, \ldots, X_n\) is a Dirichlet process on
\( (X,A) \) with parameter \( \beta = \alpha + \sum_{i=1}^{n} \delta_{X_i} \), where, for \( x \in X, A \subseteq A \)

\[
\delta_x(A) = \begin{cases} 
1 & \text{if } x \in A \\
0 & \text{if } x \notin A.
\end{cases}
\]

Recall Definitions 1.4, 1.5 on the mixture of Dirichlet processes. If we regard \( Y \), the index of the transition measure \( \alpha \), as a random variable with prior distribution \( H \), then Theorem 1.3 gives the posterior distribution of \( Y \), given a sample from the mixture \( P \) of Dirichlet processes. Theorem 1.3 will be used in Section 3 to derive a Bayes estimator of the index of the transition measure of a mixture of Dirichlet processes. In Theorem 1.3, and in the sequel, we find it necessary to restrict various spaces to be standard Borel spaces so that certain conditional distributions exist.

**DEFINITION 1.7** (cf. Parthasarathy (1967)). A **standard Borel space** is a measurable space \( (X,A) \) in which \( A \) is countably generated, and for which there exists a bi-measurable mapping between \( (X,A) \) and some complete separable metric space \( (Z,C) \).

**THEOREM 1.3** (Antoniak). Let \( P \) be a mixture of Dirichlet processes on a standard Borel space \( (X,A) \) with standard Borel index space \( (Y,B) \) and transition measure \( \alpha \) on \( Y \times A \). Let \( \bar{X} = (X_1, \ldots, X_n) \) be a sample of size \( n \) from \( P \), and suppose there exists a \( \sigma \)-finite \( \sigma \)-additive measure \( \mu \) on \( (X,A) \) such that for each \( y \in Y \)

(i) \( \alpha_y = \alpha(y, \cdot) \) is \( \sigma \)-additive and absolutely continuous with respect to \( \mu \), and
(ii) The measure $\mu$ has mass one at each atom of $\alpha_y$. Then, the conditional distribution $H_{X|Y}(y)$ of $Y$ given $X$ is given by

$$dH_{X|Y}(y) = \frac{1}{M_y^n} \prod_{i=1}^{r} \frac{\alpha'_y(X_i')(m_y(X_i') + 1)}{\prod_{i=1}^{n(X_i')-1} dH(y)}$$

where $\alpha'_y(\cdot)$ is the Radon-Nikodym derivative of $\alpha_y(\cdot)$ with respect to $\mu$; $X_i'(X)$ is the $i$-th distinct value of $X$ in $X$; $n(X_i')$ is the number of times the value $X_i'$ occurs in $X$; $M_y = \alpha_y(X)$; $m_y(X_i') = \alpha_y(X_i')$ if $X_i'$ is an atom of $\alpha_y$, zero otherwise.

2. Distribution theory for the $r_i$'s and $X_i$'s. Let $X_1, \ldots, X_n$ be a sample of size $n$ from a Dirichlet process on a standard Borel space $(X, \mathcal{A})$ with parameter $\alpha$. Consider the $r_i$'s defined by (0.1) and set $r = \sum_{i=1}^{n} r_i$, the number of distinct observations in the sample.

Propositions 2.1-2.3 will be used to establish Theorem 2.6, a weak law of large numbers for samples from a Dirichlet process.

PROPOSITION 2.1. Assume $\alpha$ is non-atomic. Then the $\{r_i\}$ are independent Bernoulli random variables with success probabilities $p_i = \alpha(X)/(\alpha(X) + i - 1)$. The random variable $r$ is thus generalized Binomial with parameters $(n, p_1, \ldots, p_n)$.

PROOF. We first show that

$$p_i = \mathbb{P}(r_i = 1) = \frac{\alpha(X)}{\alpha(X) + i - 1}, \quad i = 1, \ldots, n,$$
and then show that the \( r_i \)'s are independent. Now,

\[
Q(r_i = 1) = EQ(r_i = 1 | X_1 = x_1, \ldots, X_{i-1} = x_{i-1})
\]

\[
= EQ(X_i \in X \setminus \{x_1, \ldots, x_{i-1}\} | X_1 = x_1, \ldots, X_{i-1} = x_{i-1})
\]

\[
= E[a(X \setminus \{x_1, \ldots, x_{i-1}\}) + \sum_{k=1}^{i-1} \delta_{x_k}(X \setminus \{x_1, \ldots, x_{i-1}\})]/\{a(X) + i - 1\}.
\]

Here \( A - B \) denotes \( A \cap B^c \), and the last step follows from Theorems 1.2 and 1.1.

Thus, since \( a \) is non-atomic,

\[
Q(r_i = 1) = E[a(X)/\{a(X) + i - 1\}] = a(X)/\{a(X) + i - 1\},
\]

establishing (2.1). To show \( r_1, \ldots, r_n \) are independent, it is sufficient to show that

\[
(2.2) \quad Q(r_i = 1 | r_{i_1}, \ldots, r_{i_k}) = Q(r_i = 1) \quad a.s.
\]

for any \( 1 \leq k \leq n \) and \( 1 \leq i_1 < \ldots < i_k \leq n \). Now, for \( i_k > 1 \),

\[
(2.3) \quad Q(r_i = 1 | r_{i_1}, \ldots, r_{i_k}) = E[Q(r_i = 1 | X_{i_1} = x_{i_1}, j = 1, \ldots, i_{k-1}) | r_{i_1}, \ldots, r_{i_k}] \quad a.s.,
\]

since the \( \sigma \)-field \( \sigma(r_{i_1}, \ldots, r_{i_k}) \) generated by \( r_{i_1}, \ldots, r_{i_k} \) is a sub-\( \sigma \)-field of the \( \sigma \)-field \( \sigma(X_{i_1}, \ldots, X_{i_k}) \) generated by \( X_{i_1}, \ldots, X_{i_k} \). Thus,
\[ Q(r_{i_k} = 1 | X_j = x_j, j = 1, \ldots, i_k - 1) \]

\[ = Q(x_{i_k} \in X - \{x_1, \ldots, x_{i_k - 1}\} | X_j = x_j, j = 1, \ldots, i_k - 1) \]

\[ = \alpha(X - \{x_1, \ldots, x_{i_k - 1}\}) + \sum_{j=1}^{i_k-1} \delta_{X_j} = (X - \{x_1, \ldots, x_{i_k - 1}\})/(\alpha(X) + i_k - 1) \text{ a.s.,} \]

by Theorems 1.2 and 1.1. Now, in (2.4), condition further on \( r_{i_1}, \ldots, r_{i_{k-1}} \), use the fact that \( \alpha \) is non-atomic, and apply (2.3) to find

\[ Q(r_{i_k} = 1 | r_{i_1}, \ldots, r_{i_{k-1}}) = \alpha(X)/(\alpha(X) + i_k - 1) \text{ a.s.} \]

Equation (2.2) is a direct consequence of (2.5) and (2.1). The fact that \( r \) has a generalized Binomial distribution with the specified parameters follows from (2.1) and the independence of the \( r_i \)'s.

**PROPOSITION 2.2.** Assume \( \alpha \) is non-atomic. Then \( \text{p-lim } r^{-1} = 0. \)

**PROOF.** Consider the variables \( r/\log n, n \geq 2 \). We have, by Proposition 2.1,

\[ 0 < E(r/\log n) = [\log n]^{-1} \sum_{i=1}^{n} \alpha(X)/(\alpha(X) + i - 1). \]

Thus,
\[ 0 < \frac{E(r/\log n)}{\log n} < (\log n)^{-1} + (\log n)^{-1} \alpha(X) \sum_{i=1}^{n-1} i^{-1} \]

(2.6) \[ < (\log n)^{-1} + (\log n)^{-1} \alpha(X) \left( \sum_{i=1}^{n} i^{-1} \log n \right) + \alpha(X) \]

\[ \rightarrow 0 + 0 \cdot \alpha(X) \cdot \gamma + \alpha(X) = \alpha(X), \]

where \( \gamma \) is Euler's constant. Also, by Proposition 2.1, we find

(2.7) \[ \text{Var}(r/\log n) = (\log n)^{-2} \left\{ \sum_{i=1}^{n} \frac{\alpha(X)}{\alpha(X) + i - 1} - \sum_{i=1}^{n} \frac{\alpha^2(X)}{(\alpha(X) + i - 1)^2} \right\}. \]

Now,

(2.8) \[ 0 < (\log n)^{-2} \sum_{i=1}^{n} \frac{\alpha(X)}{\alpha(X) + i - 1} = (\log n)^{-1} E(r/\log n) \rightarrow 0, \]

using (2.6). Similarly,

(2.9) \[ 0 < (\log n)^{-2} \sum_{i=1}^{n} \frac{\alpha^2(X)}{\alpha(X) + i - 1} \leq (\log n)^{-2} \alpha^2(X) (\log n)^{-2} \sum_{i=1}^{n} i^{-2} \rightarrow 0. \]

Thus, by (2.7) - (2.9), \( \text{Var}(r/\log n) \rightarrow 0 \), and hence, by Chebychev's inequality, and (2.6), it follows that

(2.10) \[ p\text{-lim}(r/\log n) = \alpha(X). \]

The proposition follows directly from (2.10). ||
PROPOSITION 2.3. Assume $a$ is non-atomic. Then

\[(2.11) \quad Q\{X_i \in A_i | r_1 = 1, r_k = \rho_k, \rho_k = 0 \text{ or } 1, \rho_i = 1, k=2, \ldots, n\} = Q_{\{X_i \in A_i\}} \text{ a.s.,}\]

\[(2.12) \quad Q\{X_i \in A_i, X_j \in A_j | r_1 = 1, r_k = \rho_k, \rho_k = 0 \text{ or } 1, \rho_i = \rho_j = 1, k=2, \ldots, n\} = Q_{\{X_i \in A_i\}}Q_{\{X_j \in A_j\}} \text{ a.s.,}\]

\[(2.13) \quad Q\{X_i \in A_i, X_j \in A_j | r_j = 1\} = Q_{\{X_i \in A_i\}}Q_{\{X_j \in A_j\}} \text{ a.s.,}\]

where $A_i, A_j \in A$, and $j > i$. (Note that $r_1 = 1$ a.s.; in all future conditions "$r_1$" should be understood as "$|r_1 = 1$".)

PROOF. We give the proof of (2.11); (2.12) and (2.13) follow by similar arguments. We have

\[Q\{X_i \in A_i, r_i = 1 | r_1, \ldots, r_{i-1}\} = E\{Q\{X_i \in A_i, r_i = 1 | x_k, k=1, \ldots, i-1\} | r_1, \ldots, r_{i-1}\} \text{ a.s.,}\]

\[= E \left( \frac{\alpha(A_i \cap \{x_1, \ldots, x_{i-1}\}) + \sum_{k=1}^{i-1} \delta_{x_k} (A_i \cap \{x_1, \ldots, x_{i-1}\})}{\alpha(X) i-1} \right) | r_1, \ldots, r_{i-1} \text{ a.s.,}\]

the last step following from Theorems 1.2 and 1.1. Thus

\[(2.14) \quad Q\{X_i \in A_i, r_i = 1 | r_1, \ldots, r_{i-1}\} = E(\alpha(X) / (\alpha(X) i-1) | r_1, \ldots, r_{i-1}) \text{ a.s.}\]

\[= \alpha(A_i) / (\alpha(X) i-1) \text{ a.s.}\]
Again, for $k = i + 1, \ldots, n$,

$$Q\{r_k = \rho_k | r_1, \ldots, r_{i-1}, x_i, r_i, \ldots, r_{k-1}\}$$

$$= E(Q\{r_k = \rho_k | X_j = x_j, j = 1, \ldots, k-1\} | r_1, \ldots, r_{i-1}, x_i, r_i, \ldots, r_{k-1}\) \text{ a.s.}$$

$$= \begin{cases} 
\frac{\alpha(\{x_1, \ldots, x_{k-1}\}) + \sum_{j=1}^{k-1} \delta_{x_j}(\{x_1, \ldots, x_{k-1}\})}{\alpha(X+k-1)} | r_1, \ldots, r_{i-1}, x_i, r_i, \ldots, r_{k-1}\) \text{ a.s., if } \rho_k = 0, \\
\frac{\alpha(X-\{x_1, \ldots, x_{k-1}\}) + \sum_{j=1}^{k-1} \delta_{x_j}(X-\{x_1, \ldots, x_{k-1}\})}{\alpha(X+k-1)} | r_1, \ldots, r_{i-1}, x_i, r_i, \ldots, r_{k-1}\) \text{ a.s., if } \rho_k = 1
\end{cases}$$

by Theorems 1.2 and 1.1.

Thus,

$$Q\{r_k = \rho_k | r_1, \ldots, r_{i-1}, x_i, r_i, \ldots, r_{k-1}\}$$

$$= \begin{cases} 
E(\{0-1\}/(\alpha(X)+k-1)) | r_1, \ldots, r_{i-1}, x_i, r_i, \ldots, r_{k-1}\) \text{ a.s., if } \rho_k = 0 \\
E(\alpha(X)/(\alpha(X)+k-1)) | r_1, \ldots, r_{i-1}, x_i, r_i, \ldots, r_{k-1}\) \text{ a.s., if } \rho_k = 1
\end{cases}$$

(2.15)

$$= \begin{cases} 
(k-1)/\alpha(X)+k-1) \text{ a.s., if } \rho_k = 0 \\
\alpha(X)/(\alpha(X)+k-1) \text{ a.s., if } \rho_k = 1
\end{cases}$$

$$= Q\{r_k = \rho_k\} \text{ a.s.}$$
Also,

(2.16) \[ Q\{r\_k = \rho\_k, k = 1, \ldots, i-1\} = \prod_{k=1}^{i-1} Q\{r\_k = \rho\_k\}, \]

since, by Proposition 2.1, \( r_1, \ldots, r_n \) are independent.

Now, let \( C_1 = \{ r_1 = \rho_1, \ldots, r_{i-1} = \rho_{i-1}, X\_i \in A\_i, r_i = 1, r_{i+1} = \rho_{i+1}, \ldots, r_n = \rho_n\}, \)
\( C_2 = \{ r_1 = \rho_1, \ldots, r_{i-1} = \rho_{i-1}, X\_i \in A\_i, r_i = 1, r_{i+1} = \rho_{i+1}, \ldots, r_{n-1} = \rho_{n-1}\}. \)

Then,

\[ Q\{C_1\} = \int_{C_2} Q\{r\_n = \rho\_n | r_1, \ldots, r_{i-1}, X\_i, r_i, \ldots, r_{n-1}\} dQ \]
\[ = Q\{r\_n = \rho\_n\} \int_{C_2} dQ \cdot \]

Continuing in this fashion yields

\[ Q\{C_1\} = \left\{ \prod_{k=i+1}^{n} Q\{r\_k = \rho\_k\} \cdot \int_{C_3} Q\{X\_i \in A\_i, r_i = 1 | r_1, \ldots, r_{i-1}\} dQ \right\} \]
\[ = \left\{ \prod_{k=i+1}^{n} Q\{r\_k = \rho\_k\} \cdot \left\{ a(A\_i)/(a(X) + i - 1) \right\} \cdot \int_{C_3} dQ \right\}, \]

where \( C_3 = \{ r_k = \rho_k, k = 1, \ldots, i-1\} \), and the last step follows from

(2.14). Hence, by (2.16),

\[ Q\{C_1\} = \left\{ \prod_{k=i+1}^{n} Q\{r\_k = \rho\_k\} \cdot \left\{ a(A\_i)/(a(X) + i - 1) \right\} \cdot \prod_{k=1}^{i-1} Q\{r\_k = \rho\_k\} \right\}, \]

and thus

\[ Q\{C_1\} = \left\{ \prod_{k=i+1}^{n} Q\{r\_k = \rho\_k\} \cdot \left\{ a(X)/(a(X) + i - 1) \right\} \cdot \left\{ a(A\_i)/a(X) \right\} \cdot \prod_{k=1}^{i-1} Q\{r\_k = \rho\_k\} \right\}, \]

(2.17)
\[ = \left\{ \prod_{k=1}^{n} Q\{r\_k = \rho\_k\} \right\} Q\{X\_i \in A\_i\}, \]
where \( \rho_i = 1 \). Hence,

\[ Q(\xi_{1} \in A_i \mid \xi_1 = 1, \xi_k = \rho_k, \xi_k = 0 \text{ or } 1, \rho_i = 1, k=2, \ldots, n) = Q(\xi_{1} \in A_i) \quad \text{a.s.,} \]

which is (2.11).

PROPOSITION 2.4. \( X_i \) and \( r_i \) are independent.

PROOF. Sum both sides of (2.17) over all possible 0, 1 values of \( \rho_1, \ldots, \rho_{i-1}, \rho_{i+1}, \ldots, \rho_n \).

REMARK 2.5. It can be shown (see Theorem 2.7) that

\[ Q(\xi_{1} \in A_i, \xi_j \in A_j) = \frac{\alpha(A_i) + \alpha(A_j)}{\alpha(X)(1 + \alpha(X))}, \tag{2.18} \]

where \( A_i, A_j \epsilon A \). Since \( Q(\xi_k \in A_k) = \alpha(A_k)/\alpha(X), k = i, j, \) we see that \( X_i, X_j \) are not independent. But we can also show that, a.s.,

\[ Q(\xi_{1} \in A_i \mid \xi_j = 1) = Q(\xi_{1} \in A_i), \quad Q(\xi_j \in A_j \mid \xi_j = 1) = Q(\xi_j \in A_j), \tag{2.19} \]

where \( j > i, A_i, A_j \epsilon A \). Thus by (2.13) and (2.19) we have

\[ Q(\xi_{1} \in A_i, \xi_j \in A_j \mid \xi_j = 1) = Q(\xi_{1} \in A_i \mid \xi_j = 1) \cdot Q(\xi_j \in A_j \mid \xi_j = 1), \tag{2.20} \]

and hence given \( \xi_j = 1, X_i \) and \( X_j \) are (conditionally) independent.

Theorem 2.7 gives the joint distribution of a sample \( X_1, \ldots, X_n \).
from a Dirichlet process. The theorem is a generalization of Theorem 1.1 due to Ferguson. For the proof of Theorem 2.7 we utilize Lemma 2.6. The proof of Lemma 2.6 is straightforward but tedious, and hence is omitted. The Z's defined in the statement of Lemma 2.6 follow an ordered Dirichlet distribution (cf. Wilks (1962, p. 182)). Lemma 2.6 gives the mixed moments of such a distribution.

**Lemma 2.6.** Let $a_i > 0$ for some $i=1,\ldots,k+1$ and let $t_1,\ldots,t_k$ be positive integers. Let $(Y_1,\ldots,Y_{k+1})$ have the Dirichlet distribution with parameter $(a_1,\ldots,a_{k+1})$, $Z_r = \sum_{i=1}^s Y_i$, $v_s = \sum_{j=1}^s a_j$, $r=1,\ldots,k$, $s=1,\ldots,k+1$. Then,

$$E(Z_1^{t_1}\cdots Z_k^{t_k}) = \frac{v_1(v_1+1)\cdots(v_1+t_1-1)\cdots(v_k+\sum_{i=1}^k t_i)\cdots(v_k+t_k-1)}{v_{k+1}(v_{k+1}+1)\cdots(v_{k+1}+\sum_{i=1}^k t_i-1)}.$$

**Theorem 2.7.** Let $X_1,\ldots,X_n$ be a sample of size $n$ from a Dirichlet process $(\mathbb{R}, \mathcal{B})$, where $\mathbb{R}$ is the real line and $\mathcal{B}$ is the $\sigma$-field of Borel subsets of $\mathbb{R}$. Assume $\alpha$ to be $\sigma$-additive.

Then,

$$Q(X_{1,1},\ldots,X_{1,n}) = \frac{\alpha(A_{X_{1}})\cdots(\alpha(A_{X_{n}})+1)\cdots(\alpha(A_{X_{n}})+n-1)}{\alpha(\mathbb{R})\cdots(\alpha(\mathbb{R})+n-1)},$$

where $x_{(1)} \leq \cdots \leq x_{(n)}$ is an arrangement of $x_1,\ldots,x_n$ in increasing order of magnitude, and $A_x = (-\infty, x]$. 
PROOF. Observe that $A_{x(k)} \subset A_{x(k+1)}$, $k=1,\ldots,n-1$. We define events $B_1,\ldots,B_{n+1}$ by

$$A_{x(1)} = B_1,$$

$$A_{x(k)} = A_{x(1)} + A_{x(2)} + \cdots + A_{x(k-1)} - A_{x(k)} = B_1 + B_2 + \cdots + B_k, \quad k=2,\ldots,n,$$

$$A_{x(n)} = B_{n+1}.$$

Then,

$$Q(x_1 < x_1, \ldots, x_n < x_n) = Q(X_{i_1} \in A_{x(i_1)}, \ldots, X_n \in A_{x(i_n)}),$$

where $(i_1, \ldots, i_n)$ is a permutation of $(1, 2, \ldots, n)$. Thus,

$$Q(x_1 < x_1, \ldots, x_n < x_n) = E(Q(X_{i_1} \in A_{x(i_1)}, \ldots, X_n \in A_{x(i_n)}) \mid P(A_{x(i_1)}), \ldots, P(A_{x(i_n)})�)

= E(P(A_{x(i_1)}) \cdots P(A_{x(i_n)}),$$

by Definition 1.3. That is,

$$Q(x_1 < x_1, \ldots, x_n < x_n) = E(P(B_1)(P(B_1) + P(B_2)) \cdots (P(B_1) + \cdots + P(B_n))$$

$$= \frac{a(A_{x(1)})(a(A_{x(2)})+1) \cdots (a(A_{x(n)})+n-1)}{a(R)(a(R)+1) \cdots (a(R)+n-1)}.$$

THEOREM 2.6. Let \( X_1, \ldots, X_n \) be a sample of size \( n \) from a Dirichlet process on a standard Borel space \((X, \mathcal{A})\) with parameter \( \alpha \), where here \( X \) is the \( k \)-dimensional Euclidean space, \( \mathcal{A} \) is a \( \sigma \)-field of subsets of \( X \), and \( \alpha \) is assumed non-atomic and \( \sigma \)-additive. Let \( V = (\sigma_{ij}), i, j = 1, \ldots, k \) exist, where

\[
\sigma_{ij} = \frac{\int (x_i - \mu_i)(x_j - \mu_j) \text{d}a(x_1, \ldots, x_k)}{a(X)},
\]

and

\[
\mu_i = \frac{\int x_i \text{d}a(x_1, \ldots, x_k)}{a(X)},
\]

for \( i = 1, \ldots, k \). Let \( Y_i, i = 1, \ldots, r \) be the \( r \) distinct observations among \( X_i, i = 1, \ldots, n \). (If \( X_i = (x_{i1}, \ldots, x_{ik}) \) and \( X_j = (x_{j1}, \ldots, x_{jk}) \) are \( k \)-vectors, we say \( X_i = X_j \) if and only if \( X_{it} = X_{jt} \), for \( t = 1, \ldots, k \).)

Then, letting \( Z_n = \sum_{i=1}^{r} Y_i / r \), we have

\[
p\text{-lim } Z_n = \mu = (\mu_1, \ldots, \mu_k)'.
\]

PROOF. Write \( Z_n \) as \( Z_n = \sum_{i=1}^{n} r_i x_i / \sum_{i=1}^{n} r_i \) (note that \( r = \sum_{i=1}^{n} r_i > 1 \) w.p. 1) and set \( \tau_n = E(Z_n) \). Then,

\[
\tau_n = E(r_1, \ldots, r_n) E(\{ \sum_{i=1}^{n} r_i x_i / \sum_{i=1}^{n} r_i \}| r_1, \ldots, r_n) =
\]

\[
E(r^{-1} \sum_{i=1}^{n} r_i E(x_i | r_1, \ldots, r_n)) = E(r^{-1} \sum_{i=1}^{n} r_i \mu)
\]

\[
= E(r^{-1} \sum_{i=1}^{n} r_i \mu) = \mu,
\]

the last step being a consequence of Lemma 2.6.
where \( \sum \) denotes the summation over those \( i \) indices for which \( r_i = 1 \), and where the equality on the middle line of (2.23) follows from (2.11) and (2.22). Similarly,

\[
E_{n} Z_n Z_n' = \sum_{r_1, \ldots, r_n} \{ r^{-2} E( \sum_{i,j=1}^{n} r_i r_j X_i X_j' | r_1, \ldots, r_n) \}
\]

\[
= E[r^{-2} \{ \sum_{i} r_i^2 E(X_i X_i' | r_1, \ldots, r_n ) + \sum_{i \neq j} r_i r_j E(X_i X_j' | r_1, \ldots, r_n) \}]
\]

(2.24)

\[
= E(r^{-2} \{ \sum_{i=1}^{n} r_i^2 (v + uu') + \sum_{i \neq j} r_i r_j uu' \})
\]

\[
= E(\sum_{i=1}^{n} r_i^2 (v + uu') + \sum_{i \neq j} r_i r_j uu')
\]

\[
= E(uu' + r^{-1}v) = uu' + vE(r^{-1})
\]

where \( \sum \) denotes the summation over those \( (i,j) \) pairs for which \( i \neq j \) and \( r_i = r_j = 1 \), and where the equality on the middle line of (2.24)
follows from (2.12), (2.21), and (2.22). Hence,

\[(2.25) \quad E(Z_n - \tau_n)(Z_n - \tau_n)' = \lambda E(r^{-1}),\]

and thus by (2.23), (2.25), Chebychev's inequality, and Proposition 2.1, we have p-lim $Z_n = \mu$.||

The weak law of large numbers proved in Theorem 2.8 is for a Dirichlet process on $(X,A)$ when $X$ is the $k$-dimensional Euclidean space.

Since Propositions 2.1-2.3 were established for a more abstract space $X$, it may be possible to extend Theorem 2.8 to other spaces. Note also that if $X_1, \ldots, X_n$ is a sample from a Dirichlet process, then $X_1, \ldots, X_n$ are not independent although they are identically distributed.

3. Consistency of a Bayes estimator of the index of the transition measure of a mixture of Dirichlet processes. Antoniak (1969) has considered the one-dimensional version of the following problem. Recall Definitions 1.4-1.6 and let $X_1, \ldots, X_n$ be a sample of size $n$ from a mixture of Dirichlet processes on $(X,A)$ with mixing distribution $H$ on $(\mathcal{V}, \mathcal{B})$ and transition measure $\alpha$ on $\mathcal{V} \times A$. Here we take $X, \mathcal{V}$ to be $k$-dimensional Euclidean space $\mathbb{R}^k$, $A = \mathcal{B}$ is the $\sigma$-field of Lebesgue measurable subsets of $\mathbb{R}^k$, $H$ is the distribution function of a $k$-variate normal distribution with mean vector $\mu$ and covariance matrix $V$, and $\alpha$ is given by

\[(3.1) \quad \alpha_\theta = \alpha(\theta, A) = \int_A (\sqrt{2\pi})^{-k} \exp \left[ -\sum_{i=1}^k (z_i - \theta_i)^2 / 2 \right] dz_1 \ldots dz_k,\]
where $\theta = (\theta_1, \ldots, \theta_k)'$, $A \in A$. We wish to estimate $\theta$, the index of the transition measure, with squared error loss $L(\theta, \hat{\theta}) = \sum_{i=1}^{k} (\theta_i - \hat{\theta}_i)^2$, for $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)'$. We may view the situation (roughly) as follows:

First $\theta$ is chosen from a $k$-variate normal distribution with mean vector $\mu$ and covariance matrix $V$; then we sample from a Dirichlet process with parameter $a_\theta$, that is a process whose expected distribution is multivariate normal with mean $\theta$ and covariance matrix $I$. (This interpretation was offered by Antoniak for the one-dimensional case.)

We next exhibit the Bayes estimator of $\theta$. From Theorem 1.3, we find that the conditional distribution $H_X(\theta)$ of $\theta$, given the sample $X_1, \ldots, X_n$, is

$$H_X(t_1, \ldots, t_k) = C \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_k} \exp(-\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{r} (Y_{ij} - \theta_j)^2 - \frac{1}{2} (\theta - \mu)'V^{-1}(\theta - \mu)) d\theta_1 \cdots d\theta_k,$$

(3.2)

where $Y_{i1}, \ldots, Y_{ik}$, and $Y_1, \ldots, Y_r$ are the $r$ distinct observations among $X_1, \ldots, X_n$. We rewrite equation (3.2) in the form

$$H_X(t_1, \ldots, t_k) = C \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_k} \exp\left(-\frac{1}{2} \sum_{j=1}^{k} (\theta_j - \mu_j)^2 + (\theta - \mu)'V^{-1}(\theta - \mu)ight)$$

$$- 2 \sum_{j=1}^{k} (\theta_j - \mu_j)'T_j d\theta_1 \cdots d\theta_k,$$

(3.3)

where $T_j = \sum_{i=1}^{r} (Y_{ij} - \mu_j)$, $j = 1, \ldots, k$. We also rewrite equation (3.2) in the form

$$H_X(t_1, \ldots, t_k) = C \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_k} \exp\left(-\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij}(\theta_i - \mu_i - \nu_i)(\theta_j - \mu_j - \nu_j)\right) d\theta_1 \cdots d\theta_k.$$

(3.4)
Comparing coefficients of \((\theta_i - u_i), (\theta_i - u_i)(\theta_j - u_j),\) and \((\theta_i - u_i)^2\) in the exponents of equations (3.3) and (3.4), we find

\[
\begin{align*}
a_{ii} &= c_{ii} + r, \quad i = 1, \ldots, k, \\
\sum_{j=1}^{k} a_{ij} v_j &= T_i, \quad i = 1, \ldots, k,
\end{align*}
\]

where \(V^{-1} = (c_{ij})\). Thus we obtain

\[
(V^{-1} + rI)v = T,
\]

where \(v = (v_1, \ldots, v_k)', T = (T_1, \ldots, T_k)',\) and \(I\) is the identity matrix of order \(k\). Solving (3.6) for \(v\) yields

\[
v = (I + r^{-1} V^{-1})^{-1} T / r.
\]

(Note that \((I + r^{-1} V^{-1})^{-1}\) exists, since otherwise \(-r\) would be a characteristic root of \(V^{-1}\), contradicting the positive definiteness of \(V^{-1}\).) The Bayes estimator of \(\theta\) is the mean of the distribution given by (3.4), namely

\[
\hat{\theta}_n = \mu + (I + r^{-1} V^{-1})^{-1} T / r.
\]

For \(k = 2\), \(\hat{\theta}_n\) reduces to

\[
\hat{\theta}_n = \mu + (I + r^{-1} V^{-1})^{-1} T / r.
\]
\( \hat{\theta}_i(n) = \mu_i + M^{-1}(r + [\sigma_i^2, (1-\rho^2)]^{-1}) \sum_{j=1}^{r} (Y_{ij} - \mu_i) + M^{-1} \rho [\sigma_1 \sigma_2 (1-\rho^2)]^{-1} \sum_{j=1}^{r} (Y_{ij} - \mu_1), \)

where

\[ M = M(r, \sigma_1, \sigma_2, \rho) = r^2 + [\sigma_2^2 (1-\rho^2)]^{-1} (r \sigma_1^2 + \sigma_2^2 + 1), \]

\[ V = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \]

and \((i, i')\) can be either \((1,2)\) or \((2,1)\). For \(k = 1\), \(\hat{\theta}_n\) reduces to

\( \hat{\theta}_1(n) = \mu_1 + (1 + r^{-1} \sigma_1^{-2})^{-1} (\sum_{i=1}^{r} Y_{i1} - \mu_1)/r, \)

which agrees with Antoniak's result for \(k = 1\). (Antoniak took the mean of \(H\) to be \(\mu_1 = 0\).)

We now establish the consistency of \(\hat{\theta}_n\) (3.9).

**THEOREM 3.1.** Assume \(\alpha\) is non-atomic. Then \(p\)-\(\lim\) \(\hat{\theta}_n = \theta\).

**PROOF.** Note that by Definitions 1.2-1.6, we have that, given \(\theta\), \(X_1, \ldots, X_n\) is a sample of size \(n\) from a Dirichlet process on \((X, A)\) with parameter \(\alpha_\theta\) given by (3.1). Hence, given \(\theta\), we may apply the "in-probability" results obtained in Proposition 2.2 and Theorem 2.8.

We first show

\( \text{(3.11)} \quad p\text{-lim} (I + r^{-1} V^{-1})^{-1} = I, \)
Let \((b_{ij}) = (I + r^{-1} v^{-1})^{-1}\). Then, by the definition of the inverse of a matrix, \(b_{ij}\) is the ratio of two determinants and hence is a ratio of two polynomials in \((1/r)\). By Proposition 2.2, \(r^{-1}\) (and hence any positive power of \(r^{-1}\)) converges in probability to 0 as \(n \to \infty\). Thus \(p\lim b_{ij} = c_{ij}\) where \(c_{ij}\) is the ratio of the terms free from \(r^{-1}\) in these polynomials. But this ratio is precisely the \((i,j)\)-th element of \(I\). Thus (3.11) is established.

Now, by Theorem 2.6,

\[(3.12) \quad p\lim T/r = \theta,\]

and thus by (3.8), (3.11), and (3.12), we have

\[p\lim \hat{\theta}_n = \mu + I(\theta - \mu) = \theta. \quad \|\]

Acknowledgment. The authors are grateful to Professor Jayaram Sethuraman for valuable discussions. This paper is based on part of a dissertation written by R. M. Korwar, under the direction of M. Hollander, to be submitted to Florida State University.
REFERENCES


We derive some basic properties of a sample $X_1, \ldots, X_n$ from a Dirichlet process. Let $r_i = 0$ if $X_i = X_k$ for some $k = 1, \ldots, i-1$, and 1 otherwise. We first establish i) the distribution of $\sum_{i=1}^{n} r_i$, the number of
distinct observations in the sample, and ii) certain conditional and unconditional joint distributions of the $X_i$'s and $r_i$'s. These results are used to prove a weak law of large numbers for \( Z_n = \left( \frac{\sum_{i=1}^{n} r_i X_i}{\sum_{i=1}^{n} r_i} \right) \). The weak law is then applied to obtain the consistency of a Bayes estimator of the index of the transition measure of a mixture of Dirichlet processes.

<table>
<thead>
<tr>
<th>Key Words</th>
<th>Link A Role Wt.</th>
<th>Link B Role Wt.</th>
<th>Link C Role Wt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dirichlet processes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixtures of Dirichlet processes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonparametric Bayesian statistics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonparametric Bayesian estimation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distribution theory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weak law</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>