The Annals of Probability
1973, Vol. 1, No. 4, 705-711

CONTRIBUTIONS TO THE THEORY OF DIRICHLET PROCESSES[1]

BY RAMESH M. KORWAR[2] AND MYLES HOLLANDER[3]

Florida State University

Consider a sample $X_1, \cdots, X_n$ from a Dirichlet process P on an uncountable standard Borel space $(\mathscr{X}, \mathscr{A})$ where the parameter α of the process is assumed to be non-atomic and σ-additive. Let $D(n)$ be the number of distinct observations in the sample and denote these distinct observations by $Y_1, \cdots, Y_{D(n)}$. Our main results are (1) $D(n)/\log n \rightarrow_{\text{a.s.}} \alpha(\mathscr{X})$, $n \rightarrow \infty$, and (2) given $D(n)$, $Y_1, \cdots, Y_{D(n)}$ are independent and identically distributed according to $\alpha(\cdot)/\alpha(\mathscr{X})$. Result (1) shows that $\alpha(\mathscr{X})$ can be consistently estimated from the sample, and result (2) leads to a strong law for $\sum_{i=1}^{D(n)} Y_i/D(n)$.

0. Summary. Ferguson (1973) 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. Here we show that when the parameter α of the Dirichlet process is nonatomic and σ-additive, $\alpha(\mathscr{X})$ can be estimated from a sample from the process. Specifically, $D(n)/\log n \rightarrow_{\text{a.s.}} \alpha(\mathscr{X})$, $n \rightarrow \infty$, where $D(n)$ is the number of distinct observations in the sample $X_1, \cdots, X_n$. Furthermore, we show that in the nonatomic and σ-additive case, given $D(n)$, the $D(n)$ distinct sample values are independent and identically distributed (i.i.d.) with distribution $\alpha(\cdot)/\alpha(\mathscr{X})$. This yields a strong law of large numbers for samples from a Dirichlet process.

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

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

DEFINITION 1.2 (Ferguson). Let $(\mathscr{X}, \mathscr{A})$ be a measurable space. Let α be a non-null finite measure (nonnegative and finitely additive) on $(\mathscr{X}, \mathscr{A})$. We say

Received October 15, 1971; revised October 13, 1972.

[1] Research 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 No. AFOSR-71-2058. The United States Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation hereon.

[2,3] Visiting Stanford University, 1972-73 academic year.

AMS 1970 *subject classifications*. Primary 60K99; Secondary 62G05.

Key words and phrases. Dirichlet process, consistent estimation, strong law of large numbers, distribution theory.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Probability.* www.jstor.org

P is a *Dirichlet process* on $(\mathscr{X}, \mathscr{A})$ with parameter α if for every $k = 1, 2, \cdots$, and measurable partition $(B_1, \cdots, B_k)$ of $\mathscr{X}$, the distribution of $(P(B_1), \cdots, P(B_k))$ is Dirichlet with parameter $(\alpha(B_1), \cdots, \alpha(B_k))$.

DEFINITION 1.3 (Ferguson). The $\mathscr{X}$-valued random variables $X_1, \cdots, X_n$ constitute *a sample of size n from a Dirichlet process P* on $(\mathscr{X}, \mathscr{A})$ with parameter α if for any $m = 1, 2, \cdots$ and measurable sets $A_1, \cdots, A_m, C_1, \cdots, C_n$, $Q\{X_1 \in C_1, \cdots, X_n \in C_n \mid P(A_1), \cdots, P(A_m), P(C_1), \cdots, P(C_n)\} = \prod_{i=1}^n P(C_i)$ a.s., where Q denotes probability.

THEOREM 1.1 (Ferguson). *Let P be a Dirichlet process on $(\mathscr{X}, \mathscr{A})$ with parameter α, and let X be a sample of size 1 from P. Then for $A \in \mathscr{A}$, $Q\{X \in A\} = \alpha(A)/\alpha(\mathscr{X})$.*

THEOREM 1.2 (Ferguson). *Let P be a Dirichlet process on $(\mathscr{X}, \mathscr{A})$ with parameter α, and let $X_1, \cdots, X_n$ be a sample of size n from P. Then the conditional distribution of P given $X_1, \cdots, X_n$ is a Dirichlet process on $(\mathscr{X}, \mathscr{A})$ with parameter $\beta = \alpha + \sum_{i=1}^n \delta_{X_i}$, where, for $x \in \mathscr{X}$, $A \in \mathscr{A}$, $\delta_x(A) = 1$ if $x \in A$, 0 otherwise.*

The following representation (Theorem 1.3) of the Dirichlet process, also due to Ferguson, will be used in the proof of Theorem 2.6. Let $(\mathscr{X}, \mathscr{A})$ be a measurable space and $\alpha(\cdot)$ a finite, non-null measure on $(\mathscr{X}, \mathscr{A})$. Denote $\alpha(\mathscr{X})$ by β. Let $N(x) = -\beta \int_x^\infty e^{-y} y^{-1} dy$, $0 < x < \infty$, and let $J_1, J_2, \cdots$, be a sequence of random variables with distributions given by $P\{J_1 \leqq x_1\} = \exp(N(x_1))$, $x_1 > 0$, and $P\{J_j \leqq x_j \mid J_{j-1} = x_{j-1}, \cdots, J_1 = x_1\} = \exp\{N(x_j) - N(x_{j-1})\}$, $0 < x_j < x_{j-1}$. Set $Z_1 = \sum_{j=1}^\infty J_j$. Ferguson shows that Z_1 converges with probability one and that the distribution of Z_1 is the gamma distribution with characteristic function $\varphi(t) = (1 - it)^{-\beta}$. Let $P_j = J_j/Z_1$, $j = 1, 2, \cdots$. Then $P_j \geqq 0$ and $\sum_{j=1}^\infty P_j = 1$ w.p. 1. Now let $V_1, V_2, \cdots$ be a sequence of i.i.d. variables taking values in $\mathscr{X}$, independent of the sequence $J_1, J_2, \cdots$ each with distribution $\alpha(\cdot)/\alpha(\mathscr{X})$.

THEOREM 1.3 (Ferguson). *The random measure P on $(\mathscr{X}, \mathscr{A})$, given by, for $A \in \mathscr{A}$, $P(A) = \sum_{j=1}^\infty P_j \delta_{V_j}(A)$, is a Dirichlet process with parameter α.*

In the sequel we find it necessary to restrict various spaces to be standard Borel spaces so that certain conditional distributions exist.

2. A consistent estimator of $\alpha(\mathscr{X})$ and a strong law for the sample mean of the distinct observations. Let $X_1, \cdots, X_n$ be a sample of size n from a Dirichlet process on an uncountable standard Borel space [cf. Parthasarathy (1967) page 133 for the definition of a standard Borel space] $(\mathscr{X}, \mathscr{A})$ with parameter α. *Throughout this section we assume α is σ-additive and nonatomic.* We can view the observations $X_1, \cdots, X_n$ as being obtained sequentially as follows: Let X_1 be a sample of size 1 from P; having obtained X_1, let X_2 be a sample of size 1 from the conditional distribution of P (see Theorem 1.2) given X_1; and so on until $X_1, \cdots, X_n$ are obtained. Set $D_1 = 1$ and for $i = 2, \cdots, n$, set $D_i = 0$ if $X_i = X_j$ for some

$j = 1, \cdots, i-1$ and 1 otherwise, and let $D(n) = \sum_{i=1}^n D_i$. Let $Y_1, \cdots, Y_{D(n)}$ denote the distinct observations among $X_1, \cdots, X_n$. (Since the distribution chosen by a Dirichlet process is discrete with probability one (cf. Ferguson (1973), Blackwell (1973)), the sample values need not be distinct.) Lemma 2.1 is basic to our development.

LEMMA 2.1. *$Q\{D_i = 1\} = \alpha(\mathscr{X})/\{\alpha(\mathscr{X}) + i - 1\}, i = 1, \cdots, n$, and the D_i's are independent.*

PROOF. We have

$$
\begin{aligned}
&Q\{D_i = 1 \mid X_j = x_j, j = 1, \cdots, i-1\} \\
&\quad = Q\{X_i \in \mathscr{X} - \{x_1, \cdots, x_{i-1}\} \mid X_j = x_j, j = 1, \cdots, i-1\} \\
&\quad = \{\alpha(\mathscr{X} - \{x_1, \cdots, x_{i-1}\}) \\
&\qquad + \textstyle\sum_{j=1}^{i-1} \delta_{x_j}(\mathscr{X} - \{x_1, \cdots, x_{i-1}\})\}/\{\alpha(\mathscr{X}) + i - 1\} \quad \text{a.s.} \\
&\quad = \alpha(\mathscr{X})/\{\alpha(\mathscr{X}) + i - 1\} \quad \text{a.s.}
\end{aligned} \tag{2.1}
$$

Here $A - B$ denotes AB^c, the second equality of (2.1) follows from Theorems 1.2 and 1.1, and the final equality uses the nonatomicity of α. Taking expectations on both sides of (2.1) yields the desired expression for $Q\{D_i = 1\}$. To show $D_1, \cdots, D_n$ are independent, it suffices to show that $Q\{D_k = 1 \mid D_j, j = 1, \cdots, k-1\} = Q\{D_k = 1\}$ a.s., for $1 < k \leqq n$. Now

$$
\begin{aligned}
&Q\{D_k = 1 \mid D_j, j = 1, \cdots, k-1\} \\
&\quad = E\{Q\{D_k = 1 \mid X_j, j = 1, k-1\} \mid D_j, j = 1, \cdots, k-1\} \quad \text{a.s.} \\
&\quad = E\{\alpha(\mathscr{X})/(\alpha(\mathscr{X}) + k - 1) \mid D_j, j = 1, \cdots, k-1\} \quad \text{a.s.} \\
&\quad = Q\{D_k = 1\} \quad \text{a.s.},
\end{aligned} \tag{2.2}
$$

where the middle equality of (2.2) follows from (2.1). □

One consequence of Lemma 2.1 is that $D(n)$ has a generalized binominal distribution with parameters $(n, p_1, \cdots, p_n)$ where $p_i = \alpha(\mathscr{X})/\{\alpha(\mathscr{X}) + i - 1\}$, $i = 1, \cdots, n$. The distribution of $D(n)$ could be obtained from Proposition V of Antoniak (1969). Antoniak obtains an expression for the probability that simultaneously there are m_i observations in the sample which repeat exactly i times, $i = 1, \cdots, n$, and $Q\{D(n) = m\}$ could then be obtained by summing this expression over all m_i subject to $m = \sum_{i=1}^n m_i$, $0 \leqq m_i \leqq m$. However, the fine structure (and in particular, the independence) of the D_i's, as given in Lemma 2.1, is not available via Antoniak's result. We use this structure repeatedly in the sequel.

COROLLARY 2.2. *$Q\{D_n = 1 \text{ i.o.}\} = 1$ and $D(n) \rightarrow_{\text{a.s.}} +\infty$, $n \rightarrow \infty$.*

PROOF. From Lemma 2.1 we have

$$
\begin{aligned}
\textstyle\sum_{i=1}^n Q\{D_i = 1\} &= \alpha(\mathscr{X}) \textstyle\sum_{i=1}^n (\alpha(\mathscr{X}) + i - 1)^{-1} \\
&\geqq \alpha(\mathscr{X}) \textstyle\sum_{i=1}^n (k + i)^{-1} \rightarrow \infty, \qquad n \rightarrow \infty,
\end{aligned} \tag{2.3}
$$

where k is the greatest integer in $\alpha(\mathscr{X})$. Since the events $\{D_i = 1\}$ are independent, from (2.3) and the Borel-Cantelli lemma we obtain $Q\{D_n = 1 \text{ i.o.}\} = 1$. Since $\sum_{i=1}^n Q\{D_i = 1\}$ diverges to $+\infty$ we also have $D(n) \to_{\text{a.s.}} \infty$. □

Since $Q\{D_n = 1 \text{ i.o.}\} = 1$, we are assured of an infinite number of distinct observations. Theorem 2.3 shows how these observations can be used to obtain a strongly consistent estimator of $\alpha(\mathscr{X})$.

THEOREM 2.3. $D(n)/\log n \to_{\text{a.s.}} \alpha(\mathscr{X})$, $n \to \infty$.

To prove Theorem 2.3, we use the following lemma.

LEMMA 2.4. (cf. Loève (1963) page 238). *If $U_1, U_2, \cdots$ are independent integrable random variables, then $\sum \operatorname{Var}(U_i)/b_i^2 < \infty$ where $b_i \uparrow \infty$ implies*

$$(S_n - ES_n)/b_n \to_{\text{a.s.}} 0,$$

where $S_n = \sum_{i=1}^n U_i$.

PROOF OF THEOREM 2.3. By Lemmas 2.4 and 2.1, it is enough to show

(i) $\sum_{i=2}^n \operatorname{Var}(D_i)/(\log i)^2$ is bounded, and

(ii) $E(D(n)/\log n) \to \alpha(\mathscr{X})$.

Now, by Lemma 2.1,

$$\begin{aligned}(2.4)\qquad \sum_{i=2}^n \operatorname{Var}(D_i)/(\log i)^2 &= \alpha(\mathscr{X}) \sum_{i=2}^n (i-1)[(\alpha(\mathscr{X}) + i - 1)\log i]^{-2} \\ &< \alpha(\mathscr{X}) \sum_{i=2}^n [(i-1)(\log i)^2]^{-1} \\ &< \{(\log 2)^{-2} + \sum_{i=2}^{n-1} [i(\log i)^2]^{-1}\}\alpha(\mathscr{X})\end{aligned}$$

and the term on right-hand side of (2.4) is bounded since $\sum_{i=2}^\infty [i(\log i)^2]^{-1}$ is convergent. Again, by Lemma 2.1,

$$\begin{aligned}E(D(n)/\log n) &= (\log n)^{-1}\alpha(\mathscr{X}) \sum_{i=1}^n (\alpha(\mathscr{X}) + i - 1)^{-1} \\ &= (\log n)^{-1} + \alpha(\mathscr{X}) + a_n \alpha(\mathscr{X}),\end{aligned}$$

where $a_n =_{\text{def}} (\log n)^{-1}\{[\sum_{i=2}^n (\alpha(\mathscr{X}) + i - 1)^{-1}] - \log n\}$. The proof will be complete when we show $a_n \to 0$. Now, since

$$\alpha(\mathscr{X}) > 0,\ a_n < (\log n)^{-1}\{[\sum_{i=2}^n (i-1)^{-1}] - \log n\} < s_n,$$

where $s_n =_{\text{def}} (\log n)^{-1}\cdot \{[\sum_{i=1}^n i^{-1}] - \log n\}$. Now $s_n \to 0$ since $s_n \log n \to \gamma$, Euler's constant. Furthermore, $a_n \geqq (\log n)^{-1}\{[\sum_{i=2}^n (k+i)^{-1}] - \log n\} =_{\text{def}} c_n$, where k is the greatest integer in $\alpha(\mathscr{X})$. Rewriting c_n as

$$\begin{aligned}c_n &= (\log n)^{-1}\{[\sum_{i=1}^{n+k} i^{-1}] - \log(n+k)\} \\ &\quad - (\log n)^{-1} \sum_{i=1}^k i^{-1} + \{\log(n+k)/\log n\} - 1,\end{aligned}$$

it is easily seen that $c_n \to 0$. □

We note that Lemma 2.1 suggests a number of different estimators of $\alpha(\mathscr{X})$. For example, from Lemma 2.1 the likelihood of the D_i's is readily seen to be $L = \prod_{i=1}^n p_i^{D_i}(1-p_i)^{1-D_i}$ with $p_i = \alpha(\mathscr{X})/\{\alpha(\mathscr{X}) + i - 1\}$, $i = 1, \cdots, n$. Differentiating $\log L$ with respect to $\alpha(\mathscr{X})$ and setting this derivative equal to zero yields the estimator $\hat{\alpha}(\mathscr{X})$ defined by the solution to the equation $D(n) = \sum_{i=1}^n \alpha(\mathscr{X})/\{\alpha(\mathscr{X}) + i - 1\}$. Tables for this estimator can be found in Ewens

(1972); Ewens was led to the estimator via a sampling model arising in genetics. Another possibility is (assume that n is even so that $n = 2N$, say) to randomly divide the sample into N sets of pairs and let N_d denote the number of pairs in which the two observations are distinct. Since $Q\{D_2 = 1\} = \alpha(\mathscr{X})/\{\alpha(\mathscr{X}) + 1\}$, we could estimate $\alpha(\mathscr{X})$ by $\tilde{\alpha}(\mathscr{X})$, the solution of the equation $N_d/N = \alpha(\mathscr{X})/\{\alpha(\mathscr{X}) + 1\}$.

A virtue of Theorem 2.3 is that it shows that $\alpha(\mathscr{X})$ can be estimated using a *finite* sample from the Dirichlet process. This result is new. (Antoniak (1969) showed that $\alpha(\mathscr{X})$ could be estimated using (essentially) an *infinite* sample from the process.) We have not compared the efficiency properties of various estimators of $\alpha(\mathscr{X})$ but note that $D(n)/\log n$ and $\hat{\alpha}(\mathscr{X})$, for example, will have the same asymptotic properties (though $\hat{\alpha}(\mathscr{X})$ may be preferred in small samples).

Theorem 2.5, which follows from Ferguson's gamma process definition of the Dirichlet process, leads to a strong law for samples from a Dirichlet process.

THEOREM 2.5. *Given* $D(n)$, $Y_1, \cdots, Y_{D(n)}$ *are* i.i.d. *with distribution* $\alpha(\cdot)/\alpha(\mathscr{X})$.

PROOF. We will in fact prove a stronger result. Let $D(k) = \sum_{i=1}^k D_i$, $k = 1, \cdots, n$. We will show that given $D(k)$, $k = 1, \cdots, n$, $Y_1, \cdots, Y_{D(n)}$ are i.i.d. with distribution $\alpha(\cdot)/\alpha(\mathscr{X})$. Let $d(k)$, $k = 1, \cdots, n$ be a realization of the $D(k)$'s. Since $D_k = D(k) - D(k-1)$, these values of $D(k)$ uniquely determine values for the D_k's. Let these latter values be $D_{i_k} = 1, k = 1, \cdots, d(n)$, and $D_j = 0$ otherwise, where $1 = i_1 < \cdots < i_{d(n)} \leqq n$. Then from Theorem 1.3 we have for $A_k \in \mathscr{A}$, $k = 1, \cdots, d(n)$,

$$\begin{aligned} &Q\{X_{i_k} \in A_k, D_{i_k} = 1, D_j = 0, \\ (2.5)\qquad &\quad k = 1, \cdots, d(n), j = 2, \cdots, i_2 - 1, i_2 + 1, \cdots, n\} \\ &\quad = E \textstyle\sum_{\pi(J)} P_{j_1} \cdots P_{j_n} \delta_{V_{j_{i_1}}}(A_1) \cdots \delta_{V_{j_{i_{d(n)}}}}(A_{d(n)}) \\ &\quad = \{\textstyle\prod_{k=1}^{d(n)} \alpha(A_k)/\alpha(\mathscr{X})\} \textstyle\sum_{\pi(J)} E(P_{j_1} \cdots P_{j_n}), \end{aligned}$$

where in the summation $\sum_{\pi(J)}$ we allow positive integer values for $j_1, \cdots, j_n$ such that (i) the j_{i_k}'s are distinct and (ii) for t other than $i_1, \cdots, i_k$, j_t is equal to one of the j_{i_k}'s for which $i_k < t$. The interchange of $\sum$ and E is justified by the monotone convergence theorem and the final equality of (2.5) uses the mutual independence of the V_j's and the fact that the V_j's are independent of the P_j's. Setting $A_k = \mathscr{X}$, $k = 1, \cdots, d(n)$, in (2.5) yields

$$\begin{aligned} (2.6)\qquad &Q\{D_{i_k} = 1, D_j = 0, k = 1, \cdots, d(n), j = 2, \cdots, i_2 - 1, i_2 + 1, \cdots, n\} \\ &\quad = Q\{D(k) = d(k), k = 1, \cdots, n\} = \textstyle\sum_{\pi(J)} E(P_{j_1} \cdots P_{j_n}). \end{aligned}$$

From (2.5) and (2.6) we obtain

$$\begin{aligned} (2.7)\qquad &Q\{Y_1 \in A_1, \cdots, Y_{D(n)} \in A_{D(n)} \mid D(k) = d(k), k = 1, \cdots, n\} \\ &\quad = \textstyle\prod_{k=1}^{d(n)} \{\alpha(A_k)/\alpha(\mathscr{X})\} \quad \text{a.s.} \end{aligned}$$

The theorem follows by noting that $Q\{Y_1 \in A_1, \cdots, Y_{D(n)} \in A_{D(n)} \mid D(n) = d(n)\} = E\{Q\{Y_1 \in A_1, \cdots, Y_{D(n)} \in A_{D(n)} \mid D(k) = d(k), k = 1, \cdots, n\} \mid D(n) = d(n)]$ a.s. □

Note also that (2.7) yields the following result. Let $m \leqq n$, then given $D(k) = d(k)$, $k = 1, \cdots, n$, $Y_1, \cdots, Y_{D(m)}$ are i.i.d. with distribution $\alpha(\cdot)/\alpha(\mathscr{X})$. This result is obtained by setting $A_k = \mathscr{X}$ for $k \in \{1, \cdots, d(n)\} - \{1, \cdots, d(m)\}$.

COROLLARY 2.6. *Let* $(\mathscr{X}, \mathscr{A}) = (\mathscr{R}, \mathscr{B})$, *where* $\mathscr{R}$ *is the real line and* $\mathscr{B}$ *is the* σ*-field of Borel sets, and assume that* $\mu =_{\text{def}} \int x d\alpha(x)/\alpha(\mathscr{R})$ *exists. Then* $\sum_{i=1}^{D(n)} Y_i/D(n) \to_{\text{a.s.}} \mu$, $n \to \infty$.

PROOF. We can, without loss of generality, take $\mu = 0$. Let $m \leqq M \leqq N$ be arbitrary positive integers and let $S_{D(n)} = \sum_{i=1}^{D(n)} Y_i$. Then, if $\varepsilon > 0$,

$$\begin{aligned} Q\{\max_{M \leq n \leq N} |S_{D(n)}/D(n)| \geqq \varepsilon\} & \\ \leqq Q\{\max_{M \leq n \leq N} |S_{D(n)}/D(n)| \geqq \varepsilon, D(M) \geqq m\} + Q\{D(M) < m\} & \\ = \int_{\{D(M) \geqq m\}} Q\{\max_{M \leq n \leq N} |S_{D(n)}/D(n)| & \\ \geqq \varepsilon \mid D(k), k = 1, \cdots, N\} \, dQ + Q\{D(M) < m\} & \\ = \int_{\{D(M) \geqq m\}} Q\{\max_{M \leq n \leq N} |S'_{D(n)}/D(n)| \geqq \varepsilon\} \, dQ + Q\{D(M) < m\} & \\ \leqq Q\{\max_{m \leq n \leq N} |S_n'/n| \geqq \varepsilon\} + Q\{D(M) < m\} \, . & \end{aligned} \tag{2.8}$$

In (2.8), $S_n' = \sum_{i=1}^n Z_i$ where $Z_1, Z_2, \cdots$ is a sequence of i.i.d. random variables, with distribution $\alpha(\cdot)/\alpha(\mathscr{R})$, that are defined on $(\mathscr{R}^\infty, \mathscr{B}^\infty)$. The second equality of (2.8) follows from Theorem 2.5 (see the comment following the proof of Theorem 2.5). Letting $N \to \infty$ in (2.8), we obtain

$$Q\{\sup_{n \geqq M} |S_{D(n)}/D(n)| \geqq \varepsilon\} \leqq Q\{\sup_{n \geqq m} |S_n'/n| \geqq \varepsilon\} + Q\{D(M) < m\} \, . \tag{2.9}$$

Now let $\delta > 0$ be given. Choose m sufficiently large so that the first term on the right of (2.9) is less than $\delta/2$; this is possible by Kolmogorov's strong law. Then for this value of m, choose M sufficiently large so that the second term on the right of (2.9) is less than $\delta/2$; this can be done since $D(n) \to_{\text{a.s.}} + \infty$. Thus, for large M, $Q\{\sup_{n \geqq M} |S_{D(n)}/D(n)| \geqq \varepsilon\} < \delta$. □

A stronger result than Theorem 2.5 is true. Define the sequence $Y_1, Y_2, \cdots$ of random variables as follows: $Y_1 = X_1$, and for $j = 2, 3, \cdots, Y_j = X_k$, where k is the smallest positive integer for which $D(k) = j$. Note that $Y_1, \cdots, Y_{D(n)}$ are none other than the $D(n)$ distinct observations in a sample of size n. Then we have

THEOREM 2.7. $Y_1, Y_2, \cdots$ *are independently and identically distributed with common distribution* $\alpha(\cdot)/\alpha(\mathscr{X})$.

PROOF. Let t be a fixed positive integer. Then, for $n \geqq t$, and $A_i \in \mathscr{A}$, $i = 1, \cdots, t$,

$$\begin{aligned} Q\{Y_1 \in A_1, \cdots, Y_t \in A_t\} &= \sum_{j=t}^n Q\{Y_1 \in A_1, \cdots, Y_t \in A_t, D(n) = j\} \\ &\quad + Q\{Y_1 \in A_1, \cdots, Y_t \in A_t, D(n) < t\} \\ &= (\textstyle\prod_{i=1}^t \alpha(A_i)/\alpha(\mathscr{X}))Q\{D(n) \geqq t\} \\ &\quad + Q\{Y_1 \in A_1, \cdots, Y_t \in A_t, D(n) < t\} \, , \end{aligned} \tag{2.10}$$

the last equality following from Theorem 2.5. Let $n \to \infty$ in (2.10) and note that the left-hand side of (2.10) does not depend on n. By (2.10) and Corollary 2.2, it follows that

$$Q\{Y_1 \in A_1, \cdots, Y_t \in A_t\} = \prod_{i=1}^{t} [\alpha(A_i)/\alpha(\mathscr{X})] . \qquad \square$$

Note that a different proof of Corollary 2.6 can be obtained by utilizing Theorem 2.7 in conjunction with Kolmogorov's strong law, the result $D(n) \to +\infty$ a.s., $n \to \infty$, and Theorem 1 of Richter (1965).

Acknowledgments. The authors are grateful to Professors T. S. Ferguson, F. Proschan, I. R. Savage, J. Sethuraman, and a referee for valuable comments. This paper is based on part of a dissertation written by R. M. Korwar, under the direction of M. Hollander, at The Florida State University.

REFERENCES

[1] ANTONIAK, C. E. (1969). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. Ph. D. Dissertation. Univ. of California at Los Angeles.

[2] BLACKWELL, D. (1973). Discreteness of Ferguson selections. *Ann. Statist.* **1** 356–358.

[3] EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3** 87–112.

[4] FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

[5] LOEVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton.

[6] PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces.* Academic Press, New York.

[7] RICHTER, W. (1965). Limit theorems for sequences of random variables with sequences of random indices. *Theor. Probability Appl.* **10** 73–84.

RAMESH M. KORWAR
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

MYLES HOLLANDER
DEPARTMENT OF COMMUNITY AND
PREVENTIVE MEDICINE
STANFORD UNIVERSITY
SCHOOL OF MEDICINE
STANFORD, CALIFORNIA 94305