The Annals of Statistics
1978, Vol. 6, No. 1, 142-153

RANK ORDER ESTIMATION WITH THE DIRICHLET PRIOR[1]

BY GREGORY CAMPBELL AND MYLES HOLLANDER

Purdue University and Florida State University

Suppose that a sample of size n from a distribution function F is obtained. However, only $r (< n)$ values from the sample are observed, say $X_1, \cdots, X_r$. Without loss of generality, we can consider $X_1, \cdots, X_r$ to be the first r values in the (unordered) sample. The problem is to estimate the rank order G of X_1 among $X_1, \cdots, X_n$. The situations of interest include F nonrandom, either known or unknown, and F random. The random case assumes that F is a random distribution function chosen according to Ferguson's (*Ann. Statist.* **1** (1973) 209-230) Dirichlet process prior. Since this random distribution function is discrete with probability one, average ranks are used to resolve ties. A Bayes estimator (squared-error loss) of G is developed for the random model. For the nonrandom distribution function model, optimal non-Bayesian estimators are developed in both the case where F is known and the case where F is unknown. These estimators are compared with the Dirichlet estimator on the basis of average mean square errors under both the random and nonrandom models.

1. Introduction and summary. Let $X_1, X_2, \cdots, X_n$ be a sample of size n from the distribution function F. The problem is to estimate the rank order G of X_1 among $X_1, \cdots, X_n$ from the knowledge of $r (< n)$ observed values $X_1, \cdots, X_r$. Without loss of generality we can consider $X_1, \cdots, X_r$ to be the first r values in the (unordered) sample. Situations in which the model is applicable include the following:

(i) The Mantilla River has flooded four times in this decade with the severity of each flood measured by X, the height of the river. On the basis of the observations $X_1, \cdots, X_4$, how can we estimate the severity of the first flood, in the group of these four and the next five that occur? Equivalently, how can we estimate the rank order of X_1 among $X_1, \cdots, X_9$? (Note that we could, for example, interchange the roles of X_1 and X_4 and pose the question in terms of estimating the severity of the fourth flood.) Here $r = 4$ and $n = 9$.

(ii) An astronaut (WW, say) undergoes, as one of a pilot group of 15 astronaut trainees, extensive preparation for a space mission. Each astronaut earns a score X, a measure of overall performance. WW's score is X_1. Based on the observed values $X_1, \cdots, X_{15}$, we wish to estimate WW's rank in the total pool of 50 trainees. (Only the best ten astronauts, as measured by X, will be chosen for the mission.) Here $r = 15$ and $n = 50$.

Received January 1976; revised April 1977.

[1] Research sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Grants AFOSR-74-2581B and AFOSR-76-3109. The United States Government is authorized to reproduce and distribute reprints for governmental purposes.

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

Key words and phrases. Rank order estimation, Dirichlet process, Bayes procedure.

(iii) A swimmer (SS, say) competes in the first heat (six swimmers to a heat) of a two-heat class A event of fast swimmers. SS swims the required distance in X_1 seconds. We observe $X_1, \cdots, X_6$, the times of the heat-1 swimmers, and we wish to estimate SS' rank order among $X_1, \cdots, X_{12}$. (The six fastest swimmers in the two heats combined earn individual awards and also score points for their swim teams.) Here $r = 6$ and $n = 12$.

Of course example (i) is easily generalized to cover other undersirable (or desirable) events, example (ii) is applicable in other situations where a subgroup is selected to be on a team or perform a mission, and example (iii) can be stated in the context of other sports competitions.

This paper emphasizes the case where F is a *random* distribution function chosen according to Ferguson's (1973) Dirichlet process prior with parameter $\alpha(\cdot)$, a (completely specified) measure on the real line with the Borel σ-field. In this Dirichlet model, care must be taken in the definition of a rank order since the distribution chosen by a Dirichlet process is discrete with probability one (see Ferguson (1973), Blackwell (1973), Blackwell and MacQueen (1973), and Berk and Savage (1977)). To resolve the issue of ties with regard to the rank order, average ranks are used.

DEFINITION 1.1. Let K, L, and M denote the number of observations of X_1, $X_2, \cdots, X_n$ that are less than, equal to, and greater than X_1, respectively. Then the rank order G of X_1 among $X_1, X_2, \cdots, X_n$ is the average value of the ranks that would be assigned to the L values tied at X_1, in a joint ranking from least to greatest, if those values could be distinguished; namely,

$$G = \{(K + 1) + (K + 2) + \cdots + (K + L)\}/L = K + (L + 1)/2 . \tag{1.1}$$

Similarly, for K', L', and M' defined, respectively, to be the number of observations of $X_1, X_2, \cdots, X_r$ less than, equal to, and greater than X_1, the rank order G' of X_1 among $X_1, X_2, \cdots, X_r$ is given by $G' = K' + (L' + 1)/2$.

Section 2 contains a brief description of the Dirichlet process. Let $X_1, \cdots, X_n$ be a sample from the Dirichlet process. Given $X_1, \cdots, X_r$, the problem is to estimate G, which is a function of K, L, and M. The prior distribution of (K, L, M) given X_1 corresponds to the no data situation for this problem. In Section 3, the posterior distribution of (K, L, M), given $X_1, \cdots, X_r$, is obtained (Theorem 3.2). For squared error loss, the mean of the posterior distribution of G is the Bayes estimator. This mean, denoted $\hat{G}$, is found (Theorem 3.3) to be:

$$\hat{G} = G' + (n - r)\{\alpha'(-\infty, X_1) + \tfrac{1}{2}\alpha'(\{X_1\})\}/\alpha'(\mathscr{R}) , \tag{1.2}$$

where $\mathscr{R}$ is the real line and $\alpha' = \alpha + \sum_{i=1}^{r} \delta_{X_i}$, where δ_z is that measure which concentrates its entire mass of one at the point z. The remainder of Section 3 furnishes the necessary distribution theory to compute the mean (3.15) and the variance (3.16) of $\hat{G}$, conditional on the configuration (K, L, M).

For purposes of comparison with $\hat{G}$, two non-Bayesian competitors are

introduced; viz.,

$$G_F = G' + (n - r)F(X_1), \tag{1.3}$$

for the case where F is known and continuous, and

$$G_u = \{(n + 1)/(r + 1)\}G' \tag{1.4}$$

in the case where F is unknown. Note that G_u can be obtained from G_F by replacing $F(X_1)$ with $G'/(r + 1)$ in (1.3). In Section 4 it is shown that, if $X_1, \cdots, X_n$ is a sample from the nonrandom distribution function F, G_F has minimum average mean square error in the class of estimators of the form $aG' + bF(X_1) + c$, and G_u has minimum average mean square error in the class of estimators of the form $aG' + c$. We note that if, for the Dirichlet estimator $\hat{G}$, the measure $\alpha(\cdot)$ is nonatomic with $\alpha(-\infty, x) = \alpha(\mathscr{R})F(x)$, then as $\alpha(\mathscr{R})$ tends to infinity, $\hat{G}$ approaches G_F. (It is helpful to think of the Dirichlet framework as intermediate to the cases of F unknown and F known. In the Dirichlet model F is random, but partial information is supplied through specification of the parameter $\alpha(\cdot)$.)

The estimators $\hat{G}$, G_F, and G_u are compared on the basis of mean square errors for two models; viz., (I): the nonrandom model where $X_1, \cdots, X_n$ is a random sample from a known, continuous (nonrandom) distribution function F and, (II): the Dirichlet model where $X_1, \cdots, X_n$ is a sample of size n from a Dirichlet process with parameter $\alpha(\cdot)$, where $\alpha(\cdot)$ is assumed to be known. The estimator G_F has the smallest average mean square error for model I and $\hat{G}$ is so preferred in model II. However, for moderate $\alpha(\mathscr{R})$, the estimator $\hat{G}$ performs remarkably well in model I; in average mean square error it is positioned between G_u and G_F but as $\alpha(\mathscr{R})$ increases it approaches G_F.

2. Dirichlet process preliminaries. This section contains the basic definitions and results concerning the Dirichlet process that will be used in the sequel.

DEFINITION 2.1. 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 $Y_j = Z_j / \sum_{i=1}^{k} Z_i$, $j = 1, \cdots, k$.

PROPOSITION 2.2 (Wilks (1962), page 179). *The $\mu_{r_1, \cdots, r_l}$ moment of the Dirichlet distribution $\mathscr{D}(\alpha_1, \cdots, \alpha_k)$ is, for $l \leqq k$ and r_i a nonnegative integer such that r_i positive implies α_i positive for $i = 1, \cdots, l$,*

$$\mu_{r_1, \cdots, r_l} = \frac{\Gamma(\alpha_1 + r_1) \cdots \Gamma(\alpha_l + r_l)\Gamma(\alpha)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_l)\Gamma(\alpha + r)}, \tag{2.1}$$

where $\alpha = \sum_{i=1}^{k} \alpha_i$ and $r = \sum_{j=1}^{l} r_j$.

For k a positive integer, let $y^{[k]}$ denote the ascending factorial $y(y + 1) \cdots (y + k - 1)$ and define $y^{[0]} \equiv 1$. Then it is convenient to rewrite (2.1):

$$\mu_{r_1, \cdots, r_l} = \alpha_1^{[r_1]} \cdots \alpha_l^{[r_l]} / \alpha^{[r]}. \tag{2.1'}$$

DEFINITION 2.3 (Ferguson (1973)). Let $(\mathscr{X}, \mathscr{A})$ be a measurable space and let α denote a nonnull, finite measure on $(\mathscr{X}, \mathscr{A})$. Then P is a Dirichlet process on $(\mathscr{X}, \mathscr{A})$ with parameter α if, for every $k = 1, 2, \cdots$, and every 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 2.4 (Ferguson (1973)). 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 every $m = 1, 2, \cdots$, and measurable sets $A_1, \cdots, A_m, C_1, \cdots, C_n$,

$$\Pr\{X_1 \in C_1, \cdots, X_n \in C_n \mid P(A_1), \cdots, P(A_m), P(C_1), \cdots, P(C_n)\} = \prod_{i=1}^{n} P(C_i) \quad \text{a.s.} \tag{2.2}$$

where Pr denotes probability.

THEOREM 2.5 (Ferguson (1973)). *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 with updated parameter $\alpha + \sum_{i=1}^{n} \delta_{X_i}$, where δ_z denotes the measure which concentrates a mass of 1 at z, mass 0 elsewhere.*

3. The rank order problem. Assume that, for $X_1, \cdots, X_n$ a sample of size n from a Dirichlet process on $(\mathscr{R}, \mathscr{B})$ ($\mathscr{R}$ is the real line, $\mathscr{B}$ the Borel σ-field), only the first r variables are observed. The problem is to estimate the rank order (recall Definition 1.1) of X_1, based on the realizations $X_1 = x_1$, $X_2 = x_2$, $\cdots$, $X_r = x_r$ and the knowledge of the parameter $\alpha(\cdot)$ of the Dirichlet process. To obtain the proposed Bayes estimator, we begin by deriving (Theorem 3.2) the posterior distribution of (K, L, M), given $X_1, X_2, \cdots, X_r$. In this regard, it is helpful to first establish Proposition 3.1 which will be used in the proof of Theorem 3.2.

PROPOSITION 3.1. *If $X_1, \cdots, X_r$ is a sample of size r from a Dirichlet process on $(\mathscr{R}, \mathscr{B})$ with parameter α and if $A \in \mathscr{B}^n$, the n-dimensional Borel σ-field, then*

$$\Pr\{(X_1, \cdots, X_n) \in A \mid X_1 = x_1, \cdots, X_r = x_r\} = \int \Pr\{(X_1, \cdots, X_n) \in A \mid X_1 = x_1, \cdots, X_r = x_r, F\}\, dQ_{\alpha'}(F)$$

where $Q_{\alpha'}$ denotes the Dirichlet process prior with updated parameter $\alpha' = \alpha + \sum_{i=1}^{r} \delta_{X_i}$.

PROOF. Recall (cf. Breiman (1968), page 74) that if $E|Y| < \infty$ and if σ-fields $\mathscr{D}$ and $\mathscr{E}$ are such that $\mathscr{D} \subset \mathscr{E}$, then

$$E(Y \mid \mathscr{D}) = E(E(Y \mid \mathscr{E}) \mid \mathscr{D}) \quad \text{a.s.} \tag{3.1}$$

Now, let $Y = I_A$, $\mathscr{D} = \sigma(X_1 = x_1, \cdots, X_r = x_r)$ and $\mathscr{E}$ the σ-field generated by $X_1, \cdots, X_r$ and F. The outer expectation on the right of equation (3.1) is the integral over the conditional distribution of F given $X_1 = x_1, \cdots, X_r = x_r$, which, by Theorem 2.5, is the properly updated Dirichlet process. □

THEOREM 3.2. *Let $X_1, \cdots, X_n$ be a sample of size n from a Dirichlet process with parameter α. Then*

$$\begin{aligned}\Pr\{(K, L, M) = (k, l, m) \mid X_1 = x_1, \cdots, X_r = x_r\} \\ = \binom{n-r}{k-k',\ l-l',\ m-m'} \alpha'(-\infty, x_1)^{[k-k']} \alpha'(\{x_1\})^{[l-l']} \\ \times \alpha'(x_1, \infty)^{[m-m']} / \alpha'(\mathscr{R})^{[n-r]},\end{aligned} \tag{3.2}$$

where $K' = k'$, $L' = l'$, and $M' = m'$.

PROOF. Given $X_1, \cdots, X_r$, and F, the probability of the configuration $(K, L, M) = (k, l, m)$ is:

$$\begin{aligned}\Pr\{(K, L, M) = (k, l, m) \mid X_1 = x_1, \cdots, X_r = x_r, F\} \\ = \binom{n-r}{k-k',\ l-l',\ m-m'} F(x_1^-)^{k-k'} (F(x_1) - F(x_1^-))^{l-l'} (1 - F(x_1))^{m-m'},\end{aligned} \tag{3.3}$$

where $K' = k'$, $L' = l'$, and $M' = m'$.

By Proposition 3.1, the desired probability can be obtained by integrating the right-hand side of (3.3) with respect to the probability $Q_{\alpha'}(F)$. The integral is readily evaluated, using Proposition 2.2, to yield equation (3.2). □

The problem is to estimate the rank order of X_1, having observed $X_1 = x_1, \cdots, X_r = x_r$, where $X_1, \cdots, X_n$ is a sample from a Dirichlet process with parameter $\alpha(\cdot)$. The Dirichlet process prior on the space of distribution functions induces a prior distribution on the random variable G. The posterior distribution of G given $X_1, \cdots, X_r$ is obtainable from Theorem 3.2. Let $L(g, a)$ denote the loss incurred by taking action a (an estimate for the rank order) when g is the true state of nature (the rank order). Our development is for squared-error loss, $L(g, a) = (g - a)^2$. Recall that (K, L, M) is definitionally dependent on X_1. Consider first the "no-sample" or "no-data" problem. The "no-sample" problem is to estimate the rank order of X_1 based on the single observation X_1 $(r = 1)$. (If there is really no sample, the problem is not defined.) Then the solution of the "data" problem, with $X_1, \cdots, X_r$ provided $(r > 1)$, can be obtained by estimating $G - G'$ by merely updating the Dirichlet parameter and adding G' to it. The Bayes solution to both the "no-sample" and the "data" problem, since the loss is quadratic, is given by the mean of the posterior distribution of G given the r X's. This conditional mean is obtained in the next theorem.

THEOREM 3.3. *The mean of $G = K + (L + 1)/2$, conditional on $X_1, \cdots, X_r$, is given by the right-hand side of equation* (1.2).

PROOF. Given $X_1 = x_1, \cdots, X_r = x_r$ such that $(K', L', M') = (k', l', m')$, the random vector $(K - K', L - L', M - M')$ has a Dirichlet compound multinomial distribution (cf. Johnson and Kotz (1969), page 309) with parameters $n - r$, $\alpha'(-\infty, x_1)$, $\alpha'(\{x_1\})$, and $\alpha'(x_1, \infty)$. Thus,

$$\begin{aligned}E(K - K' + (L - L')/2 \mid X_1 = x_1, \cdots, X_r = x_r) \\ = (n - r)\{\alpha'(-\infty, x_1) + \tfrac{1}{2}\alpha'(\{x_1\})\}/\alpha'(\mathscr{R}),\end{aligned}$$

and the result follows. □

It is convenient to rewrite equation (1.2) as:

$$\hat{G} = \{(n + \alpha(\mathscr{R}))/(r + \alpha(\mathscr{R}))\}G' - \tfrac{1}{2}(n - r)/(r + \alpha(\mathscr{R})) + (n - r)[\alpha(-\infty, X_1) + \tfrac{1}{2}\alpha(\{X_1\})]/(r + \alpha(\mathscr{R}))\,. \tag{3.4}$$

Note that $\hat{G}$ depends on $X_1, \cdots, X_r$ only through X_1 and G'.

The subsequent distributional results of this section culminate in a derivation of the conditional mean and variance of $\hat{G}$, given (K, L, M).

THEOREM 3.4. (i) *If $X_1, \cdots, X_r, \cdots, X_n$ is a sample of size n from distribution function F, the distribution of $(K', L' - 1, M')$ given $(K, L, M) = (k, l, m)$ is multivariate hypergeometric with parameters $r - 1, k, l - 1, m$.* (ii) *If $X_1, \cdots, X_r, \cdots, X_n$ is a sample of size n from a Dirichlet process, then $(K', L' - 1, M')$ given $(K, L, M) = (k, l, m)$ is again multivariate hypergeometric with parameters $r - 1, k, l - 1, m$.*

PROOF. The first follows from a direct hypergeometric argument and (ii) is obtained by integration of the mass function of the multivariate hypergeometric, conditioned on F, with respect to F, and noting that the original mass function of (i) does not depend on F. □

COROLLARY 3.5. *If $X_1, \cdots, X_n$ is a sample of size n from a Dirichlet process, the mean and variance of G' given $(K, L, M) = (k, l, m)$ are:*

$$E(G' \mid (K, L, M) = (k, l, m)) = \{(r - 1)(g - 1)/(n - 1)\} + 1\,, \tag{3.5}$$

$$\begin{aligned}\operatorname{Var}(G' \mid (K, L, M) = (k, l, m)) &= (r - 1)(n - r)(n - 1)^{-2}(n - 2)^{-1} \\ &\quad \times \{k(n - k - l) + (4)^{-1}(l - 1)(n - l)\}\,.\end{aligned} \tag{3.6}$$

To compute the conditional mean and variance of $\hat{G}$, we first determine the mean and variance of $\alpha(-\infty, X_1) + \frac{1}{2}\alpha(\{X_1\})$ given $(K, L, M) = (k, l, m)$.

THEOREM 3.6. *The conditional probability of X_1 given $(K, L, M) = (k, l, m)$ is, for C a Borel subset of $\mathscr{R}$:*

$$\begin{aligned}&\Pr\{X_1 \in C \mid (K, L, M) = (k, l, m)\} \\ &\quad = \int_C \alpha(-\infty, x)^{[k]}(\alpha(\{x\}) + 1)^{[l-1]}\alpha(x, \infty)^{[m]}\, d\alpha(x)/\psi_{k,l,m}(\alpha)\,,\end{aligned} \tag{3.7}$$

where

$$\psi_{k,l,m}(\alpha) = \int_{\mathscr{R}} \alpha(-\infty, x)^{[k]}(\alpha(\{x\}) + 1)^{[l-1]}\alpha(x, \infty)^{[m]}\, d\alpha(x)\,. \tag{3.8}$$

PROOF. The distribution of (K, L, M) given X_1 and F is given by:

$$\begin{aligned}&\Pr\{(K, L, M) = (k, l, m) \mid X_1 = x, F\} \\ &\quad = \binom{n-1}{k, l-1, m} F(x^-)^k(F(x) - F(x^-))^{l-1}(1 - F(x))^m\,.\end{aligned} \tag{3.9}$$

By Proposition 3.1, integration of the right-hand side of equation (3.9) over the updated Dirichlet prior distribution for F with parameter $\alpha + \delta_x$ yields:

$$\begin{aligned}&\Pr\{(K, L, M) = (k, l, m) \mid X_1 = x\} \\ &\quad = \binom{n-1}{k, l-1, m}\alpha(-\infty, x)^{[k]}(\alpha(\{x\}) + 1)^{[l-1]}\alpha(x, \infty)^{[m]}/(\alpha(\mathscr{R}) + 1)^{[n-1]}\,.\end{aligned}$$

The joint probability of X_1 and (K, L, M) is therefore:

$$\Pr\{(K, L, M) = (k, l, m), X_1 \in C\} = \binom{n-1}{k, l-1, m} \int_C \alpha(-\infty, x)^{[k]}(\alpha(\{x\}) + 1)^{[l-1]}\alpha(x, \infty)^{[m]}\, d\alpha(x)/\alpha(\mathscr{R})^{[n]}. \tag{3.10}$$

Thus, the marginal of (K, L, M) is:

$$\Pr\{(K, L, M) = (k, l, m)\} = \binom{n-1}{k, l-1, m}\psi_{k,l,m}(\alpha)/\alpha(\mathscr{R})^{[n]}. \tag{3.11}$$

Dividing the joint probability (3.10) by the marginal of (3.11) yields the conditional probability as given in (3.7). □

COROLLARY 3.7. *The conditional mean and variance of* $\alpha(-\infty, X_1) + \frac{1}{2}\alpha(\{X_1\})$ *are given by*:

$$E(\alpha(-\infty, X_1) + \tfrac{1}{2}\alpha(\{X_1\}) \mid (K, L, M) = (k, l, m)) = \psi_{k+1,l,m}(\alpha)/\psi_{k,l,m}(\alpha) + [\tfrac{1}{2}\psi_{k,l+1,m}(\alpha)/\psi_{k,l,m}(\alpha)] - k - l/2. \tag{3.12}$$

$$\begin{aligned}\operatorname{Var}(\alpha(-\infty, X_1) + \tfrac{1}{2}\alpha(\{X_1\}) \mid (K, L, M) = (k, l, m)) &= \psi_{k+2,l,m}(\alpha)/\psi_{k,l,m}(\alpha) - \psi_{k+1,l,m}(\alpha)/\psi_{k,l,m}(\alpha)\\ &\quad - [\psi_{k+1,l,m}(\alpha)/\psi_{k,l,m}(\alpha)]^2 + \tfrac{1}{4}\psi_{k,l+2,m}(\alpha)/\psi_{k,l,m}(\alpha)\\ &\quad - \tfrac{1}{4}\psi_{k,l+1,m}(\alpha)/\psi_{k,l,m}(\alpha) - \tfrac{1}{4}[\psi_{k,l+1,m}(\alpha)/\psi_{k,l,m}(\alpha)]^2\\ &\quad + \psi_{k+1,l+1,m}(\alpha)/\psi_{k,l,m}(\alpha) - \psi_{k+1,l,m}(\alpha)\psi_{k,l+1,m}(\alpha)/[\psi_{k,l,m}(\alpha)]^2.\end{aligned} \tag{3.13}$$

PROOF. Write $\alpha(-\infty, X_1)$ as $(\alpha(-\infty, X_1) + k) - k$ and $\alpha(-\infty, X_1)^2$ as $(\alpha(-\infty, X_1) + k)^{[2]} - (2k + 1)\alpha(-\infty, X_1) - k(k + 1)$ and integrate with respect to the conditional distribution of X_1 given $(K, L, M) = (k, l, m)$. □

THEOREM 3.8. *X_1 and (K', L', M') are conditionally independent, given $(K, L, M) = (k, l, m)$.*

PROOF. The result will be established if, for any Borel set C in $\mathscr{R}$, the joint probability of $X_1 \in C$ and (K', L', M') given $(K, L, M) = (k, l, m)$ is the product of the probability of (K', L', M') given $(K, L, M) = (k, l, m)$ and the probability of $X_1 \in C$ *given* $(K, L, M) = (k, l, m)$. Now, with $\alpha' = \alpha + \sum_{i=1}^{r}\delta_{X_i}$,

$$\begin{aligned}&\Pr\{(K', L', M') = (k', l', m'), (K, L, M) = (k, l, m) \mid X_1 = x\}\\ &\quad = \Pr\{(K, L, M) = (k, l, m) \mid (K', L', M') = (k', l', m'), X_1 = x\}\\ &\qquad \times \Pr\{(K', L', M') = (k', l', m') \mid X_1 = x\}\\ &\quad = \binom{n-r}{k-k', l-l', m-m'}\alpha'(-\infty, x)^{[k-k']}\alpha'(\{x\})^{[l-l']}\alpha'(x, \infty)^{[m-m']}/\alpha'(\mathscr{R})^{[n-r]}\\ &\qquad \times \binom{r-1}{k', l'-1, m'}\alpha(-\infty, x)^{[k']}(\alpha(\{x\}) + 1)^{[l'-1]}\\ &\qquad \times \alpha(x, \infty)^{[m']}/(\alpha(\mathscr{R}) + 1)^{[r-1]}.\end{aligned} \tag{3.14}$$

The last equality of (3.14) follows from (3.2). Thus, the joint probability is given by:

$$\begin{aligned}&\Pr\{(K', L', M') = (k', l', m'), (K, L, M) = (k, l, m), X_1 \in C\}\\ &\quad = \binom{n-r}{k-k', l-l', m-m'}\binom{r-1}{k', l'-1, m'}\\ &\qquad \times \int_C \alpha(-\infty, x)^{[k]}(\alpha(\{x\}) + 1)^{[l-1]}\alpha(x, \infty)^{[m]}\, d\alpha(x)/\alpha(\mathscr{R})^{[n]}.\end{aligned}$$

Division by the marginal of (K, L, M) from (3.11) yields:

$$\begin{aligned}\Pr\{(K', L', M') &= (k', l', m'), X_1 \in C \mid (K, L, M) = (k, l, m)\} \\ &= \{\tbinom{k}{k'}\tbinom{l-1}{l'-1}\tbinom{m}{m'}/\tbinom{n-1}{r-1}\} \\ &\quad \times \int_C \alpha(-\infty, x)^{[k]}(\alpha(\{x\}) + 1)^{[l-1]}\alpha(x, \infty)^{[m]}\, d\alpha(x)/\psi_{k,l,m}(\alpha)\,. \quad \square\end{aligned}$$

Consider the special case where α is a nonatomic measure. Then, since $\alpha(\{x\}) \equiv 0$ for all x,

$$\begin{aligned}\psi_{k,l,m}(\alpha) &= \int_{\mathscr{R}} \alpha(-\infty, x)^{[k]}1^{[l-1]}\alpha(x, \infty)^{[m]}\, d\alpha(x) \\ &= (l-1)! \int_0^{\alpha(\mathscr{R})} y(\alpha(\mathscr{R}) - y)\, dy\,.\end{aligned}$$

Note that $\psi_{k,l,m}(\alpha)$ in this special case depends on $\alpha(\cdot)$ only through $\alpha(\mathscr{R})$ and depends on the l ties at X_1 only by the factor $(l-1)!$ The mean of $\hat{G}$ given $(K, L, M) = (k, l, m)$ is given by

$$\begin{aligned}&E(\hat{G} \mid (K, L, M) = (k, l, m)) \\ &\quad = [\{n + \alpha(\mathscr{R})\}/\{r + \alpha(\mathscr{R})\}][1 + \{(r-1)(g-1)/(n-1)\}] \\ &\qquad - [\tfrac{1}{2}\{n-r\}/\{r + \alpha(\mathscr{R})\}] + [\{n-r\}/\{r + \alpha(\mathscr{R})\}] \\ &\qquad \times E\{\alpha(-\infty, X_1) + \tfrac{1}{2}\alpha(\{X_1\}) \mid (K, L, M) = (k, l, m)\}\,,\end{aligned} \tag{3.15}$$

where the conditional mean of $\alpha(-\infty, X_1) + \frac{1}{2}\alpha(\{X_1\})$ is given by (3.12). The variance of $\hat{G}$ given $(K, L, M) = (k, l, m)$ is obtained by applying Theorem 3.8:

$$\begin{aligned}&\operatorname{Var}(\hat{G} \mid (K, L, M) = (k, l, m)) \\ &\quad = [\{n + \alpha(\mathscr{R})\}/\{r + \alpha(\mathscr{R})\}]^2[(r-1)(n-r)(n-1)^{-2}(n-2)^{-1} \\ &\qquad \times \{k(n-k-l) + (4)^{-1}(l-1)(n-l)\}] \\ &\qquad + [(n-r)/(r + \alpha(\mathscr{R}))]^2 \\ &\qquad \times \operatorname{Var}[\alpha(-\infty, X_1) + \tfrac{1}{2}\alpha(\{X_1\}) \mid (K, L, M) = (k, l, m)]\,,\end{aligned} \tag{3.16}$$

where the conditional variance of $\alpha(-\infty, X_1) + \frac{1}{2}\alpha(\{X_1\})$ is given in (3.13).

4. Average mean square error comparisons of rank order estimators. In this section optimal properties are developed for the competitors G_u and G_F of the estimator $\hat{G}$ and the Dirichlet estimator is then compared with these two estimators under the nonrandom and the Dirichlet models. The comparisons are on the basis of average mean square errors (mean square errors conditioned on configurations $(K, L, M) = (k, l, m)$ and then averaged over the rank configurations). In these comparisons the nonatomic measure α is related to the continuous distribution function F by the equation $\alpha(-\infty, x) = \alpha(\mathscr{R})F(x)$, in order that the distributions of samples of size one, from the distribution $F(x)$ and from the Dirichlet process with parameter α, agree.

Model I: *The nonrandom model.* Assume F is a nonrandom, continuous distribution function from which a sample $X_1, \cdots, X_r, \cdots, X_n$ is generated. For the rank configuration (K, L, M) based on X_1, it is clear that $L = 1$ with probability one and further that all n configurations $(k, 1, m)$ are equally likely. Let

(K', L', M') denote the rank configuration of X_1 among $X_1, \cdots, X_r$. Theorem 4.1 is useful in the proof of Theorem 4.2, the latter providing optimal properties of G_u and G_F. The proof of Theorem 4.1 is omitted because part (i) is well known, part (ii) is straightforward, and part (iii) is readily obtained from Theorem 3.4.

THEOREM 4.1. *If $X_1, \cdots, X_r, \cdots, X_n$ is a sample from a continuous distribution function F and if $G = K + \frac{1}{2}(L + 1)$ and $G' = K' + \frac{1}{2}(L' + 1)$, then:*

(i) *The distribution of $F(X_1)$ given $G = g$ is beta with parameters $(g, n - g + 1)$.*

(ii) *X_1 and G' are conditionally independent, given $G = g$.*

(iii)

$$E(G' \mid G = g) = \{(r - 1)(g - 1)/(n - 1)\} + 1 , \tag{4.1}$$

$$\text{Var}\,(G' \mid G = g) = (r - 1)(n - r)(n - 1)^{-2}(n - 2)^{-1}(n - g)(g - 1) . \tag{4.2}$$

Theorem 4.2 establishes an optimality property of the estimators G_u and G_F defined in equations (1.4) and (1.3), respectively.

THEOREM 4.2. (i) *In the class of linear rank order estimators of the form $aG' + c$, the estimator G_u minimizes the average mean square error under model* I.

(ii) *In the class of linear rank order estimators of the form $aG' + bF(X_1) + c$, the estimator G_F minimizes the average mean square error under model* I.

PROOF. For the general estimator $aG' + bF(X_1) + c$, the average mean square error $\mathscr{S}_1$ (say) is given (using part (ii) of Theorem 4.1) by:

$$\begin{aligned} \mathscr{S}_1 = \frac{1}{n} \sum_{g=1}^{n} \{ & a^2E(G'^2 \mid G = g) + b^2E(F^2(X_1) \mid G = g) + c^2 + g^2 \\ & - 2agE(G' \mid G = g) - 2bgE(F(X_1) \mid G = g) - 2cg \\ & + 2abE(G' \mid G = g)E(F(X_1) \mid G = g) + 2acE(G' \mid G = g) \\ & + 2bcE(F(X_1) \mid G = g)\} . \end{aligned} \tag{4.3}$$

The first and second conditional moments of G' and $F(X_1)$ are obtained from Theorem 4.1. With $b = 0$, solving the two equations $(\partial\mathscr{S}_1/\partial a) = 0$, $(\partial\mathscr{S}_1/\partial c) = 0$, yields $a = (n + 1)/(r + 1)$, $c = 0$, corresponding to G_u. Similarly, solving the three simultaneous equations $(\partial\mathscr{S}_1/\partial a) = 0$, $(\partial\mathscr{S}_1/\partial b) = 0$, $(\partial\mathscr{S}_1/\partial c) = 0$, yields $a = 1$, $b = n - r$, $c = 0$, corresponding to G_F. Since the second partial derivatives of (4.3) are positive, $G_u[G_F]$ is the desired minimum for part (i) [(ii)] of the theorem. □

Johnson (1974) considered the rank order estimation problem when F is non-random, for the case where F is unknown and the case where F is known. When F is unknown, he showed that for $r > 1$,

$$\tilde{T} = (r - 1)^{-1}(n - 1)(G' - 1) + 1 \tag{4.4}$$

is, conditional on $G = g$, an unbiased estimator of g. When F is known, Johnson

showed that

$$\hat{\tilde{T}} = [G' + (n - r + 1)F(X_1) - \{(n - r)/(n - 1)\}]/[1 + \{(2r - n - 1)/(n^2 - 1)\}] \tag{4.5}$$

is, conditional on $G = g$, an unbiased estimator of g.

Tables 4.1 and 4.2, for the nonrandom model and Dirichlet model (to be discussed below), respectively, give average mean square errors for $\hat{G}$, G_u, G_F, $\tilde{T}$ and $\hat{\tilde{T}}$. Comparisons are for the cases $3 \leqq n \leqq 5$, $r < n$, and $\alpha(\mathscr{R}) = 1.0$ and 10.0. The estimator $\tilde{T}$ is not defined for $r = 1$; this is indicated by an asterisk in Tables 4.1 and 4.2. Average mean square errors in Table 4.1 are obtained via (4.3), and average mean square errors in Table 4.2 are obtained from expression (4.6).

Model II: *The Dirichlet model.* Assume $X_1, \cdots, X_n$ is a sample of size n from a Dirichlet process with unknown parameter $\alpha(\cdot)$, where, for convenience of mean square error calculations, $\alpha(\cdot)$ is assumed to be nonatomic. With the relationship $\alpha(-\infty, x) = \alpha(\mathscr{R})F(x)$, it suffices to write the average mean square error $\mathscr{S}_2$ (say) for estimators of the form $aG' + b\alpha(-\infty, X_1) + c$, where a, b, and c can depend on n, r, and $\alpha(\mathscr{R})$. In this case, for the rank order configuration (K, L, M), the event $\{L = 1\}$ does not occur with probability one, so that averaging, over all possible nonnegative integer triples (k, l, m) such that $k + l + m = n$, is necessary. Let A denote the event $\{(K, L, M) = (k, l, m)\}$. Then, using Theorem 3.8 we find the average mean square error is

$$\begin{aligned}\mathscr{S}_2 = \sum_{(k,l,m);k+l+m=n} \Pr\{(K, L, M) = (k, l, m)\}[&a^2E_A((G')^2) \\ &+ b^2E_A(\alpha^2(-\infty, X_1)) + c^2 + (k + \tfrac{1}{2}(l + 1))^2 \\ &+ 2abE_A(G')E_A(\alpha(-\infty, X_1)) + 2acE_A(G') + 2bcE_A(\alpha(-\infty, X_1)) \\ &- 2a(k + \tfrac{1}{2}(l + 1))E_A(G') - 2b(k + \tfrac{1}{2}(l + 1))E_A(\alpha(-\infty, X_1)) \\ &- 2c(k + \tfrac{1}{2}(l + 1))]\,.\end{aligned} \tag{4.6}$$

For particular values of a, b, and c, $\mathscr{S}_2$ can be calculated by using (3.5), (3.6) and (3.11)—(3.13).

TABLE 4.1

Average mean square errors for model I

n	r	AMSE $(\hat{G})$		AMSE (G_u)	AMSE (G_F)	AMSE $(\tilde{T})$	AMSE $(\hat{\tilde{T}})$
		$\alpha(\mathscr{R}) = 1.0$	$\alpha(\mathscr{R}) = 10.0$				
3	1	.417	.336	.667	.333	*	.667
3	2	.194	.168	.222	.167	.333	.250
4	1	.688	.506	1.250	.500	*	.833
4	2	.444	.340	.556	.333	1.000	.472
4	3	.193	.169	.208	.167	.250	.215
5	1	1.000	.678	2.000	.667	*	1.000
5	2	.750	.516	1.000	.500	2.000	.678
5	3	.438	.343	.500	.333	.667	.417
5	4	.190	.170	.200	.167	.222	.201

TABLE 4.2
Average mean square errors for model II

$\alpha(\mathscr{R})$	n	r	AMSE $(\hat{G})$	AMSE (G_u)	AMSE (G_F)	AMSE $(\tilde{T})$	AMSE $(\hat{\tilde{T}})$
1.0	3	1	.278	.361	.361	*	1.028
1.0	3	2	.093	.102	.125	.139	.292
1.0	4	1	.521	.708	.708	*	1.540
1.0	4	2	.231	.269	.361	.417	.690
1.0	4	3	.087	.092	.125	.104	.257
1.0	5	1	.833	1.167	1.167	*	2.167
1.0	5	2	.417	.500	.708	.833	1.208
1.0	5	3	.208	.229	.361	.278	.611
1.0	5	4	.083	.087	.125	.093	.250
10.0	3	1	.343	.619	.346	*	.740
10.0	3	2	.157	.198	.159	.290	.258
10.0	4	1	.554	1.174	.561	*	.985
10.0	4	2	.339	.501	.346	.871	.518
10.0	4	3	.156	.184	.159	.218	.224
10.0	5	1	.792	1.894	.803	*	1.258
10.0	5	2	.545	.909	.561	1.742	.795
10.0	5	3	.335	.445	.346	.581	.460
10.0	5	4	.156	.176	.159	.194	.212

In Table 4.1, G_F has uniformly the smallest average mean square error and AMSE $(G_u) \leqq$ AMSE $(\tilde{T})$, results that follow from Theorem 4.2. Note that for the value $\alpha(\mathscr{R}) = 10$, AMSE $(\hat{G})$ is close to (but greater than) AMSE (G_F). (Recall that as $\alpha(\mathscr{R}) \to \infty$, $\hat{G} \to G_F$.) In view of the prior sample size interpretation of $\alpha(\mathscr{R})$ (see, for example, Ferguson (1973)), Table 4.1 suggests that with just one prior observation, $\hat{G}$ is to be preferred to G_u and with as few as ten prior observations, $\hat{G}$ performs well relative to the optimal estimator G_F. Since Johnson's estimators were developed using an unbiasedness criterion it is not surprising that they do not do well in terms of average mean square errors.

Table 4.2 reflects the (Bayesian) optimality of $\hat{G}$. Note that, as in Table 4.1, as $\alpha(\mathscr{R})$ increases from 1 to 10, the gap between AMSE $(\hat{G})$ and AMSE (G_F) narrows (but of course in model II, AMSE $(\hat{G})$ is less than AMSE (G_F)). Again, not unexpectedly, Johnson's estimators lag behind.

Acknowledgments. We are indebted to Professor Jayaram Sethuraman for many helpful comments. We also wish to acknowledge the valuable comments of the referees.

REFERENCES

[1] BERK, R. H. and SAVAGE, I. R. (1977). Dirichlet processes produce discrete measures: an elementary proof. (To appear in *Contributions to Statistics—Jaroslav Hájek Memorial Volume*, Academia, North-Holland, Prague.)

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

[3] BLACKWELL, D. and MACQUEEN, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1** 353–355.

[4] BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.
[5] FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.
[6] JOHNSON, N. L. (1974). Estimation of rank order. Univ. of North Carolina Institute of Statistics, Mimeo Series 931.
[7] JOHNSON, N. L. and KOTZ, S. (1969). *Discrete Distributions*. Houghton Mifflin, Boston.
[8] WILKS, S. S. (1962). *Mathematical Statistics*. Wiley, New York.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306