*The Annals of Statistics*
1976, Vol. 4, No. 3, 581–588

# EMPIRICAL BAYES ESTIMATION OF A DISTRIBUTION FUNCTION[1]

BY RAMESH M. KORWAR AND MYLES HOLLANDER

*University of Massachusetts and Florida State University*

A sequence of empirical Bayes estimators is defined for estimating a distribution function. The sequence is shown to be asymptotically optimal relative to a Ferguson Dirichlet process prior. Exact risk expressions are derived and the rate, at which the overall expected loss approaches the minimum Bayes risk, is exhibited. The empirical Bayes approach, based on the Dirichlet process, is also applied to the problem of estimating the mean of a distribution.

**1. Introduction.** Let $(P_i, \mathbf{X}_i)$, $i = 1, 2, \cdots$ be a sequence of pairs of independent random elements. The $P_i$ are probability measures which are taken to have a common prior distribution given by a Dirichlet process on $(\mathscr{R}, \mathscr{B})$, where $(\mathscr{R}, \mathscr{B})$ is the measurable space of the real line and the $\sigma$-field $\mathscr{B}$ of Borel subsets of $\mathscr{R}$. Given $P_i = P$, $\mathbf{X}_i = (X_{i1}, \cdots, X_{im})$ is random sample of size $m$ from $P$. *Throughout this paper, the parameter $\alpha(\cdot)$ of the Dirichlet process is taken to be a $\sigma$-additive nonnull finite measure on $(\mathscr{R}, \mathscr{B})$.* [We assume the reader is familiar with Ferguson's basic paper [10] that introduces the Dirichlet process. Related work includes that of Antoniak [2], Blackwell [3], Blackwell and MacQueen [4], Doksum [6], Ferguson [11], Ferguson and Klass [12], Goldstein [13], [14], Korwar and Hollander [16] and Savage [20].]

In this empirical Bayes framework, we consider the problem of estimating $F_{n+1}(t) = P_{n+1}((-\infty, t])$ on the basis of $\mathbf{X}_1, \cdots, \mathbf{X}_{n+1}$. Assume $\alpha(\mathscr{R})$ is known. Let the loss function be $L(P, \hat{F}) = \int_{\mathscr{R}} (F(t) - \hat{F}(t))^2 \, dW(t)$, where $W$ is a given finite measure (a weight function) on $(\mathscr{R}, \mathscr{B})$, $F(t) = P((-\infty, t])$, and $\hat{F}$ is an estimator of $F$. Let the parameter and action spaces be the set of all distributions $P$ on $(\mathscr{R}, \mathscr{B})$. Define, for $n = 1, 2, \cdots$, the sequence of estimators $G_{n+1}$ by

$$G_{n+1}(t) = p_m \sum_{i=1}^{n} \hat{F}_i(t)/n + (1 - p_m)\hat{F}_{n+1}(t) , \tag{1.1}$$

where

$$p_m = \alpha(\mathscr{R})/(\alpha(\mathscr{R}) + m) , \tag{1.2}$$

and $\hat{F}_i$ is the sample distribution function of $\mathbf{X}_i$, $i = 1, \cdots, n + 1$. We propose $G_{n+1}$ as a sequence of empirical Bayes estimators of $F_{n+1}$.

Received March 1974; revised October 1975.

[1] Research supported by the Office of Naval Research Contract No. NONR-998 (08) Task Order NR 042–004, by Air Force Office of Scientific Research Grant No. AFOSR-74-2581, and by National Research Council of Canada Grant A-3061. The United States Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation hereon.

*AMS* 1970 *subject classifications*. Primary 62G05; Secondary 62C99.

*Key words and phrases*. Distribution function, empirical Bayes estimator, Dirichlet process.

In Section 2 the asymptotic optimality of $G = \{G_{n+1}\}$ is established. Thus even though one need only specify $\alpha(\mathscr{R})$, the procedure is asymptotically as good as though $\alpha$ were known exactly. Exact risk expressions are given. Also, the rate at which the overall expected loss, incurred by using $G_{n+1}$, approaches the minimum Bayes risk is exhibited.

In Section 3 we compare the performance of the empirical Bayes estimator $G_{n+1}$ with that of the sample distribution function $\hat{F}_{n+1}$ based on $\mathbf{X}_{n+1}$. We show that for all $n \geqq 2$, the Bayes risk of $\hat{F}_{n+1}$, with respect to the Dirichlet process prior, is larger than the overall expected loss using $G_{n+1}$.

The results in Sections 2 and 3 suggest that the empirical Bayes approach based on the Dirichlet process can be successfully applied to other problems. Results for such an application—to the problem of estimating the mean of a distribution—are given in Section 4.

**2. Asymptotic optimality of $G_{n+1}$.** Theorems 2.1 and 2.2 below are used repeatedly in the sequel. Theorem 2.1 is Theorem 1 of Ferguson [10]. Theorem 2.2 is a direct generalization of Ferguson's [10] Proposition 4; its proof is omitted.

THEOREM 2.1 (Ferguson). *Let $P$ be a Dirichlet process on $(\mathscr{X}, \mathscr{A})$ with parameter $\alpha$, and let $X_1, \cdots, X_m$ be a sample of size $m$ from $P$. Then the conditional distribution of $P$ given $X_1, \cdots, X_m$ is a Dirichlet process on $(\mathscr{X}, \mathscr{A})$ with parameter $\beta = \alpha + \sum_{i=1}^m \delta_{X_i}$, where, for $x \in \mathscr{X}$, $A \in \mathscr{A}$, $\delta_x(A) = 1$ if $x \in A$, 0 otherwise.*

THEOREM 2.2. *Let $P$ be a Dirichlet process on $(\mathscr{R}, \mathscr{B})$ with parameter $\alpha$ and let $X_1, \cdots, X_m$ be a sample of size $m$ from $P$. Then*

$$Q\{X_1 \leqq x_1, \cdots, X_m \leqq x_m\} = \{\alpha(A_{x_{(1)}}) \cdots (\alpha(A_{x_{(m)}}) + m - 1)\}/\{\alpha(\mathscr{R}) \cdots (\alpha(\mathscr{R}) + m - 1)\},$$

*where $x_{(1)} \leqq \cdots \leqq x_{(m)}$ is an arrangement of $x_1, \cdots, x_m$ in increasing order of magnitude, $A_x = (-\infty, x]$, and $Q$ denotes probability.*

We now address the asymptotic optimality of $G_{n+1}$. In our empirical Bayes framework, Ferguson's Bayes estimator ((3) of [10], page 222) of $F$ based on $\mathbf{X}_{n+1}$ is

$$\tilde{F}_m(t) = p_m F_0(t) + (1 - p_m)\hat{F}_{n+1}(t), \tag{2.1}$$

where the dependence of $\tilde{F}_m$ on $n$ is suppressed and where $p_m$ is given by (1.2), $F_0(t)$ by

$$F_0(t) = \alpha((-\infty, t])/\alpha(\mathscr{R}), \tag{2.2}$$

and $\hat{F}_{n+1}$ is the sample distribution function of $\mathbf{X}_{n+1}$. The Bayes risks $R(\alpha)$ and $R(G_{n+1}, \alpha)$ of the estimators (2.1) and (1.1), respectively, with respect to the Dirichlet prior, are

$$R(\alpha) \equiv_{\text{def}} R(\tilde{F}_m, \alpha) = E_{\mathbf{X}_{n+1}}[\textstyle\int \{E_{F(t)|\mathbf{X}_{n+1}}(F(t) - \tilde{F}_m(t))^2\}\, dW(t)], \tag{2.3}$$

and

$$(2.4) \qquad R(G_{n+1}, \alpha) = E_{\mathbf{X}_{n+1}}[\textstyle\int \{E_{F(t)|\mathbf{X}_{n+1}}(F(t) - G_{n+1}(t))^2\}\, dW(t)]\,.$$

Let $R_{n+1}(G, \alpha)$ be the expectation of $R(G_{n+1}, \alpha)$ with respect to $\mathbf{X}_1, \cdots, \mathbf{X}_n$ (the past observations).

DEFINITION 2.3. The sequence $G = \{G_{n+1}\}$ is said to be *asymptotically optimal relative to* $\alpha$ *if* $R_{n+1}(G, \alpha)$ converges to the minimum Bayes risk $R(\alpha)$, as $n \to \infty$.

Definition 2.3 of asymptotic optimality is given here in the specific setting of the problem under discussion. For a more general definition see Section 2 of Robbins [19].

THEOREM 2.4. *Let* $\alpha(\mathscr{R})$ *be known. Then the sequence* $\{G_{n+1}\}$ *is asymptotically optimal relative to* $\alpha$.

Our proof of Theorem 2.4 uses the following lemma.

LEMMA 2.5. *Let the hypotheses be those of Theorem* 2.2 *and let* $F(t) = P((-\infty, t])$ *and* $\hat{F}(t)$ *be the sample distribution function of* $\mathbf{X} = (X_1, \cdots, X_m)$. *Then, for each* $t \in \mathscr{R}$,

$$(2.5) \qquad E(F(t)\,|\,\mathbf{X}) = \tilde{F}_m(t)\,,$$

$$(2.6) \qquad E(F^2(t)\,|\,\mathbf{X}) = \tilde{F}_m(t)(\tilde{F}_m(t)\beta(\mathscr{R}) + 1)/(\beta(\mathscr{R}) + 1)\,,$$

$$(2.7) \qquad E(\hat{F}(t)) = F_0(t)\,,$$

$$(2.8) \qquad E(\hat{F}^2(t)) = F_0(t)/m + (m-1)F_0(t)\{F_0(t)\alpha(\mathscr{R}) + 1\}/\{m(\alpha(\mathscr{R}) + 1)\}\,,$$

*where* $\tilde{F}_m(t)$ *is the analog of* (2.1) *for* $\mathbf{X}$, $F_0(t)$ *is given by* (2.2), *and* $\beta(\mathscr{R}) = \alpha(\mathscr{R}) + m$.

SKETCH OF PROOF OF LEMMA 2.5. Results (2.5) and (2.6) follow from Theorem 2.1, the definition of a Dirichlet process, and the moments of a Dirichlet distribution (cf. DeGroot [5], page 51). Results (2.7) and (2.8) follow in a straightforward way from Theorem 2.2 and the definition of $\hat{F}(t)$. □

PROOF OF THEOREM 2.4. Using (2.5) of Lemma 2.5, we can rewrite $R(G_{n+1}, \alpha)$ as

$$(2.9) \qquad R(G_{n+1}, \alpha) = R(\tilde{F}_m, \alpha) + E_{\mathbf{X}_{n+1}}[\textstyle\int \{E_{F(t)|\mathbf{X}_{n+1}}(\tilde{F}_m(t) - G_{n+1}(t))^2\}\, dW(t)]\,,$$

where $\tilde{F}_m(t)$ is given by (2.1). Now

$$(2.10) \qquad \tilde{F}_m(t) - G_{n+1}(t) = p_m\{F_0(t) - \textstyle\sum_{i=1}^n \hat{F}_i(t)/n\}\,,$$

and this is independent of $\mathbf{X}_{n+1}$ and $F(t)$. Thus

$$(2.11) \qquad R(G_{n+1}, \alpha) = R(\tilde{F}_m, \alpha) + \textstyle\int (\tilde{F}_m(t) - G_{n+1}(t))^2\, dW(t)\,.$$

It follows by (2.11), and the definition of $R_{n+1}(G, \alpha)$, that

$$(2.12) \qquad R_{n+1}(G, \alpha) = R(\tilde{F}_m, \alpha) + \textstyle\int E_{\mathbf{X}_1, \cdots, \mathbf{X}_n}(\tilde{F}_m(t) - G_{n+1}(t))^2\, dW(t)\,.$$

We evaluate each term on the right of (2.12) separately. We have, by (2.10),

$$(2.13)\qquad E_{\mathbf{X}_1,\cdots,\mathbf{X}_n}(\tilde{F}_m(t) - G_{n+1}(t))^2$$
$$= p_m{}^2 E_{\mathbf{X}_1,\cdots,\mathbf{X}_n}(F_0^2(t) - 2F_0(t)\textstyle\sum_{i=1}^n \hat{F}_i(t)/n + \sum_{i,i'=1}^n \hat{F}_i(t)\hat{F}_{i'}(t)/n^2)\,.$$

Now use Lemma 2.5 to evaluate the right side of (2.13). Note that $\hat{F}_i$ depends only on $\mathbf{X}_i$ and $\hat{F}_i$, $\hat{F}_{i'}$ are independent when $i \neq i'$. Hence (2.13), after simplification, becomes

$$(2.14)\qquad E_{\mathbf{X}_1,\cdots,\mathbf{X}_n}(\tilde{F}_m(t) - G_{n+1}(t))^2$$
$$= p_m{}^2\{(\alpha(\mathscr{R}) + m)/(mn(\alpha(\mathscr{R}) + 1))\}F_0(t)(1 - F_0(t))\,.$$

Next, let us compute $R(\tilde{F}_m, \alpha)$ which is defined by (2.3). We first obtain, for every $t \in \mathscr{R}$, $E_{F(t)|\mathbf{X}_{n+1}}(F(t) - \tilde{F}_m(t))^2$. By specializing (2.7) and (2.8) to $\mathbf{X}_{n+1}$, it follows that

$$(2.15)\qquad E_{F(t)|\mathbf{X}_{n+1}}(F(t) - \tilde{F}_m(t))^2 = \tilde{F}_m(t)(1 - \tilde{F}_m(t))/(\beta(\mathscr{R}) + 1)\,,$$

$$(2.16)\qquad E_{\mathbf{X}_{n+1}}(\tilde{F}_m(t)) = F_0(t)\,,$$

and

$$E_{\mathbf{X}_{n+1}}(\tilde{F}_m{}^2(t))$$
$$(2.17)\qquad = p_m{}^2F_0{}^2(t) + 2p_m(1 - p_m)F_0^2(t) + (1 - p_m)^2$$
$$\times [F_0(t) + (m - 1)F_0(t)(F_0(t)\alpha(\mathscr{R}) + 1)/(\alpha(\mathscr{R}) + 1)]/m\,.$$

Hence, from (2.15) to (2.17) we obtain, after some straightforward algebra,

$$(2.18)\qquad E_{\mathbf{X}_{n+1}}E_{F(t)|\mathbf{X}_{n+1}}(F(t) - \tilde{F}_m(t))^2$$
$$= [\alpha(\mathscr{R})/\{(\alpha(\mathscr{R}) + 1)(\alpha(\mathscr{R}) + m)\}]F_0(t)(1 - F_0(t))\,.$$

From (2.3) and (2.18), we then have

$$(2.19)\qquad R(\alpha) = [\alpha(\mathscr{R})/\{(\alpha(\mathscr{R}) + 1)(\alpha(\mathscr{R}) + m)\}] \textstyle\int F_0(t)(1 - F_0(t))\,dW(t)\,.$$

Then by (2.12), (2.14) and (2.19), we have, after simplification,

$$(2.20)\qquad R_{n+1}(G, \alpha) = (1 + \alpha(\mathscr{R})/mn)R(\alpha)\,.$$

Hence $\lim_{n\to\infty} R_{n+1}(G, \alpha) = R(\alpha)$. □

Note that (2.20) exhibits the rate, $1/n$, at which $R_{n+1}(G, \alpha)$ converges to $R(\alpha)$.

REMARK 2.5. Note that if one were getting $F$'s from a Dirichlet process with known $\alpha(\cdot)$, the Bayes risk for estimating $F$ without taking any sample from $F$ would be, from (2.19) with $m = 0$, $\int F_0(t)(1 - F_0(t))\,dW(t)/\{\alpha(\mathscr{R}) + 1\}$. If one planned to take a sample of size $m$ and then estimate $F$, the Bayes risk would be decreased by the factor $\alpha(\mathscr{R})/(\alpha(\mathscr{R}) + m)$, as seen from (2.19). Now, if instead only $\alpha(\mathscr{R})$ is known, but there exists one previous sample of size $m$ from another $F$ chosen by the process, then the ratio of the Bayes risk of $G_{n+1}$ to the Bayes risk of the no-sample estimator is, from (2.20), found to be $\alpha(\mathscr{R})/m$. This gives additional justification to the interpretation of $\alpha(\mathscr{R})$ as the "prior sample size" of the process.

**3. The performance of $G_{n+1}$ relative to the sample distribution function.** In most empirical Bayes situations there exists a non-Bayesian estimator which is better, in the sense of having a smaller Bayes risk, than the empirical Bayes estimator for small $n$, but which is inferior to the empirical Bayes estimator for all $n$ larger than some integer $n_0$. See, for example, Maritz [17]. Here, as a non-Bayesian estimator, we consider the sample distribution function. This function is known to possess may desirable properties. For example, denoting the sample distribution function based on $X_1, \cdots, X_n$ by $\hat{F}_n$, the Glivenko–Cantelli theorem states that, as $n \to \infty$, $\sup_{-\infty<t<\infty} |F(t) - \hat{F}_n(t)| \to 0$ a.s. Furthermore, consider the group $\mathscr{G}$ of transformations $\phi$, where $\phi$ is a continuous strictly increasing function from the real line onto the real line. Let $F$ be continuous. Aggarwal [1] has shown that for the group $\mathscr{G}$, the sample distribution function $\hat{F}_n$ is a minimax invariant estimator of $F$ under the loss function $L(F, \hat{F}) = \int [(F(t) - \hat{F}(t)]^2/\{F(t)(1 - F(t))\}]\, dF(t)$. (Also see Ferguson [9], page 191.) Dvoretzky, Kiefer and Wolfowitz [7] have shown that the sample distribution is asymptotically minimax for a wide class of loss functions. Phadia [18] establishes that $\hat{F}_n$ is minimax under the loss function $L(F, \hat{F}) = \int \{[F(t) - \hat{F}(t)]^2/[F(t)(1 - F(t))]\}\, dW(t)$.

The following theorem shows that the empirical Bayes estimator $G_{n+1}$ is better than the sample distribution function in the sense that for all $n \geqq 2$, $G_{n+1}$ has a smaller overall expected loss.

THEOREM 3.1. *Let $\alpha(\mathscr{R})$ be known. Let $\hat{F}_{n+1}$ be the sample distribution function based on $\mathbf{X}_{n+1} = (X_{n+1,1}, \cdots, X_{n+1,m})$. Then, for all $n \geqq 2$, $R(\hat{F}_{n+1}, \alpha)$, the Bayes risk of $\hat{F}_{n+1}$ with respect to the Dirichlet process prior, is larger than $R_{n+1}(G, \alpha)$, the overall expected loss using $G_{n+1}$.*

PROOF. We first compute $\int \{E_{\mathbf{X}_{n+1}}(\tilde{F}_m(t) - \hat{F}_{n+1}(t))^2\}\, dW(t)$. We find, from (2.1) and Lemma 2.5,

$$E_{\mathbf{X}_{n+1}}(\tilde{F}_m(t) - \hat{F}_{n+1}(t))^2 = [\alpha^2(\mathscr{R})/\{m(\alpha(\mathscr{R}) + 1)(\alpha(\mathscr{R}) + m)\}]F_0(t)(1 - F_0(t))\,. \tag{3.1}$$

Hence,

$$\begin{aligned} R(\hat{F}_{n+1}, \alpha) &= E_{\mathbf{X}_{n+1}}\{\int [E_{F(t)|\mathbf{X}_{n+1}}(F(t) - \tilde{F}_m(t))^2]\, dW(t)] \\ &\quad + \int \{E_{\mathbf{X}_{n+1}}(\tilde{F}_m(t) - \hat{F}_{n+1}(t))^2\}\, dW(t) \\ &= R(\alpha) + [\alpha^2(\mathscr{R})/\{(\alpha(\mathscr{R}) + 1)(\alpha(\mathscr{R}) + m)\}] \\ &\quad \times \int F_0(t)(1 - F_0(t))\, dW(t)\,. \end{aligned} \tag{3.2}$$

The second equality of (3.2) is a consequence of (3.1) and (2.3). Thus, by (3.2) and (2.19), we obtain

$$R(\hat{F}_{n+1}, \alpha) = (1 + \alpha(\mathscr{R})/m)R(\alpha)\,. \tag{3.3}$$

Comparing (3.3) with (2.20), we conclude that

$$R(\hat{F}_{n+1}, \alpha) > R_{n+1}(G, \alpha),\ n \geqq 2\,. \tag{3.4}$$

□

REMARK 3.2. Recall that the James–Stein estimator for simultaneous estimation of $k$ normal means does better, when $k \geqq 3$, in terms of mean squared error, than the classical rule which estimates each population mean by its sample mean. (See Stein [21], James and Stein [15], and, for more recent references and results, Efron and Morris [8].) Inequality (3.4) has a similar interpretation for the problem of simultaneous estimation of $k$ distribution functions. For example, in the case $k = 3$, by symmetry, if one uses $\mathbf{X}_1$ and $\mathbf{X}_2$ to estimate $F_3$, then one can use $\mathbf{X}_2$ and $\mathbf{X}_3$ to estimate $F_1$, and similarly for $F_2$. Inequality (3.4) shows that if there are at least three distribution functions to be estimated, one can do better than using, for each distribution, the corresponding sample distribution function.

**4. Empirical Bayes estimation of the mean of a distribution.** In this section we define a sequence of empirical Bayes estimators of the mean of a distribution. We assume the empirical Bayes framework of Section 2. Let the parameter space be the set of all distributions on $(\mathscr{R}, \mathscr{B})$ and the action space be the real line. Take the loss function to be $L(\mu, \hat{\mu}) = (\mu - \hat{\mu})^2$, where $\mu = \int x\,dP(x)$ is the mean of the distribution $P$ and $\hat{\mu}$ is an estimator of $\mu$. Assume also that $\int x\,d\alpha(x)$ exists and is finite. To estimate $\mu_{n+1} = \int x\,dP_{n+1}(x)$ on the basis of $\mathbf{X}_1, \cdots, \mathbf{X}_{n+1}$ we define, for $n = 1, 2, \cdots$, the sequence of estimators $\nu_{n+1}$ by

$$\nu_{n+1} = p_m \textstyle\sum_{i=1}^{n} \bar{X}_i/n + (1 - p_m)\bar{X}_{n+1}\,, \tag{4.1}$$

where $\bar{X}_i$, $i = 1, \cdots, n+1$, is the mean for the sample $\mathbf{X}_i = (X_{i1}, \cdots, X_{im})$, and $p_m$ is given by (1.2). We propose $\nu_{n+1}$ as an empirical Bayes estimator of $\mu_{n+1}$. The analogues of Theorems 2.4 and 3.1 are stated, without proof, below.

THEOREM 4.1. *Let $\alpha(\mathscr{R})$ be known. Suppose $\int x^2\,d\alpha(x)/\alpha(\mathscr{R})$ exists and is finite. Let $\mu_0 = \int x\,d\alpha(x)/\alpha(\mathscr{R})$ and $\sigma_{11} = \int x^2\,d\alpha(x)/\alpha(\mathscr{R}) - \mu_0^2$. Then*

$$R(\alpha) = [\alpha(\mathscr{R})/\{(\alpha(\mathscr{R}) + 1)(\alpha(\mathscr{R}) + m)\}]\sigma_{11}\,, \tag{4.2}$$

$$R_{n+1}(M, \alpha) = (1 + \alpha(\mathscr{R})/mn)R(\alpha)\,, \tag{4.3}$$

*where $R(\alpha)$ and $R_{n+1}(M, \alpha)$ are the analogues of (2.3) and (2.12) defined for $\hat{\mu}_{n+1}$ (the Bayes estimator of $\mu_{n+1}$) and $\nu_{n+1}$, respectively. In particular, $M = \{\nu_{n+1}\}$ is asymptotically optimal relative to $\alpha$.*

THEOREM 4.2. *Let the hypotheses of Theorem 4.1 hold. Set $\bar{X}_{n+1} = \sum_{j=1}^{m} X_{n+1,j}/m$. Then $R(\bar{X}_{n+1}, \alpha)$, the Bayes risk of $\bar{X}_{n+1}$ with respect to the Dirichlet process prior, is*

$$R(\bar{X}_{n+1}, \alpha) = (1 + \alpha(\mathscr{R})/m)R(\alpha)\,, \tag{4.4}$$

*where $R(\alpha)$ is given by (4.2). In particular, $R(\bar{X}_{n+1}, \alpha)$ is, for all $n \geqq 2$, greater than $R_{n+1}(M, \alpha)$, the overall expected loss using $\nu_{n+1}$.*

REMARK 4.3. Throughout our paper, $\alpha$ has been a $\sigma$-additive nonnull finite measure on $(\mathscr{R}, \mathscr{B})$. In particular, in proving the asymptotic optimality of $G_{n+1}$ and $\nu_{n+1}$ we did not have to impose the condition that $\alpha$ be nonatomic. However, when $\alpha$ is nonatomic, for fixed $\alpha(\mathscr{R})$ the distinct observations in a sample from a Dirichlet process are sufficient for $\alpha(-\infty, t]/\alpha(\mathscr{R})$. This sufficiency is a direct consequence of Theorem 2.5 of [16].) Motivated by this fact, and noting that $G_{n+1}$ gives extra weight to duplicated observations among the $X_{11}, \cdots, X_{nm}$, a referee has suggested that under the additional restriction that $\alpha$ be nonatomic, estimators of $F_{n+1}$ (and $\mu_{n+1}$) should give equal weight to all distinct values in the earlier samples through $X_{nm}$. This is an interesting point for future investigation.

**Acknowledgments.** We are grateful to I. R. Savage and to the referees for careful readings and many useful comments and suggestions.

## REFERENCES

[1] AGGARWAL, O. P. (1955). Some minimax invariant procedures for estimating a cumulative distribution function. *Ann. Math. Statist.* **26** 450–463.

[2] ANTONIAK, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2** 1152–1174.

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

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

[5] DEGROOT, M. H. (1970). *Optimal Statistical Decisions.* McGraw-Hill, New York.

[6] DOKSUM, K. A. (1972). Decision theory for some nonparametric models. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 331–341, Univ. of California Press.

[7] DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

[8] EFRON, B. and MORRIS, C. (1975). Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.* **70** 311–319.

[9] FERGUSON, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* Academic Press, New York.

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

[11] FERGUSON, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2** 615–629.

[12] FERGUSON, T. S. and KLASS, M. J. (1972). A representation of independent increment processes without Gaussian components. *Ann. Math. Statist.* **43** 1634–1643.

[13] GOLDSTEIN, M. (1975). Approximate Bayes solutions to some nonparametric problems. *Ann. Statist.* **3** 512–517.

[14] GOLDSTEIN, M. (1975). A note on some Bayesian nonparametric estimates. *Ann. Statist.* **3** 736–740.

[15] JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 361–379, Univ. of California Press.

[16] KORWAR, R. M. and HOLLANDER, M. (1973). Contributions to the theory of Dirichlet processes. *Ann. Probability* **1** 705–711.

[17] MARITZ, J. S. (1970). *Empirical Bayes Methods.* Methuen, London.

[18] PHADIA, E. G. (1973). Minimax estimation of a cumulative distribution function. *Ann. Statist.* **1** 1149–1157.

[19] ROBBINS, H. E. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.

[20] SAVAGE, I. R. (1969). Nonparametric statistics: A personal review. *Sankhyā Ser. A* **31** 107–144.

[21] STEIN, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 197-206, Univ. of California Press.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MASSACHUSETTS
AMHERST, MASSACHUSETTS 01002

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306