The Annals of Statistics
1983, Vol. 11, No. 2, 632–639

ROBUSTNESS OF FERGUSON'S BAYES ESTIMATOR OF A DISTRIBUTION FUNCTION[1]

BY ROBERT HANNUM AND MYLES HOLLANDER

University of Denver and Florida State University

We derive an explicit expression for the Bayes risk (using weighted squared error loss) of Dalal's Bayes estimator of a symmetric distribution under a $\mathscr{G}$-invariant Dirichlet process prior. We compare this risk to the risk of Ferguson's estimator of an arbitrary distribution under the $\mathscr{G}$-invariant prior. This enables us to (i) assess the savings in risk attained by incorporating known symmetry structure in the model and (ii) provide information about the robustness of Ferguson's estimator against a prior for which it is not Bayes.

1. Introduction and summary. Ferguson (1973) has developed Bayesian nonparametric estimators of various parameters by introducing a class of priors, called Dirichlet process priors, on a set of probability distributions. Dalal (1979a, 1979b) has further advanced Bayesian nonparametric methods by introducing $\mathscr{G}$-invariant Dirichlet process priors, whose realizations are probability measures which are invariant under a finite group $\mathscr{G}$ of transformations. $\mathscr{G}$-invariant Dirichlet process priors are useful because they allow additional information, pertaining to the structure of the underlying distribution, to be incorporated into the Bayesian estimation procedure. For example, it may be known a priori that the distribution to be estimated is exchangeable in its coordinates, or symmetric about a given point, and thus one wants the corresponding Bayes estimator to reflect this knowledge.

Dalal's (1979a) Bayes estimator of a symmetric distribution with known center of symmetry, $\tilde{F}_{\mu,n}$, is given by expression (3.1). In Section 3 we derive an explicit expression for the Bayes risk $R_g(\tilde{F}_{\mu,n}, \alpha)$ of $\tilde{F}_{\mu,n}$ under the $\mathscr{G}$-invariant Dirichlet process prior with parameter α (see Definition 2.1). We then compare this risk to the risk $R_g(\tilde{F}_n, \alpha)$ of Ferguson's (1973) estimator $\tilde{F}_n$ (defined by (3.5)) under the $\mathscr{G}$-invariant prior. In this way we assess the savings in risk attained by incorporating the known symmetry structure through the use of $\tilde{F}_{\mu,n}$. One way to measure the savings is via the behavior of the quantity $E^n_{\alpha,W} = R_g(\tilde{F}_n, \alpha)/R_g(\tilde{F}_{\mu,n}, \alpha)$. $E^n_{\alpha,W}$ depends on the choice of $\alpha(\cdot)$, including the size of $\alpha(\mathscr{R})$, W, and n. The comparison of $\tilde{F}_{\mu,n}$ and $\tilde{F}_n$ can be viewed as a Bayesian analogue of comparisons given by Schuster (1973), who considered, in a non-Bayesian framework, the problem of estimating a symmetric distribution with known center of symmetry. Our comparison thus increases the utility of Ferguson's estimator $\tilde{F}_n$ by providing information about its robustness against a prior for which it is not Bayes. We find (roughly speaking) that in the $\mathscr{G}$-invariant model where Dalal's $\tilde{F}_{\mu,n}$ is optimal, when $\alpha(\mathscr{R})$ is large relative to n, then Ferguson's $\tilde{F}_n$ performs nearly as well as Dalal's $\tilde{F}_{\mu,n}$. However, $\tilde{F}_n$ lags far behind when $\alpha(\mathscr{R})$ is very small compared to n.

Section 2 contains definitions and preliminaries relating to Dalal's $\mathscr{G}$-invariant Dirichlet process. (We have omitted preliminaries relating to Ferguson's Dirichlet process; see Ferguson (1973) for such background.)

Received November 1981; revised November 1982.

[1] This research was sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Grant 81-0038 to Florida State University and the National Institute of General Medical Sciences under Grant R01 GM21215 to Stanford University. Part of this research was completed when M. Hollander was on sabbatical leave visiting Stanford University.

AMS 1980 *subject classifications.* Primary 62G05; secondary 62G35.

Key words and phrases. $\mathscr{G}$-invariant Dirichlet process; Bayes estimator; symmetric distribution.

2. $\mathscr{G}$-Invariant Dirichlet process preliminaries. Although the results in this section hold for ℓ-dimensional Euclidean space, where ℓ is any positive integer, we will, for the sake of notational simplicity, restrict ourselves to the case $\ell = 1$.

Let $(\mathscr{R}, \mathscr{B})$ denote the real line with the Borel σ-field, and let $\mathscr{G} = \{g_1, \cdots, g_k\}$ be any finite group of transformations $\mathscr{R} \to \mathscr{R}$. A measure ν is invariant with respect to the group $\mathscr{G}$ if for all $B \in \mathscr{B}$, $\nu(B) = \nu(g_i B)$ for all $i = 1, \cdots, k$. A measurable partition $(B_1, \cdots, B_m)$ of $\mathscr{R}$ is a $\mathscr{G}$-invariant if for any $j \in \{1, \cdots, m\}$, $B_j = g_i B_j$ for all $i = 1, \cdots, k$.

DEFINITION 2.1. (Dalal, 1979a). Let α be a finite, non-negative, $\mathscr{G}$-invariant measure on $(\mathscr{R}, \mathscr{B})$. We say P_g is a $\mathscr{G}$-invariant Dirichlet process on $(\mathscr{R}, \mathscr{B})$ with parameter α, denoted by $P_g \in \mathscr{D}\mathscr{G}(\alpha)$, if P_g is invariant a.s., and if for every $k = 1, 2, \cdots$ and measurable $\mathscr{G}$-invariant partition $(B_1, \cdots, B_k)$ of $\mathscr{R}$, the distribution of $(P_g(B_1), \cdots, P_g(B_k))$ is Dirichlet with parameter $(\alpha(B_1), \cdots, \alpha(B_k))$.

Since it is not obvious how to obtain the joint distribution of $(P_g(A_1), \cdots, P_g(A_m))$ for arbitrary measurable $A_1, \cdots, A_m$ via Definition 2.1, we illustrate this procedure for the case when $\mathscr{G} = \{e, g\}$ where e is the identity transformation. Given arbitrary measurable sets $A_1, \cdots, A_m$, form the 2^m sets $B_{\nu_1,\ddagger\ddagger\ddagger,\nu_m}$ defined by $B_{\nu_1,\ddagger\ddagger\ddagger,\nu_m} = \cap_{\iota=1}^m A_j^{\nu_j}$ where $\nu_j = 0$ or 1, and $A_j^1 = A_j$ and $A_j^0 = A_j^c$, for $j = 1, \cdots, m$. Then the partition formed by distinct $[(B_{\nu 1,\cdots,\nu m} \cap gB_{\mu 1,\cdots,\mu m}) \cup (gB_{\nu 1,\cdots,\nu m} \cap B_{\mu 1,\cdots,\mu m})]$, $\nu_i = 0$ or 1 and $\mu_i = 0$ or 1, is a $\mathscr{G}$-invariant partition. Thus, given the joint distribution of this invariant partition, Dalal defines the joint distribution of $(P_g(A_1), \cdots, P_g(A_m))$ by

$$P_g(A_i) = \textstyle\sum_{\nu_\iota=1} P_g(B_{\nu_1,\cdots,\nu_m}),$$

where

$$P_g(B_{\nu 1,\cdots,\nu m}) = P_g(B_{\nu 1,\cdots,\nu m} \cap gB_{\nu 1,\cdots,\nu m})$$
$$+ \tfrac{1}{2} \textstyle\sum_{\mu\neq\nu} P_g\{(B_{\nu 1,\cdots,\nu m} \cap gB_{\mu 1,\cdots,\mu m}) \cup (B_{\mu 1,\cdots,\mu m} \cap gB_{\nu 1,\cdots,\nu m})\}.$$

DEFINITION 2.2. (Dalal, 1979a). Let $P_g \in \mathscr{D}\mathscr{G}(\alpha)$ on $(\mathscr{R}, \mathscr{B})$. Then $X_1, \cdots, X_n$ is said to be a random sample of size n from P_g if for any $m = 1, 2, \cdots$, and measurable sets $A_1, \cdots, A_m, C_1, \cdots, C_n$, $\Pr\{X_1 \in C_1, \cdots, X_n \in C_n \mid P_g(A_1), \cdots, P_g(A_m), P_g(C_1), \cdots, P_g(C_n)\} = \Pi_{\iota=1}^n P_g(C_\iota)$ a.s., where Pr denotes probability.

The following theorem shows that the posterior distribution of a $\mathscr{G}$-invariant Dirichlet process given a random sample from that process is again a $\mathscr{G}$-invariant Dirichlet process.

THEOREM 2.1. (Dalal, 1979a). *Let $P_g \in \mathscr{D}\mathscr{G}(\alpha)$ on $(\mathscr{R}, \mathscr{B})$ and let $X_1, \cdots, X_n$ be a random sample of size n from P_g. Then the conditional distribution of P_g, given $X_1, \cdots, X_n$, is a $\mathscr{G}$-invariant Dirichlet process with parameter $\alpha + \sum_{i=1}^n \delta_{X_\iota}^g$, where $\delta_{X_\iota}^g = \sum_{j=1}^k \delta_{g_j X_\iota}/k$, for $i = 1, \cdots, n$, and δ_X is a measure degenerate at X.*

The $\mathscr{G}$-invariant Dirichlet process construction given by Definition 2.1 yields a measure on $(\mathscr{R}, \mathscr{B})$ which is invariant with respect to the group $\mathscr{G}$ of transformations on $\mathscr{R}$. For example, if $\mathscr{G} = \{e, g\}$ where $e(x) = x$ and $g(x) = 2\mu - x$ for some known μ, then $F_g(\cdot) = P_g\{(-\infty, \cdot]\}$ is a random df symmetric about the point μ. An alternative method for obtaining a random symmetric df is to choose a df randomly according to Ferguson's Dirichlet process and then symmetrize this resulting distribution. More specifically, if P is a (Ferguson) Dirichlet process on $(\mathscr{R}, \mathscr{B})$, and if $F(x) = P\{(-\infty, x]\}$ and $F(x^-) = P\{(-\infty, x)\}$, then $F^*(x) = \tfrac{1}{2}\{F(x) + 1 - F([2\mu - x]^-)\}$ is clearly a random df which is symmetric about the point μ. How, if at all, do F^* and F_g differ? The following result, which we state here for the general case of a k-element group (for k a positive integer), shows that F^* and F_g have identical distributions.

THEOREM 2.2. (Dalal, 1979a). *Let P be a Dirichlet process on $(\mathscr{R}, \mathscr{B})$ with param-*

eter α, and let $P_g \in \mathscr{D}\mathscr{G}(\alpha)$ on $(\mathscr{R}, \mathscr{B})$ where $\mathscr{G} = \{g_1, \cdots, g_k\}$ is any group of transformations $\mathscr{R} \to \mathscr{R}$. Define

$$P^*(\cdot) = \textstyle\sum_{j=1}^{k} P_{g_j}(\cdot)/k, \tag{2.1}$$

where

$$P_{g_j}(\cdot) = P(g_j(\cdot)) \quad \text{for} \quad j = 1, \cdots, k. \tag{2.2}$$

Then the distribution of P^* is identical to the distribution of P_g.

A proof of Theorem 2.2 based on Ferguson's (1973) original definition of the Dirichlet process can be found, for the case $k = 2$, in Hannum (1979). See Dalal (1979a) for a proof for the general k-element group based on Ferguson's (1973) (also see Ferguson and Klass, 1972) gamma process representation of the Dirichlet process. Another proof based on Sethuraman's (1978) constructive definition of the Dirichlet process can be found in Tiwari (1980).

3. Risk analysis of Bayes estimators of a symmetric distribution. Let $\mathbf{X}_n = (X_1, \cdots, X_n)$ be a random sample of size n from an unknown symmetric df F_μ, where the point of symmetry μ is assumed known. In this section we consider Bayesian estimation of F_μ based on $\mathbf{X}_n$. We take the parameter space to be the set of all df's on $(\mathscr{R}, \mathscr{B})$ which are symmetric about μ, while the action space $\mathscr{A}$ is the set of all df's on $(\mathscr{R}, \mathscr{B})$. We assume that the loss incurred by using $\hat{F} \in \mathscr{A}$ as an estimate of F_μ is of the form $L(\hat{F}, F_\mu) = \int \{\hat{F}(t) - F_\mu(t)\}^2 \, dW(t)$, where W is a given finite weight (measure) on $(\mathscr{R}, \mathscr{B})$. Using a $\mathscr{G}$-invariant Dirichlet process prior with $\mathscr{G} = \{e, g\}$ where $e(x) = x$ and $g(x) = 2\mu - x$, Dalal (1979a) has shown that the Bayes estimator of F_μ is given by

$$\tilde{F}_{\mu,n}(t) = p_n F_0(t) + (1 - p_n) \textstyle\sum_{i=1}^{n} \{\delta_{X_i}(t) + \delta_{2\mu - X_i}(t)\}/(2n), \tag{3.1}$$

where

$$p_n = \alpha(\mathscr{R})/\{\alpha(\mathscr{R}) + n\}, \tag{3.2}$$

$$F_0(t) = \alpha\{(-\infty, t]\}/\alpha(\mathscr{R}), \tag{3.3}$$

and

$$\delta_X(t) = \begin{cases} 1 & \text{when } X \in \{(-\infty, t]\} \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

In addition to the fact that $\tilde{F}_{\mu,n}$ has minimum Bayes risk with respect to the $\mathscr{G}$-invariant Dirichlet prior, it is also a natural choice as an estimator of a symmetric df. Ferguson (1973) has shown that the Bayes (with respect to the Dirichlet process prior) estimator of an arbitrary df on $(\mathscr{R}, \mathscr{B})$ can be written

$$\tilde{F}_n(t) = p_n F_0(t) + (1 - p_n) \textstyle\sum_{i=1}^{n} \delta_{X_i}(t)/n. \tag{3.5}$$

It is straightforward then to see that Dalal's proposed estimator of a symmetric df, $\tilde{F}_{\mu,n}$, is a symmetrized version of Ferguson's Bayes estimator of an arbitrary df, $\tilde{F}_n$:

$$\tilde{F}_{\mu,n}(t) = \tfrac{1}{2}\{\tilde{F}_n(t) + 1 - \tilde{F}_n([2\mu - t]^-)\}, \tag{3.6}$$

where the notation $F([x]^-)$ denotes the probability that X is less than x if X has df F. This form (expression (3.6)) for the Bayes estimator of a symmetric df is analogous to the non-Bayesian estimator suggested by Schuster (1973) in which he symmetrizes the empirical df. In fact, replacing $\tilde{F}_n(t)$ by $\hat{F}_n(t)$ on the right-hand side of (3.6), where $\hat{F}_n(t) = \sum_{i=1}^{n} \delta_{X_i}(t)/n$ is the empirical df of the sample, yields Schuster's non-Bayesian estimator of a symmetric df:

$$\hat{F}_{\mu,n}(t) = \tfrac{1}{2}\{\hat{F}_n(t) + 1 - \hat{F}_n([2\mu - t]^-)\}. \tag{3.7}$$

Schuster examines the virtues of the estimator $\hat{F}_{\mu,n}$ by computing the variance of $\hat{F}_{\mu,n}(t)$ and comparing this variance to that of the usual non-Bayesian estimator $\hat{F}_n(t)$. We now undertake an analogous comparison in the Bayesian framework utilizing the risks of $\tilde{F}_{\mu,n}$ and $\tilde{F}_n$.

By interchanging orders of integration, the Bayes risk of $\tilde{F}_{\mu,n}$ with respect to the $\mathscr{G}$-invariant Dirichlet prior can be written as

$$R_g(\tilde{F}_{\mu,n}, \alpha) = E_{\mathbf{X}_n} \int [E_{F_\mu|\mathbf{X}_n}\{\tilde{F}_{\mu,n}(t) - F_\mu(t)\}^2]\, dW(t). \tag{3.8}$$

Now in order to evaluate the expectations in (3.8) we use algebraic manipulations similar to those in Korwar and Hollander (1976). Proofs of Lemma 3.1 and Theorem 3.2 are given in Hannum and Hollander (1982).

LEMMA 3.1. *Let $\mathscr{G}$ be the symmetry group on $(\mathscr{R}, \mathscr{B})$ with the point of symmetry μ known. Let $P_g \in \mathscr{D}\mathscr{G}(\alpha)$ on $(\mathscr{R}, \mathscr{B})$, let $F_\mu(t) = P_g\{(-\infty, t]\}$, and let $\mathbf{X}_n = (X_1, \cdots, X_n)$ be a random sample of size n from P_g. Then*

(i) $E_{F_\mu|\mathbf{X}_n} F_\mu(t) = \tilde{F}_{\mu,n}(t)$,

(ii) $E_{F_\mu}|_{\mathbf{X}_n} F_\mu^2(t) = \tilde{F}_{\mu,n}(t)[\{\tilde{F}_{\mu,n}(t)\beta(\mathscr{R}) + \tfrac{1}{2}\} + \{\tilde{F}_{\mu,n}(t) - \tfrac{1}{2}\}\delta_\mu(t)]/(\beta(\mathscr{R}) + 1)$,

(iii) $E_{\mathbf{X}_n} \hat{F}_{\mu,n}(t) = F_0(t)$,

(iv) $E_{\mathbf{X}_n} \hat{F}^2_{\mu,n}(t) = \{F_0(t)/(2n)\} + [(n-1)/\{\alpha(\mathscr{R}) + 1\}n] F_0(t)\{F_0(t)\alpha(\mathscr{R}) + \tfrac{1}{2}\}$
$$+ [\{\alpha(\mathscr{R}) + n\}/\{\alpha(\mathscr{R}) + 1\}n]\{F_0(t) - \tfrac{1}{2}\}\delta_\mu(t),$$

where $\beta(\mathscr{R}) = \alpha(\mathscr{R}) + n$, $\hat{F}_{\mu,n}(t) = \sum_{i=1}^n \{\delta_{X_i}(t) + \delta_{gX_i}(t)\}/(2n)$, and where $\delta_\mu(t) = 1$ if $t \geq \mu$ and $\delta_\mu(t) = 0$ if $t < \mu$.

Using Lemma 3.1 to evaluate the expectations in (3.8) yields the Bayes risk of $\tilde{F}_{\mu,n}$:

$$\begin{aligned} R_g(\tilde{F}_{\mu,n}, \alpha) &= [\alpha(\mathscr{R})/\{\alpha(\mathscr{R}) + 1\}\{\alpha(\mathscr{R}) + n\}] \\ &\quad \cdot \left[\int_{-\infty}^{\infty} F_0(t)\left\{\frac{1}{2} - F_0(t)\right\} dW(t) + \int_\mu^\infty \left\{F_0(t) - \frac{1}{2}\right\} dW(t)\right]. \end{aligned} \tag{3.9}$$

Analogous calculations for the risk of Ferguson's $\tilde{F}_n$ against the $\mathscr{G}$-invariant Dirichlet prior yield

$$\begin{aligned} R_g(\tilde{F}_n, \alpha) &= R_g(\tilde{F}_{\mu,n}, \alpha) + [n/2\{\alpha(\mathscr{R}) + n\}^2] \\ &\quad \left[\int_{-\infty}^{u} F_0(t)\, dW(t) + \int_\mu^\infty \{1 - F_0(t)\}\, dW(t)\right]. \end{aligned} \tag{3.10}$$

Thus (3.9) gives an exact expression for the Bayes risk of the symmetrized estimator $\tilde{F}_{\mu,n}$, and (3.10) relates this Bayes risk to the risk of the ordinary estimator $\tilde{F}_n$ against the $\mathscr{G}$-invariant prior. Using these results we can now determine the savings obtained when using $\tilde{F}_{\mu,n}$ instead of $\tilde{F}_n$ as an estimator of a symmetric distribution function. Since the underlying distribution, F_μ, is assumed to be symmetric about μ we will henceforth assume that the weight function, $W(\cdot)$, is also symmetric about μ. With this condition the right hand sides of (3.9) and (3.10) simplify so that the ratio of the risk of $\tilde{F}_n$ to the Bayes risk of $\tilde{F}_{\mu,n}$ can be written as

$$
\begin{aligned}
E^n_{\alpha,W} &= R_g(\tilde{F}_n,\ \alpha)/R_g(\tilde{F}_{\mu,n},\ \alpha) \\
&= 1 + [n\{\alpha(\mathscr{R}) + 1\}/\alpha(\mathscr{R})\{\alpha(\mathscr{R}) + n\}] \qquad (3.11) \\
&\quad \cdot \left[\int_{-\infty}^{\mu} F_0(t)\, dw(t) \Big/ \int_{-\infty}^{\mu} F_0(t)\{1 - 2F_0(t)\}\, dw(t)\right],
\end{aligned}
$$

where $w(\cdot) = W(\cdot)/W(\mathscr{R})$ is the normalized version of $W(\cdot)$. Using (3.11) we obtain:

THEOREM 3.2. *Let $E^n_{\alpha,W} = R_g(\tilde{F}_n,\ \alpha)/R_g(\tilde{F}_{\mu,n},\ \alpha)$ where $R_g(\tilde{F}_n,\ \alpha)$ and $R_g(\tilde{F}_{\mu,n},\ \alpha)$ represent the risk against the $\mathscr{G}$-invariant Dirichlet prior of Ferguson's $\tilde{F}_n$ and Dalal's symmetrized $\tilde{F}_{\mu,n}$ respectively. Then*

(i) *$E^n_{\alpha,W}$ is monotonically increasing in n, with $E^0_{\alpha,W} = 1$ and $\lim_{n\to\infty} E^n_{\alpha,W} = 1 + [\{\alpha(\mathscr{R}) + 1\}/\alpha(\mathscr{R})]D_{\alpha,W}$ where $D_{\alpha,W} = \int_{-\infty}^{\mu} F_0(t)\, dw(t)/\int_{-\infty}^{\mu} F_0(t)\{1 - 2F_0(t)\}\, dw(t)$,*

and

(ii) *$E^n_{\alpha,W}$ is monotonically decreasing in $\alpha(\mathscr{R})$, with $\lim_{\alpha(\mathscr{R})\to 0} E^n_{\alpha,W} = \infty$ and $\lim_{\alpha(\mathscr{R})\to\infty} E^n_{\alpha,W} = 1$.*

We mention some implications of Theorem 3.2. Since $\tilde{F}_{\mu,n}$ is Bayes, $\tilde{F}_{\mu,n}$ is always as good as, and usually better than, $\tilde{F}_n$. Note that when $n = 0$, $\tilde{F}_{\mu,n} = \tilde{F}_n$ and so $E^0_{\alpha,W} = 1$. Part (i) says that when n is increased, the performance of $\tilde{F}_{\mu,n}$ can only get better relative to that of $\tilde{F}_n$ for a fixed $\alpha(\mathscr{R})$. Part (ii) sheds light on the performance of $\tilde{F}_n$ compared to that of $\tilde{F}_{\mu,n}$ for changes in the value of $\alpha(\mathscr{R})$. It is clear from the forms of $\tilde{F}_n$ and $\tilde{F}_{\mu,n}$ given in (3.5) and (3.1) respectively that when $\alpha(\mathscr{R})$ is large relative to n, both estimators give much weight to the prior guess at F_μ which, for a given symmetric α, is the same ($F_0(t)$) for both $\tilde{F}_n$ and $\tilde{F}_{\mu,n}$. Roughly speaking then, $\tilde{F}_n$ and $\tilde{F}_{\mu,n}$ will tend to agree (both will be "near" F_0) when n is small compared to $\alpha(\mathscr{R})$ and so $R_g(\tilde{F}_n,\ \alpha)$ will be close to $R_g(\tilde{F}_{\mu,n},\ \alpha)$. Thus it would seem reasonable that there is little to be gained in using the symmetrized $\tilde{F}_{\mu,n}$ to estimate F_μ instead of Ferguson's $\tilde{F}_n$ in the case that $\alpha(\mathscr{R})$ is large relative to n. This is reflected by Theorem 3.2 in the fact that when n is fixed, $E^n_{\alpha,W}$ decreases to 1 as $\alpha(\mathscr{R}) \to \infty$. These remarks support an interpretation by Ferguson (1973), later advanced from a different viewpoint by Korwar and Hollander (1976), that $\alpha(\mathscr{R})$ be viewed as the "prior sample size" of the process. (But see Sethuraman and Tiwari (1982), for a probabilistic context in which the case $\alpha(\mathscr{R}) \to 0$ corresponds to much "information.") If one has a great amount of faith in the prior guess, F_0, then $\alpha(\mathscr{R})$ should be chosen very large relative to the size of the sample to be taken, n. In this case it will matter little which of $\tilde{F}_n$ and $\tilde{F}_{\mu,n}$ is used as an estimator of F_μ, since both will have risk values against the $\mathscr{G}$-invariant Dirichlet prior near $R_g(F_0,\ \alpha)$. The magnitude of the difference between $R_g(\tilde{F}_n,\ \alpha)$ and $R_g(\tilde{F}_{\mu,n},\ \alpha)$ in this situation will depend on the particular choice of $\alpha(\cdot)$ (including the size of $\alpha(\mathscr{R})$), the sample size n, and the exact weight, $W(\cdot)$, which is used in the loss function (including the size of $W(\mathscr{R})$).

In order to get some idea of how close $R_g(\tilde{F}_n,\ \alpha)$ can be to $R_g(\tilde{F}_{\mu,n},\ \alpha)$ in the case that $\alpha(\mathscr{R})$ is large relative to n, the reader is referred to Tables 3.1 through 3.3 which list values of $E^n_{\alpha,W}$ for some particular choices of α and W, and for selected values of $\alpha(\mathscr{R})$ and n. These tables will also illustrate the extent to which $\tilde{F}_{\mu,n}$ performs better than $\tilde{F}_n$ when $\alpha(\mathscr{R})$ is chosen small relative to the sample size n. In this situation more weight is placed on the sample observations and hence the difference between $R_g(\tilde{F}_n,\ \alpha)$ and $R_g(\tilde{F}_{\mu,n},\ \alpha)$ may, depending on the particular α and W chosen, be quite large. In fact, when $\alpha(\mathscr{R}) = 0$, both $\tilde{F}_n$ and $\tilde{F}_{\mu,n}$ place all the weight on the sample observations and no weight is given to the prior guess F_0 (see (3.1) and (3.5)). Thus, when $\alpha(\mathscr{R})$ is small relative to n (little faith in the prior guess F_0), the value of $E^n_{\alpha,W}$ may be much larger than 1. A further analysis of the behavior of $E^n_{\alpha,W}$ is given in the examples which follow.

EXAMPLE 3.3. Let $W(\cdot)$ be the uniform measure on the interval $\{[-c, c]\}$ for some $c > 0$, and let $F_0(\cdot) = \alpha(\cdot)/\alpha(\mathscr{R})$ be a double exponential distribution centered at zero with scale parameter $\lambda > 0$. In this case (3.11) becomes

$$E^n_{\alpha,W} = 1 + \frac{2(1 - e^{-\lambda c})n\{\alpha(\mathscr{R}) + 1\}}{\alpha(\mathscr{R})\{\alpha(\mathscr{R}) + n\}(1 - e^{-\lambda c})^2}. \tag{3.12}$$

In studying the behavior of (3.12) for different values of λ and c we note that these two parameters enter expression (3.12) only in their product form $\lambda \cdot c$. Thus we denote $\theta = \lambda c$ and consider the behavior of $E^n_{\alpha,W}$ as a function of θ. Standard derivative arguments yield the following proposition.

PROPOSITION 3.4. *If $W(\cdot)$ is uniform on $\{[-c, c]\}$ and $F_0(\cdot)$ is double exponential centered at zero with scale parameter $\lambda > 0$, then $E^n_{\alpha,W}$ is strictly monotonically decreasing in θ, with $\lim_{\theta\to 0} E^n_{\alpha,W} = \infty$ and*

$$\lim_{\theta\to\infty} E^n_{\alpha,W} = 1 + [2n\{\alpha(\mathscr{R}) + 1\}/\alpha(\mathscr{R})\{\alpha(\mathscr{R}) + \alpha(\mathscr{R}) + n\}].$$

The fact that $E^n_{\alpha,W}$ is strictly monotone as a function of $\theta = \lambda c$ gives the Bayesian decision maker who wishes to estimate a possibly symmetric df a rough guide as to how to choose an appropriate value for the scale parameter λ, depending on his degree of belief that the underlying distribution really is symmetric. For a given value of c, the value of

TABLE 3.1
Selected values of $E^n_{\alpha,W}$ when W is uniform $\{[-c, c]\}$ and α is double exponential (λ).

$\theta = \lambda c = 0.1$

$\alpha(\mathscr{R})$	$n = 1$	$n = 2$	$n = 10$	$n = 25$	$\lim_{n\to\infty}$
0.01	2103.98	2114.45	2122.89	2124.16	2125.01
0.10	211.30	221.31	230.04	231.41	232.33
1	22.03	29.04	39.24	41.44	43.06
2	11.51	16.77	27.29	30.21	32.54
10	3.10	4.86	12.57	17.52	24.13
100	1.21	1.42	2.93	5.25	22.24

$\theta = \lambda c = 1.0$

$\alpha(\mathscr{R})$	$n = 1$	$n = 2$	$n = 10$	$n = 25$	$\lim_{n\to\infty}$
0.01	317.39	318.97	320.24	320.43	320.56
0.10	32.64	34.15	35.46	35.66	35.80
1	4.16	5.22	6.75	7.08	7.33
2	2.58	3.37	4.95	5.39	5.75
10	1.32	1.58	2.74	3.49	4.48
100	1.03	1.06	1.29	1.64	4.20

$\lim_{\theta\to\infty}$

$\alpha(\mathscr{R})$	$n = 1$	$n = 2$	$n = 10$	$n = 25$	$\lim_{n\to\infty}$
0.01	201.00	202.00	202.80	202.92	203.00
0.10	21.00	21.95	22.78	22.91	23.00
1	3.00	3.67	4.64	4.85	5.00
2	2.00	2.50	3.50	3.78	4.00
10	1.20	1.37	2.10	2.57	3.20
100	1.02	1.04	1.18	1.40	3.02

$E^n_{\alpha,W}$ is smallest when λ is very large and the value of $E^n_{\alpha,W}$ is largest for λ near zero. Thus, for example, if one is quite certain that F_μ is symmetric (about zero), it would be appropriate to choose a relatively small scale parameter λ in the double exponential prior, since it is in this situation (when θ is small) that the symmetrized estimator $\tilde{F}_{\mu,n}$ performs best against $\tilde{F}_n$. Proposition 3.4 implies that the smaller the value of λ (for a fixed c) the larger the value of $E^n_{\alpha,W}$. On the other hand, choosing a relatively large value of λ would be appropriate when the degree of belief in the assumption of symmetry is not as strong. To get an idea of the increase in performance of $\tilde{F}_{\mu,n}$ relative to that of $\tilde{F}_n$, measured by the ratio of their risks against the $\mathscr{G}$-invariant Dirichlet prior, the reader is referred to Table 3.1 which lists the values of $E^n_{\alpha,W}$ for selected choices of n and $\alpha(\mathscr{R})$ for the cases $\theta = .1$, $\theta = 1.0$, and $\lim_{\theta\to\infty}$, respectively. More complete versions of Table 3.1 (and of Tables 3.2 and 3.3 which follow) can be found in Hannum and Hollander (1982).

EXAMPLE 3.5. In this example we assume only that $W(\cdot)$ and $\alpha(\cdot)$ are the same

TABLE 3.2

Selected values of $E^n_{\alpha,W}$ when $F_0(\cdot) = w(\cdot)$

$\alpha(\mathscr{R})$	$n = 1$	$n = 2$	$n = 10$	$n = 25$	$\lim_{n\to\infty}$
0.01	301.00	302.49	303.70	303.88	304.00
0.10	31.00	32.43	33.67	33.87	34.00
1	4.00	5.00	6.45	6.77	7.00
2	2.50	3.25	4.75	5.17	5.50
10	1.30	1.55	2.65	3.36	4.30
100	1.03	1.06	1.28	1.61	4.03

TABLE 3.3

Selected values of $E^n_{\alpha,W}$ when W is normal ($\mu = 0, \sigma^2$) and α is double exponential (λ).

$\omega = \sigma\lambda = 0.1$

$\alpha(\mathscr{R})$	$n = 1$	$n = 2$	$n = 10$	$n = 25$	$\lim_{n\to\infty}$
0.01	1389.57	1396.48	1402.06	1402.90	1403.46
0.10	139.86	146.47	152.23	153.13	153.74
1	14.89	19.51	26.25	27.70	28.77
2	7.94	11.41	18.36	20.29	21.83
10	2.39	3.55	8.64	11.91	16.27
100	1.14	1.28	2.28	3.80	15.02

$\omega = \sigma\lambda = 1.0$

$\alpha(\mathscr{R})$	$n = 1$	$n = 2$	$n = 10$	$n = 25$	$\lim_{n\to\infty}$
0.01	281.80	283.20	284.32	284.49	284.61
0.10	29.08	30.42	31.58	31.76	31.89
1	3.81	4.74	6.11	6.40	6.62
2	2.40	3.11	4.51	4.90	5.21
10	1.28	1.51	2.54	3.21	4.09
100	1.03	1.06	1.26	1.57	3.84

$\omega = \sigma\lambda = 1.5$

$\alpha(\mathscr{R})$	$n = 1$	$n = 2$	$n = 10$	$n = 25$	$\lim_{n\to\infty}$
0.01	232.88	234.03	234.96	235.10	235.20
0.10	24.19	25.29	26.25	26.40	26.51
1	3.32	4.09	5.22	5.46	5.64
2	2.16	2.74	3.90	4.22	4.48
10	1.23	1.43	2.28	2.82	3.55
100	1.02	1.05	1.21	1.47	3.34

measure. This entails both that $\alpha\{(-\infty, t]\} = W\{(-\infty, t]\}$ for all $t \in \mathscr{R}$ and that $\alpha(\mathscr{R}) = W(\mathscr{R})$. Then (3.11) reduces to

$$E^n_{\alpha,W} = 1 + [3n\{\alpha(\mathscr{R}) + 1\}/\alpha(\mathscr{R})\{\alpha(\mathscr{R}) + n\}]. \tag{3.13}$$

Table 3.2 provides a list of values of $E^n_{\alpha,W}$ given by (3.13) for various choices of n and $\alpha(\mathscr{R})$. We note that in this example the value of $E^n_{\alpha,W}$ does not depend on the particular measure chosen for $\alpha(\cdot)$ and $W(\cdot)$, except through the values of $\alpha(\mathscr{R})$ and $W(\mathscr{R})$. As expected, Table 3.2 shows that when $\alpha(\mathscr{R})$ is large relative to the sample size n, $\tilde{F}_n$ is quite robust (in terms of risk) compared to the symmetrized $\tilde{F}_{\mu,n}$. This is reflected by the near-unity values of $E^n_{\alpha,W}$ in the lower left part of the table. Also note that in this example $\lim_{n\to\infty} E^n_{\alpha,W} = 1 + [3\{\alpha(\mathscr{R}) + 1\}/\alpha(\mathscr{R})]$, agreeing with result (i) of Theorem 3.2, since in this example, $\int_{-\infty}^{\mu} F_0(t)\, dw(t) = 1/8$ and $\int_{-\infty}^{\mu} F_0^2(t)\, dw(t) = 1/24$.

EXAMPLE 3.6. Let $w(\cdot)$ be the normal probability distribution with mean zero and variance σ^2, and let $F_0(\cdot)$ be the double exponential distribution centered at zero with scale parameter $\lambda > 0$. Then (3.11) becomes

$$E^n_{\alpha,W} = 1 + [n\{\alpha(\mathscr{R}) + 1\}/\alpha(\mathscr{R})\{\alpha(\mathscr{R}) + n\}]/[1 - \exp(3\omega^2/2)\Phi(-2\omega)/\Phi(-\omega)], \tag{3.14}$$

where $\omega = \sigma\lambda$, and $\Phi(\cdot)$ denotes the standard normal *df*. Table 3.3 lists the values of $E^n_{\alpha,W}$ as given in (3.14) for selected choices of n and $\alpha(\mathscr{R})$ for the cases $\omega = 0.1$, $\omega = 1.0$, and $\omega = 1.5$ respectively.

Acknowledgment. We are grateful to the referees for helpful comments and for detecting an error in the original version of Lemma 3.1. We also thank the Editor and the Associate Editor for helpful suggestions.

REFERENCES

DALAL, S. R. (1979a). Dirichlet invariant processes and applications to nonparametric estimation of symmetric distribution functions. *Stochastic Process. Appl.* **9** 99–107.

DALAL, S. R. (1979b). Nonparametric and robust Bayes estimation of location. *Optimizing Methods in Statistics* (ed. by J. S. Rustagi) 141-166. Academic, New York.

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

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

HANNUM, R. C. (1979). Contributions to Bayesian nonparametric statistics. Ph.D. dissertation, Florida State University.

HANNUM, R. C. and HOLLANDER, M. (1982). Robustness of Ferguson's Bayes Estimator of a Distribution Function. Florida State University, Department of Statistics, Report No. M595-R.

KORWAR, R. M., and HOLLANDER, M. (1976). Empirical Bayes estimation of a distribution function. *Ann. Statist.* **4** 581-588.

SCHUSTER, E. F. (1973). On the goodness-of-fit problem for continuous symmetric distributions. *J. Amer. Statist. Assoc.* **68** 713–715.

SETHURAMAN, J. (1978). Unpublished manuscript. Department of Statistics, Florida State University.

SETHURAMAN, J., and TIWARI, R. C. (1982). Convergence of Dirichlet measures and the interpretation of their parameters. *Statistical Decision Theory and Related Topics III*, Vol. 2, (edited by S. S. Gupta and James O. Berger) 305–315. Academic, New York.

TIWARI, R. C. (1980). A mathematical study of the Dirichlet process. Ph.D. dissertation, Florida State University.

DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH
UNIVERSITY OF DENVER
DENVER, COLORADO 80208

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306