Robustness of Ferguson's Bayes Estimator of a Distribution Function

by

Robert Hannum² and Myles Hollander³

FSU Statistics Report No. M595
APOS Report No. 81-128

November, 1981

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

¹This research was sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Contract F49620-82-K-0007 and the National Institute of General Medical Sciences under Public Health Service Grant 5 RO1 GM21215-07. This technical report has also appeared as Stanford University Division of Biostatistics Report No. 72. Reproduction in whole or in part is permitted for any purpose of the United States Government.

²Department of Statistics, University of Denver, Denver, Colorado.

³Currently on leave at the Division of Biostatistics, Stanford University, Stanford, California.
Robustness of Ferguson's Bayes Estimator
of a Distribution Function

By
Robert Hannum and Myles Hollander
University of Denver and Florida State University and Stanford University

Abstract

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 Q-invariant Dirichlet process prior. We compare this risk to the risk of Ferguson's estimator of an arbitrary distribution under the Q-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.

1This research was sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Contract F49620-82-K-0077 and the National Institute of General Medical Sciences under Public Health Service Grant 5 R01 GM21215-07. This technical report has also appeared as Stanford University Division of Biostatistics Report No. 72.

AMS 1970 subject classification numbers: Primary 62G05; Secondary 62G35.

Key words and phrases: Q-invariant Dirichlet process; Bayes estimator; symmetric distribution.
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 $\mathcal{G}$-invariant Dirichlet process priors, whose realizations are probability measures which are invariant under a finite group $\mathcal{G}$ of transformations. $\mathcal{G}$-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.

The contributions of this paper are as follows:

In Section 3 we derive a Bayesian nonparametric estimator of a $\mathcal{G}$-invariant distribution function $(df)$, $F_{\mathcal{G}}$, where $\mathcal{G} = \{g_1, \ldots, g_k\}$ is any finite group of transformations on the space $(\mathcal{X}, \mathcal{F})$, for $k$ a positive integer. Here $\mathcal{X}$ denotes $k$-dimensional Euclidean space and $\mathcal{F}$ denotes the associated Borel $\sigma$-field. The loss, when using the estimator $\hat{F}$, is taken to be $L(F, \hat{F}) = \int (F(t) - \hat{F}(t))^2 dW(t)$, where $W$ is a given finite measure on $(\mathcal{X}, \mathcal{F})$. It is shown that the Bayes estimator $\hat{F}_{\mathcal{G}, n}$, based on a sample of size $n$ (see Definition 2.2) and given by (3.1), can be viewed as a $\mathcal{G}$-symmetrized version (averaged with respect to the group $\mathcal{G}$) of a related Ferguson Bayes estimator. Examples for various choices of the group $\mathcal{G}$ are given.

In Section 4 we consider in detail the special case where $k = 2$, $l = 1$, and $\mathcal{G}$ is the symmetric group $(\mathcal{G} = \{e, s\}$ where $e(x) = x$ and
\( g(x) = 2\mu - x \), where \( \mu \) is the known center of symmetry. In this case \( \tilde{F}_{\mu, n}^{\alpha} \) reduces to \( \tilde{F}_{\mu, n} \) (see (3.8)), Dalal's (1979a) Bayes estimator of a symmetric distribution with known center of symmetry. We derive an explicit expression for the Bayes risk \( R_{g}(\tilde{F}_{\mu, n}^{\alpha}) \) of \( \tilde{F}_{\mu, n} \) under the \( G \)-invariant Dirichlet process prior with parameter \( \alpha \) (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 (3.3), under the \( G \)-invariant prior. In this way we assess the savings in risk attained by incorporating the known symmetry structure through the use of \( \tilde{F}_{\mu, n}^{\alpha} \). One way to measure the savings is via the behavior of the quantity \( E_{\alpha, \omega}^{\mu, n} \) defined as \( R_{g}(\tilde{F}_{\mu, n}^{\alpha})/R_{g}(\tilde{F}_{\mu, n}^{\alpha}) \). \( E_{\alpha, \omega}^{\mu, n} \) depends on the choice of \( \alpha(*)\), including the size of \( \alpha(\mathbb{R}) \), \( \omega \), and \( n \). The comparison of \( \tilde{F}_{\mu, n}^{\alpha} \) and \( \tilde{F}_{n}^{\alpha} \) 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 also provides information about the robustness of Ferguson's estimator \( \tilde{F}_{n}^{\alpha} \) against a prior for which it is not Bayes. We find (roughly speaking) that in the \( G \)-invariant model where Dalal's \( \tilde{F}_{\mu, n}^{\alpha} \) is optimal, when \( \alpha(\mathbb{R}) \) is very large relative to \( n \), then Ferguson's \( \tilde{F}_{n}^{\alpha} \) performs nearly as well as Dalal's \( \tilde{F}_{\mu, n}^{\alpha} \). However, \( \tilde{F}_{n}^{\alpha} \) lags far behind when \( \alpha(\mathbb{R}) \) is very small relative to \( n \).

Section 2 contains definitions and preliminaries relating to Dalal's \( G \)-invariant Dirichlet process. (We have omitted preliminaries relating to Ferguson's (more well-known) Dirichlet process; see Ferguson (1973) for such background.) In Section 2 we describe an alternative construction of the \( G \)-invariant Dirichlet process which differs from Dalal's construction as given in Definition 2.1. Theorem 2.2 shows that the \( G \)-invariant measure \( P_{g} \) (say)
can be considered as Ferguson's Dirichlet measure $\mathbb{P}$ which has been "averaged" with respect to $\mathcal{Q}$. This theorem facilitates proofs of certain results for Dalal's $\mathcal{Q}$-invariant Dirichlet process through the use of analogous results pertaining to Ferguson's Dirichlet process. For example, checking the Kolmogorov consistency conditions for Dalal's $\mathcal{Q}$-invariant Dirichlet process, as defined in Definition 2.1, is now unnecessary since by Theorem 2.2 the conditions are inherited from Ferguson's Dirichlet process.

2. Alternative Construction for the $\mathcal{Q}$-Invariant Dirichlet Process

Let $(\mathbb{R}_+, \mathcal{B}_+)$, for $l$ a positive integer, be $l$-dimensional Euclidean space with the associated Borel $\sigma$-field, and let $\mathcal{Q} = \{\mathcal{g}_1, \ldots, \mathcal{g}_K\}$ be any finite group of transformations $\mathbb{R}^l \rightarrow \mathbb{R}^l$. A measure $\nu$ is invariant with respect to the group $\mathcal{Q}$ if for all $B \in \mathcal{B}_l$, $\nu(B) = \nu(\mathcal{g}_i B)$ for all $i = 1, \ldots, k$. A partition $(B_1, \ldots, B_m)$ of $\mathbb{R}^l$ is a $\mathcal{Q}$-invariant partition if for any $j \in \{1, \ldots, m\}$, $B_j = \mathcal{g}_i B_j$ for all $i = 1, \ldots, k$.

**DEFINITION 2.1** (Dalal, 1979a). Let $\alpha$ be a finite, non-negative, $\mathcal{Q}$-invariant measure on $(\mathbb{R}^l, \mathcal{B}^l)$. We say $\mathbb{P}_\alpha$ is a $\mathcal{Q}$-invariant Dirichlet process on $(\mathbb{R}^l, \mathcal{B}^l)$ with parameter $\alpha$ (denoted $\mathbb{P}_\alpha \in \mathcal{M}_2(\alpha)$) if for every $k = 1, 2, \ldots$, and measurable $\mathcal{Q}$-invariant partition $(B_1, \ldots, B_k)$ of $\mathbb{R}^l$, the distribution of $(\mathbb{P}_{\alpha} B_1, \ldots, \mathbb{P}_{\alpha} B_k)$ is Dirichlet with parameter $(\alpha(B_1), \ldots, \alpha(B_k))$.

Since it is not obvious how to obtain the joint distribution of $(\mathbb{P}_{\alpha}(A_1), \ldots, \mathbb{P}_{\alpha}(A_m))$ for arbitrary measurable $A_1, \ldots, A_m$ via Definition 2.1, we illustrate this procedure for the case when $\mathcal{Q} = \{e, g\}$ where $e$ is the identity transformation. Given arbitrary measurable sets $A_1, \ldots, A_m$, form the $2^m$ sets $B_{v_1}, \ldots, B_{v_m}$ defined by

$$B_{v_1}, \ldots, B_{v_m} = \bigcap_{j=1}^m A_{v_j}^j$$

where
\( v_j = 0 \) or \( l \), and \( A_j^1 = A_j \) and \( A_j^0 = A_j^c \), for \( j = 1, \ldots, m \). Then the partition formed by distinct
\[
\{(B_{v_1}, \ldots, B_{v_m} \cap gB_{\mu_1}, \ldots, \mu_m} \cup (gB_{1}, \ldots, \mu_m)\}, \quad v_1 = 0 \text{ or } 1 \text{ and } \mu_i = 0 \text{ or } 1,
\]
is a \( G \)-invariant partition. Thus given the joint distribution of this invariant partition, Dalal defines the joint distribution of \( (P_g(A_1), \ldots, P_g(A_m)) \) by
\[
P_g(A_i) = \sum_{v_1=1}^{\nu_m} P_g(B_{v_1}, \ldots, v_m)
\]
where
\[
P_g(B_{v_1}, \ldots, v_m) = P_g(A_{v_1}, \ldots, v_m \cap gB_{\mu_1}, \ldots, \mu_m)
\]
\[
+ \frac{1}{2} \sum_{(\mu_1, ..., \mu_m) \neq (v_1, ..., v_m)} P_g\left(B_{\mu_1}, \ldots, \mu_m \cap gB_{\mu_1}, \ldots, \mu_m\right)
\]
\[
\cup \{B_{\mu_1}, \ldots, \mu_m \cap gB_{\mu_1}, \ldots, \mu_m\}
\]

**DEFINITION 2.2** (Dalal, 1979a). Let \( P_g \in \mathcal{G}(a) \) on \((\mathbb{R}_+, \mathbb{B})\). Then \( X_1, \ldots, X_n \) is said to be a random sample of size \( n \) from \( P_g \) if for any \( m = 1, 2, \ldots \), and measurable sets \( A_1, \ldots, A_m, C_1, \ldots, C_n \),
\[
\Pr[X_1 \in C_1, \ldots, X_n \in C_n | P_g(A_1), \ldots, P_g(A_m), P_g(C_1), \ldots, P_g(C_n)] = \prod_{i=1}^{n} P_g(C_i) \text{ a.s.}
\]
where \( \Pr \) denotes probability.

The following theorem shows that the posterior distribution of a \( G \)-invariant Dirichlet process given a random sample from that process is again a \( G \)-invariant Dirichlet process.

**THEOREM 2.1** (Dalal, 1979a). Let \( P_g \in \mathcal{G}(a) \) on \((\mathbb{R}_+, \mathbb{B})\) and let \( X_1, \ldots, X_n \) be a random sample of size \( n \) from \( P_g \). Then the conditional
distribution of \( P_\beta \), given \( X_1, \ldots, X_n \), is a \( G \)-invariant Dirichlet process with parameter \( \alpha + \sum_{i=1}^{n} \delta_{X_i}^\beta \), where \( \delta_{X_i}^\beta = \frac{\delta(\delta_{X_i}^\beta \in \beta)}{k} \), for \( i = 1, \ldots, n \), and \( \delta_{X} \) is a measure degenerate at \( X \).

The \( G \)-invariant Dirichlet process construction given by Definition 2.1 yields a measure on \( (\mathbb{R}^k, \mathcal{B}^k) \) which is invariant with respect to the group \( G \) of transformations on \( \mathbb{R}^k \). For example if \( P_\mu \in \mathcal{M}(\alpha) \) on \( (\mathbb{R}^1, \mathcal{B}^1) \), and if \( G = \{e, g\} \) where \( e(x) = x \) and \( g(x) = 2\mu - x \) for some known \( \mu \), then \( P_\mu(\cdot) = P_\mu((-\infty, \cdot]) \) is a random df symmetric about the point \( \mu \). 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 \( (\mathbb{R}^1, \mathcal{B}^1) \), and if \( P(x) = P((-\infty, x]) \) and \( P(x^-) = P((-\infty, x)) \), then \( P^*(x) = \frac{1}{2}[P(x) + 1 - P([2\mu-x^-])] \) is clearly a random df which is symmetric about the point \( \mu \). How, if at all, do \( P^* \) and \( P_\mu \) differ? The following theorem, which we state for the general case of a \( k \)-element group (for \( k \) a positive integer), shows that \( P^* \) and \( P_\mu \) do not differ.

**THEOREM 2.2.** Let \( P \) be a Dirichlet process on \( (\mathbb{R}^1, \mathcal{B}^1) \) with parameter \( \alpha \), and let \( P_\mu \in \mathcal{M}(\alpha) \) on \( (\mathbb{R}^1, \mathcal{B}^1) \) where \( G = \{\mathbb{S}_1, \ldots, \mathbb{S}_k\} \) is any group of transformations \( \mathbb{R}^k \rightarrow \mathbb{R}^k \). Define

\[
(2.1) \quad P^*(\cdot) = \frac{1}{k} \sum_{j=1}^{k} P_{\mathbb{S}_j}(\cdot),
\]

where

\[
(2.2) \quad P_{\mathbb{S}_j}(\cdot) = P(\mathbb{S}_j(\cdot)) \quad \text{for} \quad j = 1, \ldots, k.
\]

Then the distribution of \( P^* \) is identical to the distribution of \( P_\mu \).
PROOF. A proof based on Ferguson's (1973) original definition of the Dirichlet process can be found, for the case $k = 2$, in Handum (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).

The above theorem suggests an alternative way in which to view the random $\mathbb{Q}$-invariant Dirichlet measure. This perspective is quite intuitive and may be roughly stated as "the $\mathbb{Q}$-invariant Dirichlet measure, $\mathbb{P}_{\mathbb{Q}}$, can be considered as Ferguson's Dirichlet measure which has been 'averaged' with respect to the group $\mathbb{Q}$".

3. Bayesian Estimation of a $\mathbb{Q}$-Invariant Distribution Function

Let $\mathbb{G} = \{g_1, \ldots, g_k\}$ be any finite group of transformations defined on $\ell$-dimensional Euclidean space $\mathbb{R}^\ell$, where $\ell$ is a positive integer. We wish to estimate the distribution function $F_\epsilon$ defined on $\mathbb{R}^\ell$ and known to be invariant under the group of transformations $\mathbb{G}$. The parameter space is taken to be the set of all $\mathbb{Q}$-invariant distribution functions on $(\mathbb{R}^\ell, \mathcal{B}^\ell)$, where $\mathcal{B}^\ell$ denotes the Borel $\sigma$-field associated with $\mathbb{R}^\ell$. We take the action space to be the space of all cumulative distribution functions on $(\mathbb{R}^\ell, \mathcal{B}^\ell)$, and we assume the loss incurred by using $\hat{F}$ as an estimate of $F_\epsilon$ to be of the form $L(F_\epsilon, \hat{F}) = \int (F_\epsilon(t) - \hat{F}(t))^2 dW(t)$, where $W$ is a given finite weight (measure) on $(\mathbb{R}^\ell, \mathcal{B})$. The following theorem gives the Bayes estimator of $F_\epsilon$ assuming a $\mathbb{Q}$-invariant Dirichlet process prior.

THEOREM 3.1. Let $F_\epsilon$ be a $\mathbb{Q}$-invariant Dirichlet process on $(\mathbb{R}^\ell, \mathcal{B}^\ell)$ where $\mathbb{G} = \{g_1, \ldots, g_k\}$ is any finite group of transformations on $\mathbb{R}^\ell$ and
define \( F_\infty(t) = \frac{1}{1+\sum_{i=1}^{k} \mathbb{I}_{\{t \leq t_i\}} \delta_{X_i}(t)} \) where \( t = (t_1, \ldots, t_k) \). Let the parameter space, the action space, and the loss function be defined as above. Then the Bayes estimator of \( F_\infty \), based on the sample \( X_1, \ldots, X_n \), is

\[
(3.1) \quad \tilde{F}_{\infty,n}(t) = p_n F_\infty(t) + (1-p_n) \sum_{i=1}^{n} \sum_{j=1}^{k} \delta_{X_j}(t)/n ,
\]

where \( p_n = \alpha(\infty)/[\alpha(\infty) + n] \), \( F_0(t) = \alpha(\infty) / \alpha(\infty) \), and where \( \delta_{X_i}(t) = 1 \) when \( X_i \leq t_i \) for all \( i = 1, \ldots, k \), and \( \delta_{X_i}(t) = 0 \) otherwise.

PROOF. The proof uses methods similar to those in Section 4.2 of Dalal (1979a) and thus is omitted.

We note that in addition to the fact that the estimator \( \tilde{F}_{\infty,n} \), given by (3.1), has minimum Bayes risk with respect to the \( \mathcal{G} \)-invariant Dirichlet prior, it is also a natural choice as an estimator of a \( \mathcal{G} \)-invariant distribution as can be seen by the following argument. Ferguson (1973) has shown that the Bayes (with respect to the Dirichlet process prior) estimator of an arbitrary df on \( (\infty, \infty) \) is given by

\[
(3.3) \quad \tilde{F}_n(t) = p_n F_0(t) + (1-p_n) \left( \sum_{i=1}^{n} \delta_{X_i}(t) / n \right) ,
\]

where

\[
(3.4) \quad p_n = \alpha(\infty)/[\alpha(\infty) + n] ,
\]

\[
(3.5) \quad F_0(t) = \alpha(\infty) / \alpha(\infty) ,
\]

and
\((3.6)\)

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

When constructing an estimator for an unknown \(G\)-invariant df, \(F_{\tilde{G}}\) based on the random sample \(X_1, \ldots, X_n\) from a \(G\)-invariant Dirichlet process on \((\mathbb{R}, \mathcal{B})\), an intuitive approach would be to average the Ferguson estimator \((3.3)\) with respect to the group of transformations \(G\) and use this resulting averaged version of \(F_n\) to estimate the unknown \(G\)-invariant df. That is, use \(\frac{1}{k} \left[ \sum_{j=1}^{k} \frac{\tilde{F}_{\tilde{G}, n}^{(j)}(t)}{k} \right] / k\) as the estimator of \(F_{\tilde{G}}\), where for each \(j = 1, \ldots, k\), \(\tilde{F}_{\tilde{G}, n}^{(j)}\) denotes Ferguson's estimator \((3.3)\) based on \(g_j X_1, g_j X_2, \ldots, g_j X_n\). However, since

\[
\frac{1}{k} \left[ \sum_{j=1}^{k} \tilde{F}_{\tilde{G}, n}^{(j)}(t) \right] / k = \frac{1}{k} \left\{ \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}^{g_j}(t) \right\} / k
\]

\[
= \frac{p_n F_0(t) + (1-p_n) \sum_{i=1}^{n} \delta_{X_i}^{g_j}(t)}{kn}
\]

\[
= \tilde{F}_{\tilde{G}, n}(t),
\]

this intuitive approach leads to the Bayes (with respect to the \(G\)-invariant Dirichlet prior) estimator of a \(G\)-invariant df given in Theorem 3.1.

It is easy to see that when \(k = 2, \ell = 1\), and \(G = \{e, g\}\) where \(e(x) = x\) and \(g(x) = 2u - x\) for some known point \(u\), the estimator \(\tilde{F}_{\tilde{G}, n}\) can be written as

\[(3.8)\]

\[
\tilde{F}_{h,n}(t) = p_n F_0(t) + (1-p_n) \sum_{i=1}^{n} \delta_{X_i} + \delta_{2u-X_i}/2n ,
\]
where \( p, F, \delta \) are defined by (3.4)–(3.6). The estimator given by (3.8) is Dalal's (1979a) proposed estimator of a symmetric df with known center of symmetry. In this case expression (3.7) reduces to

\[
(3.9) \quad \tilde{F}_{\mu, n}(t) = \frac{1}{2} \{ F_n(t) + 1 - \tilde{F}_n(2\mu - t) \}
\]

where the notation \( F([xℑ]) \) denotes the probability that \( X \) is less than \( x \) if \( X \) has df \( F \). This form (expression (3.9)) for the Bayes estimator of a symmetric df is analogous to the non-Bayesian method of symmetrizing the empirical df, as suggested by Schuster (1973). In fact replacing \( \tilde{F}_n(t) \) by \( F_n(t) \) on the right hand side of (3.9), where \( F_n(t) = \frac{\sum_{i=1}^{n} \delta_X(t)}{n} \)

is the empirical df of the sample, yields Schuster's (non-Bayesian) estimate of a symmetric df:

\[
(3.10) \quad \hat{F}_{\mu, n}(t) \overset{\text{def.}}{=} \frac{1}{2} \{ F_n(t) + 1 - \hat{F}_n(2\mu - t) \}
\]

Schuster (1973) 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 \( \hat{F}_n(t) \). An analogous comparison in the Bayesian framework utilizing the risks of \( \tilde{F}_{\mu, n} \) and \( \tilde{F}_n \) is done in Section 4 of this paper.

Other cases in which Theorem 3.1 is useful include Bayesian estimation of a multivariate symmetric df and Bayesian estimation of a multivariate df which is exchangeable in the coordinates. As an example, when estimating a bivariate df which is known to be exchangeable in the coordinates, use the estimator (3.1) with \( \lambda = 2 \) and \( Q = \{ e, g \} \) where \( e(x, y) = (x, y) \) and \( g(x, y) = (y, x) \). This will yield the Bayes estimator.
\[ \hat{F}_{g,n}(t) = p_n \hat{F}_0(t) + (1-p_n) \sum_{i=1}^{n} [\delta(x_i, Y_i(t)) + \delta(y_i, x_i(t))] / 2n, \]

where \((X_1, Y_1), \ldots, (X_n, Y_n)\) is a sample of size \(n\) from the unknown df \(F_g\). Although the cases of multivariate symmetry and exchangeability in the coordinates are the most natural examples of the possible uses of Theorem 3.1, Bayes estimators can be obtained for distributions which are invariant under any (finite) group of transformations. The Klein four group (cf. Rotman (1973), p. 38) on \(R^7\) exchanges pairs of coordinates and can be defined as \(Q = \{e, g_1, g_2, g_3, g_4\}\), where \(g_1(w, x, y, z) = (w, x, y, z)\), \(g_2(w, x, y, z) = (x, w, z, y)\), \(g_3(w, x, y, z) = (y, z, w, x)\), and \(g_4(w, x, y, z) = (z, y, x, w)\). The Bayes estimator of a df which is invariant with respect to the Klein four group is obtained via Theorem 3.1 and is given by

\[ \hat{F}_{g,n}(t) = p_n \hat{F}_0(t) + (1-p_n) \sum_{i=1}^{n} [\delta((W_i, X_i, Y_i, Z_i), (X_i, W_i, Z_i, Y_i)) + \delta((Y_i, Z_i, W_i, X_i), (X_i, Y_i, W_i, Z_i)) / 4n, \]

where \((W_1, X_1, Y_1, Z_1), \ldots, (W_n, X_n, Y_n, Z_n)\) is a random sample of size \(n\) from the unknown df.

4. Risk Analysis of Bayes Estimators of a Symmetric Distribution

Let \(X_n = (X_1, \ldots, X_n)\) be a random sample of size \(n\) from an unknown df, \(F_\mu\), where the point of symmetry \(\mu\) is assumed known. In this section we will consider Bayesian estimation of \(F_\mu\) based on \(X_n\).

In the case \(G\) is the symmetry group on \((\mathbb{R}, \mathbb{R})\) (so that \(Q = \{e, g\}\) with \(e(x) = x\) and \(g(x) = 2\mu - x\)), Theorem 3.1 gives a Bayes estimator of
\( F_\mu \) when the prior distribution for \( F_\mu \) is a \( \mathcal{G} \)-invariant Dirichlet process. 

This estimator is Dalal's \( \tilde{F}_{\mu,n} \) given in (3.8). We have shown that this estimator is a symmetrized version of Ferguson's estimator of an arbitrary distribution function, \( \bar{F}_n \) (given in expression (3.3)), as is evidenced by expression (3.9). Clearly then \( \tilde{F}_{\mu,n} \) is a natural choice as an estimator of a symmetric distribution function since (i) \( \tilde{F}_{\mu,n} \) possesses the minimum Bayes risk with respect to the \( \mathcal{G} \)-invariant Dirichlet prior, and (ii) \( \tilde{F}_{\mu,n} \) is the estimator obtained by symmetrizing the usual (Ferguson's) Bayes (with respect to Ferguson's prior) estimator of an arbitrary distribution, \( \bar{F}_n \).

It is unclear, however, exactly how much is gained when the statistician utilizes the symmetry structure of \( F_\mu \) and (correctly) employs \( \tilde{F}_{\mu,n} \) instead of \( \bar{F}_n \). Although \( \tilde{F}_{\mu,n} \) had been previously proposed by Dalal (1979a) as an estimator of \( F_\mu \), he did not study the risk of \( \tilde{F}_{\mu,n} \). In the following analysis we will investigate these problems by first deriving an exact expression for the Bayes risk of \( \tilde{F}_{\mu,n} \) and then, for a given \( \alpha \), comparing this risk to the risk (against the \( \mathcal{G} \)-invariant Dirichlet prior) of Ferguson's estimator \( \bar{F}_n \).

The parameter space is the set of all \( \mathcal{G} \)-invariant (symmetric) df's on \((\mathbb{R}, \mathcal{B})\) while the action space \( \mathcal{G} \) is the set of all df's on \((\mathbb{R}, \mathcal{B})\). When using \( F \in \mathcal{G} \) as an estimator of \( F_\mu \) the loss function is taken to be of the form 
\[ L(\tilde{F}, F) = \int (\tilde{F}(t) - F(t))^2 dW(t), \]
where \( W \) is a given finite measure on \((\mathbb{R}, \mathcal{B})\). Then by definition and interchanges of orders of integration, the Bayes risk of \( \tilde{F}_{\mu,n} \) with respect to the \( \mathcal{G} \)-invariant Dirichlet prior is given by

\[
R_{\mathcal{G}}(\tilde{F}_{\mu,n}, \alpha) = \mathbb{E}_{\mathcal{X}_n} \left[ \mathbb{E}_{\mathcal{F}_\mu} \left[ \frac{1}{n} \sum_{i=1}^{n} \left( \tilde{F}_{\mu,n}(t) - F_\mu(t) \right)^2 \right] dW(t) \right].
\]
In developing an explicit expression for \( R_{\mu,n}(\hat{\mu},n,\alpha) \), the Bayes risk of Dalal's estimator, and \( R_{\mu,n}(\hat{\mu},n,\alpha) \), the risk of Ferguson's estimator (against the prior for which \( \hat{\mu},n \) is Bayes), we use algebraic manipulations similar to those in Korwar and Hollander (1976). Thus we state these results without proof. Detailed proofs are given in Hannum (1979).

**Lemma 4.1.** Let \( G \) be the symmetry group on \((\mathbb{R},\mathbb{R})\) with the point of symmetry \( \mu \) known. Let \( P_{\mu} \in \mathcal{L}G(\alpha) \) on \((\mathbb{R},\mathbb{R})\), let \( F_{\mu}(t) = P_{\mu}((-\infty,t]} \), and let \( X_{n} = (X_1, \ldots, X_n) \) be a random sample of size \( n \) from \( P_{\mu} \). Then

1. \[ E_{X_{n} \mid \hat{X}_{n}}(F_{\mu}(t)) = \hat{F}_{\mu,n}(t), \]
2. \[ E_{X_{n} \mid \hat{X}_{n}}(P_{\mu}(t)) = \hat{F}_{\mu,n}(t)[(\hat{F}_{\mu,n}(t)\beta(\mu)+1)/(\beta(\mu)+1), \]
3. \[ E_{X}(F_{\mu}(t)) = F_{\mu}(t), \]

and

4. \[ E_{X_{n}}(\hat{P}_{\mu,n}(t)) = (F_{\mu}(t)/2 + (n-1)F_{0}(t)[F_{0}(t)\alpha(\mu) + 1/2]/(\alpha(\mu)+1))/n + ([F_{0}(t) - 1/2)/n]I_{t,\mu}, \]

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

Using Lemma 4.1 to evaluate the expectations in (4.1) yields
\[
R_g(\tilde{F}_{\mu,n}, \alpha) = \{\alpha(\alpha+1)/(\alpha(\alpha)+1)\} \int F_0(t)(1-F_0(t))dW(t)
+ \left\{ n/(\alpha(\alpha)+1)(\alpha(\alpha)+1) \right\} \left\{ (2(\alpha(\alpha)+1))^{-1} \int_{-\infty}^{\mu} F_0(t)dW(t) \right. \\
+ \left\{ [(n-\alpha(\alpha)-2)/2(\alpha(\alpha)+1)(\alpha(\alpha)+1)] \int_{\mu}^{\infty} F_0(t)dW(t) \right. \\
+ \left. [(2(\alpha(\alpha)+1))^{-1}W([\mu, \infty)) \right\}.
\]

(4.2)

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

\[
R_g(\tilde{F}_n, \alpha) = R_g(\tilde{F}_{\mu,n}, \alpha) + [n/2(\alpha(\alpha)+1)] \left\{ \int_{-\infty}^{\mu} F_0(t)dW(t) \right. \\
+ \left. \int_{\mu}^{\infty} (1-F_0(t))dW(t) \right\}.
\]

(4.3)

Thus we have an exact expression for the Bayes risk of the symmetrized estimator \( \tilde{F}_{\mu,n} \), given by (4.2), and an expression which relates this Bayes risk to the risk of the ordinary estimator \( \tilde{F}_n \) against the \( G \)-invariant prior, given by (4.3). 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, \( F_{\mu} \), is assumed to be symmetric about \( \mu \) we will henceforth assume that the weight function, \( W(\cdot) \), is also symmetric about \( \mu \). With this condition the right hand sides of (4.2) and (4.3) 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
\[ E_{\alpha, W}^n = R_{g}(\hat{F}_n, \alpha)/R_{g}(\hat{F}_n, n, \alpha) \]

\[ = 1 + \left\{ 2n(\alpha(\hat{R})+1)(\alpha(\hat{R})+n+1) \int_{-\infty}^{\mu} F_0(t)dw(t) \right\}/n(n-1) \]

\[ + [4\alpha(\hat{R})(\alpha(\hat{R})+n)(\alpha(\hat{R})+n+1) + 2n(\alpha(\hat{R})+1)] \int_{-\infty}^{\mu} F_0(t)dw(t) \]

\[ - [4\alpha(\hat{R})(\alpha(\hat{R})+n)(\alpha(\hat{R})+n+1)] \int_{-\infty}^{\mu} F_0^2(t)dw(t) \],

where \( w(\cdot) = W(\cdot)/W(\hat{R}) \) is the normalized version of \( W(\cdot) \). Using (4.4) we obtain:

**Theorem 4.2.** Let \( W \) be a nondegenerate finite measure on \((R, \mathcal{B})\) and assume that \{support of \( W \)\} \subseteq \{support of \( \alpha \)\}. Define \( w(\cdot) = W(\cdot)/W(\hat{R}) \),

and let \( E_{\alpha, W}^n = R_{g}(\hat{F}_n, \alpha)/R_{g}(\hat{F}_n, n, \alpha) \) where \( R_{g}(\hat{F}_n, \alpha) \) and \( R_{g}(\hat{F}_n, n, \alpha) \) represent the risk against the \( G \)-invariant Dirichlet prior of Ferguson's \( \hat{F}_n \) and Dalal's symmetrized \( \hat{F}_n \) respectively. Then

(i) if \( n > 1 \), then \( E_{\alpha, W}^n > 1 \); if \( n = 0 \), then \( E_{\alpha, W}^n = 1 \),

(ii) \( \lim_{\alpha(\hat{R}) \to \infty} E_{\alpha, W}^n = 1 \), for \( n(\neq 0) \) fixed,

and

(iii) \( \lim_{n \to \infty} E_{\alpha, W}^n = l + D_{\alpha, W} \), for \( c(\hat{R})(\neq 0) \) fixed, where

\[ D_{\alpha, W} = \left[ (\alpha(\hat{R})+1) \int_{-\infty}^{\mu} F_0(t)dw(t) \right]/\left[ 2\alpha(\hat{R}) \int_{-\infty}^{\mu} F_0(t)(1-F_0(t))dw(t) \right]. \]

**Proof.** See Hannum (1979).
We briefly mention some implications of Theorem 4.2. Part (i) says that \( \tilde{F}_{\mu,n} \) is always as good, and usually better, than \( \tilde{F}_n \). Of course this is because \( \tilde{F}_{\mu,n} \) is Bayes. Note that when \( n = 0 \), \( \tilde{F}_{\mu,0} = \tilde{F}_0 \). Part (ii) sheds light on the performance of \( \tilde{F}_n \) compared to that of \( \tilde{F}_{\mu,n} \) when \( n \) is small relative to \( \alpha(\theta) \). It is clear from the forms of \( \tilde{F}_n \) and \( \tilde{F}_{\mu,n} \) given in (3.3) and (3.6) respectively that when \( n \) is small relative to \( \alpha(\theta) \), both estimators give much weight to the prior guess at \( F_\mu \) which, for a given (symmetric) \( \alpha \), 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 \( \alpha(\theta) \) 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 \( F_\mu \) instead of Ferguson's \( \tilde{F}_n \) in the case that \( n \) is small relative to \( \alpha(\theta) \). This is reflected by Theorem 4.2 in the fact that when \( n \) is fixed, \( \frac{R_g(\tilde{F}_n, \alpha)}{R_g(\tilde{F}_{\mu,n}, \alpha)} \to 1 \) as \( \alpha(\theta) \to 0 \). These remarks support an interpretation by Ferguson (1973), later advanced from a different viewpoint by Korwar and Hollander (1976), that \( \alpha(\theta) \) be viewed as the "prior sample size" of the process. (But see Tiwari (1980), for a probabilistic context in which the case \( \alpha(\theta) \to 0 \) corresponds to much "information".) If one has a great amount of faith in the prior guess, \( F_0 \), then \( \alpha(\theta) \) 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 \( F_\mu \) since both will have risk values against the \( G \)-invariant Dirichlet prior near \( R_g(F_\mu, \alpha) \). 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 \( \alpha(\theta) \)), the sample size \( n \), and the exact weight, \( W(\cdot) \), which is used in the loss.
function (including the size of $W(\mathcal{G})$). In order to get some idea of how close $R_\mathcal{G} (\tilde{F}_n, \alpha)$ can be to $R_\mathcal{G} (\tilde{F}_{\mu_n}, \alpha)$ in the case that $\alpha(\mathcal{G})$ is large relative to $n$, the reader is referred to Tables 4.1 through 4.6 which list values of $E^n_{\alpha, W}$ for some particular choices of $\alpha$ and $W$, and for selected values of $\alpha(\mathcal{G})$ and $n$. These tables will also illustrate the extent to which $\tilde{F}_{\mu_n}$ performs better than $\tilde{F}_n$ when $\alpha(\mathcal{G})$ 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 $R_\mathcal{G} (\tilde{F}_n, \alpha)$ and $R_\mathcal{G} (\tilde{F}_{\mu_n}, \alpha)$ may, depending on the particular $\alpha$ and $W$ chosen, be quite large. Thus when $\alpha(\mathcal{G})$ 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 4.3.** Let $W(\cdot)$ be the uniform measure on the interval $\{[-c, c]\}$ for some $c > 0$, and let $F_0(\cdot) = \alpha(\cdot)/\alpha(\mathcal{G})$ be a double exponential distribution centered at zero with scale parameter $\lambda > 0$. In this case (4.4) becomes

$$E^n_{\alpha, W} = 1 + \frac{2n(\alpha(\mathcal{G})+1)(\alpha(\mathcal{G})+n+1)(1-e^{-\lambda c})}{4\lambda c n(n-1)}$$

$$+ (1-e^{-\lambda c})[4\alpha(\mathcal{G})(\alpha(\mathcal{G})+n)(\alpha(\mathcal{G})+n+1) + 2n(\alpha(\mathcal{G})+1)]$$

$$- (1-e^{-2\lambda c})[\alpha(\mathcal{G})(\alpha(\mathcal{G})+n)(\alpha(\mathcal{G})+n+1)]].$$

In studying the behavior of (4.5) for different values of $\lambda$ and $c$ we note that these two parameters enter expression (4.5) only in the form $\lambda \cdot c$. Thus we denote $\theta = \lambda c$ and consider the behavior of $E^n_{\alpha, W}$ as a function of $\theta$. Standard derivative arguments yield Proposition 4.4.
PROPOSITION 4.4. If $W(\cdot)$ is uniform on $\{-c, c\}$ and if $P_0(\cdot)$ is double exponential centered at zero with scale parameter $\lambda > 0$, then $E^n_{\alpha, W}$ is given by (4.5). Further, if $\theta = \lambda c$, then

1. $E^n_{\alpha, W}$ is strictly monotonically decreasing in $\theta$,

2. $\lim_{\theta \to \infty} E^n_{\alpha, W} = 1$,

and

3. $\lim_{\theta \to 0} E^n_{\alpha, W} = 1 + \frac{2n(\alpha(\tilde{R})+1)(\alpha(\tilde{R})+n+1)}{\ln(n-1)} + 2[\alpha(\tilde{R})(\alpha(\tilde{R})+n)(\alpha(\tilde{R})+n+1) + n(\alpha(\tilde{R})+1)]$.

Note that by Proposition 4.4,

\[
1 < E^n_{\alpha, W} < 1 + \frac{2n(\alpha(\tilde{R})+1)(\alpha(\tilde{R})+n+1)}{\ln(n-1)} + 2[\alpha(\tilde{R})(\alpha(\tilde{R})+n)(\alpha(\tilde{R})+n+1) + n(\alpha(\tilde{R})+1)] .
\]

Expression (4.6) coupled with 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 distribution function a rough guide as to how to choose an appropriate value for the scale parameter $\lambda$, depending on his degree of belief that the underlying distribution really is symmetric. For a given value of $c$, the value of $E^n_{\alpha, W}$ is smallest when $\lambda$ is very large and the value of $E^n_{\alpha, W}$ is largest for $\lambda$ near zero. The value

\[
\sup_{\theta} E^n_{\alpha, W} = 1 + \frac{2n(\alpha(\tilde{R})+1)(\alpha(\tilde{R})+n+1)}{\ln(n-1)} + 2[\alpha(\tilde{R})(\alpha(\tilde{R})+n)(\alpha(\tilde{R})+n+1) + n(\alpha(\tilde{R})+1)]
\]

can be considered an upper bound on the gain in overall performance (measured by the risk with respect to the $G$-invariant Dirichlet prior) of the symmetrized estimator $\tilde{P}_{\mu, n}$ relative to that of Ferguson's
Thus for example, if one is quite certain that $F_\mu$ is symmetric (about zero) it would be appropriate to choose a relatively small scale parameter $\lambda$ in the double exponential prior, since it is in this situation (when $\theta$ is small) that the symmetrized estimator $\tilde{F}_{\mu,n}$ performs best against the usual (Ferguson) estimator $\tilde{F}_n$. Because of part (i) of Proposition 4.4, the smaller the value of $\lambda$ (for a fixed $c$) the larger the value of $E^{n}_{\alpha,W}$. On the other hand, choosing a relatively large value of $\lambda$ would be appropriate when the degree of belief in the assumption of symmetry is not as strong. To get an idea of the best possible increase in performance of $\tilde{F}_{\mu,n}$ relative to that of $\tilde{F}_n$, measured by the ratio of their risks against the $G$-invariant Dirichlet prior, the reader is referred to Table 4.1 which lists the values of $\sup_{\theta} E^{n}_{\alpha,W}$ for selected choices of $n$ and $\alpha(\mathcal{R})$. For purposes of comparison we have provided an analogous listing of values of $E^{n}_{\alpha,W}$ for the case that $\theta = \lambda c = 1$ in Table 4.2.

**EXAMPLE 4.5.** In this example we assume only that $W(\cdot)$ and $\alpha(\cdot)$ are the same measure. This will entail both that $\alpha([-\infty,t]) = W([-\infty,t])$ for all $t \in \mathcal{R}$ and that $\alpha(\mathcal{R}) = W(\mathcal{R})$. Then (1.4) reduces to

$$E^{n}_{\alpha,W} = 1 + [3n(\alpha(\mathcal{R})+1)(\alpha(\mathcal{R})+n+1)/12n(n-1) + 3n(\alpha(\mathcal{R})+1)]$$

$$+ \lambda \alpha(\mathcal{R})(\alpha(\mathcal{R})+n)(\alpha(\mathcal{R})+n+1]).$$

Table 4.3 provides a list of values of $E^{n}_{\alpha,W}$ given by (4.7) for various choices of $n$ and $\alpha(\mathcal{R})$. 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 value of $\alpha(\mathcal{R})$ (and $W(\mathcal{R})$). As expected, Table 4.3 shows
TABLE 4.1

Selected values of $\sup_\theta E_\alpha^W$ when $W$ is uniform $[-c,c]$ and $\alpha$ is double exponential $(\lambda)$, where $\theta = \lambda c$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\alpha(\lambda)$</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>25</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>3.0000</td>
<td>2.0000</td>
<td>1.6667</td>
<td>1.5789</td>
<td>1.5306</td>
<td>1.5000</td>
</tr>
<tr>
<td>2</td>
<td>0.01</td>
<td>2.9704</td>
<td>1.9999</td>
<td>1.6692</td>
<td>1.5816</td>
<td>1.5332</td>
<td>1.5025</td>
</tr>
<tr>
<td>3</td>
<td>0.1</td>
<td>2.7355</td>
<td>1.9955</td>
<td>1.6902</td>
<td>1.6038</td>
<td>1.5551</td>
<td>1.5238</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1.7500</td>
<td>1.8000</td>
<td>1.7609</td>
<td>1.7229</td>
<td>1.6916</td>
<td>1.6667</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1.4444</td>
<td>1.6000</td>
<td>1.7185</td>
<td>1.7471</td>
<td>1.7535</td>
<td>1.7500</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>1.2184</td>
<td>1.3243</td>
<td>1.5323</td>
<td>1.6667</td>
<td>1.7750</td>
<td>1.8571</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>1.0992</td>
<td>1.1803</td>
<td>1.3527</td>
<td>1.5145</td>
<td>1.7034</td>
<td>1.9167</td>
</tr>
<tr>
<td>8</td>
<td>25</td>
<td>1.3553</td>
<td>1.4338</td>
<td>1.5153</td>
<td>1.5728</td>
<td>1.6630</td>
<td>1.9630</td>
</tr>
</tbody>
</table>

TABLE 4.2

Selected values of $E_\alpha^W$ when $W$ is uniform $[-c,c]$ and $\alpha$ is double exponential $(\lambda)$, with $\theta = \lambda c = 1$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\alpha(\lambda)$</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>25</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>3.0000</td>
<td>1.7205</td>
<td>1.4394</td>
<td>1.3732</td>
<td>1.3379</td>
<td>1.3161</td>
</tr>
<tr>
<td>2</td>
<td>0.01</td>
<td>2.9582</td>
<td>1.7215</td>
<td>1.4416</td>
<td>1.3753</td>
<td>1.3399</td>
<td>1.3179</td>
</tr>
<tr>
<td>3</td>
<td>0.1</td>
<td>2.6453</td>
<td>1.7267</td>
<td>1.4604</td>
<td>1.3932</td>
<td>1.3567</td>
<td>1.3338</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1.6663</td>
<td>1.6125</td>
<td>1.6223</td>
<td>1.5016</td>
<td>1.4700</td>
<td>1.4464</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1.3470</td>
<td>1.4617</td>
<td>1.5317</td>
<td>1.5377</td>
<td>1.5299</td>
<td>1.5176</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>1.1487</td>
<td>1.2482</td>
<td>1.2238</td>
<td>1.4990</td>
<td>1.5693</td>
<td>1.6157</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>1.0755</td>
<td>1.1374</td>
<td>1.1442</td>
<td>1.3901</td>
<td>1.5278</td>
<td>1.6738</td>
</tr>
<tr>
<td>8</td>
<td>25</td>
<td>1.0304</td>
<td>1.0584</td>
<td>1.1309</td>
<td>1.2229</td>
<td>1.3835</td>
<td>1.7209</td>
</tr>
</tbody>
</table>

19
<table>
<thead>
<tr>
<th>n</th>
<th>3.0000</th>
<th>1.6000</th>
<th>1.4444</th>
<th>1.3529</th>
<th>1.2300</th>
<th>1.2973</th>
<th>1.2807</th>
<th>1.2680</th>
<th>1.2680</th>
<th>1.2680</th>
<th>1.2680</th>
<th>1.2680</th>
<th>1.2680</th>
<th>1.2680</th>
<th>1.2680</th>
</tr>
</thead>
<tbody>
<tr>
<td>α(μ) = 0.0</td>
<td>2.9957</td>
<td>1.6002</td>
<td>1.4447</td>
<td>1.3532</td>
<td>1.2302</td>
<td>1.2975</td>
<td>1.2809</td>
<td>1.2682</td>
<td>1.2682</td>
<td>1.2682</td>
<td>1.2682</td>
<td>1.2682</td>
<td>1.2682</td>
<td>1.2682</td>
<td>1.2682</td>
</tr>
<tr>
<td>α(μ) = 0.01</td>
<td>2.9575</td>
<td>1.6020</td>
<td>1.4469</td>
<td>1.3552</td>
<td>1.3221</td>
<td>1.2993</td>
<td>1.2826</td>
<td>1.2698</td>
<td>1.2698</td>
<td>1.2698</td>
<td>1.2698</td>
<td>1.2698</td>
<td>1.2698</td>
<td>1.2698</td>
<td>1.2698</td>
</tr>
<tr>
<td>α(μ) = 0.1</td>
<td>2.6406</td>
<td>1.6162</td>
<td>1.4666</td>
<td>1.3742</td>
<td>1.3401</td>
<td>1.3164</td>
<td>1.2988</td>
<td>1.2854</td>
<td>1.2854</td>
<td>1.2854</td>
<td>1.2854</td>
<td>1.2854</td>
<td>1.2854</td>
<td>1.2854</td>
<td>1.2854</td>
</tr>
<tr>
<td>α(μ) = 1</td>
<td>1.6000</td>
<td>1.5714</td>
<td>1.5294</td>
<td>1.4795</td>
<td>1.4532</td>
<td>1.4317</td>
<td>1.4137</td>
<td>1.3987</td>
<td>1.3987</td>
<td>1.3987</td>
<td>1.3987</td>
<td>1.3987</td>
<td>1.3987</td>
<td>1.3987</td>
<td>1.3987</td>
</tr>
<tr>
<td>α(μ) = 2</td>
<td>1.3429</td>
<td>1.4455</td>
<td>1.4779</td>
<td>1.4911</td>
<td>1.4895</td>
<td>1.4839</td>
<td>1.4762</td>
<td>1.4676</td>
<td>1.4676</td>
<td>1.4676</td>
<td>1.4676</td>
<td>1.4676</td>
<td>1.4676</td>
<td>1.4676</td>
<td>1.4676</td>
</tr>
<tr>
<td>α(μ) = 3</td>
<td>1.2381</td>
<td>1.3529</td>
<td>1.4118</td>
<td>1.4639</td>
<td>1.4843</td>
<td>1.4965</td>
<td>1.5026</td>
<td>1.5045</td>
<td>1.5045</td>
<td>1.5045</td>
<td>1.5045</td>
<td>1.5045</td>
<td>1.5045</td>
<td>1.5045</td>
<td>1.5045</td>
</tr>
<tr>
<td>α(μ) = 5</td>
<td>1.1469</td>
<td>1.2441</td>
<td>1.3103</td>
<td>1.3913</td>
<td>1.4368</td>
<td>1.4752</td>
<td>1.5067</td>
<td>1.5314</td>
<td>1.5314</td>
<td>1.5314</td>
<td>1.5314</td>
<td>1.5314</td>
<td>1.5314</td>
<td>1.5314</td>
<td>1.5314</td>
</tr>
<tr>
<td>α(μ) = 10</td>
<td>1.0745</td>
<td>1.1355</td>
<td>1.1860</td>
<td>1.2639</td>
<td>1.3205</td>
<td>1.3806</td>
<td>1.4436</td>
<td>1.5083</td>
<td>1.5083</td>
<td>1.5083</td>
<td>1.5083</td>
<td>1.5083</td>
<td>1.5083</td>
<td>1.5083</td>
<td>1.5083</td>
</tr>
<tr>
<td>α(μ) = 15</td>
<td>1.0499</td>
<td>1.0935</td>
<td>1.1319</td>
<td>1.1963</td>
<td>1.2477</td>
<td>1.3077</td>
<td>1.3780</td>
<td>1.4607</td>
<td>1.4607</td>
<td>1.4607</td>
<td>1.4607</td>
<td>1.4607</td>
<td>1.4607</td>
<td>1.4607</td>
<td>1.4607</td>
</tr>
<tr>
<td>α(μ) = 25</td>
<td>1.0300</td>
<td>1.0576</td>
<td>1.0833</td>
<td>1.1291</td>
<td>1.1689</td>
<td>1.2196</td>
<td>1.2861</td>
<td>1.3765</td>
<td>1.3765</td>
<td>1.3765</td>
<td>1.3765</td>
<td>1.3765</td>
<td>1.3765</td>
<td>1.3765</td>
<td>1.3765</td>
</tr>
<tr>
<td>α(μ) = 100</td>
<td>1.0075</td>
<td>1.0149</td>
<td>1.0221</td>
<td>1.0361</td>
<td>1.0495</td>
<td>1.0688</td>
<td>1.0987</td>
<td>1.1511</td>
<td>1.1511</td>
<td>1.1511</td>
<td>1.1511</td>
<td>1.1511</td>
<td>1.1511</td>
<td>1.1511</td>
<td>1.1511</td>
</tr>
</tbody>
</table>

that when \( \alpha(\mu) \) 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 \( \text{En}_{\alpha,\tilde{W}} \) in the lower left part of the table. However, \( \text{En}_{\alpha,\tilde{W}} \) is not monotone in \( \alpha(\mu) \) in this example so that in some cases an increase in the value of \( \alpha(\mu) \) will not necessarily result in an improved performance of \( \tilde{F}_n \) compared to \( \tilde{F}_{\mu,n} \). Also note that in this example

\[
\lim_{n \to \infty} \text{En}_{\alpha,\tilde{W}} = 1 + \{(3(\alpha(\mu)+1)/(4(\alpha(\mu)+3)) \},
\]

agreeing with result (iii) of Theorem 4.2 since in this example, \( \int_{-\infty}^{\mu} F_{\tilde{m}}(t) dt(t) = 3^{-1} \) and
\[ \int_{-\infty}^{\infty} F_0(t)(1-F_0(t))\,dT = (8^{-1} - 2^{-1}) = 1 e^{-1}. \] Finally, expression (4.7) yields that \( \lim_{\alpha(\bar{R}) \to 0} \mu_{n, n} = 1 + \{(n+1)/(ln-3)\} \) so that values in the upper left portion of Table 4.3 (when \( \alpha(\bar{R}) \) is small compared to \( n \)) will be near \( 1 + \{(n+1)/(ln-3)\} \).

**Example 4.6.** Let \( \omega(\cdot) \) be the normal probability distribution with mean zero and variance \( \sigma^2 \), and let \( F_0(\cdot) \) be the double exponential distribution centered at zero with scale parameter \( \lambda > 0 \). Then (4.4) becomes

\[
E_n^{\alpha, \omega} = 1 + \{n(\alpha(\bar{R})+1)(\alpha(\bar{R})+n+1)\}^{1/2} \omega^2 \Phi(-\omega) / \{n(n-1)
\]

\[
+ [2\alpha(\bar{R})(\alpha(\bar{R})+n)(\alpha(\bar{R})+n+1) + n(\alpha(\bar{R})+1)]e^{3/2} \omega^2 \Phi(-\omega)
\]

\[- [\alpha(\bar{R})(\alpha(\bar{R})+n)(\alpha(\bar{R})+n+1)]e^{2\omega^2} \Phi(-2\omega) \],
\]

where \( \omega = \sigma \lambda \), and \( \Phi(\cdot) \) denotes the standard normal df. From (4.8) it follows that

\[ \lim_{\omega \to 0} E_n^{\alpha, \omega} = 1 + \{n(\alpha(\bar{R})+1)(\alpha(\bar{R})+n+1)\}/\{2n(n-1)
\]

\[ + \alpha(\bar{R})(\alpha(\bar{R})+n+1)(\alpha(\bar{R})+n) + n(\alpha(\bar{R})+1) \],
\]

which sheds some light on the behavior of \( E_n^{\alpha, \omega} \) for values of \( \omega \) near zero. Table 4.4 gives selected values of \( \lim_{\omega \to 0} E_n^{\alpha, \omega} \), Table 4.5 gives selected values of \( E_n^{\alpha, \omega} \) when \( \omega = 0.5 \), and Table 4.6 gives selected values of \( E_n^{\alpha, \omega} \) when \( \omega = 1.0 \).
TABLE 4.4

Selected values of \( \lim_{\omega \to 0} E_n^{\alpha, \omega} \) when \( \omega \) is normal \((\mu = 0, \sigma^2)\) and \( \alpha \) is double exponential \((\lambda)\), where \( \omega = \sigma \lambda \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \alpha(\omega) = 0.00 )</th>
<th>( \alpha(\omega) = 0.01 )</th>
<th>( \alpha(\omega) = 0.1 )</th>
<th>( \alpha(\omega) = 1 )</th>
<th>( \alpha(\omega) = 2 )</th>
<th>( \alpha(\omega) = 5 )</th>
<th>( \alpha(\omega) = 10 )</th>
<th>( \alpha(\omega) = 25 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.0000</td>
<td>2.9704</td>
<td>2.7355</td>
<td>1.7500</td>
<td>1.4444</td>
<td>1.1981</td>
<td>1.0992</td>
<td>1.0399</td>
</tr>
<tr>
<td>2</td>
<td>2.0000</td>
<td>1.9999</td>
<td>1.9955</td>
<td>1.8000</td>
<td>1.6000</td>
<td>1.3333</td>
<td>1.1803</td>
<td>1.0768</td>
</tr>
<tr>
<td>5</td>
<td>1.6667</td>
<td>1.6692</td>
<td>1.6902</td>
<td>1.7639</td>
<td>1.7136</td>
<td>1.7181</td>
<td>1.3527</td>
<td>1.1721</td>
</tr>
<tr>
<td>10</td>
<td>1.5789</td>
<td>1.5816</td>
<td>1.6038</td>
<td>1.7229</td>
<td>1.7471</td>
<td>1.7481</td>
<td>1.5145</td>
<td>1.2930</td>
</tr>
<tr>
<td>25</td>
<td>1.5306</td>
<td>1.5332</td>
<td>1.5551</td>
<td>1.6916</td>
<td>1.7535</td>
<td>1.7881</td>
<td>1.7034</td>
<td>1.5053</td>
</tr>
<tr>
<td>( \infty )</td>
<td>1.5000</td>
<td>1.5025</td>
<td>1.5238</td>
<td>1.6667</td>
<td>1.7500</td>
<td>1.8571</td>
<td>1.9167</td>
<td>1.9630</td>
</tr>
</tbody>
</table>

TABLE 4.5

Selected values of \( E_n^{\alpha, \omega} \) when \( \omega \) is normal \((\mu = 0, \sigma^2)\) and \( \alpha \) is double exponential \((\lambda)\), where \( \omega = \sigma \lambda = .5 \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \alpha(\omega) = 0.00 )</th>
<th>( \alpha(\omega) = 0.01 )</th>
<th>( \alpha(\omega) = 0.1 )</th>
<th>( \alpha(\omega) = 1 )</th>
<th>( \alpha(\omega) = 2 )</th>
<th>( \alpha(\omega) = 5 )</th>
<th>( \alpha(\omega) = 10 )</th>
<th>( \alpha(\omega) = 25 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.0000</td>
<td>2.9607</td>
<td>2.6630</td>
<td>1.6311</td>
<td>1.3633</td>
<td>1.1563</td>
<td>1.0794</td>
<td>1.0221</td>
</tr>
<tr>
<td>2</td>
<td>1.7780</td>
<td>1.3894</td>
<td>1.3908</td>
<td>1.6470</td>
<td>1.4857</td>
<td>1.2608</td>
<td>1.1414</td>
<td>1.0614</td>
</tr>
<tr>
<td>5</td>
<td>1.4829</td>
<td>1.0970</td>
<td>1.1099</td>
<td>1.5849</td>
<td>1.5653</td>
<td>1.4256</td>
<td>1.2825</td>
<td>1.1376</td>
</tr>
<tr>
<td>10</td>
<td>1.4119</td>
<td>1.0414</td>
<td>1.0433</td>
<td>1.5427</td>
<td>1.5757</td>
<td>1.5279</td>
<td>1.4108</td>
<td>1.2344</td>
</tr>
<tr>
<td>25</td>
<td>1.3739</td>
<td>1.0150</td>
<td>1.0157</td>
<td>1.5110</td>
<td>1.5706</td>
<td>1.6053</td>
<td>1.5576</td>
<td>1.4036</td>
</tr>
<tr>
<td>( \infty )</td>
<td>1.3501</td>
<td>1.3521</td>
<td>1.3690</td>
<td>1.4869</td>
<td>1.5598</td>
<td>1.6584</td>
<td>1.7157</td>
<td>1.7616</td>
</tr>
</tbody>
</table>
### TABLE 4.6

Selected values of $E_{a,W}^n$ when $W$ is normal ($\mu = 0$, $\sigma^2$) and $a$ is double exponential ($\lambda$), where $\omega = a\lambda = 1$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$n = 1$</th>
<th>$n = 2$</th>
<th>$n = 5$</th>
<th>$n = 10$</th>
<th>$n = 25$</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha(\hat{a}) = 0.00$</td>
<td>3.0000</td>
<td>1.6231</td>
<td>1.3690</td>
<td>1.3113</td>
<td>1.2809</td>
<td>1.2621</td>
</tr>
<tr>
<td>$\alpha(\hat{a}) = 0.01$</td>
<td>2.9568</td>
<td>1.6248</td>
<td>1.3713</td>
<td>1.3133</td>
<td>1.2827</td>
<td>1.2638</td>
</tr>
<tr>
<td>$\alpha(\hat{a}) = 0.1$</td>
<td>2.6357</td>
<td>1.6368</td>
<td>1.3901</td>
<td>1.3305</td>
<td>1.2985</td>
<td>1.2785</td>
</tr>
<tr>
<td>$\alpha(\hat{a}) = 1$</td>
<td>1.5935</td>
<td>1.5745</td>
<td>1.4893</td>
<td>1.4430</td>
<td>1.4106</td>
<td>1.3871</td>
</tr>
<tr>
<td>$\alpha(\hat{a}) = 2$</td>
<td>1.3386</td>
<td>1.4431</td>
<td>1.4945</td>
<td>1.4905</td>
<td>1.4765</td>
<td>1.4603</td>
</tr>
<tr>
<td>$\alpha(\hat{a}) = 5$</td>
<td>1.1449</td>
<td>1.2411</td>
<td>1.3894</td>
<td>1.4740</td>
<td>1.5330</td>
<td>1.5675</td>
</tr>
<tr>
<td>$\alpha(\hat{a}) = 10$</td>
<td>1.0735</td>
<td>1.1336</td>
<td>1.2607</td>
<td>1.3768</td>
<td>1.5052</td>
<td>1.6347</td>
</tr>
<tr>
<td>$\alpha(\hat{a}) = 25$</td>
<td>1.0296</td>
<td>1.0569</td>
<td>1.1274</td>
<td>1.2168</td>
<td>1.3720</td>
<td>1.6914</td>
</tr>
</tbody>
</table>

### REFERENCES


Robustness of Ferguson's Bayes Estimator of a Distribution Function

Robert Hannum and Myles Hollander

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

The U.S. Air Force
Air Force Office of Scientific Research
Bolling Air Force Base, D.C. 20332

Approved for public release; distribution unlimited.

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 G-invariant Dirichlet process prior. We compare this risk to the risk of Ferguson's estimator of an arbitrary distribution under the 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.