EMPIRICAL BAYES ESTIMATION OF A DISTRIBUTION FUNCTION

by

Ramesh M. Korwar and Myles Hollander

FSU Statistics Report M288
AFOSR Technical Report No. 24

March, 1974
The Florida State University
Department of Statistics
Tallahassee, Florida

\(^1\)Research supported by the Office of Naval Research Contract No. NONR-998(08) Task Order NR 042-004, by Air Force of Scientific Research Grant No. AFOSR-74-2581, and NRC grant A-30-6A. The United States Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation hereon.
Empirical Bayes Estimation of a Distribution Function

by

Ramesh M. Korwar and Myles Hollander
University of Alberta and Florida State University

Abstract

Let \((P_i, X_i), i = 1, 2, \ldots\) be a sequence of independent pairs of random variables, the \(P_i\) having a common prior distribution given by a Dirichlet process (Ferguson, Ann. Statist. 1 (1973) 209-230) on \((R, \mathcal{B})\) with a parameter \(\alpha\) (assumed \(\sigma\)-additive), and given \(P_i = P, X_i = (X_{i1}, \ldots, X_{im})\) is a sample of size \(m\) from \(P\). Consider the problem of estimating \(F_{n+1}(t) = P_{n+1}((-\infty, t])\) on the basis of \(X_1, \ldots, X_{n+1}\). Assume \(\alpha\) is non-atomic, \(\alpha(R)\) is known, and take the loss in estimating \(F(t)\) by \(a(t)\) to be \(\int (F(t) - a(t))^2 dW(t)\) where \(W\) is a weight function. Define \(G_n(t) = \frac{1}{n} \sum_{i=1}^{n} \hat{F}_i(t) + (1 - p_m) \hat{F}_{n+1}(t)\) where \(p_m = [\alpha(R)/(\alpha(R) + m)]\) and \(\hat{F}_i\) is the empirical distribution function for the sample \(X_{i1}, \ldots, X_{im}\). Then the sequence \(\{G_n\}\) is shown to be asymptotically optimal in the sense of Robbins (cf. Ann. Math. Statist. 35 (1964) 1-20). The performance of \(G_n\) is compared with that of a corresponding Bayes estimator derived by Ferguson, and with the sample distribution function \(\hat{F}_{n+1}\). The empirical Bayes approach based on the Dirichlet process is also applied to the problem of estimating the mean of a distribution.
EMPIRICAL BAYES ESTIMATION OF A DISTRIBUTION FUNCTION

Ramesh M. Korwar and Myles Hollander

University of Alberta and Florida State University

1. Introduction. Given a sample \( X_1, \ldots, X_n \) of size \( n \) from an unknown distribution, consider the problem of estimating the distribution function. Ferguson [9] obtains a solution to this problem by deriving an estimator that is Bayes with respect to a Dirichlet process prior. Specifically, he takes \( X_1, \ldots, X_n \) to be a sample from a distribution which is chosen by a Dirichlet process. (We assume the reader is familiar with Ferguson's basic paper that introduces the Dirichlet process. Related work includes that of Antoniak [2], Blackwell [3], Blackwell and MacQueen [4], Doksum [6], Korwar and Hollander [10] and Savage [14].) The parameter \( \alpha \) of the process, is assumed to be \( \sigma \)-additive. Let the parameter and the action spaces be the set of all the distributions \( P \) on \((R,B)\), the measurable space of the real line \( R \) and the \( \sigma \)-field of Borel subsets of \( R \). Take the loss function to be

\[
L(P, \hat{F}) = \int_{R} (F(t) - \hat{F}(t))^2 dW(t),
\]

where \( W \) is given finite measure on \((R,B)\) (a weight function) and where \( F(t) = P((\infty, t]) \) and \( \hat{F} \) is an estimator of \( F \). When \( P \) is a Dirichlet process on \((R,B)\) with parameter \( \alpha \), the Bayes estimator of \( F \), which was derived by Ferguson, is given by

\[
\hat{F}_n(t|X_1, \ldots, X_n) = \frac{1}{n} F_0(t) + \frac{1}{n} \hat{F}_n(t|X_1, \ldots, X_n),
\]

\(1\) Research supported by the Office of Naval Research Contract No. NONR-998(08) Task Order NR 042-004, by Air Force of Scientific Research Grant No. AFOSR-74-2501, and by NRC grant A-30-6A. The United States Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation hereon.
(1.3) \( p_n = a(R)/(a(R)+n) \),

(1.4) \( F_0(t) = a((-\infty, t])/a(R) \),

and \( \hat{F}_n(t|X_1, \ldots, X_n) = (# \text{ of } X_i \text{'s } \leq t)/n \) is the empirical distribution function of the sample \( X_1, \ldots, X_n \). The Bayes estimator given by (1.2) is a weighted average of \( F_0 \), the prior "guess" at \( F \), and the sample distribution function \( \hat{F}_n(t|X_1, \ldots, X_n) \). Ferguson notes this brings out the role of \( a(R) \) as a measure of faith in the prior. In order to use (1.2) as an estimator of \( F \) one must specify \( a(\cdot) \) completely.

In this paper we propose and investigate an empirical Bayes estimator which requires less information about \( a \). One need only specify \( a(R) \). Furthermore, the estimator is shown to be asymptotically optimal in the sense of Robbins [13].

Consider the following framework for the empirical Bayes estimation problem. Let \( (P_i, X_i), i = 2, \ldots \) be a sequence of pairs of independent random elements. We assume that the probability measures \( P_i \) have a common prior distribution given by a Dirichlet process on \((R, B)\) with parameter \( a \) (assumed \( \sigma \)-additive), and given \( P_i = P, X_i = (X_{i1}, \ldots, X_{im}) \) is a random sample of size \( m \) from \( P \). Consider the problem of estimating \( F_{n+1}(t) = P_{n+1}((-\infty, t]) \) on the basis of \( X_1, \ldots, X_{n+1} \). Assume \( a(R) \) is known. Let the loss function be given by (1.1), and let the parameter and action spaces be the set of all distributions \( P \) on \((R, B)\). Define, for \( n = 1, 2, \ldots \), the sequence of estimators \( G_n \) of \( F \) by

(1.5) \( G_n(t) = p_m \sum_{i=1}^{n} \hat{F}_i(t)/n + (1-p_m) \hat{F}_{n+1}(t) \),
where \( p_m \) is given by (1.3) and \( \hat{F}_i \) is the sample distribution function of \( X_i, i = 1, \ldots, n+1 \). We propose \( G_n \) as an empirical Bayes estimator of \( F_{n+1} \).

In Section 2 we establish the asymptotic optimality of
\[ G = \{ G_n \}. \]
Exact risk expressions are given. The rate at which the overall expected loss, incurred by using \( G_n \), approaches the minimum Bayes risk is exhibited.

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

To the best of our knowledge, this paper represents the first use of Ferguson's Dirichlet process in an empirical Bayes context.

The results in Sections 2 and 3 show that empirical Bayes estimation of a distribution function, via the Dirichlet process, is useful (\( G_n \) requires substantially less prior information than Ferguson's Bayes estimator and—riskwise—may be preferred to \( \hat{F}_{n+1} \) and tractable (the estimator is readily computed and exact expressions for its risk are obtained). This suggests 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 \( G_n \). Theorems 2.1 and 2.2 below are used repeatedly in the sequel. Theorem 2.1 is Theorem 1 of Ferguson [9]. Theorem 2.2 is a direct generalization of Ferguson's [9] Proposition 4; its proof is omitted.
THEOREM 2.1 (Ferguson). Let $P$ be a Dirichlet process on $(X, A)$ with parameter $\alpha$, and let $X_1, \ldots, X_n$ be a sample of size $n$ from $P$. Then the conditional distribution of $P$ given $X_1, \ldots, X_n$ is a Dirichlet process on $(X, A)$ with parameter $\beta = \alpha + \sum_{i=1}^{n} \delta_{X_i}$, where, for $x \in X$, $A \in A$, $\delta_x(A) = 1$ if $x \in A$, 0 otherwise.

THEOREM 2.2. Let $P$ be a Dirichlet process on $(R, B)$ with parameter $\alpha$. Assume $\alpha$ to be countably additive, and let $X_1, \ldots, X_n$ be a sample of size $n$ from $P$. Then

$$Q(X_1 \leq x_1, \ldots, X_n \leq x_n) = (\alpha(A_x(1)) \ldots (\alpha(A_x(n)) + n-1))/\{\alpha(R) \ldots (\alpha(R)+n-1)\},$$

where $x_1 \leq \cdots \leq x_n$ is an arrangement of $x_1, \ldots, x_n$ in increasing order of magnitude, $A_x = (-\infty, x]$, and $Q$ denotes probability.

We now address the asymptotic optimality of $G_n$. In our empirical Bayes framework, Ferguson's Bayes estimator (1.2) of $F$ based on $X_{n+1}$ now becomes

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

where the dependence of $\tilde{F}_{n+1}^{\cdot}$ on $n$ is suppressed and where $p_m$ is given by (1.3), $F_0(t)$ by (1.4), and $\hat{F}_{n+1}$ is the sample distribution function of $X_{n+1}$. The Bayes risks $R(\alpha)$ and $R(G_n, \alpha)$ of the estimators (2.1) and (1.5), respectively, with respect to the Dirichlet prior, are

$$ (2.2) \quad R(\alpha) \equiv R(\tilde{F}_m, \alpha) = E_{\tilde{F}_m}E_{X_{n+1}} |F| (F(t) - F_0(t))^2 dW(t)), $$

and

$$ (2.3) \quad R(G_n, \alpha) = E_{\tilde{F}_m}E_{X_{n+1}} |F| (F(t) - G_n(t))^2 dW(t)), $$
where the first expectation is taken with respect to the (conditional) distribution of $X_{n+1}$ under a distribution $P$ specified by the Dirichlet prior and the second is with respect to the distribution of $P$ which is a Dirichlet process on $(R,B)$ with parameter $\alpha$. Since

$$\int (\tilde{F}(t) - \tilde{P}_m(t))^2 dW(t) \leq \int dW(t) = W(R) < \infty$$

and

$$\int (\tilde{F}(t) - \tilde{P}_m(t))^2 dW(t) \leq W(R) < \infty,$$

and the risks are non-negative, we may interchange the order of integration in (2.2) and (2.3) and write

$$R_{m}(\tilde{F}, \alpha) = E_{X_{n+1}} E_{P|X_{n+1}} \{\int (\tilde{F}(t) - \tilde{P}_m(t))^2 dW(t)\},$$

and

$$R_{n}(\tilde{G}, \alpha) = E_{X_{n+1}} E_{P|X_{n+1}} \{\int (\tilde{F}(t) - \tilde{G}_n(t))^2 dW(t)\}.$$
Let \( R_n(G, \alpha) \) be the expectation of \( R(G_n, \alpha) \) with respect to 
\( X_1, \ldots, X_n \) (the past observations).

**DEFINITION 2.3.** The sequence \( G = \{G_n\} \) is said to be **asymptotically optimal relative to \( \alpha \)** if \( R_n(G, \alpha) \) 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 [13].

**THEOREM 2.4.** Let \( \alpha \) be \( \sigma \)-additive and \( \alpha(R) \) be known. Then the sequence \( G = \{G_n\} \) is asymptotically optimal relative to \( \alpha \).

**PROOF.** We first prove that for every \( t \in R \), 
\[
E_{m}(t | X_{n+1} = \bar{F}_m(t),
\]
where \( F_m(t) \) is given by (2.1). By Theorem 2.1 the conditional distribution of \( P \), given the sample \( X_{n+1} \), is a Dirichlet process on \( (R, B) \) with parameter 
\[
\beta = \alpha + \sum_{j=1}^{m} \delta_{X_{n+1}, j},
\]
and thus, for every \( t \in R \), the conditional distribution of \( (F(t), 1-F(t)) \), given \( X_{n+1} \), is \( D(\beta((\infty, t]), \beta(R) - \beta((\infty, t])) \). (The notation \( D(p, q) \) is used to denote the Dirichlet distribution with parameter \( (p, q) \); cf. Ferguson [9].)

It follows that
\[
E_{m}(t | X_{n+1} = \bar{F}_m(t), (F(t)) = \beta((\infty, t]) / \beta(R) = (\alpha((\infty, t]) + \sum_{j=1}^{m} \delta_{X_{n+1}, j} / (\alpha(R) + m))
\]
(2.8)
\[
= \bar{F}_m(t),
\]
where the last equality in (2.8) follows from (2.1), (1.3) and (1.4).

Thus we can write (2.7) as
\[
R(G_n, \alpha) = E_{X_{n+1}} \left[ \int [E_{m}(F(t) - \bar{F}_m(t))^2] dW(t) \right] + E_{X_{n+1}} \left[ \int [E_{m}(F_m(t) - \bar{F}_m(t))^2] dW(t) \right]
\]
(2.9)
\[
= R(\bar{F}_m, \alpha) + E_{X_{n+1}} \left[ \int [E_{m}(F_m(t) - \bar{F}_m(t))^2] dW(t) \right],
\]
the last equality following by (2.6). Now,

\[ (2.10) \quad \tilde{F}_m(t) - G_n(t) = p_m(F_0(t) - \sum_{i=1}^{n} \hat{F}_i(t)/n) \]

and this is independent of \( X_{n+1} \) and \( F(t) \). Thus

\[ (2.11) \quad R(G_n, \alpha) = \tilde{R}(F_m, \alpha) + \int (\tilde{F}_m(t) - G_n(t))^2 \, dW(t). \]

As required for establishing asymptotic optimality of \( G = \{ G_n \} \), we take the expectation of \( R(G_n, \alpha) \) with respect to the independent, identically distributed random vectors \( X_1, \ldots, X_n \) (which have a common distribution given by Theorem 2.2), and denote this expectation by \( R_n(G, \alpha) \). We have

\[ (2.12) \quad R_n(G, \alpha) = \tilde{R}(F_m, \alpha) + \int \{ \sum_{i=1}^{n} \hat{X}_i \} (\tilde{F}_m(t) - G_n(t))^2 \, dW(t), \]

where the interchange of order of integration is easily justified as before. Let us evaluate separately each of the terms on the right of (2.12). We have, by (2.10),

\[ (2.13) \quad \sum_{i=1}^{n} \hat{X}_i (\tilde{F}_m(t) - G_n(t))^2 = \sum_{i=1}^{n} \hat{F}_i(t)/n + \sum_{i \neq j} \hat{F}_i(t) \hat{F}_j(t)/n^2. \]

Now, by definition of \( \hat{F}_i \) and using the fact that \( \hat{F}_i \) depends only on \( X_i \), for \( i = 1, \ldots, n \), we have

\[ (2.14) \quad \sum_{i=1}^{n} \hat{X}_i (\hat{F}_i(t)) = \hat{X}_i (\hat{F}_i(t)) = \hat{X}_i \left( \sum_{j=1}^{m} \delta_{x_j} ((-\infty, t])/m \right) = F_0(t). \]

Also, for \( i = 1, \ldots, n \),
\[ E_{X_1, \ldots, X_n}(\hat{F}_i(t)) = E_{X_1}(\hat{F}_i^2(t)) \]
\[ = \sum_{j=1}^{m} E_{X_i} \{ \delta_{X_{ij}}^2((-\infty, t])/m^2 + \sum_{j \neq j'}^{m} E_{X_i} \{ \delta_{X_{ij}}((-\infty, t)] \cdot \delta_{X_{ij'}}((-\infty, t]) / m^2 \) \]
\[ = \sum_{j=1}^{m} Q(X_{ij} \leq t)/m^2 + \sum_{j \neq j'}^{m} Q(X_{ij} \leq t, X_{ij'}, \leq t)/m^2 \]
\[ = \sum_{j=1}^{m} F_0(t)/m^2 + \sum_{j \neq j'}^{m} \frac{\alpha((-\infty, t]) \cdot \alpha((-\infty, t]) + 1}{m^2 \alpha(\alpha(R)+1)} \]
\[ = F_0(t)/m + (m-1)F_0(t)\{F_0(t)\alpha(R) + 1\}/\{m(\alpha(R) + 1)\}, \]

where the equality between the third and fourth lines of (2.15)
follows from Theorem 2.2. Furthermore, for \( i \neq i', i, i' = 1, \ldots, n, \)
we have, since \( \hat{F}_i \cdot \hat{F}_{i'} \) depends only on \( X_i \) and \( X_{i'}, \)
\[ E_{X_1, \ldots, X_n}(\hat{F}_i(t) \cdot \hat{F}_{i'}(t)) = \{ E_{X_i}(\hat{F}_i(t)) \} \cdot \{ E_{X_{i'}}(\hat{F}_{i'}(t)) \} = F_0^2(t). \]
Hence, by (2.13) - (2.16), we have, after simplification,
\[ E_{X_1, \ldots, X_n}(\tilde{G}_m(t) - \tilde{G}_n(t))^2 = \frac{p^2((\alpha(R)+m)/(mn(\alpha(R) + 1))}{\{m(\alpha(R) + 1)\}}F_0(t)(1-F_0(t)). \]

Next, let us compute \( R(F_m, \alpha)(2.6). \) We first obtain, for every \( t \in R, \)
\[ E_{F(t)|X_{n+1}}(\{ F(t)-\tilde{F}_m(t) \}^2). \]
The conditional distribution of \( P, \)
given the sample \( X_{n+1}, \) is a Dirichlet process on \( (\alpha(R,B) \) with parameter \( \beta = \alpha + \sum_{j=1}^{m} \delta_{X_{n+1,j}} \), and thus the conditional distribution of \( (F(t), \)
\[ 1-F(t)), \) given \( X_{n+1}, \) is \( D(\beta((-\infty, t]), \beta(\beta((-\infty, t]))). \)
From the moments of the Dirichlet distribution (cf. DeGroot [5], p. 51)
we obtain
(2.18) \[ E_F(t|X_n, F^2(t)) = \{ \beta((-\infty, t]) \beta((-\infty, t]+1) \}/\{ \beta(R) \beta(R)+1 \} \]
\[ = \tilde{F}_m(t) \tilde{F}_m(t) \beta(R)+1)/(\beta(R)+1). \]

It follows from (2.8) and (2.18) that

(2.19) \[ E_F(t|X_n, \{ (F(t) - \tilde{F}_m(t))^2 \}) = \tilde{F}_m(t) (1 - \frac{\tilde{F}_m(t)}{\beta(R)+1}). \]

Also,

(2.20) \[ E_{X_n+1}(\tilde{F}_m(t)) = E_{X_n+1}(p_m F_0(t) + (1-p_m) \tilde{F}_{n+1}(t)) = F_0(t), \]

and

\[ E_{X_n+1}(\tilde{F}_m^2(t)) = E_{X_n+1}(p_m^2 F_0^2(t) + 2p_m (1-p_m) \tilde{F}_{n+1}(t) F_0(t) + (1-p_m)^2 \tilde{F}_{n+1}^2(t)) \]
\[ = p_m^2 F_0^2(t) + 2p_m (1-p_m) F_0(t) F_0(t) \]
\[ + (1-p_m)^2 [F_0(t)+(m-1)F_0(t)(F_0(t)\alpha(R)+1)/(\alpha(R)+1)]/m, \]

where we have used the results

(2.22) \[ E_{X_n+1}(\tilde{F}_{n+1}(t)) = F_0(t), \]

and

(2.23) \[ E_{X_n+1}(\tilde{F}_{n+1}^2(t)) = \{ F_0(t) + (m-1) F_0(t) (F_0(t) \alpha(R)+1)/(\alpha(R)+1) \}/m. \]
Equations (2.22) and (2.23) are obtained in the same manner as (2.14) and (2.15) respectively. Hence, by (2.19) - (2.21) we obtain, after some straightforward algebra,

\[ (2.24) \quad E_{X_{n+1}} E_{F(t)|X_{n+1}} (F(t) - \tilde{F}_m(t))^2 = \frac{\alpha R}{\{(\alpha R) + 1\}(\alpha R) + m} F_0(t)(1 - F_0(t)). \]

Hence

\[ R(\alpha) = E_{X_{n+1}} \left[ \int \{E_{F(t)|X_{n+1}} (F(t) - \tilde{F}_m(t))^2\} dW(t) \right] \]

\[ (2.25) \quad = \frac{\alpha R}{\{(\alpha R) + 1\}(\alpha R) + m} \int F_0(t)(1 - F_0(t)) dW(t). \]

Then, by (2.12), (2.17) and (2.25), we have, after simplification,

\[ (2.26) \quad R_n(G, \alpha) = (1 + \frac{\alpha R}{mn}) R(\alpha). \]

Hence \( \lim_{n \to \infty} R_n(G, \alpha) = R(\alpha). \)

Note that (2.26) exhibits the rate, \( \frac{1}{n} \), at which \( R_n(G, \alpha) \) converges to \( R(\alpha) \).

3. The performance of \( G_n \) 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 [11]. Here, as a non-Bayesian estimator, we consider the sample distribution function. This function is known to possess many desirable properties. For example, denoting the sample distribution function based on \( X_1, \ldots, X_n \) by \( \tilde{F}_n \), the Glivenko-Cantelli theorem states that, as \( n \to \infty \), \( \sup_{-\infty < t < \infty} |F(t) - \tilde{F}_n(t)| \to 0 \) a.s. Furthermore, consider the group \( G \) 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 G, the sample distribution function \( \hat{F}_n \) is a minimax invariant estimator of F under the loss function 
\[
L(F, \hat{F}) = \int \frac{[(F(t) - \hat{F}(t))^2/(F(t)(1-F(t)))]}{dF(t)}.
\]
(Also see Ferguson [8], p. 191). Dvoretzky, Kiefer and Wolfowitz [7] have shown that the sample distribution is asymptotically minimax for a wide class of loss functions. Phadia [12] establishes that \( \hat{F}_n \) is minimax under the loss function 
\[
L(F, \hat{F}) = \int \frac{[(F(t) - \hat{F}(t))^2/(F(t)(1-F(t)))]}{dW(t)}.
\]

The following theorem shows that the empirical Bayes estimator \( G_n \) is better than the sample distribution function in the sense that for all \( n \geq 2 \), \( G_n \) has a smaller overall expected loss.

**THEOREM 3.1.** Let \( \alpha \) be \( \sigma \)-additive and \( \alpha(R) \) be known. Let

\[
(3.1) \quad T_n(t) = \hat{F}_{n+1}(t),
\]

where \( \hat{F}_{n+1} \) is the sample distribution function based on \( X_{n+1} = (X_{n+1,1}, \ldots, X_{n+1,n}) \). Then, for all \( n \geq 2 \), \( R(T_n, \alpha) \), the Bayes risk of \( T_n \) with respect to the Dirichlet process prior, is larger than \( R_n(G, \alpha) \), the overall expected loss using \( G_n \).

**PROOF.** We first compute \( \int_{X_{n+1}} (\hat{F}_m(t) - T_n(t))^2 \, dW(t) \). We find

\[
(3.2) \quad E_{X_{n+1}} (\hat{F}_m(t) - T_n(t))^2 = [\alpha^2(R)/(m(\alpha(R)+1)(\alpha(R)+m))]F_0(t)(1-F_0(t)).
\]

Hence,

\[
R(T_n, \alpha) = E_{F} E_{X_{n+1}} \left\{ \int (F(t) - T_n(t))^2 dW(t) \right\}
\]

\[
= E_{X_{n+1}} E_{F} \left\{ \int (F(t) - \hat{F}_m(t))^2 dW(t) \right\} + \int (E_{X_{n+1}} (\hat{F}_m(t) - T_n(t))^2 dW(t)
\]

\[
= R(\alpha) + [\alpha^2(R)/(m(\alpha(R)+1)(\alpha(R)+m))]F_0(t)(1-F_0(t))dW(t).
\]
The last equality of (3.3) is a consequence of (3.2) and (2.4). Thus, by (3.3) and (2.25), we obtain

$$ R(T_n, \alpha) = (1 + \alpha(R)/m) R(\alpha). $$

Comparing (3.4) with (2.26), we conclude that

$$ R(T_n, \alpha) > R_n(G, \alpha), \ n \geq 2. \quad || $$

4. **Empirical Bayes estimation of the mean of a distribution.**

Given a sample $X_1, \ldots, X_n$ of size $n$ from an (unknown) distribution $P$, one may desire to estimate the mean $\mu$ of the distribution. Ferguson [9] has obtained a Bayes estimator of the mean using a Dirichlet prior. Let the parameter space be the set of all the distributions on $(R, B)$ and the action space be the real line $R$. Take the loss function to be $L(\mu, \hat{\mu}) = (\mu - \hat{\mu})^2$, where $\mu = \int x dP(x)$ and $\hat{\mu}$ is an estimator of $\mu$. Suppose $P$ is a Dirichlet process on $(R, B)$ with parameter $\alpha$, which we assume to be $\sigma$-additive. Assume also that $\int x d\alpha(x)$ exists and is finite. Then Ferguson's estimator, which is Bayes with respect to the Dirichlet process prior, is

$$ (4.1) \quad \hat{\mu}_n = p_n \mu_0 + (1-p_n)\bar{X}, $$

where

$$ (4.2) \quad \mu_0 = \int x d\alpha(x)/\alpha(R), $$

$$ \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i/n, \text{ and } p_n \text{ is given by (1.3). Note that in order to use (4.1) we must specify not only } \alpha(R), \text{ but also } \mu_0. \text{ An empirical Bayes estimator, which does not require specification of } \mu_0, \text{ is given by (4.3).} $$
We assume the empirical Bayes framework of Section 2. To estimate
\( u_{n+1} = \int x dP_{n+1}(x) \) on the basis of \( X_1, \ldots, X_{n+1} \) we define, for \( n = 1, 2, \ldots \), the sequence of estimators \( v_n \) by

\[
(4.3) \quad v_n = p_m \sum_{i=1}^{n} \frac{X_i}{n} + (1-p_m) \frac{\bar{X}}{n+1},
\]

where \( \bar{X}_i, i = 1, \ldots, n+1, \) is the mean for the sample \( X_i = (X_{11}, \ldots, X_{1m}) \), and \( p_m \) is given by (1.3). We propose \( v_n \) as an empirical Bayes estimator of \( u_{n+1} \). The analogues of Theorems 2.4 and 3.1 are stated, without proof, below.

**Theorem 4.1.** Let \( \sigma \) be \( \sigma \)-additive and \( \alpha(R) \) be assumed known. Let

\[
\int x^2 da(x)/\alpha(R)
\]

exist and be finite. Then

\[
(4.4) \quad R(\alpha) = \left[ \frac{\alpha(R)}{(\alpha(R)+1)(\alpha(R)+m)} \right] \sigma_{11},
\]

and

\[
(4.5) \quad R_n(M,\alpha) = (1 + \alpha(R)/mn) R(\alpha),
\]

where \( R(\alpha) \) and \( R_n(M,\alpha) \) are the analogues of (2.2) and (2.12) defined for \( \hat{u}_n \) and \( v_n \), respectively. Here the loss function is \( L(u, \hat{u}) = (u-\hat{u})^2 \), and \( \sigma_{11} = \int x^2 da(x)/\alpha(R) - \mu_0^2 \). In particular, \( M = \{v_n\} \) is asymptotically optimal with respect to \( \alpha \).

**Theorem 4.2.** Let the hypotheses of Theorem 4.1 hold. Set

\[
U_n = \bar{X}_{n+1} = \frac{\sum_{j=1}^{m} X_{n+1,j}}{m}.
\]

Then \( R(U_n, \alpha) \), the Bayes risk of \( U_n \) with respect to the Dirichlet process prior, is

\[
(4.6) \quad R(U_n, \alpha) = (1 + \alpha(R)/m) R(\alpha),
\]

where \( R(\alpha) \) is given by (4.4). In particular, \( R(U_n, \alpha) \) is, for all \( n \geq 2 \), greater than \( R_n(M, \alpha) \), the overall expected loss using \( v_n \).
Ferguson's Bayes estimators (1.2) and (4.1), and the corresponding empirical Bayes estimators (1.5) and (4.3), depend on the value of \( a(R) \). If \( a(R) \) is unknown (or if one is unwilling to specify a value for \( a(R) \)), then \( a(R) \) can be consistently estimated from the data. Various estimators can be found in [10]. In particular let \( X_1, \ldots, X_n \) be a sample of size \( n \) from a Dirichlet process on \((R, \mathcal{B})\) with parameter \( a \). Assume \( a \) to be \( \sigma \)-additive and non-atomic. Let \( D(n) \) denote the number of distinct observations on the sample. Then it is shown in [10] that \( D(n)/\log n \xrightarrow{a.s.} a(R) \), \( n \to \infty \).
REFERENCES


Let \((P_i, X_i), i = 1, 2, \ldots\) be a sequence of independent pairs of random variables, the \(P_i\) having a common prior distribution given by a Dirichlet process (Ferguson, Ann. Statist. 1(1973) 209-230) on \((R, \mathcal{B})\) with a parameter \(\alpha\) (assumed \(\sigma\)-additive), and given \(P_i = P, X_i = (X_{i1}, \ldots, X_{im})\) is a sample of size \(m\) from \(P\). Consider the problem of estimating \(F_{n+1}(t) = P_{n+1}\{(-\infty, t]\}\) on the
basis of $X_1, \ldots, X_{n+1}$. Assume $\alpha$ is non-atomic, $\alpha(R)$ is known, and take the loss in estimating $F(t)$ by $a(t)$ to be $\int (F(t)-a(t))^2 dW(t)$ where $W$ is a weight function. Define $G_n(t) = p_{m}(\sum_{i=1}^{n} \hat{F}_i(t)/n) + (1-p_{m})\hat{F}_{n+1}(t)$ where $p_{m} = [\alpha(R)/(\alpha(R)+m)]$ and $\hat{F}_i$ is the empirical distribution function for the sample $X_{i1}, \ldots, X_{im}$. Then the sequence $\{G_n\}$ is shown to be asymptotically optimal in the sense of Robbins (cf. *Ann Math. Statist.* 35 (1964) 1-20). The performance of $G_n$ is compared with that of a corresponding Bayes estimator derived by Ferguson, and with the sample distribution function $\hat{F}_{n+1}$. The empirical Bayes approach based on the Dirichlet process is also applied to the problem of estimating the mean of a distribution.

### Key Words

<table>
<thead>
<tr>
<th>Link A</th>
<th>Link B</th>
<th>Link C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Role Wt.</td>
<td>Role Wt.</td>
<td>Role Wt.</td>
</tr>
</tbody>
</table>

- Distribution function
- Empirical Bayes estimator
- Dirichlet process