NONPARAMETRIC EMPIRICAL BAYES ESTIMATION
OF THE PROBABILITY THAT \( X \leq Y \)

by

Myles Hollander\(^1\) and Ramesh M. Korwar

FSU Statistics Report M389
APOS-N Technical Report No. 1
APOS-N Technical Report No. 64

October, 1976
The Florida State University
Department of Statistics
Tallahassee, Florida

---

NONPARAMETRIC EMPIRICAL BAYES ESTIMATION
OF THE PROBABILITY THAT $X \leq Y$

Myles Hollander\textsuperscript{1}
Florida State University
Tallahassee, Florida

Ramesh M. Korwar
University of Massachusetts
Amherst, Massachusetts

ABSTRACT

A sequence of empirical Bayes estimators is defined for estimating, in a two-sample problem, the probability that $X \leq Y$. The sequence is shown to be asymptotically optimal relative to a Ferguson Dirichlet process prior.

1. INTRODUCTION

Let $X$ and $Y$ be two real valued independent random variables with distribution functions $F$ and $G$, respectively. We consider the problem of estimating the probability that $X \leq Y$, denoted by $\Delta$,

$$\Delta = \int F dG.$$ 

\textsuperscript{1}This research was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grants AFSOR-74-2581B and AFSOR-76-3109. The United States Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation hereon.
This problem was recently treated by Ferguson (1973) who used a nonparametric Bayesian approach based on the Dirichlet process. Ferguson's nonparametric Bayes estimator of $\Delta$ is as follows. [For Dirichlet process definitions see Section 2, and for more details, see Ferguson's (1973) paper.] To estimate $\Delta$, Ferguson lets $X_1, \ldots, X_{n_1}$ be a sample from $F$ where it is assumed $F$ is a random distribution function chosen by a Dirichlet process $P_1$ with parameter $\alpha_1$. Furthermore, $Y_1, \ldots, Y_{n_2}$ is a sample from $G$ where $G$ is chosen by a Dirichlet process $P_2$ with parameter $\alpha_2$, and $P_1$ and $P_2$ are independent. For squared error loss, Ferguson's Bayes estimator of $\Delta$ is given by

$$
\hat{\Delta}^* = p_{1,n_1} p_{2,n_2} \Delta_0 + p_{1,n_1} (1 - p_{2,n_2}) \frac{1}{n_2} \sum_{j=1}^{n_2} F_0(Y_j)
+ (1 - p_{1,n_1}) p_{2,n_2} \frac{1}{n_1} \sum_{i=1}^{n_1} (1 - G_i(X_i))
+ (1 - p_{1,n_1}) (1 - p_{2,n_2}) \frac{1}{n_1 n_2} U
$$

where

$$p_{1,n_1} = \frac{\alpha_1(R)}{\alpha_1(R) + n_1}, \quad p_{2,n_2} = \frac{\alpha_2(R)}{\alpha_2(R) + n_2},$$

(1.2)

$$F_0(x) = \frac{\alpha_1((-\infty,x])}{\alpha_1(R)}, \quad G_0(y) = \frac{\alpha_2((-\infty,y])}{\alpha_2(R)},$$

(1.3)

$$\Delta_0 = \int F_0 dG_0,$$

(1.4)

$R$ is the real line and $U$, the number of pairs $(X_i, Y_j)$ for which $X_i \leq Y_j$,

$$U = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I_{(-\infty,Y_j]}(X_i)$$

(1.5)

is the Mann-Whitney statistic. Here

$$I_A(x) = \begin{cases} 
1 & \text{if } x \in A \\
0 & \text{if } x \notin A.
\end{cases}$$

(1.6)
Ferguson notes that the estimator $\hat{\Delta}^*$ is a simple mixture of four separate estimators of $\Delta$. As both $a_1(R)$ and $a_2(R)$ tend to zero, $\hat{\Delta}^*$ converges to $(n_1 n_2)^{-1} U$, the usual nonparametric estimator.

Motivated by Ferguson's $\hat{\Delta}^*$, we propose an empirical Bayes estimator of $\Delta$ which requires less prior information about $a_1(\cdot)$ and $a_2(\cdot)$. Only $a_1(R)$ and $a_2(R)$ need be specified. Consider, then, the following set up appropriate for an empirical Bayes estimation problem. Let $\{X_1^{(1)}\}, \{X_2^{(2)}\}, i = 1, 2, \ldots$, be two independent sequences of independent vectors of observations from $F$ and $G$ respectively. Here $X_i^{(j)} = (X_{1i}^{(j)}, \ldots, X_{ni}^{(j)})$, $j = 1, 2$ and $i = 1, 2, \ldots$. Assume independent Dirichlet priors on $(R, B)$ with parameters $a_1$ and $a_2$ respectively for $F$ and $G$. Here $R$ is the real line and $B$ is the $\sigma$-field of Borel subsets of $R$. Let the action space be the closed interval $[0, 1]$, and the loss function be

$$L(\hat{\Delta}, \Delta) = (\Delta - \hat{\Delta})^2,$$  \hspace{1cm} (1.7)

where $\hat{\Delta}$ is an estimator of $\Delta$. We assume $a_1(R)$ and $a_2(R)$ are known.

We then propose the estimator $\hat{\Delta}_n$ below as an estimator of $\Delta$ on the $(n+1)$th occasion. The estimator is given by

$$\hat{\Delta}_n = \hat{\Delta}(X_1^{(1)}, \ldots, X_{n+1}^{(1)}, X_1^{(2)}, \ldots, X_{n+1}^{(2)})$$

$$= p_1, n_1 p_2, n_2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_1} \sum_{\ell=1}^{n_2} I(\infty, X_{ij}^{(1)}) (X_{k\ell}^{(2)})$$

$$+ p_1, n_1 (1-p_2, n_2) \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_1} \sum_{\ell=1}^{n_2} \delta_{X_{ij}^{(1)}}((-\infty, X_{n+1,j}^{(2)}])$$

$$+ (1-p_1, n_1) p_2, n_2 \sum_{i=1}^{n_1} \left\{ \frac{1}{n_2} \sum_{j=1}^{n_2} \sum_{k=1}^{n_1} \delta_{X_{ij}^{(2)}}((-\infty, X_{n+1,j}^{(2)}]) \right\}$$

$$+ (1-p_1, n_1) (1-p_2, n_2) \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(\infty, X_{ij}^{(1)}) (X_{n+1,i}^{(2)}),$$  \hspace{1cm} (1.8)
where $p_{1, n}, i = 1, 2,$ are given by (1.2), and

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases} \quad (1.9)$$

Note that the first three terms in (1.8) are the natural estimators of corresponding terms in the Bayes estimator based on all the observations or only the past observations.

In Section 3 we prove that the sequence $D = \{\delta_n\}$ is asymptotically optimal in the sense of Robbins (1964). Thus even though one need only specify $\alpha_1(R)$ and $\alpha_2(R)$, the procedure is asymptotically as good as though $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ were known exactly.

Empirical Bayes methods, based on the Dirichlet process, have also been used by Antoniak (1974) for a model based on mixtures of Dirichlet processes, by Korwar and Hollander (1976) for estimating the mean of a distribution, by Korwar and Hollander (1976) and Hollander and Korwar (1976) for estimating a distribution function, and by Susarla and Van Ryzin (1976) for estimating a distribution function when the observations are censored on the right.

2. DIRICHLET PROCESS PRELIMINARIES

In this section we present some Dirichlet process definitions and results that will be used in our proof of asymptotic optimality. See Ferguson (1973), (1974) for more comprehensive coverage of results pertaining to the Dirichlet process.

**Definition 2.1** (Ferguson, 1973). Let $Z_1, \ldots, Z_k$ be independent random variables with $Z_j$ having a gamma distribution with shape parameter $\alpha_j \geq 0$ and scale parameter $1, j = 1, \ldots, k$. Let $\alpha_j > 0$ for some $j$. The Dirichlet distribution with parameter $(\alpha_1, \ldots, \alpha_k)$, denoted by $D(\alpha_1, \ldots, \alpha_k)$, is defined as the distribution of

$$(Y_1, \ldots, Y_k), \text{ where } Y_j = Z_j / \sum_{i=1}^k Z_i, \ j = 1, \ldots, k.$$  

This distribution is always singular with respect to Lebesgue measure on $k$-dimensional Euclidean space. Also, if any $\alpha_i = 0$, the corresponding $Y_i$ is degenerate at 0. However, if $\alpha_i > 0$ for
all \( i = 1, \ldots, k \), the \((k-1)\)-dimensional distribution of \((Y_1, \ldots, Y_{k-1})\) has density, with respect to Lebesgue measure on the \((k-1)\)-dimensional Euclidean space, given by

\[
f(y_1, \ldots, y_{k-1} | \alpha_1, \ldots, \alpha_k) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_k)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k) \left( \Pi y_i \right)^{k-1} \left( 1 - \sum_{i=1}^{k-1} y_i \right)^{\alpha_i-1}} I_S(y_1, \ldots, y_{k-1}),
\]

where \( S \) is the simplex

\[ S = \{(y_1, \ldots, y_{k-1}): y_i \geq 0, \sum_{i=1}^{k-1} y_i \leq 1\}. \]

For \( k = 2 \), (2.1) becomes the density of a Beta distribution, \( \text{Be}(\alpha_1, \alpha_2) \). Note that the condition \( \alpha_i > 0 \) for some \( i = 1, \ldots, k \), is required in Definition 2.1 so that \( \sum_{i=1}^{k} \alpha_i \) is not degenerate at 0.

Let \((X, A)\) be a measurable space. Ferguson defined the following stochastic process \( \{P(A), A \in A\} \).

**Definition 2.2** (Ferguson, 1973). Let \((X, A)\) be a measurable space. Let \( \alpha \) be a non-null finite measure (nonnegative and finitely additive) on \((X, A)\). Then \( P \) is a Dirichlet process on \((X, A)\) with parameter \( \alpha \) if for every \( k = 1, 2, \ldots \), and measurable partition \( (B_1, \ldots, B_k) \) of \( X \), the distribution of \( (P(B_1), \ldots, P(B_k)) \) is Dirichlet with parameter \( (\alpha(B_1), \ldots, \alpha(B_k)) \).

If \( F \) is chosen by a Dirichlet process, then \( F \) is discrete with probability one [see Ferguson (1973), Berk and Savage (1975), Blackwell (1973), and Blackwell and MacQueen (1973)].

A sample from a Dirichlet process is next defined.

**Definition 2.3** (Ferguson, 1973). The \( X \)-valued random variables \( X_1, \ldots, X_m \) constitute a sample of size \( m \) from a Dirichlet process \( P \) on \((X, A)\) with parameter \( \alpha \) if for any \( \ell = 1, 2, \ldots \) and measurable sets \( A_1, \ldots, A_\ell, C_1, \ldots, C_m \), \( Q(X_1 \in C_1, \ldots, X_m \in C_m | P(A_1), \ldots, P(A_\ell)) = \prod_{i=1}^{m} P(C_i) \) a.s., where \( Q \) denotes probability.
Roughly speaking, we may view a sample of size m from a
Dirichlet process as follows. The process chooses a random distri-
bution F, say, and then given F, \( X_1, \ldots, X_m \) is a random sample from
F.

Theorem 2.4 gives the posterior distribution of a Dirichlet
process P, given a sample \( X_1, \ldots, X_m \) from the process.

**Theorem 2.4** (Ferguson, 1973). Let P be a Dirichlet process on
\((X, A)\) with parameter \( \alpha \), and let \( X_1, \ldots, X_m \) be a sample of size m
from P. Then the conditional distribution of P given \( X_1, \ldots, X_m \)
is a Dirichlet process on \((X, A)\) with parameter \( \beta = \alpha + \sum_{i=1}^{m} \delta_{X_i} \),
where, for \( x \in X \), \( A \in A \), \( \delta_{X}(A) \) is given by (1.9).

Theorem 2.5 is a generalization of Ferguson's (1973)
Proposition 4.

**Theorem 2.5.** Let P be a Dirichlet process on \((R, B)\) with parameter
\( \alpha \) and let \( X_1, \ldots, X_m \) be a sample of size m from P. Then

\[
Q\{X_1 \leq x_1, \ldots, X_m \leq x_m\} = \left\{ \alpha(A_{x_1}) \ldots (\alpha(A_{x_m}) + m-1) \right\} / \{\alpha(R) \ldots (\alpha(R) + m-1)\},
\]

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

**Proof.** Observe that \( A_{x}^{(k)} \subset A_{x}^{(k+1)} \), \( k = 1, \ldots, m-1 \). We write

\[
A_{x}^{(1)} = B_1,
\]

\[
A_{x}^{(k)} = A_{x}^{(1)} + A_{x}^{(2)} + \ldots + A_{x}^{(k)}
= B_1 + B_2 + \ldots + B_k,
\]

\( k = 2, \ldots, m \).

\[
A_{x}^{(m)} = B_{m+1}.
\]
Here $A^C$ denotes the complement of the set $A$. Then

$$Q\{X_1 \leq x_1, \ldots, X_m \leq x_m\} = Q\{X_1 \in A \times (i_1), \ldots, X_m \in A \times (i_m)\},$$

where $(i_1, \ldots, i_m)$ is a permutation of $(1, \ldots, m)$. Now

$$Q\{X_1 \in A \times (i_1), \ldots, X_m \in A \times (i_m)\}
\begin{align*}
&= E[Q\{X_1 \in A \times (i_1), \ldots, X_m \in A \times (i_m)\} | P(A \times (i_1), \ldots, P(A \times (i_m))]
\end{align*}
\begin{align*}
&= E[P(A \times (i_1)) \cdots P(A \times (i_m))],
\end{align*}

by Definition 2.3. Since $(P(B_1), \ldots, P(B_{m+1}))$ is $D(a(B_1), \ldots, a(B_{m+1}))$, using the moments of the Dirichlet distribution we can obtain

$$E[P(A \times (i_1)) \cdots P(A \times (i_m))]
\begin{align*}
&= E[P(B_1)(P(B_1)+P(B_2)) \cdots (P(B_1) + \ldots + P(B_m))]
\end{align*}
\begin{align*}
&= \{a(A \times (1))a(A \times (2)) \cdots (a(A \times (m)+m-1))\}/\{a(R)(a(R)+1)\cdots (a(R)+m-1)\},
\end{align*}

Theorem 2.6 (Ferguson, 1973). Let $P$ be the Dirichlet process (Definition 2.2) with parameter $\alpha$ and let $Z_1$ and $Z_2$ be measurable real valued functions defined on $(X, A)$. If $\int |Z_1|d\alpha < \infty$, $\int |Z_2|d\alpha < \infty$ and $\int |Z_1Z_2|d\alpha < \infty$, then

$$E[\int Z_1dP]Z_2dP = \{\sigma_{12}/(\alpha(X)+1)\} + \mu_1\mu_2,$$

where

$$\mu_i = \int Z_id\alpha/\alpha(X), \quad i = 1, 2$$

and

$$\sigma_{12} = \{\int Z_1Z_2d\alpha/\alpha(X)\} - \mu_1\mu_2.$$
3. ASYMPOTIC OPTIMALITY OF \{\hat{\Delta}_n\}.

We now establish the asymptotic optimality of \( D = \{\hat{\Delta}_n\} \). In our empirical Bayes framework, Ferguson's Bayes estimator of \( \Delta \) based on \( (X_{n+1}^{(1)}, X_{n+1}^{(2)}) \) is

\[
\hat{\Delta}_n = p_1 p_2 \Delta_0^* \sum_{j=1}^{n_2} F_0(X_{n+1,j}^{(2)})/n_2 + (1-p_1)p_2 \sum_{i=1}^{n_1} (1-G_0(X_{n+1,i}^{(1)})/n_1

+ (1-p_1)(1-p_2) \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(-\infty, X_{n+1,i}^{(2)} / n_1 n_2),
\]

(3.1)

where \( p_1 \equiv p_1, n_1 \) and \( p_2 \equiv p_2, n_2 \) are given by (1.2), \( F_0 \) and \( G_0 \) by (1.3), \( \Delta_0 \) by (1.4) and \( I_A(x) \) by (1.6). The Bayes risks \( R(\hat{\Delta}_n^{*}, \alpha_1, \alpha_2) \) and \( R(\hat{\Delta}_n, \alpha_1, \alpha_2) \) of (3.1) and (1.8) respectively, with respect to the Dirichlet priors, are

\[
R(\hat{\Delta}_n^{*}, \alpha_1, \alpha_2) \overset{\text{def}}{=} R(\hat{\Delta}_n, \alpha_1, \alpha_2) = E_{X_{n+1}^{(1)}, X_{n+1}^{(2)}} E_{F,G|X_{n+1}^{(1)}, X_{n+1}^{(2)}} (\Delta - \hat{\Delta}_n)^2,
\]

(3.2)

and

\[
R(\hat{\Delta}_n, \alpha_1, \alpha_2) = E_{X_{n+1}^{(1)}, X_{n+1}^{(2)}} E_{F,G|X_{n+1}^{(1)}, X_{n+1}^{(2)}} (\Delta - \hat{\Delta}_n)^2.
\]

(3.3)

Let \( R_n(D, \alpha_1, \alpha_2) \) be the expectation of \( R(\hat{\Delta}_n, \alpha_1, \alpha_2) \) with respect to \( X_{1}^{(1)}, X_{1}^{(2)}, \ldots, X_{n}^{(1)}, X_{n}^{(2)} \) (the past observations).

Definition 3.1. The sequence \( D = \{\hat{\Delta}_n\} \) is said to be asymptotically optimal relative to \((\alpha_1, \alpha_2)\) if \( R_n(D, \alpha_1, \alpha_2) \) converges to the minimum Bayes risk \( R(\hat{\Delta}^{*}, \alpha_1, \alpha_2) \), as \( n \rightarrow \infty \).

Definition 3.1 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 (1964).

Theorem 3.2. Let \( \alpha_1(R) \) and \( \alpha_2(R) \) be known. Then

\[
R(\hat{\Delta}^{*}, \alpha_1, \alpha_2) = E\Delta^2 - E\hat{\Delta}^{*2},
\]

(3.4)
\[ R_n(D, \alpha_1, \alpha_2) = (E\Delta^2_n - E\Delta^*2_n) + (E\Delta^*2_n - E\Delta^*2_n), \]  
\hspace{1cm} (3.5)

where

\[ E\Delta^2_n = p_{1,1}p_{2,1}\Delta_0^2 + (1-p_{1,1})(1-p_{2,1})\Delta_0 \]
\[ + (1-p_{1,1})p_{2,1}I_1 + p_{1,1}(1-p_{2,1})I_2, \]  
\hspace{1cm} (3.6)

\[ E\Delta^*2_n = g_1g_2\Delta_0^2 + (1-g_1)(1-g_2)\Delta_0 + (1-g_1)g_2I_1 \]
\[ + g_1(1-g_2)I_2, \]  
\hspace{1cm} (3.7)

\[ E\Delta^2_n = h_1h_2\Delta_0^2 + (1-h_1)(1-h_2)\Delta_0 + (1-h_1)h_2I_1 \]
\[ + h_1(1-h_2)I_2, \]  
\hspace{1cm} (3.8)

where

\[ I_1 = \int F_0(y_1) dG_0(y_1) dG_0(y_2), \]  
\hspace{1cm} (3.9)

\[ I_2 = \int F_0^2(y) dG_0(y), \]  
\hspace{1cm} (3.10)

and

\[ g_1 \equiv g_1,n_1 = \frac{p_{1,1}p_{1,n_1}}{p_{1,n_1+1}}, \quad g_2 \equiv g_2,n_2 = \frac{p_{2,1}p_{2,n_2}}{p_{2,n_2+1}}, \]  
\hspace{1cm} (3.11)

\[ h_1 \equiv h_1,n_1 = g_1\left[1 - \frac{p_{1,n_1+1}}{nn_1}\right], \quad h_2 \equiv h_2,n_2 = g_2\left[1 - \frac{p_{2,n_2+1}}{nn_2}\right], \]

In particular \( D = \{\Delta_n\} \) is asymptotically optimal relative to \((\alpha_1, \alpha_2)\).

Our proof of Theorem 3.2 uses the following lemma.

**Lemma 3.3.** Let \( X_i^{(1)} = (x_{i1}^{(1)}, \ldots, x_{i n_1}^{(1)}), \ i = 1, 2, \) be two independent samples, each of size \( n_1, \) from a Dirichlet process on \((R, \mathcal{B})\) with parameter \( \alpha_1, \) and let \( X_i^{(2)} = (x_{i1}^{(2)}, \ldots, x_{i n_2}^{(2)}), \ i = 1, 2, \) be two independent samples, each of size \( n_2, \) independent of \((X_1^{(1)}, X_2^{(1)}), \)
from an independent Dirichlet process on \((R, \mathcal{B})\) with parameter \(\alpha_2\). Assume that each of \(\alpha_1, \alpha_2\) is \(\sigma\)-additive. Then

\[
\mathbb{E} I_{(-\infty, X^{(2)}_{i,j})} = \mathbb{E} F_0(X^{(2)}_{i,j}) \quad (3.12)
\]

\[
\mathbb{E} \delta_{X^{(1)}_{k,\ell}}((-\infty, X^{(2)}_{i,j})) = \mathbb{E} F_0(X^{(2)}_{i,j}), \quad (3.13)
\]

\[
\mathbb{E} \delta_{X^{(2)}_{i,j}}([-\infty, \alpha_1]) \mathbb{E} F_0(X^{(2)}_{i,j}) = \mathbb{E} F_0(X^{(2)}_{i,j}) \mathbb{E} F_0(X^{(2)}_{i,j}), \quad (3.14)
\]

\[
\mathbb{E} F_0(X^{(2)}_{i,j}) = \Delta_0, \quad (3.15)
\]

\[
\mathbb{E}(1-G_0(X^{(1)}_{k,\ell}-)) = \Delta_0, \quad (3.16)
\]

\[
\mathbb{E} F_0(X^{(2)}_{i,j}) I_{(-\infty, X^{(2)}_{i',j'})} = \mathbb{E} F_0(X^{(2)}_{i,j}) \mathbb{E} F_0(X^{(2)}_{i',j'}), \quad (3.17)
\]

\[
\mathbb{E}(1-G_0(X^{(1)}_{k,\ell}-)) I_{(-\infty, X^{(2)}_{i,j})} = \mathbb{E}(1-G_0(X^{(1)}_{k,\ell}-)) \mathbb{E}(1-G_0(X^{(1)}_{k,\ell'}-)), \quad (3.18)
\]

\[
\mathbb{E} \delta_{X^{(1)}_{k,\ell}}((-\infty, X^{(2)}_{i,j}) \delta_{X^{(1)}_{k',\ell'}}((-\infty, X^{(2)}_{i',j'})) = \left\{
\begin{array}{ll}
\mathbb{E} F_0(X^{(2)}_{i,j}) F_0(X^{(2)}_{i',j'}) & , \quad k \neq k' \\
\{\alpha_1(R)\mathbb{E} F_0(X^{(2)}_{i,j}) F_0(X^{(2)}_{i',j'}) + \mathbb{E} F_0(X^{(2)}_{(1)})\}/(\alpha_1(R) + 1) & , \quad k=k', \ell \neq \ell' \\
\mathbb{E} F_0(X^{(2)}_{(1)}) & , \quad k=k', \ell = \ell'
\end{array}\right. \quad (3.19)
\]
\[ EI \quad (X^{(1)}_{i,j})_{k'\ell'} \quad (X^{(1)}_{i',j'}) = \]

\[
\begin{cases} 
EF_0(X^{(2)}_{i,j})F_0(X^{(2)}_{i',j'}) \quad \text{if } k \neq k', \\
\{ \alpha_1(R)EF_0(X^{(2)}_{i,j})F_0(X^{(2)}_{i',j'}) + EF_0(X^{(2)}_{i,j}) \} / (\alpha_1(R)+1), \\
k=k', \end{cases}
\]

(3.20)

\[ E\delta \quad (X^{(1)}_{i,j}, \infty)_{k'\ell'} \quad \delta \quad (X^{(1)}_{i',j'}, \infty) = \]

\[
\begin{cases} 
\{ \alpha_1(R)EF_0(X^{(2)}_{i,j})F_0(X^{(2)}_{i',j'}) + EF_0(X^{(2)}_{i,j}) \} / (\alpha_1(R)+1), \\
\quad \ell=\ell', \\
EF_0(X^{(2)}_{i,j}), \\
\quad \ell=\ell'. 
\end{cases}
\]

(3.21)

\[ EI \quad (X^{(1)}_{i,j}, \infty)_{k'\ell'} \quad (X^{(1)}_{i',j'}) \quad \delta \quad (X^{(1)}_{i',j'}, \infty) = \]

\[
\begin{cases} 
EF_0(X^{(2)}_{i,j})F_0(X^{(2)}_{i',j'}) \quad k=k', \\
\{ \alpha_1(R)EF_0(X^{(2)}_{i,j})F_0(X^{(2)}_{i',j'}) + EF_0(X^{(2)}_{i,j}) \} / (\alpha_1(R)+1), \\
k=k', \quad \ell \neq \ell', \\
EF_0(X^{(2)}_{i,j}), \\
k=k', \quad \ell=\ell'. 
\end{cases}
\]

(3.22)
$$\text{ET}_{(-\infty, X_{ij}^{(2)})} \delta_{X_{i'j}', X_{i'j}'} \left( [X_{k',}\ell'], \infty) \right) =$$

$$\left\{ \begin{array}{l}
\text{EF}_0(X_{ij}^{(2)}) F_0(X_{i'j}^{(2)}), \quad k=k', \\
\left\{ \alpha_1(R) \text{EF}_0(X_{ij}^{(2)}) F_0(X_{i'j}^{(2)}) + \text{EF}_0(X_{i'j}) \right\} / (\alpha_1(R) + 1), \quad k=k', \ell=\ell' \\
\text{EF}_0(X_{i'j}') \quad , \quad k=k', \ell=\ell'.
\end{array} \right.$$

(3.23)

In (3.19) - (3.23), $X_{ij}^{(2)}$ is the smaller of $X_{ij}$ and $X_{i'j}'$. Also,

$$\text{EF}_0(X_{i'j}') = \left\{ \begin{array}{l}
\int F_0(y) dG_0(y) dG_0(y'), \quad i\neq i' \\
\int F_0(y) dK_0(y_1, y_2), \quad i=i',
\end{array} \right.$$  

(3.24)

$$\text{EF}_0(X_{i'j}') F_0(X_{i'j}') = \left\{ \begin{array}{l}
\Delta_0^2 \quad i\neq i' \\
\int F_0(y) F_0(y) dK_0(y_1, y_2) \quad i=i', j\neq j' \\
\int F_0(y) dG_0(y) \quad i=i', j=j'.
\end{array} \right.$$

(3.25)

where $y_{ij}$ is the smaller of $y_1$ and $y_2$, and

$$E(1-G_0(X_{k}\ell^{-})) (1-G_0(X_{k'}\ell'^{-})) =$$

$$\left\{ \begin{array}{l}
\Delta_0^2 \quad , \quad k=k' \\
\int (1-G_0(x_1^-)) (1-G_0(x_2^-)) dH_0(x_1, x_2), \quad k=k', \ell=\ell' \\
\int (1-G_0(x^-))^2 dF_0(x) \quad k=k', \ell=\ell'.
\end{array} \right.$$

(3.26)
In (3.24) - (3.26), $H_0(x_1, x_2)$ and $K_0(y_1, y_2)$ are the distribution functions of $(X^{(1)}_{k\ell}, X^{(1)}_{i'j'})$, $\ell \neq \ell'$ and of $(X^{(2)}_{ij}, X^{(2)}_{ij'})$, $j \neq j'$ respectively which are given by Theorem 2.5.

**Sketch of Proof of Lemma 3.3.** To prove (3.12) - (3.14), use the independence of $X^{(1)}_k$ and $X^{(2)}_i$ and Theorem 2.5. We prove (3.12):

$$
E \int_{(-\infty, X^{(2)}_{ij})} \left( \frac{X^{(1)}_{k\ell}}{X^{(2)}_{ij}} \right) = E(E(I \mid X^{(1)}_{k\ell} \mid X^{(2)}_{ij})) = EF_0(X^{(2)}_{ij}).
$$

Equations (3.15) - (3.16) are readily verified from Theorem 2.5 and the definition of $A_0$. Equation (3.17) follows from Theorem 2.5 and the independence of $X^{(1)}_k$ from $X^{(2)}_i$ and $X^{(2)}_{i'}$, and (3.18) follows from Theorem 2.5 and the independence of $X^{(2)}_i$ from $X^{(1)}_k$ and $X^{(1)}_{k'}$.

To prove (3.19) - (3.23), use the independence of $(X^{(1)}_k, X^{(1)}_{k'})$ and $(X^{(2)}_i, X^{(2)}_{i'})$ and Theorem 2.5. We prove (3.19). It follows from the independence of $(X^{(1)}_k, X^{(1)}_{k'})$, $(X^{(2)}_i, X^{(2)}_{i'})$, and Theorem 2.5, that

$$
E \delta_{X^{(1)}_k((-\infty, X^{(2)}_{ij}))} \delta_{X^{(1)}_{k'}}((-\infty, X^{(2)}_{ij'})), \kappa \neq \kappa'
$$

$$
= E(E(\delta_{X^{(1)}_k((-\infty, X^{(2)}_{ij}))} \delta_{X^{(1)}_{k'}}((-\infty, X^{(2)}_{ij'})) | X^{(2)}_i, X^{(2)}_{i'}))
$$

$$
= \begin{cases}
EF_0(X^{(2)}_{ij})F_0(X^{(2)}_{ij'}), \kappa \neq \kappa' & \text{(by the independence of $X^{(1)}_k, X^{(1)}_{k'}$)} \\
\frac{E \alpha_1((-\infty, X^{(2)}_{ij}))}{\alpha_1(R)}(\alpha_1((-\infty, X^{(2)}_{ij})) + 1), \kappa = \kappa', \ell \neq \ell' \\
\frac{E \alpha_1((-\infty, X^{(2)}_{ij})}{\alpha_1(R)}, \kappa = \kappa', \ell = \ell',
\end{cases}
$$

which reduces to (3.19). Equations (3.24) - (3.26) are readily verified.
Remark 3.4. By using the conditional distribution \( L_0(y_2 | y_1) = \frac{\alpha_2(-\infty, y_2] + \delta_{y_1}((-\infty, y_2])}{\alpha_2(R) + 1} \) of \( X_{i'j'}^{(2)} \), given \( X_{ij}^{(2)} \) \((j \neq j')\), we can show that

\[
\int F_0(y_1) dK_0(y_1, y_2) = \frac{[\alpha_2(R) I_1 + A_0]}{(\alpha_2(R) + 1)}, \quad (3.27)
\]

\[
\int F_0(y_1) F_0(y_2) dK_0(y_1, y_2) = \frac{[\alpha_2(R) \Delta_0^2 + I_2]}{(\alpha_2(R) + 1)}. \quad (3.28)
\]

The function \( K_0(y_1, y_2) \) is the distribution function of \( X_{i'j'}^{(2)} \), \( X_{ij}^{(2)} \) \((j \neq j')\), and \( y_1 \) is \( \min(y_1, y_2) \). Similarly, by considering the conditional distribution of \( X_{kl'}^{(1)} \), given \( X_{kl}^{(1)} \) \((l \neq l')\) we can show

\[
\int (1 - G_0(x_1^-)) (1 - G_0(x_2^-)) dH_0(x_1, x_2) = \frac{[\alpha_1(R) \Delta_0^2 + I_1]}{(\alpha_1(R) + 1)}. \quad (3.29)
\]

Proof of Theorem 3.2. To prove (3.4), expand \((\Delta - \Delta^*)^2\), and note that \( \Delta^* \), being the Bayes estimator with squared error loss, is

\[
E(\Delta | X_{n+1}^{(1)}, X_{n+1}^{(2)}). \quad (3.5)
\]

To prove (3.5), expand \((\Delta - \Delta_n^*)^2\), and note that

\[
E(\Delta_n^*) = E(E(\Delta_n^* | X_{n+1}^{(1)}, X_{n+1}^{(2)})) = E(E(\Delta_n^* | X_{n+1}^{(1)}, X_{n+1}^{(2)})) = E(\Delta_n^*)^2. \quad (3.6)
\]

(Here we used the fact that \( E(\Delta_n^* | X_{n+1}^{(1)}, X_{n+1}^{(2)}) = \Delta^* \), which is easily verified.) To prove (3.6), use Theorem 2.6 twice. Thus,

\[
E\Delta^2 = E(\int F(y) dG(y))^2
\]

\[
= E(E(\int F(y) dG(y))^2 | F)
\]

\[
= E((\int F(y) dG_0(y) / (\alpha_2(R) + 1) + \alpha_2(R) (\int F(y) dG_0(y))^2 / (\alpha_2(R) + 1))
\]

\[
= \frac{[\int F_0(y) (\alpha_1(R) F_0(y) + 1) dG_0(y) / (\alpha_1(R) + 1) + \alpha_2(R) E(\int (1 - G_0(x^-)) dF(x))^2]}{\alpha_2(R) + 1}
\]

\[
= \frac{[(I_2 \alpha_1(R) + A_0^1 / (\alpha_1(R) + 1)) + \alpha_2(R) \int (1 - G_0(x^-))^2 dF_0(x) + \alpha_1(R) \Delta_0^2 / (\alpha_1(R) + 1)]}{\alpha_2(R) + 1},
\]

14
which is (3.6). (In the above derivation, use the second moment of
a beta distribution to get from the third equality to the fourth.)
To prove (3.7), expand \( \hat{\Delta}^* \) into ten sums, take expectations and use
Lemma 3.3 repeatedly. For example one of the sums in the expansion
\[
\sum_{j=1}^{n_2} \sum_{j'=1}^{n_2} F_0(x_{n+1,j}) F_0(x_{n+1,j'}) (\text{apart from a multiplicative constant}).
\]
To evaluate its expectation use (3.25) of Lemma 3.3. Equation (3.8)
is similarly proved by using Lemma 3.3. To simplify \( \hat{\Delta}_n^2 \) and \( \hat{\Delta}^{*2} \),
use Remark 3.4. The asymptotic optimality of \( D = \{ \Delta_n \} \) follows from
the fact that \( \lim_{n \to \infty} \hat{\Delta}_n^2 = \hat{\Delta}^{*2} \), which follows by letting \( n \to \infty \) in (3.8).

Note, that from (3.11), it is seen that the rate at which \( R_n(D, \alpha_1, \alpha_2) \)
converges to the minimum Bayes risk is \( 1/n \).

**BIBLIOGRAPHY**


discrete measures: An elementary proof. Unpublished manuscript.

Statist.* 1, 356-358.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via


distribution functions. (To appear in the Proceedings of the
Air Force Office of Scientific Research Conference on the Theory
and Applications of Reliability, Tampa, Florida, 1975.)


**Report Documentation Page**

<table>
<thead>
<tr>
<th>1. REPORT NUMBERS</th>
<th>2. GOVT. ACCESSION NO.</th>
<th>3. RECIPIENT'S CATALOG NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFSOR-76-3109 No. 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFSOR-74-2581B No. 64</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. TITLE</th>
<th>5. TYPE OF REPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>NONPARAMETRIC EMPirical BAYES ESTIMATION OF THE PROBABILITY THAT X ≤ Y</td>
<td>Technical Report</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7. AUTHOR(s)</th>
<th>8. CONTRACT OR GRANT NUMBER(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myles Hollander and Ramesh M. Korwar</td>
<td>AFSOR-76-3109 No. 1</td>
</tr>
<tr>
<td></td>
<td>AFSOR-74-2581B No. 64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9. PERFORMING ORGANIZATION NAME &amp; ADDRESS</th>
<th>10. PROGRAM ELEMENT, PROJECT, TASK AREA AND WORK UNIT NOS.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Florida State University</td>
<td></td>
</tr>
<tr>
<td>Department of Statistics</td>
<td></td>
</tr>
<tr>
<td>Tallahassee, Florida 32306</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>11. CONTROLLING OFFICE NAME &amp; ADDRESS</th>
<th>12. REPORT DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Force Office of Scientific Research</td>
<td>October, 1976</td>
</tr>
<tr>
<td>Bolling Air Force Base, D. C. 20332</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>13. NUMBER OF PAGES</th>
<th>14. MONITORING AGENCY NAME &amp; ADDRESS (if different from Controlling Office)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>15. SECURITY CLASS</th>
<th>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNCLASSIFIED</td>
<td></td>
</tr>
</tbody>
</table>

**6. DISTRIBUTION STATEMENT (of this Report)**
Approved for public release: distribution unlimited.

**7. DISTRIBUTION STATEMENT (of the abstract, if different from Report)**

**8. SUPPLEMENTARY NOTES**

**9. KEY WORDS**
Nonparametric estimation  Dirichlet process
Empirical Bayes estimator  Probability that X ≤ Y

**0. ABSTRACT**
A sequence of empirical Bayes estimators is defined for estimating, in a two-sample problem, the probability that X ≤ Y. The sequence is shown to be asymptotically optimal relative to a Ferguson Dirichlet process prior.