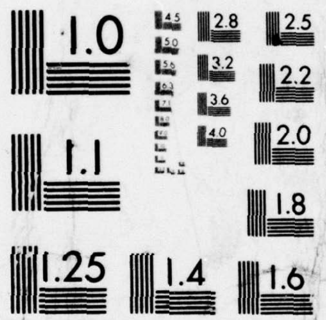
AD-A033 309 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1
NONPARAMETRIC EMPIRICAL BAYES ESTIMATION OF THE PROBABILITY THA--ETC(U)
OCT 76 M HOLLANDER, R M KORWAR AF-AFOSR-2581-76
UNCLASSIFIED FSU-STATISTICS-M389 AFOSR-TR-76-1214 NL

| OF |

AD
A033309

END

DATE
FILMED
1-77



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

The Florida State University

Department of Statistics

Tallahassee, Florida
32306

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is approved for publ.. release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

(4)

(2)

NONPARAMETRIC EMPIRICAL BAYES ESTIMATION OF THE PROBABILITY THAT [redacted]

X is less than or Equal to Y.

by

(10)

Myles Hollander[1] [redacted] Ramesh M. Korwar

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

(12) 19p.

(9) Technical rept.

(16) 2304

(17) A5

(11)

Octo[redacted]76

The Florida State University
Department of Statistics
Tallahassee, Florida

(14) FSU-Statistics-M389, TR-1

(15) AF-AFOSR-2581-76, AF-AFOSR-3109-76

(18) AFOSR

(19) TR-76-1217

[1]Research sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Grants AFOSR-74-2581B and AFOSR-76-3109.

400277

1B

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

Myles Hollander[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 Δ,

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

[1] This research was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grants AFOSR-74-2581B and AFOSR-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 Δ is as follows. [For Dirichlet process definitions see Section 2, and for more details, see Ferguson's (1973) paper.] To estimate Δ, 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 α_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 α_2, and P_1 and P_2 are independent. For squared error loss, Ferguson's Bayes estimator of Δ is given by

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

$$+ (1 - p_{1,n_1})p_{2,n_2}\frac{1}{n_1}\sum_{i=1}^{n_1}(1 - G_0(X_i^-))$$

$$+ (1 - p_{1,n_1})(1 - p_{2,n_2})\frac{1}{n_1 n_2} U \qquad (1.1)$$

where

$$p_{1,n_1} = \frac{\alpha_1(R)}{\alpha_1(R) + n_1}, \quad p_{2,n_2} = \frac{\alpha_2(R)}{\alpha_2(R) + n_2}, \qquad (1.2)$$

$$F_0(x) = \alpha_1((-\infty,x])/\alpha_1(R), \; G_0(y) = \alpha_2((-\infty,y])/\alpha_2(R), \qquad (1.3)$$

$$\Delta_0 = \int F_0 dG_0, \qquad (1.4)$$

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

$$U = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} I_{(-\infty,Y_j]}(X_i) \qquad (1.5)$$

is the Mann-Whitney statistic. Here

$$I_A(X) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases} \qquad (1.6)$$

Ferguson notes that the estimator $\hat{\Delta}^*$ is a simple mixture of four separate estimators of Δ. As both $\alpha_1(\mathcal{R})$ and $\alpha_2(\mathcal{R})$ tend to zero, $\hat{\Delta}^*$ converges to $(n_1n_2)^{-1}U$, the usual nonparametric estimator.

Motivated by Ferguson's $\hat{\Delta}^*$, we propose an empirical Bayes estimator of Δ which requires less prior information about $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$. Only $\alpha_1(\mathcal{R})$ and $\alpha_2(\mathcal{R})$ need be specified. Consider, then, the following set up appropriate for an empirical Bayes estimation problem. Let $\{X_i^{(1)}\}$, $\{X_i^{(2)}\}$, $i = 1, 2, \ldots$, be two independent sequences of independent vectors of observations from F and G respectively. Here $X_i^{(j)} = (X_{i1}^{(j)}, \ldots, X_{in_j}^{(j)})$, $j = 1, 2$ and $i = 1, 2, \ldots$. Assume independent Dirichlet priors on $(\mathcal{R}, \mathcal{B})$ with parameters α_1 and α_2 respectively for F and G. Here $\mathcal{R}$ is the real line and $\mathcal{B}$ is the σ-field of Borel subsets of $\mathcal{R}$. Let the action space be the closed interval [0, 1], and the loss function be

$$L(\Delta, \hat{\Delta}) = (\Delta - \hat{\Delta})^2, \tag{1.7}$$

where $\hat{\Delta}$ is an estimator of Δ. We assume $\alpha_1(\mathcal{R})$ and $\alpha_2(\mathcal{R})$ are known. We then propose the estimator $\hat{\Delta}_n$ below as an estimator of Δ on the (n+1)th occasion. The estimator is given by

$$\begin{aligned}
\hat{\Delta}_n &= \hat{\Delta}(X_1^{(1)}, \ldots, X_{n+1}^{(1)}; X_1^{(2)}, \ldots, X_{n+1}^{(2)}) \\
&= p_{1,n_1} p_{2,n_2} \frac{1}{n^2 n_1 n_2} \sum_{i=1}^{n} \sum_{j=1}^{n_2} \sum_{k=1}^{n} \sum_{\ell=1}^{n_1} I_{(-\infty, X_{ij}^{(2)}]}(X_{k\ell}^{(1)}) \\
&+ p_{1,n_1}(1-p_{2,n_2}) \frac{1}{n n_1 n_2} \sum_{j=1}^{n_2} \sum_{k=1}^{n} \sum_{\ell=1}^{n_1} \delta_{X_{k\ell}^{(1)}}((-\infty, X_{n+1,j}^{(2)}]) \\
&+ (1 - p_{1,n_1}) p_{2,n_2} \frac{1}{n_1} \sum_{i=1}^{n_1} \left[1 - \frac{1}{n n_2} \sum_{j=1}^{n} \sum_{k=1}^{n_2} \delta_{X_{jk}^{(2)}}((-\infty, X_{n+1,i}^{(1)})) \right] \\
&+ (1 - p_{1,n_1})(1 - p_{2,n_2}) \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I_{(-\infty, X_{n+1,j}^{(2)}]}(X_{n+1,i}^{(1)}),
\end{aligned} \tag{1.8}$$

where p_{1,n_i}, $i = 1, 2$, are given by (1.2), and

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

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 $D = \{\hat{\Delta}_n\}$ 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 $\mathcal{D}(\alpha_1,\ldots,\alpha_k)$, is defined as the distribution of $(Y_1,\ldots,Y_k)$, 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

$$f(y_1,\ldots,y_{k-1}|\alpha_1,\ldots,\alpha_k) \tag{2.1}$$

$$= \frac{\Gamma(\alpha_1+\ldots+\alpha_k)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_k)} \left(\prod_{i=1}^{k-1} y_i^{\alpha_i-1}\right)\left(1 - \sum_{i=1}^{k-1} y_i\right)^{\alpha_k-1} I_S(y_1,\ldots,y_{k-1}),$$

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, $\mathcal{B}e(\alpha_1, \alpha_2)$. Note that the condition $\alpha_i > 0$ for some $i = 1,\ldots,k$, is required in Definition 2.1 so that $\sum_{i=1}^{k} Z_i$ is not degenerate at 0.

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

Definition 2.2 (Ferguson, 1973). Let $(\mathcal{X}, \mathcal{A})$ be a measurable space. Let α be a non-null finite measure (nonnegative and finitely additive) on $(\mathcal{X}, \mathcal{A})$. Then P is a Dirichlet process on $(\mathcal{X}, \mathcal{A})$ with parameter α if for every $k = 1, 2,\ldots$, and measurable partition $(B_1,\ldots,B_k)$ of $\mathcal{X}$, 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 $\mathcal{X}$-valued random variables $X_1,\ldots,X_m$ constitute a sample of size m from a Dirichlet process P on $(\mathcal{X}, \mathcal{A})$ with parameter α if for any $\ell = 1, 2,\ldots$ and measurable sets $A_1,\ldots A_\ell$, $C_1,\ldots,C_m$, $Q\{X_1 \in C_1,\ldots,X_m \in C_m|P(A_1),\ldots,P(A_\ell),$

$P(C_1),\ldots,P(C_m)\} = \prod_{i=1}^{m} P(C_i)$ 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 distribution 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.

<u>Theorem 2.4</u> (Ferguson, 1973). Let P be a Dirichlet process on $(\mathcal{X}, \mathcal{A})$ with parameter α, 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 $(\mathcal{X}, \mathcal{A})$ with parameter $\beta = \alpha + \sum_{i=1}^{m} \delta_{X_i}$, where, for $x \in \mathcal{X}$, $A \in \mathcal{A}$, $\delta_x(A)$ is given by (1.9).

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

<u>Theorem 2.5</u>. Let P be a Dirichlet process on $(\mathcal{R}, \mathcal{B})$ with parameter α and let $X_1,\ldots,X_m$ be a sample of size m from P. Then

$$Q\{X_1 \le x_1,\ldots,X_m \le x_m\} = \{\alpha(A_{x_{(1)}})\ldots(\alpha(A_{x_{(m)}}) + m-1)\}/\{\alpha(\mathcal{R})\ldots(\alpha(\mathcal{R}) + m-1)\},$$

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

<u>Proof.</u> Observe that $A_{x_{(k)}} \subset A_{x_{(k+1)}}$, $k = 1,\ldots,m-1$. We write

$$A_{x_{(1)}} = B_1,$$

$$A_{x_{(k)}} = A_{x_{(1)}} + A^C_{x_{(1)}} A_{x_{(2)}} +\ldots+ A^C_{x_{(k-1)}} A_{x_{(k)}}$$

$$= B_1 + B_2 +\ldots+ B_k,\ k = 2,\ldots,m,$$

$$A^C_{x_{(m)}} = B_{m+1}.$$

Here A^C denotes the complement of the set A. Then

$$Q\{X_1 \le x_1, \ldots, X_m \le x_m\} = Q\{X_1 \epsilon A_{x_{(i_1)}}, \ldots, X_m \epsilon A_{x_{(i_m)}}\},$$

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

$$Q\{X_1 \epsilon A_{x_{(i_1)}}, \ldots, X_m \epsilon A_{x_{(i_m)}}\}$$

$$= E[Q\{X_1 \epsilon A_{x_{(i_1)}}, \ldots, X_m \epsilon A_{x_{(i_m)}} | P(A_{x_{(i_1)}}), \ldots, P(A_{x_{(i_m)}})\}]$$

$$= E\{P(A_{x_{(i_1)}}) \cdots P(A_{x_{(i_m)}})\},$$

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

$$E\{P(A_{x_{(i_1)}}) \cdots P(A_{x_{(i_m)}})\}$$

$$= E\{P(B_1)(P(B_1)+P(B_2)) \cdots (P(B_1) + \ldots + P(B_m))\}$$

$$= \{\alpha(A_{x_{(1)}})(\alpha(A_{x_{(2)}})+1) \cdots (\alpha(A_{x_{(m)}})+m-1)\}/\{\alpha(\mathcal{R})(\alpha(\mathcal{R})+1) \cdots$$

$$(\alpha(\mathcal{R})+m-1)\}. \quad ||$$

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

$$E \int Z_1 dP \int Z_2 dP = \{\sigma_{12}/(\alpha(\mathcal{X})+1)\} + \mu_1 \mu_2,$$

where

$$\mu_i = \int Z_i d\alpha / \alpha(\mathcal{X}), \quad i = 1,2$$

and

$$\sigma_{12} = \{\int Z_1 Z_2 d\alpha / \alpha(\mathcal{X})\} - \mu_1 \mu_2.$$

3. ASYMPTOTIC 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 Δ based on $(X_{n+1}^{(1)}, X_{n+1}^{(2)})$ is

$$\hat{\Delta}_n^* = p_1 p_2 \Delta_0 + p_1(1-p_2)\sum_{j=1}^{n_2} F_0(X_{n+1,j}^{(2)})/n_2 + (1-p_1)p_2 \sum_{i=1}^{n_1}(1-G_0(X_{n+1,i}^{(1)}\,^-))/n_1$$
$$+ (1-p_1)(1-p_2)\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} I_{(-\infty, X_{n+1,j}^{(2)}]}(X_{n+1,i}^{(1)})/n_1 n_2, \tag{3.1}$$

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), Δ_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

$$R(\hat{\Delta}^*, \alpha_1, \alpha_2) \overset{\text{def}}{\equiv} R(\hat{\Delta}_n, \alpha_1, \alpha_2) = E_{X_{n+1}^{(1)}, X_{n+1}^{(2)}} E_{F,G|X_{n+1}^{(1)}, X_{n+1}^{(2)}} (\Delta - \hat{\Delta}_n^*)^2, \tag{3.2}$$

and

$$R(\hat{\Delta}_n, \alpha_1, \alpha_2) = E_{X_{n+1}^{(1)}, X_{n+1}^{(2)}} E_{F,G|X_{n+1}^{(1)}, X_{n+1}^{(2)}} (\Delta - \hat{\Delta}_n)^2. \tag{3.3}$$

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 (α_1, α_2) if $R_n(D, \alpha_1, \alpha_2)$ converges to the minimum Bayes risk $R(\hat{\Delta}^*, \alpha_1, \alpha_2)$, as $n \to \infty$.

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

$$R(\hat{\Delta}^*, \alpha_1, \alpha_2) = E\Delta^2 - E\hat{\Delta}^{*2}, \tag{3.4}$$

$$R_n(D,\alpha_1,\alpha_2) = (E\Delta^2 - E\hat{\Delta}^{*2}) + (E\hat{\Delta}_n^2 - E\hat{\Delta}^{*2}), \tag{3.5}$$

where

$$\begin{aligned} E\Delta^2 = {} & p_{1,1}p_{2,1}\Delta_0^2 + (1-p_{1,1})(1-p_{2,1})\Delta_0 \\ & + (1-p_{1,1})p_{2,1}I_1 + p_{1,1}(1-p_{2,1})I_2, \end{aligned} \tag{3.6}$$

$$\begin{aligned} E\hat{\Delta}^{*2} = {} & g_1g_2\Delta_0^2 + (1-g_1)(1-g_2)\Delta_0 + (1-g_1)g_2I_1 \\ & + g_1(1-g_2)I_2, \end{aligned} \tag{3.7}$$

$$\begin{aligned} E\hat{\Delta}_n^2 = {} & h_1h_2\Delta_0^2 + (1-h_1)(1-h_2)\Delta_0 + (1-h_1)h_2I_1 \\ & + h_1(1-h_2)I_2, \end{aligned} \tag{3.8}$$

where

$$I_1 = \int F_0(y_{(1)})dG_0(y_1)dG_0(y_2), \tag{3.9}$$

$$I_2 = \int F_0^2(y)dG_0(y), \tag{3.10}$$

and

$$\begin{aligned} & g_1 \equiv g_{1,n_1} = \frac{p_{1,1}p_{1,n_1}}{p_{1,n_1+1}}, \quad g_2 \equiv g_{2,n_2} = \frac{p_{2,1}p_{2,n_2}}{p_{2,n_2+1}}, \\ & h_1 \equiv h_{1,n_1} = g_1\left(1-\frac{p_{1,n_1+1}}{nn_1}\right), \quad h_2 \equiv h_{2,n_2} = g_2\left(1-\frac{p_{2,n_2+1}}{nn_2}\right), \end{aligned} \tag{3.11}$$

In particular $D = \{\hat{\Delta}_n\}$ is asymptotically optimal relative to (α_1,α_2).

Our proof of Theorem 3.2 uses the following lemma.

<u>Lemma 3.3</u>. Let $X_i^{(1)} = (X_{i1}^{(1)},\ldots,X_{in_1}^{(1)})$, $i = 1,2$, be two independent samples, each of size n_1, from a Dirichlet process on $(\mathcal{R}, \mathcal{B})$ with parameter α_1, and let $X_i^{(2)} = (X_{i1}^{(2)},\ldots,X_{in_2}^{(2)})$, $i = 1,2$, be two independent samples, each of size n_2, independent of $(X_1^{(1)}, X_2^{(1)})$,

from an independent Dirichlet process on $(\mathcal{R}, \mathcal{B})$ with parameter α_2. Assume that each of α_1, α_2 is σ-additive. Then

$$EI_{(-\infty, X_{ij}^{(2)}]}(X_{k\ell}^{(1)}) = EF_0(X_{ij}^{(2)}) \tag{3.12}$$

$$E\delta_{X_{k\ell}^{(1)}}((-\infty, X_{ij}^{(2)}]) = EF_0(X_{ij}^{(2)}), \tag{3.13}$$

$$E\delta_{X_{ij}^{(2)}}([X_{k\ell}^{(1)}, \infty)) = EF_0(X_{ij}^{(2)}), \tag{3.14}$$

$$EF_0(X_{ij}^{(2)}) = \Delta_0, \tag{3.15}$$

$$E(1-G_0(X_{k\ell}^{(1)-})) = \Delta_0, \tag{3.16}$$

$$EF_0(X_{ij}^{(2)})I_{(-\infty, X_{i'j'}^{(2)}]}(X_{k\ell}^{(1)}) = EF_0(X_{ij}^{(2)})F_0(X_{i'j'}^{(2)}), \tag{3.17}$$

$$E(1-G_0(X_{k\ell}^{(1)-}))I_{(-\infty, X_{ij}^{(2)}]}(X_{k'\ell'}^{(1)}) = E(1-G_0(X_{k\ell}^{(1)-}))(1-G_0(X_{k'\ell'}^{(1)-})), \tag{3.18}$$

$$E\delta_{X_{k\ell}^{(1)}}((-\infty, X_{ij}^{(2)}])\delta_{X_{k'\ell'}^{(1)}}((-\infty, X_{i'j'}^{(2)}] = \begin{cases} EF_0(X_{ij}^{(2)})F_0(X_{i'j'}^{(2)}), & k \neq k' \\ \{\alpha_1(\mathcal{R})EF_0(X_{ij}^{(2)})F_0(X_{i'j'}^{(2)}) + EF_0(X_{(1)}^{(2)})\}/(\alpha_1(\mathcal{R}) + 1), & k=k',\ \ell\neq\ell' \\ EF_0(X_{(1)}^{(2)}) & ,k=k',\ \ell=\ell' \end{cases} \tag{3.19}$$

$$EI_{(-\infty, X_{ij}^{(2)}]}(X_{k\ell}^{(1)}) I_{(-\infty, X_{i'j'}^{(2)}]}(X_{k'\ell'}^{(1)}) =$$

$$\begin{cases} EF_0(X_{ij}^{(2)}) F_0(X_{i'j'}^{(2)}) & k \neq k' \\ \{\alpha_1(R) EF_0(X_{ij}^{(2)}) F_0(X_{i'j'}^{(2)}) + EF_0(X_{(1)}^{(2)})\} / (\alpha_1(R)+1), & k = k', \end{cases} \quad (3.20)$$

$$E\delta_{X_{ij}^{(2)}}([X_{k\ell}^{(1)}, \infty)) \delta_{X_{i'j'}^{(2)}}([X_{k\ell'}^{(1)}, \infty)) = \quad (3.21)$$

$$\begin{cases} \{\alpha_1(R) EF_0(X_{ij}^{(2)}) F_0(X_{i'j'}^{(2)}) + EF_0(X_{(1)}^{(2)})\} / (\alpha_1(R)+1), & \ell \neq \ell' \\ EF_0(X_{(1)}^{(2)}), & \ell = \ell' \end{cases}$$

$$EI_{(-\infty, X_{ij}^{(2)}]}(X_{k\ell}^{(1)}) \delta_{X_{k'\ell'}^{(1)}}((-\infty, X_{i'j'}^{(2)}]) =$$

$$\begin{cases} EF_0(X_{ij}^{(2)}) F_0(X_{i'j'}^{(2)}), & k \neq k' \\ \{\alpha_1(R) EF_0(X_{ij}^{(2)}) F_0(X_{i'j'}^{(2)}) + EF_0(X_{(1)}^{(2)})\} / (\alpha_1(R)+1), & k = k', \ \ell \neq \ell' \\ EF_0(X_{(1)}^{(2)}), & k = k', \ \ell = \ell', \end{cases} \quad (3.22)$$

$$EI_{(-\infty,X_{ij}^{(2)}]}(X_{k\ell}^{(1)})\delta_{X_{i'j'}^{(2)}}([X_{k'\ell'},\infty)) =$$

$$\begin{cases} EF_0(X_{ij}^{(2)})F_0(X_{i'j'}^{(2)}), & k\neq k', \\ \{\alpha_1(R)EF_0(X_{ij}^{(2)})F_0(X_{i'j'}^{(2)})+EF_0(X_{(1)}^{(2)})\}/(\alpha_1(R)+1), & k=k',\ \ell\neq\ell' \\ EF_0(X_{(1)}^{(2)}) & ,\ k=k',\ \ell=\ell'. \end{cases} \quad (3.23)$$

In (3.19) - (3.23), $X_{(1)}^{(2)}$ is the smaller of $X_{ij}^{(2)}$ and $X_{i'j'}^{(2)}$.

Also,

$$EF_0(X_{(1)}^{(2)}) = \begin{cases} \int F_0(y_{(1)})dG_0(y_1)dG_0(y_2), & i\neq i' \\ \int F_0(y_{(1)})dK_0(y_1,y_2), & i=i', \end{cases} \quad (3.24)$$

$$EF_0(X_{ij}^{(2)})F_0(X_{i'j'}^{(2)}) = \begin{cases} \Delta_0^2 & i\neq i' \\ \int F_0(y_1)F_0(y_2)dK_0(y_1,y_2) & i=i',j\neq j' \\ \int F_0^2(y)dG_0(y) & i=i',j=j', \end{cases} \quad (3.25)$$

where $y_{(1)}$ is the smaller of y_1 and y_2, and

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

$$\begin{cases} \Delta_0^2 & ,\ k\neq k' \\ \int(1-G_0(x_1^-))(1-G_0(x_2^-))dH_0(x_1,x_2), & k=k',\ell\neq\ell' \\ \int(1-G_0(x^-))^2 dF_0(x) & k=k',\ell=\ell'. \end{cases} \quad (3.26)$$

In (3.24) - (3.26), $H_0(x_1,x_2)$ and $K_0(y_1,y_2)$ are the distribution functions of $(X_{k\ell}^{(1)},X_{k\ell'}^{(1)})$, $\ell\neq\ell'$ and of $(X_{ij}^{(2)},X_{ij'}^{(2)})$, $j\neq j'$ 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 $X_k^{(1)}$ and $X_i^{(2)}$ and Theorem 2.5. We prove (3.12):

$$EI_{(-\infty,X_{ij}^{(2)}]}(X_{k\ell}^{(1)}) = E(E(I_{(-\infty,X_{ij}^{(2)}]}(X_{k\ell}^{(1)})|X_{ij}^{(2)})) = EF_0(X_{ij}^{(2)}).$$

Equations (3.15) - (3.16) are readily verified from Theorem 2.5 and the definition of Δ_0. Equation (3.17) follows from Theorem 2.5 and the independence of $X_k^{(1)}$ from $X_i^{(2)}$ and $X_{i'}^{(2)}$, and (3.18) follows from Theorem 2.5 and the independence of $X_i^{(2)}$ from $X_k^{(1)}$ and $X_{k'}^{(1)}$.

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

$$E\delta_{X_{k\ell}^{(1)}}((-\infty,X_{ij}^{(2)}])\delta_{X_{k'\ell'}^{(1)}}((-\infty,X_{i'j'}^{(2)}])$$

$$= E(E(\delta_{X_{k\ell}^{(1)}}((-\infty,X_{ij}^{(2)}])\delta_{X_{k'\ell'}^{(1)}}((-\infty,X_{i'j'}^{(2)}])|X_{ij}^{(2)},X_{i'j'}^{(2)}))$$

$$= \begin{cases} EF_0(X_{ij}^{(2)})F_0(X_{i'j'}^{(2)}), \ k\neq k' \text{ (by the independence of } X_k^{(1)},X_{k'}^{(1)}) \\ \dfrac{E\alpha_1((-\infty,X_{(1)}^{(2)}])(\alpha_1((-\infty,X_{(2)}^{(2)}])+1)}{\alpha_1(R)(\alpha_1(R)+1)}, \ k=k',\ \ell\neq\ell' \\ \dfrac{E\alpha_1((-\infty,X_{(1)}^{(2)}])}{\alpha_1(R)} \quad k=k',\ \ell=\ell', \end{cases}$$

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 $X^{(2)}_{ij'}$, given $X^{(2)}_{ij}$ $(j\neq j')$, we can show that

$$\int F_0(y_{(1)})dK_0(y_1,y_2) = [\alpha_2(R)I_1+\Delta_0]/(\alpha_2(R)+1), \qquad (3.27)$$

$$\int F_0(y_1)F_0(y_2)dK_0(y_1,y_2) = [\alpha_2(R)\Delta_0^2+I_2]/(\alpha_2(R)+1). \qquad (3.28)$$

The function $K_0(y_1,y_2)$ is the distribution function of $X^{(2)}_{ij}$, $X^{(2)}_{ij'}$ $(j\neq j')$, and $y_{(1)}$ is $\min(y_1,y_2)$. Similarly, by considering the conditional distribution of $X^{(1)}_{k\ell'}$ given $X^{(1)}_{k\ell}$ $(\ell\neq\ell')$ we can show

$$\int(1-G_0(x_1^-))(1-G_0(x_2^-))dH_0(x_1,x_2) = [\alpha_1(R)\Delta_0^2+I_1]/(\alpha_1(R)+1). \qquad (3.29)$$

Proof of Theorem 3.2. To prove (3.4), expand $(\Delta - \hat{\Delta}^*)^2$, and note that $\hat{\Delta}^*$, being the Bayes estimator with squared error loss, is $E(\Delta|X^{(1)}_{n+1}, X^{(2)}_{n+1})$. To prove (3.5), expand $(\Delta - \hat{\Delta}_n)^2$, and note that

$E(\Delta\hat{\Delta}_n) = E(E(\Delta\hat{\Delta}_n|X^{(1)}_1,X^{(2)}_1,\ldots,X^{(1)}_{n+1},X^{(2)}_{n+1})) = E(\hat{\Delta}_n E(\Delta|X^{(1)}_{n+1},X^{(2)}_{n+1}))$

$= E(\hat{\Delta}_n\hat{\Delta}^*) = E(\hat{\Delta}^* E(\hat{\Delta}_n|X^{(1)}_{n+1},X^{(2)}_{n+1})) = E\hat{\Delta}^{*2}$. (Here we used the fact

that $E(\hat{\Delta}_n|X^{(1)}_{n+1},X^{(2)}_{n+1}) = \hat{\Delta}^*$, which is easily verified.) To prove (3.6), use Theorem 2.6 twice. Thus,

$$
\begin{aligned}
E\Delta^2 &= E(\int F(y)dG(y))^2 \\
&= E(E(\int F(y)dG(y))^2|F) \\
&= E\{(\int F^2(y)dG_0(y)/(\alpha_2(R)+1))+\alpha_2(R)(\int F(y)dG_0(y))^2/(\alpha_2(R)+1)\} \\
&= \frac{\{(\int F_0(y)(\alpha_1(R)F_0(y)+1)dG_0(y)/(\alpha_1(R)+1))+\alpha_2(R)E(\int(1-G_0(x^-))dF(x))^2\}}{\alpha_2(R)+1} \\
&= \frac{[(\{I_2\alpha_1(R)+\Delta_0\}/(\alpha_1(R)+1))+\alpha_2(R)\{\int(1-G_0(x^-))^2dF_0(x)+\alpha_1(R)\Delta_0^2\}/(\alpha_1(R)+1)]}{\alpha_2(R)+1},
\end{aligned}
$$

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 $\hat{\Delta}^{*2}$ into ten sums, take expectations and use Lemma 3.3 repeatedly. For example one of the sums in the expansion is $\sum_{j=1}^{n_2} \sum_{j'=1}^{n_2} F_0(X_{n+1,j}^{(2)})F_0(X_{n+1,j'}^{(2)})$ (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 $E\hat{\Delta}_n^2$ and $E\hat{\Delta}^{*2}$, use Remark 3.4. The asymptotic optimality of $D = \{\hat{\Delta}_n\}$ follows from the fact that $\lim_{n\to\infty} E\hat{\Delta}_n^2 = E\hat{\Delta}^{*2}$, 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

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. Ann. Statist. 2, 1152-1174.

Berk, R. H. and Savage, I. R. (1975). Dirichlet processes produce discrete measures: An elementary proof. Unpublished manuscript.

Blackwell, D. (1973). Discreteness of Ferguson selections. Ann. Statist. 1, 356-358.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes. Ann. Statist. 1, 353-355.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. Ann. Statist. 1, 209-230.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. Ann. Statist. 2, 615-629.

Hollander, M. and Korwar, R. M. (1976). Nonparametric estimation of 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.)

Korwar, R. M. and Hollander, M. (1976). Empirical Bayes estimation of a distribution function. Ann. Statist. 4, 580-587.

Robbins, H. E. (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Statist. 35, 1-20.

Susarla, V. and Van Ryzin, J. (1976). Empirical Bayes estimation of a distribution (survival) function from right censored observations. University of Wisconsin Technical Report.

UNCLASSIFIED

Security Classification of This Page

REPORT DOCUMENTATION PAGE

1. REPORT NUMBERS AFOSR-76-3109 No. 1 AFOSR-74-2581B No. 64	2. GOVT. ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE NONPARAMETRIC EMPIRICAL BAYES ESTIMATION OF THE PROBABILITY THAT X ≤ Y		5. TYPE OF REPORT Technical Report 6. PERFORMING ORGANIZATION REPORT NO. FSU Statistical Report No. M389
7. AUTHOR(s) Myles Hollander and Ramesh M. Korwar		8. CONTRACT OR GRANT NUMBER(s) AFOSR-76-3109 No. 1 AFOSR 74-2581B No. 64
9. PERFORMING ORGANIZATION NAME & ADDRESS The Florida State University Department of Statistics Tallahassee, Florida 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA AND WORK UNIT NOS.
11. CONTROLLING OFFICE NAME & ADDRESS Air Force Office of Scientific Research Bolling Air Force Base, D. C. 20332		12. REPORT DATE October, 1976 13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited.

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

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Nonparametric estimation
Empirical Bayes estimator
Dirichlet process
Probability that X ≤ Y

20. 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.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	AFOSR - TR - 76 - 1214
2. GOVT ACCESSION NO.	
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle)	NONPARAMETRIC EMPIRICAL BAYES ESTIMATION OF THE PROBABILITY THAT X IS LESS THAN OR EQUAL TO Y
5. TYPE OF REPORT & PERIOD COVERED	Interim
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s)	Myles Hollander and Ramesh M. Korwar
8. CONTRACT OR GRANT NUMBER(s)	AFOSR 76-3109
9. PERFORMING ORGANIZATION NAME AND ADDRESS	Florida State University Department of Statistics Tallahassee, FL 32306
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS	Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332
12. REPORT DATE	October 1976
13. NUMBER OF PAGES	16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	
15. SECURITY CLASS. (of this report)	UNCLASSIFIED
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report)	Approved for public release; distribution unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)	
18. SUPPLEMENTARY NOTES	
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)	Nonparametric estimation Empirical Bayes estimator Dirichlet process Probability that $X \leq Y$

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

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.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)