NONPARAMETRIC BAYES ESTIMATION WITH INCOMPLETE DIRICHLET PRIOR INFORMATION*

BY

Gregory Campbell and Myles Hollander
Purdue University and Florida State University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #498

July 1977

*Research sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Grants AFOSR-74-2581B and AFOSR-76-3109. The United States Government is authorized to reproduce and distribute reprints for governmental purposes.

NONPARAMETRIC BAYES ESTIMATION WITH INCOMPLETE DIRICHLET PRIOR INFORMATION*

by

Gregory Campbell and Myles Hollander

Purdue University and Florida State University

1. Introduction and Summary

In this paper we treat the topic of incomplete information regarding the parameter α of a Dirichlet process prior. Ferguson [4] introduced the Dirichlet process for the incorporation of prior information into the analysis of nonparametric problems. The process can be viewed as a prior on the set of all distributions on a measureable space $(\mathcal{X},\mathcal{A})$. The process is parameterized by α, a non-negative, non-null finite measure on $(\mathcal{X},\mathcal{A})$. (In this paper we restrict to situations where $\mathcal{X} = \mathcal{R}$, the real line, and $\mathcal{A} = \mathcal{B}$, the Borel σ-field.) Typically, to use estimators which are Bayes with respect to a Dirichlet process with parameter α, the statistician must provide a complete specification of the measure α. This paper develops some estimators that rely only on partial information concerning α.

One approach to incomplete information concerning α is that initiated by Doksum [3]. Doksum assumes that $\alpha(t_i,t_{i+1}]$, $i=1,\ldots,k-1$, are known with $\alpha(\mathcal{R} - (t_1,t_k]) = 0$. That is, the values that α assigns to the k-1 intervals $(t_1,t_2],\ldots,(t_{k-1},t_k]$ are known, and $\alpha(\mathcal{R}) = \alpha(t_1,t_k]$. In Section 3 of this paper, Doksum's technique for obtaining a mixed rule (Definition 3.1) is considered and shown also to yield a $\mathcal{G}$-minimax rule (Definition 3.2) for a suitable choice of $\mathcal{G}$.

*Research sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Grants AFOSR-74-2581B and AFOSR-76-3109. The United States Government is authorized to reproduce and distribute reprints for governmental purposes.

Section 4 considers the estimation of $\Delta = \Pr\{X \leq Y\}$, when $X_1,\ldots,X_m$ is a sample from a Dirichlet process with parameter α and $Y_1,\ldots,Y_n$ is a sample from a second, independent Dirichlet process with parameter β. A mixed rule is found to be

$$\hat{\Delta}_k = \frac{\sum_{i=1}^{k-1}\{\alpha(t_1,t_i] + M_1+\ldots+M_i + \tfrac{1}{2}\alpha(t_i,t_{i+1}] + \tfrac{1}{2}M_{i+1}\}\{\beta(t_i,t_{i+1}] + N_i\}}{(\alpha(\mathcal{R})+m)(\beta(\mathcal{R})+n)}, \tag{1.1}$$

where M_j and N_j denote the number of X's and Y's, respectively, in the interval $(t_j,t_{j+1}]$.

In Section 5 the problem considered is the estimation of the rank order (Definition 5.1) of X_1 among $X_1,\ldots,X_n$ based on $X_1,\ldots,X_r (r < n)$, where $X_1,\ldots,X_n$ is a sample of size n from a Dirichlet process on $(\mathcal{R},\mathcal{B})$ with parameter α. For the case where α is completely specified, a Bayes estimator was developed by Campbell and Hollander [2]. Here a mixed rule is obtained for the case where α is not completely known but instead only the $\alpha(t_i,t_{i+1}]$ values, i = 1,...,k-1 (with $\alpha(\mathcal{R}) = \sum_{i=1}^{k-1} \alpha(t_i,t_{i+1}]$), are specified.

Section 2 contains some Dirichlet process preliminaries.

2. Dirichlet Process Preliminaries

Let $\mathcal{G}(\alpha,\beta)$ denote the gamma distribution with shape parameter $\alpha \geq 0$ and scale parameter $\beta > 0$. If $\alpha = 0$, the distribution is degenerate at 0. If $\alpha > 0$, it has a density with respect to Lebesgue measure on the real line given by:

$$f(z|\alpha,\beta) = (\Gamma(\alpha)\beta^{\alpha})^{-1} z^{\alpha-1} \exp(-z/\beta) I_{(0,\infty)}(z), \tag{2.1}$$

where $I_A(\cdot)$ denotes the indicator function on the set A.

Definition 2.1. The Dirichlet distribution with parameter $(\alpha_1,\dots,\alpha_k)$ where $\alpha_j \geq 0$ for all j and $\sum_{j=1}^{k} \alpha_j > 0$, denoted $\mathcal{D}(\alpha_1,\dots,\alpha_k)$, is defined as the distribution of $(Y_1,\dots,Y_k)$, where

$$Y_j = Z_j / \sum_{\ell=1}^{k} Z_\ell, \qquad j = 1,\dots,k,$$

and the Z_i's are independent random variables with gamma distributions $\mathcal{G}(\alpha_i,1)$, for $i = 1,\dots,k$.

If $\alpha_j > 0$ for all $j = 1,\dots,k$, the (k-1)-dimensional distribution of $(Y_1,\dots,Y_{k-1})$ is absolutely continuous with respect to Lebesgue measure on the (k-1)-dimensional Euclidean space with density

$$f(y_1,\dots,y_{k-1}|\alpha_1,\dots,\alpha_k) = \frac{\Gamma(\alpha_1+\dots+\alpha_k)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_k)} \left[\prod_{i=1}^{k-1} y_i^{\alpha_i-1}\right] \left[1 - \sum_{i-1}^{k-1} y_i\right]^{\alpha_k-1} I_S(y_1,\dots,y_{k-1}), \tag{2.2}$$

where S is the simplex

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

For k = 2, (2.2) becomes the density of a beta distribution with parameters α_1 and α_2.

Proposition 2.2. (Wilks [6],p. 179). The $r_1,\dots,r_\ell$ moment of the Dirichlet distribution $\mathcal{D}(\alpha_1,\dots,\alpha_k)$ is, for $\ell \leq k$ and r_i, a nonnegative integer such that r_i positive implies α_i positive, for $i = 1,\dots,\ell$:

$$\mu_{r_1,\dots,r_\ell} = \frac{\Gamma(\alpha_1+r_1)\dots\Gamma(\alpha_\ell+r_\ell)\Gamma(\alpha)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_\ell)\Gamma(\alpha+r)}, \tag{2.3}$$

where $\alpha = \sum_{i=1}^{k} \alpha_i$ and $r = \sum_{j=1}^{\ell} r_j$.

For k a positive integer, let $y^{[k]}$ denote the ascending factorial $y(y+1)\dots(y+k-1)$ and define $y^{[0]} \equiv 1$. Then it is convenient to rewrite (2.3) as

$$\mu_{r_1,\dots,r_\ell} = \alpha_1^{[r_1]} \dots \alpha_\ell^{[r_\ell]} / \alpha^{[r]}. \tag{2.4}$$

For a more complete treatment of the Dirichlet distribution, the reader is referred to Wilks [6].

Let $(\mathcal{X},\mathcal{A})$ denote a measurable space. A particular stochastic process $\{P(A)\colon\ A \in \mathcal{A}\}$ is defined.

<u>Definition 2.3</u>. (Ferguson [4]). Let α denote a non-negative, non-null, finite measure on $(\mathcal{X},\mathcal{A})$. P is a Dirichlet process on $(\mathcal{X},\mathcal{A})$ with parameter α if, for every $k = 1,2,\dots,$ and every measurable partition $(B_1,\dots,B_k)$ of $\mathcal{X}$, the distribution of $(P(B_1),\dots,P(B_k))$ is Dirichlet with parameter $(\alpha(B_1),\dots,\alpha(B_k))$.

Ferguson [4] shows, using the Kolmogorov extension theorem, that there exists a probability measure, call it Q_α, on $([0,1]^{\mathcal{A}}, \mathcal{B}^{\mathcal{A}})$ yielding the above finite-dimensional marginal Dirichlet distributions. Here $[0,1]^{\mathcal{A}}$ represents the space of all functions from $\mathcal{A}$ into $[0,1]$ (which thus includes $\mathcal{P}$, the set of all probability measures on $(\mathcal{X},\mathcal{A})$) and $\mathcal{B}^{\mathcal{A}}$ the σ-field generated by fields of cylinder sets.

<u>Definition 2.4</u>. (Ferguson [4]). The collection of random variables $X_1,\dots,X_n$ is said to be a sample of size n from the Dirichlet process P on $(\mathcal{X},\mathcal{A})$ with parameter α if, for any $m = 1,2,\dots,$ and measurable sets $A_1,\dots,A_m$, $C_1,\dots,C_n$,

$$\Pr\{X_1 \in C_1,\dots,X_n \in C_n | P(A_1),\dots,P(A_m),P(C_1),\dots,P(C_n)\} = \prod_{j=1}^{n} P(C_j), \tag{2.5}$$

where Pr denotes probability.

Intuitively, $X_1,\ldots,X_n$ is a sample of size n from a Dirichlet process if P is randomly selected according to Q_α and then, given P, $X_1,\ldots,X_n$ is a sample from the probability measure P.

Using Kolmogorov's extension theorem once again, Ferguson shows that there exists a probability measure on $(\mathcal{X}^n\times[0,1]^{\mathcal{G}},\mathcal{G}^n\times\mathcal{B}\mathcal{F}^{\mathcal{G}})$ with marginal probability on $([0,1]^{\mathcal{G}},\mathcal{B}\mathcal{F}^{\mathcal{G}})$ given by the above Q_α. Since this probability also depends on α, it also will be called Q_α. It can be shown (cf. Berk and Savage [1]) that Q_α concentrates all its mass on $(\mathcal{X}^n\times\mathcal{P},\mathcal{G}^n\times\sigma(\mathcal{P}))$, where $\sigma(\mathcal{P})$ is the inherited σ-field for $\mathcal{P}$ from $\mathcal{B}\mathcal{F}^{\mathcal{G}}$. Thus, P is a random probability measure. If $F(x) = P(-\infty,x]$, then F is a random distribution function, a sample path of the Dirichlet process.

Theorem 2.5. (Ferguson [4]). If P is a Dirichlet process on $(\mathcal{X},\mathcal{G})$ with parameter α, and if $X_1,\ldots,X_n$ is a sample of size n from P, then the conditional distribution of P given $X_1,\ldots,X_n$ is also a Dirichlet process on $(\mathcal{X},\mathcal{G})$ with parameter $\alpha + \sum_{i=1}^{n} \delta_{X_i}$, where δ_z denotes the measure with mass one at z, zero elsewhere.

3. Mixed Rules and $\mathcal{G}$-Minimax Rules

Doksum [3] considered the problem of partial prior information in the decision theoretic framework, in particular, as applied to nonparametric problems with Dirichlet parameters incompletely specified. It is assumed throughout this section that $\alpha(t_i, t_{i+1}]$, $i = 1,\ldots,k-1$, are known and that $\alpha(\mathcal{R}-(t_1,t_k]) = 0$.

Let Ω be a class of distribution functions of $(\mathcal{R},\mathcal{B})$, where $\mathcal{R}$ is the real line and $\mathcal{B}$ the Borel σ-field. Suppose that Q, the probability on Ω, is not completely specified but that, for fixed real numbers $t_1,\ldots,t_k$, the distribution of $(F(t_1),\ldots,F(t_k))$ is known, where F is a random distribution function from Q. Let L(F,a) denote the loss function for action a for distribution function $F \in \Omega$ and d a decision rule from the observation space $\mathcal{R}$ to the action space $\mathcal{A}$. Then the risk function R(F,d), associated with distribution function $F \in \Omega$ when decision rule d is taken, is defined by

$$R(F,d) = EL(F,d(X)),$$

where the expectation is over X, where X has distribution F. The maximum risk, R(d), is given by

$$R(d) = \sup_{F \in \Omega} R(F,d).$$

A rule (if one exists) which minimizes the maximum risk over all decision rules is called a minimax rule. The average risk, R(Q,d), for completely specified probability Q on Ω, is given by

$$R(Q,d) = \int_{\Omega} R(F,d)dQ(F).$$

A rule (if one exists) is called a Bayes rule if it minimizes the average risk over all decision rules.

<u>Definition 3.1</u>. (Doksum [3]). Let $\Omega(q,k) = \{F \in \Omega: F(t_i) = q_i\}$ for $q = (q_1,\ldots,q_k) \in \mathcal{R}^k$. Let the measure λ on $\mathcal{R}^k$, dependent on Q, be given by

$$\lambda(q;\, Q,k) = Q\{F \in \Omega: F(t_i) \le q_i,\ i = 1,\ldots,k\}.$$

λ is then the distribution of $F(t_1),\ldots,F(t_k)$ under the probability measure Q. The average maximum risk, $r_k(Q,d)$, associated with probability Q and decision rule d, is

$$r_k(Q,d) = \int_{\mathfrak{R}^k} [\sup_{F\in\Omega(q,k)} R(F,d)]d\lambda(q).$$

A rule is said to be mixed (or mixed Bayes-minimax) if it minimizes the average maximum risk over all decision rules.

Definition 3.2. Let $\mathcal{G}$ denote a set of probability measures on Ω. Define the $\mathcal{G}$-maximum risk for rule d as $\sup_{Q\in\mathcal{G}} R(Q,d)$. A rule (if it exists) is said to be $\mathcal{G}$-minimax if the rule minimizes the $\mathcal{G}$-maximum risk over all decision rules.

If Q_F denotes the probability on Ω which is the distribution function F with probability one, then a $\mathcal{G}$-minimax rule is minimax if $\mathcal{G}$ contains Q_F for all $F \in \Omega$.

A natural question is what are the relationships between these various risks and their associated rules. Doksum [3] provides a partial answer.

Lemma 3.3 (Doksum [3]). For any decision rule d and prior Q on Ω, the following hold:

(i) $R(d) \geq r_k(Q,d) \geq R(Q,d)$ $\quad (k \geq 1)$;

(ii) if $\{\Pi_m: \; t_{m,1}<\dots<t_{m,k_m}\}_{m=1}^{\infty}$ is a sequence of partitions such that each partition is a refinement of the previous one, then

$$r_{k_m}(Q,d) \geq r_{k_\ell}(Q,d) \qquad \text{for} \quad m < \ell.$$

Definition 3.4. The carrier of a given distribution is the smallest compact set whose probability under the given distribution is one. For example, for $F \in \Omega$, the carrier of F, denoted C(F), is the smallest compact set on $\mathfrak{R}$ whose probability under distribution F is one.

Definition 3.5. The support of Ω, $S(\Omega)$, is given by

$$S(\Omega) = \bigcup_{F \in \Omega} C(F).$$

Proposition 3.6. If $Q \in \mathcal{G}$, then, for every d,

$$R(d) \geq \sup_{Q' \in \mathcal{G}} R(Q',d) \geq R(Q,d).$$

Proof. Clearly, $\sup_{Q' \in \mathcal{G}} R(Q',d) \geq R(Q,d)$ since $Q \in \mathcal{G}$. But also, for Q_F as defined previously, if $\mathcal{G}^* = \mathcal{G} \cup \{Q_F: F \in \Omega\}$, then $\sup_{Q' \in \mathcal{G}^*} R(Q',d) = R(d) \geq \sup_{Q' \in \mathcal{G}} R(Q',d)$.||

Doksum defines a rule, which, in some cases, is a mixed rule. Let $t_1 = \inf\{t: t \in S(\Omega)\}$ and let $t_k = \sup\{t: t \in S(\Omega)\}$ and assume $-\infty < t_1 < t_k < \infty$. Let $F_{q,k}$ denote the polygonal distribution function with $F(t_i) = q_i$ for $i = 1,\ldots,k$ and $F_{q,k}$ linear on $[t_i,t_{i+1}]$ for $i = 1,\ldots,k-1$. Let F_k denote the random distribution function obtained by letting q in $F_{q,k}$ have distribution $\lambda = \lambda(\cdot; Q, k)$, for Q a prior on Ω. Assume F_k is measurable. Let Q_k denote the distribution of F_k and d_k the Bayes rule for Q_k (if it exists).

Theorem 3.7. (Doksum [3]). If $F_{q,k} \in \Omega$ for almost all q in $C(\lambda)$, if such a d_k exists, and if $r_k(Q,d_k) = R(Q_k,d_k)$, then d_k is a mixed procedure.
Theorem 3.7 provides a method for obtaining a mixed rule; i.e., one finds the Bayes rule for prior Q_k, and, if the hypotheses are satisfied, the Bayes rule is a mixed rule.

Let $\mathcal{G}_k$ = {Q a probability on Ω: $(F(t_1), F(t_2) - F(t_1),\ldots,F(t_k)-F(t_{k-1}))$ has a fixed, known distribution}.

Proposition 3.8. For any decision rule d and for $Q \in \mathcal{G}_k$,

$$r_k(Q,d) \geq \sup_{Q' \in \mathcal{G}_k} R(Q',d).$$

Proof. For $Q', Q'' \in \mathcal{G}_k$, $\lambda(q; Q',k) = \lambda(q; Q'',k)$ for all $q \in \mathcal{R}^k$, in that λ depends on F at $t_1,\ldots,t_k$, for F a random distribution function. Therefore, $r_k(Q',d) = r_k(Q'',d)$ for all rules d. Taking sups over $Q' \in \mathcal{G}_k$ on both sides of the inequality

$$r_k(Q',d) \geq R(Q',d),$$

obtained by Proposition 3.6, yields, for any $Q'' \in \mathcal{G}_k$,

$$r_k(Q'',d) \geq \sup_{Q' \in \mathcal{G}} R(Q',d).$$

In particular, $Q \in \mathcal{G}_k$ and the proof is complete. ||

Corollary 3.9. If Proposition 3.8 holds and if, for distribution Q on Ω, the Bayes risk equals the mixed risk associated with mixed rule d, then d is also a $\mathcal{G}_k$-minimax rule.

Proof. For a Bayes rule δ,

$$R(d) \geq r_k(Q,d) \geq \sup_{Q' \in \mathcal{G}_k} R(Q',d) \geq R(Q,d) \geq R(Q,\delta),$$

by Lemma 3.3 and Propositions 3.6 and 3.8. Now note by assumption, $r_k(Q,d) = R(Q,\delta)$, so $\sup_{Q' \in \mathcal{G}_k} R(Q',d) = R(Q,\delta)$. Therefore, d is $\mathcal{G}_k$-minimax. ||

The significance of Corollary 3.9 is that, in certain special instances, a $\mathcal{G}_k$-minimax rule can be found by finding a Bayes rule.

Let $\{\Pi_m\colon\ t_{m,1} < \ldots < t_{m,k_m}\}_{m=1}^{\infty}$ be a sequence of partitions such that each partition is a refinement of the preceding one and such that $|t_{m,i+1} - t_{m,i}| \to 0$ as $m \to \infty$. Further, suppose the t's are from the space [0,1]. Let C[0,1] denote the continuous distribution functions defined on [0,1]. For partition Π_k, let d_k denote a mixed rule for the given probability Q on C[0,1].

Theorem 3.10. (Doksum [3]). Let $F_{q,k}$ denote the polygonal distribution function with $F(t_{k,i}) = q_i$ and $F_{q,k}$ linear on $[t_{k,i}, t_{k,i+1}]$ for $i = 1,\ldots,\ell_k$. If $F_{q,k} \in \Omega$ for almost all q in $C(\lambda)$, if d_k denotes the mixed rule for probability Q on Ω associated with partition k, and if d is a Bayes rule such that d has continuous bounded risk R(Q,d), then, for $\Omega \subset C[0,1]$,

$$\lim_{k\to\infty} r_k(Q,d_k) = \lim_{k\to\infty} R(Q,d_k) = R(Q,d).$$

Theorem 3.11. Under the conditions of Theorem 3.10, if $\mathcal{G}_k$-minimax rules δ_k exist for $k = 1,2,\ldots$, then, for $Q \in \mathcal{G}_k$ for $k = 1,2,\ldots$,

$$\lim_{k\to\infty} \sup_{Q'\in\mathcal{G}_k} R(Q',\delta_k) = R(Q,d).$$

Proof. It follows from Propositions 3.6 and 3.8 and the definition of a $\mathcal{G}_k$-minimax rule that

$$R(Q,d) \le R(Q,\delta_k) \le \sup_{Q'\in\mathcal{G}_k} R(Q',\delta_k) \le \sup_{Q'\in\mathcal{G}_k} R(Q',d_k) \le r_k(Q,d_k).$$

Thus, by Theorem 3.10,

$$\lim_{k\to\infty} \sup_{Q'\in\mathcal{G}_k} R(Q',\delta_k) = R(Q,d). \quad ||$$

The importance of Theorem 3.11 is that, if $\mathcal{G}_k$-minimax rules exist and the conditions of the theorem are satisfied, the associated $\mathcal{G}_k$-minimax risk approaches the Bayes risk.

The application of this development to the Dirichlet situation will become apparent immediately. Let $\mathcal{G}_k = \{$Q a probability measure on Ω: $(F(t_2) - F(t_1),\ldots,F(t_k) - F(t_{k-1}))$ has a Dirichlet distribution with parameters $(\alpha(t_1,t_2],\ldots,\alpha(t_{k-1},t_k]\}$. Then $\mathcal{G}_k$-minimax rules are exactly those rules for which α is known only on (k-1) intervals. The search for $\mathcal{G}_k$-minimax rules will be conducted by means of Corollary 3.9. The behavior of such rules as $k \to \infty$, under the conditions enumerated, is given in Theorem 3.11.

The remaining two sections contain applications of this development. Section 4 treats estimation of $\Pr(X \leq Y)$ under incomplete Dirichlet prior information. Section 5 considers estimation of a rank order under Dirichlet incomplete prior information.

4. Estimation of $\Pr\{X \leq Y\}$ Under Partial Prior Information

Consider the problem of estimating $\Pr\{X \leq Y\}$ in the two-sample situation under incomplete Dirichlet prior information. In particular, assume $X_1,\ldots,X_m$ is a sample of size m from a Dirichlet process on $(\mathcal{R},\mathcal{B})$ with parameter α and $Y_1,\ldots,Y_n$ a sample of size n from a second Dirichlet process (independent of the first process) on $(\mathcal{R},\mathcal{B})$ with parameter β. Further, assume that $t_1,\ldots,t_k$ are fixed such that $\alpha(t_i, t_{i+1}]$ and $\beta(t_i, t_{i+1}]$ are known for $i = 1,\ldots,k-1$ and that $\alpha(\mathcal{R} - (t_1,t_k]) = \beta(\mathcal{R} - (t_1,t_k]) = 0$. The parameter of interest is $\Delta(F,G) = \Pr\{X \leq Y\} = \int F dG$ where F is the random distribution function from the first Dirichlet process and G the random distribution function from the second process. Let F_k and G_k denote the polygonal random distribution functions with $F_k(t_i) = F(t_i)$, $G_k(t_i) = G(t_i)$ for $i = 1,\ldots,k$ and F_k and G_k linear on $[t_i,t_{i+1}]$ for $i = 1,\ldots,k-1$. Then

$\Delta(F_k,G_k) = \int F_k dG_k = \frac{1}{2}\sum_{i=1}^{k-1} [F(t_i) + F(t_{i+1})][G(t_{i+1}) - G(t_i)]$. For squared error loss function, the Bayes estimate $\hat{\Delta}_k$ of $\Delta(F_k,G_k)$ is

$$\hat{\Delta}_k = E_\pi(\Delta(F_k,G_k)\,|\,X_1,\ldots,X_m,Y_1,\ldots,Y_n),$$

where π denotes that F(t) is a Dirichlet process with updated parameter $\alpha + \sum_{i=1}^{m} \delta_{X_i}$ and G(t) is a Dirichlet process with updated parameter $\beta + \sum_{j=1}^{n} \delta_{Y_j}$. Let $p_i = F(t_{i+1}) - F(t_i)$ and $p_i' = G(t_{i+1}) - G(t_i)$ for $i = 1,\ldots,k-1$. By Theorem 2.5 and Definition 2.4, $p = (p_1,\ldots,p_{k-1})$ has a Dirichlet distribution with parameters $\{\alpha(t_i,t_{i+1}] + M_i\}_{i=1}^{k-1}$ and $p' = (p_1',\ldots,p_{k-1}')$ has a Dirichlet distribution with parameters

$\{\beta(t_i, t_{i+1}] + N_i\}_{i=1}^{k-1}$, where M_i and N_i denote the number of X's and Y's, respectively, which fall into $(t_i, t_{i+1}]$ for $i = 1,\ldots,k-1$. It is easy to see by independence of the processes, therefore, that $\hat{\Delta}_k$ is given by the right-hand-side of (1.1). The estimator $\hat{\Delta}_k$ may be rewritten as

$$\begin{aligned}\hat{\Delta}_k = {} & \alpha_m \beta_n \sum_{i=1}^{k-1} \frac{\alpha(t_1,t_i] + \tfrac{1}{2}\alpha(t_i,t_{i+1}]}{\alpha(\mathcal{R})} \cdot \frac{\beta(t_i,t_{i+1}]}{\beta(\mathcal{R})} \\ & + (1-\alpha_m)\beta_n \sum_{i=1}^{k-1} \frac{M_1+\ldots+M_i + \tfrac{1}{2}M_{i+1}}{m} \cdot \frac{\beta(t_i,t_{i+1}]}{\beta(\mathcal{R})} \\ & + \alpha_m(1-\beta_n) \sum_{i=1}^{k-1} \frac{\alpha(t_1,t_i] + \tfrac{1}{2}\alpha(t_i,t_{i+1}]}{\alpha(\mathcal{R})} \cdot \frac{N_i}{n} \\ & + (1-\alpha_m)(1-\beta_n) \sum_{i=1}^{k-1} \frac{M_1+\ldots+M_i + \tfrac{1}{2}M_{i+1}}{m} \cdot \frac{N_i}{n},\end{aligned} \tag{4.1}$$

where $\alpha_m = \alpha(\mathcal{R})/(\alpha(\mathcal{R}) + m)$ and $\beta_n = \beta(\mathcal{R})/(\beta(\mathcal{R}) + n)$. Note that this estimator with the squared error loss function is both a mixed rule (by Theorem 3.7) and a $\mathcal{G}_k$-minimax rule (by Corollary 3.9) for $\Omega = \{(F,G):$ p and p' are independent Dirichlet distributions with parameters $(\alpha(t_1,t_2], \ldots,\alpha(t_{k-1},t_k])$ and $(\beta(t_1,t_2],\ldots,\beta(t_{k-1},t_k])$, respectively$\}$.

As the t_i's become dense, $\hat{\Delta}_k$ is seen to approach Ferguson's [4] estimator for $\Pr(X \leq Y)$ for complete Dirichlet prior information. As $\alpha(\mathcal{R})$ and $\beta(\mathcal{R}) \to 0$, $\hat{\Delta}_k$ approaches the Mann-Whitney U' statistic for grouped data (as given in Putter [5]):

$$U' = \sum_{i=1}^{k-1} \frac{M_1+\ldots+M_i + \tfrac{1}{2}M_{i+1}}{m} \cdot \frac{N_i}{n} .$$

As $\alpha(\mathcal{R})$ and $\beta(\mathcal{R})$ get large,

$$\hat{\Delta}_k \sim \sum_{i=1}^{k-1} \frac{\alpha(t_1,t_i] + \tfrac{1}{2}\alpha(t_i,t_{i+1}]}{\alpha(\mathcal{R})} \cdot \frac{\beta(t_i,t_{i+1}]}{\beta(\mathcal{R})} .$$

The estimator $\hat{\Delta}_k$ would be useful, for example, in the following situation. Suppose there are two middle-sized towns for which one wishes to compare the cholesterol rates, in particular to estimate $\Pr(X \leq Y)$ where X is the cholesterol level of a randomly selected person from town A and Y is the cholesterol level of a randomly selected person in town B. Town B could be undergoing a program designed to lower cholesterol rates with Town A serving as a control. There is prior knowledge about the cholesterol levels in the two towns. The prior knowledge is qualified by specifying the weights $\alpha(t_i,t_{i+1}]$ and $\beta(t_i,t_{i+1}]$ for $i = 1,\ldots,k-1$. The values $\alpha(\mathbb{R})$ and $\beta(\mathbb{R})$ reflect the degrees of confidence held in these weights. The estimator $\hat{\Delta}_k$ is then a combination of the priors and the actual data tabulated by intervals.

5. Rank Order Estimation Under Partial Prior Information

Let $X_1,\ldots,X_n$ be a sample of size n from the distribution F. Assuming F is a random distribution function chosen according to the Dirichlet process prior with parameter α, Campbell and Hollander [2] derive the Bayes estimator of the rank order G of X_1 among $X_1,\ldots,X_n$ based on knowledge of $r(<n)$ observed values $X_1,\ldots,X_r$. In this Dirichlet model, care must be taken in the definition of a rank order since the distribution chosen by a Dirichlet process is discrete with probability one (cf. Berk and Savage [1]). To resolve the issue of ties with regard to the rank order, average ranks are used.

Definition 5.1. Let K, L, and M denote the number of observations of $X_1, X_2,\ldots,X_n$ that are less than, equal to, and greater than X_1, respectively. Then the rank order G of X_1 among $X_1, X_2,\ldots,X_n$ is the average value of the ranks that would be assigned to the L values tied at X_1, in a joint ranking from least to greatest, if those values could be distinguished; namely,

$$G = \{(K+1) + (K+2)+\ldots+(K+L)\}/L = K + (L+1)/2.$$

Similarly, for K', L', and M' defined, respectively, to be the number of observations of $X_1, X_2, \ldots, X_r$ less than, equal to, and greater than X_1, the rank order G' of X_1 among $X_1, X_2, \ldots, X_r$ is given by $G' = K' + (L'+1)/2$.

For squared error loss, the Bayes estimator is (see equation (1.2) of [2])

$$\hat{G} = G' + (n-r)\{\alpha'(-\infty, X_1) + \tfrac{1}{2}\alpha'(\{X_1\})\}/\alpha'(\mathcal{R}), \tag{5.1}$$

where $\mathcal{R}$ is the real line and $\alpha' = \alpha + \sum_{i=1}^{r} \delta_{X_i}$, where δ_z is that measure which concentrates its entire mass of one at the point z.

In this section it is assumed that α is not completely known; instead α is specified only on k-1 intervals $(t_i, t_{i+1}]$ for $i = 1,\ldots,k-1$ with $\alpha(\mathcal{R}) = \sum_{i=1}^{k-1} \alpha(t_i, t_{i+1}]$. Let F_k denote the polygonal random distribution function from the Dirichlet process. What is the Bayes estimate for the true rank order g if F is known and $X_1, \ldots, X_r$ have been observed? It is easy to appeal to equation (3.3) of [2] for $\Pr\{(K,L,M) = (k,\ell,m) | X_1, \ldots, X_r, F\}$. The mean G_F of G, given $X_1, \ldots, X_r$ and F, is obtained from the mean of a multinomial. We find

$$G_F = G' + (n-r)\{F(X_1^-) + \tfrac{1}{2}[F(X_1) - F(X_1^-)]\}.$$

Restricting just to polygonal distribution functions, it is clear that G_F using squared error loss function depends on F_k not just at $F_k(t_i)$, $i = 1,\ldots,k$. This makes finding a mixed rule for the rank order problem most difficult.

Suppose the observations have simply been grouped into intervals where the values α assigns to these intervals are known. Rather than take the loss function $L(g,d) = (g-d)^2$, we use the following modified loss function. For $g(F,X_1,\ldots,X_r) = g' + (n-r)F(t_i) + \frac{1}{2}(n-r)[F(t_{i+1}) - F(t_i)]$ if $X_1 \in (t_i,t_{i+1}]$, the loss is given by $[g(F,X_1,\ldots,X_r)-d]^2$. The mixed Bayes minimax rule is then easily shown to be

$$\hat{G} = G' + (n-r)[\alpha'(t_1,t_i] + \tfrac{1}{2}\alpha'(t_i,t_{i+1}]]/\alpha'(\mathcal{R}) \tag{5.2}$$

if $X_1 \in (t_i,t_{i+1}]$ for $i = 1,\ldots,k-1$. Note that this rule is really just the Dirichlet estimator with complete information concerning the parameter α, but where α is concentrated at (k-1) atoms $\{t_i\}_{i=2}^{k}$ so that $\sum_{i=2}^{k} \alpha(\{t_i\}) = \alpha(\mathcal{R})$.

An example in which such an estimator could be of use is as follows. An automobile driver is passing through a town in need of regular gas. The driver knows there are n stations in town and all n clearly post their prices for gas. From past experience at the gas pump, the driver has some idea of the distribution of prices in the region. The model tends to be contagious in that if one station advertises a particular price, competition (or lack of it) will cause others to be more likely to adopt that price also. Hence the Dirichlet model is not unreasonable here. The problem is for the driver to estimate, as he passes the $r^{\underline{th}}$ station, the rank of that station's gas price among all n stations, on the basis of the prices at the first r stations and his prior information. Then, the estimator $\hat{G}$ could be used, with the parameter $\alpha(\mathcal{R})$ reflecting the weight or confidence attached to the driver's prior knowledge of regional gasoline prices.

REFERENCES

[1] Berk, R. H. and Savage, I. R. (1977). Dirichlet processes produce discrete measures: An elementary proof. (To appear in Contributions to Statistics - Jaroslav Hájek Memorial Volume, Academia, Prague; North Holland Publishing Co.)

[2] Campbell, G. and Hollander, M. (1978). Rank order estimation with the Dirichlet prior. Ann. Statist. 6 (to appear).

[3] Doksum, K. A. (1972). Decision theory for some nonparametric models. Proc. Sixth Berkeley Symp. Math. Statist. Prob. 1 331-343.

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

[5] Putter, J. (1955). The treatment of ties in some nonparametric tests. Ann. Math. Statist. 26 368-386.

[6] Wilks, S. S. (1962). Mathematical Statistics. Wiley, New York.