The Annals of Mathematical Statistics
1972, Vol. 43, No. 6, 1993–1999

BAYES ESTIMATION OF THE MIXING DISTRIBUTION, THE DISCRETE CASE

BY GLEN MEEDEN

Iowa State University

Let $X_1, X_2, \cdots$ be independent identically distributed random variables taking on values in the positive integers with a family of possible probability distributions indexed by $G \in \mathscr{G}$, the class of all probability distribution functions on $[0, +\infty)$. Under the assumption that the family is identifiable we wish to estimate the true but unknown G_0. This is done by constructing a prior probability distribution on $\mathscr{G}$ and showing that the Bayes estimate corresponding to the prior is consistent.

1. Introduction and summary. Let X be a random variable taking values in the positive integers with $\mathscr{P} = \{P_t : t \in [0, +\infty)\}$, a family of possible probability distributions. t is a realization of a random variable taking on values in $[0, +\infty)$ and

$$P_t(X = x) = q_x(t) \qquad \text{for} \quad x = 1, 2, \cdots$$

is the conditional distribution of X given t. If G, a probability distribution function on $[0, +\infty)$, is the distribution of t then the unconditional distribution of X is a G-mixture over $\mathscr{P}$ and

$$P_G(X = x) = \textstyle\int_0^\infty q_x(t)\, dG(t) = q_x(G) .$$

We assume that the family $\mathscr{P}$ is identifiable, that is, if $q_x(G_1) = q_x(G_2)$ for $x = 1, 2, \cdots$ then $G_1 = G_2$. In this paper we define a prior distribution on $\mathscr{G}$, the class of all probability distribution functions on $[0, +\infty)$ and show the posterior distribution, based on independent observations of X, is consistent.

This was done in Rolph [3] for the case where the parameter set is $[0, 1]$ and $q_x(\cdot)$ is a continuous function. Our approach is to define the prior in a way suggested by Dubins and Freedman [1] and Kraft and Van Eeden [2] and then use Theorem 2 of Rolph [3] to prove consistency of the posterior when $q_x(\cdot)$ is continuous on $[0, +\infty)$, of bounded variation on any finite sub-interval and $\lim_{t\to\infty} q_x(t) = 0$. Finally, we use the posterior distribution to construct consistent Bayes estimates for the mixing distribution G.

2. Defining a prior distribution. We begin by defining a prior probability distribution on $\mathscr{G}$, the set of all probability distribution functions on $[0, +\infty)$. Let $\{r_i : i = 1, 2, \cdots\}$ be a dense set in $[0, +\infty)$. Since a distribution function is uniquely determined by its values on a dense set there is a one to one correspondence between $\mathscr{G}$ and $D = (\{G(r_i)\} : G \in \mathscr{G})$, a subset of the infinite dimensional unit cube $[0, 1]^\infty$. Since given the values $G(r_1), \cdots, G(r_N)$ there exist

Received June 17, 1971; revised February 9, 1972.

AMS 1970 *subject classifications.* Primary 62 C10; Secondary 62 E20 and 62 G05.

Key words and phrases. Weak* convergence, consistency, Bayes estimation, mixing distribution.

unique numbers $0 \leqq \underline{G}_N \leqq \bar{G}_N \leqq 1$ such that $G(r_{N+1}) \in [\underline{G}_N, \bar{G}_N]$, a distribution V will be defined on D via the coordinates and then mapped to $\mathscr{G}$. Let $h_1, h_2, \cdots$ be a sequence of everywhere positive densities with respect to Lebesgue measure on $[0, 1]$. On the first coordinate V_1, the density with respect to Lebesgue measure, is

$$(1) \qquad \begin{aligned} V_1(G(r)) &= h_1(G(r)) && \text{if } 0 \leqq G(r) \leqq 1 \\ &= 0 && \text{elsewhere} \end{aligned}$$

and the density V_k on the kth given $(G(r_1), \cdots, G(r_{k-1}))$ is

$$(2) \qquad \begin{aligned} &V_k(G(r) \mid G(r_1), \cdots, G(r_{k-1})) \\ &\quad = h_k(G(r))/\textstyle\int_{\underline{G}_{k-1}}^{\bar{G}_{k-1}} h_k(u)\,du && \text{if } G(r) \in [\underline{G}_{k-1}, \bar{G}_{k-1}] \\ &\quad = 0 && \text{elsewhere}\,. \end{aligned}$$

$\{G(r_i)\} \to (G(r_1), \cdots, G(r_N))$ maps $\mathscr{G}$ onto D^N, the projection of D onto the first N coordinates. Let $\mathscr{B}_D{}^N$ be the Borel sigma field on D^N in the product topology. We derive the posterior distribution on $(D^N, \mathscr{B}_D{}^N)$ and appeal to the Kolmogorov extension theorem to define it on $(D, \mathscr{B}_D)$ with $\mathscr{B}_D$ being the Borel sigma field on D.

Now put the weak topology on D. Since the weak topology is contained in the product topology $\mathscr{B}_D{}^*$, the Borel sigma field on D generated by the weak open sets, is contained in $\mathscr{B}_D$. Hence V induces a measure μ on $(\mathscr{G}, \mathscr{A}^*)$ where $\mathscr{A}^*$ is the sigma algebra generated by the weak open sets.

Since we are interested in only proper distributions on $[0, +\infty)$ we desire that the probability measure V gives probability one to the G's with $G(+\infty) = 1$. To see that this need not be the case consider the following example.

For the sequence $\{r_i\}$ we say that $r_{i'}$ is an upper point if $r_i \leqq r_{i'}$ for $i \leqq i'$. Let $\{r_{i_j}\}$ be the strictly increasing subsequence of upper points of $\{r_i\}$. Note that the behavior of G at infinity depends only on the values $\{G(r_{i_j})\}$. Let $\{b_j : j = 0, 1, \cdots\}$ be a strictly increasing sequence of positive numbers bounded above by $\frac{1}{2}$. Let $h_{i_1}, h_{i_2}, \cdots$ be chosen so that

$$\int_{b_{j-1}}^{b_j} h_{i_j}(u)\,du \geqq 1 - (\tfrac{1}{2})^j \qquad \text{for } j = 1, 2, \cdots,$$

then we have

$$\begin{aligned} P(G(r_{i_n}) \leqq \tfrac{1}{2}) &\geqq P(b_{j-1} \leqq G(r_{i_j}) \leqq b_j) \qquad \text{for } j = 1, \cdots, n \\ &\geqq \textstyle\prod_{j=1}^{n} (1 - (\frac{1}{2})^j)\,. \end{aligned}$$

Since $\sum_{j=1}^{\infty} (\frac{1}{2})^j < \infty$ we have that $\prod_{j=1}^{\infty} (1 - (\frac{1}{2})^j) > 0$ and so with positive probability $G(+\infty) \leqq \frac{1}{2}$.

In the rest of the paper we assume that $\{h_i\}$ is chosen so that the distribution V assigns probability one to the class of proper distributions. It is easy to give various conditions which imply that the previous assumption is true. One such condition which would usually hold in any problem for which one wished to compute the estimate is that $\{h_i\}$ contains only finitely many different densities.

By different choices of $\{r_i\}$ and $\{h_i\}$ this procedure yields a large class of priors. In particular, one may wish the prior to reflect his state of knowledge about the true mixing distribution, G_0. This can be done quite conveniently if one has information about the quantiles of G_0. For example, if it is known that the median of G_0 is about 17 one can choose $r_1 = 17$ and for h_1 a density concentrated about $\frac{1}{2}$. One may continue in this fashion. But the calculations necessary to compute the posterior distribution and the estimates will be simplified if one takes h_i to be the uniform density for $i \geqq i_0$.

Let $X_1, \cdots, X_n$ be independent random variables defined on a probability space $(\Omega, \mathscr{F})$ with $P_G(X_i = x) = q_x(G)$ for $i = 1, \cdots, n$ and $x = 1, 2, \cdots$. Given $X_1(\omega), \cdots, X_n(\omega)$ let $n_j(\omega)$ be the number of $X_i(\omega) = j$. The sample can be written $(n_1, n_2, \cdots, n_c, 0, \cdots)$ where c is the largest of the n observations.

The joint frequency function of $(n_1, n_2, \cdots)$ given n and G is

$$f(n_1, \cdots, n_c, \cdots \mid n, G) = \binom{n}{n_1 \cdots n_c} \prod_{x=1}^{c} (q_x(G))^{n_x} .$$

The posterior distribution of G given $(n_1, \cdots, n_c, \cdots)$ is

$$d\mu_{n,\omega}(G) = \prod_{x=1}^{c} (q_x(G))^{n_x} \, d\mu(G)/I(n_1, \cdots, n_c, \cdots) \tag{3}$$

where

$$I(n_1, \cdots, n_c, \cdots) = \int_{\mathscr{G}} \prod_{x=1}^{c} (q_x(G))^{n_x} \, d\mu(G) .$$

We wish to prove the consistency of the posterior distribution, that is, $\mu_{n,\omega}$ converges to the probability measure which concentrates its mass at the true mixing distribution G_0.

3. Consistency. Before proving consistency we need some preliminaries. Following [3] we let S be the space functions from I, the positive integers, to $[0, 1]$ with the product topology. Let $L = \{\lambda : \lambda \in S, \Sigma\lambda_i \leqq 1\}$ with the relative topology and $\Lambda = \{\lambda : \lambda \in L, \Sigma\lambda_i = 1\}$ with μ a probability measure on $\mathscr{B}$, the Borel sigma field of L. Let $X_1, X_2, \cdots$ be a sequence of independent, I valued, random variables on $(\Omega, \mathscr{F})$ with common distribution $P_\lambda\{\omega : \omega \in \Omega, X_n(\omega) = i\} = \lambda_i$, $i \in I$. Let $\mu_{n,\omega}$ denote the posterior distribution λ given $X_1(\omega), \cdots, X_n(\omega)$.

We use the weak* topology on the space of probability measures on $\mathscr{B}$ so that $\mu_n \to \mu$ means that $\int_L f \, d\mu_n \to \int_L f \, d\mu$ for every continuous function f on L. Let δ_λ be a point mass at λ. We say that the pair (λ, μ) is consistent if and only if $\mu_{n,\omega} \to \delta_\lambda$ for P_λ-almost all ω. That is, the $\mu_{n,\omega}$ measure of every L neighborhood of λ converges to 1 for all but a P_λ-null set of ω. Theorem 1 is Theorem 2 of [3].

THEOREM 1. *If μ is a probability on $\mathscr{B}$ such that $\mu\{\lambda : \lambda \in L, \sum_{i=0}^{m} p_i \log (p_i/\lambda_i) < \mathscr{E}$ where $p_0 = \sum_{i=m+1}^{\infty} p_i$ and $\lambda_0 = \sum_{i=m+1}^{\infty} \lambda_i\} > 0$ for all m, then (p, μ) is consistent.*

The following Lemma will be used in the proof.

LEMMA. *Let f be a continuous function on $[0, +\infty)$ with $0 \leqq f \leqq 1$ and suppose that f is of bounded variation on every finite interval. Let G be a probability function on $[0, +\infty)$. Given $\mathscr{E} > 0$ there exist numbers $0 \leqq s_1 < \cdots < s_N$ and $\delta > 0$ such* *that for each probability distribution function H on $[0, +\infty)$ with* $\max_{i=1,\cdots,N} |G(s_i) - H(s_i)| < \delta$ *it follows that* $|\int_0^\infty f\,dG - \int_0^\infty f\,dH| < \mathscr{E}$.

PROOF. First suppose G is continuous. Given $\mathscr{E} > 0$ choose $0 \leqq s_1 < \cdots < s_N$ such that $1 - G(s_N) < \mathscr{E}/4$ and $\max_{i=1,\cdots,N-1} |G(s_{i+1}) - G(s_i)| < \mathscr{E}/4c$ where c is the total variation of f on $[0, s_N]$. Without loss of generality we assume $s_1 = 0$. For a given distribution function H let $\delta = \max_{i=1,\cdots,N} |G(s_i) - H(s_i)|$ then

$$\begin{aligned}|\textstyle\int_0^\infty f\,dG - \int_0^\infty f\,dH| &\leqq \mathscr{E}/4 + \mathscr{E}/4 + \delta + \textstyle\int_0^{s_N} f\,d(G-H)\\ &\leqq \mathscr{E}/2 + \delta + 2\delta + [\mathscr{E}/4c + 2\delta]c\\ &\leqq \mathscr{E}\end{aligned}$$

for δ sufficiently small, where the second inequality follows by integration by parts.

For the general case choose s such that $1 - G(s) < \mathscr{E}/10$ and let $t_1, \cdots, t_k$ denote the points at which G jumps at least $\mathscr{E}' = \mathscr{E}/\{10(\text{total variation of } f \text{ on } [0, s])\}$. Next we choose $0 = s_1 < \cdots < s_N = s$ to satisfy the following:

(i) The total variation of f on the k sub-intervals containing $t_1, \cdots, t_k$ is less than or equal to $\mathscr{E}/10$.

(ii) If $t_j \notin (s_i, s_{i+1}]$ for $j = 1, \cdots, k$ then $G(s_{i+1}) - G(s_i) \leqq \mathscr{E}'$.

The result follows by noting that $\int_0^{s_N} f\,d(G - H)$ can be separated into two integrals. The first integral is over the k intervals where G jumps at least $\mathscr{E}'$, which is small by (i). The integral over the remaining intervals is small by (ii).

We are now ready to prove the consistency for the measure μ which was defined on $\mathscr{G}$ in Section 2. Before stating the Theorem we briefly outline the proof. For each $G \in \mathscr{G}$ $q_x(G) = \int_0^\infty q_x(t)\,dG(t)$ is a probability measure on I, say $q(G)$. The map $Q: G \to q(G)$ maps $\mathscr{G}$ into a subset of Λ, say W. The measure μ induces through the map Q a measure on W. It will be shown that the consistency of the induced measure follows from Theorem 1 and the Lemma. Then the consistency of μ will be demonstrated.

Consistency on $(\mathscr{G}, \mathscr{A}^*)$ is defined analogously to consistency on $(L, \mathscr{B})$. The pair (G, μ) is consistent if the $\mu_{n,\omega}$ measure of every $\mathscr{G}$ neighborhood of G converges to 1 for all but a $P_{Q(G_0)}$-null set of ω.

THEOREM 2. *If the family $\mathscr{P}$ is identifiable and if for each x $q_x(t)$ is continuous on $[0, +\infty)$, of bounded variation on $[0, \gamma]$ for every $\gamma > 0$ and $\lim_{t\to+\infty} q_x(t) = 0$ then (G_0, μ) is consistent.*

PROOF. The map Q of $\mathscr{G}$ into L is continuous since each $q_x(G)$ is a continuous function. Let $W =$ Range of Q with the relative topology $\mathscr{B}_W$. Since the family $\mathscr{P}$ is identifiable, Q is one to one, so $Q^{-1}: W \to \mathscr{G}$ exists. Let μQ^{-1} denote the measure induced on W by Q. We will show using Theorem 1 that $(Q(G_0), \mu Q^{-1})$ is consistent.

For an arbitrary fixed m, we let

$$U^m = \{G : G \in \mathscr{G}, \sum_{x=0}^{m} q_x(G_0) \log (q_x(G_0)/q_x(G)) < \mathscr{E}\}$$

where $q_0(G) = \sum_{i=m+1}^{\infty} q_x(G)$. Since $q_x(G)$ sufficiently close to $q_x(G_0)$ for $x = 1, \cdots, m$ implies that $G \in U^m$ we have by the Lemma that there exist $r_{i_1}, \cdots, r_{i_k}$ such that

$$\{G : G \in \mathscr{G} \text{ and } \max_{j=1,\cdots,k} |G(r_{i_j}) - G_0(r_{i_j})| < \delta\} \subset U^m$$

for δ sufficiently small. Set $V^m = Q(U^m)$, then $\mu Q^{-1}(V^m) = \mu(U^m) > 0$ for each m and $(Q(G_0), \mu Q^{-1})$ is consistent by Theorem 1.

The next step is to show that Q^{-1} is a continuous function. It is enough to show that if $\{G_n\}$ and G satisfy

$$\int_0^\infty q_x(t)\, dG_n(t) \to \int_0^\infty q_x(t)\, dG(t) \qquad \text{for } x = 1, 2, \cdots$$

then $\{G_n\}$ converges weakly to G. Let $\{G_{n_j}\}$ be a subsequence which converges weakly to a right continuous non-decreasing function G' with $0 = G'(0-) \leqq G'(+\infty) \leqq 1$. Since for each x, $\lim_{t\to+\infty} q_x(t) = 0$ we have that

$$\int_0^\infty q_x(t)\, dG_{n_j}(t) \to \int_0^\infty q_x(t)\, dG'(t)$$

which implies that

$$\int_0^\infty q_x(t)\, dG(t) = \int_0^\infty q_x(t)\, dG'(t)$$

for $x = 1, 2, \cdots$. By interchanging the summation and the integral we have that $1 = \sum_{x=1}^{\infty} \int_0^\infty q_x(t)\, dG'(t) = G'(+\infty) - G'(0-)$. Therefore $G'(+\infty) = 1$ and by the identifiability assumption $G = G'$ and $\{G_{n_j}\}$ converges weakly to G. Since every weakly convergent subsequence converges weakly to G it follows that $\{G_n\}$ converges weakly to G.

Finally we show the consistency of (G_0, μ). Let U be a neighborhood of G_0 in $(\mathscr{G}, \mathscr{A}^*)$. By the continuity of Q^{-1} there exists a neighborhood V of $Q(G_0)$ in $(W, \mathscr{B}_W)$ so that $Q^{-1}(V) \subset U$.

$$\mu_{n,\omega}(U) \geqq \mu_{n,\omega}(Q^{-1}(V)) = \mu Q_{n,\omega}^{-1}(V) \to 1$$

as $n \to \infty$ for a.e. $[P_{Q(G_0)}]\omega$. But $P_{Q(G_0)}$ is the distribution on $(\Omega, \mathscr{F})$ corresponding to G_0 which completes the proof.

4. Bayes estimates. For estimating the mixing distribution let $L(G, G')$ be the loss incurred when G' is the true parameter value and G is estimated. The Bayes estimate, based on $X = (X_1, \cdots, X_n)$, is the estimate which minimizes the Bayes risk

$$R(\hat{G}(X)) = \int_{\mathscr{G}} [\sum_X L(\hat{G}(X), G') q_X(G')\, \mathrm{d}\mu(G') \tag{4}$$

where μ is the prior distribution on $\mathscr{G}$. A convenient and perhaps not unreasonable loss function is

$$L(\hat{G}, G') = \sum_{i=1}^{\infty} \lambda_i [q_i(\hat{G}) - q_i(G')]^2$$

where $\lambda_i > 0$ and $\Sigma\lambda_i$ is finite. By interchanging the integral and summation in (4) it is easily seen that for $X = x$ the Bayes estimate $\hat{G}_n$ is a distribution function with $q_j(\hat{G}_n) = \Phi(j)$ for $j = 1, 2, \cdots$ where

$$\Phi(j) = \int_{\mathscr{G}} q_j(G')\, d\mu_{n,\omega}(G') \tag{5}$$

is the average probability of j under the posterior distribution. Since $\{\Phi(j)\} \in W$, the range of Q, $\hat{G}_n$ exists and is unique by the identifiability assumption.

Let G_0 denote the true but unknown mixing distribution and μ_0 the probability measure which concentrates all of its mass at G_0. Since a.e. $[P_{G_0}]\omega$ $\mu_{n,\omega} \to \mu_0$ and $q_j(G)$ is a continuous function of G we have that $q_j(\hat{G}_n) \to q_j(G_0)$ for all j and as in the proof of Theorem 2 we have that $\hat{G}_n$ converges weakly to G_0 and our estimates are consistent.

It is not possible to calculate the preceding estimate since it is necessary to compute an infinite number of integrals each of which is over an infinite dimensional space. It is, however, possible to construct a sequence of approximate Bayes estimates which are consistent and computable.

As a first step consider the posterior density $d\mu_{n,\omega}$ given in (3), which is a density on an infinite dimensional space. This density will be approximated by considering just its first N coordinates. Let $\mathscr{G}^N$ denote the class of probability distribution function which concentrate their mass on $r_1, \cdots, r_N$. The transformation $G \to (G(r_1), \cdots, G(r_N))$ maps $\mathscr{G}^N$ onto a subset of D^N, say $D_\alpha{}^N$. Letting $\lambda_i = G(r_i)$, $D_\alpha{}^N = \{(\lambda_1, \cdots, \lambda_N)\colon 0 \leqq \lambda_{i_1} \leqq \cdots \leqq \lambda_{i_{N-1}} \leqq \lambda_{i_N} = 1$ for $i = 1, \cdots, N$ where $r_{i_1} < r_{i_2} \cdots < r_{i_N}\}$ then as in (1) and (2) $h_1, \cdots, h_N$ define a probability density $\gamma^N(\lambda_1, \cdots, \lambda_N)$ on $D_\alpha{}^N$. If we consider the parameter G to be restricted to the class $\mathscr{G}^N$ then, given n and $(n_1, \cdots, n_c, \cdots)$, the posterior density of $(\lambda_1, \cdots, \lambda_N)$ on $D_\alpha{}^N$ is

$$\begin{aligned} &g^N(\lambda_1, \cdots, \lambda_N \mid n_1, \cdots, n_c, \cdots) \\ &\quad = \textstyle\prod_{x=1}^{c} \left(\sum_{j=1}^{N} (\lambda_{i_j} - \lambda_{i_{j-1}}) q_x(r_{i_j})\right)^{n_x} \gamma^N(\lambda_1, \cdots, \lambda_N) / I^N(n_1, \cdots, n_c, \cdots) \end{aligned}$$

where

$$\begin{aligned} &I^N(n_1, \cdots, n_c, \cdots) \\ &\quad = \textstyle\int_{D_\alpha{}^N} \prod_{x=1}^{c} \left(\sum_{j=1}^{N} (\lambda_{i_j} - \lambda_{i_{j-1}}) q_x(r_{i_j})\right)^{n_x} \gamma^N(\lambda_1, \cdots, \lambda_N)\, d\lambda_1, \cdots, d\lambda_N\,. \end{aligned}$$

To show consistency of the approximate estimates that will be proposed it is necessary to permit N to depend on n. Let $\{N(n)\}$ be a sequence of positive integers. The form of our Bayes estimates in (5) suggests that in constructing our approximate Bayes estimates we use

$$\begin{aligned} \Phi^{N(n)}(x) = &\textstyle\int_{D_\alpha{}^{N(n)}} \left(\sum_{i=1}^{N(n)} (\lambda_{i_j} - \lambda_{i_{j-1}}) q_x(r_{i_j})\right) \\ &\times g^{N(n)}(\lambda_1, \cdots, \lambda_{N(n)} \mid n_1, \cdots, n_c, \cdots)\, d\lambda_1, \cdots, d\lambda_{N(n)} \end{aligned}$$

for $x = 1, 2, \cdots$. To accomplish this we choose a positive integer $k(n)$ and a distribution $G \in \mathscr{G}^{N(n)}$ such that

$$q_j(G) = \Phi^{N(n)}(j) \qquad \text{for } j = 1, 2, \cdots, k(n)\,.$$

This is a system of $k(n)$ liner equations in the unknowns $\lambda_1, \cdots, \lambda_{N(n)}$ which has at least one solution in $D_\alpha^{N(n)}$. Denote one such solution by $\hat{G}_n^{N(n)}$.

We will show that the sequence $\{\hat{G}_n^{N(n)}\}$ of approximate Bayes estimates is consistent provided that $N(n) \to \infty$ sufficiently fast and $\lim_{n\to\infty} k(n) = \infty$. The only additional assumption necessary to ensure consistency is that $q_x(t)$ is positive on $[0, \infty)$ for all x.

For $\{r_1, \cdots, r_{N(n)}\}$ let $r_{i_1} < r_{i_2} < \cdots < r_{i_{N(n)}}$ be the ordering of the r_i's. For $G \in \mathscr{G}$ let

$$F_x^{N(n)}(G) = G(r_{i_1})q_x(r_{i_1}) + \sum_{j=2}^{N(n)-1} [G(r_{i_j}) - G(r_{i_{j-1}})]q_x(r_{i_j}) + [1 - G(r_{i_{N(n)-1}})]q_x(r_{i_{N(n)}}) .$$

Then

$$q_{\hat{G}_n^{N(n)}}(j) = \int_{\mathscr{G}} F_j^{N(n)}(G)\, d\mu_{n,\omega}(G) \tag{6}$$

for $j = 1, \cdots, k(n)$. For $\mathscr{E} > 0$ there exists a $N_{x,\mathscr{E}}$ such that $|F_x^{N(n)}(G)/q_x(G) - 1| < \mathscr{E}$ for all $G \in \mathscr{G}$ whenever $N(n) \geqq N_{x,\mathscr{E}}$. So for N($n$) sufficiently large we have by (5) and (6) that

$$(1 - \mathscr{E})^{n+1}/(1 + \mathscr{E})^n < q_j(\hat{G}_n^{N(n)})/q_j(\hat{G}_n) < (1 + \mathscr{E})^{n+1}/(1 - \mathscr{E})^n$$

for $j = 1, \cdots, k(n)$. If $\mathscr{E}(n)$ is chosen so that $[(1 + \mathscr{E}(n))/(1 - \mathscr{E}(n))]^n$ approaches one as $n \to +\infty$ then $q_j(\hat{G}_n^{N(n)}) \to q_j(\hat{G}_n)$ for all j and the approximate Bayes estimates are consistent.

REMARK. These results can be extended to the case with $t \in (-\infty, +\infty)$ if for each x $q_x(t)$ is continuous, of bounded variation on any finite interval, and $\lim_{t\to+\infty} q_x(t) = \lim_{t\to-\infty} q_x(t) = 0$.

Acknowledgment. I wish to thank Barry Arnold, Dean Isaacson and the referee for some helpful comments.

REFERENCES

[1] DUBINS, L. E. and FREEDMAN, DAVID A. (1963). Random distribution functions. *Bull. Amer. Math. Soc.* 69 548-551.

[2] KRAFT, CHARLES H. and VAN EEDEN, Constance (1964). Bayesian bio-assay. *Ann. Math. Statist.* **35** 886-890.

[3] ROLPH, JOHN E. (1968). Bayesian estimation of mixing distributions. *Ann. Math. Statist.* **39** 1289-1302.

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 50010