The Annals of Statistics
1981, Vol. 9, No. 1, 221–224

A NOTE ON THE ASYMPTOTIC OPTIMALITY OF THE EMPIRICAL BAYES DISTRIBUTION FUNCTION

BY BENJAMIN ZEHNWIRTH

Macquarie University, N.S.W., Australia

This paper establishes the asymptotic optimality (in the sense of Robbins) of the empirical Bayes distribution function created from the Bayes rule relative to the Dirichlet process prior with unknown parameter $\alpha(.)$. It will follow that the same result applies to the estimation of the mean of a distribution function.

0. Summary and introduction. Korwar and Hollander [3] demonstrate that the empirical Bayes distribution function satisfies a criterion that is intimately related to Robbins' [4] concept of asymptotic optimality of empirical Bayes procedures. Their analysis presumes, inter alia, that the parameter $\alpha(\mathbb{R})$ of Ferguson's [2] Dirichlet process is known. It is natural to extend the optimality result to the case of unknown $\alpha(\mathbb{R})$. In order to carry out this extension effectively we present a consistent estimator of $\alpha(\mathbb{R})$ that is based on an identity found in Zehnwirth [5] and moreover make extensive use of the results contained in Ferguson [2].

1. Framework and preliminaries. For the necessary preliminaries relating to the Dirichlet process, the reader is referred to Ferguson [2]. The terminology and definitions which are met within the present paper without explanation may be found in his paper. For the formulation of our problem the reader is referred to the paper of Korwar and Hollander [3]. Briefly, the setup is as follows:

A random sample of distribution functions $F_1, F_2, \cdots, F_{n+1}$ is chosen from a prior distribution given by a Dirichlet process on $(\mathbb{R}, B)$ with parameter $\alpha(\cdot)$. Next, random samples $\mathbf{X}_i = (X_{i1}, X_{i2}, \cdots, X_{im})$ are chosen from the distribution functions F_i, $i = 1, 2, \cdots, n+1$. The interest is on estimating F_{n+1} using the squared error loss function, $L(F_{n+1}, \hat{F}) = \int_{\mathbb{R}} (\hat{F}(t) - F_{n+1}(t))^2 \, dW(t)$, where $W(\cdot)$ is a given finite measure on $(\mathbb{R}, B)$ and $\hat{F}$ is an estimate of F_{n+1}. The parameter and action spaces are the set of all distributions on $(\mathbb{R}, B)$. Recall that Ferguson's [2] Bayes estimator of F_{n+1} based on $\mathbf{X}_{n+1}$ is given by

$$\tilde{F}_m(t) = (1 - p_m) F_0(t) + p_m \hat{F}_{n+1}(t), \tag{1.1}$$

where

$$p_m = m/(m + \alpha(\mathbb{R})) \tag{1.2}$$

$$F_0(t) = \alpha((-\infty, t])/\alpha(\mathbb{R}) \tag{1.3}$$

and $\hat{F}_i(t)$ is the sample distribution function of $\mathbf{X}_i$, $i = 1, 2, \cdots, n+1$. The Bayes risk of $\tilde{F}_m(t)$ is given by

$$r(\tilde{F}_m, \alpha) = E\left[\int_{\mathbb{R}} (\tilde{F}_m(t) - F(t))^2 \, dW(t)\right] = \int_{\mathbb{R}} E[(\tilde{F}_m(t) - F(t))^2] \, dW(t), \tag{1.4}$$

where F is a realization of the Dirichlet process, and the last equality follows from Fubini's theorem. The expectation operator E denotes expectation with respect to distributions of the random variables $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_{n+1}, F$. Define for $n = 1, 2, \cdots$ the sequence of estimators H_{n+1} by

$$H_{n+1}(t) = (1 - \hat{p}_{mn}) \hat{F}_{0n}(t) + \hat{p}_{mn} \hat{F}_{n+1}(t) \tag{1.5}$$

Received February 1979; revised October 1979.
AMS 1970 *subject classifications.* Primary 62G05; Secondary 62C99.
Key words and phrases. Empirical Bayes, Dirichlet process, distribution function.

where

$$\hat{F}_{0n}(t) = \textstyle\sum_{i=1}^{n} \hat{F}_i(t)/n, \tag{1.6}$$

$$\hat{p}_{mn} = m/(m + \hat{\alpha}_n(\mathbb{R})) \tag{1.7}$$

and $\hat{\alpha}_n(\mathbb{R})$ is an estimator of $\alpha(\mathbb{R})$ based on the data $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$. The overall Bayes risk of H_{n+1} based on the data $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_{n+1}$ is given by

$$r(H_{n+1}) = E\left[\int_{\mathbb{R}} (H_{n+1}(t) - F(t))^2\, dW(t)\right] = \int_{\mathbb{R}} E[(H_{n+1}(t) - F(t))^2]\, dW(t). \tag{1.8}$$

2. Asymptotic optimality of $\{H_{n+1}\}$. In this section operators with subscripts will represent conditional expectations (variances) given the subscripted variables. Before we present the main result we give a lemma which aids the demonstration of the result.

LEMMA. *Under the set-up of the foregoing section the following identities hold,*

$$E_{F_i}[\hat{F}_i(t)] = F_i(t). \tag{2.1}$$

$$\mathrm{Var}_{F_i}[\hat{F}_i(t)] = F_i(t)(1 - F_i(t))/m. \tag{2.2}$$

$$E[F_i(t)(1 - F_i(t))] = \alpha((-\infty, t])\alpha((t, \infty))/[\alpha(\mathbb{R})(\alpha(\mathbb{R}) + 1)]. \tag{2.3}$$

$$E[\hat{F}_{0n}(t)] = F_0(t). \tag{2.4}$$

Identities (2.1) and (2.2) are well-known, being based on the moment properties of the binomial distribution, whereas identities (2.3) and (2.4) are based on the moment properties of the beta distribution since $F_i(t)$ is distributed as $Be[\alpha((-\infty, t]), \alpha((t, \infty))]$.

THEOREM. *If the sequence $\{\hat{p}_{mn}\}$ is consistent for p_m as n approaches ∞, then the sequence $\{r(H_{n+1})\}$ converges to $r(F_m, \alpha)$ as n approaches ∞.*

PROOF. The proof is split into three parts. Let $f_n(t) = E[(H_{n+1}(t) - F(t))^2]$ and let $f(t) = E[(\tilde{F}_m(t) - F(t))^2]$. By virtue of the bounded convergence theorem it suffices to show that $f_n(t) \to f(t)$ as n approaches ∞, for each t. Now, $f_n(t)$ may be recast in the form,

$$\begin{aligned} f_n(t) &= E[\{F(t) - F_0(t) - \hat{p}_{mn}(\hat{F}_{n+1}(t) - F_0(t))\}^2] \\ &\quad + 2E[\{F(t) - F_0(t) - \hat{p}_{mn}(\hat{F}_{n+1}(t) - F_0(t))\} \\ &\qquad \{F_0(t) - \hat{F}_{0n}(t) + \hat{p}_{mn}(\hat{F}_{0n}(t) - F_0(t))\}] \\ &\quad + E[\{F_0(t) - \hat{F}_{0n}(t) + \hat{p}_{mn}(\hat{F}_{0n}(t) - F_0(t))\}^2]. \end{aligned} \tag{2.5}$$

First, consider the third term on the right of expression (2.5).

$$\begin{aligned} E[\{F_0(t) &- \hat{F}_{0n}(t) + \hat{p}_{mn}(\hat{F}_{0n}(t) - F_0(t))\}^2] \\ &= E[(\hat{p}_{mn} - 1)^2(\hat{F}_{0n}(t) - F_0(t))^2] \\ &\le 2\mathrm{Var}[\hat{F}_{0n}(t)] \\ &= 2\mathrm{Var}[\hat{F}_i(t)]/n \\ &\le 2/n \\ &\to 0 \quad \text{as} \quad n \to \infty. \end{aligned}$$

Next, we show that the first term on the right of expression (2.5) converges to $f(t)$ (for every t). Define

$$g_n(t) = (F(t) - F_0(t) - \hat{p}_{mn}\,(\hat{F}_{n+1}(t) - F_0(t)))^2$$

and

$$g(t) = (F(t) - F_0(t) - p_m\,(\hat{F}_{n+1}(t) - F_0(t)))^2$$

We require $\lim_{n\to\infty} E[g_n(t)] = E[g(t)]$ for every t. Assume the equality is not valid for t_0, say. We can therefore find an $\epsilon > 0$ and a subsequence $\{g_{n'}(t_0)\} \subset \{g_n(t_0)\}$ for which

$$E[g_{n'}(t_0)] \geq E[g(t_0)] + \epsilon \qquad \text{for all } n'. \tag{2.6}$$

But since $g_{n'}(t_0) \to_{\Pr} g(t_0)$ as $n' \to \infty$ there exists a subsequence $\{g_{n''}(t_0)\} \subset \{g_{n'}(t_0)\}$ for which $g_{n''}(t_0) \to g(t_0)$ almost surely. So, by the bounded convergence theorem $\lim_{n''\to\infty} E[g_{n''}(t_0)] = E[g(t_0)]$, which contradicts inequality (2.6). Alternatively, for any t every subsequence of $\{E[g_n(t)]\}$ has a subsequence that converges to $E[g(t)]$. Finally, we note that the second term on the right of expression (2.5) approaches 0 as n approaches ∞, as a consequence of Schwartz inequality and the result obtained concerning the third term. This completes the proof.

We now exhibit a consistent estimator for $\alpha(\mathbb{R})$. Invoke the identity,

$$\alpha(\mathbb{R}) = E[\mathrm{Var}_{F_i}(X_{ij})]/\mathrm{Var}[E_{F_i}(X_{ij})], \tag{2.7}$$

found in Zehnwirth [5]. This identity suggests the estimator

$$\tilde{\alpha}_n(\mathbb{R}) = mF^{-1} \tag{2.8}$$

where F denotes the F-ratio statistic in the one-way ANOVA based on the treatments $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$. However, this estimator is not consistent because Var $E_{F_i}(X_{ij})$ is being estimated by a statistic that is biased upwards as seen from (2.10) below. Adopting the usual notation for the one-way ANOVA setup we have

$$\begin{aligned} MS_W = \textstyle\sum_{i=1}^{n} \sum_{j=1}^{m} (X_{ij} - \bar{X}_{i\cdot})^2/(n(m-1)) &\xrightarrow[n\to\infty]{\Pr} \\ E[\textstyle\sum_{j=1}^{m} (X_{ij} - \bar{X}_{i\cdot})^2/(m-1)] &= E[\mathrm{Var}_{F_i}(X_{ij})]. \end{aligned} \tag{2.9}$$

Similarly,

$$\begin{aligned} MS_B/m = \textstyle\sum_{i=1}^{n} (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2/(n-1) &\xrightarrow[n\to\infty]{\Pr} \mathrm{Var}[\bar{X}_{i\cdot}] \\ &= \mathrm{Var}[E_{F_i}(X_{ij})] + \frac{1}{m} E[\mathrm{Var}_{F_i}(X_{ij})]. \end{aligned} \tag{2.10}$$

Therefore, one has

$$(MS_B - MS_W)/m \xrightarrow[n\to\infty]{\Pr} \mathrm{Var}[E_{F_i}(X_{ij})]. \tag{2.11}$$

Thus, $m(F-1)^{-1}$ is consistent for $\alpha(\mathbb{R})$. But then F can be less than or equal to one, making the above estimator possibly negative. If one defines

$$\begin{aligned} \hat{\alpha}_n(\mathbb{R}) &= 0 && \text{if } F \leq 1 \\ &= m(F-1)^{-1} && \text{if } F > 1, \end{aligned} \tag{2.12}$$

then

$$\hat{\alpha}_n(\mathbb{R}) \xrightarrow[n\to\infty]{\Pr} \alpha(\mathbb{R}).$$

Note that the consistency argument does not rely on any special properties peculiar to the Dirichlet process. Indeed, $\hat{\alpha}_n(\mathbb{R})$ is consistent for the ratio $E[\mathrm{Var}_{F_i}(X_{ij})]/\mathrm{Var}[E_{F_i}(X_{ij})]$ in the case where $F_1, F_2, \cdots, F_{n+1}$ belong to a parametric family and the parameter is assigned a prior distribution (provided second order moments exist).

FINAL COMMENTS. Perhaps it is worthwhile noting that the present framework does not fall within Robbins' [4] or Deely and Zimmer's [1] general framework for proving asymptotic optimality. Our proof is facilitated by the boundedness property of the estimators. Convergence in probability plus uniform boundedness implies convergence in L^2-space.

ACKNOWLEDGEMENT. I am indebted to the referee for pointing out the inconsistency of $\tilde{\alpha}_n(\mathbb{R})$and for suggesting the alternative $\hat{\alpha}_n(\mathbb{R})$. His proof was based on a special property of the Dirichlet process.

REFERENCES

[1] DEELY, J. J. and ZIMMER, W. J. (1976). Asymptotic optimality of the empirical Bayes procedure. *Ann. Statist.* **4** 576–580.

[2] FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

[3] KORWAR, R. M. and HOLLANDER, M. (1976). Empirical Bayes estimation of a distribution function. *Ann. Statist.* **4** 581–588.

[4] ROBBINS, M. E. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–10.

[5] ZEHNWIRTH, B. (1977). The mean credibility formula is a Bayes rule. *Scand. Actuar. J.* **4** 213–216.

[6] ZEHNWIRTH, B. (1978). The credible distribution function is an admissible Bayes rule. *Scand. Actuar. J.* **3** 121–127.

LABORATORY OF ACTUARIAL MATHEMATICS
UNIVERSITY OF COPENHAGEN
UNIVERSITETSPARKEN 5
DK 2100 COPENHAGEN
DENMARK