The Annals of Statistics
1976, Vol. 4, No. 6, 1290–1293

SOME EMPIRICAL BAYES RESULTS IN THE CASE OF COMPONENT PROBLEMS WITH VARYING SAMPLE SIZES FOR DISCRETE EXPONENTIAL FAMILIES[1]

By Thomas E. O'Bryan

University of Wisconsin-Milwaukee

Consider a modified version of the empirical Bayes decision problem where the component problems in the sequence are not identical in that the sample size may vary. In this case there is not a single Bayes envelope $R(\cdot)$, but rather a sequence of envelopes $R^{m(n)}(\cdot)$ where $m(n)$ is the sample size in the nth problem. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots)$ be a sequence of i.i.d. G random variables and let the conditional distribution of the observations $\mathbf{X}_n = (X_{n,1}, \cdots, X_{n,m(n)})$ given $\boldsymbol{\theta}$ be $(P_{\theta_n})^{m(n)}$, $n = 1, 2, \cdots$. For a decision concerning θ_{n+1}, where θ indexes a certain discrete exponential family, procedures t_n are investigated which will utilize all the data $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_{n+1}$ and which, under certain conditions, are asymptotically optimal in the sense that $E|t_n - \theta_{n+1}|^2 - R^{m(n+1)}(G) \rightarrow 0$ as $n \rightarrow \infty$ for all G.

1. Introduction. Empirical Bayes decision theory, as introduced by Robbins (1956), deals with a sequence of independent repetitions of a given Bayesian statistical decision problem, called the component problem, where each problem in the sequence has the same unknown prior distribution G. The history of the empirical Bayes problem is such that the only case that seems to have been considered thus far is where the sequence of problems consists of identical repetitions of a given component problem. One could ask whether it is possible to apply empirical Bayes procedures to sequences of independent but not identical decision problems all having the same unknown G. To answer this question in part, this paper considers the case where the sequence of problems are identical except for sample size and concern squared error loss estimation involving certain discrete exponential families. This component problem has been discussed by Robbins (1964), Johns (1957), Macky (1966), and Hannan and Macky (1971).

In the situation considered here, there is a sequence of independent random vectors $\{(\theta_i, \mathbf{X}_i)\}$, $i = 1, 2, \cdots$, where $\mathbf{X}_i = (X_{i,1}, \cdots, X_{i,m(i)})$ is the sample of size $m(i)$ from the ith problem. The random variables θ_i are unobservable and i.i.d. with distribution G. Conditional on $\theta_i = \theta$, $X_{i,1}, \cdots, X_{i,m(i)}$ are i.i.d. with probability function

$$f_\theta(x) = \theta^x z(\theta) g(x)\,, \qquad x \in \chi = \{0, 1, \cdots\} \tag{1.1}$$

Received January 1973; revised December 1975.

[1] Research supported by National Science Foundation, Grants No. GP-31123X and GP33677X. This paper represents part of the author's dissertation at Michigan State University. The revision of this paper was supported by a grant from the University of Wisconsin.

AMS 1970 *subject classifications.* Primary 62C99; Secondary 62C99, 62C25.

Key words and phrases. Decision theory, empirical Bayes, squared error loss estimation.

where $g(x) > 0$, and $\theta \in \Theta \subset \pi \equiv \{\theta \geqq 0 : \sum_{x=0}^{\infty} \theta^x g(x) < \infty\}$. The function z defined by

$$z(\theta) \equiv (\sum_{x=0}^{\infty} \theta^x g(x))^{-1}, \qquad \theta \geqq 0, \tag{1.2}$$

is continuous on $[0, \infty)$. The random variable Y_i, where

$$Y_i = \sum_{j=1}^{m(i)} X_{i,j}, \tag{1.3}$$

is sufficient for θ_i and, with $m(i)$ and θ_i abbreviated by m and θ, has probability function

$$f_{\theta,m}(y) = \theta^y z^m(\theta) g_m(y), \qquad y \in \chi \tag{1.4}$$

where

$$g_m(y) = \sum_{A_m(y)} \prod_{i=1}^{m} g(x_i) > 0, \qquad y \in \chi \tag{1.5}$$

with $A_m(y) = \{(x_1, \cdots, x_m) : \sum_{i=1}^{m} x_i = y\}$.

Consider a decision rule t_n for use in the $(n+1)$st problem which depends on $\mathbf{X}_1, \cdots, \mathbf{X}_n$. The risk of t_n conditional on $\mathbf{X}_1, \cdots, \mathbf{X}_n$ is

$$R^{m(n+1)}(t_n, G) \equiv E(t_n(\mathbf{X}_{n+1}) - \theta)^2. \tag{1.6}$$

With the overall expected loss for the decision concerning θ_{n+1} denoted by

$$R_n(t_n, G) \equiv ER^{m(n+1)}(t_n, G), \tag{1.7}$$

it follows that $R_n(t_n, G) \geqq R^{m(n+1)}(G)$, the Bayes envelope in the $(n+1)$st problem, which motivates the following definition paralleling Robbins (1964).

DEFINITION. A sequence of decision rules $\{t_n\}$ is said to be asymptotically optimal (a.o.) relative to G if

$$\lim_{n\to\infty} \{R_n(t_n, G) - R^{m(n+1)}(G)\} = 0. \tag{1.8}$$

2. An a.o. sequence $\{t_n\}$. Letting $m = m(n+1)$, a nonrandomized rule which is Bayes with respect to G in the $(n+1)$st problem is given by

$$t_G{}^m(y) = q_m(y+1)/q_m(y) \tag{2.1}$$

where

$$q_m(y) = \int_\Theta \theta^y z^m(\theta) G(d\theta) \tag{2.2}$$

and ratios 0/0 are to be interpreted as 0 throughout the paper. The following lemma, motivated by the approach of Robbins (1964), is a consequence of the fact that the difference in (1.8) can be expressed as $E(t_n(Y_{n+1}) - t_G{}^m(Y_{n+1}))^2$. P_y denotes probability conditional on $Y_{n+1} = y$.

LEMMA. *Suppose*

$$m(i) \leqq M < \infty, \qquad i = 1, 2, \cdots \tag{A1}$$

and

$$\Theta \subset [0, \beta] \quad \textit{for some} \quad \beta < \infty. \tag{A2}$$

With t_n^m *defined for each m to be a y-measurable decision rule truncated to* $[0, \beta]$ *in the sample size m problem depending on* $\mathbf{X}_1, \cdots, \mathbf{X}_n$, *then*

$$t_n^m(y) - t_G^m(y) \to_{P_y} 0 \qquad \text{for all} \quad y \in \chi\,, \quad 1 \leqq m \leqq M \tag{2.3}$$

implies that the sequence $\{t_n\}$ *with* $t_n = t_n^{m(n+1)}$ *is* a.o. *relative to all G.*

This lemma shows that in order to find a.o. sequences under (A1) and (A2) it suffices to approximate t_G^m as $n \to \infty$ for each m. With G unknown, q_m (and hence t_G^m) is unknown. Under (A2) since z is continuous, for each m and every $\varepsilon > 0$, there exists a polynomial $\sum \gamma_k \theta^k$ such that

$$|z^{m-1}(\theta) - \textstyle\sum \gamma_k \theta^k| < \varepsilon \qquad \text{for all} \quad \theta \in [0, \beta]\,. \tag{2.4}$$

Defining

$$q_{m,\varepsilon}(y) = \textstyle\sum \gamma_k q_1(y + k)\,, \qquad y \in \chi\,, \tag{2.5}$$

it follows that

$$|q_m(y) - q_{m,\varepsilon}(y)| \leqq \varepsilon q_1(y)\,, \qquad y \in \chi\,. \tag{2.6}$$

With $g_0(u) = 1$ if $u = 0$ and 0 if $u \neq 0$, and with $g_m(u) = 0$ if $m \geqq 1$ and $u < 0$, an unbiased estimate of $q_{m,\varepsilon}(y)$ is given by

$$\bar{q}_{m,\varepsilon}(y) = n^{-1} \textstyle\sum_{i=1}^n (\sum \gamma_k g_{m(i)-1}(Y_i - y - k)/g_{m(i)}(Y_i)) \tag{2.7}$$

since, for $Y \sim g_m q_m$,

$$\begin{aligned} E[g_{m-1}(Y - u)/g_m(Y)] &= \textstyle\sum_{y=0}^\infty g_{m-1}(y-u) q_m(y) \\ &= \textstyle\int_\Theta z(\theta)\theta^u (\sum_{y=0}^\infty g_{m-1}(y-u) z^{m-1}(\theta)\theta^{y-u})\, dG(\theta) \\ &= q_1(u)\,. \end{aligned}$$

Under (A1), the summand in (2.7) is bounded in absolute value for each i by

$$\rho(\varepsilon, y) \equiv \sup_{1 \leqq m \leqq M} \{\textstyle\sum |\gamma_k|/g(y + k)\} \tag{2.8}$$

which is independent of G. From (2.6) and (2.8),

$$E_y(\bar{q}_{m,\varepsilon}(y) - q_m(y))^2 \leqq n^{-1}\rho^2(\varepsilon, y) + \varepsilon^2 q_1^2(y)\,. \tag{2.9}$$

With $\varepsilon \to 0$ there exist $n = n(\varepsilon, y)$, a function of ε and y, such that $n^{-1}\rho^2(\varepsilon, y) \to 0$. By inverting the function for each fixed y, a choice $\varepsilon = \varepsilon(y, n) \to 0$ is obtained such that $n^{-1}\rho^2(\varepsilon, y) \to 0$. For such choices

$$\bar{q}_{m,\varepsilon}(y) \to_{P_y} q_m(y)\,, \qquad y \in \chi,\ 1 \leqq m \leqq M\,. \tag{2.10}$$

Hence an application of the lemma yields the following theorem.

THEOREM. *Let* (A1) *and* (A2) *obtain. With* $t_n = t_n^{m(n+1)}$ *where* t_n^m *is given by*

$$t_n^m(y) = \min\{\bar{q}_{m,\varepsilon}^+(y + 1)/\bar{q}_{m,\varepsilon}^+(y)),\ \beta\} \tag{2.11}$$

where $a^+ = \max\{a, 0\}$ with a choice $\varepsilon = \varepsilon(y, n)$ such that (2.10) obtains, the sequence $\{t_n\}$ is a.o. relative to every G.

3. Remarks. In case z is a polynomial in θ, (A2) must hold and a choice corresponding to $\varepsilon \equiv 0$ exists. In case (A1) does not hold, i.e., the sample sizes are unbounded, and

$$\text{(A2}^+\text{)} \qquad \Theta \subset [0, \beta] \quad \text{where} \quad \beta \in \pi$$

holds, sequences of a.o. rules do exist (see O'Bryan (1972)) but such procedures are difficult to calculate and consequently would lack practical significance.

A similar technique of estimating $z^{m-1}(\theta)$ and $\theta z^{m-1}(\theta)$ by polynomials in $e^{-\theta}$ has been employed by O'Bryan and Susarla (1975) to handle a more general situation involving continuous exponential families with density

$$f_\theta(x) = e^{-\theta x} z(\theta) g(x) \tag{3.1}$$

which requires a modification of the estimate in (2.7).

Acknowledgment. The author wishes to thank his thesis advisor Professor Dennis C. Gilliland for the introduction to empirical Bayes decision theory and for many useful discussions concerning the problem treated here. Also thanks are due to Professor James Hannan for suggestions which led to improvements of some of the results discussed here.

REFERENCES

[1] HANNAN, JAMES and MACKY, DAVID W. (1971). Empirical Bayes squared error loss estimation of unbounded functionals in exponential families. Unpublished.

[2] JOHNS, M. V., JR. (1957). Non-parametric empirical Bayes procedures. *Ann. Math. Statist.* **28** 649–669.

[3] MACKY, DAVID W. (1966). Empirical Bayes estimation in an exponential family. RM-176, Department of Statistics and Probability, Michigan State Univ.

[4] O'BRYAN, THOMAS E. (1972). Empirical Bayes results in the case of non-identical components. RM-306, Department of Statistics and Probability, Michigan State Univ.

[5] O'BRYAN, THOMAS E. and SUSARLA, V. (1975). Empirical Bayes estimation with nonidentical components. Continuous case. Submitted for publication.

[6] ROBBINS, HERBERT (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 157–163, Univ. of California Press.

[7] ROBBINS, HERBERT (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WISCONSIN
MILWAUKEE, WISCONSIN 53201