The Annals of Statistics
1976, Vol. 4, No. 3, 576-580

ASYMPTOTIC OPTIMALITY OF THE EMPIRICAL BAYES PROCEDURE[1]

BY J. J. DEELY AND W. J. ZIMMER

University of Canterbury, Christchurch, and University of New Mexico

The purpose of this paper is to obtain weaker sufficient conditions than those given by Robbins [6] for the asymptotic optimality of empirical Bayes procedures. In so doing a general method for providing asymptotic optimality is given which because of its simplicity should prove quite useful not only for the empirical Bayes problem but for other related asymptotic problems as well. Three areas of application are indicated herein.

0. Introduction. In a fundamental paper, Robbins [6] formulated the empirical Bayes problem and gave two basic theorems which could be used to prove that an empirical Bayes procedure was asymptotically optimal (a.o.). One of these theorems gave conditions under which a consistent estimater of the Bayes procedure was indeed a.o. Unfortunately the conditions of this theorem did not admit a quadratic loss function for an unbounded parameter space. Rutherford and Krutchkoff [8] recognized this fact and with weaker conditions obtained what they called "ε asymptotic optimality" for the quadratic loss case. The theorem given in Section 1 has weaker conditions than either Robbins or Rutherford and Krutchkoff, admits the quadratic loss function and obtains the a.o. property exactly. This result is a straightforward application of a theorem given in Pratt [4] who subsequently (see Pratt [5]) acknowledges priority to earlier authors, W. H. Young in 1910 and Hans Hahn and A. Rosenthal in 1948.

In Section 2 the special case of quadratic loss is treated. Using a result by Verbeek [9] on bounds for the posterior mean and the theorem of Section 1, we obtain the a.o. property for an empirical Bayes estimator of the mean of a normal distribution considered by Miyasawa [3] and for all the empirical Bayes estimators of Rutherford and Krutchkoff [7] for the location parameter case.

In Section 3 other applications of the main theorem are given, one being a confidence interval problem of Deely and Zimmer [1] and the other concerning asymptotic G-minimax estimators as indicated by Jackson et al. [2].

1. Notation and theorem. The usual decision theory model is assumed, namely a state of nature, ω, taking values in Ω, an action a taking values in A the set of possible actions, a loss function $L(a, \omega) \geqq 0$, an observable random

Received November 1974; revised September 1975.

[1] Research supported in part by the Office of Naval Research Grant No. N00014-68-A-0155-0003.

AMS 1970 *subject classifications*. Primary 62F15.

Key words and phrases. Asymptotic optimality, empirical Bayes application of a dominated convergence theorem.

variable X taking values in $\mathscr{X}$, a decision procedure t which maps $\mathscr{X}$ into A, and a probability distribution G on Ω. It is assumed that X has a conditional probability density function with respect to some measure μ on $\mathscr{X}$ given by $f(x \mid \omega)$. The Bayes risk, simply risk, associated with a decision procedure t is given by

$$R(t, G) = \int_\Omega \int_{\mathscr{X}} L(t(x), \omega) f(x \mid \omega)\, d\mu(x)\, dG(\omega)$$

and the optimal Bayes decision procedure t_G is obtained by minimizing over A for each x the function

$$\varphi(a, x) = \int_\Omega L(a, \omega) f(x \mid \omega)\, dG(\omega)\,.$$

It is assumed the desired minimum can be attained with some $a \in A$. When G is unknown but assumed to exist, we assume past history is available in the form of independent observations $(x_1, \omega_1), \cdots, (x_n, \omega_n)$ on X but not ω, the second element ω_i indicating that the conditional distribution governing the values of x_i was given by $f(\cdot \mid \omega_i)$, the ω_i itself having first been drawn from G. We will call x_i an observation of the random variable X_i which is identically distributed as X taking values in $\mathscr{X}$ and having unconditional probability density function

$$f_G(x) = \int_\Omega f(x \mid \omega)\, dG(\omega)\,.$$

A function $t_n(x) = t_n(x_1, \cdots, x_n; x)$ based on the past and present observations and taking values in A is called an *empirical Bayes* decision procedure if $t_n(x) \to_p t_G(x)$ at each x, where p is taken with respect to the past history $x_1, \cdots, x_n$. With t_n we associate a global risk

$$\tilde{R}(t_n, G) = \int_\Omega \int_{\mathscr{X}} E[L(t_n(x), \omega)] f(x \mid \omega)\, d\mu(x)\, dG(\omega)$$

where E indicates the expectation with respect to $\prod_{i=1}^n f_G(x_i)$, the joint unconditional distribution of $X_1, \cdots, X_n$. If an empirical Bayes procedure has the additional property that $\tilde{R}(t_n, G) \to R(t_G, G)$ as $n \to \infty$ then t_n is said to be *asymptotically optimal* (a.o.). A more detailed explanation of the above and other relevant material including an example of an empirical Bayes procedure which is not a.o. can be found in Robbins [6].

We will need the following.

LEMMA. *Let (Y, B, P) be a probability space and let $\{f_n\}$ and $\{g_n\}$ be two sequences of measurable functions such that*

(i) $f_n \to_p f$, $g_n \to_p g$,
(ii) $0 \leqq f_n \leqq g_n$ *for* $n = 1, 2, \cdots$,
(iii) $\lim \int g_n\, dP = \int g\, dP < \infty$.

Then $\lim \int f_n\, dP = \int f\, dP < \infty$.

This lemma is contained in Theorem 1 of Pratt [4].

THEOREM. *Let t_n be an empirical Bayes procedure and let the loss function be such that $L(t_n, \omega) \to_p L(t_G, \omega)$ for each ω in Ω. Suppose there exists a sequence of* *functions* $h_n(x, \omega) = h_n(x_1, \cdots, x_n; x, \omega)$ *such that*

(i) $h_n(x, \omega) \to_p h(x, \omega)$ *for each* (x, ω),
(ii) $L(t_n(x), \omega) \leqq h_n(x, \omega)$ *for* $n = 1, 2, \cdots$ *and for each* (x, ω), *and*
(iii) $\lim_{n\to\infty} E[h_n(x, \omega)] = E[\lim_n h_n(x, \omega)] < \infty$.

Then t_n *is* a.o.

PROOF. The proof follows from the lemma by taking $f_n = L(t_n(x), \omega)$ and $g_n = h_n(x, \omega)$. Note that the convergence in probability with respect to X_1, $X_2, \cdots$ at each (x, ω) gives convergence in probability on the product space $\mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X} \times \Omega$.

REMARKS. (i) The appropriate measurability of all the functions above is assumed. In practice the form of both t_n and h_n will be known by construction and the measurability is easily verified.

(ii) The statement "for all (x, ω)" can be replaced with "almost everywhere" where the measure implied is the product probability measure on the product space $\mathscr{X} \times \Omega$.

(iii) The condition on the loss function to insure that $L(t_n, \omega) \to_p L(t_G, \omega)$ can also easily be checked in practice. For example continuity in the first argument is sufficient.

(iv) As is often the case in practice one can find a consistent estimator of the Bayes procedure, i.e. an empirical Bayes procedure. The theorem implies that if in addition the loss structure preserves this convergence, then one looks for the bounding sequence h_n for which the computations of $\lim E[h_n]$ and $E[\lim h_n]$ can easily be made. If equality obtains, the a.o. property of the empirical Bayes procedure is then insured. It should be noted that finding a bounding *sequence* is much less restrictive in practical problems than finding a bounding *function*, the usual approach in applying the dominated convergence theorem. The examples in Sections 2 and 3 illustrate these ideas and indicate how a bounding sequence may be found.

(v) Finally if $\int \sup_{a \in A} L(a, \omega)\, dG(\omega) < \infty$ (Robbins' condition, see [6], page 4), then a bounding sequence can be found by taking $h_n \equiv \sup L(a, \omega)$ for each n. It is in this sense that the conditions are weaker than those of Robbins even though we do require an additional condition on the loss function.

2. We now turn our attention to the estimation problem ($A = \Omega$) under quadratic loss. If the posterior mean $E[\omega \mid x]$ exists, it is the Bayes estimator in this case, and we assume that $t_n(x)$ is a known consistent estimator of the posterior mean. That is, we assume the existence of an empirical Bayes estimator. It has been shown by Verbeek [9] that if ω is a location parameter and f and G are unimodal and symmetric then

$$|E[\omega \mid x]| \leqq |x| + |\lambda|$$

where λ is the mean of G. From t_n we form the estimator

$$\begin{aligned} \bar{t}_n(x) &= |x| + |\bar{x}_n| \,, &&\text{if} \quad t_n(x) > |x| + |\bar{x}_n| \\ &= t_n(x) \,, &&\text{otherwise} \\ &= -|x| - |\bar{x}_n| \,, &&\text{if} \quad t_n(x) < -|x| - |\bar{x}_n| \end{aligned}$$

where $\bar{x}_n = (1/n) \sum_{i=1}^n x_i$. Note that $\bar{x}_n \to_{\text{a.e.}} \lambda$ by the SLLN and hence $\bar{t}_n(x) \to_p E[\omega \mid x]$. Let

$$h_n(x, \omega) = (|x| + |\bar{x}_n|)^2 + 2|\omega|(|x| + |\bar{x}_n|) + \omega^2 \,.$$

Then

$$L(\bar{t}_n, \omega) = (\bar{t}_n(x) - \omega)^2 \leqq h_n(x, \omega) \,.$$

Now

$$h_n(x, \omega) \to_p (|x| + |\lambda|)^2 + 2|\omega|(|x| + |\lambda|) + \omega^2$$

and

$$\begin{aligned} E[h_n(x, \omega)] = E[X^2] &+ 2E[|X|]E[|\bar{X}_n|] + E[\bar{X}_n{}^2] \\ &+ 2E[|X|\,|\omega|] + 2E[|\bar{X}_n|]E[|\omega|] + E[\omega^2] \,. \end{aligned}$$

Thus it follows that

$$E[\lim h_n(x, \omega)] = \lim E[h_n(x, \omega)] \,,$$

and by the theorem we have that $\bar{t}_n$ is a.o.

REMARKS. (i) If $f(x \mid \omega)$ is normal with mean ω and known variance σ^2, then the above method simplifies both the proof and construction of an a.o. estimator for ω as given in Miyasawa [3]. Miyasawa found a consistent estimator t_n for the posterior mean using the fact that

$$E[\omega \mid x] = x + \sigma^2(f_G'(x)/f_G(x)) \,.$$

(ii) In [7] empirical Bayes estimators were derived for various families of $f(x \mid \omega)$ and in [8] ε asymptotic optimality was proved for a truncated version of any empirical Bayes estimator. This truncation and proof required that the prior moments of order greater than two be bounded by a known number. The above technique for the location parameter case can be used to obtain the a.o. precisely without knowledge of the prior moments.

3. We indicate in this section the way in which the main theorem can be applied to related convergence problems. Deely and Zimmer [1] considered a quality control problem and made rigorous use of past history to find a shorter than usual confidence interval on the mean of a normal random variable. Besides dealing with confidence intervals, they also suggested a point estimator t_n for the present lot mean ω. If ω is assumed to have a normally distributed prior with unknown mean λ and unknown variance β^2, the estimator t_n, a straightforward function of $x_1, \cdots, x_n$ past observations and a present observation x, is asymptotically optimal and has smaller mean squared error than the usual estimator when the expectation is taken over the joint distribution of the random variables $X_1, \cdots, X_n, X$ and ω. The estimator is of the form

$$t_n(x) = a_n \bar{x}_n + (1 - a_n)x$$

where $a_n = s_\sigma^2/(s_\sigma^2 + rs_\beta^2) \leqq 1$, r is the number of observations taken from each lot and s_σ^2, s_β^2 are the consistent estimates of the variance within each lot and the variance of the prior on ω respectively. The proof that $t_n(x) \to_p t_G(x)$ was straightforward. In this case the main theorem can be applied directly by noting that

$$(t_n(x) - \omega)^2 \leqq h_n(x, \omega)$$

for h_n as given in Section 2 and thus the fact that t_n is a.o. is obtained.

Another problem related to empirical Bayes is that which arises when one assumes only that the prior distribution belongs to a family with the same known first two moments. Then one can seek a procedure which is minimax over such a family, say G_2. See, for example, [2]. Such a procedure, say t^*, would have a risk given by $R(t^*, G)$. If these two moments exist but are unknown then one uses prior observations to estimate these moments and take the minimax procedure t_n^* with respect to these estimates. To prove that t_n^* is asymptotically G_2 minimax, it is required to show that $R(t_n^*, G) \to R(t^*, G)$ for any $G \in G_2$. In [2] such estimators were produced and the existence of the function h_n was indicated such that

$$(t_n^* - \omega)^2 \leqq h_n(x, \omega) .$$

This function h_n can be seen to satisfy (iii) of the theorem and thus the desired convergence is obtained.

REFERENCES

[1] Deely, J. J. and Zimmer, W. J. (1969). Shorter confidence intervals using prior observations. *J. Amer. Statist. Assoc.* **64** 378–386.

[2] Jackson, D. A., O'Donovan, T. M., Zimmer, W. J. and Deely, J. J. (1970). G_2-minimax estimators in the exponential family. *Biometrika* **57** 439–443.

[3] Miyasawa, Koichi (1961). An empirical Bayes estimator for the mean of a normal population. *Bull. Inst. Internat. Statist.* **38** 181–188.

[4] Pratt, J. W. (1960). On interchanging limits and integrals. *Ann. Math. Statist.* **31** 74–77.

[5] Pratt, J. W. (1966). Acknowledgement of priority. *Ann. Math. Statist.* **37** 1407.

[6] Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.

[7] Rutherford, J. R. and Kruchkoff, R. G. (1969). Some empirical Bayes techniques in point estimation. *Biometrika* **56** 133–137.

[8] Rutherford, J. R. and Krutchkoff, R. G. (1969). ε asymptotic optimality of empirical Bayes estimators. *Biometrika* **56** 220–223.

[9] Verbeek, A. (1973). Bounds for the posterior mean of a location parameter. Unpublished.

University of Canterbury
Department of Statistics
Christchurch
New Zealand

Department of Mathematics and Statistics
University of New Mexico
Albuquerque, New Mexico 87131