Ann. Inst. Statist. Math.
Vol. 40, No. 1, 187-193 (1988)

ON EMPIRICAL BAYES WITH SEQUENTIAL COMPONENT

DENNIS C. GILLILAND[1] AND ROHANA KARUNAMUNI[2]

[1]*Department of Statistics and Probability, Michigan State University, East Lansing, Michigan 48824, U.S.A.*
[2]*Department of Statistics and Applied Probability, University of Alberta, Edmonton, Alberta, Canada T6G 2G1*

(Received January 28, 1986; revised July 4, 1986)

Abstract. Laippala (1979, *Scand. J. Statist.*, **6**, 113–118, correction note, 7, 105; 1985, *Ann. Inst. Statist. Math.*, **37**, 315–327) has defined a concept within the empirical Bayes framework that he calls "floating optimal sample size". We examine this concept and show that it is one of many possibilities resulting from restricting the class of component sampling procedures in the empirical Bayes decision problem with a sequential component. All ideas are illustrated with the finite state component.

Key words and phrases: Empirical Bayes, sequential component, asymptotic optimality.

1. Introduction

We assume that the reader is familiar with the empirical Bayes decision problem (see, e.g., Maritz (1970), Robbins (1956, 1964), Susarla (1982) and Suzuki (1975)). As our component we take the m-truncated sequential decision problem (see Berger (1985)). Specifically, let $X_1,\ldots, X_m$ i.i.d. $P_\theta, \theta \in \Theta$, be the observable random variables taking values in the sample space $\mathscr{X}$. The component problem has actions $a \in \mathscr{A}$, loss function $L(\theta, a) \geq 0$, stopping rules $\tau \in \mathscr{T}$, (terminal) decision rules $\delta \in \mathscr{D}$, $\boldsymbol{d}=(\tau, \delta)$, constant cost per observation $c \geq 0$, (terminal) decision risk $r(\theta, \boldsymbol{d})$, Bayes terminal decision risk $r(G, \boldsymbol{d})$ for priors $G \in \mathscr{G}$, and infimum Bayes risk $r(G)$.

We take $\mathscr{T}$ to be the class of nonrandomized stopping rules that take at least one observation, i.e., that result in sample size N where $1 \leq N \leq m$. (In the empirical Bayes application, this ensures that at least one observation is made at each repetition of the component which allows for an updating of the empirical Bayes estimates.) With our notation, $\tau=(\tau_1,\ldots, \tau_m)$ where τ_k: $\mathscr{X}^k \to \{0,1\}$, $k=1,\ldots, m$, and the random variable N is defined by

$$[N = 1] = [\tau_1 = 1] ,$$

$$[N = k] = [\tau_1 = 0,\ldots, \tau_{k-1} = 0, \tau_k = 1] , \quad k = 2,\ldots, m .$$

The overall risk associated with $\boldsymbol{d}$, including cost for observations and terminal loss, is $R(\theta, \boldsymbol{d}) = r(\theta, \boldsymbol{d}) + c\, E_\theta N$; we let $R(G, \boldsymbol{d})$ denote its expectation with respect to the G measure on Θ. The Bayes envelope risk associated with a subclass $\mathscr{T}_* \times \mathscr{D}_* \subset \mathscr{T} \times \mathscr{D}$ is

$$R_*(G) = \inf\{R(G, \boldsymbol{d}) | \boldsymbol{d} \in \mathscr{T}_* \times \mathscr{D}_*\} . \tag{1.1}$$

The idea of restricting the class of component decision rules to a subclass appears in Gilliland and Hannan (1974, 1985).

If $\mathscr{T}_* = \mathscr{T}_F = \{\tau_1,\ldots, \tau_m\}$ where τ_k is the fixed-sample-size-k stopping rule (i.e., the associated stopping time satisfies $N_k = k$, $k = 1,\ldots, m$) and $\mathscr{D}_* = \mathscr{D}$, we denote the envelope risk by $R_F(G)$. We call R_F the fixed sample size envelope. The usual envelope in the empirical Bayes problem is the Bayes envelope resulting from the largest class of procedures. For $\mathscr{T}_* = \mathscr{T}$ and $\mathscr{D}_* = \mathscr{D}$, we denote this envelope risk by $R(G)$. Of course, $R_F(G) \geq R(G)$ for all G.

With $R_k = r_k + kc$ denoting the envelope associated with $\{\tau_k\} \times \mathscr{D}$, $k = 1, \ldots, m$, we have

$$R_F(G) = \min\{R_k(G) | k = 1,\ldots, m\} . \tag{1.2}$$

Laippala (1985, (3.11)) denotes $r_k(G)$ by W_G^k. However, as Laippala (1985, pp. 325–326) states, his "optimization rule" (3.7) does not define a minimizer for (1.2) since the smallest k such that $r_{k+1}(G) - r_k(G) + c \geq 0$ does not, in general, define a minimizer for (1.2) when $m \geq 3$. Of course, an optimum fixed sample size is defined by

$$n_G = \min\{k | R_k(G) = R_F(G), k = 1,\ldots, m\} , \tag{1.3}$$

or, in fact, any function that maps G into a minimizer of $R_k(G)$. This sample size together with a fixed sample size Bayes terminal decision rule with respect to G achieves minimum risk among all fixed sample size procedures. Berger (1985, Subsection 7.2) shows how to approximate n_G in some examples of untruncated sequential decision problems.

Example 1.1 (Testing Simple vs. Simple). Let $\Theta = \{0,1\}$, $\mathscr{A} = \{0,1\}$ and $L(0,0) = L(1,1) = 0$, $L(0,1) = L(1,0) = L > 0$, a constant. We identify a prior G on θ by the mass π it puts on the state 1 so that $\mathscr{G}$ can be identified with the unit interval. Let P_0 be $N(-1,1)$ and P_1 be $N(1,1)$. (Our example is the sequential version of that used by Robbins (1951) to introduce the idea of compound decision theory.) The posterior probability of $\theta = 1$ given $X_1 = x_1,\ldots, X_k = x_k$ is

$\pi_k=\pi \exp(\sum_1^k x_j)/\{\pi \exp(\sum_1^k x_j)+(1-\pi)\exp(-\sum_1^k x_j)\}$ and a Bayes nonrandomized decision rule is

$$\delta_k(\pi) = \begin{cases} 1 & \text{if} \quad \pi_k \geq 1/2 \\ 0 & \text{if} \quad \pi_k < 1/2 \,. \end{cases} \tag{1.4}$$

The event $\pi_k \geq 1/2$ is equivalent to $\sum_1^k x_j \geq c(\pi)$ where $c(\pi)=1/2 \ln((1-\pi)/\pi)$.

Consider the case of truncation at $m=2$. Let the loss for misclassification be $L=1$ and the cost per observation be $c=.05$.

A Bayes stopping rule $\tau(\pi)$ is defined by $\tau_2(\pi)=1$ and

$$\tau_1(\pi) = \begin{cases} 1 & \text{if} \quad r_1(\pi_1) + .05 - r_0(\pi_1) \geq 0 \\ 0 & \text{if} \quad r_1(\pi_1) + .05 - r_0(\pi_1) < 0 \,. \end{cases} \tag{1.5}$$

Here

$$r_0(\pi) = \pi[\pi \leq 1/2] + (1 - \pi)[\pi > 1/2] \,,$$

and

$$r_1(\pi) = \pi\Phi(c(\pi) - 1) + (1 - \pi)\{1 - \Phi(c(\pi) + 1)\} \,,$$

where Φ denotes the standard normal cdf. Calculations show that (1.5) is equivalent to

$$\tau_1(\pi) = \begin{cases} 1 & \text{if} \quad |\pi_1 - .5| \leq .361567 \\ 0 & \text{if} \quad |\pi_1 - .5| > .361567 \,, \end{cases} \tag{1.6}$$

and

$$\tau_1(\pi) = \begin{cases} 1 & \text{if} \quad |x_1 - c(\pi)| \leq c(.138433) \\ 0 & \text{if} \quad |x_1 - c(\pi)| > c(.138433) \,. \end{cases} \tag{1.7}$$

The envelope risk $R(\pi)$ resulting from the Bayes procedure $\boldsymbol{d}(\pi)=(\tau(\pi), \delta(\pi))$ was calculated for selected values of π and is plotted in Fig. 1 along with the fixed sample size envelopes $R_k(\pi)$, $k=1, 2$, where

$$\begin{aligned} R_k(\pi) = {} & \pi\Phi((c(\pi) - k)/\sqrt{k}) \\ & + (1 - \pi)\{1 - \Phi((c(\pi) + k)/\sqrt{k})\} + .05k \,. \end{aligned} \tag{1.8}$$

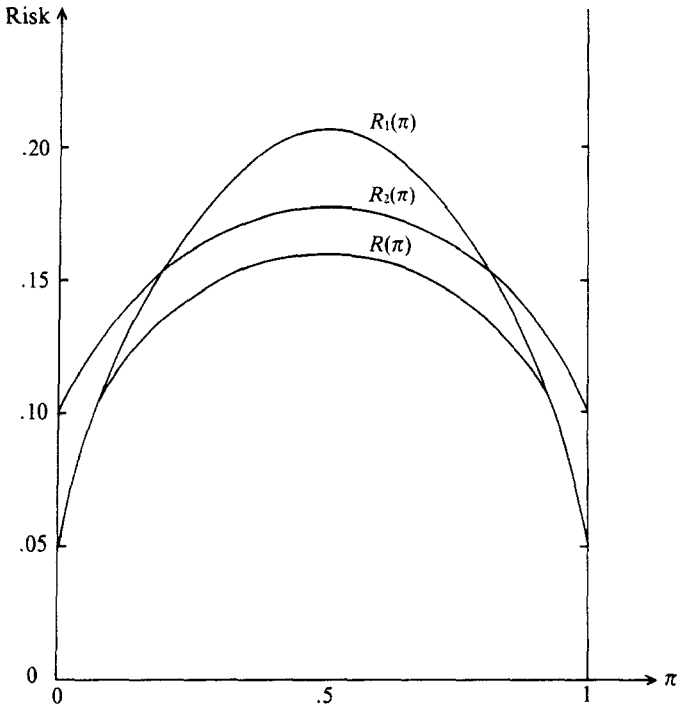

Fig. 1. Envelope risk functions—testing $N(-1,1)$ vs. $N(1,1)$.

Of course, the optimal fixed sample size risk envelope is $R_F(\pi)=\min\{R_1(\pi), R_2(\pi)\}$. Note that $R(\pi)$ is considerably less than $R_F(\pi)$ for priors π near .5.

Theorem 2.1 of the next section shows that the Bayes envelope risk $R(\pi)$ for the truncated sequential component is achieved in the limit by empirical Bayes decision procedures $\boldsymbol{d}^n=(\tau^n, \boldsymbol{\delta}^n)$ where τ^n and $\boldsymbol{\delta}^n$ are Bayes with respect to consistent estimates of π. The theorem as stated and proved also subsumes the usual fixed sample size case since, through restriction of the class of component stopping rules, one can produce the envelope R_k for any desired k or the envelope R_F.

2. Empirical Bayes

Let N_n denote the random sample size and $\boldsymbol{X}_n=(X_{n1},\ldots, X_{nN_n})$ denote the observed random vector in the n-th repetition of the sequential component problem. Of course, the event $[N_1=k]$ is $(X_{11},\ldots, X_{1k})$—measurable, $k=1,\ldots, m$, and $[N_n=k]$ is $(\boldsymbol{X}_1,\ldots, \boldsymbol{X}_{n-1}; X_{n1},\ldots, X_{nk})$—measurable, $k=1,\ldots, m$; $n=2, 3,\ldots$. This formalizes the empirical Bayes setup where data accumulated in stages $1, 2,\ldots, n-1$ are available to the decision maker going into stage n. The empirical Bayes rule determines the sample size sequence in contrast to the

nonrandom varying sample size problem of O'Bryan (1972, 1976, 1979) and O'Bryan and Susarla (1977).

A goal in empirical Bayes theory is asymptotic optimality. For an envelope $R*$, this means the construction of a sequence of stopping rules $(\tau^1, \tau^2, \ldots)$ and decision rules $(\delta^1, \delta^2, \ldots)$ where $\boldsymbol{d}^n=(\tau^n, \delta^n)$ depends upon $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n-1}$, $n=2, 3, \ldots$, such that

$$\lim_n E\, R(G, \boldsymbol{d}^n) = R_*(G) \quad \text{for all} \quad G \in \mathcal{G}. \tag{2.1}$$

The fact that the sequence $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ is not the usual i.i.d. sequence (as it is in the fixed sample size component) raises interesting and difficult questions concerning the efficient use of the data. With our restriction to stopping rules resulting in sample sizes $N_n \geq 1$, the first components $X_{11}, X_{21}, \ldots$ do form an i.i.d. sequence with common distribution being the G-mixture of $\{P_\theta | \theta \in \Theta\}$. This makes possible at least consistent estimation of G or of Bayes stopping rules and decision rules for the standard loss structures and distributions. For the case $\Theta=\{0, 1, \ldots, b\}$ finite with the family of mixtures $\Sigma\, g_i P_i$ identifiable, it is easy to construct consistent estimators of $G=(g_0, \ldots, g_b)$.

THEOREM 2.1. *Suppose that $\theta=\{0, 1, \ldots, b\}$ and that the loss function L is bounded. Suppose that $\mathcal{T}*$ is a specified subset of T, $R*$ denotes the associated envelope risk function and that $R*(G)$ is attained by the $\mathcal{T}*\times\mathcal{D}$-valued $\boldsymbol{d}(G)=(\tau(G), \delta(G))$, $G \in \mathcal{G}$. Suppose further that $\hat{G}_n=G_n(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n-1})$, $n=2, 3, \ldots$ is a $\mathcal{G}$-valued a.s. consistent estimator of $G \in \mathcal{G}$. (Here we identify $\mathcal{G}$ with the b-dimensional simplex of probability vectors on Θ and we denote the sup norm on Enclidean $(b+1)$-dimensional space by $\| \|$.) Then the empirical Bayes procedure $\boldsymbol{d}^n=\boldsymbol{d}(\hat{G}_n)$ is a.o. on $\mathcal{G}$, that is, satisfies* (2.1).

PROOF. We abbreviate $\hat{G}_n$ by $\hat{G}$. Using the definition of $R*$ and adding subtracting the nonnegative quantity $R(\hat{G}, \boldsymbol{d}(G))-R(\hat{G}, \boldsymbol{d}(\hat{G}))$ results in

$$\begin{aligned} 0 &\leq R(G, \boldsymbol{d}(\hat{G})) - R_*(G) \\ &\leq R(G, \boldsymbol{d}(\hat{G})) - R(G, \boldsymbol{d}(G)) + R(\hat{G}, \boldsymbol{d}(G)) - R(\hat{G}, \boldsymbol{d}(\hat{G})) \\ &\qquad\qquad \text{for} \quad G, \hat{G} \in \mathcal{G}. \end{aligned} \tag{2.2}$$

Bounds like (2.2) are basic in empirical Bayes analysis and go back at least to Hannan (1957). Hence,

$$\begin{aligned} 0 &\leq R(G, \boldsymbol{d}(\hat{G})) - R_*(G) \\ &\leq 2 \sup \{|R(\hat{G}, \boldsymbol{d}) - R(G, \boldsymbol{d})| : \boldsymbol{d} \in \mathcal{T}_* \times \mathcal{D}\}. \end{aligned} \tag{2.3}$$

Thus, if the convergence of the estimator is in the metric defined in RHS (2.3), the empirical Bayes procedure based on $\hat{G}$ will be a.o. Since L is bounded, the risk set associated with the component is bounded; specifically, $0 \leq R(\theta,$

$\boldsymbol{d}) \leq L + c\ m$ for $\theta = 0, 1, \ldots, b$. Hence, RHS (2.3) $\leq 2(L + c\ m) \| \hat{G} - G \|$ which together with (2.3) and the assumed consistency of $\hat{G}$ implies $R(G, \boldsymbol{d}^n) \to R*(G)$ a.s., $G \in \mathscr{G}$. Taking expectation with respect to $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n-1}$ establishes the L_1 convergence (2.1).□

Of course, choices $\mathscr{T}* = \{\tau_k\}$, $\mathscr{T}* = \mathscr{T}_F$, $\mathscr{T}* = \mathscr{T}$ result in envelopes $R* = R_k$, $R* = R_F$, $R* = R$ defined in the last section and displayed for a particular component in Fig. 1. For a component where $\mathscr{T}*$ includes truly sequential stopping rules, the implementation of the empirical Bayes stopping rule $\tau(\hat{G}_n)$ may require considerable calculation.

When $\mathscr{T}* = \mathscr{T}_F$, an optimal sample size n_G is defined by (1.3). The random sample size sequence N_n resulting from implementation of an a.o. empirical Bayes rule need not converge to n_G in any sense. This is clear in the context of the finite state component, in particular, the two state problem in Example 1.1. At a π where $R_1(\pi) = R_2(\pi)$, say π_0, $n_{\pi_0} = 1$. Of course, when $\hat{\pi}_n$ takes values in the region where $R_2(\pi) < R_1(\pi)$, then $N_n = 2$ so that at $\pi = \pi_0$, N_n can equal 2 infinitely often with positive probability.

Laippala (1985, Theorem 1) claims convergence of N_n to n_G in probability in his case. The condition that appears in the first line of this proof essentially removes boundary sets that prevent the convergence. This condition should be given in the hypothesis of the theorem since it is not without loss of generality.

This note was written in order to present the idea of the empirical Bayes model with a sequential component in clear and concise way. Our theorem shows that asymptotic optimality is achieved by simple empirical Bayes procedures in this general setting for a finite state component. Of course, small and moderate sample size behavior is of critical concern in most applications. We will not attempt to review the large literature involving fixed sample size components in regard to applications and the smoothing and admissibility and small to moderate sample size risk behavior of empirical Bayes procedures.

Acknowledgement

The authors wish to thank the referee for helpful comments and suggestions.

REFERENCES

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer, New York.

Gilliland, D. C. and Hannan, J. (1974). The finite state compound decision problem, equivariance and restricted risk components, RM-317, Statistics and Probability, Michigan State University.

Gilliland, D. C. and Hannan, J. (1985). The finite state compound decision problem, equivariance and restricted risk components, *Proc. Robbins' Symp.*, Brookhaven, New York.

Hannan, J. (1957). Approximation to Bayes risk in repeated play, *Contributions to the Theory of Games*, Vol. 3, 97–139, Princeton University Press.

Laippala, P. (1979). The empirical Bayes approach with floating optimal sample size in binomial experimentation, *Scand. J. Statist.*, **6**, 113–118; correction note, **7**, 105.

Laippala, P. (1985). The empirical Bayes rules with floating optimal sample size for exponential conditional distributions, *Ann. Inst. Statist. Math.*, **37**, 315–327.

Maritz, J. S. (1970). *Empirical Bayes Methods*, Methuen, London.

O'Bryan, T. (1972). Empirical Bayes results in the case of non-identical components, Ph. D. Dissertation, RM-306, Statistics and Probability, Michigan State University.

O'Bryan, T. (1976). Some empirical Bayes results in the case of component problem with varying sample sizes for discrete exponential families, *Ann. Statist.*, **4**, 1290–1293.

O'Bryan, T. (1979). Rates of convergence in a modified empirical Bayes estimation involving Poisson distributions, *Comm. Statist. A—Theory Methods*, **8**, 167–174.

O'Bryan, T. and Susarla, V. (1977). Empirical Bayes estimation with non-identical components, Continuous case, *Austral. J. Statist.*, **19**, 115–125.

Robbins, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems, *Proc. 3rd Berkeley Symp. Math. Statist. Prob.*, 131–148, University of California Press.

Robbins, H. (1956). The empirical Bayes approach to statistics, *Proc. 3rd Berkeley Symp. Math. Statist. Prob.*, Vol. 1, 157–163, University of California Press.

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

Susarla, V. (1982). Empirical Bayes theory, *Encyclopedia of Statistical Science*, (eds. Kotz and Johnson), **2**, 490–503, Wiley, New York.

Suzuki, Y. (1975). On empirical Bayes models, *Proc. 40th ISI Session*, 372–385.