Ann. Inst. Statist. Math.
37 (1985), Part A, 315-327

THE EMPIRICAL BAYES RULES WITH FLOATING OPTIMAL SAMPLE SIZE FOR EXPONENTIAL CONDITIONAL DISTRIBUTIONS

PEKKA LAIPPALA

(Received July 5, 1982)

Summary

We consider the empirical Bayes solution in such a situation where the sample size is successively determined by a rule which includes the Bayes risks and the observation costs. The empirical Bayes floating optimal sample size depends on current as well as on previous information assumed to be collected from earlier performances of similar decisions. The sampling is done from an exponential conditional distribution, with a single parameter. The proofs, which show the asymptotic optimality of the empirical Bayes solution, are presented for a hypotheses-testing problem. A straight generalization to a multiple decision problem is also given.

1. Introduction

Bayesian analysis presupposes knowledge of an adequate prior distribution, before it is possible to derive Bayesian decision functions and make use of Bayesian techniques in a classical sense. It is often difficult to assert what the prior distribution should be. This might lead to the application of a rule, the risk of which will be greater than the minimum risk attainable when the prior distribution is known. When we abandon the assumption of a known prior, the classical Bayes solution is unattainable. In practice, however, one easily finds fields of applications where the is no conceptual difficulty in postulating the existence of the prior. This alone does not make the Bayesian solution practicable, but if we deal with a repeated decision process, where similar decisions are made successively, the empirical Bayes solution may be attained.

The 'standard' empirical Bayes theory deals with repetitions of

Key words and phrases: Asymptotic optimality, empirical Bayes, exponential conditional distributions, floating optimal sample size, hypotheses testing, monotone multiple decisions.

independent and identical Bayesian decision problem where the prior is the same from one decision to another, i.e. the sample size is fixed. During the recent years, however, there have been papers which have dealt with independent repetitions of a given problem with varying sample sizes. When we apply the empirical Bayes techniques to a repeated decision process, the information available increases as the process continues. In the present paper we will utilize this property in such a way that when we decide the sample size of a certain decision, we allow the sample size to depend on information collected from earlier performances of the same decision. We should then be in a situation where the sample size may change from decision to decision. Consequently, we will call the approach the 'floating sample size approach'.

The criterion, on which the approach is based, includes the Bayes risks and the observation costs. We will look for such a sample size that the expected utility gained by one more observation is less than the cost of this observation. The sample size for each decision is determined, based on the above principle, before the sampling. This means that the sample size for the classical Bayes solution, with a known prior, is fixed, and that of the empirical Bayes solution, especially for the first decisions, tends to be high, but later it will properly approximate the classical Bayes solution.

The detailed presentation will be given to a hypotheses-testing problem where two composite hypotheses are tested against each other. The presentation includes the proofs of the asymptotic optimality. The generalization to a monotone multiple decision problem is also given, but without proofs, because the latter is a straight generalization of the former. For the conditional distribution of the observable random variable we assume that the sampling is done from an exponential family, with a single parameter.

The reference on the general empirical Bayes theory can be made to Robbins [9], [11], and Maritz [3]. The empirical Bayes hypotheses-testing has been dealt with in Robbins [10] and Samuel [13], where a fixed sample size is used. O'Bryan [5], [6] and O'Bryan and Susarla [7] have discussed the empirical Bayes rules with varying sample sizes. The determination of necessary sample size has been discussed in Suzuki [14]. The floating sample size approach in empirical Bayes context is introduced in Laippala [2], where the particular case of binomial experimentation is studied. In the present paper we generalize those results to exponential families in general.

2. General background

The decision problem may be characterized as follows. We deal with independent performances of the same decision problem with one and the same prior distribution. We have to make a decision the consequences of which will depend on an unknown realization λ of a random variable Λ. Λ has a prior distribution G which in empirical Bayes situation is assumed to remain unknown. Furthermore, we assume that the realizations λ remain unknown, but that we get information about their distribution by observing the values of x which are realizations of a random variable X. X has a conditional distribution depending on λ, say $f(x|\lambda)$.

We assume that $f(x|\lambda)$ may be given in the general form

$$f(x|\lambda)=\lambda^x g(x)h(\lambda) . \tag{2.1}$$

This form covers, e.g. the Poisson and the negative binomial distributions, and after a simple one to one transformation such continuous distributions as the normal and the gamma. The binomial distribution is not entirely covered by (2.1), and a slight modification is needed, see Laippala [2]. What we will observe with the first performances of the decision problem is the sequence of pairs of random variables $\{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n$ where i_k^* ($\leqq \hat{i}=i_0^*$) is the sample size at the kth stage, and z_{k,i_k^*} is the sum of i_k^* observations. By $\hat{i}$ we denote the maximum number of observations that we are willing to make, and by i_0^* the sample size of the first stage, called the 0-stage, because then no previous information is available and the optization rule of the sample size cannot be applied. The conditional distribution of Z, given λ and i_k^*, has the probability mass or density function given by

$$f(z|\lambda, i_k^*)=\lambda^z g_{i_k^*}(z)\{h(\lambda)\}^{i_k^*} . \tag{2.2}$$

When using the accumulated observations, the exponential form and the value of λ remain unaltered, but g is a function of i_k^*, which is added to emphasize the dependence.

The hypotheses to be tested against each other are of the type

$$\begin{aligned} &H_0: \lambda\leqq\lambda^* \\ &H_1: \lambda>\lambda^* \end{aligned} \tag{2.3}$$

where λ^* is a fixed critical constant. In the hypotheses-testing, the action space A consists of two actions, namely the action a_0 pertaining to the hypotheses H_0, and the action a_1 to H_1. We apply the following loss function:

$$(2.4)\qquad L(a_0, \lambda)=\begin{cases} 0 & \text{when } \lambda \leqq \lambda^* \\ \lambda-\lambda^* & \text{otherwise} \end{cases}$$
$$L(a_1, \lambda)=\begin{cases} 0 & \text{when } \lambda > \lambda^* \\ \lambda^*-\lambda & \text{otherwise}\,. \end{cases}$$

For the prior distribution G we make a standard assumption that $\int L(a_m, \lambda)dG(\lambda)<\infty$ to ensure that the Bayes risk is finite. The loss function above has been widely applied in empirical Bayes and also in classical Bayes literature.

For the floating optimal sample size we assume that at each stage, i.e. when making each decision, at least one observation has to be made, i.e. $i_k^* \geqq 1\ \forall k=1, 2, \cdots$. For simplicity, we let the cost for each observation be a constant c.

3. Classical Bayes solution: fixed optimal sample size

We begin with the explicit derivation of the classical Bayes solution: we operate with the assumption that the prior is known to be G. From hence on, the subscript G indicates a classical Bayes quantity.

Any decision function $\delta=\delta(z)$ has with respect to the prior G the risk

$$(3.1)\qquad \begin{aligned} W_G(\delta) &= \mathrm{E}\,[\delta(Z)L(a_1, \lambda)+\{1-\delta(Z)\}\,L(a_0, \lambda)] \\ &= \mathrm{E}_G\,[L(a_0, \lambda)]-\mathrm{E}\,[\delta(Z)\,\mathrm{E}_G\,\{L(a_0, \lambda)-L(a_1, \lambda)\,|\,Z\}] \end{aligned}$$

where E is the expectation with respect to the joint distribution of (Z, Λ) and E_G is the expectation with respect to the prior G. The decision function minimizing the risk (3.1)—and making it the Bayes risk—for fixed G is

$$(3.2)\qquad \delta_G(z)=\begin{cases} 1\,, & \text{when } \mathrm{E}_G\,\{L(a_0, \lambda)-L(a_1, \lambda)\,|\,z\}>0 \\ 0\,, & \text{otherwise}\,. \end{cases}$$

For a more detailed discussion of (3.1) and (3.2), see Laippala [2]. Recalling (2.2), and denoting $\Lambda=\lambda$, $i_k^*=i$, we have for the unconditional probability of any outcome Z the marginal probability

$$(3.3)\qquad f_i^G(z)=\int f(z\,|\,\lambda, i)dG(\lambda)\,.$$

According to (3.2), the Bayes rule δ_G^i is based on the conditional expectation, which, when applying the loss function (2.4), may be written in the form

$$(3.4) \qquad K_G(z, i) = \mathrm{E}_{G,i}\{L(a_0, \lambda) - L(a_1, \lambda)\,|\,z\} = \frac{f_i^G(z+1)/g_i(z+1)}{f_i^G(z)/g_i(z)} - \lambda^*$$

where the first term of the r.h.s. is the expectation of the posterior of Λ, given $Z=z$. When we define the sets

$$(3.5) \qquad A_j^G = \{z\,|\,K_G(z, j) > 0\}\,, \qquad j=1,\cdots,\hat{i}\,.$$

we may write that

$$(3.6) \qquad \mathrm{E}\,[\delta_G^i K_G(z, i)] = \int_z \delta_G^i K_G(z, i)\,d\mu(z) = \sum_{A_i^G} \left[\frac{f_i^G(z+1)/g_i(z+1)}{f_i^G(z)/g_i(z)} - \lambda^*\right] f_i^G(z)$$
$$= \sum_{A_i^G} [g_i(z) f_i^G(z+1)/g_i(z+1) - \lambda^* f_i^G(z)] = \sum_{A_i^G} [\varphi_G(z, i)]\,.$$

The optimization rule for the sample size may then be formally given in the form

$$(3.7) \qquad i_G^* = \inf\{i\,|\,h_{G,i} \geqq 0\}$$

where

$$(3.8) \qquad h_{G,i} = W_G(\delta_G^{i+1}) - W(\delta_G^i) + c = \sum_{A_i^G} [\varphi_G(z, i)] - \sum_{A_{i+1}^G} [\varphi_G(z, i+1)] + c\,.$$

The Bayes decision rule with optimal sample size i_G^* is

$$(3.9) \qquad \delta_G^{i_G^*} = \begin{cases} 1\,, & \text{when } K_G(z, i_G^*) > 0 \\ 0 & \text{otherwise} \end{cases}$$

and its Bayes risk is

$$(3.10) \qquad W_G(\delta_G^{i_G^*}) = \mathrm{E}_G\,[L(a_0, \lambda)] - \mathrm{E}\,[\delta_G^{i_G^*} K_G(z, i_G^*)]$$
$$= \int_{\lambda > \lambda^*} (\lambda - \lambda^*)\,dG(\lambda) - \sum_{A_{i_G^*}^G} [\varphi_G(z, i_G^*)]\,.$$

O'Bryan [5] has pointed out that when the sample size varies there is no single Bayes envelope functional of the prior, but rather a sequence of envelopes W_G^j. Then the Bayes rule (3.9) satisfies the condition

$$(3.11) \qquad W_G^j = W_G(\delta_G^j) = \min_{\delta^j} \{W_G(\delta_G^j)\}\,, \qquad j=1,\cdots,\hat{i}\,.$$

4. Empirical Bayes solution: floating optimal sample size

If we give up the assumption of a known prior, the classical Bayes solution is unattainable. In the empirical Bayes situation we suppose only that the prior is fixed, even if unknown. Suppose that the same decision has been made n times, i.e. we are at the stage $n+1$. Then we have at our disposal the information $\{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n$. We estimate the marginal probability $f_j^G(z)$ by

$$(4.1) \qquad f_j^n(z)=(n+1)^{-1}\sum_{k=0}^{n}\frac{g_j(z)g_{i_k^*-1}(z_{k,i_k^*}-z)}{g_{i_k^*}(z_{k,i_k^*})}I_{[z,\infty]}(z_{k,i_k^*})$$

where $g_0(u)=1$ iff $u=0$. The estimator is motivated by O'Bryan [5] and employed effectively in O'Bryan [6]. It uses all the past data and satisfies the condition

$$(4.2) \qquad \lim_{n\to\infty} f_j^n(z)=f_j^G(z)\ , \qquad (P)\ ,$$

provided that one assumes $\lambda\in[0, \beta]$ where $\beta<\infty$. The additional assumption of the bounded parameter space cannot be regarded as a practical limitation in applications. In continuous case, the marginal density may be estimated as follows. Let q_j^n be the number of those stages where at least j observations are made. Then $f_j^G(z)$ is estimated by

$$(4.3) \qquad f_j^n(z)=[F_j^n(z+h_n)-F_j^n(z-h_n)]/2h_n$$

where $h_n=dn^{-0.5}$ with $d>0$ a constant and where $F_j^n(z)$ is the empirical distribution function defined by

$$(4.4) \qquad F_j^n(z)=(1/q_j^n)\ \#\{k\,|\,z_{kj}\leqq z,\ 0\leqq k\leqq n\}\ , \qquad q_j^n>0\ .$$

The empirical density $f_j^n(z)$ satisfies also the condition (4.2), see Rosenblatt [12] and Parzen [8]. Considering the applicability of the procedure, we now agree that every now and then, e.g. successively after a fixed number stages, the maximum number of observations is made. Then the formula (4.2) holds for every j, $j=1,\cdots,\hat{i}$. For other motivation of the modification, we refer to the Section 7.

Substituting now the estimator $f_j^n(z)$ for $f_j^G(z)$ in the classical Bayes formulae of the previous section, we get their empirical Bayes counterparts

$$(4.5) \qquad K_n(z, i\,; \{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n)=\frac{f_i^n(z+1)/g_i(z+1)}{f_i^n(z)/g_i(z)}-\lambda^*$$

$$(4.6) \qquad A_i^n=\{z\,|\,K_n(z, i\,; \{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n)>0\}$$

$$(4.7) \qquad \varphi_n(z, i\,; \{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n)=g_i(z)f_i^n(z+1)/g_i(z+1)-\lambda^* f_i^n(z)$$

$$(4.8) \qquad i_{n+1}^*=\inf\{i\,|\,h_{n,i}\geqq 0\}$$

$$(4.9) \qquad h_{n,i}=\sum_{A_i^n}[\varphi_n(z, i\,; \{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n)]-\sum_{A_{i+1}^n}[\varphi_n(z, i+1\,; \{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n)]+c$$

$$(4.10)\quad \delta_n^{i_{n+1}^*}=\begin{cases}1, & \text{when } K_n(z, i_{n+1}^*; \{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n)>0\\ 0 & \text{otherwise}.\end{cases}$$

In the empirical Bayes situation the Bayes risk is estimated by

$$(4.11)\qquad W_n(\delta_n^{i_{n+1}^*})=\mathrm{E}_G\,[L(a_0, \lambda)]-\sum_{A_{i_{n+1}^*}^n}[\varphi_n(z, i_{n+1}^*; \{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n)]\,.$$

The actual Bayes risk of the empirical Bayes rule is the conditional Bayes risk, given the prior G:

$$(4.12)\qquad \begin{aligned} W_G(\delta_n^{i_{n+1}^*})&=\mathrm{E}_G\,[L(a_0, \lambda)]-\mathrm{E}\,[\delta_n^{i_{n+1}^*}K_G(z, i_{n+1}^*)]\\ &=\mathrm{E}_G\,[L(a_0, \lambda)]-\sum_{A_{i_{n+1}^*}^n}[\varphi_G(z, i_{n+1}^*)]\end{aligned}$$

on the convergence of which the asymptotic optimality depends.

5. Asymptotic properties

Recalling the respective definitions, it is seen that for fixed j and z

$$(5.1)\qquad \lim_{n\to\infty} K_n(z, j; \{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n)=K_G(z, j)\,, \qquad (P)\,,$$

$$(5.2)\qquad \lim_{n\to\infty} h_{n,j}=h_{G,j}\,, \qquad (P)\,,$$

$$(5.3)\qquad \lim_{n\to\infty} \varphi_n(z, j)=\varphi_G(z, j)\,, \qquad (P)\,.$$

THEOREM 1.

$$(5.4)\qquad \lim_{n\to\infty} i_{n+1}^*=i_G^*\,, \qquad (P)\,.$$

PROOF. Suppose that h_{G,i_G^*} is not exactly 0. The event $i_{n+1}^*\neq i_G^*$ can come about either through $i_{n+1}^*<i_G^*$ or $i_{n+1}^*>i_G^*$. Accordingly,

$$(5.5)\qquad \{i_{n+1}^*\neq i_G^*\}=\left[\bigcup_{j=1}^{i_G^*-1}\{h_{n,j}\geqq 0\}\right]\cup[\{h_{n,i_G^*}<0\}]\,.$$

Because $h_{G,j}<0$, $j=1,\cdots, i_G^*-1$,

$$(5.6)\qquad \{h_{n,j}\geqq 0\}\subseteqq\{h_{n,j}-h_{G,j}\geqq\varepsilon\}\subseteqq\{|h_{n,j}-h_{G,j}|\geqq\varepsilon\}$$

for $\varepsilon>0$, $\varepsilon\leqq\min\{-h_{G,j}, j=1,\cdots, i_G^*-1\}$. On the other hand, $h_{G,i_G^*}>0$, and thus

$$(5.7)\qquad \{h_{n,i_G^*}<0\}\subseteqq\{h_{n,i_G^*}-h_{G,i_G^*}\leqq-\varepsilon\}\subseteqq\{|h_{n,i_G^*}-h_{G,i_G^*}|\geqq\varepsilon\}$$

for $\varepsilon>0$, $\varepsilon\leqq h_{G,i_G^*}$. From (5.5)-(5.7) it follows that

$$(5.8)\qquad \{i_{n+1}^*\neq i_G^*\}=\left[\bigcup_{j=1}^{i_G^*-1}\{h_{n,j}\geqq 0\}\right]\cup[\{h_{n,i_G^*}<0\}]\subseteqq\bigcup_{j=1}^{i_G^*}\{|h_{n,j}-h_{G,j}|\geqq\varepsilon\}$$

for $\varepsilon>0$, $\varepsilon\leqq\min\{|h_{G,j}|, j=1,\cdots, i_G^*\}$. Recalling (5.2), and letting $n\to\infty$, the last line (5.8) tends to ϕ, and the result follows.

THEOREM 2.

$$\lim_{n\to\infty} W_G(\delta_n^{i_{n+1}^*})=W_G(\delta_G^{i_G^*}) . \tag{5.9}$$

PROOF. From (3.10) and (4.12) it follows that it suffices to show that

$$\lim_{n\to\infty} \mathrm{E}\,[\delta_n^{i_{n+1}^*}K_G(z, i_{n+1}^*)]=\mathrm{E}\,[\delta_G^{i_G^*}K_G(z, i_G^*)] . \tag{5.10}$$

$$\begin{aligned}&\{\lim_{n\to\infty} \mathrm{E}\,[\delta_n^{i_{n+1}^*}K_G(z, i_{n+1}^*)]\neq\mathrm{E}\,[\delta_G^{i_G^*}K_G(z, i_G^*)]\}\\ &\qquad\subseteqq\{\lim_{n\to\infty} i_{n+1}^*\neq i_G^*\}\cup\{\lim_{n\to\infty} \mathrm{E}\,[\delta_n^{i_G^*}K_G(z, i_G^*)]\neq\mathrm{E}\,[\delta_G^{i_G^*}K_G(z, i_G^*)]\} .\end{aligned} \tag{5.11}$$

We restrict ourselves to the set $\{i_{n+1}^*=i_G^*\}$, i.e. to the second term of the r.h.s. of (5.11):

$$\begin{aligned}\mathrm{E}\,[\delta_n^{i_G^*}K_G(z, i_G^*)]&=\int_\lambda\int_{\{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n}\int_z \delta_n^{i_G^*}(\{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n, z)[L(a_0, \lambda)-L(a_1, \lambda)]\\ &\qquad\cdot f_{i_G^*}^G(z_{1,i_G^*})\cdots f_{i_G^*}^G(z_n, i_G^*)f(z\,|\,\lambda, i_G^*)d\mu^{n+1}dG(\lambda)\\ &=\int_{\{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n}\int_z \delta_n^{i_G^*}(\{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n, z)\\ &\qquad\cdot f_{i_G^*}^G(z_{1,i_G^*})\cdots f_{i_G^*}^G(z_{n,i_G^*})K_G(z, i_G^*)d\mu^{n+1}\\ &=\int_z \mathrm{P}\,[K_n(z, i_G^*; \{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n)>0]K_G(z, i_G^*)d\mu(z) .\end{aligned} \tag{5.12}$$

When $z\in A_{i_G^*}^G$ is fixed, it follows from (5.1), as $n\to\infty$, that

$$\lim_{n\to\infty} \mathrm{P}\,[K_n(z, i_G^*; \{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n)>0]=\mathrm{P}\,[K_G(z, i_G^*)>0] . \tag{5.13}$$

$$|\mathrm{P}\,[K_n(z, i_G^*; (z_{k,i_k^*}, i_k^*)_{k=0}^n)>0]K_G(z, i_G^*)|\leqq|K_G(z, i_G^*)| , \tag{5.14}$$

$$\begin{aligned}\int_z |K_G(z, i_G^*)|d\mu(z)&\leqq\int_z\int_\lambda |L(a_0, \lambda)-L(a_1, \lambda)|f(z\,|\,\lambda, i_G^*)dG(\lambda)d\mu(z)\\ &=\int_\lambda |L(a_0, \lambda)-L(a_1, \lambda)|dG(\lambda)<\infty .\end{aligned} \tag{5.15}$$

Then, from the Lebesgue dominated convergence theorem, using (5.12), (5.13), (5.14) and (5.15), follows the second term of the r.h.s. of (5.11). (5.10) and the theorem follow from (5.11) using Theorem 1 and the result above.

6. Monotone multiple decision problem

In the empirical Bayes context, the monotone multiple decision problem has been discussed, e.g. in Van Ryzin and Susarla [15]. Let now the action space consists of a finite number of distinct actions, say $A=\{a_0,\cdots,a_s\}$ and let $-\infty=\lambda_{-1}<\lambda_0<\cdots<\lambda_{s-1}<\lambda_s=\infty$. The decision corresponding to a_m is that '*the value of $\Lambda=\lambda$ is in the interval* $[\lambda_{m-1}, \lambda_m]$, $m=0,\cdots,s$'. The loss function $L(a_m, \lambda)$ is such that for m, $m=0,\cdots,s-1$,

$$(6.1)\qquad \begin{aligned} &L(a_{m+1}, \lambda)-L(a_m, \lambda)=d(\lambda_m-\lambda)\\ &L(a_0, \lambda)=\begin{cases} 0\,, & \text{when } \lambda\leqq\lambda_0\\ d\sum_{t=1}^{m}(\lambda-\lambda_{t-1})\,, & \text{when } \lambda_{m-1}<\lambda\leqq\lambda_m \end{cases}\end{aligned}$$

where d is a known positive constant. The loss function (6.1) is actually a generalized version of the loss function (2.4) used in two-action case, which is modified to cover multiple decision problem. Without loss of generality, we can simplify the loss function by taking $d=1$. The loss function (6.1) is monotone, because $L(a_{m+1}, \lambda)-L(a_m, \lambda)\geqq 0$, when $\lambda\leqq\lambda_m$, and $\leqq 0$, when $\lambda\geqq\lambda_m$, see Van Ryzin and Susarla [15].

The form of the decision rule is $\delta^i(z)=\{\delta^i(0|z),\cdots,\delta^i(s|z)\}$ where $\delta^i(m|z)=\Pr[\text{choose } a_m|z, i]$. The risk w.r.t. the prior distribution G may be given in the form

$$(6.2)\qquad W_G(\delta^i)=\sum_{m=0}^{s}\int\delta(m|z)\left\{\int L(a_m, \lambda)f(z|\lambda, i)dG(\lambda)\right\}d\mu(z)$$

which is minimized by defining $\delta(m|z)=\delta_G(m|z)$, $m=0,\cdots,s$, where $\delta_G(m|z)$ is the indicator function of the set

$$(6.3)\qquad A_{i,m}^G=\{z|\lambda_{m-1}<\mathrm{E}\{\Lambda|z, i\}\leqq\lambda_m\}$$

where

$$(6.4)\qquad \mathrm{E}\{\Lambda|z, i\}=\frac{\int\lambda f(z|\lambda, i)dG(\lambda)}{\int f(z|\lambda, i)dG(\lambda)}=\frac{f_i^G(z+1)/g_i(z+1)}{f_i^G(z)/g_i(z)}\,.$$

Accordingly, $\delta_G^i(z)=\{\delta_G^i(0|z),\cdots,\delta_G^i(s|z)\}$ is the Bayes rule w.r.t. the prior G.

Denote now

$$(6.5)\qquad K_G(z, i, m)=\int L(a_m, \lambda)f(z|\lambda, i)dG(\lambda)=\lambda_m-\frac{f_i^G(z+1)/g_i(z+1)}{f_i^G(z)/g_i(z)}\,.$$

We may now write

$$W_G(\delta_G^i)=\sum_{m=0}^{s}\int \delta_G^i(m\,|\,z)\left\{\int L(a_m,\lambda)f(z\,|\,\lambda,i)dG(\lambda)\right\}d\mu(z) \tag{6.6}$$
$$=\sum_{m=0}^{s}\int_{A_{i,m}^G}\left\{\int L(a_m,\lambda)f(z\,|\,\lambda,i)dG(\lambda)\right\}d\mu(z)$$
$$=\sum_{m=0}^{s}\int_{A_{i,m}^G} K_G(z,i,m)d\mu(z)$$
$$=\sum_{m=0}^{s}\int_{A_{i,m}^G}\{\lambda_m f_i^G(z)-g_i(z)f_i^G(z+1)/g_i(z+1)\}dz$$
$$=\sum_{m=0}^{s}\int_{A_{i,m}^G}\varphi_G(z,i,m)dz\ .$$

The fixed optimal sample size of the classical Bayes solution is based on the function

$$h_{G,i}=W_G(\delta_G^{i+1})-W_G(\delta_G^i)+c \tag{6.7}$$
$$=\sum_{m=0}^{s}\int_{A_{i+1,m}^G}\varphi_G(z,i+1,m)dz-\sum_{m=0}^{s}\int_{A_{i,m}^G}\varphi_G(z,i,m)dz+c$$

such that

$$i_G^*=\inf\{i\,|\,h_{G,i}\geqq 0\}\ . \tag{6.8}$$

The Bayes rule $\delta_G^{i_G^*}$ with optimal sample size i_G^* has the Bayes risk

$$W_G(\delta_G^{i_G^*})=\sum_{m=0}^{s}\int_{A_{i_G^*,m}^G}\varphi_G(z,i_G^*,m)dz\ . \tag{6.9}$$

In the empirical Bayes solution, we may again use the estimators $f_j^n(z)$ defined by (4.1) and (4.3), and analogously to Section 4 we get the empirical Bayes quantities

$$K_n(z,i,m)=\lambda_m-\frac{f_i^n(z+1)/g_i(z+1)}{f_i^n(z)/g_i(z)} \tag{6.10}$$

$$A_{i,m}^n=\left\{z\,|\,\lambda_{m-1}<\frac{f_i^n(z+1)/g_i(z+1)}{f_i^n(z)/g_i(z)}\leqq\lambda_m\right\} \tag{6.11}$$

$$\varphi_n(z,i,m)=\lambda_m f_i^n(z)-g_i(z)f_i^n(z)/g_i(z+1) \tag{6.12}$$

$$h_{n,i}=\sum_{m=0}^{s}\int_{A_{i+1,m}^n}\varphi_n(z,i+1,m)dz-\sum_{m=0}^{s}\int_{A_{i,m}^n}\varphi_n(z,i,m)dz+c \tag{6.13}$$

$$i_{n+1}^*=\inf\{i\,|\,h_{n,i}\geqq 0\}\ . \tag{6.14}$$

At the stage $n+1$, the actual Bayes risk of the empirical Bayes rule $\delta_n^{i_{n+1}^*}$, evaluated conditionally for fixed past $\{(z_{k,i_k^*}, i_k^*)\}_{k=0}^n$ and i_{n+1}^*, and regarding G as known, is

$$
\begin{aligned}
(6.15)\qquad W_G(\delta_n^{i_{n+1}^*}) &= \sum_{m=0}^{s} \int \delta_n^{i_{n+1}^*}(m\,|\,z)\left\{\int L(a_m, \lambda)f(z\,|\,\lambda, i_{n+1}^*)dG(\lambda)\right\}d\mu(z) \\
&= \sum_{m=0}^{s} \int_{A_{i_{n+1}^*,m}^n} K_G(z, i_{n+1}^*, m)d\mu(z) \\
&= \sum_{m=0}^{s} \int_{A_{i_{n+1}^*,m}^n} \varphi_G(z, i_{n+1}^*, m)dz\ .
\end{aligned}
$$

The corresponding asymptotic properties of Section 5, where the two action rules were considered, may be shown to cover also the multiple monotone case. The proofs for the theorems of the present case are parallel, but more complicated, and they are not repeated here.

7. Discussion

In the present paper we have tried to keep our approach as non-parametric as possible. The approach has been chosen with purpose, because the theory in the present form gives possibilities to a wide selection of the appropriate probability mass or density estimator $f_j^n(z)$. In the text, we gave only one estimator for discrete and one for continuous case, and both of them are such that they are applicable to any situation where the probability mass or density function may be given or reparametrized to the form (2.2). In principle, there are two tendencies in empirical Bayes estimation of $f_j^G(z)$. First, one might use a general estimator, as we have done. Second, one might use an estimator which takes into account of the particular family, to which the conditional distribution belongs. As an example of the second tendency, we refer to Martz and Lian [4] where several empirical Bayes estimators for binomial parameter are presented, and their properties studied. It should, however, be noted that all the estimators given in empirical Bayes literature, do not allow varying sample sizes.

When we consider the criterion for optimizing the sample size, it has to be realized that it is based on look-ahead idea, and particularly in our case, on one-step-look-adead idea. The procedure, however, is easily generalized to an m-step-look-ahead-procedure. The presentation is based on one-step procedure, because it is easier to handle. Furthermore, we are aware that look-ahead procedures often fail to give optimal solution, that is one might find such priors that the rule stops the sampling too early in classical Bayes solution. So there might be situations where the total costs, including the Bayes risks and the observation costs, are lower than in our classical Bayes solution. However, it has been said that the look-ahead procedures in any case are a satisfactory approximation, and they also have computational advantages. Look-ahead procedures have been discussed, e.g. in Berger [1].

The modification that successively after a fixed period maximum sample size is taken might also be used to control, that the prior distribution actually remains the same as it was at the beginning. Furthermore, the modification should improve the convergence properties, because under some certain priors, the optimization rule in the empirical Bayes situation tends at early stages to give too small sample sizes.

Some particular exponential conditional distributions, having the form (2.2), are summarized in Table 1, where transformations, if needed, and the functions $g_i(z)$ and $\{h(\lambda)\}^i$ are displayed.

Table 1. The functions $g_i(\lambda)$ and $\{h(\lambda)\}^i$ and possible transformations for some exponential conditional distributions

Distribution	Transformation	$g_i(z)$	$\{h(\lambda)\}^i$
Binomial r fixed	*not needed*	$\binom{ir}{z}$	$(1-\lambda)^{ir-z}$
Poisson	*not needed*	$i^z/i!$	$e^{-i\lambda}$
Negative binomial r fixed	*not needed*	$\binom{ir+z-1}{z}$	$(1-\lambda)^{ir}$
Normal σ_0^2 fixed	$\lambda=\exp(\theta/\sigma_0^2)$	$\exp(-z^2/2i\sigma_0^2)$	$\lambda^{-i\sigma_0^2 \log \lambda/2}$
Gamma α fixed	$\lambda=\exp(-\theta)$	$z^{i\alpha-1}$	$(-\log \lambda)^{i\alpha}$

Acknowledgement

This article is dedicated to the memory of Professor Gustav Elfving (1908–1984) with deepest appreciation.

UNIVERSITY OF TURKU

REFERENCES

[1] Berger, J. O. (1980). *Statistical Decision Theory Foundations, Concepts, Methods*, Springer, New York.

[2] Laippala, P. (1979). The empirical Bayes approach with floating optimal sample size in binomial experimentation, *Scand. J. Statist.*, **6**, 113-338, correction note *Scand. J. Statist.*, **7**, 105.

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

[4] Martz, H. F. and Lian, M. G. (1974). Empirical Bayes estimation of binomial parameter, *Biometrika*, **61**, 517-523.

[5] 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.

[6] O'Bryan, T. (1979). Rates of convergence in a modified empirical Bayes estimation involving Poisson distributions, *Comm. Statist.*, A8, 167-174.
[7] O'Bryan, T. and Susarla, V. (1977). Empirical Bayes estimation with non-identical components, Continuous case, *Aust. J. Statist.*, **19**, 115-125.
[8] Parzen, E. (1962). On estimation of a probability density function and mode, *Ann. Math. Statist.*, **33**, 1065-1076.
[9] Robbins, H. (1956). The empirical Bayes approach to statistics, *Proc. 3rd Berkeley Symp. Math. Statist. Prob.*, **1**, 157-163, University of California Press.
[10] Robbins, H. (1963). The empirical Bayes approach to testing statistical hypotheses, *Rev. Inst. Int. Statist.*, **35**, 195-208.
[11] Robbins, H. (1964). The empirical Bayes approach to statistical decision problems, *Ann. Math. Statist.*, **35**, 1-20.
[12] Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function, *Ann. Math. Statist.*, **27**, 832-837.
[13] Samuel, E. (1963). An empirical Bayes approach to the testing of certain parametric hypotheses, *Ann. Math. Statist.*, **34**, 1370-1384.
[14] Suzuki, Y. (1975). On empirical Bayes models, *Proc. 40th ISI Session*, 372-385.
[15] Van Ryzin, J. and Susarla, V. (1977). On the empirical Bayes approach to multiple decision problems, *Ann. Statist.*, **5**, 172-182.