한국통계학회논문집 제2권 1호
1995년 4월 pp 145-154

Empirical Bayes Problems with Dependent and Nonidentical Components[1)]

Inha Jung[2)], Jee Chang Hong[2)] , Kang Sup Lee[3)]

Abstract

Empirical Bayes approach is applied to estimation of the binomial parameter when there is a cost for observations. Both the sample size and the decision rule for estimating the parameter are determined stochastically by the data, making the result more useful in applications.

Our empirical Bayes problems with non-iid components are compared to the usual empirical Bayes problems with iid components. The asymptotic optimal procedure with a computer simulation is given.

1. Introduction

Suppose that the rate θ at which defectives are produced by a given production process varies from day-to-day. On each day a random sample of at least two parts is taken at a cost of \$.50 per part and an estimate $\hat{\theta}$ made with loss \$100$(\hat{\theta}-\theta)^2$. If the sequence $\theta_1, \theta_2, \cdots$ is modeled as a stochastic sequence with independent and identically G-distributed variables with G unknown, then the empirical Bayes method is appropriate. When G is restricted to the Beta(α, β) family and the sampling is two-at-a-time, we show how to construct a decision procedure with risk plus cost for observations converging to the lowest possible risk, whatever be α and β. In Section 3 we find that in this case the envelope risk plus cost is no greater than \$18.00 per day, the minimax risk plus cost. Against the least favorable $\alpha=\beta=2$, the empirical Bayes risk is estimated to be below \$20.00 after 15 days. The empirical Bayes sample size converges to the optimal $8\times2=16$ parts here. Other α, β values are tested in the computational work of Section 3. In this section and the next we dovelop the empirical Bayes procedure and prove its asymptotic optimality.

1) Research of the first author was supported by Korea Science and Engineering Foundation Grant(KOSEF 901-0105-024-1)
2) Department of Mathematics, Ajou University, Suwon, 441-749, KOREA.
3) Department of Computer Sciences and Statistics, Dankook University, Seoul, 140-714, KOREA.

Let $X_1, X_2, \cdots$ be i.i.d. B(m, θ), where m is a given positive integer and the parameter θ has prior distribution G in the beta family $\{\text{Beta}(\alpha, \beta) | \alpha > 0, \beta > 0\}$. Estimation of θ is considered under the squared-error loss. Here parameter space Θ and action space A are $[0,1]$. Let $c>0$ be a constant cost per observation.

Let $d \in D_n$ be a decision rule based on the observation $\underline{X}^n = (X_1, \cdots, X_n)$. The decision loss plus cost for observation is given by $[\theta - d(\underline{X}^n)]^2 + cn$. Let R_n denote the risk and Bays risk of the decision rule $d \in D_n$, i.e.,

$$R_n(\theta, d) = E_\theta[\theta - d(\underline{X}^n)]^2 \tag{1.1}$$

$$R_n(G, d) = E_G R_n(\theta, d) \tag{1.2}$$

and r_n denote the risk and Bayes risk of the decision rule $d \in D_n$ including cost for observations. Then

$$r_n(\theta, d) = R_n(\theta, d) + cn \ , \tag{1,3}$$

$$r_n(G, d) = R_n(G, d) + cn \ . \tag{1.4}$$

For each G and $n = 1, 2, \cdots$ let $d_G^n \in D_n$ be a Bayes rule. Thus

$$\inf_{d \in D_n} R_n(G, d) = R_n(G, d_G^n) \ . \tag{1.5}$$

Let $R_n(G) = R_n(G, d_G^n)$ and

$$r_n(G) = R_n(G) + cn \ . \tag{1.6}$$

For $G = \text{Beta}(\alpha, \beta)$ we let ξ and η denote the first two moments, that is,

$$\xi = E_G \theta = \frac{\alpha}{\alpha + \beta}$$

$$\eta = E_G \theta^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \ , \tag{1.7}$$

and note that $0<\xi^2<\eta<\xi<1$ since $\alpha>0,\ \beta>0$. Also

$$EX_i = m\xi$$
$$EX_i^2 = m(\xi-\eta)+m^2\eta\ , \tag{1.8}$$

and from (1.7), it follows that

$$\alpha=\frac{\xi(\xi-\eta)}{\eta-\xi^2}\ ,\qquad \beta=\frac{(1-\xi)(\xi-\eta)}{\eta-\xi^2}\ . \tag{1.9}$$

In the empirical Bayes application, (1.8) and (1.9) will be useful in the construction of consistent estimates for α and β. The posterior distribution of θ, given $\underline{X}^n$, is Beta$(\alpha+n\overline{X}_n,\ \beta+mn-n\overline{X}_n)$. Hence a Bayes rule $d_G^n\in D_n$ is

$$d_G(\underline{X}^n)=\frac{\alpha}{\alpha+\beta+mn}+\frac{n}{\alpha+\beta+mn}\overline{X}_n \tag{1.10}$$

if $G=$Beta$(\alpha,\ \beta)$.

Let $G=$Beta$(\alpha,\ \beta)$, $G'=$ Beta(α',β'). It can be shown that

$$R_n(G,d_{G'})=\frac{1}{(\alpha'+\beta'+mn)^2}\{[\ (\alpha'+\beta')^2-mn]\ \eta-[\ 2\alpha'(\alpha'+\beta')-mn]\ \xi+(\alpha')^2\}\ , \tag{1.11}$$

$$|R_n(G,d_{G'})-R_n(G',d_{G'})|\le 2|\xi-\xi'|+|\eta-\eta'| \tag{1.12}$$

and

$$R_n(G)=\frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)(\alpha+\beta+mn)} \tag{1.13}$$

From (1.13), the minimum Bayes risk including cost for observation is

$$r_n(G)=\frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}(\alpha+\beta+mn)^{-1}+cn \tag{1.14}$$

We seek the optimal sample size n^*, a minimizer of $r_n(G)$ among $n=1,2,\cdots$. $r_n(G)$ is a continuous and convex function of real $n>\frac{-(\alpha+\beta)}{m}$.

Consider the equation

$$0=\frac{d}{dn}\, r_n(G)=-\frac{m\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}(\alpha+\beta+mn)^{-2}+c$$

Its larger solution is

$$\nu=\left\{\left(\frac{m}{c}\,\frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}\right)^{\frac{1}{2}}-(\alpha+\beta)\right\}\Big/ m \qquad (1.15)$$

and an optimal fixed sample size $n^*=n^*(\alpha,\ \beta)$ is given by

$$n^*=\begin{cases} 1 & ,\ \text{if } \nu<1 \\ \nu & ,\ \text{if } \nu\in\{1,2,3,\ldots\} \\ [\ \nu]\ \text{ or } [\ \nu]\ +1 & \\ \text{depending on which integer} & ,\ \text{otherwise} \\ \text{minimizes } r_n(G) & \end{cases} \qquad (1.16)$$

Here [ ] denotes the greatest integer function and we take $n^*=[\ \nu]$ if both $[\ \nu]$ and $[\ \nu]\ +1$ minimize $r_n(G)$.

Since $R_1(G)\le .25$ for all G, it follows that $n^*\le(.25+c)/c$ for all G. If $\alpha,\ \beta$ were known constants, we can use $d_G\in D_{n^*}$ to achieve minimum Bayes risk, i.e.,

$$r(G)=\min\{r_n(G)\,|\,n=1,\ 2,\cdots\}$$

In section 2, we show how $(\alpha,\ \beta)$ is estimated in the empirical Bayes problem with this component and establish the asymptotic optimality for the resulting procedure.

In section 3, we give the results of computer simulations that provide estimates of risk behavor for small to moderate number of component problems.

2. An Empirical Bayes Decision Procedure

Consider the binomial component problem of the last section. Let $\hat{\alpha}_0,\ \hat{\beta}_0$ be initial nonrandom estimates of $\alpha,\ \beta$ and the $N_1=n^*(\hat{\alpha}_0,\ \hat{\beta}_0)$ be the sample size chosen for the first component. (See(1.16) for the definition of the optimal fixed sample size function n^*.) Let

$\underline{X}^1 = (X_{11}, X_{12}, \ldots, X_{1N_1})$ denote the vector of observations from the first component.

We will define a sequence of estimates $\hat{\alpha}_i$, $\hat{\beta}_i$ based on $(\underline{X}^1, \underline{X}^2, \ldots, \underline{X}^i)$. Then for component $i+1$, the empirical Bayes sample size is $N_{i+1} = n^*(\hat{\alpha}_i, \hat{\beta}_i)$ and the empirical Bayes estimator of θ_{i+1} is

$$d_{i+1}(\underline{X}^{i+1}) = \frac{\hat{\alpha}_i + N_{i+1} Y_{i+1}}{\hat{\alpha}_i + \hat{\beta}_i + mN_{i+1}}, \quad i = 0, 1, \ldots \tag{2.1}$$

(see (1.10)), where

$$Y_i = \frac{1}{N_i} \sum_{j=1}^{N_i} X_{ij}, \quad i = 1, 2, \ldots \tag{2.2}$$

We will give estimates based on the method of moments and will find it useful to consider

$$Z_i = \frac{1}{N_i} \sum_{j=1}^{N_i} X_{ij}^2, \quad i = 1, 2, \ldots \tag{2.3}$$

and denote average of Y_j, Z_j, $j = 1, 2, \ldots, i$ as $\overline{Y}_i$, $\overline{Z}_i$, $i = 1, 2, \ldots$.

Let $\mathcal{F}_0$ be the trivial σ-field and let $\mathcal{F}_j = \sigma(\underline{X}^1, \underline{X}^2, \ldots, \underline{X}^j)$, $j = 1, 2, \ldots$. The sample size N_j is $\mathcal{F}_{j-1}$ measurable, $j = 1, 2, \ldots$, and we see that

$$\begin{aligned} E(Y_j | \mathcal{F}_{j-1}) &= m\xi, & j &= 1, 2, \ldots \\ E(Z_j | \mathcal{F}_{j-1}) &= m(\xi - \eta) + m^2 \eta, & j &= 1, 2, \ldots \end{aligned} \tag{2.4}$$

follow from(1.8).

Since $Y_j \le m$ and $Z_j \le m^2$, $j = 1, 2, \ldots$, the strong law for centerings at conditional expectation(see Hall and Heyde(1980, Theorem 2.19)) implies

$$\begin{aligned} \overline{Y}_i - \frac{1}{i} \sum_{j=1}^{i} E(Y_j | \mathcal{F}_{j-1}) &\to 0 \text{ a.s.} \\ \overline{Z}_i - \frac{1}{i} \sum_{j=1}^{i} E(Z_j | \mathcal{F}_{j-1}) &\to 0 \text{ a.s.} \end{aligned} \tag{2.5}$$

From (2.4) and (2.5) we have

$$\overline{Y}_i \rightarrow m\xi \qquad \text{a.s.}$$
$$\overline{Z}_i \rightarrow m(\xi-\eta)+m^2\eta \qquad \text{a.s.} \tag{2.6}$$

Lemma 2.1. Let $m\geq 2$. The estimators defined for $i=1,2,\ldots$ by

$$\hat{\xi}_i \equiv \frac{\overline{Y}_i}{m}$$
$$\hat{\eta}_i \equiv \frac{\overline{Z}_i-\overline{Y}_i}{m(m-1)} \tag{2.7}$$

and

$$\hat{\alpha}_i \equiv \left[\frac{\hat{\xi}_i(\hat{\xi}_i-\hat{\eta}_i)}{\hat{\eta}_i-\hat{\xi}_i^2}\right]^+$$
$$\hat{\beta}_i \equiv \left[\frac{(1-\hat{\xi}_i)(\hat{\xi}_i-\hat{\eta}_i)}{\hat{\eta}_i-\hat{\xi}_i^2}\right]^+ \tag{2.8}$$

are a.s. consistent. (In (2.8), take ratios 0/0 to be 0.)

Proof. The a.s. convergence of the estimates (2.7) follows from (2.8). The a.s. convergence of the estimates (2.8) follows from (1.9).

Let the sample size sequence $\underline{N}=(N_1, N_2, \ldots)$ be defined by $N_{i+1}=n^*(\hat{\alpha}_i, \hat{\beta}_i)$, $i=0,1,2,\ldots$, where n^* is defined by (1.16). Let the empirical Bayes decision rule $\underline{d}=(d_1, d_2, \ldots)$ be defined by (2.1) and (2.2). The following lemma is used to establish the asymptotic optimality of our empirical Bayes procedure $(\underline{N}, \underline{d})$.

Lemma 2.2. For priors ω and ν, let $n=n^*(\omega)$, $m=n^*(\nu)$ be optimal fixed sample sizes and let $d_\omega^k, d_\nu^k \in D_k$ denote Bayes decision rules with respect to ω, ν for $k=1,2,\ldots$. Then

$$0 \leq r_m(\omega, d_\nu^m)-r(\omega)$$
$$\leq \sup_k |R_k(\omega, d_\nu^k)-R_k(\nu, d_\nu^k)| + \sup_k |R_k(\omega, d_\omega^k)-R_k(\nu, d_\omega^k)| \tag{2.9}$$

Proof. This is an immediate consequence of the well-known triangle inequality.

Theorem 2.3. Let $m\geq 2$. The empirical Bayes procedure $(\underline{N}, \underline{d})$ defined above is asymptotically optimal at each $G=(\alpha, \beta)$.

Proof. By lemma 2.1 and (2.6),

$$0 \leq r_{N_{i+1}}(G, d_{i+1}) - r(G) \leq 4|\hat{\xi}_i - \xi| + 2|\hat{\eta}_i - \eta| \tag{2.10}$$

Since $|\hat{\xi}_i - \xi| \leq 1$ and $|\hat{\eta}_i - \eta| \leq 2$ for all i, the Dominated Convergence Theorem and (2.9) in lemma 2.2 imply that $Er_{N_{i+1}}(G, d_{i+1}) \rightarrow r(G)$.

3. Some Empirical Bayes Risk Calculations

In this section we treat the empirical Bayes problem of the last section. All risks are multiplied by 1000, which corresponds to a component with loss function $1000(a-\theta)^2$ and cost $1000c$ per observation.

We calculate the envelope risk $r(\alpha, \beta)$ and the optimal sample size(s) for various m, c, α, and β and present some of the results in Table 1. We include the mean and standard deviation of the Beta(α, β) prior in each case.

Figure 1. below is a graph of the envelope risk function $r(\alpha, \alpha)$ plotted against α on a log scale. For this we choose $m=2$ and $c=.001$.

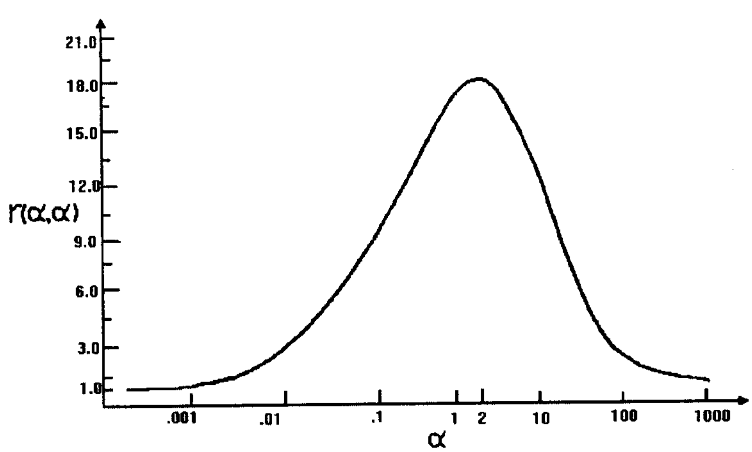

Figure 1. A Risk Envelope

Table 1. $n^*(\alpha, \beta)$ and $r(\alpha, \beta)$

Prior				$m=2$				$m=3$			
				$c=.001$		$c=.002$		$c=.001$		$c=.002$	
α	β	μ	σ	n^*	r	n^*	r	n^*	r	n^*	r
0.1	0.1	0.50	0.456	4	9.081	3	12.720	4	7.415	3	10.529
0.1	0.3	0.25	0.366	5	10.151	3	14.371	4	8.320	3	11.699
0.1	0.9	0.10	0.212	4	9.000	3	12.429	4	7.462	2	10.429
0.1	1.9	0.05	0.126	3	6.958	2	9.278	3	5.879	2	7.958
0.2	0.2	0.50	0.423	6	11.760	4	16.503	5	9.638	3	10.529
0.2	0.6	0.25	0.323	6	12.510	4	17.470	3	10.274	3	11.699
0.2	1.2	0.14	0.226	5	11.266	4	15.599	4	9.330	2	10.429
0.2	1.8	0.10	0.173	4	10.000	3	13.500	4	8.286	3	11.455
0.3	0.3	0.50	0.395	7	13.421	5	18.844	5	11.010	4	15.440
0.3	0.6	0.33	0.342	7	14.065	5	19.657	6	11.569	4	16.160
0.3	1.2	0.20	0.253	6	13.111	4	18.105	5	10.818	4	15.111
0.3	1.8	0.14	0.199	5	11.855	4	16.213	5	9.851	3	13.473
0.5	0.5	0.50	0.354	7	15.333	5	21.364	6	12.579	4	17.615
0.5	1.0	0.33	0.298	7	15.602	5	21.594	6	12.838	4	17.877
0.5	1.5	0.25	0.250	7	14.812	5	20.417	6	12.250	4	16.929
1.0	1.0	0.50	0.289	8	17.259	5	23.889	7	14.246	5	19.804
1.0	1.5	0.40	0.262	8	17.266	5	23.714	7	14.295	5	19.796
1.0	2.0	0.33	0.236	8	16.772	5	22.821	6	13.937	4	19.111
1.5	1.5	0.50	0.250	8	17.868	5	24.423	7	14.813	5	20.417
1.5	2.0	0.43	0.233	8	17.768	5	24.109	7	14.775	4	20.289
2.0	2.0	0.50	0.224	8	18.000	5	24.286	7	15.000	4	20.500
3.0	3.0	0.50	0.189	7	17.714	4	23.306	6	14.929	4	19.905
4.0	4.0	0.50	0.167	7	17.101	3	21.873	6	14.547	3	19.072
5.0	5.0	0.50	0.151	6	16.331	3	20.205	5	14.091	3	17.962
10.0	10.0	0.50	0.109	1	11.823	1	12.823	2	11.158	1	12.352

For $m=2$, $c=.001$ and selected α, β values, we obtain Monte Carlo estimates of the empirical Bayes risk of our procedure with inital starting estimates $\hat{\alpha}_0=\hat{\beta}_0=1$. This is done for stages $i=10, 15, 20, 25, 50$ and 100 and the result are presented in Table 2 along with the standard errors of the estimates.

Table 2. Estimated Empirical Bayes Risks $(m=2k,\ c=.001)$

		Estimated Empirical Bayes Risks(Standard Errors)						
α	β	10	15	20	25	50	100	Envelope Risk
0.1	0.1	10.22 (0.18)	9.83 (0.07)	10.13 (0.14)	10.00 (0.14)	9.28 (0.05)	9.13 (0.01)	9.081
0.5	0.5	17.31 (0.67)	15.97 (0.10)	15.68 (0.05)	15.56 (0.03)	15.40 (0.01)	15.37 (0.00)	15.333
1.0	1.0	21.27 (0.73)	19.05 (0.43)	18.26 (0.25)	18.05 (0.28)	17.41 (0.02)	17.32 (0.00)	17.259

<표계속>

α	β	Estimated Empirical Bayes Risks(Standard Errors) 10	15	20	25	50	100	Envelope Risk
2.0	2.0	21.26 (0.43)	19.67 (0.25)	19.89 (0.30)	19.44 (0.25)	19.09 (0.20)	18.27 (0.04)	18.000
3.0	3.0	20.43 (0.28)	19.73 (0.24)	19.36 (0.21)	19.75 (0.25)	18.73 (0.14)	18.47 (0.17)	17.714
4.0	4.0	19.98 (0.29)	19.34 (0.19)	19.05 (0.16)	18.95 (0.16)	18.66 (0.15)	18.10 (0.12)	17.101
0.1	0.9	12.25 (0.27)	12.58 (0.34)	13.12 (0.42)	13.05 (0.44)	10.69 (0.31)	9.41 (0.31)	9.000
0.2	1.8	12.79 (0.19)	13.34 (0.24)	13.24 (0.29)	13.28 (0.29)	12.38 (0.28)	10.86 (0.17)	10.000

4. Concluding Remarks

In discussing the empirical Bayes decision problems, problems with independent and identical components were usually considered. (Gilliland, Dennis and Hannan, James(1977), Johns, M.V. and Van Ryzin(1972), Morris, Carl(1985), Robbins, H(1951), Robbins, H(1956)). O′ Bryan(1972,1976) considered a variant of the standard empirical Bayes problem where the sequence of component problems are not identical in that the sample size may vary with the component problem. In O′ Bryan(1972,1976), sample sizes are given nonrandom numbers and component problems are independent. In our empirical Bayes approach data accumulated over past component problems are used in selecting both sample size and decision rule to be used in the current component problem. The component problems are neither independent nor identical. Our method of estimation in the empirical Bayes version requires that $m \geq 2$. This assumption can be removed if we require $N_i \geq 2$ and use the estimators based on the pooled data. Requiring that $N_i \geq 2$ i.o. would suffice.

References

[1] Gilliland, D. and Hannan, J.(1977). Improved rates in the empirical Bayes monotone multiple decision problem with MLR family, *Annals of Statistics*, Vol.5, 516-521.

[2] Hall, P. and Heyde, C.C.(1980). *Martingale limit theory and its applications,* Academic Press, New York.

[3] Johns, M.V. and Van Ryzin,J.(1972). Convergence rates for empirical Bayes two-action problems II. Continuous case, *Annals of Mathematical Statistics,* Vol.43, 934-947.

[4] Morris, C.(1983). Parametric Bayes inference : Theory and Applications, *Journal of the*

American Statistical Association, Vol.78, 47-65.
[5] O'Bryan, T.(1972). *Empirical Bayes results in the case of non-identical components,* Ph.D. Thesis, RM-306, Statistics and Probability, Michigan State University.
[6] O'Bryan, T.(1976). Some empirical Bayes results in the case of component problem with varying sample sizes for discrete exponential families, *Annals of Statistics,* Vol.4, 1290-1293.
[7] Robbins, H.(1951). Asymptotically subminimax solutions of compound statistical decision problems, *Proceedings of 2nd Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 131-148.
[8] Robbins, H.(1956). The empirical Bayes approach to statistics, *Proceedings of 3rd Berkeley Symposium on Mathematical Statistics and Probability,* Berkeley and Los Angeles, University of California Press, Vol.1, 157-163.