응용통계연구 제 6권 1호
1993년 3월 pp. 173 - 181

Nonparametric Empirical Bayes Estimation of a Distribution Function with respect to Dirichlet Process Prior in Case of the Non-identical Components[3)]

Inha Jung[4)]

Abstract

Nonparametric empirical Bayes estimation of a distribution function with respect to Dirichlet process prior is considered when sample sizes are varying from component to component.

Zehnwirth's estimate of $\alpha(\mathbf{R})$ is modified to be used in our empirical Bayes problem with non-identical components.

1. Introduction

In nonparametric ststistical decision problem with respect to Dirichlet process prior $D(\alpha)$ originated from Ferguson(1973), parameter space is the set of all probability measures P on a measurable space (X,A). Statistician chooses an action a and thereby incurring loss $L(P,a)$. Infimum Bayes risk is denoted by $R(\alpha)$. Even if α is unknown, one can construct a statistical procedure based on the data gathered from n independent repetitions of the decision problem for which the risk converges to $R(\alpha)$ as $n\rightarrow\infty$ for all α. Only sequences of identical components have been treated in the nonparametric decision problems. However, it is clear that when the only difference from component to component is sample size, empirical Bayes methods should still be useful. In this case there is not a single Bayes envelope $R(\alpha)$ but rather a sequence of envelopes $R^{k_n}(\alpha)$, where k_n is the sample size in the n-th component problem.

3) This Research wes supported by the Research Foundation of Ajou University 1991.

4) Department of Mathematics, Ajou University, Suwon, 441-749, Korea.

Let $P_1, P_2, \cdots$ be a sequence of i.i.d. random probability measures on $(\mathbb{X}, \mathbb{A})$. from a prior distribution given by Dirichlet process $D(\alpha)$ and let $\underline{X_i} = (X_{i1}, \cdots, X_{ik_i})$ be a random sample of size k_i from P_i, $i = 1, 2, \cdots$. For the decision on P_{n+1}, we seek a decision procedure t_{n+1} based on $\underline{X_1}, \cdots, \underline{X}_{n+1}$ such that

$$\lim_n [R^{k_n}(t_n, \alpha) - R^{k_n}(\alpha)] = 0 \text{ for all } \alpha(\cdot), \tag{1.1}$$

where $R^{k_n}(t_n, \alpha)$ denotes the Bayes risk of the rule t_n based on k_n observations.

Definition 1.1. A statistical decision procedure $\{t_n\}$ is said to be asymptotically optimal (a.o.) if it satisfies (1.1) for all $\alpha(\cdot)$.

Let $\mathbb{X} = \mathbf{R}$ = real line, $\mathbb{A} = \mathbb{B}$ = all Borel sets in $\mathbf{R}$. We consider the problem of estimating $F_{n+1}(t) = P_{n+1}((-\infty, t])$ using $\underline{X_1}, \cdots, \underline{X}_{n+1}$ under the loss function

$$L(F_{n+1}, \hat{F}) = \int_{\mathbf{R}} (F_{n+1}(t) - \hat{F}(t))^2 dW(t), \tag{1.2}$$

where $W(\cdot)$ is a given finite measure on $(\mathbf{R}, \mathbb{B})$ and $\hat{F}$ is an estimate of F_{n+1}. Parameter space and action space are the set of all distributions on $(\mathbf{R}, \mathbb{B})$. Let $\hat{F}_i$ denote the empirical distribution function determined by the observations from the i-th component $\underline{X_i} = (X_{i1}, \cdots, X_{ik_i})$, $i = 1, 2, \cdots, n+1$. Under the loss function given by (1.2), Bayes estimate of F_{n+1} for the no-sample problem is $F_0(t) = EF_{n+1}(t)$. Bayes estimate of F_{n+1} based on the observation $\underline{X}_{n+1} = (X_{n+1,1}, \cdots, X_{n+1,k_{n+1}})$ is $\tilde{F}_{n+1}(t) = E[F_{n+1}(t) | \underline{X}_{n+1}]$. Since $F_{n+1}(t)$ has distribution Beta $(\alpha((-\infty, t]), \alpha((t, \infty)))$ and posterior distribution of $F_{n+1}(t)$ given $\underline{X}_{n+1}$ is provided by $D(\alpha + \sum_{i=1}^{k_{n+1}} \delta_{X_{n+1,i}})$ with δ_x denoting the unit mass at x, we see that

$$F_0(t) = \frac{\alpha((-\infty, t])}{\alpha(\mathbf{R})}, \tag{1.3}$$

$$\tilde{F}_{n+1}(t) = (1 - p_{k_{n+1}}) F_0(t) + p_{k_{n+1}} \hat{F}_{n+1}(t), \tag{1.4}$$

where

$$p_{k_{n+1}} = \frac{k_{n+1}}{k_{n+1} + \alpha(\mathbf{R})}. \tag{1.5}$$

The Bayes risk of $\tilde{F}_{n+1}(t)$ is given by

$$\begin{aligned} R^{k_{n+1}}(t) &= E\int_{\mathbf{R}} (\tilde{F}_{n+1}(t) - F_{n+1}(t))^2 dW(t) \\ &= \int_{\mathbf{R}} E(\tilde{F}_{n+1}(t) - F_{n+1}(t))^2 dW(t) \end{aligned} \tag{1.6}$$

Define for $n = 1, 2, \cdots$ the sequence of estimates H_{n+1} by

$$H_{n+1}(t) = (1 - \hat{p}_{k_{n+1,n+1}}) \hat{F}_{0n}(t) + \hat{p}_{k_{n+1,n+1}} \hat{F}_{n+1}(t), \tag{1.7}$$

where

$$\hat{F}_{0n}(t) = \frac{\sum_{i=1}^{n} \hat{F}_{i(t)}}{n} \tag{1.8}$$

$$\hat{p}_{k_{n+1,n+1}} = \frac{k_{n+1}}{k_{n+1} + \hat{\alpha}_{n(\mathbf{R})}} \tag{1.9}$$

and $\hat{\alpha}_n(\mathbf{R})$ is an estimate of $\alpha(\mathbf{R})$ based on $X_1, \cdots, X_n$. The Bayes risk of H_{n+1} is given by

$$\begin{aligned} R^{k_{n+1}}(H_{n+1}, \alpha) &= E\int_{\mathbf{R}} (H_{n+1}(t) - F_{n+1}(t))^2 dW(t) \\ &= \int_{\mathbf{R}} E(H_{n+1}(t) - F_{n+1}(t))^2 dW(t). \end{aligned} \tag{1.10}$$

Empirical Bayes decision problem for estimating the distributuon function F_{n+1} has been considered by Korwar and Hollander(1976) under the loss function given by (1.2) given value of $\alpha(\mathbf{R})$. The value $\alpha(\mathbf{R})$ is interpreted as the "prior sample size". Therefore, the work of Korwar and Hollander is not fully empirical Bayes in this point of view. Zehnwirth(1981) has given a consistent estimate of $\alpha(\mathbf{R})$ in the one-way ANOVA setup and used it in his asymptotically optimal empirical Bayes estimate of F_{n+1}. We modify in section 3 Zehnwirth's estimate of $\alpha(\mathbf{R})$ for our empirical Bayes problem of unequal sample size components.

2. Asymptotic optimality of $\{H_{n+1}\}$

Let $\hat{\alpha}_n(\mathbf{R})$ denote a consistent estimate of $\alpha(\mathbf{R})$. i.e. $\hat{\alpha}_n(\mathbf{R}) \xrightarrow{Pr} \alpha(\mathbf{R})$ as $n \to \infty$. The following lemmas are useful in proving asymptotic optimality of $\{H_{n+1}\}$ given by (1.7).

Lemma 2.1.(Zehnwirth,1981) Let F_0, $\hat{F}_{0n}$ be given by (1.3), (1.8) and let $\hat{F}_i$ be the empirical distribution determined by $X_i = (X_{i1}, \cdots, X_{ik_i})$. Then,

$$E(\hat{F}_i(t)|F_i) = F_i(t) \tag{2.1}$$

$$Var(\hat{F}_i(t)|F_i) = \frac{F_i(t)(1-F_i(t))}{k_i} \tag{2.2}$$

$$EF_i(t)(1-F_i(t)) = \frac{\alpha((-\infty,t])\alpha((t,\infty))}{\alpha(\mathbf{R})(\alpha(\mathbf{R})+1)} \tag{2.3}$$

$$E\hat{F}_{0n}(t) = F_0(t). \tag{2.4}$$

Lemma 2.2. Let $\hat{F}_{0n}(t)$ be given by (1.8). Then $\lim_n Var[\hat{F}_{0n}(t)] = 0$.

Proof. From (2.1) and (2.4) we have

$$\begin{aligned} Var[\hat{F}_{0n}(t)] &= E[\hat{F}_{0n}(t) - F_0(t)]^2 \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var[\hat{F}_i(t)]. \end{aligned}$$

Let $\beta(t) = \dfrac{\alpha((-\infty,t])\alpha((t,\infty))}{\alpha(\mathbf{R})(\alpha(\mathbf{R})+1)}$. Then by lemma 2.1.,

$$\begin{aligned} Var[\hat{F}_i(t)] &= EVar(\hat{F}_i(t)|F_i) + VarE(\hat{F}_i(t)|F_i) \\ &= \frac{1}{k_i} EF_i(t)(1-F_i(t)) + Var(F_i(t)) \\ &= \frac{1}{k_i}\,\frac{\alpha((-\infty,t])\alpha((t,\infty))}{\alpha(\mathbf{R})(\alpha(\mathbf{R})+1)} + \frac{1}{\alpha(\mathbf{R})}\cdot\frac{\alpha((-\infty,t])\alpha((t,\infty))}{\alpha(\mathbf{R})(\alpha(\mathbf{R})+1)} \\ &= \left(\frac{1}{k_i} + \frac{1}{\alpha(\mathbf{R})}\right)\beta(t). \end{aligned}$$

Therefore,

$$Var[\hat{F}_{0n}(t)] = \frac{\beta(t)}{n^2}\sum_{i=1}^{n}\left(\frac{1}{k_i} + \frac{1}{\alpha(\mathbf{R})}\right) \to 0 \text{ as } n \to \infty.$$

Theorem 2.3. If $\hat{\alpha}_n(\mathbf{R})$ is a consistent estimate of $\alpha(\mathbf{R})$ based on $X_1, \cdots, X_n$, then the empirical Bayes procedure $\{H_n\}$ defined by (1.7)-(1.9) is a.o.

Proof. Using the L_2-orthogonality of $E(F_{n+1}(t)|\ X_{n+1})-F_{n+1}(t)$ and $H_{n+1}(t)-E(F_{n+1}(t)|\ X_{n+1})$,

$$\begin{aligned} 0 &\le R^{k_{n+1}}(H_{n+1},\alpha)-R^{k_{n+1}}(\alpha) \\ &= \int E[H_{n+1}(t)-\tilde{F}_{n+1}(t)]^2 dW(t). \end{aligned}$$

Since the integrand is bounded in t it suffices to show that

$$\lim_n E[H_{n+1}(t)-\tilde{F}_{n+1}(t)]^2=0 \quad \text{for all } t.$$

From (1.4), (1.5) and (1.7)-(1.9),

$$\begin{aligned} H_{n+1}(t)-\tilde{F}_{n+1}(t) &= (1-\hat{p}_{k_{n+1,n+1}})\hat{F}_{0n}(t)+\hat{p}_{k_{n+1,n+1}}\hat{F}_{n+1}(t) \\ &\quad -(1-p_{k_{n+1}})F_0(t)-p_{k_{n+1}}\hat{F}_{n+1}(t) \\ &= (1-\hat{p}_{k_{n+1,n+1}})(\hat{F}_{0n}(t)-F_{0(t))} \\ &\quad +(p_{k_{n+1}}-\hat{p}_{k_{n+1,n+1}})(F_0(t)-\hat{F}_{n+1}(t)). \end{aligned}$$

Therefore,

$$\begin{aligned} E[H_{n+1}(t)-\tilde{F}_{n+1}(t)]^2 &\le E[\hat{F}_{0n}(t)-F_0(t)]^2+2E|\hat{F}_{0n}(t)-F_0(t)| \\ &\quad + E(\hat{p}_{k_{n+1,n+1}}-p_{k_{n+1}})^2. \end{aligned}$$

It is clear that the last term converges to 0. The first two terms converge to 0 by lemma 2.2.

3. The consistent estimate of $\alpha(\mathbf{R})$

Zehnwirth(1981) has produced a consistent estimate of $\alpha(\mathbf{R})$ given by

$$\hat{\alpha}_n(\mathbf{R})=\begin{cases} 0, & \text{if } F\le 1 \\ k(F-1)^{-1}, & \text{if } F>1 \end{cases} \tag{3.1}$$

where F denotes the F-ratio statistic in the one-way ANOVA based on the treatments $X_1, \cdots, X_n$ and the integer k denotes the common sample size for each component.

This estimate has been used in Korwar and Hollander(1976) and Ghosh(1985) for their empirical Bayes problems of equal sample size components. The estimate

(3.1) based on the data of equal size is not directly applicable in our case. We modify (3.1) for the use in our empirical Bayes problem where sample sizes change.

Using the notations in one-way ANOVA let

$$W_n = MS_W = \sum_{i=1}^{n}\sum_{j=1}^{k_i} \frac{(X_{ij} - \overline{X}_{i\cdot})^2}{K_n - n}, \tag{3.2}$$

$$B_n = \frac{MS_B}{m_n} = \sum_{i=1}^{n} \frac{(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})^2}{n-1}, \tag{3.3}$$

where

$$K_n = \sum_{i=1}^{n} k_i = \text{ total sample size} \tag{3.4}$$

$$m_n = \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{k_i} \right)^{-1} = \text{harmonic mean of sample sizes.} \tag{3.5}$$

Let

$$\Phi_n = \frac{MS_B}{MS_W} = \frac{m_n B_n}{W_n}, \tag{3.6}$$

the F-ratio statistic in one-way ANOVA.

Define

$$\hat{\alpha}_n(\mathbf{R}) = \begin{cases} 0, & \text{if } \Phi_n \le 1 \\ m_n(\Phi_n - 1)^{-1}, & \text{if } \Phi_n > 1 \end{cases} \tag{3.7}$$

as an estimate of $\alpha(\mathbf{R})$ based on $X_1, \cdots, X_n$. Using Theorem3 and 4 (Ferguson (1973)) and a law of large numbers for the uncorrelated random variables with a bounded second moments we may state the following lemma.

In what follows we write $\alpha = \alpha(\mathbf{R})$. Let

$$\mu_r = EX_{ij}^r, \quad r \ge 1 \tag{3.8}$$

and

$$\sigma^2 = Var X_{ij} = \mu_2 - \mu_1^2. \tag{3.9}$$

Lemma 3.1. Let W_n be given by (3.2) with $2 \le k_i \le K$, $i = 1,2,\cdots,n$. Then

$$W_n \overset{\Pr}{\to} \frac{\alpha}{\alpha+1}\sigma^2, \tag{3.10}$$

provided that

$$\mu_4 = \alpha^{-1}\int_R x^4 d\alpha(x) < \infty. \tag{3.11}$$

Proof. Let $Y_i = \sum_{j=1}^{k_i}(X_{ij} - \overline{X}_i.)^2$, $i = 1,2,\cdots$. Then the Y_i are independent and by (3.11) second moments of the Y_i are bounded. Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - EY_i) \overset{\Pr}{\to} 0. \tag{3.12}$$

Using Theorem 3 and 4 of Ferguson(1973) we see that

$$EY_i = (k_i - 1)(\frac{\alpha}{\alpha+1}\sigma^2). \tag{3.13}$$

Combining (3.2),(3.12) and (3.13) we have

$$\frac{K_n - n}{n}[W_n - \frac{\alpha}{\alpha+1}\sigma^2] \overset{\Pr}{\to} 0. \tag{3.14}$$

Since $\frac{K_n - n}{n} \ge 1$ for $n = 1,2,\cdots$ (3.10) follows from (3.14).

Lemma 3.2. Let W_n, B_n be given by (3.2)-(3.5) with $2 \le k_i \le K$, $i = 1,2,\cdots,n$. Assume (3.11). Then

$$B_n - m_n^{-1}W_n \overset{\Pr}{\to} \frac{\sigma^2}{\alpha+1} \quad \text{as } n \to \infty. \tag{3.15}$$

Proof. By the similar arguments in the proof of lemma 3.1., we see that

$$B_n - \frac{1}{n-1}\sum_{i=1}^{n}Var(\overline{X}_i.) \overset{\Pr}{\to} 0. \tag{3.16}$$

Since $Var(\overline{X}_i.) = (\frac{\alpha}{\alpha+1})\frac{\sigma^2}{k_i} + \frac{\sigma^2}{\alpha+1}$, (3.16) becoms

$$B_n - \left[m_n^{-1}(\frac{\alpha}{\alpha+1})\sigma^2 + \frac{\sigma^2}{\alpha+1}\right]\left(\frac{n}{n-1}\right) \overset{\Pr}{\to} 0. \tag{3.17}$$

Also from lemma 3.1. we see that

$$m_n^{-1}\left[W_n - (\frac{\alpha}{\alpha+1})\sigma^2\right] \overset{\Pr}{\to} 0. \tag{3.18}$$

Combining (3.17) and (3.18) yields (3.15).

Theorem 3.3. Let $\hat{\alpha}_n(\mathbf{R})$ be given by (3.7). Then,

$$\hat{\alpha}_n(\mathbf{R}) \overset{Pr}{\to} \alpha(\mathbf{R}) \text{ as } n \to \infty. \tag{3.19}$$

Proof. Applying lemma 3.1., 3.2. to the numerator and denominator of

$$m_n(\Phi_n - 1)^{-1} = \frac{W_n}{B_n - m_n^{-1} W_n}$$

leads to (3.19).

References

[1] Chung, K. L. (1974), *A course in probability theory,* 2nd ed., Academic, New York.

[2] Ferguson, T. S. (1973), "A Bayesian analysis of some nonparametric problems. *Annals of Statistics*. 1, 209-230.

[3] Ghosh, M. (1985), "Nonparametric estimation of certain functionals," *Communications of Statistics Theory and Methods*, 14, 9, 2081-2094.

[4] Korwar, R. M. and Hollander, M. (1976), "Empirical Bayes estimation of a distribution function," *Annals of Statistics*. 4, 581-585.

[5] Zehnwirth, B. (1981), "A note on the asymptotic optimality of the empirical Bayes distribution function," *Annals of Statistics*. 9, 221-224.

분포함수의 추정및 응용에 관한연구[1]
(Dirichlet process에 의한 비모수 결정이론을 중심으로)

정 인 하[2]

요 약

각 성분 문제에서, 표본의 크기가 상이한 경우 Dirichlet process prior에 대한 경험적 베이즈에 의한 분포함수의 추정문제를 연구하였다. 특히, 위의 경험적 베이즈 문제에 사용할 수 있도록 Zehnwirth의 $\alpha(R)$을 수정하였다.

[1] 이 연구는 1991년도 아주대학교 연구비 지원에 의하여 연구되었음.
[2] (441-749) 경기도 수원시 권선구 원천동 산5번지 아주대학교 수학과