RATES OF CONVERGENCE IN EMPIRICAL BAYES ESTIMATION PROBLEMS: DISCRETE CASE*

PI-ERH LIN

(Received Nov. 14, 1969)

1. Introduction and summary

Consider the univariate discrete random variable x which is assumed to have a probability mass function of the form

$$p(x\,|\,\theta)=h(x)\beta(\theta)\theta^x\,,\qquad \theta\in(0,u),\ u>0\,;\ x=0,1,\cdots,N\,, \tag{1.1}$$

when θ is given. The numbers N, u may be finite or infinite. Assume that θ has an *a priori* distribution $G(\theta)$ over the interval $(0,u)$ such that, for all x, $x=0,\cdots,N$,

$$p(x)=\int_0^u p(x\,|\,\theta)dG(\theta)>0\,, \tag{1.2}$$

and

$$\int_0^u \theta^2 dG(\theta)<\infty\,. \tag{1.3}$$

Let D be the class of all non-randomized decision functions, when x is observed, with a generic element d. We wish to estimate θ with squared error loss $[d(x)-\theta]^2$. The Bayes estimator, relative to $G(\theta)$, and its optimal Bayes risk are, respectively

$$d_G(x)=w(x)p(x+1)/p(x)\,, \tag{1.4}$$

where

$$w(x)=h(x)/h(x+1) \tag{1.5}$$

and

$$\begin{aligned} B(G)&=\inf_{d\in D}\sum_{x=0}^{N}\int_0^u [d(x)-\theta]^2 p(x\,|\,\theta)dG(\theta)\\ &=\sum_{x=0}^{N}\int_0^u [d_G(x)-\theta]^2 p(x\,|\,\theta)dG(\theta)\,. \end{aligned} \tag{1.6}$$

---

* This paper is part of a doctoral dissertation submitted at Columbia University, November 1968. The research was supported by NSF Grant GP-7456.

In practice, however, it is found that $G(\theta)$ is usually unknown, but the estimation problem about θ with the same loss function occurs repeatedly and independently. More precisely, let $(x_1, \theta_1), \cdots, (x_n, \theta_n), \cdots$, be a sequence of pairs of random variables, each pair being independent of the others, the θ's having a common *a priori* distribution $G(\theta)$ and the conditional mass function of x_n given $\theta_n=\theta$ being $p(x|\theta)$. At the $(n+1)$st stage, when the decision is to be made about θ_{n+1}, we have observed $x_1, x_2, \cdots, x_{n+1}$, although the values of $\theta_1, \theta_2, \cdots, \theta_n$ remain unknown. From this knowledge, an empirical Bayes estimator, $d_n(x)= d_n(x_1, \cdots, x_n; x)$, depending only on $x_1, \cdots, x_n$, x, can be constructed. This is usually referred to as *the standard empirical Bayes assumption*. The Bayes risk associated with $d_n(x)$, given by

$$B_n=\sum_{x_1=0}^{N}\cdots\sum_{x_n=0}^{N}\sum_{x=0}^{N}\int_0^u [d_n(x)-\theta]^2 \prod_{i=1}^{n} p(x_i)p(x|\theta)dG(\theta) , \tag{1.7}$$

will converge to $B(G)$ as $n\to\infty$. If this can be achieved, the sequence $\{d_n(x),\ n=1, 2, \cdots\}$ is called *asymptotically optimal*.

In Section 2, a sequence of empirical Bayes estimators is proposed. This sequence is shown to be asymptotically optimal, and the rate of convergence for the Bayes risks associated with the empirical Bayes estimators to the optimal Bayes risk is obtained. Finally, some examples are exhibited in Section 3 to show the applicability of the main result.

2. Rates of convergence

Fix an arbitrary decision function $d_0 \in D$, the Bayes risks associated with d_G and d_n can be written, respectively, as

$$B(G)=c_G+\sum_{x=0}^{N} \Delta_G(d_G, x) \tag{2.1}$$

and

$$B_n=c_G+\sum_{x=0}^{N} E[\Delta_G(d_n, x)] , \tag{2.2}$$

where we have set

$$c_G=\sum_{x=0}^{N}\int_0^u (d_0-\theta)^2 p(x|\theta)dG(\theta) , \tag{2.3}$$

$$\Delta_G(d_G, x)=-p(x)(d_G-d_0)^2 , \tag{2.4}$$

$$E[\Delta_G(d_n, x)]=p(x)[E(d_n-d_G)^2-(d_G-d_0)^2] , \tag{2.5}$$

and where E denotes expectation with respect to the joint distribution of $x_1, x_2, \cdots, x_n$. From (2.1) and (2.2), it is easy to prove the following result.

LEMMA 2.1. *If $B(G)$ and B_n are given by* (2.1) *and* (2.2), *respectively, then*

$$0 \leqq B_n - B(G) = \sum_{x=0}^{N} p(x) E(d_n - d_G)^2 , \tag{2.6}$$

where E is as defined in (2.5).

We are now in a position to construct an empirical Bayes estimator of θ. Define an indicator function

$$I_j(x) = \begin{cases} 1 & \text{if } x_j = x \\ 0 & \text{otherwise} , \end{cases}$$

and let

$$p_n(x) = p_n(x_1, \cdots, x_n; x) = n^{-1} \sum_{j=1}^{n} I_j(x) \tag{2.7}$$

be an empirical estimator of $p(x)$. It is easily seen that, for each x,

$$E[p_n(x)] = p(x) \tag{2.8}$$

and

$$\text{Var}\,[p_n(x)] \leqq (4n)^{-1} . \tag{2.9}$$

With the standard empirical Bayes assumptions, we propose an empirical Bayes estimator of θ, at the $(n+1)$st stage, given by

$$d_n(x) = \begin{cases} w(x) p_n(x+1)/p_n(x) & \text{if } p_n(x) \geqq \delta_n \\ w(x) p_n(x+1)/\delta_n & \text{otherwise} , \end{cases} \tag{2.10}$$

where $\{\delta_n,\ n=1, 2, \cdots\}$ is a sequence of positive real numbers with

$$c_1 n^{-\gamma} \leqq \delta_n \leqq c_2 n^{-\gamma} , \qquad 0 < c_1 \leqq c_2 < \infty,\ \gamma = 1/3 . \tag{2.11}$$

LEMMA 2.2. *Let $p(x)$, $p_n(x)$ be given by* (1.2) *and* (2.7), *respectively. For some positive real number δ_n, defined by* (2.11), *and some t, $0 < t \leqq 1$,*

$$P[p_n(x) < \delta_n] \leqq c_3 \delta_n^t , \qquad 0 < c_3 < \infty ,$$

if and only if

$$P[p(x) < \delta_n] \leqq c_4 \delta_n^t , \qquad 0 < c_4 < \infty ,$$

PROOF. To simplify notation, we write $p_n = p_n(x)$ and $p = p(x)$. The "if" part:

$$\begin{aligned}P(p_n<\delta_n)&\leqq P(p_n<\delta_n,\,|p_n-p|>\delta_n)+P(p_n<\delta_n,\,|p_n-p|\leqq\delta_n,\,p\geqq 2\delta_n)\\&\quad+P(p_n<\delta_n,\,|p_n-p|\leqq\delta_n,\,p<2\delta_n)\\&\leqq P(|p_n-p|>\delta_n)+0+P(p<2\delta_n)\\&\leqq\delta_n^{-2t}E|p_n-p|^{2t}+c_4(2\delta_n)^t\\&\leqq c_3\delta_n^t\,.\end{aligned}$$

The last inequality is obtained by noting that

$$\begin{aligned}E|p_n-p|^{2t}&=E[E(|p_n-p|^{2t}|x)]\\&\leqq E[E(|p_n-p|^2|x)]^t\\&\leqq(4n)^{-t}\,.\end{aligned}$$

The "only if" part can similarly be proved.

We are now able to state the rate of convergence theorem.

THEOREM 2.1. *Let $p(x|\theta)$ be given by* (1.1), *and $G(\theta)$ be such that* (1.2) *and* (1.3) *hold. To estimate θ with squared error loss, the Bayes and empirical Bayes estimators are given by* (1.4) *and* (2.10), *respectively, with their corresponding Bayes risks given by* (2.1) *and* (2.2). *If,*

$$\sum_{x=0}^{N}w^2(x)p(x)p(x+1)<\infty\,, \tag{2.12}$$

$$\sum_{x=0}^{N}w^2(x)p^2(x+1)<\infty\,, \tag{2.13}$$

$$\sum_{x=0}^{N}w^2(x)p^2(x+1)/p(x)<\infty\,, \tag{2.14}$$

and if, for some t, $0<t\leqq 1$,

$$P[p(x)<\delta_n]\leqq c\delta_n^t\,, \tag{2.15}$$

where δ_n is defined by (2.11) *and $0<c<\infty$, then*

$$B_n-B(G)=0(n^{-t/3})\,, \tag{2.16}$$

and hence the sequence $\{d_n,\ n=1,2,\cdots\}$ of estimators given by (2.10) *is asymptotically optimal.*

PROOF. To simplify notation, we write $g_n=p_n(x+1)$, $g=p(x+1)$, $p_n=p_n(x)$, $p=p(x)$ and $w=w(x)$. Now for each x, $x=0,1,\cdots,N$, we have the upper bound

$$\begin{aligned}w^{-2}E(d_n-d_G)^2&\leqq 2\delta_n^{-2}E(g_n-g)^2+4(g/p)^2\delta_n^{-2}E(p_n-p)^2\\&\quad+16(g/p)^2c'\delta_n^t\,,\end{aligned} \tag{2.17}$$

where $0<c'<\infty$, by applying Lemma 2.2. From Lemma 2.1 and the inequalities (2.9), (2.17), together with assumptions (2.12) through (2.14), we have

$$B_n-B(G)=\sum_{x=0}^{N} pE(d_n-d_G)^2$$
$$\leqq c_1^{-3}\delta_n\left(2\sum_{x=0}^{N} w^2pg+4\sum_{x=0}^{N} w^2g^2\right)+16c'\delta_n^t\sum_{x=0}^{N} w^2g^2/p ,$$

which completes the proof.

Note that conditions (2.12) through (2.14) always hold when N is finite. Let us consider the case when $N=\infty$. Suppose that there exist positive integers C and M_1 such that, for all $x\geqq M_1$,

$$\sum_{x=M_1}^{\infty} w^2(x)p(x+1)\leqq C , \tag{2.18}$$

then conditions (2.12), (2.13) and (2.14) are satisfied. For, since $p(x)$ is a probability mass function, it is necessary that there exists a positive integer M_2 such that $p(x+1)\leqq p(x)$ for all $x\geqq M_2$. Setting $M=\max(M_1, M_2)$, we have

(i) $\sum_{x=M}^{\infty} w^2(x)p(x)p(x+1)\leqq Cp(M)$,

(ii) $\sum_{x=M}^{\infty} w^2(x)p^2(x+1)\leqq Cp(M)$,

(iii) $\sum_{x=M}^{\infty} w^2(x)p^2(x+1)/p(x)\leqq C$.

From which conditions (2.12), (2.13) and (2.14) follow immediately. Thus we have proved

COROLLARY 2.1. *If conditions* (2.18) *and* (2.15) *are satisfied, then the result* (2.16) *holds.*

Observe that (2.18) holds if one of the following three conditions is satisfied:

(2.19) There exists a positive integer x_0 such that, for all $x\geqq x_0$, $w^2(x+1)p(x+2)<w^2(x)p(x+1)$,

(2.20) $\sup_x |w(x)|<\infty$,

(2.21) $\lim_{x\to\infty} |w(x)|=0$.

3. Examples

(i) Consider a Poisson distribution

$$p(x\mid\theta)=e^{-\theta}\theta^x/x! ,\qquad x=0,\cdots,\theta>0 , \tag{3.1}$$

and let the prior probability density be

$$G'(\theta)=ae^{-a\theta}, \qquad \theta>0,\ 0<a<\infty . \tag{3.2}$$

Then we have

$$p(x)=(a/x!)\int_0^\infty \theta^x e^{-(a+1)\theta}d\theta=a(a+1)^{-(x+1)} .$$

Since for any real number $b>1$,

$$\sum_{x=1}^{\infty} x^2 b^{-x}=b(b+1)(b-1)^{-3}<\infty , \tag{3.3}$$

and since $w(x)=x+1$, conditions (2.12) through (2.14), which can be rewritten in the form (3.3) with $b=a+1$ or $b=(a+1)^2$, are satisfied. Moreover, letting

$$y_n=-(\log \delta_n-\log a)/\log (a+1) ,$$

we have

$$\begin{aligned} P\{p(x)<\delta_n\}=P(x\geqq y_n)&\leqq a\sum_{x=[y_n]}^{\infty}(a+1)^{-(x+1)} \\ &\leqq(1+a^{-1})\delta_n , \end{aligned}$$

where $[y_n]$ denotes the integral part of y_n, and hence condition (2.15) is satisfied. Thus we have proved

THEOREM 3.1. *Let x be a Poisson random variable with mean θ and let $G'(\theta)$ be given by* (3.2). *To estimate θ with squared error loss, the results of Theorem* 2.1 *hold with $t=1$.*

(ii) Let x have a negative binomial mass function

$$p(x\mid\theta)=\binom{x+k-1}{x}\theta^x(1-\theta)^k , \qquad \theta\in(0,1);\ x=0,1,\cdots , \tag{3.4}$$

for some fixed $k=1,2,\cdots$, when θ is given. If the prior probability density of θ is

$$G'(\theta)=\begin{cases} 1 & \text{if } \theta\in(0,1) \\ 0 & \text{otherwise} , \end{cases} \tag{3.5}$$

then

$$p(x)=k\{(x+k)(x+k+1)\}^{-1} .$$

Noting that $w(x)=(x+1)/(x+k)\leqq 1$, condition (2.18) is verified by (2.20). To check condition (2.15), we proceed as follows. Let

$$y_n = (k\delta_n^{-1})^{1/2} - k - 1 \ .$$

Then

$$\begin{aligned} P\{p(x) < \delta_n\} &\leqq P\{(x+k+1)^{-2} < k^{-1}\delta_n\} = p(x > y_n) \\ &= \sum_{x=[y_n]+1}^{\infty} k\{(x+k)(x+k+1)\}^{-1} \\ &= k\{[(k\delta_n^{-1})^{1/2}]\}^{-1} \\ &\leqq c\delta_n^{1/2} , \end{aligned}$$

where $0 < c < \infty$ and $[y_n]$ denotes the integral part of y_n. We may summarize the result as follows:

THEOREM 3.2. *Let* $p(x \mid \theta)$, $G'(\theta)$ *be given by* (3.4) *and* (3.5), *respectively. To estimate* θ *with squared error loss, Theorem* 2.1 *holds with* $t = 1/2$.

THE FLORIDA STATE UNIVERSITY

REFERENCES

[1] Loève, M. (1960). *Probability Theory*, 3rd Ed., Van Nostrand, Princeton.
[2] Robbins, H. (1963). The empirical Bayes approach to testing statistical hypotheses, *Rev. Inst. Inter. Statist.*, **31**, 195-208.
[3] Robbins, H. (1964). The empirical Bayes approach to statistical decision problems, *Ann. Math. Statist.*, **35**, 1-20.