Chin. Ann. of math.
4B (2) 1983

ON ONE CONJECTURE OF R. S. SINGH

CHEN XIRU (陈希孺)
(The University of Science and Technology of China)

Abstract

In 1979 R. S. Singh (Ann. Statist, 1979, p. 890) made a conjecture concerning the convergence rate of EB estimates of the parameter θ in an one-dimensional continuous exponential distribution family, under the square loss function, the prior distribution family being confined to a bounded interval. The conjecture asserts that the rate cannot reach o(1/n) or even O(1/n). In this article, the weaker part of this conjecture (i. e. the o(1/n) part) is shown to be correct.

1. Introduction

Suppose that there are given a sample space $\mathscr{X}$, a family of probability distributions depending on an one-dimensional parameter θ, and loss function $(\theta-a)^2$. Suppose also that there are gisven a family $\mathscr{F}$ of prior distributions. and the "historical" samples $X_1, \cdots, X_n$ and "contemporary" sample X, in a standard empirical Bayes structure. Denote by $\delta_G(X)$ the Bayes estimate of θ when the prior is G. In the case we know only $G\in\mathscr{F}$ and $\mathscr{F}$ contains more than one member. δ_G can no longer be determined, and one may resort to the empirical Bayes (EB) estimation based on $X_1, \cdots, X_n$ and X. The Bayes risk of δ_G and the "over-all" Bayes risk of an EB esfimate $\delta_n\triangleq\delta_n(X)\triangleq\delta_n(X_1, \cdots, X_n, X)$ will be denoted by $R(G)$ and $R(\delta_n, G)$, respectively. If

$$\lim_{n\to\infty} R(\delta_n, G)=R(G), \text{ for each } G\in\mathscr{F}, \tag{1}$$

then δ_n is said to be asymptotically optimal (a. o.) with respect to the prior family $\mathscr{F}$. Recently, there has been much interest in the rate of convergence in (1), in the case of exponential family. Under various (rather complicate) conditions, Lin, Singh and Zhao, among others, obtained rates of convergence in the form

$$R(\delta_n, G)-R(G)=O(n^{-t}) \tag{2}$$

(see [1] ~ [3]).

In 1979, Singh[2] considered the Lebesgue-exponential family

$$f(x, \theta)dx=C(\theta)h(x)e^{\theta x}dx. \tag{3}$$

He showed that under certain conditions, EB estimate δ_n can be constructed such that in (2), t may take values arbitrarily close to, but less than, one (Zhao established in

Manuscript received February 17, 1981.

[3] a similar result for discrete exponential family). In view of this result Singh made the conjecture that however the concrete form of (3) may be, and even if we assume a "well-behaved" prior family such as

$$\mathscr{F}_{a,b}=\{G: G \text{ has a support within } (a, b)\} \tag{4}$$

(where $-\infty<a<b<\infty$), we cannot find an EB estimate δ_n such that the left hand side of (2) is of the order $o\left(\frac{1}{n}\right)$ or even $O\left(\frac{1}{n}\right)$ for any G belonging to the prior family.

The purpose of this article is to show that the weaker part of Singh's conjecture, the $o\left(\frac{1}{n}\right)$ part, is correct. It remains unknown whether the $O\left(\frac{1}{n}\right)$ part of the conjectnre is correct or not.

2. The main theorem

For simplicity we shall consider the case of discrete exponential family. The result, together with its proof, carries over to the continuous case (3) without any difficulty.

Consider the exponential famrily

$$P(X=x\mid\theta)=h(x)\beta(\theta)\theta^x, x=0, 1, 2, \cdots, \tag{5}$$

where

$$h(x)>0, \ x=0, 1, 2, \cdots \tag{6}$$

and the parameter space is

$$\Theta=\left\{\theta: \theta>0, \ \sum_{x=0}^{\infty} h(x)\theta^x<\infty\right\}.$$

We may assume $2\in\Theta$, for otherwise we can choose suitable $b>0$ and use $b\theta$ as the new parameter. Thus $\sum\limits_{x=0}^{\infty} h(x)<\infty$, and

$$0<\left(\sum_{x=0}^{\infty} h(x)\right)^{-1}\leqslant\beta(\theta)\leqslant\frac{1}{h(0)}<\infty, 0<\theta<1. \tag{7}$$

Theorem 1. *Under the above assumptions and suppose the prior distribution family is $\mathscr{F}_{0,1}$ (see (4)), for any EB estimate δ_n there exists $G\in\mathscr{F}_{0,1}$ such that*

$$\liminf_{n\to\infty} n(R(\delta_n, G)-R(G))>0. \tag{8}$$

Hence, at least for this G, $R(\delta_n, G)-R(G)$ is not of the form $o\left(\frac{1}{n}\right)$.

Proof Take a subfamily $\mathscr{F}^*$ of $\mathscr{F}_{0,1}$ as follows

$$\mathscr{F}^*=\left\{G_\lambda: dG_\lambda(\theta)=\frac{M(\lambda)}{\beta(\theta)}\theta^\lambda I_{(0,1)}(\theta)d\theta, \ 0<\lambda<\infty\right\}, \tag{9}$$

where

$$M(\lambda)=\left(\int_0^1\frac{1}{\beta(\theta)}\theta^\lambda d\theta\right)^{-1}$$

which is continuously differentiable in $\lambda>0$.

Under the prior G_λ, X has a marginal distribution

$$p_\lambda(x)=P_\lambda^*(X=x)=M(\lambda)h(x)(1+\lambda+x)^{-1},\ x=0,\ 1,\ 2,\ \cdots \tag{10}$$

and the Bayes estimate of θ is

$$\delta_\lambda(x)=\frac{h(x)}{h(x+1)}\frac{p_\lambda(x+1)}{p_\lambda(x)}=\frac{1+\lambda+x}{2+\lambda+x}. \tag{11}$$

Now suppose in the contrary that there is an EB estimate δ_n such that

$$R(\delta_n,\ G_\lambda)-R(G_\lambda)=o\left(\frac{1}{n}\right),\ \text{for each } \lambda>0. \tag{12}$$

By the well-known formula

$$R(\delta_n, G_\lambda)-R(G_\lambda)=\sum_{x=0}^{\infty}p_\lambda(x)E_\lambda[\delta_n(X_1, \cdots, X_n,\ x)-\delta_\lambda(x)]^2 \tag{13}$$

and the fact that $p_\lambda(x)>0$ (which is an easy consequence of (6)), one finds that

$$\lim_{n\to\infty} E_\lambda[\delta_n(X_1,\ \cdots,\ X_n,\ x)-\delta_\lambda(x)]^2=0 \tag{14}$$

for each $\lambda>0$ and $x=0,\ 1,\ 2,\ \cdots$, and this in turn contains that as $n\to\infty$

$$g_n(\lambda,\ x)\triangleq E_\lambda[\delta_n(X_1,\ \cdots,\ X_n,\ x)]\to\delta_\lambda(x) \tag{15}$$

for each $\lambda>0$ and $x=0,\ 1,\ 2,\cdots$. Take arbitrarily positive integer x_0, and for simplicity write $\delta(\lambda)$ for $\delta_\lambda(x_0)$. We can view $\delta_n(X_1,\ \cdots,\ X_n\ x_0)$ as an estimate of $\delta(\lambda)$, with the *iid*. samples $X_1,\ X_2,\ \cdots$ drawn from a population possessing probability distribution (10).

It is an easy maeter to verify the following facts:

a) The parameter space of distribution family (10) is an interval (i. e., $(0,\ \infty)$).

b) $p_\lambda(x)>0$, $\partial p_\lambda(x)/\partial\lambda$ exists and is a continuous function of λ, for $\lambda>0$ and $x=0,\ 1, 2,\ \cdots$.

c) For each $\lambda>0$

$$\sum_{x=0}^{\infty}\partial p_\lambda(x)/\partial\lambda=0. \tag{16}$$

Since, as pointed out earlier, $\sum_{x=0}^{\infty}h(x)<\infty$ and $M(\lambda)$ is continuously differentiable, it follows easily that the series $\sum_{x=0}^{\infty}\ \partial p_\lambda(x)/\partial\lambda$ converges uniformly in $a\leqslant\lambda\leqslant b$ for any $0<a<b<\infty$. This, together with $\sum_{x=0}^{\infty}p_\lambda(x)=1$, establishes (16).

d) $d\delta(\lambda)/d\lambda$ is continuous for $\lambda>0$, and

$$I(\lambda)=E_\lambda[\partial\log p_\lambda(X)/\partial\lambda]^2=\sum_{x=0}^{\infty}p_\lambda(x)\left(\frac{M'(\lambda)}{M(\lambda)}-\frac{1}{1+\lambda+x}\right)^2.$$

Hence $I(\lambda)$ is continuous and positive for $\lambda>0$. So

$$0<\frac{1}{A}\triangleq\sup_{1\leqslant\lambda\leqslant2}\ I(\lambda)<\infty. \tag{17}$$

Now we observe that, under any prior distribution $G_\lambda\in\mathscr{F}^*$, the Bayes estimate $\delta_\lambda(\cdot)$ can take values only in $(0,\ 1)$. Hence, if we introduce a new EB estimate $\bar{\delta}_n=\bar{\delta}_n(x_1,\ \cdots,\ x_n,\ x)$ as follows

$$\bar{\delta}_n = \begin{cases} 0, & \text{for} \quad \delta_n \leqslant 0, \\ \delta_n, & \text{for} \quad 0 < \delta_n < 1, \\ 1, & \text{for} \quad \delta_n \geqslant 1, \end{cases}$$

then we have

$$0 \leqslant R(G_\lambda) \leqslant R(\bar{\delta}_n, G_\lambda) \leqslant 1, \quad R(\bar{\delta}_n, G_\lambda) \leqslant R(\delta_n, G_\lambda) \tag{18}$$

for any $\lambda > 0$. From (12), (18), it follows that

$$R(\bar{\delta}_n, G_\lambda) - R(G_\lambda) = o\left(\frac{1}{n}\right), \text{ for any } \lambda > 0. \tag{19}$$

The argument leading to (15) applies to $\bar{\delta}_n$, rendering

$$g_n(\lambda) \triangleq E_\lambda[\bar{\delta}_n(X_1, \cdots, X_n, x_0)] \to \delta(\lambda) \tag{20}$$

for any $\lambda > 0$ as $n \to \infty$. Since $0 \leqslant \bar{\delta}_n \leqslant 1$, it follows easily that the series

$$g_n(\lambda) = \sum_{x_1=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} \bar{\delta}_n(x_1, \cdots, x_n, x_0) M^n(\lambda) \prod_{i=1}^{n} \frac{h(x_i)}{1+\lambda+x_i}$$

(an be diffrentiated under the summation sign, and hence $g_n(\lambda)$ is continuously diffrentiable in $\lambda > 0$. From this and the facts a)—d) proved above, it follows that (see [4]) the conditions for applying the Cramer-Rao inequality are fulfilled. Hence, noticing (17), we get

$$\operatorname{var}_\lambda(\bar{\delta}_n(X_1, \cdots, X_n, x_0)) \geqslant \frac{(g_n'(\lambda))^2}{nI(\lambda)} \geqslant \frac{A}{n}(g_n'(\lambda))^2, \ 1 \leqslant \lambda \leqslant 2. \tag{21}$$

From (18) and (19), using Fatou's lemma, we have

$$\begin{aligned} &\limsup_{n\to\infty} \int_1^2 n(R(\bar{\delta}_n, G_\lambda) - R(G_\lambda))d\lambda \\ &\qquad \leqslant \int_1^2 \limsup_{n\to\infty} [n(R(\bar{\delta}_n, G_\lambda) - R(G_\lambda))]d\lambda = 0. \end{aligned} \tag{22}$$

On the other hand, denote

$$d \triangleq \inf_{1\leqslant\lambda\leqslant 2} p_\lambda(x_0) \geqslant \frac{h(x_0)}{3+x_0} \inf_{1\leqslant\lambda\leqslant 2} M(\lambda) > 0,$$

by (13) (replacing δ_n by $\bar{\delta}_n$), (20) and Cauchy-Schwarz inequality, we also have as $n \to \infty$

$$\begin{aligned} &\int_1^2 n(R(\bar{\delta}_n, G_\lambda) - R(G_\lambda))d\lambda \\ &\geqslant \int_1^2 np_\lambda(x_0) E_\lambda[\bar{\delta}_n(X_1, \cdots, X_n, x_0) - \delta(\lambda)]^2 d\lambda \\ &\geqslant \int_1^2 nd \operatorname{var}_\lambda[\bar{\delta}_n(X_1, \cdots, X_n, x_0)]d\lambda \geqslant Ad\int_1^2 [g_n'(\lambda)]^2 d\lambda \\ &\geqslant Ad\left(\int_1^2 g_n'(\lambda)d\lambda\right)^2 = Ad(g_n(2) - g_n(1))^2 \\ &\to Ad(\delta(2) - \delta(1))^2 = Ad/[(4+x_0)^2(3+x_0)^2] > 0 \end{aligned} \tag{23}$$

which contradicts evidently (22), and the theorem is proved.

3. Some further remarks.

a) It is easily seen that the conclusion and the proof of the theorem remain valid when the prior family is changed from $\mathscr{F}_{0,1}$ to $\mathscr{F}_{a,b}$ for any $a\in\Theta$, $b\in\Theta$, and $a<b$.

b) In case of (6), not being satisfied, it has been shown in [5] that there exists no a.o. EB estimate of θ, unless the prior distribution family satisfies a condition of very peculiar nature. With some necessary modifications to look after this point, the whole argument in section 2 goes through without difficulty.

c) The method of proof employed in this paper applies to a wide class of distributions, not only the exponential. Supposing that the distribution family of X, $\{P_\theta,\ \theta\in\Theta\}$, is rather "well-behaved", one can usually find a prior family $\{G_\lambda,\ \lambda\in I\}$, where I is an interval of $(-\infty,\ \infty)$, such that

(i) The family of marginal distribution of X, i. e.

$$P_\lambda^*(X\in A)=\int_I P_\theta(A)dG_\lambda(\theta),\ \lambda\in I$$

satisfies the conditions for employing the Cramer-Rao inequality.

(ii) The Bayes estimate of $\theta, \delta_\lambda(x)$, under the prior distribution G_λ, is a sufficiently smooth function of λ. for each $x\in\mathscr{X}$ with the possible exception of a $\mathscr{P}^*$-null set.

If this is the case, then, starting from the basic equality

$$R(\delta_n,\ G_\lambda)-R(G_\lambda)=\int_{\mathscr{X}} E_\lambda[\delta_n(X_1,\ \cdots,\ X_n,\ x)-\delta_\lambda(x)]^2dP_\lambda^*(x)$$

one has not much difficulty in carring through the arguments of section 2 to reach the conclusion that it is impossible to have

$$R(\delta_n,\ G_\lambda)-R(G_\lambda)=o\left(\frac{1}{n}\right),\ \text{for all } \lambda\in I,$$

with any possible choice of δ_n. Particularily, in case of exponential family this can be easily done.

However, the $O\left(\frac{1}{n}\right)$ part of Singh's conjecture looks difficult.The author does not know whether an easy and satisfactory answer can be reached.

References

[1] Lin, P. E., Rates of convergence in empirical Bayes estimation problems:Continuous case, *Ann. Statist.* **3**, (1975), 155—164.

[2] Singh, R., Empirical Bayes estimation in Lebesgue-exponential families with rates near best possible rate, Ann. statist., **7** (1979), 890—901.

[3] Zhao Lincheng(赵林城), EB estimates with convergence rates about a class of discrete distribution families (in chinese), *J. math. Research & Exposition*, **1**:1 (1981), 59—69.

[4] Wijsman, R. A., On the attainment of the Cramer-Rao lower bound, *Ann. Statist.*, **1** (1973), 538—542.

[5] Chen Xiru (陈希孺), Asymptotically optimal EB estimates for the parameters of one-dimensional discrete exponential families, Chinese Annals of Mathematics, **4B**:1 (1983), 41—50.