Chin. Ann. of Math.
4B (1) 1983

ASYMPTOTICALLY OPTIMAL EMPIRICAL BAYES ESTIMATION FOR PARAMETER OF ONE-DIMENSIONAL DISCRETE EXPONENTIAL FAMILIES

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

Abstract

Consider the discrete exponential family written in the form $P_\theta(X=x)=h(x)\beta(\theta)\theta^x$, $x=0, 1, 2, \cdots$, where $h(x)>0$, $x=0, 1, 2, \cdots$. The prior distribution of θ belongs to the family $\mathscr{F}=\left\{G: \int_0^\infty \theta^2 dG<\infty\right\}$. Denote by $\delta_n(x)=\delta_n(x_1, \cdots, x_n; x)=\frac{h(x)}{h(x+1)}\frac{u_n(x+1)}{u_n(x)}$ the Robbins' *EB* Estimate of θ under the square loss $(\theta-d)^2$, where $u_n(i)$ is the number of elements equaling to i in $x_1, \cdots, x_n, x$. Under a quite general assumption imposed upon h, it is shown that δ_n is an asymptotically optimal *EB* estimate of θ relative to the whole family $\mathscr{F}$. Further, the condifion imposed on h mentioned above can be dispensed with by slightly modifying the definition of δ_n. Also the case that h assumes the value zero is discussed.

§ 1. Introduction and Summary

Consider the discrete exponential distribution written in the following form:

$$P_\theta(X=x)=h(x)\beta(\theta)\theta^x, \quad x\in\mathscr{X}=\{0, 1, \cdots, N\},$$

$$\theta\in\Theta, \ \Theta=\left\{\theta: \theta>0, \ \sum_{x=0}^{N} h(x)\theta^x<\infty\right\}, \tag{1}$$

where N is a positive integer or ∞. One may always take $N=\infty$ by letting $h(i)=0$ for $i>N$ if necessary. Θ is a finite or infinite interval. It is well-known that (1) contains many important discrete distributions, for example

Poisson: $$h(x)=\frac{1}{x!}, \ \beta(\theta)=e^{-\theta}, \ \Theta=(0, \infty), \tag{2}$$

Negative Binomial: $h(x)=\binom{x+r-1}{x}$ (r-known positive integer),

$$\beta(\theta)=(1-\theta)^r, \quad \Theta=(0, 1),$$

Logarithm: $h(0)=0$, $h(x)=\frac{1}{x}$ for $x\geqslant 1$

Manuscript received January 31, 1981.

$$\beta(\theta)=\log\frac{1}{1-\theta},\quad \Theta=(0,\ 1).$$

Also the Binomial $P_\theta(X=x)=\binom{k}{x}\theta^x(1-\theta)^{k-x},\quad x=0,\ 1,\ \cdots,\ k\ (k\ \text{known}),\quad 0<\theta<1,$ can be reduced to the form of (1) by introducing the new parameter $\varphi=\frac{\theta}{1-\theta}$.

The purpose of this artical is to study the asymptotically optimal. (a. o.) empirical Bayes (EB) estimation of θ in (1), under quadratic loss $(\theta-a)^2$. Our main result is that if the prior distributions are restricted to the class

$$\mathscr{F}=\left\{G:\ \int_\Theta \theta^2 dG(\theta)<\infty\right\} \tag{3}$$

and

$$h(x)>0,\quad x=0,\ 1,\ 2,\ \cdots, \tag{4}$$

then an a. o. EB estimator of θ can be constructed. We shall also consider the case when (4) does not hold.

Now we briefly mention the historical developments. First, in the Poisson case (2) Robbins[1] introduced the "natural" EB estimate of θ

$$\delta_n^*(x)=\delta_n^*(x_1,\ \cdots,\ x_n,\ x)=(x+1)u_n(x+1)/u_n(x), \tag{5}$$

where $x_1,\ \cdots,\ x_n$ are the "historical" samples, x is the "present" sample, and u_n is defined as

$$u_n(i)=u_n(x_1,\ \cdots,\ x_n,\ x;\ i)=\frac{1}{n+1}\{\text{number of } x_1,\ \cdots,\ x_n,\ x \text{ equaling to } i\}. \tag{6}$$

According to Robbins[3], Johns in [4] proved that this estimate is a. o. under quadratic loss and prior distribution family (3). This is the first significant result concerning a. o. EB estimation. It seems that the paper [4] has not been published openly, and we do not know the method of his proof. Another result of Johns[5] is related to this problem, but has the undesirable feature that each historical sample must be replicated at least twice, and thus in some degree violates the original meaning of EB structure.

In 1972, P. E. Lin considered in [6] the family (1). Assuming temporarily (4) holds, one shows easily that under the prior G the Bayes estimate of θ (always assuming a quadratic loss in the present artical) is

$$\delta_G(x)=W(x)f_G(x+1)/f_G(x), \tag{7}$$

where $W(x)=h(x)/h(x+1)$, and f_G is the marginal distribution of X under prior G

$$f_G(x)=\int_\Theta h(x)\beta(\theta)\theta^x dG(\theta),\ x=0,\ 1,\ 2,\ \cdots. \tag{8}$$

Hence, similar to the Poisson case, one obtained the "natural" EB estimator of θ

$$\delta_n(x)=\delta_n(x_1,\ \cdots,\ x_n,\ x)=w(x)u_n(x+1)/u_n(x) \tag{9}$$

(Notice that $u_n(x)>0$). Lin modified this estimate in attempting to obtain an a. o. EB estimate of θ with a convergence rate of $O(n^{-\frac{1}{3}})$ under a group of conditions. Unfortunately his proof contains a serious error, and the main result of [6] was incorrect, as shown by Chao in a recent paper [7] by constructing a counter-example. Also, Chaomodified the estimate (9) by truncation, obtaining an EB estimate of θ with a convergence rate $O(n^{-t})$ (t depends upon various conditions and may be arbitrarily near one).

Thus, as far as the author knows, the following two important questions have not been solved clearly in the published literature:

1. Is the natural EB estimate (9) a. o. under the prior family (3)?

2. If the above question cannot be answered in the positive unconditionally, then can we modify (9) in a suitable manner, in order to get an EB estimate $\tilde{\delta}_n$ of θ, which is a. o. under the sole condition that the prior belongs to family (3)?

This paper is devoted to these questions. The main results can be formulated in the following two theorems:

Theorem 1. *If* (4) *is true and there exists a constant A such that*

$$h^3(x) \leqslant Ah(x-2)h^2(x+1), \quad x=2, 3, \cdots, \tag{10}$$

then (9) *is an a. o. EB estimator of θ under the prior family* (3).

Modify δ_n as follows: When the present sample is x, we do not estimate $f_G(x+1)$ by $u_n(x+1)$. Instead, we use

$$v_n(x) = \begin{cases} u_n(x+1), & \text{if } \{i: 1 \leqslant i \leqslant n, x_i = x+1\} \geqslant 2, \\ 0, & \text{otherwise}, \end{cases}$$

and define

$$\tilde{\delta}_n(x) = \tilde{\delta}_n(x_1, \cdots, x_n, x) = W(x)v_n(x)/u_n(x). \tag{11}$$

Theorem 2. *If* (4) *is true, then* (11) *is an a. o. EB estimator of θ under the prior family* (3).

Section 2 is devoted to the proof of these theorems and in section 3 some related problems are considered.

§ 2. Proof of The Theorems

The Main task is to prove Theorem 1. A minor modification of this proof will be sufficient to the proof of Theorem 2.

As mentioned earlier, $X_1, \cdots, X_n$ denote the historical samples, and X is the present sample. According to the fundamental philosophy of EB procedure, $X_1, \cdots, X_n$, (X, θ) are mutually independent, each X_i possesses the same distribution as X given by (8). The distribution of θ is the prior G, while the conditional distribution of X for given θ is (1). For any EB estimator δ_n, its "over-all" Bayes risk is given by

$$\begin{aligned} R_G(\delta_n) &= E[\delta_n(X_1, \cdots, X_n, X) - \theta]^2 \\ &= E\{E[(\delta_n(X_1, \cdots, X_n, x) - \theta)^2 | X, \theta]\} \end{aligned} \tag{12}$$

and the Bayes risk of the Bayes estimator (7) is

$$R_G = E[\delta_G(X) - \theta]^2. \tag{13}$$

By definition, δ_n is said to be an a. o. EB estimate of θ with respect to (or under) the prior family $\mathscr{F}$ if

$$\lim_{n\to\infty} R_G(\delta_n) = R_G, \text{ for any } G \in \mathscr{F}. \tag{14}$$

By (12), (13) and dominant convergence theorem, one sees that in order to establish the a. o.-property of δ_n one must verify that

(a). for any $G \in \mathscr{F}$, there exists a function $M_G(x, \theta)$ not depending on n such that

$$E[(\delta_n(X_1, \cdots, X_n, X) - \theta)^2 | X, \theta] \leqslant M_G(X, \theta), \text{ for } n = 1, 2, \cdots, \tag{15}$$

$$E[M_G(X, \theta)] < \infty; \tag{16}$$

(b). for fixed $X \in \mathscr{X}$ and $\theta \in \Theta$

$$\lim_{n\to\infty} E[(\delta_n(X_1, \cdots, X_n, X) - \theta)^2 | X, \theta] = (\delta_G(X) - \theta)^2. \tag{17}$$

The following simple lemma plays a role in the proof of (a).

Lemma 1. *Suppose that* (Y_1, Y_2, Y_3) *obeys the multinomial law* $M(n; p_1, p_2, p_3)$, *where* $p_i > 0$, $i = 1, 2, 3$, *and* $\sum_1^3 p_i = 1$

$$P(Y_i = n_i, i = 1, 2, 3) = \frac{n!}{n_1! n_2! n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3},$$

then

$$E\left(\frac{Y_2}{1+Y_1}\right)^2 \leqslant 4(p_2 | p_1)^2 + 1. \tag{18}$$

Proof One has

$$E\left(\frac{Y_2}{1+Y_1}\right)^2 = \sum \frac{n!}{n_1! n_2! n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3} \left(\frac{n_2}{1+n_1}\right)^2 = \sum_{n_2 \leqslant 1} + \sum_{n_2 \geqslant 2} \triangleq I_1 + I_2. \tag{19}$$

Since $n_2/(1+n_1) \leqslant 1$ for $n_2 \leqslant 1$, one gets

$$I_1 \leqslant \sum \frac{n!}{n_1! n_2! n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3} = 1. \tag{20}$$

Also, since

$$\frac{2+n_1}{1+n_1} \leqslant 2, \quad \frac{n_2}{n_2 - 1} \leqslant 2 \text{ for } n_2 \geqslant 2,$$

we have

$$I_2 \leqslant \sum_{n_2 \geqslant 2} \frac{2+n_1}{1+n_1} \frac{n_2}{n_2-1} \frac{n!}{(n_1+2)!(n_2-2)! n_3!} p_1^{n_1+2} p_2^{n_2-2} p_3^{n_3} \left(\frac{p_2}{p_1}\right)^2$$

$$\leqslant 4\left(\frac{p_2}{p_1}\right)^2 \sum \frac{n!}{n_1! n_2! n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3} = 4\left(\frac{p_2}{p_1}\right)^2. \tag{21}$$

From (19)—(21), (18) follows. It is noticeable that the rhs. of (18) is free from n.

To prove (a), define

$$Y_1 = \#\{i : 1 \leqslant i \leqslant n, x_i = x\},$$

$$Y_2 = \#\{i : 1 \leqslant i \leqslant n, x_i = x+1\},$$

$$Y_3 = n - Y_1 - Y_2.$$

Then $(Y_1, Y_2, Y_3) \sim M(n; p_1, p_2, p_3)$, with

$$p_1 = f_G(x), \quad p_2 = f_G(x+1), \quad p_3 = 1-(p_1+p_2).$$

In this new notation, (9) has a form

$$\delta_n(x) = \delta_n(x_1, \cdots, x_n, x) = W(x)\frac{Y_2}{1+Y_1}, \tag{22}$$

hence

$$[\delta_n(x_1, \cdots, x_n, x)-\theta]^2 \leqslant 2W^2(x)\left(\frac{Y_2}{1+Y_1}\right)^2 + 2\theta^2. \tag{23}$$

Lemma 1 gives

$$E[(\delta_n(X_1, \cdots, X_n, X)-\theta)^2 | X, \theta] \leqslant 2W^2(X)[4f_G^2(X+1)/f_G^2(X)+1] + 2\theta^2 \triangleq M_G(X, \theta), \tag{24}$$

$M_G(X, \theta)$ so defined is free from n, and (15) holds. To show that (16) is true, we have to verify

$$E[W^2(X)f_G^2(X+1)/f_G^2(X)] < \infty, \tag{25}$$

$$E[W^2(X)] < \infty, \tag{26}$$

$$E(\theta^2) < \infty. \tag{27}$$

(27) is nothing else than (3). To verify (26), use assumption (10) and $G \in \mathscr{F}$, then

$$\begin{aligned} E[W^2(X)] &\leqslant W^2(0)+W^2(1)+\sum_{x=2}^{\infty} W^2(x)h(x)\int_{\Theta}\beta(\theta)\theta^x dG(\theta) \\ &= W^2(0)+W^2(1)+\sum_{x=2}^{\infty}\frac{h^3(x)}{h^2(x+1)h(x-2)}\int_{\Theta}h(x-2)\beta(\theta)\theta^x dG(\theta) \\ &\leqslant W^2(0)+W^2(1)+A\int_{\Theta}\theta^2\sum_{x=0}^{\infty}h(x)\beta(\theta)\theta^x dG(\theta) \\ &= W^2(0)+W^2(1)+A\int_{\Theta}\theta^2 dG(\theta) < \infty. \end{aligned}$$

This is (26). Now notice that

$$I \triangleq E[W^2(X)f_G^2(X+1)/f_G^2(X)] = \sum_{x=0}^{\infty} W^2(x)f_G^2(x+1)/f_G(x). \tag{28}$$

By Cauchy-Schwarz

$$\begin{aligned} f_G^2(x+1) &= h^2(x+1)\left(\int_{\Theta}\beta(\theta)\theta^{x+1}dG(\theta)\right)^2 \\ &= h^2(x+1)\left(\int_{\Theta}\beta^{1/2}(\theta)\theta^{x/2}\cdot\beta^{1/2}(\theta)\theta^{x/2+1}dG(\theta)\right)^2 \\ &\leqslant h^2(x+1)\int_{\Theta}\beta(\theta)\theta^x dG(\theta)\int_{\Theta}\beta(\theta)\theta^{x+2}dG(\theta) \\ &= \frac{W(x+1)}{W(x)}f_G(x)f_G(x+2). \end{aligned} \tag{29}$$

From (28), (29) we have

$$I \leqslant \sum_{x=0}^{\infty}\frac{h(x)}{h(x+2)}f_G(x+2) = \sum_{x=0}^{\infty}h(x)\int_{\Theta}\beta(\theta)\theta^{x+2}dG(\theta) = \int_{\Theta}\theta^2 dG(\theta) < \infty.$$

This proves (25), hence (16), and (a) is established.

To prove (b), use (22). Under (4) we have $p_1 = f_G(x) > 0$, $p_2 = f_G(x+1) > 0$.

Employing an inequality given by Hoeffding [8], we get

$$P(|Y_i/n-p_i|\geqslant n^{-1/3})=O(\exp(-n^{1/3}))=o(n^{-2}),\quad i=1,\ 2. \tag{30}$$

Also notice that for any $x_1, \cdots, x_n$

$$0\leqslant\delta_n(x_1, \cdots, x_n, x)\leqslant nW(x). \tag{31}$$

Write

$$A_1=\left(W(x)\frac{n(p_2-n^{-1/3})}{1+n(p_1+n^{-1/3})}-\theta\right)^2,$$

$$A_2=\left(W(x)\frac{n(p_2+n^{-1/3})}{1+n(p_1-n^{-1/3})}-\theta\right)^2.$$

By (22), (30), (31), we obtain for n sufficiently large

$$[1-o(n^{-2})]\min(A_1, A_2)\leqslant E[(\delta(X_1, \cdots, X_n, X)-\theta)^2|X, \theta]$$
$$\leqslant\max(A_1, A_2)+o(n^{-2})n^2W^2(x). \tag{32}$$

Evidently

$$\lim_{n\to\infty}A_1=\lim_{n\to\infty}A_2=(\delta_G(x)-\theta)^2. \tag{33}$$

From (32) and (33), (b) follows, and the proof of Theorem 1 is concluded.

Condition (10) is a mild restriction on h, which is satisfied in the three cases mentioned earlier. This gives another proof of Johns' result, and verifies that (9) gives an a. o. EB estimate in the case of negative-binomial and logarithm distributions without any restriction on the prior distribution.

Now turn to the proof of Theorem 2. This runs basically on the same line as that of Theorem 1. One needs only to notice that $\tilde{\delta}_n$ defined by (11) has a form

$$\tilde{\delta}_n(x)=\tilde{\delta}_n(x_1, \cdots, x_n, x)=W(x)\frac{Y_2'}{1+Y_1},$$

where Y_2' equals to Y_2 or 0, according to $Y_2\geqslant 2$ or otherwise. A glance at the proof of Lemma 1 will convince us that

$$E\left(\frac{Y_2'}{1+Y_1}\right)^2\leqslant 4(p_2/p_1)^2, \tag{34}$$

where p_1, p_2 are the same as defined earlier. Also it is easy to see that for $0<\varepsilon<p_2$ and n sufficiently large,

$$P(|Y_2'/n-p_2|\geqslant\varepsilon)=P(|Y_2/n-p_2|\geqslant\varepsilon).$$

So the proof of Theorem 1 can be carried here almost word by word, with the only diffrence that the constant 1, which appeared on the rhs. of (18), is missing on the rhs. of (34). Hence in the verification of (a) we have only to verify (25) and (27), which can be done under the sole condition $G\in\mathscr{F}$, without any restriction on h. Hence Theorem 2 is proved.

§ 3. Further Results

1. The case in which (4) is false.

If the present sample is x and $h(x+1)=0$, (9) and (11) become meaningless.

For such x one has

$$\delta_G(x)=h(x)\int_\Theta \beta(\theta)\theta^{x+1}dG(\theta)/f_G(x).$$

Generally speaking, $\int_\Theta \beta(\theta)\theta^{x+1}dG(\theta)$ is not determined by the marginal distribution of X, f_G. In such circumstances, there exists no a. o. EB estimate of θ. To be specific, we have the following

Theorem 3. *Denote*

$$D=\{x: x=0, 1, 2, \cdots; h(x)>0, h(x+1)=0\}$$

and let $\mathscr{F}^*$ be the prior distribution family. Then a necessary condition for the existence of an a. o. EB estimator of θ is the following: For any $G_i\in\mathscr{F}^*$, $i=1, 2$, satisfying

$$f_{G_1}(x)=f_{G_2}(x), \quad x=0, 1, 2, \cdots, \tag{35}$$

we must have

$$\int_\Theta \beta(\theta)\theta^{x+1}dG_1(\theta)=\int_\Theta \beta(\theta)\theta^{x+1}dG_2(\theta), \text{ for any } x\in D. \tag{36}$$

Proof Suppose that the condition is not satisfied, then some $G_i\in\mathscr{F}^*$, $i=1, 2$, can be found such that (35) is true, but there exists $x_0\in D$ for which

$$\int_\Theta \beta(\theta)\theta^{x_0+1}dG_1(\theta)\neq\int_\Theta \beta(\theta)\theta^{x_0+1}dG_1(\theta). \tag{37}$$

If in the contrary an a. o. EB estimator $\delta_n(x_1, \cdots, x_n, x)$ exists, then we have as $n\to\infty$

$$R_G(\delta_n)-R_G=\sum_{x=0}^{\infty}f_G(x)E[\delta_n(X_1, \cdots, X_n, x)-\delta_G(x)]^2\to 0 \tag{38}$$

for any $G\in\mathscr{F}^*$. Hence for $f_G(x)>0$ we have

$$\lim_{n\to\infty}E[\delta_n(X_1, \cdots, X_n, x)-\delta_G(x)]^2=0. \tag{39}$$

Take $G=G_1$ and $G=G_2$ respectively. By (35), under these two priors $X_1, \cdots, X_n, x$ have the same marginal distribution, and it follows easily from (39) that

$$\delta_{G_1}(x)=\delta_{G_2}(x), \text{ for any } x \text{ such that } h(x)>0.$$

But $x_0\in D$, so $h(x_0)>0$, hence $\delta_{G_1}(x_0)=\delta_{G_2}(x_0)$, which evidently contradicts (37). Hence the theorem is proved.

When the sample space is finite, generally the prior G cannot be uniquely determined by f_G, unless $\mathscr{F}^*$ is a parametric family. Hence in case $\mathscr{X}$ is finite, in general a. o. EB estimates cannot exist except for parametric prior families.

2. Lin's Estimate.

Consider the family (1). Take $N=\infty$ and suppose that (4) holds. Lin introduced in [6] the following EB estimator d_n for θ

$$d_n(x)=d_n(x_1, \cdots, x_n, x)=w(x)p_n(x+1)/\max(p_n(x), \delta_n),$$

where

$$p_n(x)=p_n(x_1, \cdots, x_n, x)=\frac{1}{n}\#\{i: 1\leqslant i\leqslant n, x_i=x\}$$

and δ_n is a constant of the order $n^{-1/3}$ (For simplicity we shall take $\delta_n=n^{-1/3}$ in the following). The main conclusion of [6] is that under certain conditions (see [6], (2.12)—(2.14)) one has $R_G(d_n)-R_G=O(n^{-1/3})$. As mentioned earlier, Chao found this conclusion to be incorrect. Our aim is to show that, by slightly modifying (and in the same time simplifying) Lin's conditions, it is possible to prove that d_n is a.o.

Theorem 4. *Suppose that there exists a constant A such that*

$$h^2(x)\leqslant Ah(x-1)h(x+1),\quad x=1,2,\cdots. \tag{40}$$

Then under the prior family (3), d_n is an a.o. EB estimate of θ.

Proof By simple manipulations, we obtain

$$\begin{aligned}(d_n(x)-\delta_G(x))^2\leqslant W^2(x)\,[&4n^{2/3}(p_n(x+1)-f_G(x+1))^2\\ &+6n^{2/3}(f_G(x+1)/f_G(x))^2(p_n(x)-f_G(x))^2\\ &+4(f_G(x+1)/f_G(x))^2I_{(p_n(x)<n^{-1/3})},\end{aligned} \tag{41}$$

where I_C denotes the indicator of the set C. As

$$E[p_n(i)-f_G(i)]^2=f_G(i)[1-f_G(i)]/n\leqslant f_G(i)/n \tag{42}$$

and $f_G(x)\geqslant 2n^{-1/3}$, we have for x fixed

$$\begin{aligned}P(p_n(x)<n^{-1/3})&\leqslant P(|p_n(x)-f_G(x)|\geqslant n^{-1/3})\\ &\leqslant n^{2/3}E[p_n(x)-f_G(x)]^2\leqslant n^{-1/3}.\end{aligned} \tag{43}$$

From (41) to (43) it follows that

$$R_G(d_n)-R_G=\sum_{x=0}^{\infty}f_G(x)E[d_n(x)-\delta_G(x)]^2\leqslant I_{1n}+I_{2n}+I_{3n}+I_{4n}, \tag{44}$$

where

$$I_{1n}=4n^{-1/3}\sum_{x=0}^{\infty}W^2(x)f_G(x)f_G(x+1),$$

$$I_{2n}=6n^{-1/3}\sum_{x=0}^{\infty}W^2(x)f_G^2(x+1),$$

$$I_{3n}=4n^{-1/3}\sum_{x=0}^{\infty}W^2(x)f_G^2(x+1)/f_G(x),$$

$$I_{4n}=4\sum_{x\in Q_n}W^2(x)f_G^2(x+1)/f_G(x),$$

and Q_n is the set

$$Q_n=\{x:\ f_G(x)<2n^{-1/3}\}.$$

By (29)

$$W^2(x)f_G^2(x+1)/f_G(x)\leqslant\frac{h(x)}{h(x+2)}f_G(x+2)=h(x)\int_{\Theta}\beta(\theta)\theta^{x+2}dG(\theta). \tag{45}$$

Using $G\in\mathscr{F}$, we get

$$\sum_{x=0}^{\infty}W^2(x)f_G^2(x+1)/f_G(x)\leqslant\int_{\Theta}\theta^2dG(\theta)<\infty. \tag{46}$$

Hence

$$I_{3n}=O(n^{-1/3}). \tag{47}$$

Since $0<f_G(x)<1$, it follows by (47) that

$$I_{2n}=O(n^{-1/3}). \tag{48}$$

Also for $x\geqslant 1$, using Cauchy-Schwarz and (40), we have

$$\begin{aligned} f_G^2(x) &= h^2(x)\left(\int_\Theta \beta(\theta)\theta^x dG(\theta)\right)^2 \\ &= h^2(x)\left(\int_\Theta \beta^{1/2}(\theta)\theta^{(x-1)/2}\beta^{1/2}(\theta)\theta^{(x+1)/2}dG(\theta)\right)^2 \\ &\leqslant h^2(x)\int_\Theta \beta(\theta)\theta^{x-1}dG(\theta)\int_\Theta \beta(\theta)\theta^{x+1}dG(\theta) \\ &= \frac{h^2(x)}{h(x-1)h(x+1)} f_G(x-1)f_G(x+1) \\ &\leqslant Af_G(x-1)f_G(x+1) \leqslant Af_G(x+1). \end{aligned}$$

Hence by (46)

$$\sum_{x=0}^{\infty} W^2(x)f_G(x)f_G(x+1) \leqslant W^2(0) + A\sum_{x=0}^{\infty} W^2(x)f_G^2(x+1)/f_G(x) < \infty$$

and

$$I_{1n} = O(n^{-1/3}). \tag{49}$$

Finally, by (45) we have

$$I_{4n} \leqslant 4\int_\Theta \theta^2 \sum_{x\in Q_n} h(x)\beta(\theta)\theta^x dG(\theta). \tag{50}$$

Since $f_G(x)>0$ for all x by (4), we see that

$$f_G(x) < 2n^{-1/3} \Rightarrow x\to\infty, \text{ as } n\to\infty. \tag{51}$$

This combined with the fact that

$$\sum_{x=0}^{\infty} h(x)\beta(\theta)\theta^x = 1 \tag{52}$$

gives us that for any $\theta\in\Theta$

$$\lim_{n\to\infty}\sum_{x\in Q_n} h(x)\beta(\theta)\theta^x = 0. \tag{53}$$

Now by (52), $G\in\mathscr{F}$, and the dominant convergence theorem, one finds that the rhs. of (50) tends to zero as $n\to\infty$, i. e.

$$\lim_{n\to\infty} I_{4n} = 0. \tag{54}$$

Combining (44), (47)—(49) and (54), we finally obtain

$$\lim_{n\to\infty} R_G(d_n) = R_G$$

for any $G\in\mathscr{F}$, and the proof of Theorem 4 is concluded.

The condition (40), similar in nature to (10), is satisfied for most common distributions.

Also we note that the two conditions (i. e. (40) and (3)) possess the pleasing feature that each one involves only a single element (i.e., h or G) of the problem.

References

[1] Robbins, H., An empirical Bayes approach to statistics, *Proc. Third Berkeley Symp. Math. Statist. Prob. Univ. California Press*, **1** (1955), 157—163.

[2] Robbins, H., The empirical Bayes approach to statistical decision problems, *Ann. Math. Statist.*, **35** (1964), 1—20.

[3] Maritz, J. S., Empirical Bayes Methods, Methuen Monograph, Methuen London, 1970.

[4] Johns, M. V. Jr., Contributions to the theory of non-parametric empirical Bayes procedures in statistics, *Columbia Univ. Dissertation*, (1956).

[5] Johns, M. V. Jr., Nonparametric empirical Bayes procedures, *Ann. Math. Statist.*, **28** (1957), 649—669.

[6] Lin, P. E., Rates of convergence in empirical Bayes estimation problems, *Discrete case. Ann. Inst. Statist. Math.*, **24** (1972), 319—325.

[7] Chao Lincheng, EB estimates for a class of discrete distributions with convergence rates (in chinese) *J. Math. Research & Exposition*, **1** (1981), 59—69.

[8] Hoeffding, W., Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.*, **58** (1863), 11—30.