数理解析研究所講究録
507巻 1983年 117-134

RATE $O((n^{-1} \log n)^{\frac{1}{2}})$ OF RISK CONVERGENCE IN THE EMPIRICAL BAYES BOOTSTRAP ESTIMATION : CASE OF RETRACTED DISTRIBUTIONS

by Yoshiko Nogami

University of Tsukuba

1. Introduction.

The empirical Bayes (EB) problem involved here is the same one as stated i Nogami (1983). We use the same notational conventions. We often let P(h) or P(h(w)) denote $\int h(w)dP(w)$.

For the uniform distribution $U[\theta,\theta+1)$ where $\theta\in(-\infty,\infty)$ Fox (1970) exhibited a distribution-valued Lévy consistent estimate $\hat{G}_n$ for G. In the empirical Baye problem where the θ_i are iid with a common unknown prior distribution G, Fox (1 the second Remark after Theorem 3.1) indicated (without rates) a convergence of the expected risks to the Bayes envelope R(G)(=R) for a bootstrap decision rule based on component procedures Bayes vs $\hat{G}_n$. In this paper we consider the squar error loss estimation and exhibit two distribution-valued Lévy consistent estima $\hat{G}_n$ and $\hat{\hat{G}}_n$ for G involving the family $p(f)$ of retracted distributions into the interval $[\theta,\theta+1)$. We furthermore obtain two bootstrap estimates $\hat{\phi}_n$ and $\hat{\hat{\phi}}_n$ (for θ) with a convergence rate $O((n^{-1}\log n)^{\frac{1}{2}})$ for both $R(\hat{\phi}_n,G)-R$ and $R(\hat{\hat{\phi}}_n,G)-R$.

This paper is essentially dealt with an application of the compound estima in Nogami (1982) for the empirical Bayes problem. Differently from Nogami (198 we here exhibits two kernel type EB estimates and the proof of Lemma 4.5 is slightly simpler than that in Nogami (1982, Proof of Lemma 2.5).

In Section 2 we obtain an upper bound of $R(\hat{t}_n,G)-R$ for any bootstrap estimate $\hat{t}_n$ (for θ). In Section 3 we introduce two EB estimates $\hat{\phi}_n$ and $\hat{\hat{\phi}}_n$. We shall obtain, in Section 4, a rate $O((n^{-1}\log n)^{\frac{1}{2}})$ for $R(\hat{\phi}_n,G)-R$ and, in Section 5, the same rate for $R(\hat{\hat{\phi}}_n,G)-R$.

2. An Upper Bound for $R(\hat{t}_n,G)-R$.

Let $\Omega=(-\infty,\infty)$. Assume $(0<)m^{-1}\le f\le 1$ throughout the paper. Let $\hat{G}_n$ be any distribution-valued random variable which is an estimate of the unknown prior distribution G, obtained from $X_1,\dots,X_n,X\doteq X_{n+1}$ where each $X_i\sim P\doteq\int P_\theta\, dG(\theta)$ for $P_\theta\varepsilon p(f)$ and $X\sim P_\theta(\theta\doteq\theta_{n+1})$. Define an estimate for θ by

$$(1.1)\qquad \hat{t}_n(X)=\int_{X'+}^{X}\theta\, q(\theta)\, d\hat{G}_n(\theta)/\int_{X'+}^{X} q(\theta)\, d\hat{G}_n(\theta).$$

When $R(\hat{t}_n,G)$ and R are finite, we have

$$(1.2)\qquad (0<)\ R(\hat{t}_n,G)-R = \underset{\sim}{E}(\hat{t}_n(X)-\phi_G(X))^2$$

where $\phi_G(X)$ is the Bayes estimate vs G at X of the form

$$(1.3)\qquad \phi_G(X) = \int_{X'+}^{X}\theta\, q(\theta)\, dG(\theta)/\int_{X'+}^{X} q(\theta)\, dG(\theta)$$

where the affix + means the right limit and hereafter we abbreviate it.

To get a bound we need to set the following assumptions on G:

A i) For some $(0<)\varepsilon<1$ and a positive constant B

$$\sup_s \sup_{y\varepsilon(s,s+\varepsilon)}^{a.e.} |G'(y)|\le B\ (<+\infty).$$

A ii) For a positive constant C,

$$\sup_{(0<)\varepsilon<1/3} \int\{G(y+1-\varepsilon)-G(y+\varepsilon)\}^{-1}\, dy \le C\ (<+\infty).$$

Lévy distance for two distribution functions F and H of random variables (cf. Feller (1971), p.285) is defined by

$$L(F,H) = \inf\{\varepsilon>0:\ F(y-\varepsilon)-\varepsilon\le H(y)\le F(y+\varepsilon)+\varepsilon \text{ for all } y\varepsilon(-\infty,\infty)\}.$$

Remark that the infimum in the definition attains (see Appendix of Nogami (1975)).

In this section we shall find an upper bound for (1.2). Since by (1.1) and (1.3) $X'<\hat{t}_n(X)$, $\phi_G(X)\le X$, we have $|\hat{t}_n-\phi_G|\le 1$. Hence, for any $0<\varepsilon<1$,

$$E(\hat{t}_n(X)-\phi_G(X))^2 \le E[L(G,\hat{G}_n)>\varepsilon] + E\{(\hat{t}_n(X)-\phi_G(X))^2[L(G,\hat{G}_n)\le\varepsilon]\}. \tag{1.4}$$

The main development of this section is Lemma 1.6 in which we show that two terms in the rhs(1.6) below are both no more than a constant times ε^2 for $0<\varepsilon<1$ which in turn the second term of rhs(1.4) is no more than a constant times ε^2. To reach the bound in Lemma 1.6 we shall use Lemma A.2 of Singh (1974), Proposition A of Nogami (1975) and Lemma A.3 of Nogami (1975). Hence, we shall state them here beforehand.

Lemma 1.1. (R. S. Singh (1974)) Let y, z and B be in $(-\infty,\infty)$ with $z\neq 0$ and $B>0$. If Y and Z are two real valued random variables, then for every $\gamma>0$

$$E(|(Y/Z) - (y/z)|\wedge B)^\gamma \le 2^{\gamma+(\gamma-1)^+}|z|^{-\gamma}\{E|y-Y|^\gamma + ((y/z)^\gamma + 2^{-(\gamma-1)^+}B^\gamma)E|z-Z|^\gamma\}$$

where $a^+=a\vee 0$.

Lemma 1.2. (Proposition A of Nogami (1975)) Let I be a finite interval and let F_I be the retraction of a distribution function F into the closed interval $[F(a+),F(b+)]$. Then,

$$L(F_I,G_I) \le |(F-G)(a+)|\vee|(F-G)(b+)|\vee L(F,G)$$

where we use + on the line to denote the right limit.

For Lemma 1.3 we need to introduce the following definition:

Definition 1.1. With h, a function defined on a real interval I, the modulu of continuity of h is the function given by

$$\alpha(\varepsilon) = \sup\{h(\omega_1)-h(\omega_2): \omega_1, \omega_2\varepsilon I, |\omega_1-\omega_2|<\varepsilon\}$$

for every $\varepsilon>0$.

In Lemma 1.3 below, the natural generalization of the inverse probability integral transformation is used to develop bounds for the same difference of integrals in Lemma 8' of Oaten (1969, Appendix) without partitioning as in Oaten's proof.

Lemma 1.3. (Nogami (1975, Lemma A.3)) Let I be a finite interval $\{a,b\}$ supporting finite measures μ and ν and let h be measurable on I into a finite interval $[c,d]$. Let F and G be distribution functions inducing μ and ν with $F(a-)\vee G(a-)\le F(b+)\wedge G(b+)$. Then $|\int h\, d(\mu-\nu)|$ has the following family of bounds

$$\frac{d-c}{2}\{|(F-G)(a-)|+|(F-G)(b+)|\}$$
$$+\ \alpha(L(F,G)+)\{F(b+)\wedge G(b+)-F(a-)\vee G(a-)\}+\frac{d+c}{2}|\mu(I)-\nu(I)|.$$

We first apply Lemma 1.1 of Nogami (1981) for $\hat{t}_n(X)$ and $\phi_G(X)$. (This lemma is essentially the Fubini Theorem.) Since by this lemma $\phi_G(X) = X - \{\int_0^1\int_{X'}^{X'+t} q(\theta)\, dG(\theta)\, dt/\int_{X'}^{X} q(\theta)\, dG(\theta)\}$,

$$(1.5)\qquad |t_n(X)-\phi_G(X)|^2 = |\{\int_0^1\int_{X'}^{X'+t} q(\theta)\, dG(\theta)\, dt/\int_{X'}^{X} q\, dG\}$$
$$-\{\int_0^1\int_{X'}^{X'+t} q(\theta)\, d\hat{G}_n(\theta)\, dt/\int_{X'}^{X} q\, d\hat{G}_n(\theta)\}|^2.$$

Letting rhs(1.5) $= |(v/w)-(V/W)|^2$ and * denoting conditioning on X and $\{L(G,\hat{G}_n)\le\varepsilon\}$ we apply Lemma 1.1.

(1.6) the second term of rhs(1.4) $\le 8E\{|w|^{-2}\underset{\sim}{E}_*|v-V|^2\}+12E\{|w|^{-2}\underset{\sim}{E}_*|w-W|^2\}$.

To get a bound for the second term of rhs(1.4) we bound rhs(1.6). To do so, we shall furnish the following two lemmas:

Lemma 1.4. For any $0<\varepsilon<1/3$ and with $z=X'+t$ for $0\le t\le 1$,

$$(1.7)\qquad \underset{\sim}{E}\{|(G-\hat{G}_n)(z)|/w^2\} \le C(1+2B)\varepsilon$$

where B and C are positive constants in A i) and A ii).

Proof.) By the definition of $L(G,\hat{G}_n)$ and by the fact that the infimum in the definition of Lévy distance is attained, we have

(1.8) $$|(G-\hat{G}_n)(z)|\,[L(G,\hat{G}_n)\le\varepsilon] \le \varepsilon + G(z+\varepsilon) - G(z-\varepsilon).$$

Thus,

(1.9) $$\text{lhs}(1.7) \le \varepsilon E(w^{-2}) + E[\{G(z+\varepsilon)-G(z-\varepsilon)\}/w^2].$$

But, since the second term of rhs(1.9)$=\int\{G(x'+t+\varepsilon)-G(x'+t-\varepsilon)\}f(x)/w\,dx$, by the Taylor expansion applied to $G(z+\varepsilon)-G(z-\varepsilon)$ and by two applications of $f \le 1$ for the first inequality and an application of Aii) for the second inequality,

(1.10) $$\text{the second term of rhs}(1.9) = \int\{\int_{x'+t-\varepsilon}^{x'+t+\varepsilon}|G'(y)|\,dy f(x)/w\}\,dx$$
$$\le 2\varepsilon B\int\{G(x)-G(x')\}^{-1}\,dx \le 2\varepsilon BC.$$

On the other hand, since $E(w^{-2})=\int w^{-1} f(x)\,dx$, by two applications of $f\le 1$ an an application of Aii) we have

(1.11) $$E(w^{-2}) \le C.$$

This and (1.10) gives us the asserted lemma.

Lemma 1.5. For any $0<\varepsilon<1/3$ and with $z=X'+t$ and $u=X'+s$ for $0\le t,\ s\le 1$,

(1.12) $$EE_{\underset{\sim}{*}}\{|((G-\hat{G}_n)(z))((G-\hat{G}_n)(u))|/w^2\} \le \varepsilon^2 C(1+2B)^2.$$

Proof.) Let $h]_a^b=h(b)-h(a)$ until the end of the proof. As in the proof of Lemma 1.4, for $0<\varepsilon<1/3$,

(1.13) $$|(G-\hat{G}_n)(z)|\,|(G-\hat{G}_n)(u)|\,[L(G,\hat{G}_n)\le\varepsilon] \le \varepsilon^2 + \varepsilon\{G]_{z-\varepsilon}^{z+\varepsilon} + G]_{u-\varepsilon}^{u+\varepsilon}\} + G]_{z-\varepsilon}^{z+\varepsilon}$$

Thus, in the similar fashion to (1.9) and by applying (1.10) and (1.11), we hav

(1.14) $$\text{lhs}(1.12) \le \varepsilon^2 C + 2\varepsilon^2 BC + E\{G]_{z-\varepsilon}^{z+\varepsilon}\ G]_{u-\varepsilon}^{u+\varepsilon}/w^2\}.$$

But, again, applying the similar method to obtaining (1.10) leads to

$$(\text{the third term of the extreme rhs}(1.14)) \le 4\varepsilon^2 B^2 C$$

which completes the proof.

We shall now find an upper bound for both terms in the rhs(1.6).

Lemma 1.6. For $0 \le t \le 1$,

$$(1.15) \qquad E\underset{\sim}{E}_*\{(\int_{X'}^{X'+t} q(\theta)\, d\hat{G}_n(\theta) - \int_{X'}^{X'+t} q(\theta)\, dG(\theta))^2/w^2\}$$

$$\le m^2 C\varepsilon^2\{m + 2(m+1)(1+2B)\}^2.$$

Proof.) Let the numerator in the curly bracket be $|Y-y|^2$. By letting $I=(X',X'+t]$, define by G_I the retraction of G into the closed interval $[G(X'),G(X)]$. Then, by Lemma 1.2, $L(G_I,\hat{G}_I) \le L(G,\hat{G}_n) \vee S \vee T$ where $S=|(G-\hat{G}_n)(X')|$ and $T=|(G-\hat{G}_n)(X'+t)|$. Thus,

$$(1.16) \qquad L(G,\hat{G}_n) \le \varepsilon \implies L(G_I,\hat{G}_I) \le \varepsilon \vee S \vee T \ (\doteq \lambda).$$

By applying Lemma 1.3 with $h(\theta)$, the retraction of $q(\theta)$ to I, and weakening the resulted bound,

$$(1.17) \qquad L(G_I,\hat{G}_I) \le \lambda \implies |Y-y|^2 \le \{\alpha(\lambda+) + m(S + T)\}^2.$$

To bound $\alpha(\lambda+)$, pick ω_1, $\omega_2 \varepsilon I$ such that $0<\omega_2-\omega_1<\lambda$. Now, by the definition of h and q, $h(\omega_2)-h(\omega_1)=q(\omega_2)-q(\omega_1)=q(\omega_2)q(\omega_1)(\int_{\omega_1}^{\omega_2} f(s)\, ds - \int_{\omega_1+1}^{\omega_2+1} f(s)\, ds)$ and since $q \le m$ and $f \le 1$, $|h(\omega_2)-h(\omega_1)| \le m^2\lambda$. Thus, by Definition 1.1, $\alpha(\lambda) \le m^2\lambda$. Thus, the same bound applies for $\alpha(\lambda+)$. Therefore, by (1.17) and weakening the bound

$$(1.18) \qquad L(G_I,\hat{G}_I) \le \lambda \implies |Y-y|^2 \le \{m^2\varepsilon + m(m+1)(S+T)\}^2.$$

Thus, in view of (1.16) and (1.18),

$$\text{lhs}(1.15) \le m^4\varepsilon^2 E(w^{-2}) + 2m^3(m+1)\varepsilon\{E(S/w^2)+E(T/w^2)\} + m^2(m+1)^2\{E(S^2/w^2)$$

$$+ 2E(ST/w^2) + E(T^2/w^2)\}.$$

Thus, applying (1.11), Lemma 1.4 twice and Lemma 1.5 three times. leads to the asserted bound.

From (1.4), (1.6) and Lemma 1.6 we obtain

Theorem 1.1. If $P_\theta \varepsilon \mathcal{P}(f)$ for $(0<)m^{-1} \leq f \leq 1$ and G satisfies the assumptions Ai) and Aii), then for $0<\varepsilon<1/3$,

$$R(\hat{t}_n, G) - R < \underset{\sim}{E}[L(G,\hat{G}_n)>\varepsilon] + c_0\varepsilon^2$$

where $c_0 = 20m^2 C\{m+2(m+1)(1+2B)\}^2$.

3. EB bootstrap estimates $\hat{\phi}_n$ and $\hat{\hat{\phi}}_n$.

We first construct normalized (but not monotonized) estimates $\tilde{G}$ and G^* of an unknown prior G and then get a distribution-valued (monotonized) estimates $\hat{G}_1$ and $\hat{\hat{G}}_n$ for G. We then play Bayes vs $\hat{G}_n$ and $\hat{\hat{G}}_n$ to get $\hat{\phi}_n$ and $\hat{\hat{\phi}}_n$, respectively.

Let $\Omega=(-\infty,\infty)$. Let Q be the distribution function defined by

$$Q(y) = \int_{-\infty}^{y} q(\theta)\, dG(\theta) \quad \text{for every } y.$$

Since $q \geq 1$ and q is the density of Q wrt G, it follows by Theorem 32.B of Halmos (1950) that

$$G(y) = \int_{-\infty}^{y} (q(\theta))^{-1}\, dQ(\theta). \tag{3.1}$$

This form gives an estimate of G by inserting an estimate for Q in place of Q.

Letting $p(y)=\int p_\theta(y)\, dG(\theta)$, we have by the definition of p_θ that $p(y) = f(y)(Q(y)-Q(y'))$ and by a telescopic series,

$$Q(y) = \Sigma_{r=0}^{\infty} \frac{p(y-r)}{f(y-r)}. \tag{3.2}$$

We remark that if the r-th term of rhs(3.2) is nonzero, then

$$r-1 \leq \text{range of } \{\theta_1, \theta_2, \ldots, \theta_n\}. \tag{3.3}$$

We shall first exhibit $\hat{\phi}_n$. To estimate Q, we introduce a kernel function which is a nonnegative bounded-variation function vanishing outside the interva (0,1) and satisfying

(3.4) $$0 \le K(y) \le 1 \quad \text{and} \quad \int_0^1 K(y)\,dy = 1.$$

Let us first estimate $p(y)$ by $\hat{p}(y)=(nh)^{-1}\Sigma_{j=1}^{n}K(h^{-1}(X_j-y))$ where $0<h=h_n(<1)\to 0$ as $n\to+\infty$. In view of (3.2), $Q(y)$ is estimated by

(3.5) $$\tilde{Q}(y) = \Sigma_{r=0}^{\infty} \frac{\hat{p}(y-r)}{f(y-r)}.$$

Thus, by (3.1) we estimate $G(y)$ by

(3.6) $$\overline{\tilde{W}}(y) = \int_{-\infty}^{y}(q(\theta))^{-1}\,dQ(\theta).$$

We use a raw estimate of the empiric distribution G_n of n parameters $\theta_1,\dots,\theta_n$ from n populations as an estimate of G. Let $F_n(y)=n^{-1}\Sigma_{j=1}^{n}[X_j\le y]$. Since $F_n(y)\le G_n(y)\le F_n(y+1)$, we get a raw estimate

(3.7) $$\tilde{G}_n(y) = (F_n(y)\vee\overline{\tilde{W}}(y))\wedge F_n(y+1).$$

We let $\delta=N^{-1}$, N being a positive integer depending on n, and consider the following grid points on the real line: $\dots < -2\delta<-\delta<0<\delta<2\delta< \dots$. We finally estimate G at y by

(3.8) $$\hat{G}_n(y) = \sup\{\tilde{G}_n(j\delta):\ j\delta\le y,\ j=0,\ \pm1,\ \pm2,\dots\}.$$

Our EB bootstrap estimate $\hat{\phi}_n(X)$ at X (for θ) is of form

(3.9) $$\hat{\phi}_n(X) = \int_{X'}^{X} \theta q(\theta)\,d\hat{G}_n(\theta)/\int_{X'}^{X} q(\theta)\,d\hat{G}_n(\theta).$$

Similarly, we shall exhibit $\hat{\hat{\phi}}_n$. To estimate Q, we use above kernel function without the assumption $\int_0^1 K(y)\,dy = 1$ in (3.4). In view of (3.2) $Q(y)$ is estimated by

(3.10) $$Q^*(y) = \Sigma_{r=0}^{\infty}(nh)^{-1}\Sigma_{j=1}^{n}\{K(h^{-1}(X_j-y+r))/f(X_j)\}.$$

Thus, by (3.1) we estimate $G(y)$ by

(3.11) $$\overline{W}^*(y) = \int_{-\infty}^{y}(q(\theta))^{-1}\,dQ^*(\theta).$$

Hence, as above we get a raw estimate

(3.12) $$G_n^*(y) = (F_n(y) \vee \overline{W}^*(y)) \wedge F_n(y+1)$$

and finally estimate G at y by

(3.13) $$\hat{\hat{G}}_n(y) = \sup\{G_n^*(j\delta): j\delta \le y,\ j=0,\ \pm 1,\ \ldots\}.$$

The second EB estimate (for θ) $\hat{\hat{\phi}}_n(X)$ at X is of form

(3.14) $$\hat{\hat{\phi}}_n(X) = \int_{X'}^{X} \theta q(\theta)\, d\hat{\hat{G}}_n(\theta) / \int_{X'}^{X} q(\theta)\, d\hat{\hat{G}}_n(\theta).$$

4. A Rate $O((n^{-1}\log n)^{\frac{1}{2}})$ for $R(\hat{\phi}_n, G) - R$.

To get a rate of convergence for $R(\hat{\phi}_n, G) - R$ we use the bound of Theorem 1.1 we shall get an upper bound for $\underset{\sim}{E}[L(G, \hat{G}_n) > \varepsilon]$ (forthcoming Lemma 4.6). This bound is essentially given by bounding $1 - \underset{\sim}{E}(\{G(y-\varepsilon) - \varepsilon \le \tilde{G}_n(y) \le G(y+\varepsilon) + \varepsilon\})$ (forthcoming Lemma 4.5). Therefore, main part in this section is Lemma 4.5. But, since we will apply Theorem 2 of Hoeffding (1963) for its proof, we need to get the bounds for $\underset{\sim}{E}\tilde{\overline{W}}(y)$; with a positive constant b_1, $G(y+h) + b_1 h$ for an upper boun and $G(y) - b_1 h$ for a lower bound (Lemma 4.4). To do so we shall furnish Lemmas 4.1, 4.2 and 4.3.

Throughout this section we assume $\Omega = [c,d]$ where $-\infty < c < d < +\infty$. In addition tc the assumption on f in Section 2 we also assume that 1/f satisfies the Lipshitz condition:

(4.1) $$\sup\{(v-u)^{-1} |(f(v))^{-1} - (f(u))^{-1}| : u < v\} \le M$$

for a finite nonnegative constant M. This is equivalent to

(4.2) $$|(f(s)/f(t)) - 1| \le M|s-t|.$$

By the definition (3.7) of $\tilde{\overline{W}}$, $\tilde{\overline{W}} = n^{-1}\Sigma_{j=1}^{n} \tilde{W}_j$ where for each j

$$(4.3)\qquad h\,\tilde{W}_j = \Sigma_{r=0}^{\infty} \int_{-\infty}^{y} (q(t))^{-1}\, d_t\{K(h^{-1}(X_j-t+r))/f(t-r)\}$$

where the subscript t in d_t denotes the variable of integration. To find bounds for $\underset{\sim}{E}\tilde{\tilde{W}}(y)$ we shall find an upper and a lower bound for $\tilde{W}_j$. Fix j and use the corresponding notations without subscript j until the end of the proof of Lemma 4.4. Hereafter, we abbreviate g(b)-g(a) to $g]_a^b$ until the end of the proof of Lemma 4.5.

<u>Lemma 4.1.</u> For $0<h<1$,

$$(4.4)\qquad |h(q(\theta))^{-1}P_\theta\tilde{W} - I(y) + S(y)| \le Mh^2(1+(M\wedge 1)N)$$

where $N\doteq d-c+2$, $I(y)=(h/q(y))\int_0^1 K(u)[\theta-y\le hu]\,du$ and $S(y)=h\int_{-\infty}^{y}\int_0^1 K(v)[\theta-t\le hv]\,dv\; f]_t^{t+1}dt$.

<u>Proof.</u>) Because a function satisfying the Lipshitz condition is absolutely continuous (cf. Ryden (1968), p.108)) and 1/q is clearly absolutely continuous, $1/f(.-r)$, 1/q and K are all of bounded variation in the definition (4.3) of $\tilde{W}$. Applying integration by parts (Saks (1937), Theorem III. 14. 1) and using $d(q(t))^{-1}=f]_t^{t+1}dt$ gives us that

$$(4.5)\qquad h\tilde{W} = \Sigma_{r=0}^{\infty}\frac{K(h^{-1}(X-y+r))}{f(y-r)q(y)} - \int_{-\infty}^{y}\Sigma_{r=0}^{\infty}\frac{K(h^{-1}(X-t+r))}{f(t-r)}\, f]_t^{t+1}dt.$$

Taking expectation wrt P_θ and multiplying $(q(\theta))^{-1}$ on both sides gives

$$(4.6)\qquad h(q(\theta))^{-1}P_\theta\tilde{W} = (q(y))^{-1}\Sigma_{r=0}^{\infty}\int_{y-r}^{y-r+h}[\theta\le s<\theta+1]K(h^{-1}(s-y+r))\frac{f(s)}{f(y-r)}\,ds$$

$$- \int_{-\infty}^{y}\Sigma_{r=0}^{\infty}\int_{t-r}^{t-r+h}[\theta\le s<\theta+1]K(h^{-1}(s-t+r))\frac{f(s)}{f(t-r)}\,ds\; f]_t^{t+1}dt.$$

Changing the variables $u=h^{-1}(s-y+r)$ and $v=h^{-1}(s-t+r)$ for the integrals in the first term and the second term, respectively, we get

(4.7) $$\text{rhs}(4.6) = h(q(y))^{-1}\Sigma_{r=0}^{\infty}\int_0^1 K(u)[\theta-(y-r)\le hu<\theta+1-(y-r)]\frac{f(y-r+hu)}{f(y-r)}\,du$$

$$- h\int_{-\infty}^{y}\Sigma_{r=0}^{\infty}\int_0^1 K(v)[\theta-(t-r)\le hv<\theta+1-(t-r)]\frac{f(t-r+hv)}{f(t-r)}\,dv\; f]_t^{t+1}dt$$

Since by (4.2) $|(f(t-r+hv)/f(t-r))-1|\le Mh$ for $0\le v\le 1$, applying telescopic series gives

(4.8) $$|h(q(\theta))^{-1}P_\theta\tilde{W} - I(y) + S(y)|\le Mh^2\{(q(y))^{-1}\int_0^1 K(u)[\theta-y\le hu]\,du$$

$$+ \int_{-\infty}^{y}\int_0^1 K(v)[\theta-t\le hv]\,dv|f]_t^{t+1}|dt\}.$$

Since $(0\le)K\le 1$, $[\theta-t\le hv]\le[\theta-h\le t]$, $|f|_t^{t+1}|\le M\wedge]$ and $(q(y))^{-1}\le 1$, rhs(4.8) $\le Mh^2(1+(M\wedge 1)(y-\theta+h))$ which proves the lemma.

Following Lemma 4.2 is straightforward, so we omit the proof.

Lemma 4.2. For $I(y)$ defined in Lemma 4.1,

(4.9) $$h(q(\theta))^{-1}[\theta\le y] \le I(y) \le h(q(\theta))^{-1}[\theta\le y+h].$$

Lemma 4.3. For $S(y)$ defined in Lemma 4.1

(4.10) $$|S(y) - h[\theta\le y]((q(y))^{-1} - (q(\theta))^{-1})| \le (M\wedge 1)h^2.$$

Proof.) Since by (3.4) $\int_0^1 K(u)du=1$,

(4.11) $$S(y) = h\{\int_{-\infty}^{y}[\theta\le t]f]_t^{t+1}dt + \int_{-\infty}^{y}[\theta-h\le t<\theta]\int_{h^{-1}(\theta-t)}^{1} K(v)dv\; f]_t^{t+1}dt\ .$$

Since $\int_{-\infty}^{y}[\theta\le t]\; f]_t^{t+1}dt = [\theta\le y]\{1/q(y) - 1/q(\theta)\}$, using $|f]_t^{t+1}|\le M\wedge 1$ and $(0\le)K(\cdot)\le$ and weakening the resulted bound leads to the asserted bound.

We are now ready to prove the following lemma:

Lemma 4.4. For any $y\varepsilon[c-1,d+1]$,

(4.12) $$G(y) - b_1h \le \underset{\sim}{E}\tilde{\tilde{W}}(y) \le G(y+h) + b_1h$$

where $b_1=\{(M\wedge 1)+M(1+(M\wedge 1)N)+h\wedge((1+h)M\wedge m)\}m$.

Proof.) From Lemmas 4.1, 4.2 and 4.3 and after a little computation,

(4.13) $$[\theta\le y]/q(\theta)-b_2h \le P_\theta\tilde{W}/q(\theta)\le[\theta\le y+h]/q(\theta)+[y\le\theta<y+h]\{(q(y))^{-1}-(q(\theta))^{-1}\}+b_2$$

where $b_2=\{(M\wedge 1)+M(1+N(M\wedge 1))\}$. From the definition and $f\le 1$,

$$(4.14)\qquad [y\le\theta<y+h]\,|(q(y))^{-1}-(q(\theta))^{-1}| \le h.$$

On the other hand, $\text{lhs}(4.14)=[y\le\theta<y+h]\{|(q(\theta)/q(y))-1|/q(\theta)\}\le((1+h)M)\wedge m$ where the last inequality follows because $q^{-1}\le 1$ and the Lipshitz condition (4.2). Thus, putting these together gives $\text{lhs}(4.14)\le h\wedge((1+h)M\wedge m)$. Therefore, noticing $q\le m$ leads to

$$[\theta\le y] - b_2h \le \underset{\sim}{E}_\theta\, \overline{\tilde{W}} \le [\theta\le y+h] + b_1h$$

where $\underset{\sim}{E}_\theta$ means the expectation wrt n products of the conditional probability distribution P_θ and $b_1=b_2+1\wedge((1+h)M\wedge m)h^{-1})$. Taking the expectation wrt G gives the asserted bound.

Following Lemma 4.5 becomes a direct generalization of Lemma 3.1 of Fox (1970) in the empirical problem in the sense that if $f\equiv 1$, then $m=1$ and $M=0$ and hence we get his bound $2\exp(-2nh^2\varepsilon^2)$.

<u>Lemma 4.5.</u> If $0<h\le\varepsilon\le 1$, then for each y

$$(4.15)\qquad 1 - \underset{\sim}{E}(\{G(y-\varepsilon)-\varepsilon\le\tilde{G}(y) \le G(y+\varepsilon)+\varepsilon\}) \le 2\exp\{-\frac{2nh^2((\varepsilon-b_1h)^+)^2}{(b_3+2b_4)^2}\}$$

where $b_3=(1+(N+1)M)\wedge m(N+1)$ and $b_4=(m(M\wedge 1))\wedge((N+1)(1+MN/2))$.

<u>Proof.)</u> For any y, it is sufficient to prove the lemma for the raw estimate $\overline{\tilde{W}}$, for if $G(y-\varepsilon)-\varepsilon \le \overline{\tilde{W}}(y) \le G(y+\varepsilon)+\varepsilon$, it is easily checked that $G(y-\varepsilon)-\varepsilon \le \overline{\tilde{W}}(y)\wedge F_n(y+1)\le\tilde{G}(y) \le \overline{\tilde{W}}(y)\vee F_n(y) \le G(y+\varepsilon)+\varepsilon$ a.e. $(GP_\theta)^n$.

As in the proof of Lemma 3.1 of Fox (1970) we shall apply Theorem 2 of Hoeffding (1963). To do so we shall use the bounds for $\underset{\sim}{E}\overline{\tilde{W}}(y)$ in Lemma 4.4 and furthermore need to get an upper and a lower bound of $\tilde{W}_j$ for each j.

In the definition (4.5) of $\tilde{W}$ since $K(h^{-1}(X-y+r))\ge 0$ when $X-h\le y-r<X$, there is at most one positive term in the first term of rhs(4.5). Applying (4.2) and the

fact that $r \le N$ and $K \le 1$ gives that

$$0 \le \text{the first term of rhs}(4.5) \le 1+(N+1)M.$$

Since $f^{-1} \le m$, $q^{-1} \le 1$ and $K \le 1$, we also have the first term of rhs$(4.5) \le m(N+1)$. Hence, putting these two bounds together gives

(4.16) $$0 \le \text{the first term of rhs}(4.5) \le (1+(N+1)M) \wedge m(N+1).$$

In the same way,

(4.17) $$\text{the second term of the rhs}(4.5)$$
$$= \int_{-\infty}^{y} \Sigma_{r=0}^{\infty} [X+r-h \le t < X+r] \frac{K(h^{-1}(X-t+r))}{f(t-r)} \, f]_t^{t+1} dt.$$

Since the summation in r has at most one positive term and since $|f]_t^{t+1}| \le M \wedge 1$, $f^{-1} \le m$ and $K \le 1$, we have

(4.18) $$|(4.17)| \le hm(M \wedge 1).$$

On the other hand, by changing a variable $u = h^{-1}(X-t+r)$

(4.19) $$(4.17) = h\Sigma_{r=0}^{\infty} \int_{-\infty}^{h^{-1}(X-y+r)} [0<u\le 1] \frac{K(u)}{f(X-hu)} (f(X+r-hu+1)-f(X+r-hu))du$$

Since $f(X+r-hu+1)/f(X-hu) \le 1+(r+1)M$ and $K \le 1$, extending the range of integratio leads to

(4.20) $$|(4.17)| \le h(N+1)(1+MN/2).$$

Therefore, putting two bounds (4.18) and (4.20) together and using $h \le 1$ gives $|(4.17)| \le b_4$.

Therefore, in view of (4.5), (4.16) and (4.17), $-b_4 \le h\tilde{W} \le b_3+b_4$.

We now apply Theorem 2 of Hoeffding (1963). Since $h \le \varepsilon$, using the second inequality of (4.12) in Lemma 4.4 and applying Theorem 2 of Hoeffding (1963) gi

(4.21) $$\underset{\sim}{E}[\tilde{\tilde{W}}(y) > G(y+\varepsilon)+\varepsilon] \le \underset{\sim}{E}[\tilde{\tilde{W}}(y) - \underset{\sim}{E}\tilde{\tilde{W}}(y) > \varepsilon - b_1 h]$$
$$\le \exp\{-\frac{2nh^2((\varepsilon-b_1h)^+)^2}{(b_3+2b_4)^2}\}.$$

Furthermore, by the first inequality of (4.12) $\{\bar{\tilde{W}}(y)<G(y-\varepsilon)-\varepsilon\}\subset\{\underset{\sim}{E}\bar{\tilde{W}}(y)-\bar{\tilde{W}}(y)>\varepsilon-b_1h\}$. Hence, by the symmetry of the tail bounds, $\underset{\sim}{E}[\bar{\tilde{W}}(y)<G(y-\varepsilon)-\varepsilon]\le$ rhs(4.21), which together with (4.21) gives us the asserted bound of Lemma 4.5.

With $\hat{G}_n$ defined by (3.8) we can use Lemma 4.6 (Nogami (1982, Lemma 2.6)) below for an upper bound of the first term of the right hand side of Theorem 1.1. By the triangular inequality applied to the first inclusion sign, $\{L(G,\hat{G}_n)>2\varepsilon\}\subset\{L(G,G_n)+L(G_n,\hat{G}_n)>2\varepsilon\}\subset\{L(G,G_n)>\varepsilon\}\cup\{L(G_n,\hat{G}_n)>\varepsilon\}$ where G_n is the empiric distribution of $\theta_1, \ldots ,\theta_n$. Hence,

$$\underset{\sim}{E}\{L(G,\hat{G}_n)>2\varepsilon\} \le \underset{\sim}{E}\{L(G,G_n)+L(G_n,\hat{G}_n)>2\varepsilon\} \le \underset{\sim}{G}\{L(G,G_n)>\varepsilon\}+\underset{\sim}{E}\{L(G_n,\hat{G}_n)>\varepsilon\}. \tag{4.22}$$

We shall use Lemma 4.6 below to get a bound for the second term of the extreme rhs(4.22). Since the proof involves only on the truncation of the raw estimate $\bar{\tilde{W}}(y)$ into $[F_n(y),F_n(y+1)]$ and the monotonicity of $\hat{G}_n$, the same proof as Lemma 2.6 of Nogami (1982) applies. Hence, we omit it.

<u>Lemma 4.6.</u> (Nogami (1982, Lemma 2.6)) For any $\varepsilon>0$, if $h\le\varepsilon$ and $\delta\le\varepsilon$, then

$$\underset{\sim}{E}[L(G_n,\hat{G}_n)>2\varepsilon] \le (\delta^{-1}+1)[\varepsilon^{-1}+1](\text{rhs}(4.15)). \tag{4.23}$$

For bounding the first term of the extreme rhs(4.22), we use Lemma 4.7 below.

<u>Lemma 4.7.</u>

$$\underset{\sim}{G}\{L(G,G_n)>\varepsilon\} \le 2\exp\{-2n\varepsilon^2\}. \tag{4.24}$$

<u>Proof.</u>) Since the lhs(4.24$=\underset{\sim}{E}(\{G(y)>G_n(y+\varepsilon)+\varepsilon\}\cup\{G(y)<G_n(y-\varepsilon)-\varepsilon\}) \le \underset{\sim}{E}\{G(y)>G_n(y+\varepsilon)+\varepsilon\}+\underset{\sim}{E}\{G(y)<G_n(y-\varepsilon)-\varepsilon\} \le \underset{\sim}{G}\{G(y)-G_n(y)>\varepsilon\}+\underset{\sim}{G}\{G_n(y)-G(y)>\varepsilon\}$. Since $\underset{\sim}{G}(G_n(y))=G(y)$, Hoeffding's inequality (1963) applied twice gives us the asserted bound of the lemma.

Lemmas 4.6 and 4.7 gives

(4.25) $$\underset{\sim}{E}[L(G,\hat{G}_n)>2\varepsilon] \le (\delta^{-1}+1)[\varepsilon^{-1}+1](\text{rhs}(4.15))+2\exp\{-2n\varepsilon^2\}.$$

To get a rate of convergence for $R(\hat{\phi}_n,G)-R$ defined in (1.2) we use the bound of Theorem 1.1. Let us assume $\Omega=[c,d]$ in the definition of $p(f)$. Since the similar proof to that for Theorem 2.1 of Nogami (1982) does work, we shall omit the proof of the following theorem:

<u>Theorem 4.1.</u> If $P_\theta \varepsilon p(f)$ with $\Omega=[c,d]$ and f satisfying the assumptions $(0<)m^{-1}\le f\le 1$ and the Lipshitz condition (4.1) and if G satisfies the assumption Ai) and Aii), then there exist positive constants a_1 and a_2 so that, for $\hat{\phi}_n$ with $a_1h=a_2\delta=(n^{-1}\log n)^{\frac{1}{4}}$,

(4.23) $$|R(\hat{\phi}_n,G) - R| = O((n^{-1}\log n)^{\frac{1}{2}}).$$

5. A Rate $O((n^{-1}\log n)^{\frac{1}{2}})$ for $R(\hat{\hat{\phi}}_n,G)-R$.

To get a convergence rate for $R(\hat{\hat{\phi}}_n,G)-R$ we shall proceed in the similar fashion to Section 4. In this section we dare not assume Lipshitz condition (4.1) or equivalently (4.2). As we can see from Section 4, Lipshitz condition (4.1) leads Lemma 4.5 to a direct generalization of Lemma 3.1 of Fox (1970). Even in this section we may be able to use this assumption (4.1) for this purpose However, we may not have much interest of it, so that we do not assume (4.1).

By the definition (3.11) of $\overline{W^*}$, $\overline{W^*} = n^{-1}\Sigma_{j=1}^n W_j^*$ where for each j

(5.1) $$hW_j^* = \Sigma_{r=0}^{\infty} \int_{-\infty}^{y}(q(t))^{-1}\, d_t\{K(h^{-1}(X_j-t+r)/f(X_j)\}$$

where the subscript t in d_t is defined in (4.3). Fix j and use the corresponding notations without subscript j until the end of the proof of Lemma 5.3. Following Lemma 5.1 corresponds to Lemma 4.1.

Lemma 5.1. For $0<h<1$,

$$h(q(\theta))^{-1}P_\theta W^* = I(y)-S(y) \tag{5.2}$$

where $I(y)$ and $S(y)$ are defined in Lemma 4.1.

Proof.) Let $g]_a^b=g(b)-g(a)$ until the end of the proof. By the similar reason to getting (4.5),

$$hW^* = \Sigma_{r=0}^{\infty} \frac{K(h^{-1}(X-y+r))}{f(X)q(y)} - \int_{-\infty}^{y}\Sigma_{r=0}^{\infty} \frac{K(h^{-1}(X-t+r))}{f(X)} f]_t^{t+1}dt. \tag{5.3}$$

Thus,

$$h(q(\theta))^{-1}P_\theta W^* = (q(y))^{-1}\Sigma_{r=0}^{\infty}\int_{y-r}^{y-r+h}K(h^{-1}(s-y+r))[\theta\le s<\theta+1]ds \tag{5.4}$$

$$-\int_{-\infty}^{y}\Sigma_{r=0}^{\infty}\int_{y-r}^{y-r+h}K(h^{-1}(s-t+r))[\theta\le s<\theta+1]ds\ f]_t^{t+1}dt.$$

Changing the variables $u=h^{-1}(s-y+r)$ and $v=h^{-1}(s-t+r)$ in the first and second integration in the rhs gives

$$\text{rhs}(5.4) = h(q(y))^{-1}\Sigma_{r=0}^{\infty}\int_0^1 K(u)[\theta-y+r\le hu<\theta+1-y+r]du \tag{5.5}$$

$$-h\int_{-\infty}^{y}\Sigma_{r=0}^{\infty}\int_0^1 K(v)[\theta-t+r\le hv<\theta+1-t+r]dv\ f]_t^{t+1}dt.$$

Two telescopic series leads to the equality in the lemma.

We use Lemmas 4.2 and 4.3 to get following Lemma 5.2 which corresponds to Lemma 4.4.

Lemma 5.2. For any y

$$G(y) - mh \le \underset{\sim}{E}\overline{W^*}(y) \le G(y+h) + 2mh.$$

Proof.) By Lemmas 5.1, 4.2 and 4.3 with $M\wedge 1$ replaced by 1,

$$[\theta\le y] - hq(\theta) \le P_\theta W^*$$

$$\le [\theta\le y+h] + q(\theta)[y<\theta\le y+h]\{(q(y)^{-1}-(q(\theta))^{-1}\} + q(\theta)h$$

Thus, applying inequalities (4.14) and $(0<)q\le m$ and taking average wrt j and then expectation wrt $\underset{\sim}{E}$ leads to the bounds of the asserted lemma.

Following Lemma 5.3 is no more a direct generalization of Lemma 3.1 of Fox (1970). However, as the author stated at the beginning of this section, we can easily check that if we have the Lipshitz assumption (4.1) on 1/f, then resulted Lemma 5.3 would be a direct generalization of Fox's lemma.

Lemma 5.3. If $0<h\le\varepsilon\le 1$, then for every y

$$1-\underset{\sim}{E}(\{G(y-\varepsilon)-\varepsilon\le G^*(y)\le G(y+\varepsilon)+\varepsilon\}) \le 2\exp\{-\frac{2nh^2((\varepsilon-2mh)^+)^2}{((1+2h)m)^2}\} \tag{5.6}$$

Proof.) The proof is similar to that for Lemma 4.5 except bounds for W_j^*. In view of (5.1), we have that $0 \le$ the first term of rhs(5.1) $\le m$ because $(0\le)K\le 1$, $f^{-1}\le m$ and $q^{-1}\le 1$. Similarly, (4.18) with $M\wedge 1$ replaced by 1, gives us |the second term of rhs(5.1)| $\le hm$. Therefore,

$$-m \le W^* \le (h^{-1}+1)m.$$

Using Lemma 5.2 and Hoeffding's bound (1963) gives the asserted bounds.

As in (4.25), Lemmas 5.3 and 4.7 gives

$$\underset{\sim}{E}[L(G,\hat{\hat{G}}_n) > 2\varepsilon] \le (\delta^{-1}+1)(\varepsilon^{-1}+1)(\text{rhs}(4.6)) + 2\exp\{-2n\varepsilon^2\}. \tag{5.7}$$

Therefore, we obtain without a proof

Theorem 5.1. If $P_\theta\in p(f)$ with $(0<)m^{-1}\le f\le 1$ and if G satisfies Ai) and Aii), then there exist positive constants k_1 and k_2 so that for $k_1h=k_2\delta=(n^{-1}\log n)^{\frac{1}{4}}$,

$$|R(\hat{\hat{\phi}}_n,G)-R| = O((n^{-1}\log n)^{\frac{1}{2}}). \tag{5.8}$$

REFERENCES

[1] Feller, W. (1971). An Introduction to Probability Theory and its Applications. Volume II (2nd ed.), Wiley, New York.

[2] Fox, R. (1970). Estimating the empiric distribution function of certain parameter sequences, Ann. Math. Statist., Vol. 41, 1845-1852.

[3] ——— (1978). Solutions to empirical Bayes squared error loss estimation problems, Ann. Statist., 6, 846-853.

[4] Halmos, P. R. (1950). Measure Theory, Litton Educational Publishingg.

[5] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, J. Amer. Statist. Ass., 58, 13-30.

[6] Nogami, Y. (1975). A non-regular squared-error loss set-compound estimation problem, RM-345, Department of Statistics and Probability, Michigan State University.

[7] ——— (1981). The set-compound one-stage estimation in the nonregular family of distributions over the interval $[\theta,\theta+1)$, Ann. Inst. Statist. Math., 33, A, 67-80.

[8] ——— (1982). A rate of convergence for the set compound estimation in a family of certain retracted distributions, Ann. Inst. Statist. Math., 34, A, 241-257.

[9] ——— (1983). Exact rate $n^{-2/3}$ in the empirical Bayes estimation: case of retracted distributions, (accepted from the Proceedings in Pacific Area Statistical Conference, North Holland).

[10] Oaten, A. (1969). Approximation to Bayes risk in compound decision problem, RM-233, Department of Statistics and Probability, Michigan State University.

[11] Royden, H. L. (1968). Real Analysis, The Macmillan Company, New York.

[12] Saks, S. (1937). Theory of the Integral (2nd ed.), Monografie Mathematyczne.

[13] Singh, Radhey S. (1974). Estimation of derivatives of average of μ-densities and sequence compound estimation in exponential families, RM-318, Department of Statistics and Probability, Michigan State University.