

No. 54 (79-18)

A Non-regular Squared-error Loss  
Set-compound Estimation Problem

by

Yoshiko Nogami

September, 1979

# A NON-REGULAR<sup>\*)</sup> SQUARED-ERROR LOSS

## SET-COMPOUND ESTIMATION PROBLEM<sup>1</sup>

By Yoshiko Nogami

University of Tsukuba

### Summary.

This paper is concerned with the set-compound decision problem when the component problem is the squared-error loss estimation for the family of distributions  $P_\theta$  specified by a density proportional to the restriction of an integrable function  $f$  to the interval  $[\theta, \theta+1)$  for  $\theta$  in a real interval  $\Omega$ . The work in this paper is a generalization and continuation of R. Fox's work (1968, 1970) where he constructed a Lévy consistent estimate  $\hat{G}_n$  of the empiric distribution  $G_n$  of the  $n$ -parameter sequence  $\theta_i$  under  $P_\theta$  being the uniform distribution over the interval  $[\theta, \theta+1)$  for  $\theta \in \Omega = (-\infty, \infty)$ . All the orders being uniform in  $\theta \in [c, d]^n$  with  $c, d$ , finite, Sections 1 and 2 show that there exists a procedure  $\hat{\theta}$  whose modified regret  $D(\theta, \hat{\theta})$  is  $O((n^{-1} \log n)^{1/4})$ . Section 3 gives a counter example to the convergence of the modified regret for  $\theta \in (-\infty, \infty)^n$ .

---

\*) The word "non-regular" was quoted from Ferguson (1968, p.130) in which he refers the exponential families of distributions to regular families.

1 This paper is a part of the author's Ph. D. Thesis at Michigan State University.

## 0. Introduction.

The set compound problem simultaneously considers  $n$  statistical decision problems each of which is structurally identical to the component problem. The loss is taken to be the average of  $n$  component losses.

Let  $\xi$  be Lebesgue measure and  $f$  an integrable function with  $0 \leq f \leq 1$ . Let  $\doteq$  denote the defining property. Define  $q(\theta) \doteq (\int_{\theta}^{\theta+1} f d\xi)^{-1}$  and assume that  $q$  is uniformly bounded by a finite constant, say  $m$ . Letting  $p_{\theta} = dP_{\theta}/d\xi$  we denote by  $p(f)$  the family of probability measures given by

$$(0.1) \quad \bar{p}(f) = \{P_{\theta} \text{ with } p_{\theta} = q(\theta)[\theta, \theta+1)f, \forall \theta \in \Omega\}$$

where  $\Omega$  is a real interval and we denote the indicator function of a set  $A$  by  $A$  itself. In this paper, the component problem considered is the squared-error loss estimation of  $\theta$  based on  $X$  with distribution  $P_{\theta} \in p(f)$ . For any prior distribution  $G$  on  $\Omega$ , let  $R(G)$  be the Bayes risk versus  $G$  in this component problem.

Let  $X_1, \dots, X_n$  be  $n$  independent random variables with  $X_j$  distributed according to  $P_{\theta_j} \in p(f)$ . Let  $\underline{t} = (t_1, t_2, \dots, t_n)$  be a set compound procedure: for each  $j = 1, 2, \dots, n$ ,  $t_j$  is an estimator of  $\theta_j$  based on  $\underline{X} \doteq (X_1, \dots, X_n)$ . Let  $G_n$  denote the empiric distribution of  $\theta_1, \dots, \theta_n$  and let

$$(0.2) \quad D(\underline{\theta}, \underline{t}) = \int n^{-1} \sum_{j=1}^n (t_j(\underline{x}) - \theta_j)^2 dP(\underline{x}) - R(G)$$

where  $\underline{P} = P_{\theta_1} \times \dots \times P_{\theta_n}$ .

With squared-error loss, let  $\theta_{G_n}$  be the procedure whose component procedures are Bayes against  $G_n$ :  $\theta_{G_n}(\underline{X}) \doteq (\theta_{1n}, \theta_{2n}, \dots, \theta_{nn})$  with,

for each  $j$ ,

$$(0.3) \quad \theta_{jn} = \int \theta p_{\theta}(X_j) dG_n(\theta) / \int p_{\theta}(X_j) dG_n(\theta) \\ = \int_{X'_j+}^{X_j} \theta q(\theta) dG_n(\theta) / \int_{X'_j+}^{X_j} q dG_n$$

where  $y'$  is an abbreviation of  $y-1$  and the affix  $+$  is intended to describe the integration as over  $(X'_j, X_j]$ . Henceforth we delete  $+$  in lower limits of  $\int$ s.

For the case where  $f \equiv 1$  and  $\Omega = (-\infty, \infty)$ , Fox(1970) exhibited a distribution-valued Lévy consistent estimate  $\hat{G}_n$  of  $G_n$ . In empirical Bayes problem where the  $\theta_i$  are iid with common distribution  $G$ , Fox(1968, §4.3) obtained a convergence rate  $o(1)$  of the expected risks to  $R(G)$  for a (bootstrap) decision procedure  $\hat{\theta}$  based on component procedures Bayes versus an estimate  $\hat{G}_n$ .

The behavior in the compound problem of the generalization of the procedure  $\hat{\theta}$  to  $p(f)$  is the subject of this paper.

If  $\sup\{|D(\theta, \hat{\theta})| : \theta \in \Omega^n\} = O(n^{-\alpha})$  then we will say that  $\hat{\theta}$  has a rate  $O(n^{-\alpha})$ .

In Section 1 (Theorem 1.1) we exhibit an upper bound of the modified regret  $D(\theta, \hat{\theta})$  (uniform wrt  $\theta \in \Omega^n$ ) in terms of Lévy metric  $L(G_n, \hat{G}_n)$  of  $G_n$  and any distribution-valued estimate  $\hat{G}_n$ , when  $\Omega$  is bounded. In Section 2 we construct a particular distribution-valued Lévy consistent estimate  $\hat{G}_n$  of  $G_n$  for  $\Omega = (-\infty, \infty)$ . Under an additional assumption that  $1/f$  satisfies a Lipschitz condition, we show in Theorem 2.1, by making use of the bound in Theorem 1.1, that the set compound decision procedure  $\hat{\theta}$  based on  $\hat{G}_n$  has a rate  $O((n^{-1} \log n)^{1/4})$ . Section 3 shows

in Theorem 3.1 that when  $\Omega = (-\infty, \infty)$ , there is no sequence of estimate of  $\theta$  for which  $D(\theta, \underline{t})$  converges to zero.

### Notational Conventions.

$P_j$  and  $P$  abbreviate  $P_{\theta_j}$  and  $X_{j=1}^n P_{\theta_j}$ , respectively. A distribution function also represents the corresponding measure. We often let  $Ph$ ,  $P(h)$  or  $P(h(\omega))$  denote  $\int h(\omega) dP(\omega)$ .  $G$  abbreviates the empiric distribution  $G_n$  of  $\theta_1, \dots, \theta_n$ .  $R$  denotes the real line. We abbreviate  $y-1$  to  $y'$ . We denote the indicator function of a set  $A$  by  $[A]$  or simply  $A$  itself. For any function  $h$ ,  $h]_a^b$  means  $h(b) - h(a)$ .  $\vee$  and  $\wedge$  denote the supremum and the infimum, respectively.  $\doteq$  denotes the defining property. When we refer to  $(a, b)$  in Section  $a$  that we are dealing with, we simply write  $(b)$ . For example, see the line just below (1.3). There we write  $(2)$  as we refer to (1.2). The symbol  $\square$  is used throughout to signal the end of a proof.

# 1. An Upper Bound of the Modified Regret.

Let  $\Omega = [c, d]$ , where  $-\infty < c \leq d < +\infty$ , throughout this section. Let  $\hat{G}$  be a distribution-valued random variable which is an estimate of the empiric distribution  $G$ , obtained from  $X_1, \dots, X_n$ . Define  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_n)$  to be the procedure such that, for each  $j$ ,  $\hat{\theta}_j(X) \doteq \hat{\theta}_{jn}$  is of form (0.3) with  $G$  replaced by  $\hat{G}$  (0/0 is understood to be  $X_j$ ).

The modified regret for a procedure  $t$  is of form

$$(1.1) \quad D(\theta, t) = n^{-1} \sum_{j=1}^n \{P(t_j(X) - \theta_j)^2 - P(\theta_{jn} - \theta_j)^2\}.$$

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

$$(1.2) \quad L(F, H) = \bigwedge \{ \varepsilon \geq 0 : -\varepsilon(F') \leq H' \leq \varepsilon(F') \}$$

where

$$(1.3) \quad F_\varepsilon(y) \doteq F(\varepsilon + y) \quad \text{and} \quad F'(y) \doteq y + F(y).$$

Remark that the infimum in the definition (2) attains (See Appendix of Nogami(1975)). Hereafter, we let

$$(1.4) \quad \hat{L} \doteq L(G, \hat{G}).$$

In this section we shall exhibit an upper bound of the modified regret  $D(\theta, \hat{\theta})$ . To do so, the main development is Lemma 1.3 in which we show that the average expectation of  $|\hat{\theta}_{jn} - \theta_{jn}|$  over the set where  $\hat{L} < \varepsilon$  is bounded by at most a constant times  $\varepsilon$  with  $0 < \varepsilon < 1$ . For the proof of Lemma 1.3 we use Lemma A.2 of R. S. Singh (1974).

Since  $X'_j < \theta_{jn} \leq X_j$  by (0.3) whatever be the distribution  $G$ ,

$|(\hat{\theta}_{jn} - \theta_j)^2 - (\theta_{jn} - \theta_j)^2| \leq 2 |\hat{\theta}_{jn} - \theta_{jn}|$ . Hence, it follows from (1) that

$$(1.5) \quad 2^{-1} |D(\theta, \hat{\theta})| \leq n^{-1} \sum_{j=1}^n P |\hat{\theta}_{jn} - \theta_{jn}|.$$

For fixed  $j$ , since  $|\hat{\theta}_{jn} - \theta_{jn}| \leq 1$ , for any  $0 < \epsilon < 1$ ,

$$(1.6) \quad P |\hat{\theta}_{jn} - \theta_{jn}| \leq P[\hat{L} > \epsilon] + P(|\hat{\theta}_{jn} - \theta_{jn}| | [\hat{L} \leq \epsilon]).$$

Before dealing with the second term of rhs(6), we introduce two lemmas.

Lemma 1.1. For any  $s, t \in \mathbb{R}$  with  $s \leq t$  and for any  $\delta \geq 0$  and  $\eta \geq 0$  with  $\delta + \eta < 1$ ,

$$(1.7) \quad n^{-1} \sum_{j=1}^n A_j \leq t - s$$

where  $\forall j$ ,

$$A_j = P_j \{ G \}_{X_j^{-t}}^{X_j^{-s}} [\theta_j + \delta \leq X_j < \theta_j + 1 - \eta] / \int_{X_j^{-t} + \eta}^{X_j^{-s} - \delta} q \, dG.$$

Proof. Since  $\forall j$ ,

$$(1.8) \quad A_j = \int (q(\theta_j) [\theta_j + \delta \leq y < \theta_j + 1 - \eta] / \int_{y' + \eta}^{y - \delta} q \, dG) f(y) G]_{y-t}^{y-s} dy,$$

and since  $[\theta_j + \delta \leq y < \theta_j + 1 - \eta] = [y' + \eta < \theta_j \leq y - \delta]$ , the average wrt  $j = 1, \dots, n$  of the numerator in the quotient equals to the denominator. Also, since  $f \leq 1$ , taking the average wrt  $j$  over (8) and interchanging the integral and average operation leads to lhs(7)  $\leq \int G]_{y-t}^{y-s} dy$ .

But, the Fubini Theorem leads to

$$\int F]_{y-t}^{y-s} dy = \int \int_{u+s}^{u+t} dy \, dF(u) = t - s$$

for an arbitrary distribution function  $F$  of a random variable and any  $s, t \in \mathbb{R}$  with  $s < t$ . This gives us the resulted bound.  $\square$

Lemma 1.2. For all  $s \in \mathbb{R}$ ,

$$(1.9) \quad n^{-1} \sum_{j=1}^n P\{|(G-\hat{G})(X_j-s)|[\hat{L} \leq \epsilon] / \int_{X_j}^{X_j'} q \, dG\} \leq (d-c+3)\epsilon.$$

Proof. For  $j$  fixed we let  $z = X_j - s$ . By the definition of  $\hat{L}$  and the fact that the infimum in the definition of Lévy distance is attained, if  $\hat{L} \leq \epsilon$ , then  $_{-\epsilon}(G^*) \leq \hat{G}^* \leq _{\epsilon}(G^*)$  where  $F^*$  and  $_{\epsilon}(F^*)$  are as defined in (3). Hence,

$$|(G-\hat{G})(z)|[\hat{L} \leq \epsilon] \leq G^*_{z-\epsilon} \vee G^*_{z+\epsilon} \leq \epsilon + G^*_{z-\epsilon}.$$

Thus,

$$\text{lhs}(9) \leq \epsilon \{n^{-1} \sum_{j=1}^n P_j \left( \int_{X_j}^{X_j'} q \, dG \right)^{-1}\} + n^{-1} \sum_{j=1}^n P_j (G)_{X_j-s-\epsilon}^{X_j-s+\epsilon} / \int_{X_j}^{X_j'} q \, dG.$$

From the proof of Lemma 1.1 we can see that the curly bracket of the rhs is no more than  $d-c+1$ . Hence, an application of Lemma 1.1 with  $(s, t, \eta, \delta) = (s-\epsilon, s+\epsilon, 0, 0)$  to the second term of the rhs completes the proof.  $\square$

We will invoke Lemma A.2 of R. S. Singh (1974) in the proof of Lemma 1.3 below which will give us an upper bound of the average wrt  $j$  of the second term of rhs(6).

Lemma 1.3. For  $\epsilon > 0$ ,

$$n^{-1} \sum_{j=1}^n P(|\hat{\theta}_{jn} - \theta_{jn}|[\hat{L} \leq \epsilon]) \leq a_0 \epsilon$$

where  $a_0 = 4m(17 + 24m + (7+12m)(d-c))$ .

Proof. Fix  $n$  and  $\theta \in [c, d]^n$ . We also fix  $j$  until (19).  $X$  abbreviates  $X_j$ . Since  $(0.3)-X' = \int_{X'}^X (\theta - X') q(\theta) dG / \int_{X'}^X q(\theta) dG$ , we abbreviate the quotient of the rhs to  $y/z$  and that with  $G$  replaced by  $\hat{G}$  to  $Y/Z$ . Then,

$$(1.10) \quad \hat{\theta}_{jn} - \theta_{jn} = Y/Z - y/z.$$

Let  $*$  denote conditioning on  $X$  and  $\{\hat{L} \leq \epsilon\}$ . Then, by Lemma A.2 of R. S. Singh (1974) with  $\gamma=1$  and  $L=1$  and by the fact that  $0 \leq Y/Z$ ,  $y/z \leq 1$  we have

$$(1.11) \quad P_* \left| \frac{Y}{Z} - \frac{y}{z} \right| \leq \frac{2}{z} P_*(|Y-y| + 2|Z-z|).$$

By letting  $I = (X', X]$ , define by  $G_I$  the retraction of  $G$  into the closed interval  $[G(X'), G(X)]$ . Then, by Proposition A of Nogami (1975),

$$L_I \doteq L(G_I, \hat{G}_I) \leq \hat{L} V S V T$$

where  $S = |(G-\hat{G})(X')|$  and  $T = |(G-\hat{G})(X)|$ . Thus,

$$(1.12) \quad \text{when } \hat{L} \leq \epsilon, L_I \leq \epsilon V S V T \doteq \lambda.$$

By applying Lemma A.2 of Nogami (1975) with  $h(\theta)$ , the retraction of  $(\theta - X')q(\theta)$  to  $I$ , and weakening the resulted bound, when  $L_I \leq \lambda$ ,

$$(1.13) \quad |Y-y| \leq 2\alpha(\lambda+) + m(S+T)$$

where we use  $+$  on the line to denote the right limit and  $\alpha$  is the

modulus of continuity of  $h$  defined on  $I$  such that

$$\alpha(\epsilon) = \sqrt{\{h\}_{w_1}^{w_2} : w_1, w_2 \in I, |w_1 - w_2| < \epsilon\}}.$$

To bound  $\alpha(\lambda+)$ , pick  $w_1, w_2 \in I$  such that  $0 < w_2 - w_1 < \lambda$ . Now, by the definition of  $h$ ,

$$(1.14) \quad h\big|_{w_1}^{w_2} = (w_2 - w_1)(q(w_2)) + \frac{w_1 - X'}{w_2 - w_1} q\big|_{w_1}^{w_2}.$$

But, since by the definition of  $q$ ,  $q\big|_{w_1}^{w_2} = q(w_2)q(w_1) \left( \int_{w_1}^{w_2} f(s) ds - \int_{w_1+1}^{w_2+1} f(s) ds \right)$  and since  $q \leq m$  and  $0 \leq f \leq 1$ ,

$$(1.15) \quad |q\big|_{w_1}^{w_2}| \leq m^2(w_2 - w_1).$$

Thus, from (14),  $|h\big|_{w_1}^{w_2}| \leq (w_2 - w_1)\{q(w_2) + (w_1 - X')m^2\}$ . Using  $q \leq m$ ,

$w_1 - X' \leq 1$  and  $w_2 - w_1 \leq \lambda$ , and applying the definition of  $\alpha(\lambda)$  gives us that

$$(1.16) \quad \alpha(\lambda) \leq \sqrt{\{h\}_{w_1}^{w_2} : \text{for } w_1, w_2 \in I \text{ such that } 0 < w_2 - w_1 < \lambda\}} \leq \lambda(m+m^2)$$

and thus the same bound applies for  $\alpha(\lambda+)$ .

Therefore, applying the bound of (16) to the first term of rhs(13) shows that when  $L_I \leq \lambda$ ,

$$(1.17) \quad |Y - y| \leq 2\lambda(m+m^2) + m(S+T).$$

Similarly, by Lemma A.2 of Nogami(1975) with  $1 \leq h \leq q \leq m$ , when  $L_I \leq \lambda$ ,  $|Z - z| \leq 2\alpha(\lambda+) + m(S+T)$ . Since by the definitions of  $\alpha(\lambda)$  and  $q$  and by (15)  $\alpha(\lambda) \leq \sqrt{\{|q\}_{w_1}^{w_2}\} : \text{for } w_1, w_2 \in I \text{ such that } 0 < w_2 - w_1 < \lambda\}}$

$\leq \lambda m^2$ ,  $\alpha(\lambda+)$  is also bounded by  $m^2 \lambda$ . Hence, as in (17), when  $L_I \leq \lambda$ ,

$$(1.18) \quad |Z-z| \leq 2\lambda m^2 + m(S+T).$$

Therefore, by (17), (18) and (12) and weakening the bound by replacing  $\lambda$  there by  $\varepsilon+S+T$ , when  $\hat{L} \leq \varepsilon$ ,

$$(1.19) \quad |Y-y| + 2|Z-z| \leq 2(m+3m^2)\varepsilon + (5m+6m^2)(S+T).$$

By this and in view of (11) and (10),

$$\begin{aligned} n^{-1} \sum_{j=1}^n P(|\hat{\theta}_{jn} - \theta_{jn}| [\hat{L} \leq \varepsilon]) &\leq 4(m+3m^2)\varepsilon (n^{-1} \sum_{j=1}^n P_j z^{-1}) \\ &+ 2(5m+6m^2)n^{-1} \sum_{j=1}^n P\{(S+T) [\hat{L} \leq \varepsilon]/z\}. \end{aligned}$$

Applying Lemma 1.1 with  $s=c-d-1$ ,  $t=d+1-c$  and  $\eta=\delta=0$  to the first term and using Lemma 1.2 twice to the second term results in the bound of the asserted lemma.  $\square$

The following theorem is an immediate consequence of (5), (6) and Lemma 1.3.

Theorem 1.1. If  $P_j \in p(f)$  with  $\Omega = [c, d]$ , for  $j=1, 2, \dots, n$ , then  $\varepsilon > 0$ ,

$$2^{-1} |D(\theta, \hat{\theta})| \leq P[\hat{L} > \varepsilon] + a_0 \varepsilon \quad \text{uniformly in } \theta,$$

where  $a_0$  is as defined in Lemma 1.3.

## 2. A Particular Procedure $\hat{G}$ with a Rate $O((n^{-1} \log n)^{1/4})$ .

We first construct a normalized (but not monotonized) estimate  $G^*$  of the empiric distribution function  $G$ . Main work in this section is, under the extra assumption on  $f$  (Lipshitz condition for  $1/f$ ), to obtain the generalization (Lemma 2.2) of Lemma 3.1 of Fox(1970). Then, we exhibit a distribution-valued estimate  $\hat{G}$  of  $G$ . Lemma 2.3 showing Lévy consistency of  $\hat{G}$  to  $G$ , will be proved as in the proof of Theorem 3.1 of Fox(1970) by using Lemma 2.2. Lemma 2.1 will be furnished to apply Hoeffding's bound (1963, Theorem 2) in the proof of Lemma 2.2. Finally, Theorem 2.1 shows that there exists a procedure  $\hat{G}$  with a rate  $O((n^{-1} \log n)^{1/4})$ .

In addition to the assumption on  $f$  in the introduction we now assume that  $1/f$  satisfies the Lipshitz condition:

$$(2.1) \quad \bigvee \{ (v-u)^{-1} |(f(v))^{-1} - (f(u))^{-1}| : u < v \} \leq M$$

for a finite constant  $M$ . By this assumption,

$$(2.2) \quad |f(s)/f(t) - 1| \leq M|s-t|.$$

Let  $\Omega = \mathbb{R}$  until the proof of Lemma 2.3 is ended. Let  $Q$  be the distribution function defined by

$$(2.3) \quad Q(y) = \int_{-\infty}^y q \, dG, \quad \forall y.$$

Then, letting  $\bar{p} \doteq \int p_{\theta} \, dG(\theta)$ , we have by the definition of  $p_{\theta}$  that  $\bar{p}(y) = f(y)(Q(y) - Q(y'))$  and thus

$$(2.4) \quad Q(y) = \sum \frac{\bar{p}(y-r)}{f(y-r)}$$

where  $\sum$  abbreviates  $\sum_{r=0}^{\infty}$  throughout this section. Letting  $[z]$  denote

the greatest integer  $\leq z$  if  $z > 0$  and  $-1$  if  $z < 0$ , we remark that if the  $r$ -th term of rhs(4) is nonzero then

$$(2.5) \quad r \leq [\theta_{(n)} - \theta_{(1)} + 1]$$

where  $\theta_{(1)} = \min_{1 \leq i \leq n} \theta_i$  and  $\theta_{(n)} = \max_{1 \leq i \leq n} \theta_i$ . Since  $q \geq 1$  and  $q$  is the density of  $Q$  wrt  $G$ , it follows by Theorem 32.B of Halmos(1950) that

$$(2.6) \quad G(y) = \int_{-\infty}^y (q(\theta))^{-1} dQ(\theta).$$

For each  $y$ , we let  $F^*(y) = n^{-1} \sum_{j=1}^n [X_j \leq y]$  and for any  $h > 0$   $\Delta F^*(y) = h^{-1} F^*]_y^{y+h}$ . We allow  $h$  to depend on  $n$  and assume  $h < 1$  for convenience. Let  $\bar{P} \doteq \int P_\theta dG$ . Then,  $\bar{p} = d\bar{P}/d\xi$  where  $\xi$  is Lebesgue measure. We estimate  $\bar{p}(y)$  by  $\Delta F^*(y)$  and  $Q(y)$  by

$$(2.7) \quad Q^*(y) = \sum (\Delta F^*(y-r)/f(y-r)).$$

As in (5), we note that if the  $r$ -th term of rhs(7) is nonzero, then

$$(2.8) \quad r \leq d-c+2 \doteq b_0-1.$$

Note that  $Q^*$  has bounded variation because of (1). From the relation (6), we obtain a raw estimate  $\bar{W}$  of  $G$  from

$$(2.9) \quad \bar{W}(y) = \int_{-\infty}^y (q(t))^{-1} dQ^*(t).$$

Since  $F^*(y) \leq G(y) \leq F^*(y+1)$  for all  $y \in R$ , we furthermore estimate  $G$  at a point  $y$  by

$$G^*(y) = (F^*(y) \vee \bar{W}(y)) \wedge F^*(y+1).$$

Following Lemma 2.1 will be used to prove forthcoming Lemma 2.2.

Lemma 2.1 For every  $y \in [\theta_{(1)}^{-1}, \theta_{(n)}^{+1}]$ ,

$$(2.10) \quad G(y) - b_1 h \leq \mathbb{P} \bar{W}(y) \leq G(y+h) + b_1 h$$

where  $b_1 = 2^{-1} m(2M + 3(1/M))$ .

Proof. Since the summation on  $r$  in (7) involves at most a finite number of non-zero terms, we shall freely interchange integral and summation on  $r$  without further comment.

For each  $j$ , let

$$(2.11) \quad w_j \doteq \int_{-\infty}^y (q(t))^{-1} d_t \{ [t-r < X_j \leq t-r+h] (h f(t-r))^{-1} \},$$

where the subscript  $t$  in  $d_t$  denotes the variable of integration.

By the definition (9) of  $\bar{W}$ ,

$$\bar{W}(y) = n^{-1} \sum_{j=1}^n w_j.$$

To find bounds of  $\mathbb{P} \bar{W}(y)$  we shall find an upper and a lower bound of  $\mathbb{P} w_j$ ,  $\forall j$ . Fix  $j$  and use the corresponding notations without subscript  $j$  until (27).

We shall start with getting an alternative form of  $\mathbb{P} W$ . Because a function satisfying Lipschitz condition is absolutely continuous (cf. Royden (1968), p.108) and  $1/q$  is clearly absolutely continuous,  $1/f(\cdot-r)$  and  $1/q$  are both of bounded variation. Applying integration by parts (Saks(1937), Theorem III.14.1) and using  $d(q(t))^{-1} = (f(t+1) - f(t))dt$  gives us that

$$\begin{aligned}
 (2.12) \quad & \int_{-\infty}^y (q(t))^{-1} d_t([t-r < X \leq t-r+h]/f(t-r)) \\
 &= \frac{[y-r < X \leq y-r+h]}{f(y-r) q(y)} - \int_{-\infty}^y \frac{[t-r < X \leq t-r+h]}{f(t-r)} f]_t^{t+1} dt.
 \end{aligned}$$

Now, with  $EX$  denoting the expectation of a random variable  $X$ , Proposition III. 2.1 of Neveu(1965) gives us a version of the relation  $E\{E(h(t)|X)\} = Eh(t)$  for an integrable function  $h$  and probability measures. But, because of its proof it holds for finite measures. Hence,

$$(2.13) \quad P_{\theta} \left\{ \int_{-\infty}^y \frac{[t-r < X \leq t-r+h]}{f(t-r)} f]_t^{t+1} dt \right\} = hq(\theta) \int_{-\infty}^y S(t-r) f]_t^{t+1} dt$$

where

$$(2.14) \quad S(t) \doteq h^{-1} \int_t^{t+h} [\theta \leq s < \theta+1] (f(s)/f(t)) ds.$$

Thus, taking expectation wrt  $X$  and then summation on  $r$  over (12) and multiplying  $(hq(\theta))^{-1}$  on both sides shows us that

$$(2.15) \quad (PW)/q(\theta) = (q(y))^{-1} \sum S(y-r) - I(S)$$

where

$$(2.16) \quad I(S) = \int_{-\infty}^y \sum S(t-r) f]_t^{t+1} dt.$$

To get bounds for  $PW$  we shall first find bounds for the first term of rhs(15) and then bounds for the second term  $I(S)$  of rhs(15). Until (25) we use the notation

$$(2.17) \quad \Delta(t) = h^{-1} \int_t^{t+1} [\theta \leq s < \theta+1] ds.$$

Applying (2) to the definition (14) of  $S(t)$  and changing a variable

leads to the inequality

$$(2.18) \quad |S(y-r) - \Delta(y-r)| \leq Mh^{-1} \int_0^h [\theta - (y-r) \leq u < \theta + 1 - (y-r)] u \, du.$$

Moreover, because

$$(2.19) \quad \begin{aligned} \Sigma \Delta(y-r) &= h^{-1} \int_y^{y+h} [\theta \leq t] \, dt \\ &= h^{-1} [\theta - h \leq y < \theta] (y+h-\theta) + [\theta \leq y], \end{aligned}$$

rhs(18)  $\leq 2^{-1} Mh$  and  $(q(y))^{-1} \leq 1$ , we obtain

$$(2.20) \quad \begin{aligned} (q(y))^{-1} [\theta \leq y] - 2^{-1} Mh &\leq \text{first term of rhs(15)} \\ &\leq (q(y))^{-1} [\theta \leq y+h] + 2^{-1} Mh. \end{aligned}$$

Since  $\int_{-\infty}^y S(t-r) f]_t^{t+1} \, dt = \int S(t) [t \leq y-r] f]_{t+r}^{t+r+1} \, dt$  and

$$(2.21) \quad \begin{aligned} \Sigma [t \leq y-r] (f(t+r+1) - f(t+r)) &= [t \leq y] (f(t+[y-t]+1) - f(t)) \\ &= f(t + [y-t] + 1) - f(t) \end{aligned}$$

(the latter because  $[y-t] = -1$  if  $t > y$ ), it follows that

$$(2.22) \quad I(S) = \int S(t) f]_t^{t+[y-t]+1} \, dt.$$

From the derivation, (22) holds for  $\Delta$  in place of  $S$ . Thus, from (18) with any  $y-r$  and then by  $0 \leq f \leq 1$ ,

$$(2.23) \quad |I(S) - I(\Delta)| \leq 2^{-1} Mh.$$

But by (2.19),  $I(\Delta)$  equals  $\int_{-\infty}^y (\text{lhs(19)} \text{ with } y=t)(f(t+1) - f(t)) \, dt$

which becomes

$$(2.24) \quad \{ [\theta-h \leq y < \theta] \int_{\theta-h}^y + [\theta \leq y] \int_{\theta-h}^{\theta} \} h^{-1} (t+h-\theta) f]_t^{t+1} dt \\ + [\theta \leq y] \left( \int_{\theta}^y f \right)_t^{t+1} dt.$$

Since  $|f(t+1) - f(t)| \leq |(f(t))^{-1} - (f(t+1))^{-1}| \wedge 1 \leq M \wedge 1$  and  $\int_{\theta-h}^{y \wedge \theta} (t+h-\theta) dt \leq 2^{-1} h^2$ , |first term of (24)|  $\leq (M \wedge 1) 2^{-1} h$ . Thus,

$$(2.25) \quad |I(\Delta) - [\theta \leq y] \left( \int_{\theta}^y f \right)_t^{t+1} dt| \leq (M \wedge 1) 2^{-1} h.$$

Therefore, by this and (23),

$$(2.26) \quad |I(S) - [\theta \leq y] \left( \int_{\theta}^y f \right)_t^{t+1} dt| \leq (M + M \wedge 1) 2^{-1} h.$$

Therefore, from this and (20) and in view of (15) we can see that

$(P \ W)/q(\theta)$  is bounded above and below by rhs(20) - (lower bound of  $I(S)$  in (26)) and lhs(20) - (upper bound of  $I(S)$  in (26)), respectively.

Since,

$$[\theta \leq y+h]/q(y) - [\theta \leq y+h]/q(\theta) = [\theta \leq y] \int_{\theta}^y f]_t^{t+1} dt \\ + [y < \theta \leq y+h] \int_y^{\theta} f]_{t+1}^t dt$$

for  $h > 0$  and  $|[y < \theta \leq y+h] \int_y^{\theta} f]_{t+1}^t dt| \leq (M \wedge 1) h$ , weakening the above

bounds by using  $q \leq m$  results in

$$(2.27) \quad [\theta \leq y] - b_1 h \leq P \ W \leq [\theta \leq y+h] + b_1 h$$

where  $b_1$  is as defined in the statement of this lemma.

Averaging (27) wrt  $j$  gives the bound of the asserted lemma.  $\square$

Following Lemma 2.2 is a direct generalization of Lemma 3.1 of Fox (1970) in the sense that if  $f \equiv 1$ , then  $m=1$  and  $M=0$ , and hence we get his bound  $2\exp(-2nh^2\epsilon^2)$ .

Lemma 2.2 If  $0 < h \leq \epsilon \leq 1$ , then for each  $y$

$$(2.28) \quad P(\{G(y-\epsilon) - \epsilon \leq G^*(y) \leq G(y+\epsilon) + \epsilon\}^c)$$

$$\leq 2\exp \left\{ - \frac{2nh^2((\epsilon - b_1 h)_+)^2}{(1+3b_0 M)^2} \right\}$$

where  $A^c$  is the complement of a set  $A$ ,  $b_0$  and  $b_1$  are as defined in (8) and Lemma 2.1, respectively.

Proof. For  $y > \theta_{(n)} + 1$ ,  $F^*(y) = G^*(y) = G(y+\epsilon) = 1$  and

for  $y < \theta_{(1)} - 1$ ,  $F^*(y+1) = G^*(y) = G(y-\epsilon) = 0$ ; in both case lhs(28)=0 and (28) holds trivially.

For  $y \in [\theta_{(1)} - 1, \theta_{(n)} + 1]$  it is sufficient to prove the lemma for the raw estimate  $\bar{W}$ . For if  $G(y-\epsilon) - \epsilon \leq \bar{W}(y) \leq G(y+\epsilon) + \epsilon$ , it follows that  $G(y-\epsilon) - \epsilon \leq \bar{W}(y) \wedge F^*(y+1) \leq G^*(y) \leq \bar{W}(y) \vee F^*(y) \leq G(y+\epsilon) + \epsilon$ .

Pick  $y \in [\theta_{(1)} - 1, \theta_{(n)} + 1]$ . 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 of  $P(\bar{W}(y))$  in Lemma 2.1 and furthermore need to get an upper and a lower bound of  $W_j, \forall j$ . By (12) and (21) applied to the definition (11) of  $W_j$

$$(2.29) \quad hW_j = (q(y))^{-1} E [y-r < X_j \leq y-r+h] / f(y-r) \\ - \int [t < X_j \leq t+h] \{ (f(t + [y-t] + 1) / f(t)) - 1 \} dt.$$

In the summation of the first term of rhs(29), there are at most two positive terms and both terms cannot be positive at the same time. Applying (2) and then (8) gives that with  $b_0$  as defined in (8)

$$0 \leq (\text{first term of rhs(29)}) \leq 1 + b_0 M.$$

In addition, by a use of (2) and the fact that  $[y - X_j + h] \leq b_0 - 1$  (because  $y \leq \theta_{(n)} + 1$ ,  $\theta_{(1)} \leq X_j$  and  $h < 1$ ),

$$|\text{second term of rhs(29)}| \leq b_0 M h \quad (< b_0 M).$$

Therefore,

$$-b_0 M \leq h W_j \leq 1 + 2b_0 M, \quad \forall j.$$

We now apply Theorem 2 of Hoeffding (1963). Since  $h \leq \varepsilon$ , using the second inequality of (10) in Lemma 2.1 and applying Theorem 2 of Hoeffding (1963) gives

$$(2.30) \quad \begin{aligned} P[\bar{W}(y) > G(y+\varepsilon) + \varepsilon] &\leq P[\bar{W}(y) - P\bar{W}(y) > \varepsilon - b_1 h] \\ &\leq \exp \left\{ - \frac{2nh^2((\varepsilon - b_1 h)_+)^2}{(1+3b_0 M)^2} \right\}. \end{aligned}$$

Furthermore, by the first inequality of (10),  $\{\bar{W}(y) < G(y-\varepsilon) - \varepsilon\} \subset \{P\bar{W}(y) - \bar{W}(y) > \varepsilon - bh\}$ . Hence by the symmetry of the tail bounds,  $P[\bar{W}(y) < G(y-\varepsilon) - \varepsilon]$  has the same upper bound, rhs(30), which together with (30) gives us the asserted bound of Lemma 2.2. ■

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

$$(2.31) \quad \hat{G}(y) = \sup\{G^*(j\delta) : j\delta \leq y, j=0, \pm 1, \dots\}.$$

Lemma 2.3. ((Fox(1970))). For any  $\epsilon > 0$ , if  $h \leq \epsilon$  and  $\delta \leq \epsilon$ , then

$$(2.32) \quad P[\hat{L} > 2\epsilon] \leq (\delta^{-1}+1)[\epsilon^{-1}+1](\text{rhs}(28))$$

where  $\hat{L}$  is as defined in (1.4).

Proof. We rely on the proof of Theorem 3.1 of Fox(1970).

For  $0 < \epsilon \leq 1$ , let  $n$  be so large that  $h \leq \epsilon$  and  $\delta \leq \epsilon$ . Let  $J$  be the largest integer such that  $F^*(j\delta+1) \leq \epsilon$ . We also let  $T = \{j : F^*((j+1)\delta + 1) - F^*(j\delta) > \epsilon, j \geq J, j=0, \pm 1, \dots\}$  and  $A_n = \bigcup_{j \in T} [j\delta, (j+1)\delta)$ . Since only retraction and monotonicity properties of his respective estimate  $G^*$  and  $G$  were used before Lemma 3.1 of Fox was applied, the following inequalities are still true for our estimates  $G^*$  and  $\hat{G}$ .

$$(2.33) \quad P[\hat{L} > 2\epsilon] = P\left(\bigcup_{y \in A_n} (\{\hat{G}(y) > G(y+2\epsilon) + 2\epsilon\} \cup \{\hat{G}(y) < G(y-2\epsilon) - 2\epsilon\})\right)$$

$$\leq P\left(\bigcup_{j\delta \in A_n} (\{G^*(j\delta) > G(j\delta+\epsilon) + \epsilon\} \cup \{G^*(j\delta) < G(j\delta-\epsilon) - \epsilon\})\right)$$

$$\leq \sum_{j\delta \in A_n} P(\{G^*(j\delta) > G(j\delta+\epsilon) + \epsilon\} \cup \{G^*(j\delta) < G(j\delta-\epsilon) - \epsilon\}).$$

Since there are at most  $(\delta^{-1}+1)[\epsilon^{-1}+1]$  grid points (see Fox(1970, p.1850) in  $A_n$ , by Lemma 2.2 the extreme rhs(33) is no larger than rhs(32).  $\square$

Let  $\hat{\theta}$  be the procedure whose component procedures are Bayes versus  $\hat{G}$  defined by (31). To get a rate of convergence of the modified regret for  $\hat{\theta}$  we use the bound of Theorem 1.1. Since this bound is valid only for  $\Omega = [c, d]$  where  $-\infty < c \leq d < +\infty$ , we assume  $p(f)$  with  $\Omega = [c, d]$ .

Theorem 2.1. If  $P_j \in p(f)$  with  $\Omega = [c, d]$ ,  $j = 1, 2, \dots, n$  where  $f^{-1}$  satisfies the Lipschitz condition (1), then there exist constants  $b_2$  and  $b_3$  so that, for  $\hat{\theta}$  with  $b_2 h = b_3 \delta = (n^{-1} \log n)^{1/4}$ ,

$$|D(\theta, \hat{\theta})| = O((n^{-1} \log n)^{1/4}), \text{ uniformly in } \theta \in [c, d]^n.$$

Proof. We use Theorem 1.1 and apply Lemma 2.3. Then, choosing  $\varepsilon = \delta = (2b_1 + 1)h < 1$  (for sufficiently large  $n$ ) and weakening the bound gives

$$(2.34) \quad |D(\theta, \hat{\theta})| \leq b_4 h + b_5 h^{-2} \exp\{- (nh^4/b_6)\}$$

where  $b_4$  and  $b_5$  are some constants, and  $b_6 = 2\{1 + 3(d-c+3)M\}^2$ .

Choose  $b_2$  and  $b_3$  so that  $b_2 \leq 4^{1/4} (3b_6)^{-1/4}$  and  $b_3 = b_2 (2b_1 + 1)^{-1}$ . Then, for  $b_2 h (= b_3 \delta) = (n^{-1} \log n)^{1/4}$ , (34) leads to the asserted rate in Theorem 2.1.  $\square$

### 3. A Counterexample to $D(\theta, \hat{t}) \rightarrow 0$ on $R^\infty$ .

In Section 2 we demonstrated a procedure  $\hat{\theta}$  such that  $|D(\theta, \hat{\theta})| = O((n^{-1} \log n)^{1/4})$  uniformly in  $\theta$  in case of a bounded parameter set  $\Omega = [c, d]$ . Here we prove that the boundedness assumption on  $\Omega$  is necessary for the modified regret to converge to zero.

Theorem 3.1. Let  $X_1, X_2, \dots$  be independent random variables where for each  $j$ ,  $X_j \sim U[\theta_j, \theta_j + 1]$ ,  $\theta_j \in \Omega = R$ . Let  $\hat{t}(X) = (t_1(X), \dots, t_n(X))$  be an estimator of  $\theta = (\theta_1, \dots, \theta_n)$ ,  $n = 1, 2, \dots$ . Then there exists a sequence  $(\theta_1, \theta_2, \dots) \in R^\infty$  such that  $\overline{\lim}_n D(\theta, \hat{t}) > 0$ .

Proof.  $P_{\underline{x}}$  denotes the conditional distribution of  $(X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n)$  given  $x \equiv X_j$ . Since for each  $j$ ,  $P(t_j(X) - \theta_j)^2 \geq P_j(P_{\underline{x}}(t_j(X)) - \theta_j)^2$ , it follows that

$$(3.1) \quad D(\underline{\theta}, \underline{t}) \geq n^{-1} \sum_{j=1}^n P_j(P_{\underline{x}}(t_j(X)) - \theta_j)^2 = R(G).$$

Now, let  $\mu$  be a joint prior measure on  $(\theta_1, \theta_2, \dots)$ . Let  $\mu_{\theta_j}$  be the conditional measure given  $\theta_j$  and let  $\mu_j$  be the marginal measure of  $\theta_j$ . Then, setting  $s_j = \mu_{\theta_j} P_{\underline{x}}(t_j(X))$ ,  $j = 1, 2, \dots, n$ , we have that

$$(3.2) \quad \mu \{ n^{-1} \sum_{j=1}^n P_j(P_{\underline{x}}(t_j(X)) - \theta_j)^2 \} \geq n^{-1} \sum_{j=1}^n \mu_j P_j(s_j - \theta_j)^2.$$

Now consider  $\mu = \mu_1 \times \mu_2 \times \dots$  where  $\mu_j$  puts mass  $1/2$  on each of the values  $2j \pm r$ ,  $j \geq 1$ , where  $r$  is some fixed number such that  $0 < r < 1/2$ . Then

$$(3.3) \quad \begin{aligned} \mu_j P_j(s_j - \theta_j)^2 &= 2^{-1} P_{2j-r}(s_j - (2j-r))^2 + 2^{-1} P_{2j+r}(s_j - (2j+r))^2 \\ &\geq \int_{2j+r}^{2j+1-r} \{ 2^{-1} (s_j - (2j-r))^2 + 2^{-1} (s_j - (2j+r))^2 \} dx \geq r^2(1-2r), \end{aligned}$$

where the last inequality follows since the integrand on the lhs is not less than  $r^2$ .

Since  $R(G) = n^{-1} \sum_{j=1}^n P_j(\theta_{jn} - \theta_j)^2$  where  $\theta_{jn}$  is defined by the posterior mean (0.3) with  $q \equiv 1$ , and since the  $\theta_j$ 's are apart from each other more than 1,  $\theta_{jn} = \theta_j$  for all  $j$  and hence  $R(G) = 0$ .

Thus,  $\mu(R(G)) = 0$ . Therefore, in view of (1), (2) and (3),

$$(3.4) \quad \mu \{ D(\underline{\theta}, \underline{t}) \} \geq r^2(1-2r)$$

for all  $n$ . The retraction  $t^*$  of  $t$  formed by taking  $t_j^* = (x_j' \wedge t_j) \vee x_j$  has modified regret bounded by 1 and satisfies (4). Therefore, using Fatou's lemma gives

$$(3.5) \quad \mu\{\overline{\lim}_n D(\theta, t^*)\} \geq \overline{\lim}_n \{\mu D(\theta, t^*)\} \geq r^2(1-2r) > 0.$$

By  $\overline{\lim}_n D(\theta, t) \geq \overline{\lim}_n D(\theta, t^*)$  and (5), there exists a  $(\theta_1, \theta_2, \dots) \in \mathbb{R}^\infty$  such that  $\overline{\lim}_n D(\theta, t) > 0$ .

#### Acknowledgement.

The author heartily expresses thanks to Professor James F. Hannan for suggesting the problem and for his excellent guidance and encouragement during this research.

## REFERENCES

- [1] Feller, W. (1971), An Introduction to Probability Theory and its Applications Volume II (2nd ed.), Wiley, New York.
- [2] Ferguson, T. S. (1967), Mathematical Statistics a Decision Theoretic Approach, Academic Press.
- [3] Fox, R. (1968), Contribution to compound decision theory and empirical Bayes squared error loss estimation, Research Memorandum RM-214, Department of Statistics and Probability, Michigan State University.
- [4] Fox, R. (1970), Estimating the empiric distribution function of certain parameter sequences, Ann. Math. Statist. 41, 1845-1852.
- [5] Halmos, P. R. (1950), Measure Theory, Litton Educational Publishing.
- [6] Hoeffding, W. (1963), Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. 58, 13-30.
- [7] Neveu, J. (1965), Mathematical Foundations of the Calculus of Probability, Holden-Day.
- [8] Nogami, y. (1975), A non-regular squared-error loss set-compound estimation problem, RM-345, Department of Statistics and Probability, Michigan State University.
- [9] Royden, H. L. (1963), Real Analysis, The Macmillan Company, New York.
- [10] Saks, S. (1937), Theory of the Integral (2nd ed.), Monografie Matematyczne.
- [11] Singh, Radhey S. (1974), Estimation of derivatives of average of  $\mu$ -densities and sequence-compound estimation in exponential families, RM-318, Department of Statistics and Probability, Michigan State University.