A NON-REGULAR SQUARED-ERROR LOSS SET COMPOUND ESTIMATION PROBLEM

Dissertation for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

YOSHIKO NOGAMI

1975

This is to certify that the

thesis entitled

A NON-REGULAR SQUARED ERROR LOSS
SET COMPOUND ESTIMATION PROBLEM

presented by

Yoshiko Nogami

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics and Probability

James Hannan
Major professor

Date June 17, 1975

O-7639

ABSTRACT

A NON-REGULAR SOUARED ERROR LOSS
SET COMPOUND ESTIMATION PROBLEM

By

Yoshiko Nogami

For an integrable function $f \geq 0$, let $\mathcal{P}(f)$ be the family of distributions P_θ specified by a density proportional to the restriction of f to the interval $[\theta, \theta + 1)$ for θ in a real interval Ω. The component problem is estimation of θ based on X distributed according to P_θ, with squared-error loss. For a prior G on Ω, let $R(G)$ denote the Bayes risk versus G in the component problem.

Let $X_1, \ldots, X_n$ be n independent random variable with each X_j having $P_{\theta_j} \in \mathcal{P}(f)$. Let G_n be the empiric distribution of $\theta_1, \ldots, \theta_n$.

The work here is a generalization and continuation of R. Fox's (1968, 1970) work. Under P_θ, the uniform distribution on $[\theta, \theta + 1)$, he constructed a Lévy consistent distribution-valued estimate $\hat{G}_n$ of G_n. When the θ are iid G, he showed the convergence to $R(G)$ of the respective expected risks for $\underset{\sim}{\hat{\theta}}$ with components Bayes versus $\hat{G}_n$ and for $\underset{\sim}{\phi}$ with components direct estimates of the posterior means wrt G.

In this work we introduce procedures $\underset{\sim}{\theta}$, $\underset{\sim}{\phi}$ and $\underset{\sim}{\theta}_T$ (another direct estimate of the posterior mean wrt G_n). We generalize

Fox's $\hat{G}_n$ to $\mathcal{O}(f)$ and , with all convergence rates for bounded Ω, show that $D(\underset{\sim}{\theta}, \underset{\sim}{\hat{\theta}}) = E\{n^{-1} \Sigma_{j=1}^{n}(\hat{\theta}_j(\underset{\sim}{X}) - \theta_j)^2\} - R(G_n)$ is $O((n^{-1}\log n)^{\frac{1}{4}})$, even when $f \equiv 1$, the boundedness of Ω is necessary for the convergence of $D(\underset{\sim}{\theta},\underset{\sim}{t})$ to zero whatever be the set compound procedure $\underset{\sim}{t}$. The proof is based on the bound obtained for the risk difference in terms of Lévy distance. In this connection we obtain a unified generalization of Lemmas 8 and 8' of Oaten (1969).

For a prior G^k on Ω^k, let $R^k(G^k)$ denote the Bayes risk against G^k in squared error loss estimation of θ_k, based on $X_1,\ldots,X_k$. Let $\underset{\sim}{\theta}{}_j^k = (\theta_{j-k+1},\ldots,\theta_j)$, $j = k,\ldots,n$ and G_n^k be the empiric distribution of $\underset{\sim}{\theta}{}_k^k,\ldots,\underset{\sim}{\theta}{}_n^k$. Then, $\underset{\sim}{\theta}_T$ for $\mathcal{O}(f)$ and $\underset{\sim}{\phi}$ for $\mathcal{O}(1)$, both have $O(n^{-1/(2k+2)})$ for $D^k(\underset{\sim}{\theta},\underset{\sim}{\theta}_T) = E\{(n-k+1)^{-1} \Sigma_{j=k}^{n}(\theta_{T,j}(\underset{\sim}{X}) - \theta_j)^2\} - R^k(G_n^k)$ and $D^k(\underset{\sim}{\theta},\underset{\sim}{\phi})$. It is shown that $D(\underset{\sim}{0}, \underset{\sim}{\theta}_T)$ has exact order $n^{-\frac{1}{2}}$.

A NON-REGULAR SQUARED-ERROR LOSS SET COMPOUND ESTIMATION PROBLEM

By

Yoshiko Nogami

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1975

TO MY PARENTS

ACKNOWLEDGEMENTS

I wish to express my deep gratitude to Professor James F. Hannan for his excellent guidance and patience during the preparation of this dissertation. His invaluable suggestions helped greatly to improve the manuscript.

Among all others I thank him and Professor V. Fabian for their encouragement and advice during my study at Michigan State University. My appreciation extends to Professor D. Gilliland for his careful reading of my rough manuscript and suggesting appropriate changes. Furthermore, I thank Professors V. Fabian, J. Stapleton and J. Shapiro for reading the thesis. I am also indebted to Noralee Burkhardt for her excellent typing and patience.

Finally, I am grateful to the Department of Statistics and Probability at Michigan State University for the financial support.

TABLE OF CONTENTS

CHAPTER 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.

For a non-negative integrable function f, let $\mathcal{P}(f)$ denote the family of probability measures $\{P_\theta | \theta \in \text{a real interval } \Omega\}$ with P_θ specified by a density proportional to the restriction of f to the interval $[\theta, \theta+1)$. In this thesis, the component problem considered is the squared-error loss estimation of θ based on X with distribution $P_\theta \in \mathcal{P}(f)$. For any prior distribution G on Ω, let $R(G)$ be the Bayes risk versus G in this component problem.

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

(1) $$D(\underset{\sim}{\theta}, \underset{\sim}{t}) = \int n^{-1} \Sigma_{j=1}^{n}(t_j(\underset{\sim}{X}) - \theta_j)^2 d\underset{\sim}{P} - R(G_n)$$

where $\underset{\sim}{P} = P_{\theta_1} \times\ldots\times P_{\theta_n}$.

A bootstrap procedure based on component procedures Bayes versus an estimate of G_n will be called a two-stage procedure, while a procedure based on a direct estimate of the component Bayes procedure versus G_n will be called a one-stage procedure.

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 the Empirical Bayes problem where the θ_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 two-stage procedure $\hat{\underset{\sim}{\theta}}$ based on $\hat{G}_n$ and for a certain one-stage procedure $\underset{\sim}{\phi}$.

The behavior in the compound problem of the generalizations of these procedures is the subject of this thesis.

If $\sup\{|D(\underset{\sim}{\theta}, \underset{\sim}{t})| : \underset{\sim}{\theta} \in \Omega^n\} = O(n^{-\alpha})$, then we will say $\underset{\sim}{t}$ has a rate α. All rates are obtained only for bounded Ω.

Chapter I is concerned with a two-stage procedure $\hat{\underset{\sim}{\theta}}$. In Section 1 an upper bound of $D(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}})$ for $\hat{\underset{\sim}{\theta}}$ based on any distribution-valued estimate $\hat{G}_n$ of G_n is obtained. In Section 2 we show that there is a $\hat{\underset{\sim}{\theta}}$ based on the generalization $\hat{G}_n$ of Fox (1970) with a rate $\frac{1}{4}$ -.

Chapter II is specialized to $f \equiv 1$. We here deal with two one-stage procedures $\underset{\sim}{\theta}_T$ and $\underset{\sim}{\phi}$ (the latter is completed in Chapter III) where $\underset{\sim}{\theta}_T$ is based on retraction of the Bayes estimate versus a raw estimate of G_n to the interval $(X-1,X]$, while $\underset{\sim}{\phi}$ is based on estimates of a modified form of the Bayes estimate versus G_n. In Section 1, a $\underset{\sim}{\theta}_T$ with a rate $\frac{1}{4}$ is displayed and in Sections 2 and 3 it is shown that $\underset{\sim}{\theta}_T$ has exact order $\frac{1}{2}$ at $\underset{\sim}{\theta} = \underset{\sim}{0}$. Section 4 shows that at $\underset{\sim}{\theta} = \underset{\sim}{0}$, $\underset{\sim}{\phi}$ is sometimes better than $\underset{\sim}{\theta}_T$. Section 5 shows that when $\Omega = (-\infty,\infty)$, there is no sequence of estimates $\underset{\sim}{t}$ of $\underset{\sim}{\theta}$ for which $D(\underset{\sim}{\theta},\underset{\sim}{t})$ converges to zero.

Chapter III considers the k-extended problem. For any prior distribution G^k on Ω^k, let $R^k(G^k)$ denote the Bayes risk against G^k in the squared-error loss estimation of θ_k based on $X_1,\ldots,X_k$. Let $\underset{\sim}{\theta}_j^k = (\theta_{j-k+1},\ldots,\theta_j)$, $j = k,\ldots,n$, and G_n^k be the empiric distribution of $\underset{\sim}{\theta}_k^k,\ldots,\underset{\sim}{\theta}_n^k$. Then,

$$(1^k) \qquad D^k(\underset{\sim}{\theta}, \underset{\sim}{t}) = \int (n-k+1)^{-1} \Sigma_{j=k}^n (t_j(\underset{\sim}{X})-\theta_j)^2 d\underset{\sim}{P} - R^k(G_n^k)$$

is used as the standard in the k-extended problem.

In Chapter III we exhibit two one-stage k-extended procedures $\underset{\sim}{\theta}_T$ for $\mathcal{P}(f)$ and $\underset{\sim}{\phi}$ for $\mathcal{P}(1)$ with Ω bounded. These are respective generalizations of $\underset{\sim}{\theta}_T$ and $\underset{\sim}{\phi}$ introduced in Chapter II, and have rate $(2k + 2)^{-1}$.

In Appendix, unified generalizati[illegible] of Lemmas 8 and 8' of Oaten (1969, Appendix) is introduced in connection with Chapter I.

<u>Notational Conventions</u>.

P_j and $\underset{\sim}{P}$ abbreviate P_{θ_j} and $\times_{i=1}^n P_{\theta_i}$, respectively. A distribution function also represents the corresponding measure. We often let $P(h)$ or $P(h(\omega))$ denote $\int h(\omega)dP(\omega)$. G abbreviates the empiric distribution G_n of $\theta_1,\ldots,\theta_n$. R denotes the real line. We often abbreviate $y-1$ to y'. We denote the indicator function of a set A by $[A]$ or simply A itself. For any function h, $h]_a^b$ or $[h(\cdot)]_a^b$ means $h(b) - h(a)$. $\vee$ and $\wedge$ denote the supremum and the infimum, respectively. $\doteq$ denotes the defining property. We also use the notations $a_+ \doteq 0 \vee a$ and $a_- \doteq (-a)_+$. When we refer to (c,d) in the same section that we are dealing with, we simply write (d). Lemma b.d means Lemma

d in Chapter b. The symbol ▮ is used throughout to signal the end of a proof. EX and $Var(X)$ mean the expectation and variance of a random variable X.

CHAPTER I

A BOUND AND RATE FOR A TWO-STAGE PROCEDURE

§1.0. Introduction

Let f be a measurable function with $0 \le f \le 1$. With ξ Lebesgue measure, we 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 $p_\theta \doteq dP_\theta/d\xi$ we denote by $\mathcal{P}(f)$ the family of probability measures given by

$$(0.1) \qquad \mathcal{P}(f) = \{P_\theta \text{ with } p_\theta = q(\theta)[\theta,\theta+1)f,\ \forall\ \theta \in \Omega\}$$

where Ω is a real interval. The above assumptions apply throughout the body of this thesis.

Let $X_1,\ldots,X_n$ be n independent random variables with each X_j distributed according to $P_j \in \mathcal{P}(f)$ where P_j abbreviates P_{θ_j}. Denote the empiric distribution function of $\theta_1,\theta_2,\ldots,\theta_n$ by G without exhibiting the subscript n. With squared-error loss, let $\underset{\sim}{\theta}_G$ be the procedure whose component procedures are Bayes against G: $\underset{\sim}{\theta}_G(\underset{\sim}{X}) \doteq (\theta_{1n},\theta_{2n},\ldots,\theta_{nn})$ with, for each j,

$$(0.2) \qquad \theta_{jn} = G(\theta\, p_\theta(X_j))/G(p_\theta(X_j))$$

$$= \int_{X_j'+}^{X_j} \theta q(\theta)\, dG(\theta) / \int_{X_j'+}^{X_j} q\, dG$$

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.

A bootstrap procedure based on component procedures Bayes versus an estimate of G will be called a two stage procedure. Let $\hat{G}$ be a distribution-valued estimate of G. Define $\hat{\underset{\sim}{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_n)$ to be the two-stage procedure such that, for each j, $\hat{\theta}_j(\underset{\sim}{X}) = \hat{\theta}_{jn}$ is of form (0.2) with G replaced by $\hat{G}$ (0/0 is understood to be X_j).

The modified regret for a procedure $\underset{\sim}{t}$ is of form

$$(0.3) \qquad D(\underset{\sim}{\theta}, \underset{\sim}{t}) = n^{-1} \Sigma_{j=1}^{n} \{\underset{\sim}{P}(t_j(\underset{\sim}{X}) - \theta_j)^2 - \underset{\sim}{P}(\theta_{jn} - \theta_j)^2\} .$$

In §1 we exhibit an upper bound of $D(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}})$ (uniform wrt $\underset{\sim}{\theta}$ in Ω^n when Ω is bounded) in terms of $L(G,\hat{G})$, relying on Proposition A and Lemma A.3 in Appendix. In §2 we construct a particular distribution-valued Lévy consistent estimate $\hat{G}$ of G for $\Omega = R$. To show consistency, Theorem 2 of Hoeffding (1963) will be used. Under the additional assumption that $1/f$ satisfies a Lipshitz condition, we show by making use of the bound in §1 that the modified regret $D(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}})$ has a rate $\frac{1}{4}$ - when Ω is bounded.

In both sections, $\hat{L}$ abbreviates $L(G,\hat{G})$, and $\widehat{(a,b)}$ or $\widehat{(b)}$ mean (a,b) or (b) with G replaced by an estimate $\hat{G}$.

§1.1. An Upper Bound of the Modified Regret for $\tilde{\hat{\theta}}$.

In this section we shall exhibit a bound of the modified regret $D(\tilde{\theta}, \tilde{\hat{\theta}})$ for the two-stage procedure $\tilde{\hat{\theta}}$. To do so, the main development is Lemma 6 in which we show that the average expectation of $|\hat{\theta}_{jn} - \theta_{jn}|$ over the set where $\hat{L} < \epsilon$ is bounded by at most a constant times ϵ. For the proof of Lemma 6 we use a special case of Lemma A.2 of Singh (1974) and Proposition A and Lemma A.3 in the Appendix. Lemma 6^*, which improves the bound of Lemma 6 in the special case where $f \equiv 1$, is included because it also illustrates a different proof.

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 $X_1, \ldots, X_n$. Since $X_j' < \theta_{jn} \leq X_j$ by (0.2) 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 (0.3) that

$$(1.1) \qquad 2^{-1}|D(\tilde{\theta}, \tilde{\hat{\theta}})| \leq n^{-1} \Sigma_{j=1}^{n} \tilde{P}|\hat{\theta}_{jn} - \theta_{jn}| .$$

Now, Lévy distance for two distribution functions is defined by (A.1.1). For fixed j, since $|\hat{\theta}_{jn} - \theta_{jn}| \leq 1$, for any $\epsilon > 0$

$$(1.2) \qquad \tilde{P}|\hat{\theta}_{jn} - \theta_{jn}| \leq \tilde{P}[\hat{L} > \epsilon] + \tilde{P}(|\hat{\theta}_{jn} - \theta_{jn}|[\hat{L} \leq \epsilon]) .$$

Before dealing with the second term of rhs(2), we introduce four lemmas.

Lemma 1. For any $\varepsilon \geq 0$ and $\delta \geq 0$ with $\varepsilon + \delta < 1$,

$$n^{-1} \Sigma_{j=1}^{n} A_j \leq 1 + d - c$$

where $\forall$ j,

$$A_j = P_j\{[\theta_j + \delta \leq X_j < \theta_j + 1 - \varepsilon](\int_{X_j'+\varepsilon}^{X_j-\delta} q \, dG)^{-1}\} .$$

Proof. Since $\forall$ j,

$$(1.3) \quad A_j = \int \quad \{q(\theta_j)[\theta_j+\delta \leq y < \theta_j+1-\varepsilon]/\int_{y'+\varepsilon}^{y-\delta} q \, dG\}f(y)\,dy$$

and the average wrt j of the numerator of the quotient in rhs(3) equals the denominator, it follows that

$$n^{-1} \Sigma_{j=1}^{n} A_j = \int[\int_{y'+\varepsilon}^{y-\delta} q \, dG > 0]f(y)\,dy \leq 1 + d - c .$$

Lemma 2. For an arbitrary distribution function F of a random variable and any $s,t \in R$ with $s < t$,

$$\int F]_{y-t}^{y-s} dy = t - s .$$

Proof. By the Fubini theorem $\int F]_{y-t}^{y-s} dy = \int \int_{u+s}^{u+t} dy \, dF(u) =$ t-s. ∎

Lemma 3. For $s,t \in R$ with $s \le t$ and for any $\eta \in R$,

$$(1.4) \qquad n^{-1} \Sigma_{j=1}^{n} B_j \le t-s$$

where $\forall\, j$,

$$B_j = P_j\{G]_{X_j-t}^{X_j-s}[\theta_j - \eta \le X_j < \theta_j + 1 - \eta] / \int_{X_j'+\eta}^{X_j+\eta} q\, dG\} .$$

Proof. As in the proof of Lemma 1, $\forall\, j$,

$$(1.5) \qquad B_j = \int \frac{q(\theta_j)[\theta_j + \eta_- \le y < \theta_j + 1 - \eta_+]}{\int_{y'+\eta}^{y+\eta} q\, dG} f(y) G]_{y-t}^{y-s}\, dy .$$

Since $[\theta_j + \eta_- \le y < \theta_j + 1 - \eta_+] = [y' + \eta_+ < \theta_j \le y - \eta_-]$, the average wrt j of the numerator in the quotient is no more than the denominator. Also, since $f \le 1$, taking the average wrt j and interchanging the integral and average operation leads to lhs(4) $\le \int (G(y-s) - G(y-t))dy$, hence does not exceed $t-s$ by Lemma 2. ▌

Lemma 4. For all $s \in R$,

$$(1.6) \qquad n^{-1} \Sigma_{j=1}^{n} \underset{\sim}{P}\{|(G-\hat{G})(X_j - s)|[\hat{L} \le \epsilon] / \int_{X_j'}^{X_j} q\, dG\} \le \epsilon(3+d-c) .$$

Proof. For j fixed we let $z = X_j - s$. By the definition of $\hat{L}$ and remark stated after (A.1.1) (that the infimum in the definition of Lévy distance is attained), if $\hat{L} \le \epsilon$, then

(1.7) $$ {}_{-\varepsilon}(G^{\cdot}) \le \hat{G}^{\cdot} \le {}_{\varepsilon}(G^{\cdot}) $$

where $\cdot$ denotes addition of the identity and the pre-subscripts denote composition with those translations. Hence

$$ |(G-\hat{G})(z)|[\hat{L} \le \varepsilon] \le G^{\cdot}]_{z-\varepsilon}^{z} \vee G^{\cdot}]_{z}^{z+\varepsilon} $$

$$ \le \varepsilon + G]_{z-\varepsilon}^{z+\varepsilon} \, . $$

Hence,

$$ \text{lhs}(6) \le \varepsilon n^{-1} \Sigma_{j=1}^{n} P_j (\int_{X_j'}^{X_j} q \, dG)^{-1} $$

$$ + n^{-1} \Sigma_{j=1}^{n} P_j (G]_{X_j-s-\varepsilon}^{X_j-s+\varepsilon} / \int_{X_j'}^{X_j} q \, dG) \, . $$

Lemma 1 with $\varepsilon = \delta = 0$ and Lemma 3 with $(s,t,\eta) = (s-\varepsilon, s+\varepsilon, 0)$ lead to the bound of Lemma 4.∎

We will invoke a special case of Lemma A.2 of Singh (1974, Appendix) in the proof of forthcoming Lemma 6, and also in later sections (§2.1 and §3.1).

<u>Lemma 5</u> (Singh (1974)). For real random variables Y and Z, and real numbers y and z,

$$ E(|\frac{Y}{Z} - \frac{y}{z}| \wedge 1) \le 2|z|^{-1}\{E|Y-y| + (|\frac{y}{z}| + 1)E|Z-z|\} \, . $$

We shall now get an upper bound of the average wrt j of the second term of rhs(2).

Lemma 6. For $\varepsilon > 0$,

$$n^{-1} \Sigma_{j=1}^{n} \underset{\sim}{P}(|\hat{\theta}_{jn} - \theta_{jn}|[\hat{L} \le \varepsilon]) \le a_0 \varepsilon$$

where $a_0 = 4m\{16 + 21m + (6+9m)(d-c)\}$.

Proof. Fix n and $\underset{\sim}{\theta} \in [c,d]^n$. We also fix j thru (19). X abbreviates X_j. Since (0.2) $- X' = \int_{X'}^{X} (\theta - X')q(\theta)dG / \int_{X'}^{X} q(\theta)dG$, we abbreviate the quotient of the rhs to y/z and that with G replaced by $\hat{G}$ to Y/Z. Then,

$$(1.8) \qquad \hat{\theta}_{jn} - \theta_{jn} = Y/Z - y/z .$$

Let $*$ denote conditioning on X and $[\hat{L} \le \varepsilon]$. Then, by Lemma 5 and by the fact that $0 \le Y/Z,\ y/z \le 1$,

$$(1.9) \qquad \underset{\sim}{P}_{*}\left|\frac{Y}{Z} - \frac{y}{z}\right| \le \frac{2}{z} \underset{\sim}{P}_{*}(|Y-y| + 2|Z-z|) .$$

By letting $I = (X',X]$, G_I and $\hat{G}_I$ are defined in Proposition A in Appendix. Then, by Proposition A,

$$(1.10) \qquad L(G_I, \hat{G}_I) \le \hat{L} \vee S \vee T ,$$

where

(1.11) $$S = |(G - \hat{G})(X')| \quad \text{and} \quad T = |(G - \hat{G})(X)| \ .$$

Thus

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

By applying Lemma A.3 in Appendix, with $h(\theta)$, the restriction of $(\theta-X')q(\theta)$ to $(X',X]$, and weakening the resulted bound, when $L_I \le \lambda$,

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

To bound $\alpha(\lambda+)$, pick $\omega_1, \omega_2 \in I$ such that $0 < \omega_2 - \omega_1 < \lambda$. Now, by the definition of h,

(1.14) $$h\big]_{\omega_1}^{\omega_2} = (\omega_2-\omega_1)(q(\omega_2) + \frac{\omega_1 - X'}{\omega_2-\omega_1} q\big]_{\omega_1}^{\omega_2}) \ .$$

But, since by the definition of q,

$$q\big]_{\omega_1}^{\omega_2} = q(\omega_2)q(\omega_1)(\int_{\omega_1}^{\omega_2} f(s)ds - \int_{\omega_1+1}^{\omega_2+1} f(s)ds)$$

and since $q \le m$ and $0 \le f \le 1$, (1.15) $$|q]_{\omega_1}^{\omega_2}| \le m^2(\omega_2 - \omega_1) .$$

Thus, from (14),

$$|h]_{\omega_1}^{\omega_2}| \le (\omega_2 - \omega_1)\{q(\omega_2) + (\omega_1 - X')m^2\} .$$

Using $q \le m$, $\omega_1 - X' \le 1$ and $\omega_2 - \omega_1 \le \lambda$, and applying the definition of $\alpha(\lambda)$ gives us that

(1.16) $$\alpha(\lambda) \le \vee\{|h]_{\omega_1}^{\omega_2}| : \text{ for } \omega_1,\omega_2 \in I \ni 0 < \omega_2-\omega_1 < \lambda\}$$

$$\le \lambda(m + m^2)$$

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 $L_I \le \lambda$,

(1.17) $$|Y-y| \le 2(m + m^2)\varepsilon + (3m + 2m^2)(S + T) .$$

Similarly, by Lemma A.3 in Appendix with $1 \le h \doteq q \le m$, when $L_I \le \lambda$,

$$|Z-z| \le 2\alpha(\lambda+) + m(S + T) .$$

Since by the definitions of $\alpha(\lambda)$ and q and by (15)

$$\alpha(\lambda) \le \vee\{|q]_{\omega_1}^{\omega_2}| : \text{for } \omega_1, \omega_2 \in I \ni 0 < \omega_2 - \omega_1 < \lambda\}$$

$$\le m^2\lambda ,$$

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

$$|Z-z| \le 2m^2\varepsilon + (m + 2m^2)(S + T) . \tag{1.18}$$

Therefore, by (17), (18) and (12), when $\hat{L} \le \varepsilon$,

$$|Y-y| + 2|Z-z| \le 2(m + 3m^2)\varepsilon + (5m + 6m^2)(S + T) . \tag{1.19}$$

By this and in view of (9) and (8),

$$n^{-1}\Sigma_{j=1}^{n} \underset{\sim}{P}(|\hat{\theta}_{jn} - \theta_{jn}|[\hat{L} \le \varepsilon])$$

$$\le 4(m + 3m^2)\varepsilon(n^{-1}\Sigma_{j=1}^{n} P_j z^{-1})$$

$$+ 2(5m + 6m^2)n^{-1}\Sigma_{j=1}^{n} \underset{\sim}{P}\{(S + T)[\hat{L} \le \varepsilon]/z\} .$$

We apply Lemma 1 to the first term and use Lemma 4 twice to the second term. The result is the bound of Lemma 6. ▮

The following theorem is an immediate consequence of (1), (2) and Lemma 6.

Theorem 1. If $P_j \in \mathcal{P}(f)$ with $\Omega = [c,d]$, for $j = 1,2,\ldots,n$ then $\forall\ \varepsilon > 0$,

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

where a_0 is as defined in Lemma 6.

We can prove a strengthened version of Lemma 6 for $\mathcal{P}(1)$ using an alternative proof. To do so we need to introduce the following lemma.

Lemma 7. Let τ be a signed measure, h be a measurable function and $I = (y',y]$ be an interval with $\int I\ h\ d\tau \neq 0$. Let τ_y be the signed measure with density $Ih/\int I\ h\ d\tau$ wrt τ. Then,

$$\int s\ d\tau_y(s) = y - \int_0^1 \tau_y(y', y'+t]dt .$$

Proof. By Fubini's theorem applied to the lhs of the second equality below,

$$y - \int s\ d\tau_y(s) = \int \int_{s-y'}^1 dt\ d\tau_y(s) = \int_0^1 \tau_y(y', y'+t]dt . \blacksquare$$

Since $G(X) - G(X') \ge n^{-1}$, it follows by two applications of Lemma 7 above with $h = 1$, $\tau = \hat{G}$ and G to $\hat{\theta}_{jn}$ and θ_{jn} (cf. (0.2) and $(\widehat{0.2})$ with $q = 1$) that if $\hat{G}(X) - \hat{G}(X') > 0$, then

$$(1.20) \qquad \hat{\theta}_{jn} - \theta_{jn} = \int_0^1 (W - \hat{W})dt$$

where for $0 \le t \le 1$,

(1.21) $$W = G]_{X'}^{X'+t} / G]_{X'}^{X}$$

and $\hat{W}$ is given by $(\widehat{21})$.

Lemma 6*. If $P_j \in \mathcal{P}(1)$, $j = 1,2,\ldots,n$, then for $0 < \varepsilon < 2^{-1}$,

(1.22) $$n^{-1} \Sigma_{j=1}^{n} \underset{\sim}{P}(|\hat{\theta}_{jn} - \theta_{jn}|[\hat{L} \le \varepsilon]) \le 8(2 + d-c)\varepsilon .$$

Proof. Fix j until (35). Since $|\hat{\theta}_{jn} - \theta_{jn}| \le 1$,

(1.23) $$|\hat{\theta}_{jn} - \theta_{jn}| \le [\hat{G}]_{X'}^{X} = 0] + |\hat{\theta}_{jn} - \theta_{jn}|[\hat{G}]_{X'}^{X} > 0] .$$

Now, if $\hat{L} \le \varepsilon$, then (7) holds. Hence, $G]_{X'-\varepsilon}^{X-\varepsilon} \le \hat{G}]_{X'}^{X} + 2\varepsilon$. Thus

(1.24) $$\underset{\sim}{P}[\hat{G}]_{X'}^{X} = 0,\ \hat{L} \le \varepsilon] \le P[G]_{X'+\varepsilon}^{X-\varepsilon} \le 2\varepsilon]$$

$$\le P_j[\theta_j \le X < \theta_j + \varepsilon] + P_j[\theta_j + 1-\varepsilon \le X < \theta_j+1]$$

$$+ 2\varepsilon\, P_j\{[\theta_j + \varepsilon \le X < \theta_j + 1-\varepsilon](G]_{X'+\varepsilon}^{X-\varepsilon})^{-1}\} .$$

Therefore,

(1.25) $$n^{-1}\ \Sigma_{j=1}^{n}(\text{lhs}(24)) \le 2(2 + d - c)\varepsilon$$

because both the first and second terms of rhs(24) do not exceed ε and $(2\varepsilon)^{-1}$(third term of rhs(24)) $\le 1 + d - c$ by Lemma 1 with $q = 1$ and $\delta = \varepsilon$.

On the other hand, by (20) and weakening the integrand,

(1.26) $$\underset{\sim}{P}(|\hat{\theta}_{jn} - \theta_{jn}|[\hat{G}]_{X'}^{X} > 0,\ \hat{L} \le \varepsilon])$$

$$\le \underset{\sim}{P}(\textstyle\int_0^1 |W - \hat{W}|[\hat{L} \le \varepsilon]dt)\ .$$

For any a, b and $z \in R$, when $a \le b$, $((a-z)/(b-z)) = 1 - ((b-a)/(b-z))$ decreases from 1 to zero as z increases from $-\infty$ to a. Applying the above analysis to the representation $(\widehat{21})$ with a, b and z defined by positional correspondence in $(\widehat{21})$ and then applying (7) at X', we obtain that for $0 \le t \le 1$ and $\hat{L} \le \varepsilon$

$$\frac{\hat{G}(X'+t) - G(X'+\varepsilon) - \varepsilon}{\hat{G}(X) - G(X'+\varepsilon) - \varepsilon} \le \hat{W} \le \frac{\hat{G}(X'+t) - G(X'-\varepsilon) + \varepsilon}{\hat{G}(X) - G(X'-\varepsilon) + \varepsilon}\ .$$

Finally, making the lower bound smaller (and the upper bound larger) wrt $\hat{G}(X'+t)$ and $\hat{G}(X)$ and weakening by another four usages of $\hat{L} < \varepsilon$, results in

$$(1.27)\qquad (G]_{X'+\epsilon}^{X'+t-\epsilon}\ -2\epsilon)/G]_{X'+\epsilon}^{X+\epsilon} \le \hat{W} \le (G]_{X'-\epsilon}^{X'+t+\epsilon} + 2\epsilon)/G]_{X'-\epsilon}^{X-\epsilon}\ .$$

Note for future use that if $u \le \hat{W} \le v$, then

$$(1.28)\qquad |W - \hat{W}| \le (W - u)_+ + (v - W)_+\ .$$

Now, for any a, b, y and $z \in R$

$$(1.29)\qquad z\{b/a - (y-2\epsilon)/z\} = 2\epsilon + (b-y) + (b/a)(z-a)\ .$$

Let $u = \text{lhs}(27)$. With $W \doteq b/a$ and $u \doteq (y-2\epsilon)/z$ where the component quantities are defined by positional correspondence in the definitions of W and u, (29) and the relationships $z-a \le G]_X^{X+\epsilon}$, $0 \le (b/a) \le 1$, and $b-y = G]_{X'+t-\epsilon}^{X'+t} + G]_{X'}^{X'+\epsilon}$ give

$$(1.30)\qquad (G]_{X'+\epsilon}^{X+\epsilon})(W-u) \le 2\epsilon + G]_{X'+t-\epsilon}^{X'+t} + G]_{X'}^{X'+\epsilon} + G]_X^{X+\epsilon}\ .$$

Similarly, for any a, b, y and $z \in R$,

$$(1.31)\qquad z((y+2\epsilon)/z - (b/a)) = 2\epsilon+y-b + (b/a)(a-z)\ .$$

Let $v = \text{rhs}(27)$. With $v = (y+2\epsilon)/z$ and $W = b/a$ where the corresponding quantities are defined by the positional correspondence in v and W, (31) and the relationships $a-z \le G]_{X-\epsilon}^{X}$, $0 \le (b/a) \le 1$ and $y-b = G]_{X'+t}^{X'+t+\epsilon} + G]_{X'-\epsilon}^{X'}$ show

$$(1.32) \qquad (G]_{X'-\epsilon}^{X-\epsilon})(v-W) \le 2\epsilon + G]_{X'+t}^{X'+t+\epsilon} + G]_{X'-\epsilon}^{X'} + G]_{X-\epsilon}^{X} .$$

(26), (27), (28), (30) and (32) together give us that

$$(1.33) \quad |W - \hat{W}|[L \le \epsilon] \le (G]_{X'+\epsilon}^{X+\epsilon})^{-1}(\text{rhs}(30)) + (G]_{X'-\epsilon}^{X-\epsilon})^{-1}(\text{rhs}(32)).$$

By Lemma 2 , $\int_0^1 G]_{X'+t-\epsilon}^{X'+t}\,dt \le \epsilon$ and $\int_0^1 G]_{X'+t}^{X'+t+\epsilon}\,dt \le \epsilon$.
Thus

$$(1.34) \qquad \int_0^1 \text{rhs}(33)\,dt \le \alpha + \beta$$

where

$$\alpha = (3\epsilon + G]_{X'}^{X'+\epsilon} + G]_{X}^{X+\epsilon})/G]_{X'+\epsilon}^{X+\epsilon} \qquad \text{and}$$

$$\beta = (3\epsilon + G]_{X'-\epsilon}^{X'} + G]_{X-\epsilon}^{X})/G]_{X'-\epsilon}^{X-\epsilon} .$$

Bounding $\int_0^1 \text{lhs}(33)\,dt$ by 1 (since $0 \le W, \hat{W} \le 1$) over the sets $X^{-1}[\theta_j, \theta_j+\epsilon)$ and $X^{-1}[\theta_j+1-\epsilon, \theta_j+1)$, and extending the set

$X^{-1}[\theta_j + \varepsilon,\ \theta_j+1-\varepsilon)$ in two different ways (as shown below) we get

$$(1.35)\qquad n^{-1}\,\Sigma_{j=1}^{n}\,\underset{\sim}{P}(\int_0^1|W - \hat{W}|[\hat{L} \le \varepsilon]dt)$$

$$\le 2\varepsilon + n^{-1}\,\Sigma_{j=1}^{n}\,P_j(\alpha[\theta_j \le X < \theta_j+1-\varepsilon])$$

$$+\ n^{-1}\,\Sigma_{j=1}^{n}\,P_j(\beta[\theta_j+\varepsilon \le X < \theta_j+1])\ .$$

By applying Lemma 1 with $q = 1$, twice and Lemma 3 four times to the second and third terms of rhs(35), the second and third terms of rhs(35) are both $\le 3\varepsilon(1+d-c) + 2\varepsilon$.

Hence, in view of (35), (26), (25) and (23) we recognize that the sum of rhs(25) and $2\varepsilon + 2\{3\varepsilon(1+d-c) + 2\varepsilon\}$ gives us rhs(22). ∎

§1.2. A Particular Procedure $\hat{\underset{\sim}{\theta}}$ with a Rate (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 8) of Lemma 3.1 of Fox (1970). Then, we exhibit a distribution-valued estimate $\hat{G}$ of G. Lemma 9, 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 8. Finally, Theorem 2 shows that there exists a procedure $\hat{\underset{\sim}{\theta}}$ with a rate (1/4)-.

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

(2.1) $$\vee\{(v-u)^{-1}|(f(v))^{-1}-(f(u))^{-1}| \;:\; u < v\} \le M$$

for a finite constant M.

Let $\Omega = R$ until the proof of Lemma 9 is ended. Let Q be the distribution function defined by

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

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

(2.2) $$Q(y) = \Sigma \frac{\bar{p}(y-r)}{f(y-r)}$$

where Σ abbreviates $\Sigma_{r=0}^{\infty}$ throughout this section. Since $q \geq 1$ and q is the density of Q wrt G, it follows by Theorem 32.B of Halmos (1950) that

$$(2.3) \qquad G(y) = \int_{-\infty}^{y} (q(\theta))^{-1} dQ(\theta) \; .$$

For each y, we let

$$F^*(y) = n^{-1} \Sigma_{j=1}^{n} [X_j \leq y]$$

and for any $h > 0$

$$(2.4) \qquad \Delta F^*(y) = h^{-1} F^*]_y^{y+h} \; .$$

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 ξ is Lebesgue measure. We estimate $\bar{p}(y)$ by $\Delta F^*(y)$ and $Q(y)$ by

$$(2.5) \qquad Q^*(y) = \Sigma(\Delta F^*(y-r)/f(y-r)) \; .$$

Note that Q^* has bounded variation because of (1). From the relation (3), we obtain a raw estimate $\bar{W}$ of G from

$$(2.6) \qquad \bar{W}(y) = \int_{-\infty}^{y} (q(t))^{-1} dQ^*(t) \; .$$

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

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

Following Lemma 8 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 $2\exp\{-2nh^2\varepsilon^2\}$.

Lemma 8. If $0 < h \le \varepsilon \le 1$, then for each y

$$(2.7) \quad \underset{\sim}{P}(\{G(y-\varepsilon)-\varepsilon \le G^*(y) \le G(y+\varepsilon)+\varepsilon\}^c)$$

$$\le 2 \exp - \frac{2nh^2((\varepsilon-bh)_+)^2}{\{1+4(\theta_{(n)}-\theta_{(1)}+3)M\}^2}$$

where $\theta_{(1)} = \min_{1\le i\le n} \theta_i$, $\theta_{(n)} = \max_{1\le i\le n} \theta_i$ and $b = 2^{-1}m(2M+3(1 \wedge M))$.

Proof. For $y > \theta_{(n)} + 1$, $F^*(y) = G^*(y) = G(y+\varepsilon) = 1$ and for $y < \theta_{(1)}-1$, $F^*(y+1) = G^*(y) = G(y-\varepsilon) = 0$; in both cases lhs(7) = 0 and (7) holds trivially.

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

Pick $y \in [\theta_{(1)}-1, \theta_{(n)}+1]$. Since the summation on r in (5) involves at most a finite number of non-zero terms, we shall freely interchange integral and summation on r without further comment. In fact, if the r-th term is non-zero, then $r \le y - \theta_{(1)} + h$ and, for $y \le \theta_{(n)} + 1$,

$$r \le \theta_{(n)} - \theta_{(1)} + 2 \doteq a - 1 \, . \tag{2.8}$$

For each j, let

$$W_j = \Sigma \int_{-\infty}^{y} (q(t))^{-1} d_t \{[t-r < X_j \le t-r+h](h\ f(t-r))^{-1}\} \, , \tag{2.9}$$

where the subscript t in d_t denotes the variable of integration. By the definition (6) of $\bar{W}$,

$$\bar{W}(y) = n^{-1} \Sigma_{j=1}^{n} W_j \, .$$

We are going to find an upper and a lower bound of $\underset{\sim}{P}\bar{W}(y)$ in order to apply Hoeffding's bound (1963, Theorem 2). To do so we shall find an upper and a lower bound of $\underset{\sim}{P}W_j$, $\forall$ j. Fix j until (24). We use the corresponding notations without subscript j until (28). Now, 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 and hence by two applications of it, it holds for finite signed measures.

Hence,

$$P_\theta\{\int_{-\infty}^{y}(q(t))^{-1}d_t([t-r < X \le t-r+h](f(t-r))^{-1})$$

$$= \int_{-\infty}^{y}(q(t))^{-1}d_t\{P_\theta([t-r < X \le t-r+h](f(t-r))^{-1})\} .$$

Thus, by the definition of W

$$\underset{\sim}{P}\, W = q(\theta)\ \Sigma \int_{-\infty}^{y}(q(t))^{-1}d_t S(t-r) \tag{2.10}$$

where

$$S(t) = (f(t))^{-1}h^{-1} \int_t^{t+h}[\theta \le s < \theta+1]f(s)ds . \tag{2.11}$$

Because a function satisfying Lipshitz condition is absolutely continuous (cf. Royden (1968), p. 108, Exercise 16(a)) and the product of two absolutely continuous functions is absolutely continuous, $S(\cdot-r)$ is absolutely continuous. Since $1/q$ is clearly absolutely continuous, $S(\cdot-r)$ 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

$$\int_{-\infty}^{y}(q(t))^{-1}d_t S(t-r) = \frac{S(y-r)}{q(y)} - \int_{-\infty}^{y} S(t-r)f]_t^{t+1}dt . \tag{2.12}$$

Now, by the assumption (1),

$$(2.13)\qquad \left|\frac{f(s)}{f(t)} - 1\right| \le M|s-t| \ .$$

Until (22), we use the notation

$$(2.14)\qquad \Delta(t) = h^{-1}\int_t^{t+h}[\theta \le s < \theta+1]ds \ .$$

Applying (13) to the definition (11) of S and doing exact integration leads to the inequalities

$$(2.15)\qquad 1 - \frac{Mh}{2} \le S(t)/\Delta(t) \le 1 + \frac{Mh}{2} \ .$$

Moreover, because $\Sigma\Delta(y-r) = h^{-1}\int_y^{y+h}[\theta < t]dt$,

$$[\theta \le y] \le \Sigma\Delta(y-r) \le [\theta \le y + h] \ .$$

Hence, weakening the bounds by a usage of $[\theta \le \cdot]/q(y) \le 1$, shows that

$$(2.16)\qquad \frac{[\theta \le y]}{q(y)} - \frac{Mh}{2} \le \frac{\Sigma\, S(y-r)}{q(y)} \le \frac{[\theta \le y+h]}{q(y)} + \frac{Mh}{2} \ .$$

On the other hand, in the integral of rhs(12) we make a change of the variable $t-r$ to t to get

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

Let $[z]$ denote the greatest integer $\le z$ if $z > 0$ and -1 if $z < 0$. Since $\Sigma[t \le y-r](f(t+r+1)-f(t+r)) = [t \le y](f(t+[y-t]+1)-f(t))$ $= f(t+[y-t]+1)-f(t)$ (the latter because $[y-t] = -1$ if $t > y$), it follows that

$$\text{(2.17)} \qquad \Sigma \int_{-\infty}^{y} S(t-r) f\,]_t^{t+1} dt = \int S(t) f\,]_t^{t+[y-t]+1} dt \;.$$

By one usage of (15) and the fact that $0 < f \le 1$ and $\int \Delta(t)\,dt = 1$,

$$\text{(2.18)} \qquad \left| \int (S(t)-\Delta(t)) f\,]_t^{t+[y-t]+1} dt \right|$$

$$\le \frac{Mh}{2} \int \Delta(t) \left| f\,]_t^{t+[y-t]+1} \right| dt \le \frac{Mh}{2} \;.$$

From the deviation, (17) holds for Δ in place of S. Thus,

$$\text{(2.19)} \qquad \int \Delta(t) f\,]_t^{t+[y-t]+1} dt = \int_{-\infty}^{y} \Sigma \Delta(t-r) f\,]_t^{t+1} dt \;.$$

But, by a change of variable $s+r = u$ in the definition (14) of Δ,

$$h\,\Sigma\,\Delta(t-r) = \Sigma \int_t^{t+h} [\theta + r \le u < \theta + r + 1]du$$

$$(2.20) \qquad = \int_t^{t+h} [\theta \le u]du$$

$$= [\theta\text{-}h \le t < \theta](t+h-\theta) + [\theta \le t]h \ .$$

Therefore, (19) equals $\int_{-\infty}^{y} h^{-1}\text{rhs}(20)(f(t+1)-f(t))dt$ which becomes

$$(2.21) \quad \{[\theta\text{-}h \le y < \theta]\int_{\theta-h}^{y} + [\theta \le y]\int_{\theta-h}^{\theta}\} \ \frac{t+h-\theta}{h} \ f]_t^{t+1} \ dt$$

$$+ [\theta \le y](\int_\theta^y f]_t^{t+1} \ dt) \ .$$

Since $|f(t+1)-f(t)| \le |(f(t))^{-1}-(f(t+1))^{-1}| \wedge 1 \le M \wedge 1$ and $\int_{\theta-h}^{y\wedge\theta} (t+h-\theta)dt \le h^2/2$, $|\text{first term of } (20)| \le (M \wedge 1)h/2$. (19) thru (21) together give us that

$$(2.22) \quad \int \Delta(t)f]_t^{t+[y-t]+1} \ dt - [\theta \le y](\int_\theta^y f]_t^{t+1} \ dt)| \le (M \wedge 1)h/2 \ .$$

Therefore, by (22) and (18), in view of (17)

$$(2.23) \quad [\theta \le y](\int_\theta^y f]_t^{t+1} \ dt) - (M + M\wedge 1)\frac{h}{2}$$

$$\le \text{lhs}(17) \le [\theta \le y] \int_\theta^y f]_t^{t+1} \ dt + (M + M\wedge 1)\frac{h}{2} \ .$$

Therefore, from the relations (10), (12) and the bounds (16), (23), we can see that $(\underset{\sim}{P}\, W)/q(\theta)$ is bounded above and below by (rhs(16) - lhs(23)) and (lhs(16)-rhs(23)), respectively. Since

$$[\theta \le y+h]/q(y) - [\theta \le y+h]/q(\theta) = [\theta \le y]\int_\theta^y f]_t^{t+1}\, dt$$

$$+ [y < \theta \le y+h]\int_y^\theta f]_{t+1}^{t}\, dt$$

for $h \ge 0$, weakening the above bounds by using

$$[y < \theta \le y+h]\int_y^\theta f]_{t+1}^{t}\, dt \le (M \wedge 1)h$$

and $q \le m$ results in

$$(2.24) \qquad [\theta \le y] - bh \le \underset{\sim}{P}\, W \le [\theta \le y+h] + bh$$

where b is as defined in the statement of this lemma.

Averaging (24) wrt j gives

$$(2.25) \qquad G(y) - bh \le \underset{\sim}{P}\, \bar{W}(y) \le G(y+h) + bh\ .$$

In order to apply Theorem 2 of Hoeffding (1963), we shall furthermore need to get an upper and a lower bound of W. As for the derivation of (12), $1/q$ and $1/f(\cdot - r)$ are of bounded variation, and integration by parts gives

$$\int_{-\infty}^{y}(q(t))^{-1}d_t([t-r < X \le t-r+h]/f(t-r))$$

(2.26)

$$= \frac{[y-r < X \le y-r+h]}{f(y-r)q(y)} - \int_{-\infty}^{y} \frac{[t-r < X \le t-r+h]}{f(t-r)} f]_t^{t+1} \, dt \ .$$

Then, in view of the definition (9) of W,

$$hW = \Sigma \ \mathrm{lhs}(26)$$

(2.27)

$$= \frac{1}{q(y)} \Sigma \frac{[y-r < X \le y-r+h]}{f(y-r)} - \int_{-\infty}^{y}[t < X \le t+h]\left(\frac{f(t+\lceil y-t\rceil+1)}{f(t)} - 1\right)dt \ .$$

In the summation of the first term of rhs(27), there are at most two positive terms. Applying (13) and then (8) gives that with a as defined in (8)

$$0 \le \text{first term of rhs(27)} \le 1+2aM \ .$$

In addition, by two applications of (13) and the fact that $y-X+h \le a-1$ (because $y \in [\theta_{(1)}-1, \theta_{(n)}+1]$, $\theta_{(1)} \le X$ and $h < 1$),

$$|\text{second term of rhs(27)}| \le aMh \ .$$

Therefore, weakening the bound by using $h < 1$ shows

$$-aM \le W \le 1 + 3aM \ .$$

Now, we go back to (25). Since $h \le \epsilon$, using the second inequality of (25) and applying Theorem 2 of Hoeffding (1963) gives

$$P[\overline{W}(y) > G(y+\epsilon)+\epsilon] \le P[\overline{W}(y) - P\overline{W}(y) > \epsilon - bh]$$

(2.28)

$$\le \exp\{-\frac{2nh^2((\epsilon-bh)_+)^2}{(1+4aM)^2}\} \quad .$$

Furthermore, by the first inequality of (25), $\{\overline{W}(y) < G(y-\epsilon)-\epsilon\} \subset \{P\,\overline{W}(y) - \overline{W}(y) > \epsilon - bh\}$. Hence by the symmetry of the tail bounds, $P[\overline{W}(y) < G(y-\epsilon)-\epsilon]$ has the same upper bound, rhs(28), which together with (28) gives us the asserted bound of Lemma 8. ∎

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

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

Let $\hat{L} = L(G, \hat{G})$ be Lévy metric of G and $\hat{G}$ (cf. (A.1.1)).

Lemma 9. (Fox (1970)). For any $\epsilon > 0$, if $h \le \epsilon$ and $\delta \le \epsilon$, then

(2.30) $$P[\hat{L} > 2\epsilon] \le (\delta^{-1} + 1)[\epsilon^{-1} + 1]\ \mathrm{rhs}(7).$$

Proof. We rely on the proof of Theorem 3.1 of Fox (1970). Pick $\epsilon > 0$ such that $h \le \epsilon$ and $\delta \le \epsilon$. Let J be the largest integer such that $F^*(J\delta + 1) \le \epsilon$. We also let $\mathcal{J} = \{j : F^*((j+1)\delta+1) - F^*(j\delta) > \epsilon,\ j \ge J,\ j = 0, \pm 1, \ldots\}$ and $A_n = \bigcup_{j \in \mathcal{J}} [j\delta, (j+1)\delta)$. Since only retraction and monotonicity properties of his respective estimates G^* and $\hat{G}$ were used before Lemma 3.1 of Fox was applied, the following inequalities are still true for our estimates G^* and $\hat{G}$.

$$(2.31)\quad \underset{\sim}{P}[\hat{L} > 2\epsilon] = \underset{\sim}{P}\Big(\bigcup_{y \in A_n} (\{\hat{G}(y) > G(y+2\epsilon)+2\epsilon\} \cup \{\hat{G}(y) < G(y-2\epsilon)-2\epsilon\})\Big)$$

$$\le \underset{\sim}{P} \bigcup_{j\delta \in A_n} (\{G^*(j\delta) > G(j\delta+\epsilon)+\epsilon\} \cup \{G^*(j\delta) < G(j\delta-\epsilon)-\epsilon\})$$

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

Since there are at most $(\delta^{-1}+1)[\epsilon^{-1}+1]$ grid points (see Fox (1970, p. 1850)) in A_n, by Lemma 8 the extreme rhs of (31) is no larger than rhs(30). ▮

Let $\hat{\underset{\sim}{\theta}}$ be the procedure whose component procedures are Bayes versus $\hat{G}$ defined by (29). To get a rate of convergence of the modified regret for $\hat{\underset{\sim}{\theta}}$ we use the bound of Theorem 1 in section 1. Since this bound is valid only for $\Omega = [c,d]$ where $-\infty < c \le d < +\infty$, we assume $\mathcal{P}(f)$ with $\Omega = [c,d]$.

<u>Theorem 2</u>. If $P_j \in \mathcal{P}(f)$ with $\Omega = [c,d]$, $j = 1,2,\ldots,n$ where f^{-1} satisfies the Lipshitz condition (1), then there exist constants b_1 and b_2 so that, for $\underset{\sim}{\hat{\theta}}$ with $b_1 h = b_2 \delta = (n^{-1}\log n)^{\frac{1}{4}}$,

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

<u>Proof</u>. We use Theorem 1 in section 1 with ε replaced by $2^{-1}\varepsilon$ and apply Lemma 9. Then, choosing $\varepsilon = \delta = (2b+1)h < 1$ (for sufficiently large n) and weakening the bound by changing $\theta_{(n)}$ and $\theta_{(1)}$ to d and c, respectively gives

$$|D(\underset{\sim}{\theta}, \underset{\sim}{\hat{\theta}})| \leq b_3 h + b_4 h^{-2} e^{-(nh^4/b_5)} \tag{2.32}$$

where b_3 and b_4 are some constants, and $b_5 = 2\{1+4(d-c+3)M\}^2$.

Choose b_1 and b_2 so that $b_1 \leq 4^{\frac{1}{4}}(3b_5)^{-\frac{1}{4}}$ and $b_2 = b_1(2b+1)^{-1}$. Then, for $b_1 h \ (= b_2\delta) = (n^{-1}\log n)^{\frac{1}{4}}$, (32) leads to the asserted rate in Theorem 2. ∎

CHAPTER II

RATES FOR ONE-STAGE PROCEDURES FOR A FAMILY OF UNIFORM DISTRIBUTIONS

§2.0. Introduction

In Chapter I two-stage procedures were developed for estimation of $\underset{\sim}{\theta}$ for the family $\mathcal{P}(f)$. For sufficiently smooth f and certain two-stage procedures $\hat{\underset{\sim}{\theta}}$, $D(\underset{\sim}{\theta}, \hat{\underset{\sim}{\theta}}) = O((n^{-1}\log n)^{\frac{1}{4}})$ uniformly in $\underset{\sim}{\theta}$ (cf. Theorem I.2). In this Chapter we consider one-stage procedures for estimating $\underset{\sim}{\theta}$ for the family $\mathcal{P}(1)$, i.e., where P_j is the uniform distribution $\mathcal{U}[\theta_j, \theta_j+1)$, and obtain $O(n^{-\frac{1}{4}})$.

Throughout this chapter let $X_1, X_2, \ldots$ be independent with X_j distributed according to $P_j = \mathcal{U}[\theta_j, \theta_j+1)$. We begin by motivating the structure of two one-stage procedures for estimating $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_n)$. For fixed j, $1 \leq j \leq n$, we abbreviate X_j to x. Then by (1.0.2) with $q = 1$ and Lemma I.7 with $h = 1$, $y = x$ and $\tau = G$, the empirical distribution of $\theta_1, \ldots, \theta_n$, $\underset{\sim}{\theta}_G(\underset{\sim}{X}) = (\theta_{1n}, \ldots, \theta_{nn})$ has jth coordinate

$$(0.1) \qquad \theta_{jn} = x - \int_0^1 G]_{x'}^{x'+t} / G]_{x'}^{x} \, dt .$$

For each y let $p_\theta(y) \doteq (dP_\theta/d\xi)(y) = [\theta \leq y < \theta+1]$ where ξ is Lebesgue measure. Then $\bar{p}(y) \doteq G(p_\theta(y)) = G(y) - G(y')$

which leads to the relationship

$$(0.2) \qquad G(y) = \Sigma_{r=0}^{\infty} \bar{p}(y-r) \ .$$

Let $\bar{P}(y) \doteq G(P_\theta(y))$ and F^* be the empiric distribution of $X_1,\ldots,X_n$. Since $\bar{p} = d\bar{P}/d\xi$, the divided difference ΔF^* (see (1.2.4)) is an estimate of $\bar{p}$. In view of (2) we estimate $G(y)$ by

$$(0.3) \qquad T_n(y) = \Sigma_{r=0}^{\infty} \Delta F^*(y-r) \ .$$

Thus, (1) suggests that to achieve small modified regret we might estimate $\theta_G(X)$ (hence $\theta = (\theta_1,\ldots,\theta_n)$) by θ_T where

$$(0.4) \qquad \theta_{T,j} = (x' \vee \varphi_{jn}) \wedge x$$

and

$$(0.5) \qquad \varphi_{jn} = x - \int_0^1 T_n]_{x'}^{x'+t} / T_n]_{x'}^{x} \, dt \ .$$

(Here and throughout this chapter quotients 0/0 are defined to be zero.) On the other hand, since $\int_0^1 (G(x'+t)-G(x'))/(G(x)-G(x'))dt = (\bar{P}(x)-G(x'))/\bar{p}(x)$, estimating $\bar{P}$ by F^*, G by T_n and $\bar{p}(x)$ by $\Delta F^*(x) = T_n(x)-T_n(x')$ leads to an alternative estimate ϕ where

$$(0.6) \qquad \phi_{jn} = (x' \vee \psi_{jn}) \wedge x$$

and

$$(0.7) \qquad \psi_{jn} = x - (F^*(x) - T_n(x'))/T_n]_{x'}^{x} .$$

Fox (1968) considered the estimate $\underset{\sim}{\phi}$ in the empirical Bayes problem and showed the convergence of the risk of ϕ_{nn} to minimum Bayes risk. In this chapter we study the risk behavior of the one-stage procedures $\underset{\sim}{\theta}_T$ and $\underset{\sim}{\phi}$.

In Section 1, with $\Omega = [c,d]$ where $-\infty < c \leq d < \infty$, we show that $\underset{\sim}{\theta}_T$ with $h = O(n^{-\frac{1}{4}})$ has modified regret converging with rate 1/4 (cf. Theorem 1).

In Sections 2 and 3 we investigate the risk behavior of $\underset{\sim}{\theta}_T$ at the parameter sequence where $\theta_j = 0$ for all j, i.e., with $X_1, X_2, \ldots$ iid $\mathcal{U}[0,1)$. Theorem 2 of Section 2 gives lower bounds for $D(\underset{\sim}{0}, \underset{\sim}{\theta}_T)$ which combined with an upper bound developed in Section 3 shows that $D(\underset{\sim}{0}, \underset{\sim}{\theta}_T)$ is of exact order $O(h^2)$ provided $n^{-\frac{1}{4}}h^{-1} = O(1)$ (cf. Theorem 3). This indicates that Theorem 1 can possibly be strengthened to rate 1/2.

In Section 4, we indicate that in the bounded parameter set case, $\underset{\sim}{\phi}$ also has modified regret $O(n^{-\frac{1}{4}})$ if $h = O(n^{-\frac{1}{4}})$. This result is not proved in Section 4; rather it is noted that it follows as a corollary to Theorem 2 of Chapter III (Section 2). Section 4 concentrates on the $\underset{\sim}{\theta} = \underset{\sim}{0}$ sequence and develops upper and lower bounds for $D(\underset{\sim}{0}, \underset{\sim}{\phi})$. By comparing the bounds established in

Sections 2, 3 and 4 we deduce that for large n the estimator $\underset{\sim}{\phi}$ is strictly better than $\underset{\sim}{\theta}_T$ at $\underset{\sim}{0}$ for certain choices of h.

Finally, in Theorem 5 (Section 5) we prove the necessity of the bounded Ω assumption for the existence of procedures $\underset{\sim}{t}$ such that $D(\underset{\sim}{\theta}, \underset{\sim}{t}) \to 0$ for all $\underset{\sim}{\theta}$.

Concerning our notation, in Sections 1 and 5 $P_{\underset{\sim}{x}}$ will denote the conditional distribution of $(X_1,\ldots,X_{j-1}, X_{j+1},\ldots,X_n)$ given $X_j = x$ and in sections 2, 3 and 4, the conditional distribution of $(X_1,\ldots,X_n)$ given $X_{n+1} = x$. $c_0, c_1, \ldots, c_4$ in Section 3 are some positive constants. Also, for convenience in writing bounds, we will take $h \le \frac{1}{2}$ throughout the chapter.

§2.1. A Procedure $\underset{\sim}{\theta}_T$ with a Rate 1/4.

Let $\Omega = [c,d]$ where $-\infty < c \le d < +\infty$, throughout this section. Since $X_j' < \theta_{T,j} \le X_j$, as in (1.1.1)

$$(1.1) \qquad 2^{-1}|D(\underset{\sim}{\theta}, \underset{\sim}{\theta}_T)| \le n^{-1}\sum_{j=1}^{n} \underset{\sim}{P}|\theta_{T,j} - \theta_{jn}| \;.$$

Fix j and let $x \doteq X_j$ until the end of the proof of Lemma 4. For the purpose of treating the estimator $\underset{\sim}{\theta}_T$ and throughout this section only, for a signed measure H and $0 \le t \le 1$, we abbreviate $H(x' + t) - H(x')$ to $\Delta_t H$ and furthermore Δ_1 to Δ. By (0.1), (0.4) and (0.5) it follows that

$$(1.2) \qquad |\theta_{T,j} - \theta_{jn}| \le |\int_0^1 \Delta_t G/\Delta G \, dt - \int_0^1 \Delta_t T_n/\Delta T_n dt| \wedge 1$$

where T_n is defined in (0.3). Applying Lemma I.5 (a special case of Lemma A.2 of Singh (1974)) and $0 \le \Delta_t G/\Delta G \le 1$, we obtain

$$(1.3) \qquad \underset{\sim x}{P}(\text{rhs}(2)) \le \frac{2}{\Delta G}\{\underset{\sim x}{P}|\int_0^1 \Delta_t G \, dt - \int_0^1 \Delta_t T_n dt| + 2\underset{\sim x}{P}|\Delta G - \Delta T_n|\} \;.$$

A series of lemmas and a proposition will be used to complete the development of a bound for rhs(1). We begin with Lemma 1 which is straightforward from the proof of Lemma 3.1 of Fox (1970).

Lemma 1. For every $s \in R$,

$$(1.4) \qquad \underset{\sim x}{P}(T_n(s)) = h^{-1}\int_s^{s+h} G_{\check{j}}(u)\,du + \frac{1}{nh}\Sigma_{r=0}^{\infty}[s - r < x \le s - r + h]$$

where $G_{\underset{\sim}{j}}(u) = G(u) - n^{-1}[\theta_j \leq u]$.

<u>Proof</u>. By the definition (0.3) of T_n,

$$(1.5) \quad T_n(s) = \frac{1}{nh} \Sigma_{i=1}^{n} \Sigma_{r=0}^{\infty}[s - r < X_i \leq s - r + h], \text{ for } s \in R .$$

The term associated with $i = j$ gives the second term of rhs(4).

The integral wrt $P_{\underset{\sim}{x}}$-measure of rhs(5) excluding the j-th term can be written as

$$(1.6) \quad h^{-1}\int n^{-1}\Sigma_{j\neq i=1}^{n}[\theta_i \leq u < \theta_i+1](\Sigma_{r=0}^{\infty}[s - r < u \leq s - r + h])du$$

$$= h^{-1}\int G_{\underset{\sim}{j}}]_{u-1}^{u}(\Sigma_{r=0}^{\infty}[s - r < u \leq s - r + h])du .$$

Since by a change of variable $v = u - 1$,

$$\int G_{\underset{\sim}{j}}(u-1)(\Sigma_{r=0}^{\infty}[s-r < u < s-r+h])du = \int G_{\underset{\sim}{j}}(v)(\Sigma_{r=1}^{\infty}[s-r < v \leq s-r+h])dv,$$

we have

$$(1.7) \qquad (6) = h^{-1}\int_{s}^{s+h} G_{\underset{\sim}{j}}(u)du$$

which gives us Lemma 1. ▌

<u>Lemma</u> <u>2</u>. For each $t \in [0,1]$

$$(1.8) \quad G]_{x'+h}^{x'+t}[t \geq h] - n^{-1} \leq P_{\underset{\sim}{x}}(\Delta_t T_n) \leq G]_{x'}^{x'+t+h} + \frac{1}{nh} .$$

Proof. Since $0 \le$ (second term of rhs(4)) $\le (nh)^{-1}$, using two applications of Lemma 1 with this fact, taking the upper and lower limits of ranges of integrations and performing simple computations gives

$$(1.9) \quad G_{Y_j}]_{x'+h}^{x'+t}[t \ge h] \le P_{\underset{\sim}{x}}(\Delta_t T_n) \le G_{Y_j}]_{x'}^{x'+t+h} + \frac{1}{nh} .$$

Weakening the bounds by three applications of inequalities $G(s) - n^{-1} \le G_{Y_j}(s) \le G(s)$ and using $[\theta_j \le x'] = 0$, results in the asserted bounds of Lemma 2. ▮

Since we are going to invoke Fubini representation of the integral here and in later sections (§2.3, §2.4 and §3.1), we assert it beforehand. For a positive integer r, a measure μ and a measurable function ζ,

$$(1.10) \quad \int \zeta^r d\mu = \int_0^\infty \mu[\zeta > s] ds^r + (-1)^r \int_0^\infty \mu[\zeta < -s] ds^r .$$

Proposition 1. Let $Y_1, \ldots, Y_r$ be independent and $a \le Y_i \le b$, $i = 1, 2, \ldots, r$. Let $\bar{Y} = r^{-1} \Sigma_{i=1}^r Y_i$. Then, for every η

$$E|\bar{Y} - \eta| \le |E\bar{Y} - \eta| + \frac{b-a}{\sqrt{r}} \cdot \sqrt{\pi/2} .$$

Proof. By the triangle inequality

$$(1.11) \quad E|\bar{Y} - \eta| \le |E\bar{Y} - \eta| + E|\bar{Y} - E\bar{Y}| .$$

Now, by Fubini representation (10) of the integral, Also, using the first inequality of Lemma 2, and weakening the bound by use of $G(x'+t) - G(x') \le 1$ and $1-h < 1$, gives us

$$(1.17)\quad \int_0^1 (\Delta_t G - \underset{\sim}{P}_x(\Delta_t T_n))dt \le \int_0^h G]_{x'}^{x'+t}\,dt + \int_h^1 G]_{x'}^{x'+h}\,dt + n^{-1}$$

$$\le h + G]_{x'}^{x'+h} + n^{-1} .$$

Thus, taking the maximum of two bounds of (16) and (17) (recall $h < 1/2$) and weakening the bound by a use of $(y+u) \vee (y+v+w) \le y+v+u\vee w$, leads to first term of rhs(15) $\le h + G]_{x'}^{x'+h} + \frac{1}{nh}$. Thus, in view of (15) and the fact $n^{-1} \le (n-1)^{-\frac{1}{2}}$, (14) is established. ▮

<u>Lemma</u> <u>4</u>.

$$(1.18)\quad \underset{\sim}{P}_x|\Delta T_n - \Delta G| \le G]_x^{x+h} + G]_{x'}^{x'+h} + (1 + \sqrt{\pi/2})\,\frac{1}{\sqrt{n-1}\,h} .$$

<u>Proof</u>. Define $W_i = ((n-1)/(nh))[x < X_i \le x+h]$ for $i \ne j$. Then, by definition (5) of T_n, we can directly verify $\Delta T_n = \overline{W}$. Since $0 \le W_i \le h^{-1}$ for all i, by Proposition 1 with $b-a = h^{-1}$,

$$(1.19)\quad \text{lhs}(18) \le |\underset{\sim}{P}_x(\Delta T_n) - \Delta G| + \frac{1}{\sqrt{n-1}\,h} \cdot \sqrt{\pi/2} .$$

But, by the second inequality of Lemma 2,

$$\underset{\sim}{P}_x(\Delta T_n) - \Delta G \le G]_x^{x+h} + (nh)^{-1},$$

and by the first inequality of Lemma 2,

$$\Delta G - P_{\underset{\sim}{x}}(\Delta T_n) \leq G]_{x'}^{x'+h} + n^{-1} .$$

Thus, taking the maximum of above two bounds and weakening the bound by use of $(u+v) \vee (s+t) \leq u+s+v\vee t$ gives us first term of $\text{rhs}(19) \leq G]_x^{x+h} + G]_{x'}^{x'+h} + (nh)^{-1}$. Therefore, in view of (19) and $n^{-1} \leq (n-1)^{-\frac{1}{2}}$ the bound (18) is obtained. ▮

We now go back to the inequality (3). Applying the bounds from Lemmas 3 and 4, we get

$$n^{-1} \Sigma_{j=1}^{n} P_j(\text{rhs}(3)) \leq 2(h+(3+\frac{7}{2}\sqrt{\pi/2})\frac{1}{\sqrt{n-1}\,h})n^{-1}\Sigma_{j=1}^{n} P_j(\Delta G)^{-1}$$

$$+ 4\cdot n^{-1}\Sigma_{j=1}^{n} P_j\{ G]_{X_j}^{X_j+h} / \Delta G\} + 6\cdot n^{-1}\Sigma_{j=1}^{n} P_j\{ G]_{X_j'}^{X_j'+h} / \Delta G\} .$$

Therefore, by Lemma I.1 with $\delta = \varepsilon = 0$ and by two applications of Lemma I.3 with $(\eta, s, t) = (0, 1-h, 1)$ and $= (0,-h,0)$, we finally obtain in view of (3), (2) and (1) that

$$|D(\underset{\sim}{\theta}, \underset{\sim}{\theta}_T)| \leq (10+2(d+1-c))h + 2(3+\frac{7}{2}\sqrt{\pi/2})(d+1-c)\frac{1}{\sqrt{n-1}\,h}$$

Setting $h = n^{-\frac{1}{4}}$ gives us

<u>Theorem</u> <u>1</u>. For $\underset{\sim}{\theta}_T$ defined by (0.4) with $h = n^{-\frac{1}{4}}$,

$$|D(\underset{\sim}{\theta}, \underset{\sim}{\theta}_T)| \leq O(n^{-\frac{1}{4}}) \text{ uniformly in } \underset{\sim}{\theta} .$$

<u>Remark</u> <u>1</u>. Theorem 1 of Chapter III is a k-extended generalization of Theorem 1 for non-regular families of distributions $\mathcal{P}(f)$. Its specialization to k=1 (unextended case) is itself a

generalization of Theorem 1 which concerned the $\mathcal{P}(1)$ family. Theorem 1 was presented because of its simplicity and its significance in the motivation of Theorem 1 of Chapter III.

§2.2. A Lower Bound of the Modified Regret $D(\underset{\sim}{0}, \underset{\sim T}{\theta})$.

Let $X_1,\ldots,X_{n+1}$ be i.i.d. random variables with the common distribution $P = \mathcal{U}[0,1)$. Let $\underset{\sim}{X} = (X_1,\ldots,X_{n+1})$. Here we consider $\underset{\sim T}{\theta}(\underset{\sim}{X}) = (\theta_{T,1},\ldots,\theta_{T,n+1})$ (for the definition, see (0.4) with n replaced by n+1). Since $\theta_{T,1},\ldots,\theta_{T,n+1}$ are identically distributed and since for all j, $\theta_{j,n+1} = 0$ (for the definition, see (0.1) with n replaced by n+1), abbreviating $\theta_{T,n+1}$ to $\hat{\theta}$ we see in view of (1.0.3) that the modified regret of $\underset{\sim T}{\theta}$ at $\underset{\sim}{\theta} = \underset{\sim}{0}$ is given by

$$(2.1) \qquad D(\underset{\sim}{0}, \underset{\sim T}{\theta}) = \underset{\sim}{P}\, \hat{\theta}^2 .$$

For fixed $x \doteq X_{n+1}$, we abbreviate $\varphi_{n+1,n+1}$ in the definition of $\hat{\theta}$ (see (0.5) with j and n replaced by n+1) to φ and exhibit an explicit form of φ in a.s. $\underset{\sim x}{P}$-sense.

Lemma 5. For every $x \in [0,1)$,

$$(2.2) \quad \varphi = (\Sigma_{j=1}^{n}(X_j-h)[x < X_j \le x+h]-h\, \Sigma_{j=1}^{n}[0 \le X_j \le x]$$

$$- h + \Sigma_{j=1}^{n}[0 \le X_j \le x'+h])/\Sigma_{j=1}^{n}[x < X_j \le x+h] \quad \text{a.e. } \underset{\sim x}{P}$$

Proof. Fix j and note that as a function of $t \in [0,1]$, $\Sigma_{r=0}^{\infty}[X_j-x'+r-h \le t < X_j-x'+r]$ is equal to zero, is equal to its first term, or is equal to the sum of its first two terms according to whether $1 < X_j-x'-h$, $X_j-x'-h \le 1 < X_j-x'$ or $X_j-x' \le 1$. Integrating over $t \in [0,1]$ for each case gives

$$\int_0^1 \Sigma_{r=0}^{\infty}[X_j-x'+r-h \le t < X_j-x'+r]dt = (x+h-X_j)[x < X_j \le x+h]$$

$$+ h[X_j \le x] .$$

Hence, it follows by the definition (see (1.5)) of T_{n+1}, abbreviated to T that

$$(2.3)\qquad ((n+1)h)\int_0^1 T]_{x'}^{x'+t}dt = \Sigma_{j=1}^{n+1}(x+h-X_j)[x < X_j \le x+h]$$

$$+ h\ \Sigma_{j=1}^{n+1}[X_j \le x] - \Sigma_{j=1}^{n+1}\ \Sigma_{r=0}^{\infty}[x'-r < X_j \le x'-r+h]\ .$$

But since $[x < x \le x+h] = 0$, $[x \le x] = 1$, $\Sigma_{r=0}^{\infty}[x'-r < x \le x'-r+h] = 0$ and a.e. $P_{\sim x}$, $\Sigma_{r=1}^{\infty}[x'-r < X_j \le x'-r+h] = 0$, we have

$$\text{rhs}(3) = x\ \Sigma_{j=1}^{n}[x < X_j \le x+h] - \Sigma_{j=1}^{n}(X_j-h)[x < X_j \le x+h]$$

$$+ h\ \Sigma_{j=1}^{n}[X_j \le x] + h - \Sigma_{j=1}^{n}[x' < X_j \le x'+h],\ \text{a.e. } P_{\sim x}\ .$$

On the other hand, since $[x < x \le x+h] = 0$,

$$((n+1)h)T]_{x'}^{x} = \Sigma_{j=1}^{n}[x < X_j \le x+h]\ .$$

Applying these to the definition of φ, we get the asserted expression for φ. ∎

In this section we need only deal with φ for $x < 1-h$, where the term $\Sigma_{j=1}^{n}[0 \le X_j \le x'+h]$ (cf. rhs(2)) vanishes. We also recognize that for $x < 1-h$, $P_{\sim x}[\varphi > x] = 0$. Hence, ϑ has the following simpler form:

$$(2.4)\qquad \vartheta = \begin{cases} x' \vee \varphi & \text{for } x \in [0,\ 1-h) \\ (x' \vee \varphi) \wedge x, & \text{for } x \in [1-h,\ 1)\ . \end{cases}$$

Now, we let

$$(2.5) \qquad J = [\varphi \geq x', \ x < 1-h]$$

and recognize by (1) and the definition of $\hat{\theta}$ that

$$(2.6) \qquad D(\underset{\sim}{0}, \underset{\sim}{\theta}_T) \geq \underset{\sim}{P}(\varphi^2 J) \ .$$

Let $\xrightarrow{\mathcal{D}}$ denote convergence in distribution. Also, $N(c,d)$ denotes the normal distribution with mean c and variance d. To get lower bounds for $D(\underset{\sim}{0}, \underset{\sim}{\theta}_T)$ (Theorem 2) we use the relation (6) and the fact that for fixed x, $h^{-1} \varphi J \xrightarrow{\mathcal{D}} -2^{-1}$ and $S_n \doteq (\sqrt{nh}\ \varphi + 2^{-1}\sqrt{nh^3}\ )J \xrightarrow{\mathcal{D}} N(0,x^2)$. We then apply a convergence theorem (cf. Loéve (1963) 11.4, A(i)):

$$(2.7) \qquad \text{If } U_n \xrightarrow{\mathcal{D}} U, \text{ then } \underline{\lim}\ E\ U_n^2 \geq E\ U^2 \ ,$$

where E means expectation, and Theorem A.1 in appendix. We shall first prepare Lemmas 6, 7 and 8 to prove the above two convergences in distribution for the proof of forthcoming Theorem 2.

Let $u = \Sigma_{j=1}^{n}[0 \leq X_j \leq x]$, $v = \Sigma_{j=1}^{n}[x < X_j \leq x+h]$ and $w = \Sigma_{j=1}^{n} X_j[x < X_j \leq x+h]$. We also define

$$X = (w-hv-xv-h)/(hv) ,$$

$$Y = (u-nx)/\sqrt{nx(1-x)} \qquad \text{and}$$

$$Z = (v-nh)/\sqrt{nh} \ .$$

Then, on the set J, φ of the form (2) is alternatively written as

$$\varphi = hX + \frac{x(nh)^{-\frac{1}{2}}Z}{1 + (nh)^{-\frac{1}{2}}Z} - \frac{\sqrt{x(1-x)}\, n^{-\frac{1}{2}}Y}{1 + (nh)^{-\frac{1}{2}}Z} \quad . \tag{2.8}$$

Lemma 6. Given $x \in (0,1)$, if h is a function of n such that $nh \to \infty$ and $h \to 0$, then

$$(Y, Z) \xrightarrow{\mathcal{D}} N(\underline{0}, I)$$

where $\underline{0}$ is the zero vector in R^2 and I is 2×2 identity matrix.

Proof. For each $x \in (0,1)$ we restrict to n such that $x < 1-h$. Pick t and s arbitrary, and let

$$V_j = n^{-\frac{1}{2}}\{s(x(1-x))^{-\frac{1}{2}}([0 \le X_j \le x]-x) + th^{-\frac{1}{2}}([x < X_j \le x+h]-h)\} ,$$

for $j = 1,2,\ldots,n$. Then, it is not hard to see that

$$\Sigma_{j=1}^{n} V_j = s\,Y + t\,Z \ .$$

Since the V_j are i.i.d., the characteristic function K of (Y, Z) at a point $(s,t) \in R^2$ is given by

$$K(s,t) = (J(1))^n \tag{2.9}$$

where J is the characteristic function of $V \doteq V_1$.

Since by XV (6.8) (Feller, 1971), for any complex numbers such that $|\alpha| \le 1$ and $|\beta| \le 1$,

$$|\alpha^n - \beta^n| \le n|\alpha - \beta|,$$

(2.10) $$|(J(1))^n - \exp(-\tfrac{1}{2}(s^2+t^2))| \le n|J(1) - \exp(-\tfrac{1}{2n}(s^2+t^2))| .$$

By the triangular inequality and by using $|1-y-e^{-y}| = O(y^2)$ as $y \to 0$,

(2.11) $$\text{rhs}(10) \le n|J(1) - 1 + \frac{s^2+t^2}{2n}| + O(n^{-1}) .$$

Now, from the Taylor development of characteristic functions by XV (4.14) (Feller, 1971) and from the fact that $J(0) = 1$, $J'(0) = i\, P_{\underset{\sim}{x}} V = 0$ and $J''(0) = -P_{\underset{\sim}{x}} V^2$, it follows that

$$|J(1) - 1 + \frac{1}{2} P_{\underset{\sim}{x}} V^2| \le \frac{1}{6} P_{\underset{\sim}{x}} |V|^3 .$$

Now, we verify that

$$P_{\underset{\sim}{x}} V^2 = n^{-1}\{(s^2+t^2) - t^2 h - 2stx(\sqrt{x}(1-x)^{-\frac{1}{2}} + \sqrt{1-x}\; x^{-\frac{1}{2}})\sqrt{h}\}$$

and

$$P_{\underset{\sim}{x}} |V|^3 = n^{-3/2}\{|s(x^{-1}-1)^{\frac{1}{2}} - th^{\frac{1}{2}}|^3 x + |t(1-h)h^{-\frac{1}{2}} - sx^{\frac{1}{2}}(1-x)^{-\frac{1}{2}}|^3 h$$

$$+ |sx^{\frac{1}{2}}(1-x)^{-\frac{1}{2}} + t\, h^{\frac{1}{2}}|^3 (1-x-h)\} .$$

Hence,

$$0 \le n^{-1}(s^2+t^2) - \underset{\sim}{P}_x V^2 \le O(n^{-1}h^{\frac{1}{2}})$$

and

$$\underset{\sim}{P}_x|V|^3 = O(n^{-3/2}\, h^{-\frac{1}{2}}) \ .$$

Hence, applying the triangular inequality leads to

$$\left|J(1) - 1 + \frac{s^2+t^2}{2n}\right| = O(n^{-1}\, h^{\frac{1}{2}} + n^{-3/2}\, h^{-\frac{1}{2}})) \ .$$

Thus, in view of (11), (10) and (9)

$$\left|K(t,s)-\exp(-\frac{s^2+t^2}{2})\right| = O(h^{\frac{1}{2}} + n^{-\frac{1}{2}}\, h^{-\frac{1}{2}} + n^{-1}) \ .$$

To get the conclusion we invoke the continuity theorem (cf. e.g. Breiman (1968), Theorem 11.6). ▌

We shall next prove $X \overset{P}{\to} -2^{-1}$ where $\overset{P}{\to} 0$ means convergence in probability $\underset{\sim}{P}_x$ for given x.

Lemma 7. Under the same assumption as Lemma 6,

$$X \overset{P}{\to} - 2^{-1} \ .$$

Proof. For given $x \in (0,1)$, we restrict to n such that $x < 1-h$. Then, X is written as

$$X = (C/(\frac{v}{nh})) - v^{-1} \qquad (2.12)$$

where $C = (nh)^{-1} \Sigma_{j=1}^{n} U_j$, where $U_j = h^{-1}(X_j - x - h) I_j$ with $I_j = [x < X_j \le x+h]$.

Since v has the binomial distribution with parameters n and h,

$$(2.13) \qquad \frac{v}{nh} \xrightarrow{P} 1 \quad \text{as} \quad nh \to \infty \quad \text{and} \quad h \to 0 \ .$$

By simple computations,

$$EU = -\frac{h}{2}$$

and

$$Var(U) = \frac{h}{12} + \frac{h(1-h)}{4} \ .$$

Thus, $EC = h^{-1}EU = -2^{-1}$ and $Var(C) = (nh^2)^{-1}Var(U) = (12^{-1}+(1-h)/4)/(nh)$. Therefore, by the Chebychev inequality,

$$(2.14) \qquad C \xrightarrow{P} -2^{-1} \quad \text{as} \quad nh \to \infty \quad \text{and} \quad h \to 0 \ .$$

Applying (14), (13), (12) and Slutsky's Theorem completes the proof of Lemma 7. ∎

Besides the above two lemmas we shall show that $P_{\underset{\sim}{x}}[\hat{\varphi} \le x']$ vanishes when $nh \to \infty$ and $h \to 0$.

Lemma 8. Under the same assumption as Lemma 6,

$$P_{\underset{\sim}{x}}[\varphi \le x'] \to 0 \quad \text{for fixed } x.$$

Proof. We restrict to n such that $x < 1-h$. Let $W_j = h[0 \le X_j \le x] - (X_j - h - x')[x < X_j \le x+h]$ for $j = 1,2,\ldots,n$. Then, by the representation (2) of φ, $[\varphi \le x'] = [\bar{W} \ge -n^{-1}h]$ where $\bar{W}$ is the average of i.i.d. W_j's. Since $P_{\underset{\sim}{x}} W_1 = h(2^{-1}h+x')$,

$$(2.15) \qquad P_{\underset{\sim}{x}}[\varphi \le x'] = P_{\underset{\sim}{x}}[\bar{W} - P_{\underset{\sim}{x}}\bar{W} \ge (1-x-n^{-1}-2^{-1}h)h] .$$

But, $\mathrm{Var}(\bar{W}) = n^{-1}\mathrm{Var}(W_1) = hn^{-1}\{1-(1-x)(2-x)h + (\frac{4}{3} - x)h^2 - 4^{-1}h^3\} \le (\frac{7}{3})hn^{-1}$. Hence, by the Chebychev inequality and for large n

$$\mathrm{rhs}(15) \le (7/3)h^{-1}n^{-1}(1-x-n^{-1}-2^{-1}h)^{-2}$$

which tends to zero when $nh \to \infty$ and $h \to 0$. ▮

We are now ready to prove

Theorem 2. (i) If h is a function of n such that $nh^3 \to \infty$ and $h \to 0$, then for any $\frac{1}{4} > \epsilon > 0$, there exists $N < +\infty$ so that for all $n \ge N$

$$D(\underset{\sim}{0}, \underset{\sim}{\theta}_T) > (\frac{1}{4} - \epsilon)h^2 .$$

(ii) If h is a function of n such that $nh \to \infty$, $h \to 0$ and $nh^3 = O(1)$, then for any $\frac{1}{3} > \epsilon > 0$, there exists $N < +\infty$ so that for all $n \ge N$

$$D(\underset{\sim}{0}, \underset{\sim T}{\theta}) > (\frac{1}{3} - \epsilon) \frac{1}{nh} .$$

<u>Proof</u>. (i) Since $nh^3 \to \infty$ and $h \to 0$ implies $nh \to \infty$ and $h \to 0$, we have by Lemmas 6, 7 and 8 that given $x \in (0,1)$,

$$(2.16) \qquad (Y,Z) \overset{\mathcal{D}}{\to} N(\underline{0},I), \; X \overset{P}{\to} -\frac{1}{2}, \; [\varphi \geq x'] \overset{P}{\to} 1 .$$

Hence, in view of (8) it follows from Slutsky's Theorem that if $x \in (0,1)$, then $h^{-1}\varphi J \overset{\mathcal{D}}{\to} -2^{-1}$ (see (5) for the definition of J). By a convergence theorem (7), we have

$$(2.17) \qquad \underline{\lim}\, \underset{\sim x}{P}(h^{-2} \varphi^2 J) \geq \frac{1}{4}[0 < x < 1],$$

and hence by Fatou's theorem applied to the lhs below

$$\underline{\lim}\, P\, \underset{\sim x}{P}(h^{-2}\varphi^2 J) \geq P(\text{lhs}(17)) \geq \frac{1}{4} .$$

Thus, by (6) we get that

$$\underline{\lim}\, h^{-2} D(\underset{\sim}{0}, \underset{\sim T}{\theta}) \geq \frac{1}{4} .$$

(i) follows because of the definition of lim inf.

To prove (ii) we first recognize that for this choice of h, (16) still holds. Let $S_n = \{\sqrt{nh}\, \varphi + 2^{-1} \sqrt{nh^3}\}\, J$. Then, in view of (8) it follows from Slutsky's theorem that if $x \in (0,1)$, then

$$S_n \xrightarrow{\mathcal{D}} N(0, x^2) \ .$$

Since $\underset{\sim}{P}_x\{(nh)\varphi^2 J\} = \underset{\sim}{P}_x(S_n - 2^{-1}\sqrt{nh^3}\, J)^2 \geq \mathrm{Var}(S_n)$, applying Theorem A.1 in Appendix to the rhs leads to

$$(2.18) \qquad \underline{\lim}\ \underset{\sim}{P}_x\{(nh)\varphi^2 J\} \geq x^2[0 < x < 1] \ .$$

Thus, by Fatou's Lemma applied to the lhs below

$$\underline{\lim}\ P\underset{\sim}{P}_x(nh\ \varphi^2 J) \geq P(\mathrm{lhs}(18))$$

$$\geq \int_0^1 y^2 dy = \frac{1}{3} \ .$$

Therefore by (6) we get that $\underline{\lim}\ (nh)D(\underset{\sim}{0}, \underset{\sim}{\theta}_T) \geq \frac{1}{3}$ and the definition of lim inf leads to (ii). ∎

Theorem 2(i) implies that at any parameter sequence $(\theta_1, \theta_2, \ldots)$ where $\theta_1 = \theta_2 = \ldots,\ \underset{\sim}{\theta}_T$ with the choice $h = n^{-\frac{1}{4}}$ has modified regret converging to zero at a rate no faster than $n^{-\frac{1}{2}}$. This leaves open the possibility of strengthening Theorem 1 of §1 by this improved rate. The next section develops a positive result in this direction by obtaining the improved rate at a fixed parameter sequence.

§2.3. Procedures $\underset{\sim}{\theta}_T$ where $D(\underset{\sim}{0}, \underset{\sim}{\theta}_T)$ is of Exact Order $O(h^2)$.

In this section we show that the modified regret $D(\underset{\sim}{0}, \underset{\sim}{\theta}_T)$ has an upper bound of order $O(h^2)$ when $n^{-\frac{1}{4}}h^{-1} = O(1)$, and by this choice of h we have a lower bound with the same order of magnitude. Specifically, if $h = n^{-\frac{1}{4}}$ (up to constants), then we get convergence of exact order $\frac{1}{2}$.

As in §2, let $X_1, \ldots, X_{n+1}$ be i.i.d. observations from $P = \mathcal{U}[0,1)$ and fix $x \doteq X_{n+1}$. $\hat{\theta}$ and φ abbreviate $\theta_{T,n+1}$ and $\varphi_{n+1,n+1}$, respectively. Let $v = \Sigma_{j=1}^n [x < X_j \leq x+h]$ so that (2.2) reads for each fixed $x \in [0,1)$,

$$(3.1) \quad \varphi = v^{-1}\{\Sigma_{j=1}^n \{(X_j - h)[x < X_j \leq x+h] - h[0 \leq X_j \leq x] + [0 \leq X_j \leq x'+h]\} - h\} \text{ a.e. } \underset{\sim}{P}_x .$$

Since $\hat{\theta}$ is the retraction of φ to $(x', x]$ where $0 \leq x < 1$, then $|\hat{\theta}| \leq 1$ and

$$(3.2) \qquad \underset{\sim}{P}\, \hat{\theta}^2 \leq \underset{\sim}{P}[v = 0] + \underset{\sim}{P}(\hat{\theta}^2[v > 0]) .$$

Note that

$$(3.3) \qquad P\, \underset{\sim}{P}_x[v = 0] = \int_0^{1-h} (1-h)^n dy + \int_{1-h}^1 (1-y)^n dy .$$

The first term on rhs(3) is $(1-h)^{n+1}$. Hence, the inequalities $1 = ((1-h)+h)^{n+2} \geq \binom{n+2}{1} h(1-h)^{n+1}$ imply that it is bounded by $((n+2)h)^{-1}$. The second term on rhs(3) is bounded by $(n+1)^{-1}$ so that (3) implies

$$(3.4)\qquad \underset{\sim}{P}[v = 0] = O((nh)^{-1}) .$$

Now, consider the last term on rhs(2). Since $x' < 0 \le x$, by Fubini representation of the integral (c f. (1.10) with $r = 2$) and (1) it follows that

$$(3.5)\qquad \underset{\sim}{P}(\hat{\theta}^2[v > 0]) \le P(\int_0^1 \underset{\sim}{P}_x[U > h]ds^2) + P(\int_0^1 \underset{\sim}{P}_x[V < h]ds^2)$$

where $U = \Sigma_j U_j$, $V = \Sigma_j V_j$, and for each $j = 1,2,\ldots,n$

$$(3.6)\qquad U_j = (X_j - s - h)[x < X_j \le x+h] - h[0 < X_j \le x] + [0 < X_j \le x'+h]$$

and

$$(3.7)\qquad V_j = U_j + 2s[x < X_j \le x+h] .$$

The U_j are i.i.d. with mean $-sh - 2^{-1}h^2$ for $x \in [0, 1-h)$ and mean $-s(1-x)-2^{-1}(1-x)^2$ for $x \in [1-h,1)$. Then, by (7), each V_j has mean

$$(3.8)\qquad \underset{\sim}{P}_x V_j = \begin{cases} sh - 2^{-1}h^2, & \text{for } x \in [0,1-h) \\ s(1-x)-2^{-1}(1-x)^2, & \text{for } x \in [1-h,1) . \end{cases}$$

We will use the bound 2 for the range of U_j and V_j for all $0 \le s \le 1$ and $0 \le x < 1$ in Theorem 2 of Hoeffding (1963) in order to bound the tail probabilities in (5). Moreover, the Hoeffding bound developed for the last term of rhs(5) will also bound the first term since $P_{\tilde{x}}V$ is closer to h than $P_{\tilde{x}}U$ is to h. Having noticed these facts we now prove

Lemma 9.

$$(3.9) \qquad \tilde{P}(\hat{\theta}^2[v > 0]) = O(h^2 + n^{-\frac{1}{2}} + (nh^2)^{-1}) .$$

Proof. In view of the comment preceding Lemma 9 it suffices to show that the last term of rhs(5) has the order indicated in (9).

Let $\bar{V} = n^{-1}V$. Using (8) and letting $a_1 = n^{-1} + \frac{1}{2}h$ and $a_2 = hn^{-1}(1-x)^{-1} + \frac{1}{2}(1-x)$ we have

$$P_{\tilde{x}}[V < h] = \begin{cases} P_{\tilde{x}}[-\bar{V} + P_{\tilde{x}}\bar{V} > h(s-a_1)], & \text{for } x \in [0,1-h) \\ P_{\tilde{x}}[-\bar{V} + P_{\tilde{x}}\bar{V} > (1-x)(s-a_2)], & \text{for } x \in [1-h,1) . \end{cases}$$

Applying Theorem 2 of Hoeffding (1963) with the bound 2 for the range of the V_j gives

$$(3.10) \qquad P_{\tilde{x}}[V < h] \le \begin{cases} \exp\{-\frac{1}{2}b_1((s-a_1)_+)^2\}, & \text{for } x \in [0, 1-h) \\ \exp\{-\frac{1}{2}b_2((s-a_2)_+)^2\}, & \text{for } x \in [1-h,1) \end{cases}$$

where $b_1 = nh^2$ and $b_2 = n(1-x)^2$. A direct calculation together with the fact that $\int_0^\infty \exp(-\frac{1}{2}y^2)\,dy = \sqrt{\pi/2}$, shows that

$$\int_0^1 \exp\{-\tfrac{1}{2}b((s-a)_+)^2\}ds^2$$

(3.11) $$\leq \int_0^a ds^2 + \int_a^\infty e^{-\frac{1}{2}b(s-a)^2} ds^2$$

$$\leq a^2 + (2b^{-1} + \sqrt{2\pi}\, a\, b^{-\frac{1}{2}})$$

where $a,b \geq 0$.

Applying (10) and (11) with $a = a_1$ and $b = b_1$ gives

(3.12) $$P\{[0 \leq x < 1-h]\int_0^1 P_{\tilde{x}}[V < h]ds^2\} = O(h^2 + n^{-\frac{1}{2}} + (nh^2)^{-1}).$$

The treatment of the $[1-h \leq x < 1]$ part involves more analysis because of the dependence of a_2 and b_2 on x. By (11) with $a = a_2$ and $b = b_2$, and by $h \leq 2^{-1}$,

(3.13) $$\int_0^1 P_{\tilde{x}}[V < h]ds^2 \leq \{\{\frac{h}{n(1-x)} + \frac{1-x}{2}\}^2 + c_1 n^{-\frac{1}{2}} + c_2 \frac{1}{n(1-x)^2}\} \wedge 1$$

$$\leq \tfrac{1}{4}h^2 + (\tfrac{1}{2} + c_1)n^{-\frac{1}{2}} + \{(\tfrac{1}{4} + c_2)\frac{1}{n(1-x)^2}\} \wedge 1,$$

for $x \in [1-h,1)$ and constants c_1 and c_2. But, for $d > 0$

(3.14) $$\int_{1-h}^1 \frac{d^2}{(1-y)^2} \wedge 1\, dy \leq \int_{(1-h)\vee(1-d)}^1 dy + \int_{1-h}^{(1-h)\vee(1-d)} \frac{d^2}{(1-y)^2}\, dy$$

$$= h \wedge d + d^2\{(h \wedge d)^{-1} - h^{-1}\} \leq 2d .$$

Thus,

(3.15) $$P([1-h \leq x < 1]\text{lhs}(13)) \leq O(h^3 + n^{-\frac{1}{2}}) .$$

Therefore, (9) follows from (15), (13), (12) and (5). ▮

Hence, by (2.1), (2), (4) and Lemma 9, we get the following lemma:

Lemma 10.

$$D(\underset{\sim}{0}, \underset{\sim T}{\theta}) = O(h^2 + n^{-\frac{1}{2}} + (nh^2)^{-1}). \tag{3.16}$$

Theorem 3. If $\underset{\sim T}{\theta}$ is defined by (0.4) with h such that $n^{-\frac{1}{4}}h^{-1} = O(1)$, then there exists a constant c_3 so that for sufficiently large n,

$$c_3^{-1} h^2 \le D(\underset{\sim}{0}, \underset{\sim T}{\theta}) \le c_3 h^2 .$$

Proof. If $n^{-\frac{1}{4}}h^{-1} = O(1)$, then there exists some constant $0 < M < +\infty$ such that

$$h \ge Mn^{-\frac{1}{4}} . \tag{3.17}$$

Therefore, by Lemma 10 there exists a constant c_4 such that $D(\underset{\sim}{0}, \underset{\sim T}{\theta}) \le c_4 h^2$ for sufficiently large n.

On the other hand, by (17), $nh^3 \ge n^{\frac{1}{4}}M^3 \uparrow +\infty$ as $n \uparrow +\infty$. Hence, (i) of Theorem 2 in section 2 holds. Letting $c_3 > c_4 \vee 4$ we pick $\varepsilon > 0$ in (i) of Theorem 2 so that $\varepsilon < 4^{-1} - c_3^{-1}$. Then, from (i) of Theorem 2 the first inequality in Theorem 3 follows. ▮

Remark. Here, we shall state values of h and of bounds of the modified regret, up to constants. That is, for example, $h = n^{-\alpha}$ means $h = c\, n^{-\alpha}$ for some positive constant c. We let large n be fixed.

The lower bounds of $D(\underset{\sim}{0}, \underset{\sim}{\theta}_T)$ in Theorem 2 describe the strictly convex curve which attains the minimum value $n^{-2/3}$ at $h = n^{-1/3}$, has the value $(nh)^{-1}$ for h less than $n^{-1/3}$ and h^2 for h greater than $n^{-1/3}$.

On the other hand, the upper bound in (16) describes the strictly convex curve which attains its minimum value $n^{-\frac{1}{2}}$ at $h = n^{-\frac{1}{4}}$, has the value $(nh^2)^{-1}$ for h less than $n^{-\frac{1}{4}}$ and h^2 for h greater than $n^{-\frac{1}{4}}$.

Hence, these two curves coincide with each other for h greater than and equivalent to $n^{-\frac{1}{4}}$ and attain the best exact order $\frac{1}{2}$ at $h = n^{-\frac{1}{4}}$. For h less than $n^{-\frac{1}{4}}$, our (upper and lower) bounds are not necessarily close.

§2.4. The One-Stage Procedure $\underset{\sim}{\phi}$.

Consider the procedure $\underset{\sim}{\phi}$ (0.6) originally considered by Fox (1968) in the empirical Bayes problem. It is a corollary to Theorem 2 of Chapter III that with bounded Ω and the choice $h = n^{-\frac{1}{4}}$, $|D(\underset{\sim}{\theta}, \underset{\sim}{\phi})| = O(n^{-\frac{1}{4}})$ uniformly in parameter sequences, the same rate established for $\underset{\sim}{\theta}_T$ in Theorem 1.

In this section we study the risk behavior of $\underset{\sim}{\phi}$ at $\underset{\sim}{0}$. For the modified regret of $\underset{\sim}{\phi}$ we find the same lower bound as established in Theorem 2(ii) for $\underset{\sim}{\theta}_T$ and an upper bound which is $O(n^{-\frac{1}{2}})$ for the choice $h = n^{-\frac{1}{4}}$. As in sections 2 and 3 let $X_1, \ldots, X_{n+1}$ be iid $P = \mathcal{U}[0,1)$ and replace n by $n+1$ in the definition (0.6) of $\underset{\sim}{\phi}$. Fix $x \doteq X_{n+1}$ and let ϕ and ψ abbreviate $\phi_{n+1,n+1}$ and $\psi_{n+1,n+1}$ in (0.6) and (0.7). As in section 2 define

$$u = \Sigma_{j=1}^{n} [0 \leq X_j \leq x] \quad \text{and} \quad v = \Sigma_{j=1}^{n} [x < X_j \leq x+h] .$$

Then, by (0.7) applying the definitions of F^* and T_{n+1} we can show as we have done for φ in (2.2) that ψ has the following explicit form; for $x \in [0,1)$,

$$(4.1) \quad \psi = v^{-1}\{xv - hu - h + \Sigma_{j=1}^{n} [0 \leq X_j \leq x'+h]\} \quad \text{a.e. } \underset{\sim}{P}_x \quad .$$

From Lemma 5 in section 2 and the above (1), we can easily see that

$$(4.2) \qquad 0 \leq \psi - \varphi \leq h \ .$$

In the same manner as (2.1) was obtained,

$$D(\underset{\sim}{0}, \underset{\sim}{\phi}) = \underset{\sim}{P}\, \phi^2 . \tag{4.3}$$

Note that $\phi = x' \vee \psi$ for $x \in [0, 1-h)$; $= (x' \vee \psi) \wedge x$, for $x \in [1-h, 1)$.

We get a lower bound first. Let

$$J = [\psi \geq x', x < 1-h] .$$

By (1) and the definition of ϕ,

$$\underset{\sim}{P}\, \phi^2 \geq \underset{\sim}{P}(\psi^2 J) . \tag{4.4}$$

Define $Y = [nx(1-x)]^{-\frac{1}{2}}(u-nx)$ and $Z = (nh)^{-\frac{1}{2}}(v-nh)$. Then, we can easily see that

$$\psi = \frac{(nh)^{-\frac{1}{2}}xZ - n^{-1}}{1 + (nh)^{-\frac{1}{2}}Z} - \frac{\sqrt{x(1-x)}\; n^{-\frac{1}{2}}Y}{1+(nh)^{-\frac{1}{2}}Z} . \tag{4.5}$$

<u>Theorem 4</u>. If h is a function of n such that $nh \to \infty$ and $h \to 0$, then for any $\frac{1}{3} > \epsilon > 0$, there exists $N < +\infty$ so that for all $n \geq N$,

$$D(\underset{\sim}{0}, \underset{\sim}{\phi}) > (\frac{1}{3} - \epsilon)\, \frac{1}{nh} .$$

Proof. Fix $x \in (0,1)$ until (7). Since by (2) $\varphi \le \psi$, it follows by Lemma 8 in section 2 that

$$(4.6) \quad P_{\underset{\sim}{x}}[\psi \le x'] \le P_{\underset{\sim}{x}}[\varphi \le x'] \to 0, \quad \text{for given } x.$$

Since by Lemma 6 in section 2, $(Y,Z) \overset{\mathcal{D}}{\to} N(\underline{0},I)$ where $\underline{0}$ is 2 dimensional zero vector and I, 2×2 identity matrix, and since by (6) $J \overset{P}{\to} 1$, it follows from Slutsky's theorem applied to rhs(5) that if $x \in (0,1)$, then

$$\sqrt{nh}\ \psi\ J \overset{\mathcal{D}}{\to} N(0,x^2)\ .$$

As a consequence of a convergence theorem (2.7) (cf. Loéve (1963) 11.4 A(i)) we have

$$(4.7) \qquad \underline{\lim}(nh)P_{\underset{\sim}{x}}(\psi^2 J) \ge x^2[0 < x < 1]\ .$$

Thus, by Fatou's Lemma applied to the lhs below

$$\underline{\lim}\ P\ P_{\underset{\sim}{x}}(nh\ \psi^2 J) \ge P(\text{lhs}(7)) \ge \int_0^1 y^2 dy = 3^{-1}\ .$$

Therefore, in view of (3) and (4),

$$\underline{\lim}\ (nh)D(\underset{\sim}{0}, \underset{\sim}{\phi}) \ge 3^{-1}$$

and the definition of lim inf leads to the conclusion. ∎

We shall now find an upper bound. In the same manner as (3.2) was obtained

$$(4.8) \qquad \underset{\sim}{P}\, \phi^2 \le \underset{\sim}{P}[v = 0] + \underset{\sim}{P}(\phi^2[v > 0]) .$$

In view of (3.4), we only consider the last term. As in (3.5),

$$(4.9) \qquad \underset{\sim}{P}(\phi^2[v > 0]) \le P\{\int_0^1 \underset{\sim}{P}_x[U > h]ds^2 + \int_0^1 \underset{\sim}{P}_x[V < h]ds^2\}$$

where $U = \Sigma_j U_j$, $V = \Sigma_j V_j$ and for each $j = 1,2,\ldots,n$

$$(4.10) \qquad U_j = (x-s)[x < X_j \le x+h]-h[0 \le X_j \le x] + [0 \le X_j \le x'+h]$$

and

$$V_j = U_j + 2s[x < X_j \le x+h] .$$

The V_j are i.i.d. with

$$(4.11) \qquad \underset{\sim}{P}_x V_j = \begin{cases} sh & , \text{ for } x \in [0, 1-h) \\ (1-x)(s-1+x+h) & , \text{ for } x \in [1-h, 1) . \end{cases}$$

Note that each U_j has mean $-sh$ for $x \in [0,1-h)$ and mean $(1-x)(-s-1+x+h)$ for $x \in [1-h,1)$. We will use the bound 3 for the range of V_j and U_j for all $0 \le s \le 1$ and $0 \le x < 1$ in Theorem 2 of Hoeffding (1963) in order to bound two tail probabilities in (9). As in §3, for $0 \le x < 1-h$ the bound developed for the first term in the curly brackets of rhs(9) will also bound the last term in the same brackets because $\underset{\sim}{P}_x V$ is closer to h than $\underset{\sim}{P}_x U$ is to h. But, for $x \in [1-h,1)$ such ordering varies according as the values of x and hence we require more treatment for this case. Using these facts we prove

<u>Lemma 11</u>.

(4.12) $$\underset{\sim}{P}(\phi^2[v > 0]) = O((nh^2)^{-1} + h^3 + hn^{-\frac{1}{2}}\log n + n^{-\frac{1}{2}}) .$$

<u>Proof</u>. Let $\bar{V} = n^{-1}V$. Using (11) and letting $a_1 = n^{-1}$ and $a_2 = x+h-1-n^{-1}h(1-x)^{-1}$, we get

(4.13) $$\underset{\sim x}{P}[V < h] = \begin{cases} \underset{\sim x}{P}[-\bar{V} + \underset{\sim x}{P}\bar{V} > h(s-a_1)], & \text{for } x \in [0,1-h) \\ \underset{\sim x}{P}[-\bar{V} + \underset{\sim x}{P}\bar{V} > (1-x)(s+a_2)], & \text{for } x \in [1-h,1) . \end{cases}$$

Fix $x \in [0,1-h)$ until (15). We shall find the bound for the second term in the curly brackets of rhs(9) and double it to bound the curly brackets of rhs(9).

By Theorem 2 of Hoeffding (1963) with the bound 3 for the range of the U_j and V_j,

(4.14) $$\underset{\sim x}{P}[U > h] + \underset{\sim x}{P}[V < h] \leq 2\exp\{-\tfrac{1}{2}\, b_1((s-a_1)_+)^2\}$$

where $b_1 = 4nh^2/9$.

Applying (14) and (3.11) with $a = a_1$ and $b = b_1$ gives

(4.15) $$P([0 \le x < 1-h]\int_0^1 \text{lhs}(14)\,ds^2) = O((nh^2)^{-1}) .$$

Fix $x \in [1-h,1)$ until (19) and let $\bar{U} = n^{-1}U$. By the statement after (11),

(4.16) $$\underset{\sim}{P}_x[U > h] \le \underset{\sim}{P}_x[\bar{U} - \underset{\sim}{P}_x\bar{U} > (1-x)(s-a_2)] .$$

Applying Hoeffding's bound (as (14) is obtained) to rhs (13) for $x \in [1-h,1)$ and rhs(16) and weakening the bound as below, shows

(4.17) $$\text{lhs}(14) \le 2 \exp\{-\tfrac{1}{2} b_2((s-a_2)_+ \wedge (s+a_2)_+)^2\}$$

where $b_2 = 4n(1-x)^2/9$.

Notice that $a_2 > 0$ iff $(1-h <)\delta_1 < x < \delta_2$ (< 1) where

$$\delta_1 = 1-2^{-1}(h + \sqrt{h^2-4n^{-1}h}) \quad \text{and} \quad \delta_2 = 1-2^{-1}(h - \sqrt{h^2-4n^{-1}h}) .$$

Hence,

(4.18) $$2^{-1}\text{rhs}(17) = \begin{cases} \exp\{-\tfrac{1}{2} b_2((s+a_2)_+)^2\}, & \text{for } x \in (\delta_1,\delta_2)^c \cap [1-h,1) \\ \exp\{-\tfrac{1}{2} b_2((s-a_2)_+)^2\}, & \text{for } x \in (\delta_1,\delta_2) . \end{cases}$$

Recognizing lhs(3.11) ≤ 1 and applying (18) and (3.11) with $b = b_2$ and $a = -a_2$ for $x \in (\delta_1,\delta_2)^c \cap [1-h,1)$; $= a_2$ for $x \notin (\delta_1,\delta_2)$, and weakening the bounds results in

$$(4.19)\qquad \int_0^1 (18)\, ds^2 \le \begin{cases} \dfrac{19+3\sqrt{2\pi}}{4n(1-x)^2} \wedge 1\ , & \text{for } x \in (\delta_1,\delta_2)^c \cap [1-h,1) \\ (x+h-1)^2+\left(\dfrac{9}{2n(1-x)^2}\right)\wedge 1+c_6 h(\sqrt{n}(1-x))^{-1}, & \text{for } x \in (\delta_1,\delta_2) \end{cases}$$

where c_6 is some positive constant.

Simple computation gives us

$$\int_{1-h}^1 (y+h-1)^2 dy = h^3/3,$$

$$\int_{\delta_1}^{\delta_2} (1-y)^{-1} dy = \log((1-\delta_1)/(1-\delta_2)) \le \log n \quad .$$

Also, by (3.14)

$$\int_{1-h}^1 \frac{d}{n(1-y)^2} \wedge 1\ dy = O(n^{-\frac{1}{2}}) \ .$$

Hence, we can easily check

(4.20) $P([1-h \le x < 1]\int_0^1 (18)\, ds^2) = O(h^3 + hn^{-\frac{1}{2}}\log n + n^{-\frac{1}{2}})$.

Thus, in view of (20), (18) and (17), $P([1-h \le x < 1]\int_0^1 \text{lhs}(14)\, ds^2)$ equals rhs(20).

Therefore, (12) follows from this, (15), (14) and (9). ▌

Applying Lemma 11 and (3.4) to (8) we obtain, in view of (8) and (3), the following upper bound of $D(\underset{\sim}{0}, \underset{\sim}{\phi})$.

Theorem 5.

$$D(\underset{\sim}{0},\underset{\sim}{\phi}) = O((nh^2)^{-1} + h^3 + hn^{-\frac{1}{2}}\log n + n^{-\frac{1}{2}}) .$$

Remark . As in the remark of section 3, we shall state values of h and of bounds of the modified regrets, up to constants. We let large n be fixed.

The lower bound of $D(\underset{\sim}{0}, \underset{\sim}{\phi})$ in Theorem 4 describes the hyperbola $(nh)^{-1}$ which coincides with (up to constants) that of the lower bound $D(\underset{\sim}{0}, \underset{\sim}{\theta}_T)$ in Theorem 2, for h less than $n^{-1/3}$, and then decreases to n^{-1} as h increases.

On the other hand, the upper bound of $D(\underset{\sim}{0}, \underset{\sim}{\phi})$ in Theorem 5 describes the strictly convex curve which attains the minimum value $n^{-\frac{1}{2}}$ for $n^{-\frac{1}{4}} \le h \le n^{-1/6}$, has the value $(nh^2)^{-1}$ for h less than $n^{-\frac{1}{4}}$ and h^3 for h greater than $n^{-1/6}$.

Thus, from the remark at the end of section 3, we can easily see that for h greater than $n^{-\frac{1}{4}}$, $D(\underset{\sim}{0}, \underset{\sim}{\phi})$ must be strictly below $D(\underset{\sim}{0}, \underset{\sim}{\theta}_T)$. Hence, for such h $\underset{\sim}{\phi}$ is strictly better than $\underset{\sim}{\theta}_T$.

§2.5. A Counterexample to $D(\underset{\sim}{\theta}, \underset{\sim}{t}) \to 0$ on R^{∞}.

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

Theorem 6. Let $X_1, X_2, \ldots$ be independent random variables where for each j, $X_j \sim U[\theta_j, \theta_j+1)$, $\theta_j \in \Omega = R$. Let $\underset{\sim}{t}(\underset{\sim}{X}) = (t_1(\underset{\sim}{X}), \ldots, t_n(\underset{\sim}{X}))$ be an estimator of $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_n)$, $n = 1,2,\ldots$ Then there exists a sequence $(\theta_1, \theta_2, \ldots) \in R^{\infty}$ such that $\overline{\lim}_n \, D(\underset{\sim}{\theta}, \underset{\sim}{t}) > 0$.

Proof. Since for each j, $\underset{\sim}{P}(t_j(\underset{\sim}{X}) - \theta_j)^2 \geq P_j(P_{\underset{\sim}{x}}(t_j(\underset{\sim}{X})) - \theta_j)^2$, it follows that

$$(5.1) \qquad D(\underset{\sim}{\theta}, \underset{\sim}{t}) \geq n^{-1} \Sigma_{j=1}^{n} P_j(P_{\underset{\sim}{x}}(t_j(\underset{\sim}{X})) - \theta_j)^2 - R(G) \, .$$

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

$$(5.2) \qquad \underset{\sim}{\mu}\{n^{-1} \Sigma_{j=1}^{n} P_j(P_{\underset{\sim}{x}}(t_j(\underset{\sim}{X})) - \theta_j)^2\} \geq n^{-1} \Sigma_{j=1}^{n} \mu_j P_j(s_j - \theta_j)^2 \, .$$

Now consider $\underset{\sim}{\mu} = \mu_1 \times \mu_2 \times \ldots$ where μ_j puts mass $\frac{1}{2}$ on each of the values $2j \pm r$, $j \geq 1$, where r is some fixed number such that $0 < r < \frac{1}{2}$. Then

$$\mu_j P_j(s_j-\theta_j)^2 = \tfrac{1}{2} P_{2j-r}(s_j-(2j-r))^2 + \tfrac{1}{2} P_{2j+r}(s_j-(2j+r))^2$$

(5.3)

$$\geq \int_{2j+r}^{2j+1-r} \{\tfrac{1}{2}(s_j-(2j-r))^2 + \tfrac{1}{2}(s_j-(2j+r))^2\}dx \geq r^2(1-2r) ,$$

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

Since $R(G) = n^{-1} \Sigma_{j=1}^n P_j(\theta_{jn} - \theta_j)^2$ where θ_{jn} is defined by the posterior mean (1.0.2) with $q = 1$, and since the θ_j's are apart from each other more than 1, $\theta_{jn} = \theta_j$ for all j and hence $R(G) = 0$. Thus, $\underset{\sim}{\mu}(R(G)) = 0$. Therefore, in view of (1), (2) and (3),

$$(5.4) \qquad \underset{\sim}{\mu}\{D(\underset{\sim}{\theta}, \underset{\sim}{t})\} \geq r^2(1-2r)$$

for all n. The retraction $\underset{\sim}{t}^*$ of $\underset{\sim}{t}$ 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

$$(5.5) \qquad \underset{\sim}{\mu}\{\overline{\lim}_n D(\underset{\sim}{\theta}, \underset{\sim}{t}^*)\} \geq \overline{\lim}_n\{\underset{\sim}{\mu}\, D(\underset{\sim}{\theta}, \underset{\sim}{t}^*)\} \geq r^2(1-2r) > 0 .$$

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

CHAPTER III

RATES FOR ONE-STAGE PROCEDURES IN THE k-EXTENDED PROBLEM

§3.0. Introduction

Let $X_1,\ldots,X_n$ be independent random variables each X_j having the distribution $P_j \in \mathcal{P}(f)$ (see (1.0.1) for the definition of $\mathcal{P}(f)$) where f satisfies the assumptions (stronger here in that f is now assumed bounded away from zero) that for given finite positive constants m (> 0) and M (≥ 0),

$$m^{-1} \leq f \leq 1 \tag{0.1}$$

and

$$\vee\{(v-u)^{-1}|(f(v))^{-1}-(f(w)^{-1}| \; : u < v\} \leq M \; . \tag{0.2}$$

Throughout the chapter, we assume $\Omega = [c,d]$ with $-\infty < c \leq d < \infty$.

For each $j = k,k+1,\ldots,n$, let $\underset{\sim}{z}_j^k = (z_{j-k+1},\ldots,z_j)$, $x \doteq X_j$, $y \doteq \underset{\sim}{X}_{j-1}^{k-1}$ and $\underline{x} = (y,x)$.

Throughout this chapter, G is interpreted as the empirical distribution of the $(n-k+1)$ k-tuples $\underset{\sim}{\theta}_k^k, \underset{\sim}{\theta}_{k+1}^k,\ldots,\underset{\sim}{\theta}_n^k$ of parameters. Let $\underset{\sim}{\theta}_G$ be the k-extended procedure with $\underset{\sim}{\theta}_G(\underset{\sim}{X}) \doteq (\theta_{kn},\ldots,\theta_{nn})$ where θ_{jn} is the posterior mean of θ_j given $\underline{x}$ wrt the prior

G on $\underset{\sim}{\theta}_j^k$, namely,

$$(0.3) \qquad \theta_{jn} = \int_{\underline{x}'+}^{x} \Theta_j \, q(\underset{\sim}{\Theta}_j^k)\, dG(\underset{\sim}{\Theta}_j^k) \Big/ \int_{\underline{x}'+}^{x} q(\underset{\sim}{\Theta}_j^k)\, dG(\underset{\sim}{\Theta}_j^k)$$

where the affix + is intended to describe the integration as over $(\underline{x}', x]$. The modified regret for any procedure $\underset{\sim}{t}$ relative to the k-extended envelope is given by

$$D^k(\underset{\sim}{\theta}, \underset{\sim}{t}) = \text{Av.} \{ \underset{\sim}{P}(t_j(\underset{\sim}{X}) - \theta_j)^2 - \underset{\sim}{P}(\theta_{jn} - \theta_j)^2 \}$$

where Av. means the average over $j = k, \ldots, n$. Since $X_j' < \theta_{jn} \le X_j$, when $X_j' < t_j(\underset{\sim}{X}) \le X_j$, we have

$$(0.4) \qquad 2^{-1} |D^k(\underset{\sim}{\theta}, \underset{\sim}{t})| \le \text{Av.} \underset{\sim}{P} |t_j(\underset{\sim}{X}) - \theta_{jn}| \, .$$

We here introduce two one-stage procedures which are respective generalizations of unextended (k = 1) procedures, $\underset{\sim}{\theta}_T$ treated for $\mathcal{P}(1)$ and $\underset{\sim}{\phi}$ treated for the uniform distribution on the unit interval [0,1) in Chapter II. We exhibit in Section 1 the k-extended version of $\underset{\sim}{\theta}_T$ for $\mathcal{P}(f)$ and in Section 2 that of $\underset{\sim}{\phi}$ for $\mathcal{P}(1)$ with the same rate $(2k + 2)^{-1}$.

§3.1 A Procedure $\underset{\sim}{\theta}_T$ with a Rate $(2k+2)^{-1}$.

Hereafter throughout this Chapter, we interpret $\underset{\sim}{\theta}_T$ as a k-extended version of the procedure treated in Sections 1, 2 and 3 of Chapter II.

We first derive $\underset{\sim}{\theta}_T$ in analogous way used in §2.0. We then bound the modified regret of $\underset{\sim}{\theta}_T$ using Lemmas 1 through 5 and Proposition 1 (analogous respectively to Lemmas I.7, I.1, I.3, II.1 and II.2 and Proposition I.1).

Let $\underline{u} = (v,u) \in R^{k-1} \times R$ and similarly $\underline{\theta} = (\omega,\theta) \in R^{k-1} \times R$. Let $f(v) = \Pi_{i=1}^{k-1} f(v_i)$, $f(\underline{u}) = f(v)f(u)$ and $q(\omega) = \Pi_{i=1}^{k-1} q(\omega_i)$, $q(\underline{\theta}) = q(\omega)q(\theta)$.

Let Q be the measure with density $q(\underline{u})$ at $\underline{u}$ wrt G. Note that, since (0.1) implies $q \le m$,

$$(1.1) \qquad q(\underline{u}) \le m^k \quad \text{for all} \quad \underline{u} \in R^k .$$

For fixed j, we abbreviate $\underset{\sim}{X}_j^k$ to $\underline{x} = (y,x)$ where y is the first $k-1$ coordinates and $x \doteq X_j$. In view of (0.3),

$$(1.2) \qquad \theta_{jn} = \int_{\underline{\theta}=\underline{x}'}^{\underline{x}} \theta \, dQ(\underline{\theta}) / \int_{\underline{\theta}=\underline{x}'}^{\underline{x}} dQ(\underline{\theta}) \quad \forall\, j = k, k+1, \ldots, n .$$

The following lemma generalizes Lemma I.7 with $h \equiv 1$.

Lemma 1. Let τ be a signed measure and $I = (\underline{u}', \underline{u}]$ be a cube with $\tau(I) \neq 0$. Let $\tau_{\underline{u}}$ be the signed measure with density $I/\tau(I)$ wrt τ. Then,

$$\int s_k d\tau_{\underline{u}}(\underline{s}) = u_k - \int_0^1 \tau_{\underline{u}}[s_k \le u_k' + t]dt \ .$$

Proof. By the Fubini theorem applied to the lhs of the second equality below,

$$\int (u_k - s_k) d\tau_{\underline{u}}(\underline{s}) = \int \int_{s_k - u_k'}^1 dt \ d\tau_{\underline{u}}(\underline{s}) = \int_0^1 \tau_{\underline{u}}[s_k \le u_k' + t]dt \ . \blacksquare$$

Applying Lemma 1 with τ the measure with density $(Q(\underline{x}',\underline{x}])^{-1}$ wrt Q, gives us that

(1.3) $$\theta_{jn} = x - \eta_{jn}$$

where

(1.4) $$\eta_{jn} = \int_0^1 Q((y',y] \times (x',x'+t])dt/Q(\underline{x}',\underline{x}] \ .$$

For every $\underline{u} \in R^k$, let

(1.5) $$\bar{p}(\underline{u}) = \int p_{\underline{\theta}}(\underline{u}) dG(\underline{\theta}) \ .$$

where

$$p_{\underline{\theta}}(\underline{u}) = p_{\omega}(v)p_{\theta}(u) = \prod_{i=1}^{k-1} p_{\omega_i}(v_i)p_{\theta}(u) .$$

Then, by the form of densities (1.0.1),

$$(1.6) \qquad \bar{p}(\underline{u}) = f(\underline{u})Q(\underline{u}',\underline{u}] .$$

Hence, $\forall\ (v,u) \in R^{k-1} \times R$,

$$(1.7) \qquad Q((v',v] \times (-\infty,u]) = \Sigma \frac{\bar{p}(v,u-r)}{f(v)f(u-r)} ,$$

where Σ abbreviates summation wrt the non-negative integer r and (also in (10) below) involves at most $\theta_{(n)} - \theta_{(1)} + 2$ terms (when $\Omega = [c,d]$, at most $d-c+2$). (7) is a generalized form of (2.0.2). Letting

$$(1.8) \qquad F^*(\underline{u}) = \text{Av.}\ [\underset{\sim}{X}_j^k \leq \underline{u}]$$

where throughout this chapter Av. means average over $j = k,\ldots,n$ and, for any $0 < h < 1$,

$$(1.9) \qquad \Delta F^*(\underline{u}) = h^{-k}F^*(\underline{u},\underline{u}+h\underline{1}],$$

we estimate $\bar{p}(\underline{u})$ by $\Delta F^{*}(\underline{u})$. Thus, in view of (7), we estimate $Q((v',v] \times (-\infty,u])$ by (a generalization of (2.0.3))

$$(1.10) \qquad T(\underline{u}) = \Sigma \frac{\Delta F^{*}(v,u-r)}{f(v)f(u-r)}$$

and $Q(\underline{u}',\underline{u}]$ by

$$(1.11) \qquad T(v,\cdot)]_{u'}^{u} = \Delta F^{*}(\underline{u})/f(\underline{u}) \ .$$

Since $0 \le \eta_{jn} \le 1$, we finally estimate θ_{jn} by

$$(1.12) \qquad \theta_{T,j} = x - (0 \vee \varphi_{jn}) \wedge 1$$

where

$$(1.13) \qquad \varphi_{jn} = \int_0^1 T(y,\cdot)]_{x'}^{x'+t} dt / T(y,\cdot)]_{x'}^{x}$$

and $0/0$ is taken to be zero.

To get an upper bound of the modified regret for $\underset{\sim}{\theta}_T$ we use an analogous method to that used in Theorem 1 in Chapter II. The following lemma is a generalization of Lemma I.1 with $\delta = 0 = \varepsilon$. Let $\underset{\sim}{P}_j^k = P_{j-k+1} \times \ldots \times P_j \quad \forall\, j = k,\ldots,n$.

Lemma 2. Let $I_{\underline{u}} = (\underline{u}', \underline{u}]$. Then

$$(1.14) \qquad \text{Av. } P_{\underset{\sim}{j}}^{k}\{I_{\underline{x}}(\underset{\sim}{\theta}_{j}^{k})/Q(I_{\underline{x}})\} \le (d+2-c)^{k} .$$

Proof. Since by the definition of Q,

$$(1.15) \qquad \text{Av. } q(\underset{\sim}{\theta}_{j}^{k}) I_{\underline{u}} = Q(I_{\underline{u}}) ,$$

it follows by k usages of $f \le 1$ that

$$\text{lhs}(14) = \int \{\underline{u} : Q(I_{\underline{u}}) > 0\} f(\underline{u}) d\underline{u} \le (d+2-c)^{k} .$$

The next lemma is a generalized analogue of Lemma I.3 with $\eta = 0$.

Lemma 3. For $\underline{s} \le \underline{t} \in R^{k}$,

$$(1.16) \qquad \text{Av. } P_{\underset{\sim}{j}}^{k}(G(\underline{x}-\underline{t}, \underline{x}-\underline{s}]/Q(\underline{x}', \underline{x}]) \le \prod_{i=1}^{k} (t_i - s_i) .$$

Proof. In view of (15) with $\epsilon = 0$ and $\underline{u} = \underline{w}$,

$$\text{lhs}(16) = \int [Q(\underline{w}', \underline{w}] > 0] G(\underline{w}-\underline{t}, \underline{w}-\underline{s}] f(\underline{w}) d\underline{w}$$

$$\le \int G(\underline{w}-\underline{t}, \underline{w}-\underline{s}] d\underline{w} .$$

By the Fubini Theorem, the rhs equals

$$\int \int_{\underline{\theta}+\underline{s}}^{\underline{\theta}+\underline{t}} d\underline{w}\; dG(\underline{\theta}) = \text{rhs}(16) \;. \blacksquare$$

By the definitions ((3) and (12), respectively) of θ_{jn} and $\theta_{T,j}$, and by the fact that $0 \le \eta_{jn} \le 1$,

$$(1.17) \qquad |\theta_{T,j} - \theta_{jn}| \le |\eta_{jn} - \varphi_{jn}| \wedge 1 \;.$$

But, by Lemma I.5 and by another usage of $0 \le \eta_{jn} \le 1$, it follows that with $\check{\underset{\sim}{P}}_{j,k} = P_1 \times \ldots \times P_{j-k} \times P_{j+1} \times \ldots \times P_n$

$$(1.18) \quad \check{\underset{\sim}{P}}_{j,k}(|\eta_{jn} - \varphi_{jn}| \wedge 1) \le 2(Q(\underline{x}',\underline{x}])^{-1}\{\check{\underset{\sim}{P}}_{j,k}|\int_0^1 T(y,\cdot)]_{x'}^{x'+t}dt - \int_0^1 Q((y',y] \times (x',x'+t])dt| + 2\check{\underset{\sim}{P}}_{j,k}|T(y,\cdot)]_{x'}^{x} - (Q(\underline{x},\underline{x}])^{-1}|\} \;.$$

The following lemma is a modified analogue to Lemma II.1. It is used to prove the forthcoming Lemma 5 which is a generalized analogue to Lemma II.2.

<u>Lemma 4</u>. For $\underline{u} = (v,u) \in R^{k-1} \times R$ and each j,

$$(1.19)\qquad h^{k}\check{P}_{\underset{\sim}{j},k}(T(\underline{u})) = \Sigma \frac{1}{f(v)f(u-r)} \int_{t=v}^{v+h1} \int_{s=u-v}^{u-r+h} \check{Q}_j(\underline{s})f(\underline{s})d\underline{s}$$

$$+ (n-k)^{-1}\, \Sigma\{[v < y < v+h1,\ u-r < x \le u-r+h]/f(v)f(u-r)\},$$

where $\check{Q}_j(B) \doteq Q(B) - (n-k)^{-1}[(y,x) \in B]q(\underset{\sim}{\theta}_j^k)$, $\forall\, B \in \mathcal{B}^k$.

Proof. By the definition (10) of T,

$$(1.20)\quad T(\underline{u}) = h^{-k}\, \mathrm{Av}.\{\Sigma[v < \underset{\sim}{X}_{j-1}^{k-1} \le v+h1,\ u-r < X_j \le u-r+h]/(f(v)f(u-r))\}.$$

The $i=j$th term gives h^{-2} (second term of rhs (19)).

Now, since for fixed r, taking the average operation inside of the integrals gives us that

$$(n-k+1)^{-1}\Sigma_{i=k,\ne j}^{n} \int_{\underset{\sim}{\theta}_i^k}^{\underset{\sim}{\theta}_i^k+\underline{1}} [v < t \le v+h1,\ u-r < s \le u-r+h]q(\underset{\sim}{\theta}_i^k)f(\underline{s})d\underline{s}$$

$$= \int_{t=v}^{v+h1} \int_{s=u-r}^{u-r+h} \check{Q}_j(\underline{s}',\underline{s}]f(\underline{s})d\underline{s}\ .$$

Multiplying by $(f(v)f(u-r))^{-1}$ and summing over r leads to the first term of rhs(19). ∎

Lemma 5. For $0 \le t \le 1$,

$$Q((y'+h1,y] \times (x'+h,x'+t])[t \geq h]-\alpha$$

$$\leq \check{P}_{\underset{\sim}{j},k} T(y,\cdot)]_{x'}^{x'+t} \leq Q((y',y+h1] \times (x',x'+t+h]) + \alpha$$

where $\alpha = O(h+n^{-1})$.

Proof. We obtain by two applications of Lemma 3 (note that the second term of rhs(19) with $s=x$ is zero) and a change of variable u' to u in the second inner integral below,

$$(1.21) \quad h^k \check{P}_{\underset{\sim}{j},k} T(y,\cdot)]_{x'}^{x'+t}$$

$$= \int_y^{y+h1} \frac{f(v)}{f(y)} \Sigma\{[\int_{\cdot-r}^{\cdot-r+h} \frac{f(u)}{f(\cdot-r)} \check{Q}_j((v',v]\times(-\infty,u])\,du]_{x'}^{x'+t}$$

$$- [\int_{\cdot-r-1}^{\cdot-r-1+h} \frac{f(u+1)}{f(\cdot-r)} \check{Q}_j((v',v]\times(-\infty,u])\,du]_{x'}^{x'+t}\}dv .$$

Now, by the assumption (0.2),

$$(1.22) \qquad \left|\frac{f(w)}{f(z)} - 1\right| \leq M|w - z| .$$

Using (22) to bound the four f-ratios in (21) and using (1), we obtain that

$$(1.23)\quad \operatorname{rhs}(21) \underset{(\geq)}{\leq} \int_y^{y+h1} \frac{f(v)}{f(y)} \Sigma\{\int_.^{.+h} \check{Q}_j((v',v]\times(-\infty,u])\,du]_{x'+t-r-1}^{x'+t-r}$$

$$- \int_.^{.+h} \check{Q}((v',v]\times(-\infty,u])\,du]_{x'-r-1}^{x'-r} \underset{(-)}{+} 4m^k Mh^2\}dv \;.$$

By our convention on Σ, and two telescopic series,

$$(1.24)\quad \operatorname{rhs}(23) \underset{(\geq)}{\leq} A \underset{(-)}{+} 4m^k Mh^2(d+2-c)(\int_y^{y+h1} \frac{f(v)}{f(y)}\,dv)$$

where

$$A = \int_y^{y+h1} \frac{f(v)}{f(y)} [\int_.^{.+h} \check{Q}_j((v',v]\times(-\infty,u])\,du]_{x'}^{x'+t}\,dv \;.$$

Making a change of variable $w = u-t$ in the positive term of A, using (22) to bound the f-ratios as needed for (21) and applying (1) gives us

$$A \leq \int_{v=y}^{y+h1} \int_{u=x'}^{x'+h} \check{Q}_j((v',v]\times(w,w+t])\,dwdv \underset{(-)}{+} O(h^{k+1}) \;.$$

By this and by the upper bound of (22) together with the assumption $h < 1$, applied to the integral of the second term of rhs(24),

$$\text{rhs}(24) \underset{(\geq)}{\leq} \int_{y}^{y+hl} \int_{x'}^{x'+h} \check{Q}_j((v',v]\times(w,w+t])\,dw\,dv \underset{(-)}{+} h^k\alpha . \tag{1.25}$$

Weakening the bounds and applying $Q-(n-1)^{-1} \leq \check{Q}_j \leq Q$ gives, in view of (25), (24), (23) and (21), the bounds of asserted Lemma 5. ∎

We finally introduce the following proposition which generalizes Proposition II.1.

Proposition 1. Let $Y_1,\ldots,Y_N$ be (k-1)-dependent random variables and $a \leq Y_i \leq b$ for all $i = 1,2,\ldots,N$. Then, for any η and every N,

$$E|\bar{Y} - \eta| \leq |E\bar{Y} - \eta| + \frac{b-a}{\sqrt{M}}\sqrt{\pi/2} , \tag{1.26}$$

where M is the greatest integer $\leq N/k$.

Proof. We prove in the same way as Proposition II.1. The extension of Theorem 2 of Hoeffding (1963, Section 5d) to (k-1)-dependent random variables gives

$$P[|\bar{Y} - E\bar{Y}| > s]ds \leq 2\exp\{-\frac{2s^2M}{(b-a)^2}\} .$$

Hence

$$E|\bar{Y} - E\bar{Y}| \leq 2\int_0^\infty \exp\{-2s^2 M(b-a)^{-2}\}ds = (b-a)\sqrt{\pi/(2M)} .$$

Thus, the triangle inequality leads to the asserted bound. ∎

We shall now go back to (18). To get an upper bound of the average expectation wrt $\underset{\sim}{P}_j^k$ of lhs(18), we apply the following Lemmas 6 and 7. These are replacements of Lemmas II.3 and II.4 combined with multiplication by $(Q(\underline{x}',\underline{x}])^{-1}$ and the average expectation. Let $\bar{K} \doteq (n-k)^{-1}\Sigma_{j\neq i=k}^{n} K_i$ for any random variables K_i .

<u>Lemma</u> <u>6</u>.

$$(1.27) \quad \text{Av.}\ \underset{\sim}{P}_j^k\{(Q(\underline{x}',\underline{x}])^{-1}(\int_0^1|\check{\underset{\sim}{P}}_{j,k}(T(y,\cdot]_{x'}^{x'+t}) - Q((y',y]\times(x',x'+t])|dt)\}$$

$$= O(h + n^{-1} + (\sqrt{n}\ h^k)^{-1}).$$

<u>Proof</u>. For notational simplicity we prove only for $k = 2$. Fix j until (32) and $t \in [0,1]$ until (30). Define $W_i = h^{-2}(f(y))^{-1}[y < X_{i-1} \leq y+h]\Sigma([x'+t-r < X_i \leq x'+t-r+h]/f(x'+t-r) -[x'-r < X_i \leq x'-r+h]/f(x'-r))$ for $i = 2,3,\ldots,n$. In view of (20), we can directly verify that $T(y,\cdot)]_{x'}^{x'+t} = \bar{W}$.

Since W_i is a function of X_{i-1} and X_i, the sequence of random variables $W_2,\ldots,W_n$ are 1-dependent. Also $|W_i| \leq m^2h^{-2}$ for all i. Thus, by Proposition 1, with $b-a = 2m^2h^{-2}$,

$$(1.28)\quad \check{P}_{\sim j,2}|T(y,\cdot)]_{x'}^{x'+t} - Q((y',y]\times(x',x'+t])|$$

$$\leq |\check{P}_{\sim j,2}(T(y,\cdot)]_{x'}^{x'+t}) - Q((y',y]\times(x',x'+t])| + \frac{2m^2}{\sqrt{\lambda}\,h^2}\sqrt{\frac{\pi}{2}}$$

where λ is the least integer greater than $\frac{1}{2}n-1$.

Let us denote the first term of rhs(28) by $|P-Q|$ where P and Q are defined by positional correspondence. Using $|P-Q| = (Q-P)_+ + (P-Q)_+$, applying the lower and upper bounds of Lemma 5 to P in $(Q-P)_+$ and $(P-Q)_+$, respectively, and performing simple computations, we bound $|P-Q|$ by

$$Q((y',y]\times(x',x'+t])[t < h] + \{Q((y',y'+h]\times(x',x'+t])$$

$$(1.29)\quad + Q((y'+h,y]\times(x',x'+h])\}[t \geq h] + Q((y',y]\times(x'+t,x'+t+h])$$

$$+ Q((y,y+h]\times(x',x'+t+h]) + O(h+n^{-1}) .$$

By (1) with $k = 2$, $Q((y',y]\times(x'+t,x'+t+h]) \leq m^2 \times G((y',y]\times(x'+t,x'+t+h])$. Also, from the Fubini Theorem

$$\int_0^1 G((y',y]\times(-\infty,\cdot])]_{x'+t}^{x'+t+h} dt = \int_{x'}^{x+h}\int_{(s-x'-h)\vee 0}^{(s-x')\wedge 1} dt\, d_s G((y',y]\times(-\infty,s])$$

$$\leq hG((y',y]\times(x',x+h]) \leq h$$

where the subscript s in d_s denotes the variable of integration.

Hence,

$$(1.30) \qquad \int_0^1 Q((y',y] \times (x'+t, x'+t+h])\,dt \le m^2 h \ .$$

Thus, taking the integral wrt Lebesgue measure over $[0,1]$, and weakening the bound in various ways (including one usage of (1) and (30)) leads to

$$(1.31) \quad \int_0^1 (29)\,dt \le Q((y',y'+h] \times (x',x]) + Q((y'+h,y] \times (x',x'+h])$$

$$+ Q((y,y+h] \times (x',x+h]) + O(h + n^{-1}) \ .$$

By Lemma 2 with $k = 2$ and by three applications of (1) and Lemma 3,

$$(1.32) \qquad \text{Av.}\ P_{j-1} P_j \{ \text{rhs}(31) / Q(\underline{x}',\underline{x}] \} = O(h + n^{-1}) .$$

Thus, in view of (32), (31) and (29),

$$(1.33) \quad \text{Av.}\ P_{\underset{\sim}{j}}^2 \{ \int_0^1 (\text{first term of rhs}(28))\,dt / Q(\underline{x}',\underline{x}] \} = O(h + n^{-1}) .$$

Since by Lemma 2 with $k = 2$, Av. $\underset{\sim}{P}_j^2(Q(\underline{x}',\underline{x}])^{-1} \le (d+2-c)^2$, it follows that

$$(1.34)\quad \text{Av. } \underset{\sim}{P}_j^2\{\int_0^1(\text{second term of rhs(28)})dt/Q(\underline{x}',\underline{x}]\} = O((\sqrt{n}\, h^2)^{-1}) .$$

In view of (33), (34) and (28) we get the $k = 2$ version of asserted Lemma 6. ∎

Lemma 7.

$$(1.35)\quad \text{Av. } \underset{\sim}{P}_j^k\{(Q(\underline{x}',\underline{x}])^{-1}\check{\underset{\sim}{P}}_{j,k}|T(y,\cdot)]_{x'}^{x} - Q((y',y] \times (x',x])|\}$$

$$= O(h + n^{-1} + (\sqrt{n}\, h^2)^{-1}) .$$

Proof. The proof proceeds in the same way as that of Lemma 6. For simplicity, we let $k = 2$. Fix j until (38). Define $Z_i = h^{-k}[\underline{x} \le \underset{\sim}{X}_i^2 \le \underline{x} + h\underline{1}]/f(\underline{x})$, $\forall\ i = 2,3,\ldots,n$. Then, in view of (11), (9) and (8) we can directly verify $T(y,\cdot)]_{x'}^{x} = \bar{Z}$.

From the definition, $Z_2,\ldots,Z_n$ are 1-dependent random variables. Also, by the assumption (0.1), $0 \le Z_i \le m^2h^{-2}$ for all i. Thus, by Proposition 1 with $b-a = m^2h^{-2}$,

$$(1.36)\quad \check{\underset{\sim}{P}}_{j,2}|T(y,\cdot)]_{x'}^{x} - Q(\underline{x}'\underline{x}]|$$

$$\le |\check{\underset{\sim}{P}}_{j,2}(T(y,\cdot)]_{x'}^{x}) - Q(\underline{x}',\underline{x}]| + \frac{m^2}{2\sqrt{\lambda}\, h^2}\sqrt{\frac{\pi}{2}}$$

where λ is as defined just below (28).

Proceeding the same way that (29) was obtained, we bound the first term of rhs(36) by

$$(1.37)\quad Q((y',y]\times(x,x+h]) + Q((y,y+h]\times(x',x+h]) + Q((y'+h,y]\times(x',x'+h])$$

$$+ Q((y',y'+h]\times(x',x]) + O(h+n^{-1}) .$$

By four applications of (1) and Lemma 3 and by Lemma 2 with $k = 2$,

$$(1.38)\qquad \text{Av. } P^2_{\underset{\sim}{j}}\{(37)/Q(\underline{x}',\underline{x}]\} = O(h+n^{-1}) .$$

Thus, in view of (38) and (37),

$$(1.39)\qquad \text{Av. } P^2_{\underset{\sim}{j}}(\text{first term of rhs}(36)) = O(h+n^{-1}) .$$

Since by Lemma 2 with $k = 2$, Av. $P^2_{\underset{\sim}{j}}(Q(\underline{x}',\underline{x}])^{-1} \le (d+2-c)^2$, we obtain

$$(1.40)\qquad \text{Av. } P^2_{\underset{\sim}{j}}(\text{second term of rhs}(36)) = O((\sqrt{n}\, h^2)^{-1}) .$$

Thus, in view of (40), (39), (36) and lhs(35), we obtain $k = 2$ version of Lemma 7. ▮

Therefore, by Lemmas 6 and 7, it follows in view of (18) and (17) that

$$\text{(1.41)} \qquad \text{Av.}\ \underset{\sim}{P}|\theta_{T,j} - \theta_{jn}| = O(h + n^{-1} + (\sqrt{n}\ h^{k})^{-1}) .$$

In view of (0.4), taking h to be exact order $n^{-1/(2k+2)}$ gives

<u>Theorem 1</u>. With h, exact order, $n^{-1/(2k+2)}$,

$$|D^{k}(\underset{\sim}{\theta}, \underset{\sim}{\theta}_{T})| \leq O(n^{-1/(2k+2)}) ,$$

uniformly in $\underset{\sim}{\theta}$.

§3.2. The Procedure $\underset{\sim}{\phi}$ for $\mathcal{P}(1)$ with a Rate $(2k + 2)^{-1}$.

Throughout this section, consider $\mathcal{P}(1)$ as the underlying family of distributions, and interpret $\underset{\sim}{\phi}$ as a k-extended version of the procedure treated in Section 4 in Chapter II. $\underset{\sim}{\phi}$ will be constructed according to the analogous method used in §2.0. We then bound the modified regret of $\underset{\sim}{\phi}$ using the fact that the difference of the jth components of $\underset{\sim}{\phi}$ and $\underset{\sim}{\theta}_T$ does not exceed h, together with (1.41).

Fix j until (5). Let $\Delta(\underline{x}) = \int p_\omega(y) P_\theta(x) dG(\underline{\theta})$. Then, it is not difficult to verify

$$(2.1) \qquad \theta_{jn} = y - \{(\Delta(\underline{x}) - G((y',y] \times (-\infty,x']))/\bar{p}(\underline{x})\}$$

where $\bar{p}(\underline{x})$ is defined by (1.5). As in §2.0, estimating $\bar{p}(\underline{u})$ by $\Delta F^*(\underline{u})$ (see (1.9) with $f \equiv 1$), $\Delta(\underline{u})$ by $h^{-(k-1)}F^*((v,v+h1] \times (-\infty,u])$ and $G((v',v] \times (-\infty,u])$ by $T(\underline{u}) \doteq \Sigma\, \Delta F^*(v,u-r)$, and noting from (1.4) that $0 \le \eta_{jn} \le 1$, gives us an estimate of θ_{jn} as

$$(2.2) \qquad \phi_{jn} = x - (0 \vee \psi_{jn}) \wedge 1$$

where

$$(2.3) \qquad \psi_{jn} = \{h^{-(k-1)}F^*((y,y+h]\times(-\infty,x]) - T(y,x')\}/T(y,\cdot)]_{x'}^{x},$$

where 0/0 is taken to be zero.

We first investigate the relation between $\underset{\sim}{\phi}$ and $\underset{\sim}{\theta}_T$. We abbreviate ψ_{jn} to ψ and φ_{jn} in the definition of $\theta_{T,j}$ to φ and will show φ lies between ψ-h and ψ. In view of (3) and (1.13)

$$(2.4) \qquad \varphi-\psi = \{\int_0^1 T(y,x'+t)dt - h^{-(k-1)}F^*(\cdot,x)]_y^{y+h}\}/T(y,\cdot)]_{x'}^{x} .$$

In the definition (1.10) of T(y,x'), for every i,

$$\int_0^1 \Sigma[y'+t-r < X_i \le y'+t-r+h]dt = h, \le h, = 0,$$

according as $X_i \le y$, $y < X_i \le y+h$, $y+h < X_i$. Applying two separate cases $X_i \le y$ and $y < X_i \le y+h$ to the rhs of the first equality below gives us

$$h^2\int_0^1 T(y,x'+t)dt = \text{Av.}[y < \underset{\sim}{X}_{i-1}^{k-1} \le y+h](\int_0^1 \Sigma[x'+t-r < X_i \le x'+t-r+h]dt)$$

$$= hF^*(\cdot,x)]_y^{y+h1}+[y < \underset{\sim}{X}_{i-1}^{k-1} \le y+h1,\ x < X_i \le x+h]\int_0^1\Sigma[x'+t-r<X_i\le x'+t-r+h]dt.$$

Hence

$$h^{-1}F^*(\cdot,x)]_y^{y+h1} \le \int_0^1 T(y,x'+t)dt \le h^{-1}F^*(\cdot,x+h)]_y^{y+h1} .$$

Thus, applying this to rhs(4) and noticing, by the definition (1.11) of $T(y,\cdot)]_{x'}^{x}$, that $T(y,\cdot)]_{x'}^{x} = h^{-1}\{F^*(\cdot,x+h)]_y^{y+h1} - F^*(\cdot,x)]_y^{y+h1}\}$, we obtain

$$(2.5) \qquad 0 \le \varphi - \psi \le h ,$$

analogous to (2.4.1) in the $k = 1$ version.

We now consider the modified regret of $\underset{\sim}{\phi}$. Since $X_j' < \phi_{jn} \le X_j$, it follows by (0.4) and by the triangular inequality with $\theta_{T,j}$ as an intermediate term that

$$(2.6) \qquad 2^{-1}|D^k(\underset{\sim}{\theta},\underset{\sim}{\phi})| \le \text{Av.}\underset{\sim}{P}|\phi_{jn} - \theta_{T,j}| + \text{Av.}\underset{\sim}{P}|\theta_{T,j} - \theta_{jn}| .$$

Since (by (2), (1.12) and (5)) $|\phi_{jn} - \theta_{T,j}| \le (\psi - \varphi) \wedge 1 \le h$,

$$(2.7) \qquad \text{first term of rhs}(6) \le h .$$

On the other hand, since the inequality (1.41) for $f \equiv 1$ bounds the second term of rhs(6), it follows by (6) that

$$|D^k(\underset{\sim}{\theta},\underset{\sim}{\phi})| = O(h + n^{-1} + (\sqrt{n}\, h^k)^{-1}) .$$

Hence, taking h to be exact order $n^{-1/(2k+2)}$ gives

<u>Theorem 2</u>. For h with exact order $n^{-1/(2k+2)}$,

$$|D^k(\underset{\sim}{\theta}, \underset{\sim}{\phi})| = O(n^{-1/(2k+2)}) ,$$

uniformly in $\underset{\sim}{\theta}$.

APPENDIX

APPENDIX

Section A.1 (the main development) relates to bounds for difference of two integrals of a bounded function in terms of extensions of Lévy metric. Section A.2 relates to convergence of a sequence of variances. Some of the results in §A.1 were used in §1.1 and the consequence in §1.2 was applied in §2.2.

§A.1. Extensions of Lévy Metric and Bounds for Difference of Two Integrals of a Bounded Function.

We first extend Lévy metric L to the family $\mathcal{F}$ of increasing real functions on R and then introduce an extension ρ to the family $\mathcal{M}$ of measures on $(R,\mathcal{B})$ determined by the variation of elements in $\mathcal{F}$.

ρ is defined as the infimum of L's and Remark A.1 shows the infimum is attained. Proposition A bounds L at retractions to an interval by the maximum of differences of values of the functions at end points of the interval and L at the unretracted functions.

We then prove strengthened generalizations (Lemma A.2) of Lemmas 8' and 8 of Oaten (1969, Appendix) giving bounds on the difference of two integrals of a bounded function. Lemma A.3 introduces another family of bounds for the same difference.

Proposition A and Lemma A.3 are used in the proof of Lemma 6 in §1.1. Although Lemma A.2 was derived for this purpose, it is now included only for its own sake (as a generalization of Oaten's results) since Lemma A.3 gives a better bound in this application.

For each $F \in \mathcal{F}$, let pre-subscripts on F denote composition with the indicated translation, that is,

$$_{\varepsilon}F(x) = F(\varepsilon + x),$$

let $F^{\cdot}(x) = x + F(x)$ and note $\varepsilon + ({}_{\varepsilon}F)^{\cdot} = {}_{\varepsilon}(F^{\cdot})$. For every $r \in R$, let S_r be the interval

$$S_r = \{\varepsilon \geq 0 : {}_{-\varepsilon}(F^{\cdot}) \leq r + G^{\cdot} \leq {}_{\varepsilon}(F^{\cdot})\} .$$

Note that (i) replacement by strict inequalities throughout would, at most, subtract an end point from S_r, (ii) replacement by restrictions to a dense subset of R would, at most add an end point to S_r. Therefore neither would affect definitions which follow. Lévy distance L of F and G in $\mathcal{F}$ is defined by

(1.1) $$L(F,G) = \wedge S_0 .$$

That L is a pseudo metric will be seen in Lemma A.1 where L is shown to be the supremum of the difference of the quantiles of modifications.

For right continuous F and G, $S_r(F,G)$ is closed. For, taking $r = 0$ without loss of generality (since $S_r(F,G) = S_0(F,r+G)$)

and letting $\varepsilon \downarrow L(F,G) = L$ through points of S_0 gives $G^{\cdot} \leq {}_L(F^{\cdot})$ and, by symmetry of Lévy distance, $F^{\cdot} \leq {}_L(G^{\cdot})$ which is equivalent to ${}_{-L}(F^{\cdot}) \leq G^{\cdot}$.

We define another distance function ρ on $\mathcal{F}$ as follows: for any F and G in $\mathcal{F}$,

$$(1.2) \qquad \rho(F,G) = \bigwedge_{r\in R} L(F, r+G) \ .$$

Note that ρ is invariant under translates of values of F and G. Since functions in $\mathcal{F}$ which differ only by a constant except at discontinuity points induce the same measure, ρ is actually a metric on $\mathcal{M}$:

$$(1.3) \qquad \rho(\mu, \nu) = \rho(F,G)$$

for any F and $G \in \mathcal{F}$, inducing the respective measures μ and ν.

Since $\bigwedge_r (\bigwedge S_r) = \bigwedge_r (\bigcup S_r)$ for any family of subsets S_r of extended real line, we see that

$$(1.4) \qquad \rho = \bigwedge_r (\bigcup S_r) \ .$$

Although we have used + (-) in the subscript position to denote the positive (negative) part, we will also use + (-) on the line to denote right (left) limit.

<u>Remark A.1</u>. The infimum in the definition of ρ is attained.

Proof. Pick a sequence $\{\epsilon_n\}$ of numbers which strictly decreases to ρ. Then, by (4) there exists r_n such that

$$-r_n + {}_{-\epsilon_n}(F^{\cdot}) \leq G^{\cdot} \leq -r_n + {}_{\epsilon_n}(F^{\cdot}) .$$

Thus, taking $\overline{\lim}$ and $\underline{\lim}$ on the lhs and the rhs respectively, leads to

$$- \underline{\lim}\, r_n + {}_{-\rho}(F^{\cdot}) - \leq G^{\cdot} \leq -\overline{\lim}\, r_n + {}_{\rho}(F^{\cdot}) + .$$

Therefore, for every $r \in [\underline{\lim}\, r_n, \overline{\lim}\, r_n]$, $L(F, r+G) = \rho$.

For each $F \in \mathcal{F}$ and $t \in R$, let t_F denote the t-th quantile of $F^{\cdot}$. Note that $t \to t_F$ maps R onto R. Define η by

$$\eta(F,G) = \bigvee_{t \in R} |t_F - t_G| . \blacksquare$$

Lemma A.1. $L = \eta$.

Proof. To show $L \leq \eta$ we first have by the definition of the t-th quantiles that

$$F^{\cdot}(t_F -) \leq G^{\cdot}(t_G +) \quad \text{and} \quad G^{\cdot}(t_G -) \leq F^{\cdot}(t_F +) .$$

Hence, for every $\delta > 0$

$$F^{\cdot}(t_G-\eta-\delta) \le F^{\cdot}(t_F-\delta) \le G^{\cdot}(t_G+\delta)$$

and

$$G^{\cdot}(t_G-\delta) \le F^{\cdot}(t_F+\delta) \le F^{\cdot}(t_G+\eta+\delta) \ .$$

Since the mapping $t \to t_G$ is onto, these inequalities show that $L(F,G) \le \eta(F,G) + 2\delta$ and thus $L \le \eta$.

On the other hand, if $L(F,G) < \varepsilon$, then $_{-\varepsilon}(F^{\cdot}) \le G^{\cdot} \le {}_{\varepsilon}(F^{\cdot})$. Noticing that the t-th quantiles have the opposite ordering and by the definition of $_s(F^{\cdot})$

$$(\text{t-th quantile of } {}_s(F^{\cdot})) = t_F - s \ ,$$

we obtain $t_F+\varepsilon \ge t_G \ge t_F-\varepsilon$. Thus $\eta(F,G) \le \varepsilon$ and therefore $\eta \le L$. ∎

We now prove the following proposition.

<u>Proposition A</u>. Let $I = (a,b]$ be a finite interval and let F_I be the retraction of F into the closed interval $[F(a+), F(b+)]$. Then,

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

<u>Proof</u>. Let

$$(1.5) \qquad \vee = F^{\cdot} \vee G^{\cdot} \quad \text{and} \quad \wedge = F^{\cdot} \wedge G^{\cdot} .$$

It is straightforward from the definition of t-th quantiles to show that according as

$$t \le \vee(a+) \quad \text{or} \quad \ge \wedge(b+) \quad \text{or} \quad \in (\vee(a+), \wedge(b+)),$$

we have

$$|t_{F_I} - t_{G_I}| \le |(F^{\cdot}-G^{\cdot})(a+)| \quad \text{or} \quad \le |(F^{\cdot}-G^{\cdot})(b+)|$$

$$\text{or} \quad = |t_F - t_G|,$$

where strict inequalities hold for $\wedge(a+) < t \le \vee(a+)$ or $\wedge(b+) \le t < \vee(b+)$. Thus,

$$\eta(F_I, G_I) \le |(F^{\cdot}-G^{\cdot})(a+)| \vee |(F^{\cdot}-G^{\cdot})(b+)| \vee \eta(F,G) .$$

Therefore, Lemma A.1 leads to the asserted inequaltiy. ∎

Definition A.1. With h, a function defined on a real interval I, the modulus of continuity of h is the function given by

$$(1.6) \qquad \alpha(\varepsilon) = \vee\{[h]_{\omega_2}^{\omega_1} : \omega_1, \omega_2 \in I, |\omega_1 - \omega_2| < \varepsilon\}$$

for every $\varepsilon > 0$.

Definition A.2. With h measurable on a real interval I supporting a finite measure τ,

$$(1.7)\qquad \tau\text{-sup } h = \wedge\{\delta: \tau[h > \delta] = 0\},$$

$$\tau\text{-inf } h = -(\tau\text{-sup }(-h))$$

and, with $\tau_{r\varepsilon}$ denoting the restriction of τ to the interval $(r-\varepsilon/2, r+\varepsilon/2)$, the τ-modulus of continuity of h is the function given by

$$(1.8)\qquad \tau\text{-}\alpha(\varepsilon) = \vee\{\tau_{r\varepsilon}\text{-sup } h - \tau_{r\varepsilon}\text{-inf } h : r \in I\}$$

for every $\varepsilon > 0$.

The following Lemma A.2 is a unified and slightly strengthened generalization of Lemmas 8' (corrected by replacing 3 by 4 in the bound) and 8 of Oaten (1969, Appendix) with proof evolving from those of Oaten.

Lemma A.2. 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 $[c,d]$. By abbreviating $\rho(\mu,\nu)$ to ρ and $L(\mu[a,\cdot], \nu[a,\cdot])$ to L, $|\int hd(\mu-\nu)|$ has the following families of upper bounds:

$$(1.9)\quad \alpha\left(\left(\frac{b-a}{k} \vee L\right)+\right)\{(k-1)L + |\mu-\nu|I+2(\mu I \wedge \nu I)\}+((-c) \vee d)|\mu I-\nu I|,$$

$$\forall \text{ positive integer } k < \frac{b-a}{L} + 1$$

$$(1.10)\quad (d-c)k\rho + (\mu+\nu)\cdot\alpha\left(\rho+\left(\frac{b-a}{k} \vee (2\rho)\right)+\right)(\mu I \wedge \nu I)+((-d) \vee c)|\mu I-\nu I|$$

$$\forall \text{ positive integer } k < \frac{b-a}{2\rho} + 1 .$$

The bounds in (9) and (10) hold for every positive integer k, but those unlisted are dominated by the bounds corresponding to the largest k listed.

Remark. The bounds of Oaten are parametrized by $\lambda \in (L,\infty)$ and $(2L,\infty)$ respectively and are improved by the $\mu I = \nu I = 1$ specialization of (9) and (10) above with k taken to be the least integer greater than $(b-a)/\lambda$.

Proof of Lemma A.2. For a given σ with $k-1 < (b-a)/\sigma < k$, let $\delta \doteq k\sigma-(b-a)$ and let $x_j \doteq a+j\delta-2^{-1}\delta$ for $j = 0,1,2,\ldots,k$. Since $\sigma < (b-a)/(k-1)$, it follows that $\delta < \sigma$ and hence $(x_0+x_1)/2$ and $(x_{k-1}+x_k)/2$ both lie inside the interval I.

Proof of (9) . Note $L < (b-a)/(k-1)$ and take $\sigma > ((b-a)/k) \vee L$. Let $h_j \doteq h((x_{j-1}+x_j)/2)$ for $j = 1,2,\ldots,k$. Then, $|h(x)-h_j| \le \alpha(\frac{1}{2}\sigma+)$ for each $x \in (x_{j-1},x_j]$, and $|h_j-h_{j+1}| \le \alpha(\sigma+)$ for each j.

Let $D_j = (\mu-\nu)(x_0,x_j]$, $j = 0,1,\ldots,k$. Then

$$\int h\,d(\mu-\nu) = \Sigma_{j=1}^{k}\{\int(x_{j-1},x_j](h-h_j)d(\mu-\nu) + h_j(D_j-D_{j-1})\}$$

(1.11)

$$\leq \alpha(\tfrac{1}{2}\sigma+) + |\mu-\nu|I + h_k D_k + \Sigma_{j=1}^{k-1}(h_j-h_{j+1})D_j .$$

From $\sigma > L$,

$$(D_j)_+ \leq (\nu(x_j, x_{j+1}]+L) \wedge (\mu(x_{j-1},x_j]+L)$$

$$\leq \nu(x_j,x_{j+1}] \wedge \mu(x_{j-1},x_j] + L$$

and, by interchange of μ and ν,

$$(D_j)_- \leq \mu(x_j, x_{j+1}] \wedge \nu(x_{j-1},x_j] + L .$$

Thus, henceforth abbreviating μI and νI to μ and ν,

$$\Sigma_{j=1}^{k-1}|D_j| \leq 2(\mu \wedge \nu) + (k-1)L .$$

Therefore,

$$\Sigma_{j=1}^{k-1}(h_j-h_{j+1})D_j \leq \alpha(\sigma+)\{2(\mu \wedge \nu) + (k-1)L\} .$$

Combining this with the inequality $h_k D_k \leq d(\mu-\nu)_+ + (-c)(\nu-\mu)_+$, we obtain that

$$(1.12)\quad \text{lhs}(11) \leq \alpha(\sigma+)\{|\mu-\nu|I+2(\mu\wedge\nu)+(k-1)L\}+d(\mu-\nu)_+ +(-c)(\nu-\mu)_+ .$$

Replacing h by $-h$ gives us

$$(1.13)\quad -\text{lhs}(11) \leq \gamma(\sigma+)\{|\mu-\nu|I+2(\mu\wedge\nu)+(k-1)L\}+(-c)(\mu-\nu)_+ +d(\nu-\mu)_+ .$$

We obtain (9) by taking the maximum of rhs(12) and rhs(13), recognizing $(d(\mu-\nu)_+ +(-c)(\nu-\mu)_+)\vee((-c)(\mu-\nu)_+ +d(\nu-\mu)_+)=((-c)\vee d)|\mu-\nu|$ and letting σ decrease to $((b-a)/k)\vee L$.

<u>Proof of (10)</u>. Since, by Remark A.2, $\rho = L(F,G)$ for some right continuous distribution functions F and G inducing μ and ν, it suffices to prove (10) with ρ replaced by $L = L(F,G)$. As in the proof of (9), note $2L < (b-a)/(k-1)$ and take $\sigma > ((b-a)/k) \vee 2L$.

By the definition of L we can find $x_0 = y_0 < y_1 < \cdots < y_k = x_k$ so that, for each j, $|x_j - y_j| \leq L$ and

$$(1.14)\qquad F(y_j-) - L \leq G(x_j) \leq F(y_j) + L$$

because

$$\bigcup_y \{[F(y-)-L,\ F(y)+L] : |y-x_j| \le L\}$$

$$= [F((x_j-L)-)-L,\ F(\ x_j+L\ )+L]$$

and the intervals $[x_j-L,\ x_j+L]$ strictly increase wrt j.

We extend the domain of h to the interval $[x_0,\ x_k]$ by defining $2h = c + d$ on complement of I.

For each j, let $\wedge_j = x_j \wedge y_j$ and $\vee_j = x_j \vee y_j$. Let $\tau = \mu+\nu$, let τ_j denote the restriction of τ to the interval $[\wedge_j,\ \vee_{j+1}]$ and let $\underline{h}_j = \tau_j\text{-}\inf h$. Then, define functions h_1 and h_2 by

$$h_1(x_j, x_{j+1}] = h_2(y_j, y_{j+1}) = \underline{h}_j \quad \text{and} \quad h_2(y_j) = h_2(y_j-) \vee h_2(y_j+) \ .$$

Now $h-h_1 \le \tau\text{-}\alpha(L+\sigma+)$ a.e. τ on $(x_j, x_{j+1}]$ because

$$\forall\ \delta > 0 \quad \tau((x_j, \vee_{j+1}]\cap[h < \underline{h}_j+\delta])+\tau([\wedge_j, x_{j+1}]\cap[h < \underline{h}_j+\delta]) > 0$$

so that if $\tau((x_j, x_{j+1}]\cap[h-\underline{h}_j > \lambda]) > 0$ then $\tau\text{-}\alpha(L+\sigma+) \ge \lambda-\delta$ and thus $\ge \lambda$.

Also $h_2-h \le 0$ a.e. τ because $h \ge \underline{h}_{j-1}\vee\underline{h}_j \ge h_2$ a.e. τ on $[\wedge_j, \vee_j]$ and $h \ge \underline{h}_j = h_2$ a.e. τ on $(\vee_j, \wedge_{j+1})$.

Let $r \in R$. If $h_2(y_{j-1}, y_{j+1}) \le r$, then $h_1(x_{j-1}, x_{j+1}] \le r$. Conversely $h_1(x_{j-1}, x_{j+1}] \le r$ implies $h_2((y_{j-1}, y_j)\cup(y_j, y_{j+1})) \le r$ and therefore $h_2(y_{j-1}, y_j) \le r$. Hence $h_2^{-1}(-\infty, r]$ is the union of at most $k/2$ intervals of the form (y_i, y_j), and $h_1^{-1}(-\infty,r]$ is the union of the corresponding intervals $(x_i, x_j]$.

We note that, by two applications of (14), $\mu(y_i, y_j) \le \nu(x_i, x_j]+2L \;\forall\; j$ so that

$$(1.15) \qquad \mu h_2^{-1}(-\infty,r] \le \nu h_1^{-1}(-\infty,r] + kL \;.$$

By two usages of the Funini representation of the integral (cf. (2.1.10)) of a nonnegative function in the rhs of the first equality below

$$(1.16) \quad \int h_1 d\nu - \int h_2 d\mu - d(\nu I - \mu I) = \int (d-h_2)\, d\mu - \int (d-h_1)\, d\nu$$

$$= \int_0^{d-c} (\mu h_2^{-1} - \nu h_1^{-1})(-\infty, d-t]dt \le (d-c)kL \;.$$

Henceforth abbreviating $\tau \cdot \alpha(\sigma + L +)$ to $\underline{\alpha}$, μI to μ and νI to ν, the triangle inequality and (16) bound $\int hd(\nu-\mu)-(d-c)kL$ by

$$(1.17) \quad \int (h-h_1)\, d\nu + \int (h_2 - h)\, d\mu + d(\nu-\mu) \le \underline{\alpha}\nu + d(\nu-\mu) \;.$$

Applying (17) to $-h$ with the measures interchanged gives the bound $\underline{\alpha}\mu + (-c)(\mu-\nu)$. The minimum of those bounds is the former or latter according as $\mu \ge$ or $\le \nu$ and therefore

$$(1.18)\qquad \int hd(\nu-\mu) \le (d-c)kL + \underline{\alpha}(\mu \wedge \nu) + c(\nu-\mu)_+ - d(\mu-\nu)_+ .$$

Applying (18) to $-h$ gives rhs(18) with c,d replaced by $-d,-c$ and gives

$$(1.19)\qquad \left|\int hd(\mu-\nu)\right| \le (d-c)kL + \underline{\alpha}(\mu \wedge \nu) + (c \vee (-d))\,|\mu-\nu| ,$$

and (10) results on letting σ decrease to $((b-a)/k) \vee 2L$. ∎

In the following lemma, the natural generalization of the inverse probability integral transformation is used to develop bounds for the same difference of integrals without recourse to partitioning.

Lemma A.3. Let I,μ,ν and h be as in Lemma A.2. Let F and G be distribution functions inducing μ and ν with $\vee(a-) \le \wedge(b+)$ where $\vee$ and $\wedge$ abbreviate $F^{\cdot} \vee G^{\cdot}$ and $F^{\cdot} \wedge G^{\cdot}$ (as in (5)). Then $\left|\int h\, d(\mu-\nu)\right|$ has the following family of bounds

$$(1.20)\qquad \frac{d-c}{2}\{|(F-G)(a-)|+|(F-G)(b+)|\}+\alpha(L(F,G)+)\{\wedge(b+)-\vee(a-)\}$$

$$+\frac{d+c}{2}|\mu I-\nu I| .$$

Proof. Without loss of generality we can assume I is open. For, $\forall\ \epsilon > 0$ $\{a,b\}\subset(a-\epsilon,b+\epsilon)$ to which h is extendible with the same modulus of continuity and, if (20) holds with a,b replaced by $a-\varepsilon$, $b+\varepsilon$, then letting $\varepsilon\downarrow 0$ gives (20) with $I = \{a,b\}$.

Let $I = (a,b)$ and let f denote the map $t \to t_F$. Since $f^{-1}\{u\} = [F^{\cdot}(u-), F^{\cdot}(u+)]$ and $F^{\cdot}$ is strictly increasing,

$$f^{-1}(\beta,\gamma) = \bigcup_{\beta<u<\gamma} f^{-1}\{u\} = (F^{\cdot}(\beta+), F^{\cdot}(\gamma-)) .$$

Thus, Lebesgue measure and f induce the measure with $F^{\cdot}$ as a distribution function on the range of f. By the transformation theorem (cf. Halmos (1950), p. 163),

$$\text{(1.21)} \qquad \int_I h \, dF^{\cdot} = \int_{F^{\cdot}(a)}^{F^{\cdot}(b)} h(t_F)dt .$$

Letting $\delta(H) = 1$ or -1 according as $H = F$ or G, the difference of (21) for F and for G results in the following representation for $\int h \, d(\mu-\nu)$:

$$\int_{\wedge(a)}^{\vee(a)} \delta(S)h(t_S)dt + \int_{\vee(a)}^{\wedge(b)} h(t_F)-h(t_G)dt + \int_{\wedge(b)}^{\vee(b)} \delta(T)h(t_T)dt$$

where S and T have values in the set $\{F,G\} \ni S^{\cdot}(a) = \wedge(a)$ and $T^{\cdot}(b) = \vee(b)$. Hence, abbreviating $L(F,G)$ to L hereafter and using Lemma A.1, $\int h \, d(\mu-\nu) \leq$

$$(1.22)\quad d((G-F)(a))_+ + (-c)((F-G)(a))_+ + \alpha(L+)(\wedge(b)-\vee(a)) + d((F-G)(b))_+$$

$$+ (-c)((G-F)(b))_+ .$$

Applied to $-h$, (22) is altered only by c,d changing to $-d$, $-c$:

$$(1.23)\quad (-c)((G-F)(a))_+ + d((F-G)(a))_+ + \alpha(L+)(\wedge(b)-\vee(a)) + (-c)((F-G)(b))_+$$

$$+ d((G-F)(b))_+ .$$

Since (22) + (23) = $(d-c)\{|(F-G)(a)| + |(F-G)(b)|\} + 2\alpha(L+)(\wedge(b)-\vee(a))$ and (22) - (23) = $(d+c)(\mu I - \nu I)$, (22) $\vee$ (23) = (20). ∎

§A.2. A Fatou Theorem for Variances.

The following theorem is used in Section 2.2.

Theorem A.1. If $\{U_n\}$ is a sequence of random variables converging in distribution to a random variable U, then $\underline{\lim}\ \mathrm{Var}(U_n) \geq \mathrm{Var}(U)$.

Proof. It suffices to show that for $\{U_n\}$ such that $\mathrm{Var}(U_n) \to$ finite.

With $\mu_n = EU_n$ and $\sigma_n^2 = \mathrm{Var}\ U_n$, the Tchebycheff inequality gives $P[|U_n - \mu_n| < \sqrt{2}\,\sigma_n] \geq 1/2$ while tightness provides a finite b independent of n for which $P[|U_n| \leq b] > 1/2$. The nonemptyness of the intersection of these events shows $|\mu_n| < b + \sqrt{2}\,\sigma_n$ so that $\{\mu_n\}$ is bounded.

Letting $\{\mu_m\}$ be a convergent subsequence with limit μ_∞, $U_m - \mu_m \xrightarrow{\mathcal{D}} U - \mu_\infty$ and hence (cf. Loéve (1963) 11.4, A(i))

$$\lim \mathrm{Var}(U_n) = \lim E(U_m - \mu_m)^2 \geq E(U - \mu_\infty)^2 \geq \mathrm{Var}\ U . \blacksquare$$

BIBLIOGRAPHY

BIBLIOGRAPHY

Breiman, L. (1968). Probability. Addison-Wesley

Deely, J.J. and Kruse, R.L. (1968). Construction of sequences estimating the mixing distribution. Ann. Math. Statist. 39, 286-288.

Feller, W. (1968). An Introduction to Probability Theory and its Applications Volume I (3rd ed.). Wiley, New York.

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

Ferguson, T.S. (1967). Mathematical Statistics a Decision Theoretic Approach. Academic Press.

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.

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

Gilliland, D.C. and Hannan, J.F. (1969). On an extended compound decision problem. Ann. Math. Statist. 40, 1536-1541.

Halmos, P.R. (1950). Measure Theory. Litton Educational Publishing.

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

Loéve, Michel (1963). Probability Theory (3rd ed.). Van Nostrand, Princeton.

Neveu, J. (1965). Mathematical Foundations of the Calculus of Probability. Holden-Day.

Oaten, A. (1969). Approximation to Bayes risk in compound decision problems. Research Memorandum RM-233, Department of Statistics and Probability, Michigan State University.

Oaten, A. (1972). Approximation to Bayes risk in compound decision problems. Ann. Math. Statist. 43, 1164-1184.

Royden, H.L. (1963). Real Analysis. The Macmillan Company, New York.

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

Sibley, D.A. (1971). A metric for weak convergence of distribution functions. Rocky Mountain Journal of Math. 1, No. 3, 427-430.

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

Swain, Donald D. (1965). Bounds and rates of convergence for the extended compound estimation problem in the sequence case. Technical Report No. 81, Department of Statistics, Stanford University.

Yu, Benito Ong (1971). Rates of convergence in empirical Bayes two-action and estimation problems and in extended sequence-compound estimation problems. Research Memorandum RM-279, Department of Statistics and Probability, Michigan State University.

MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 03169 0203