ESTIMATION OF DERIVATIVES OF AVERAGE OF μ - DENSITIES AND SEQUENCE - COMPOUND ESTIMATION IN EXPONENTIAL FAMILIES

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

RADHEY SHYAM SINGH

1974

This is to certify that the

thesis entitled

ESTIMATION OF DERIVATIVES OF AVERAGE OF μ-DENSITIES AND SEQUENCE - COMPOUND ESTIMATION IN EXPONENTIAL FAMILIES

presented by

Radhey Shyam Singh

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

James Hannan
Major professor

Date 2/21/74

O-7639

ABSTRACT

ESTIMATION OF DERIVATIVES OF AVERAGE OF μ-DENSITIES AND SEQUENCE-COMPOUND ESTIMATION IN EXPONENTIAL FAMILIES

By

Radhey Shyam Singh

Let $X_1,\ldots,X_n$ be independent random variables with μ-densities $f_1,\ldots,f_n$, where μ is a σ-finite measure dominated by Lebesgue measure on the real line R. With a fixed integer $\nu \geq 0$, we exhibit kernel estimators $\hat{\bar{f}}^{(\nu)}$ of $n^{-1}\Sigma_1^n f_j^{(\nu)}$.

For any subset D of R, we give sufficient and (somewhat) necessary conditions for asymptotic unbiasedness (asy. u.), almost sure (a.s.) and mean square (m.s.) consistencies, each uniform on D. We also prove integrated mean square (i.m.s.) consistency, and obtain convergence rates and exact rates for the asy. u., m.s. and i.m.s. consistencies. When $\bar{f}^{(r)}$, for an integer $r > \nu$, exists on D, we show that the error term is $O((n^{-1}\log n)^{(r-\nu)/2(1+r)})$ with probability one, while m.s. and i.m.s. errors are $O(n^{-2(r-\nu)/(1+2r)})$, each uniform on D. The vector $(\hat{\bar{f}}^{(\nu)}(x_1),\ldots,\hat{\bar{f}}^{(\nu)}(x_m))$ is shown to be asymptotically m-variate normal.

We extend this estimation to multivariate case. Specifically, estimation of mixed partial derivatives of the average of p-variate μ-densities has been considered.

G-5078-2

Radhey Shyam Singh

We make applications of $\hat{\tilde{f}}^{(\nu)}$ to sequence-compound squared error loss estimation (SELE). With an observation on X distributed according to $(\sim)\ P_\omega \in \mathcal{P}$, an exponential family wrt μ and $\omega \in \Omega$, the natural parameter space, we take SELE of $\theta(\omega) = \omega$, e^ω or ω^{-1} as our component problem.

With $(X_1,\ldots,X_n) \sim \underset{\sim}{P}_n = P_{\omega_1} \times\ldots\times P_{\omega_n} \in \mathcal{P}^n$, a (sequence-compound) estimator of $\underset{\sim}{\theta} = (\theta(\omega_1),\ldots,\theta(\omega_n))$ is $\underset{\sim}{\varphi} = (\varphi_1,\ldots,\varphi_n)$ with φ_i $(X_1,\ldots,X_i)$-measurable. With a $\delta > 0$, and G_n the empiric distribution function of $\omega_1,\ldots,\omega_n$, and $R(G_n)$ the Bayes risk at G_n in the component problem, we say $\underset{\sim}{\varphi}$ has a rate δ at $\underset{\sim}{\theta}$ if the modified regret $n^{-1}\Sigma_1^n \underset{\sim}{P}_i(\varphi_i - \theta(\omega_i))^2 - R(G_n) = O(n^{-\delta})$.

With $\alpha_i < \beta_i$ in Ω such that $-\alpha_i$ and β_i are increasing in i, we exhibit estimators (of $\underset{\sim}{\theta}$) having certain rates uniformly in $\underset{\sim}{\omega} \in \times_1^n[\alpha_i,\beta_i]$. These rates depend on the speed at which $|\alpha_n| \vee |\beta_n| \uparrow \infty$ as $n \uparrow \infty$. When α_i, β_i are constants wrt i and satisfy certain conditions, we exhibit a divided difference estimator of $\underset{\sim}{\omega}$ with a rate 1/5, and kernel estimators, (for each integer $r > 0$), of $\underset{\sim}{\theta}$ with rates $(r-1)/(1+2r)$, $r/(1+2r)$ or $(r-)/(1+2r)$ according as $\theta(\omega) = \omega$, e^ω or ω^{-1}, where for the case ω, $r > 1$. When $\theta(\omega) = \omega$, and $\underset{\sim}{\omega}$ has identical components, we show that rates with the divided difference and the kernel estimators of $\underset{\sim}{\omega}$ are near, but cannot be more than, 2/5 and $2(r-1)/(1+2r)$, respectively.

ESTIMATION OF DERIVATIVES OF AVERAGE OF μ-DENSITIES AND SEQUENCE-COMPOUND ESTIMATION IN EXPONENTIAL FAMILIES

By

Radhey Shyam Singh

A THESIS

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

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1974

TO MY PARENTS

ACKNOWLEDGEMENTS

I wish to express my deep gratitude to Professor James Hannan for the patience he accorded me in the preparation of this thesis. His careful criticism and invaluable suggestions aided greatly in improving and simplifying virtually all of the results in the thesis.

In addition, among many others who helped, I wish to record my thanks to Professor Dennis Gilliland for reading and commenting on a difficult rough draft; to Mrs. Noralee Barnes for her excellent typing; and to my wife for her great understanding and endurance.

Finally I would like to thank the Department of Statistics and Probability at Michigan State University for the generous support, financial and otherwise, during my stay at Michigan State University.

TABLE OF CONTENTS

Page

APPENDIX

0. INTRODUCTION

In this thesis we consider estimation of derivatives of the average of densities with applications to sequence-compound squared error loss estimation (SELE).

0.1. Estimation of Derivatives of the Average of Densities.

Estimation of a Lebesgue-density, hereafter L-density, has been studied by various authors, and a variety of methods have been used: For example, Watson and Leadbetter (1963), and Nadarya (1965) used the kernel method first introduced by Rosenblatt (1956), and studied in detail by Parzen (1962); Cencav (1962), Schwartz (1967), Kronmal and Tarter (1968), and Watson (1969) used the orthogonal series method; Weiss and Wolfowitz (1967), Rao (1969), and Wegman (1969) used maximum likelihood methods; Van Ryzin (1970) and Wahba (1971) used, respectively, histogram and polynomial (Lagrange)-interpolation methods. Estimation of derivatives of a L-density has also been considered by Bhattacharya (1967) and Schuster (1969).

Estimation of non-Lebesgue densities and their derivatives arises in Empirical Bayes problems, while that of the averages of non-Lebesgue densities and their derivatives arises in compound decision problems. Yu (1970), (Section 2 of the appendix), exhibits kernel estimators of a μ-density and its derivative, where $d\mu = u(x)dx$ and, for some $a \geq -\infty$, $u(x) > 0$ iff $x > a$. He gives rates for mean square errors (m.s.e.) at each point on the real line R. Samuel (1965), (Section 6), exhibits kernel estimators of the average of L-densities and, under uniform equicontinuity (hence, necessarily uniform equiboundedness) of densities on a subset D of R, she proves asymptotic unbiasedness (asy. u.) and weak consistency, both uniform on D. Susarla (1970), (Section 1.3), exhibits kernel estimators of the average and its first partial derivatives of m-variate normal densities with known covariance and uniformly bounded unknown means, and obtains rates for m.s.e. uniform on R^m. Samuel uses Parzen-type kernels, while Yu and Susarla use those of Johns and Van Ryzin (1972).

We consider here non-parametric estimation of derivatives of the average of non-Lebesgue densities. Let μ be a σ-finite measure with density u wrt Lebesgue measure on R. Let $X_1,\ldots,X_n$ be independent random variables with X_j having a μ-density f_j. With $r \geq \nu \geq 0$ fixed integers, we exhibit kernel estimators $\hat{\bar{f}}^{(\nu)}(x)$, depending on $X_1,\ldots,X_n$, u and r, of $\bar{f}^{(\nu)} = n^{-1}\Sigma_1^n f_j^{(\nu)}$. (If u were known to be at least as smooth as f_j were, we would estimate derivatives of the average of L-densities uf_j directly.)

In the remainder of this section we describe the main results contained in Chapter 1. Bounds obtained here are quite explicit. We make almost no assumption on u for some of the results on asy. u., a.s., m.s. and integrated mean square (i.m.s.) consistencies. For any subset D of R and any $h = h_n \downarrow 0$ as $n \uparrow \infty$, if $\sup_{x \in D} h^{-1}\int_x^{x+h} |\bar{f}^{(\nu)}(t) - \bar{f}^{(\nu)}(x)|dt \to 0$ as $n \to \infty$, and if, in case $\nu > 0$, the ν-th order Taylor expansion of $\bar{f}(x + hy)$ about x with integral form of the remainder exists for all $0 < y < 1$ and for each x in D, then, under certain boundedness conditions on $1/u$, asy. u., a,s. and m.s. consistencies, uniform on D, are proved (in Sections 2, 3 and 5, respectively). (Thus, contrary to the assumption made for similar results in most of the papers on the subject, asymptotic continuity of $\bar{f}^{(\nu)}$ at the estimation point is not needed.) In Section 4, we obtain rates and exact rate for the $\mathrm{var}(\hat{\bar{f}}^{(\nu)})$ and prove the asymptotic normality of the vector $(\hat{\bar{f}}^{(\nu)}(x_1),\ldots,\hat{\bar{f}}^{(\nu)}(x_m))$. In Section 5, we prove i.m.s. consistency.

Under certain boundedness conditions on $1/u$, the difference of $\hat{\bar{f}}^{(\nu)}$ and its expectation converges to zero a.s. and in second mean; and hence, the three properties asy. u., a.s. and m.s. consistencies of the estimator become equivalent. Sufficient and (somewhat) necessary conditions for asy. u. (and hence for a.s. or m.s. consistency) uniform on D are also given in Section 2. These, specialized to $f_j \equiv f$ and $D = (a,\infty)$ for an $a \geq -\infty$, become: If $\int_a^\infty f(x)dx < \infty$, then $\hat{\bar{f}}^{(\nu)}$ is asy. unbiased uniformly on (a,∞) iff $f^{(\nu)}$ is uniformly continuous there.

When $r > \nu$, and for all $0 < y < 1$ and for each x in D, $\bar{f}(x + hy)$, with h as indicated above, has r-th order Taylor expansion about x with integral form of the remainder, then, under certain boundedness condition of $h^{-1}\int_x^{x+h}|\bar{f}^{(r)}(t)|dt$ and of $1/u$ on D we obtain rates for various convergences.

In Section 2, we obtain rates and the exact rate for the bias term uniform on D. The result giving rates, specialized to the i.i.d. case with $r = \nu + 1$, $u \equiv 1$ and $D = R$ improves the corresponding one of Bhattacharya (1967), (see Remark 2.5).

In Section 3, we show that the error term is $0((n^{-1}\log n)^{(r-\nu)/2(1+r)})$ a.s. as $n \uparrow \infty$, uniformly on D. This result, specialized to the i.i.d. case with $r = \nu+1$, $u \equiv 1$ and $D = \dot{R}$ improves the corresponding one obtained by Schuster (1969) (see Remark 3.3).

Rates and exact rates for m.s. error uniform on D and for i.m.s. error are obtained in Section 5. These rates are shown to be $0(n^{-2(r-\nu)/(1+2r)})$ as $n \uparrow \infty$. Results, concerning bounds of m.s. and i.m.s. errors, specialized to $f_j \equiv f$, $u \equiv 1$ and $\nu = 0$ improve the corresponding ones of Parzen (1962), Schwartz (1967) and Wahba (1971), (see Remarks (5.1), (5.2) and (5.3)), though only Schwartz considered i.m.s. consistency.

In Section 6, we estimate mixed partial derivatives of the average of multivariate μ-densities. Specifically, we exhibit kernel estimators of $\bar{f}^{(\nu_1,\ldots,\nu_m)}(x) = \partial^{\nu_1+\ldots+\nu_m}\bar{f}(x)/(\Pi_1^m \partial x_i^{\nu_i})$, where $x \in R^m$, $\bar{f} = n^{-1}\Sigma_1^n f_j$ and f_j are m-variate μ-densities. These estimators have asymptotic properties analogous to those possessed by the estimators prescribed in the univariate case. We verify some of these related to asy. u., m.s. and a.s. consistencies, each with and without rates.

0.2 Sequence-Compound SELE with Applications of $\hat{\bar{f}}^{(\nu)}$.

In Chapter 2, we deal with sequence-compound SELE of certain unbounded functionals in exponential families. We use the estimators $\hat{\bar{f}}^{(\nu)}$ in order to exhibit certain sequence compound estimators whose modified regret converges to zero with certain rates.

Suppose $\mathcal{P} = \{P_\omega | \omega \in \Omega\}$ is a family of probability measures on R, and the component problem is SELE of real $\theta(\omega)$. The sequencecomp

lem

Thus

The

is

and

a

(1

(1

l

i

c

t

a

a

c

compound problem consists of n repetitions of the component problem with the loss taken to be the average of the component losses. Thus one has $\underset{\sim}{\omega} = (\omega_1, \ldots, \omega_n) \in \Omega^n$ and $(X_1, \ldots, X_n) \sim P_{\omega_1} \times \ldots \times P_{\omega_n}$. The i-th component of a (sequence compound) estimator $\underset{\sim}{\varphi} = (\varphi_1, \ldots, \varphi_n)$ of $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_n)$, where θ_j abbreviates $\theta(\omega_j)$, is allowed to depend on $(X_1, \ldots, X_i)$.

With G_n, the empiric distribution function of $\omega_1, \ldots, \omega_n$, and $R(G_n)$, the Bayes risk versus G_n in the component problem, let

$$D_n(\underset{\sim}{\omega}, \underset{\sim}{\varphi}) = n^{-1} \Sigma_1^n E(\theta_j - \varphi_j)^2 - R(G_n).$$

$D_n(\underset{\sim}{\omega}, \underset{\sim}{\varphi})$ is called the modified regret of $\underset{\sim}{\varphi}$, and is often taken as a standard for evaluating compound procedures, (e.g., Hannan (1956), (1957), Samuel (1963), (1965), Gilliland (1966), (1968), Johns (1967), and Susarla (1970); of course with varying component problems). If $\delta > 0$, and $D_n(\underset{\sim}{\omega}, \underset{\sim}{\varphi}) = O(n^{-\delta})$ as $n \to \infty$, we will say $\underset{\sim}{\varphi}$ has a rate δ (at $\underset{\sim}{\theta}$).

In the references cited in the next paragraph Ω is a bounded interval, and, except in case of Samuel, rates are uniform in $\underset{\sim}{\omega} \in \Omega^n$.

When $\theta(\omega) = e^{\omega}$ and $\mathcal{P}$ is an exponential family satisfying certain conditions, Samuel (1965) exhibits estimators $\underset{\sim}{\varphi}$ and shows that $D_n^+(\underset{\sim}{\omega}, \underset{\sim}{\varphi}) \to 0$ for each ω as $n \to \infty$. When components of $\underset{\sim}{\omega}$ are means of normal densities with variances unity, and θ is the identity, Gilliland (1966), (Chapter 3), obtains an estimator with a rate 1/5. Extending Gilliland's work to m-variate-case, Susarla (1970), (Section 1.4), exhibits, for each integer $r > 1$, estimators with rates $(r-1)/2(m+r+1)$, and thus, improves Gilliland's result.

When
com
an
est
of

wh
me

va
in
(
t
v
d

When $\mathcal{P}$ is a certain family of discrete distributions and the component problem is linear loss two-action, Johns (1967) prescribes an estimator with a rate 1/2. The same rate, 1/2, is achieved by estimators prescribed by Gilliland (1968) in sequence-compound SELE of $\underset{\sim}{\omega}$ in a certain discrete exponential families.

For our main results, $\mathcal{P}$ is an exponential family wrt μ, where μ is a σ-finite measure with density u wrt Lebesgue measure on R such that, for an $a \geq -\infty$, $u(x) > 0$ iff $x > a$.

The assumption that $u(x) > 0$ iff $x > a$ is imposed in various papers either on Empirical Bayes, or on compound problems in exponential families, (e.g., Samuel (1965), (Section 6), Yu (1970), (Chapters 1 and 2), and Johns and Van Ryzin (1972)). In the case of Gilliland (1966), (Chapter 3), and in the univariate version of Susarla (1970), (Chapter 1), u is the standard normal density function.

In each of the papers cited in the preceding paragraph, and in the paper of Hannan and Macky (1971), u is at least continuous on $\{u > 0\}$. We, instead, make certain assumptions on (local) boundedness of $1/u$. In all the papers on compound decision problems so far available in the literature, Ω is assumed to be bounded. We relax this by taking Ω as the natural parameter space. However, our assumptions restrict the speed at which $\max_{1\leq j\leq n}|\omega_j|$ grows as $n \uparrow \infty$.

We will now describe the main results of Chapter 2. We have treated only the cases $\theta(\omega) = \omega$, e^{ω} or ω^{-1}. (The cases of ω^k, $e^{\ell\omega}$ or ω^{-m}, where k and m are positive integers and $0 < \ell < \infty$, can be treated analogously). For $\alpha_i < \beta_i$ in Ω for

all
cer
de
be
ca
co

ex
of
P

se
e

W
f
o

k
r
(

e
(
(
s

all $i \geq 1$, with $\alpha_i \downarrow$ and $\beta_i \uparrow$, we exhibit estimators with certain rates. These rates are uniform in $\underset{\sim}{\omega} \in X_1^n [\alpha_i, \beta_i]$, and depend on how $\max_{1 \leq j \leq n}(|\alpha_i| \vee |\beta_i|)$ grows as $n \uparrow \infty$. Rates below are, for the sake of convenience, indicated only for the cases when α_i and β_i are constants wrt i and satisfy certain conditions.

We use the ideas of Gilliland (1966), (Chapter III), and exhibit an estimator of $\underset{\sim}{\omega}$ based on a divided difference estimator of $(\log \bar{f})^{(1)}$, where $\bar{f} = n^{-1} \Sigma_1^n f_j$ and f_j is a μ-density of P_{ω_j}. This estimator is shown, in Theorem 1, to achieve a rate 1/5.

We use the estimators $\hat{\bar{f}}^{(\nu)}$ (introduced in the preceding section) of $\bar{f}^{(\nu)}$ to obtain certain kernel estimators of $\underset{\sim}{\theta}$ when $\theta(\omega) = \omega$, e^{ω} or ω^{-1}.

For each integer $r > 1$, we exhibit kernel estimators of $\underset{\sim}{\omega}$ which are shown, in Theorem 2, to achieve a rate $(r-1)/(1+2r)$. When ω_j's are means of normal densities, our estimators are preferable, for various reasons (see Remark 4.3), to the corresponding ones of Susarla (1970), (Section 1.4).

For the case $\theta(\omega) = e^{\omega}$, we obtain, for each integer $r > 0$, kernel estimators which are shown, in Theorem 3, to have a rate $r/(1+2r)$, and thus improve (rate wise) Theorem 6 of Samuel (1965), (also, see Remark 4.5).

When $\theta(\omega) = \omega^{-1}$, we exhibit, for each integer $r > 0$, kernel estimators which are shown, in Theorem 4, to avhieve a rate $(r-\epsilon)/(1+2r)$ for any $\epsilon > 0$. The result here with $u(x) = (\Gamma(\tau))^{-1} x^{\tau-1}[x > 0]$, $\tau > 0$, generalizes and improves the main result of Section 2.1 of Susarla (1970), (see Remark 4.6).

th

2(

di

sa

s

0

v

t

v

In Theorems 5 and 6, we show that, when θ is identity and $\underset{\sim}{\omega}$ has identical components, rates with the divided difference and the kernel estimators are near, but cannot be more than, 2/5 and 2(r-1)/(1+2r), respectively.

Finally, when θ is identity, a comparison between the divided difference estimator, say $\hat{\underset{\sim}{\Psi}}$, and the kernel estimator, say $\hat{\underset{\sim}{\Psi}}_K$, is made in Section 6. $\hat{\underset{\sim}{\Psi}}_K$ with $r > 6$ is preferable to $\hat{\underset{\sim}{\Psi}}$ in the sense that $\sup_{\underset{\sim}{\omega}}|D_n(\underset{\sim}{\omega},\hat{\underset{\sim}{\Psi}}_K| \to 0$, as $n \to \infty$, faster than $\sup_{\underset{\sim}{\omega}}|D_n(\underset{\sim}{\omega},\hat{\underset{\sim}{\Psi}})| \to 0$, as $n \to \infty$.

0.3 Some Notational Conventions

We suppress the arguments of functions whenever it is convenient not to exhibit them. We denote elementary functions by their values, and, except for emphasis, do not display the dummy variables of integrations. Indicator function of a set A is denoted by A itself, or by [A]. For any measure ξ, the ξ-integral of y is denoted by ξy, $\xi(y)$ or $\xi[y]$. We abbreviate the space $L_p(R)$ to L_p, with $1 \le p \le \infty$, L_p-norm to $\|\cdot\|_p$, $g(t) - g(x)$ to $g]_x^t$ and, occasionaly, $\sup_{t\in A}|g(t)|$ to $\|g\|_A$. The symbol $\doteq$ indicates that the equation holds by the definition, or, is a defining one. The symbol ■ is used throughout to signal the end of a proof.

CHAPTER 1

NON-PARAMETRIC ESTIMATION OF DERIVATIVES OF THE AVERAGE OF n μ-DENSITIES, AND CONVERGENCE RATES IN n

1.0 Introduction.

Let μ be a σ-finite measure, dominated by Lebesgue measure on the real line R. Let $X_1,\ldots,X_n$ be independent real valued random variables with X_j distributed according to $P_j \ll \mu$. With u , a fixed determination of $d\mu/dt$, let $f_j(t) = (u(t))^{-1}\lim_{\varepsilon\downarrow 0}\varepsilon^{-1}\int_t^{t+\varepsilon} dP_j$ if the limit exists for all j and $u(t) > 0$, and 0 otherwise. (From the properties of Lebesgue points of a function, see pp. 255-256 of Natanson (1955), all of the above limits exist a.e. Moreover, if f_j' is a determination of $dP_j/d\mu$, then almost every point is a Lebesgue point of uf_j', and hence $f_j = f_j'$ a.e.) Let $f_j^{(i)}$ be the i-th order derivative of f_j. For a <u>fixed</u> $\nu \ge 0$, we want to estimate $\bar{f}^{(\nu)} = n^{-1}\Sigma_1^n f_j^{(\nu)}$.

In Section 1, we exhibit a class of kernel estimators $\hat{\bar{f}}^{(\nu)}$ of $\bar{f}^{(\nu)}$, and discuss the main assumption to be made in later sections. We obtain results on the bias in Section 2, on the error of the estimate in Section 3, and on the mean square and integrated mean square errors in Section 5. In Section 4, we prove the asymptotic normality of $(\hat{\bar{f}}^{(\nu)}(x_1),\ldots,\hat{\bar{f}}^{(\nu)}(x_m))$. In Section 6, we treat the multivariate version of the problem; specifically, for $x = (x_1,\ldots,x_m)$ in R^m, we estimate $\bar{f}^{(\nu_1,\ldots,\nu_m)}(x) = \partial^{(\nu_1+\ldots+\nu_m)}\bar{f}(x)/(\Pi_1^m \partial x_i^{\nu_i})$, where $\bar{f}$ is the average of n m-variate de

fi

1.

fu

le

(1

De

it

L

a

(1

Th

(1

Jo

sa

ex

(0

densities. Unless stated otherwise, results are obtained at a fixed point x.

1.1 Estimation of $\bar{f}^{(\nu)}$ and the Main Assumption.

Let $\mathcal{K}$ be the class of all real valued Borel-measurable functions on R vanishing off $(0,1)$. For an integer $r > \nu$, let $\mathcal{K}_\nu^r \subset \mathcal{K}$ be such that if $K \in \mathcal{K}_\nu^r$, then

$$(1.0) \qquad k_i \doteq (i!)^{-1}\int y^i K(y)\,dy = [i = \nu],\ i = 0,1,\ldots,r-1.$$

Denote $\int y^i |K(y)|dy/i!$ by $|k|_i$. The set $\mathcal{K}_\nu^r$ is non-empty, since it contains the ν-th element of the dual basis for the subspace of $L_1(0,1)$ with basis $\{1, y/1!,\ldots,y^{r-1}/(r-1)!\}$. Define $\mathcal{K}_\nu^\nu = \mathcal{K}_\nu^{\nu+1}$.

Let $0 < h \doteq h_n \le 1$ be such that $h_n \downarrow 0$ as $n \uparrow \infty$. For a fixed $r \ge \nu$ and a fixed $K \in \mathcal{K}_\nu^r$, let

$$(1.1) \qquad Y_j(\cdot) = \{K(\tfrac{X_j-\cdot}{h})/u(X_j)\}[u(X_j) > 0]\ .$$

The proposed estimator of $\bar{f}^{(\nu)}$ is

$$(1.2) \qquad \hat{\bar{f}}^{(\nu)} = (nh^{\nu+1})^{-1}\Sigma_1^n Y_j\ .$$

Hereafter we frequently abbreviate $\bar{f}^{(\nu)}$ and $\hat{\bar{f}}^{(\nu)}$ by g_n and $\hat{g}_n$, respectively.

For estimating a Lebesgue density and its derivative (in Empirical Bayes linear loss two action in exponential families) Johns and Van Ryzin (1972) introduce and use L_2-kernel functions satisfying the $\mathcal{K}_\nu^r$-conditions for $r > 1$ and $\nu = 0$ or 1, with the exception that, for the case $\nu = 1$, their kernels vanish off $(0,2)$ instead of $(0,1)$. With $f_n \equiv f$, Yu (1970), (Section 2 of his appendix), considers estimation of f and $f^{(1)}$, and uses Johns and Van Ryzin type kernels.

The orthogonality properties of $K \in \mathcal{K}_\nu^r$ and the assumption $(A_0^{(r)})$, which is introduced below only for integers $r > 0$, are used in curtailing the bias of $\hat{g}_n$, (see (1.4) - (1.7) below).

$(A_0^{(r)})$: For each $0 < y < 1$, there exists the r-th order Taylor expansion of $\bar{f}(t + hy)$ about t with integral form of the remainder:

$$\bar{f}(t + hy) = \Sigma_0^{r-1} \frac{(hy)^j}{j!} \bar{f}^{(j)}(t) + \frac{1}{(r-1)!} \int_t^{t+hy} (t+hy-z)^{r-1} \bar{f}^{(r)}(z)dz .$$

Remark 1.1. For this expansion, it is sufficient that $\bar{f}^{(r-1)}$ is continuous on $[t, t + \varepsilon]$ with $\varepsilon = hy$ and $\bar{f}^{(r)}$ exists on $(t, t + \varepsilon)$ with countably many exceptional points, and is integrable there. This follows, since by the fundamental theorem, (see Van Vleck (1973), pp. 286-7), $\bar{f}^{(r-1)}(t_2) - \bar{f}^{(r-1)}(t) = \int_t^{t_2} \bar{f}^{(r)}(t_1)dt$, $t \le t_2 \le t + \varepsilon$, and by repeated integrations of both sides, we get

$$\bar{f}(t + \varepsilon) - \Sigma_0^{r-1} \frac{\varepsilon^j}{j!} \bar{f}^{(j)}(t) = \int_t^{t+\varepsilon} \int_t^{t_r} \cdots \int_t^{t_2} \bar{f}^{(r)}(t_1)dt_1 \cdots dt_r$$
$$= \int_t^{t+\varepsilon} \frac{(t + \varepsilon - t_1)^{r-1}}{(r-1)!} \bar{f}^{(r)}(t_1)dt_1$$

where the second equality follows by Fubini theorem.

We introduce the notation

$$(1.3) \qquad \Delta_r(x) = h^{-1} \int_x^{x+h} |\bar{f}^{(r)}]_x^t| dt$$

where the dependence of Δ_r on n and h is abbreviated by omission. For some of our results, we will assume $\Delta_r(x) = o(1)$. Note that $\Delta_r(x) = o(1)$ whenever $\bar{f}^{(r)}$ is <u>asymptotically</u> <u>rt-equicontinuous</u> at x (i.e., $\bar{f}^{(r)}]_x^{x+t} \to 0$ as $t \downarrow 0$ and $n \uparrow \infty$). If only finitely many P_j are distinct and x is a rt-Lebesgue point of each of the $f_j^{(r)}$, then again $\Delta_r(x) = o(1)$.

Let B_n denote the bias of $\hat{g}_n$, i.e.,

$$(1.4) \qquad B_n = \underset{\sim}{P}_n \hat{g}_n - g_n$$

where $\underset{\sim}{P}_n = P_1 \times \ldots \times P_n$. Since X_j has Lebesgue density uf_j and the f_j's vanish wherever u vanishes, by (1.1) and (1.2)

$$(1.5) \qquad (B_n + g_n)(t) \doteq \underset{\sim}{P}_n \hat{g}_n(t) = h^{-(\nu+1)} \int K((z-t)/h)\bar{f}(z)\,dz$$

$$= h^{-\nu} \int K(y)\bar{f}(t+hy)\,dy \ .$$

If, for $\nu > 0$, $(A_0^{(\nu)})$ holds and $K \in \mathcal{K}_\nu^\nu$, then the substitution in the rhs of (1.5) of the expansion of $\bar{f}(t + hy)$ given by $(A_0^{(\nu)})$, and use of the orthogonality properties of K give

$$(1.6) \qquad (B_n + g_n)(t) = h^{-\nu} \int K(y) \int_t^{t+hy} (t + hy - z)^{\nu-1} \bar{f}^{(\nu)}(z)\,dz/(\nu-1)! \ .$$

If $(A_0^{(r)})$ for $r > \nu$ holds and $K \in \mathcal{K}_\nu^r$, then, since $k_\nu = 1$, by arguments similar to those giving (1.6),

$$(1.7) \qquad B_n(t) = h^{-\nu} \int K(y) \int_t^{t+hy} (t + hy - z)^{r-1} \bar{f}^{(r)}(z)\,dz/(r-1)! \ .$$

For $r = \nu + 1$ kernels giving (1.6) and (1.7) belong to the same class $\mathcal{K}_\nu^r$, but, since $(A_0^{(\nu+1)}) \not\Leftarrow (A_0^{(\nu)})$, the two expressions are not equivalent. We will use (1.6) and (1.7) to prove the asymptotic unbiasedness of our estimators.

In what follows $\langle a_n \rangle$ is a sequence of positive numbers, and D is a subset of R. Unless stated otherwise, all the limits, convergences and asymptotic equivalent relations (for functions depending on n) are wrt $n \to \infty$.

1.2 Asymptotic Unbiasedness and the Exact Rate

For the results of this section, $X_1, \ldots, X_n$ need not be independent. Recall that g_n and $\hat{g}_n$ stand for $\bar{f}^{(\nu)}$ and $\hat{\bar{f}}^{(\nu)}$ respectively.

We will give sufficient and (somewhat) necessary conditions for $\|B_n\|_D = o(1)$. Under different conditions we will obtain two upper bounds for $|B_n|$, and an asymptotic expression for B_n. We first prove the following, where by $(t,n) \to [0+,\infty)$ we mean t in a non-deleted rt-nbd of 0 converges to 0 and $n \to \infty$.

Theorem 1(a). Let $K \in \mathcal{K}_\nu^\nu$, and, for the case $\nu = 0$, be bounded. If $(A_0^{(\nu)})$, whenever $\nu > 0$, holds at each point in a rt-nbd of x, and if

$$(2.0) \qquad \Delta_\nu(x + t) = o(1) \quad \text{as} \quad (t,n) \to [0+,\infty)$$

then

$$(2.1) \qquad B_n(x + t) = o(1) \quad \text{as} \quad (t,n) \to [0+,\infty).$$

On the other hand, if $K \in \mathcal{K}$ (K need not be in $\mathcal{K}_\nu^\nu$) is bounded, $(g_n - g_{n'})]_x^{x+t} = o(1)$ as $(t,n),n' \to [0+,\infty),\infty$, and for a subsequence $\{m\}$, $\lim_{m\uparrow\infty, t\downarrow 0} B_m]_x^{x+t} = 0$ and $\bar{f}_m \in L_1[x, x+\tau_m]$ for some $\tau_m > h_m$, then, as $(t,n) \to [0+,\infty)$, $|g_n]_x^{x+t}| = o(1)$ (and hence, (2.0) holds).

Remark 2.1. The second part of the theorem essentially says that, in the presence of certain assumptions (which are always satisfied in the case $f_j \equiv f$ and, for some $\tau > 0$, Lebesgue-inf of the restriction to $\{t \in [x,x+\tau) | f(t) > 0\}$ of u is positive) (2.0) is necessary even for a weaker form of (2.1).

Proof. (Sufficiency of (2.0)). First consider the case $\nu = 0$. Since $K \in \mathcal{K}_0^1$, vanishes off $(0,1)$ and $\int K(y)dy = 1$, by the first equation in (1.5), we have

$$(2.2) \qquad |B_n(t)| = h^{-1}\left|\int_t^{t+h} K(\tfrac{y-t}{h})(\bar{f}]_t^y)dy\right| \le \|K\|_\infty \Delta_0(t) .$$

Thus, since K is bounded, (2.1) for the case $\nu = 0$ follows from (2.2) and (2.0).

Next consider the case $\nu \ge 1$. Since K being in $\mathcal{K}_\nu^\nu$ gives

$$(2.3) \qquad (h^{-\nu}/(\nu-1)!)\int K(y)\int_t^{t+hy}(t + hy - z)^{\nu-1}dzdy = k_\nu \doteq 1,$$

and, since, by our hypothesis, (1.6) holds at every t in $N_+(x)$, a rt-nbd of x,

$$(2.4) \qquad (\nu-1)!B_n(t) = h^{-\nu}\int K(y)\int_t^{t+hy}(t + hy - z)^{\nu-1}(g_n]_t^z)dzdy \ \forall\ t \in N_+(x).$$

Note that the integrand in (2.4) is bounded above by $|K(y)|(hy)^{\nu-1}|g_n]_t^z|$ which vanishes for $y \notin (0,1)$. Thus, by (2.4), $|B_n(t)| \le |k|_{\nu-1}\Delta_\nu(t)\ \forall\ t \in N_+(x)$, and hence, since $K \in \mathcal{K}_\nu^\nu$ implies $|k|_{\nu-1} < \infty$, (2.1) for $\nu > 0$ follows from (2.0).

Necessity of (2.0). Let m be a subsequence and $\tau_m > h_m \doteq \xi \ni \bar{f}_m \in L_1[x, x + \tau_m]$ and $\lim_{m\uparrow\infty,t\downarrow 0}(B_m]_x^{x+t}) = 0$. By (1.5), $(B_m + g_m)(\cdot) = \xi^{-\nu}\int K(y)\bar{f}_m(\cdot + \xi y)dy$. Therefore, since K vanishes off (0,1), by use of the transformation theorem

$$(2.5)\qquad |(B_m + g_m)]_x^{x+t}| \le \xi^{(\nu+1)}\|K\|_\infty \int_x^{x+\xi} |\bar{f}_m]_v^{\nu+t}|dv = o(1) \quad \text{as} \quad t \downarrow 0,$$

where, since K is bounded, $\bar{f}_m \in L_1[x, x + \tau_m]$ and $\tau_m > \xi$, the convergence in (2.5) follows by a theorem on continuity of translation of L_1-functions, (e.g., see Hewitt and Stromberg (1965), p. 199).

Since, by our hypotheses, for all sufficiently large m, $(g_n - g_m)]_x^{x+t} \to 0$ as $(t,n) \to [0+,\infty)$, and $B_m]_x^{x+t} \to 0$ as $t \downarrow 0$, the identity $g_n = (g_n - g_m) + (g_m + B_m) - B_m$ and (2.5) yield $g_n]_x^{x+t} \to 0$ as $(t,n) \to [0+,\infty)$. ∎

<u>Remark 2.2</u>. The proof of the first part of Theorem 1(a) also proves that: If $(A_0^{(\nu)})$, whenever $\nu > 0$, holds on D and if $K \in \mathcal{K}_\nu^\nu$, then

$$(2.6)\qquad \|B_n\|_D\|\Delta_\nu\|_D^{-1} \le \|K\|_\infty \quad \text{or} \quad |k|_{\nu-1} \quad \text{according as} \quad \nu = 0 \quad \text{or} \quad > 0.$$

Thus, if (2.6) holds and rhs of this is finite, then $\|\Delta_\nu\|_D = o(1)$ implies

$$(2.7)\qquad \|B_n\|_D = o(1).$$

In fact, $\|\Delta_\nu\|_D = o(1)$ is somewhat a necessary condition for (2.7): If $K \in \mathcal{K}$ (K need not be in $\mathcal{K}_\nu^\nu$) is bounded, $\sup_{x\in D}|(g_n - g_{n'})]_x^{x+t}| = o(1)$ as $(t,n),n' \to [0+,\infty),\infty$, and for a subsequence $\{m\}$, $\lim_{m\uparrow\infty, t\downarrow 0}\sup_{x\in D}|B_m]_x^{x+t}| = 0$ and $\bar{f}_m \in L_1(\cup_{x\in D}[x, x + \tau_m))$ for some $\tau_m > h_m$, then, as $(t,n) \to [0+,\infty)$, $\sup_{x\in D}|g_n]_x^{x+t}| = o(1)$ (and hence, $\|\Delta_\nu\|_D = o(1)$). Proof of this assertion follows from arguments identical to those given for that of the second part of Theorem 1(a). As an immediate corollary to this last result, we have

Corollary 1. Suppose $K \in \mathcal{K}_\nu^\nu$ is bounded and only finitely many P_j are distinct. For an $a \geq -\infty$, if each $\int_a^\infty f_j < \infty$, then $\sup_{x>a}|B_n(x)| = o(1)$ iff $\lim_{t\to 0}\sup_{x>a}|g_n]_x^{x+t}| = o(1)$.

Remark 2.3. If only finitely many P_j are distinct, then $\Delta_\nu(x) = o(1)$ whenever x is a rt-Lebesgue point of each of the $f_j^{(\nu)}$. Thus, (from the first part of Remark 2.2) with $f_j \equiv f$ our estimator of $f^{(\nu)}$ is asymptotically unbiased at x under the weaker assumption than that of the continuity of $f^{(\nu)}$ at x imposed for similar results in almost all papers on the subject. Sufficiency and necessity parts of Corollary 1 specialized to the i.i.d. case with $u \equiv 1$, $\nu = 0$ and $a = -\infty$ have been proved, respectively, by Nadarya (1964) and Schuster (1969) for their kernel estimators.

Remark 2.4. For the case $\nu = 0$ and $u \equiv 1$, (2.7), (with different kernels), has been noted by Samuel (1965), (Section 6), under the uniform equicontinuity (and necessarily uniform equiboundedness) of $f_1, f_2, \ldots$ on D.

Theorem 1(b). If, for $r > \nu$, $(A_0^{(r)})$ holds and $K \in \mathcal{K}_\nu^r$, then

$$h^{-r+\nu+1}|B_n| \leq |k|_{r-1} \int_x^{x+h} |\bar{f}^{(r)}|, \tag{2.8}$$

and

$$|h^{-r+\nu} B_n - k_r \bar{f}^{(r)}| \leq |k|_{r-1}\, \Delta_r . \tag{2.9}$$

Proof. Inequality (2.8) follows immediately from (1.7), since the absolute value of the rhs there at $t = x$ is no more than $((r-1)!)^{-1}\int y^{r-1}|K(y)|dy(\doteq |k|_{r-1})$ times $h^{r-\nu-1}\int_x^{x+h}|\bar{f}^{(r)}|$.

Also, since

$$(2.10)\quad ((r-1)!h^r)^{-1}\int K(y)\int_t^{t+hy}(t+hy-z)^{r-1}dz\,dy = (r!)^{-1}\int y^r K(y)dy \doteq k_r,$$

from (1.7), the lhs in (2.9) at t is exceeded by

$$(2.11)\quad ((r-1)!h^r)^{-1}\int |K(y)|\int_t^{t+hy}(t+hy-z)^{r-1}|\bar{f}^{(r)}]_t^z|dzdy \le |k|_{r-1}\Delta_r . \blacksquare$$

<u>Remark 2.5</u>. If only finitely many P_j are distinct, then, from the first part of theorem 1(b), the existence and the boundedness of each of the $f_j^{(r)}$ on $\cup_{x\in D}[x,x+h)$ ensure that

$$(2.12)\qquad ||B_n||_D = O(h^{r-\nu}).$$

(2.12), for the case $f_j \equiv f$, $u \equiv 1$, $D = R$ and $r = \nu + 1$, is proved by Bhattacharya (1967) for his kernel estimators under the stronger assumptions that f and its first $\nu + 1$ derivatives are bounded.

As an immediate consequence of (2.9), we have

<u>Corollary 2</u>. Let (2.9) hold uniformly on D. If $k_r \neq 0$, $||\Delta_r||_D = o(1)$ and $\lim\inf(\inf_{t\in D}|\bar{f}^{(r)}(t)|) > 0$, then

$$(2.13)\qquad h^{-(r-\nu)}B_n(t) \sim k_r\bar{f}^{(r)}(t) \text{ uniformly on D.}$$

Thus, under certain conditions, the exact rate of convergence for the bias of the estimator $\hat{g}_n$ is $h^{r-\nu}$. Theorem 1(b) describes the situations where such rate is indeed achieved by $\hat{g}_n$.

Some global properties of B_n will be obtained in Section 5. Under varying conditions, we will show that, for a fixed $a \geq -\infty$, $\int_a^\infty B_n^2 dt = o(1)$, $\int_a^\infty B_n^2 dt \leq |k|_r^2 \, h^{2(r-\nu)} \int_a^\infty |\bar{f}^{(r)}|^2 dt$ and $\int_a^\infty B_n^2 dt \sim k_r^2 h^{2(r-\nu)} \int_a^\infty |\bar{f}^{(r)}|^2 dt$.

1.3 Strong Consistency with Rates

Let E_n denote the error of the estimator $\hat{g}_n$, that is, $E_n = \hat{g}_n - g_n$, where g_n and $\hat{g}_n$ denote respectively, $\bar{f}^{(\nu)}$ and $\hat{\bar{f}}^{(\nu)}$. Unless stated otherwise, all convergences in this section will be meant with probability one. We will give sufficient and (somewhat) necessary conditions for $\|E_n\|_D = o(1)$, and prove, for $r > \nu$,

$$(3.0) \qquad \|h^{-r+\nu} E_n - k_r \bar{f}^{(r)}\|_D = o(1) \ .$$

Under conditions weaker than those used for (3.0), we will show that $\|E_n\|_D = O((n^{-1}\log n)^\alpha)$ for $2\alpha = (r-\nu)/(1+r)$.

Hereafter denote $\hat{g}_n - \underset{\sim}{P}_n \hat{g}_n$ by C_n. In view of

$$(3.1) \qquad E_n = C_n + B_n ,$$

if $\|C_n\|_D = o(1)$, then sufficient and (somewhat) necessary conditions for $\|E_n\|_D = o(1)$ can be obtained from Section 2 (Remark 2.2 and Corollary 1). Similarly, regarding rates of convergence, if $a_n\|C_n\|_D = o(1)$, then sufficient conditions for $a_n\|E_n\|_D = o(1)$ and for (3.0) (with $h^{-r+\nu} = a_n$) can be obtained from (2.8) and (2.9), respectively. Thus our objective in this section will be to obtain sufficient conditions under which $a_n\|C_n\|_D = o(1)$.

For the remainder of this chapter, let

$$(3.2)\quad u_h(\cdot) = \text{Leb-inf restriction to } \{t\in[x,x+h) \mid \vee_1^n f_j(t)>0\} \text{ of } u.$$

For the results in Theorems 2(a) and 2(b) below, K need not be in $\mathcal{K}_\nu^r$ (but $K \in \mathcal{K}$). First consider the case when $D = \{x\}$.

<u>Theorem 2(a)</u>. Let $\|K\|_\infty < \infty . \forall\ \eta > 0$

$$(3.3)\quad \underset{\sim}{P}_n[|C_n| \geq \eta] \leq 2\exp\{-\tfrac{n}{2}(h^{\nu+1}u_h\eta/\|K\|_\infty)^2\}\ .$$

<u>Proof</u>. By (1.2), the event on the lhs of (3.3) is $[|n^{-1}\Sigma_1^n(Y_j - P_jY_j)| > \eta h^{\nu+1}]$, and by (1.1) and (3.2), $|Y_j| \leq \|K\|_\infty/u_h$ a.s. for $1 \leq j \leq n$. Hence, since $Y_1,\ldots,Y_n$ are independent, Theorem 2 of Hoeffding (1963), applied to random variables Y_j and $-Y_j$ here, completes the proof. ■

Clearly, when D contains finite, m, points, $\underset{\sim}{P}_n[\|C_n\|_D > \eta] \leq m$ times the rhs of (3.3) with u_h there replaced by $\min_{t\in D}u_h(t)$. We now consider the case when D is not finite.

<u>Theorem 2(b)</u>. Let K on (0,1), and, for each t in D, 1/u on [t, t + h) be of bounded variations. Then, with $Y_\cdot(t) = K((\cdot - t)/h)[u(\cdot) > 0]/u(\cdot)$ (we may understand that by Y_j we are abbreviating Y_{X_j}), $\forall\ \eta > 0$,

$$(3.4)\quad \underset{\sim}{P}_n[\|C_n\|_D \geq \eta] \leq 4n^{\frac{1}{2}}M\exp(-2(M^2 - 1)^+),$$

where $M = n^{\frac{1}{2}}h^{\nu+1}\eta/(\sup_{t\in D}\int_t^{t+h}|dY_\cdot(t)|)$.

<u>Remark 3.1</u>. Kernel functions $K \in \mathcal{K}_\nu^r$, which are of bounded variations always exist, e.g., take those K's in $\mathcal{K}_\nu^r$ which are polynomials on (0,1).

Remark 3.2. Since $Y_{\cdot}(t)$, as a function of t, is of bounded variation, $Y_{\cdot}(t+)$ and $Y_{\cdot}(t-)$ exist for all t. Therefore, for any countable set S dense in D, $\sup_{t\in D} Y_{\cdot}(t) = \sup_{t\in S} Y_{\cdot}(t)$. Consequently, $\sup_{t\in D} Y_{X_j}(t)$ is a random variable. Similarly, $\|\hat{C}_n\|_D$ $(= \|(nh^{\nu+1})^{-1}\Sigma_1^n(Y_{X_j} - P_j Y_{X_j}\|_D$ by (1.1) and (1.2)) is a random variable, and the lhs of (3.4) is meaningful.

Proof. Fix t in D until stated otherwise. Let $\bar{F}$ be the average of distribution functions of $X_1, \ldots, X_n$, and let $2F^*(\cdot) = n^{-1}\Sigma_1^n([X_j < \cdot] + [X_j \le \cdot])$. Since Lebesgue-Steiltjes integral $\int \cdot dG$ does not depend on how G (monotone) is defined at points of discontinuity, from (1.1) and (1.2),

$$h^{(\nu+1)}C_n(t) = \int_t^{t+h} Y_{\cdot}\, d(F^* - \bar{F})(\cdot). \tag{3.5}$$

Since $Y_{\cdot}$ is of bounded variation on $[t, t+h)$, it is continuous there except on a countable subset C. But by the absolute continuity of $\bar{F}$, $\underset{\sim}{P}_n \int_C dF^* = \int_C d\bar{F} = 0$ which implies $\int_C dF^* = 0$ a.s. Consequently, (3.5) can be written as

$$2h^{(\nu+1)}C_n(t) = \int_t^{t+h} (Y_{\cdot+} + Y_{\cdot-})\, d(F^* - \bar{F})(\cdot) \quad \text{a.s.} \tag{3.5$'$}$$

Since $2(F^* - \bar{F})(\cdot) = (F^* - \bar{F})(\cdot+) + (F^* - \bar{F})(\cdot-)$, $K(y) = 0$ $\forall\, y \notin (0,1)$, and $Y_{\cdot}$ is of bounded variation (and hence is the difference of two increasing functions) on $[t, t+h)$, by (3.5)' here and (V) of Theorem 21.67 of Hewitt and Stromberg (1965), the rhs of (3.5)' is $2\int_t^{t+h} (F^* - \bar{F})(\cdot)\, d(Y_{\cdot})$. Hence, since our foregoing analysis in the proof holds good for each $t \in D$,

$$h^{\nu+1}\|C_n\|_D \le \|F^* - \bar{F}\|_\infty \sup_{t\in D} \int_t^{t+h} |dY_{\cdot}(t)|\ . \tag{3.6}$$

Now (3.4) follows from (3.6) here combined with Lemma A.1 and Remark A.1 with $c_1 = \ldots = c_n = n^{-\frac{1}{2}}$ of the appendix. ■

Let $v_h(t)$ be the total variation of $1/u$ on $[t,t+h)$, and $V(K)$ be that of K on $(0,1)$. Then $\forall\, y$ in $[t,t+h)$, $(u(y))^{-1} \leq (u_h(t))^{-1} + v_h(t)$ and $|K((y-t)/h)| \leq |K(0)| + V(K) = V(K)$; and the total variation of $Y_{\cdot}(t)$ on $[t,t+h)$ is no more than $v_h(t)\|K\|_\infty + V(K)\sup_{t\leq\cdot<t+h}(u(\cdot))^{-1}$. Consequently, since $\|K\|_\infty \leq V(K)$,

$$\text{(3.7)} \qquad \sup_{t\in D}\int_t^{t+h}|dY_{\cdot}(t)| \leq \|(u_h)^{-1} + 2v_h\|_D V(K).$$

Now we prove the following corollary to Theorems 2(a) and 2(b).

<u>Corollary 3</u>. If $\|K\|_\infty < \infty$, then

$$\text{(3.8)} \qquad u_h C_n = O((\log n)^{\frac{1}{2}}/(n^{\frac{1}{2}}h^{\nu+1}));$$

and if (3.4) and (3.7) hold with $V(K) < \infty$, then

$$\text{(3.9)} \qquad \|(u_h)^{-1} + 2v_h\|_D^{-1}\|C_n\|_D = \text{rhs of (3.8)} .$$

<u>Proof</u>. By (3.3) with $\eta = 2(\log n)^{\frac{1}{2}}\|K\|_\infty/(n^{\frac{1}{2}}h^{\nu+1}u_h)$, $\Sigma_1^\infty\, \underset{\sim}{P}_n[C_n > \eta] < \infty$. Thus (3.8) follows by Borel-Cantelli Lemma. Similarly, (3.9) follows from (3.4) and (3.7). ■

<u>Remark 3.3</u>. Suppose for $r > \nu$, $(A_0^{(r)})$ holds at each point in D, $K \in \mathcal{K}_\nu^r$ and $h^{-1}\sup_{t\in D}\int_t^{t+h}|\bar{f}^{(r)}|dt = O(1)$, then by (2.8) of Theorem 1(b),

$$\text{(3.10)} \qquad \|B_n\|_D = O(h^{r-\nu}).$$

The choice of h that balances rhs's of (3.8)-(3.10) is proportional to $\{n^{-1}(1 + \log n)\}^{1/2(r+1)}$. Thus with this h, if, for some n, $u_h > 0$ for each point in D (and $\|(u_h)^{-1} + 2v_h\|_D < \infty$, in case D is not finite), then (3.1) combined with (3-8)-(3.10) gives, with $2\alpha = (r-\nu)/(1+r)$,

$$\|E_n\|_D = O((n^{-1}\log n)^{\alpha}). \tag{3.11}$$

The result in (3.11) specialized to the case $u \equiv 1$, $f_n \equiv f$, $r = \nu+1$ and $D = R$, is proved by Schuster (1969) (for his estimators) under stronger assumptions that f and its first $\nu+1$ derivatives are bounded.

If only finitely many P_j are distinct, then (3.11) can be strengthened slightly by replacing $\log n$ there by $\log\log n$. (This follows from (3.1), (3.6) and (3.10), since $\|F^* - \bar{F}\|_\infty = O((n^{-1} \log\log n)^{\frac{1}{2}})$, see Kiefer (1961)).

1.4 Variance, Covariance and Asymptotic Normality.

In this section we prove the asymptotic normality of $(\hat{g}_n(x_1),\ldots,\hat{g}_n(x_m))$ where, as before, g_n and $\hat{g}_n$ abbreviate $\bar{f}^{(\nu)}$ and $\hat{f}^{(\nu)}$. We first obtain an upper bound for $\sigma_n^2 \doteq \operatorname{var}\hat{g}_n$ and show that $(nh^{2\nu+1})\sigma_n^2 \sim \|K\|_2^2(\bar{f}/u)$, and for $x' \neq x$, $\sigma_n(x,x') \doteq \operatorname{cov}(\hat{g}_n(x),\hat{g}_n(x')) = o((nh^{2\nu+1})^{-1})$. Throughout this section, we assume that $\|K\|_\infty < \infty$, and, unless stated otherwise, summation Σ is over $1,\ldots,n$.

Recall from (1.1) and (1.2) that $Y_j(\cdot) = \{K((x_j-\cdot)/h)/u(x_j)\}[u(x_j) > 0]$ and $\hat{g}_n = (nh^{\nu+1})^{-1}\Sigma Y_j$. Since $X_1,\ldots,X_n$ are independent, so are $Y_1,\ldots,Y_n$, it follows that

$$(nh^{\nu+1})^2\sigma_n^2 = \Sigma \operatorname{var}(Y_j) \le \Sigma P_j Y_j^2 . \tag{4.0}$$

Lemma 1. $\forall\ \xi \geq 1$,

(4.1) $$n^{-1}\Sigma\ P_j|Y_j(t) - P_jY_j(t)|^{\xi} \leq (2\|K\|_{\infty})^{\xi}\int_t^{t+h}(\bar{f}/u^{\xi-1}).$$

Proof. By c_r-equality (Loève (1963), p. 155), the lhs of (4.1) is exceeded by $2^{\xi}n^{-1}\Sigma\ P_j|Y_j(t)|^{\xi} = 2^{\xi}\int|K(y-t)/h)|^{\xi}(\bar{f}/u^{\xi-1})dy$ which is bounded above by the rhs of (4.1), since K vanishes off $(0,1)$. ∎

Inequality (4.0) and the latter part of the arguments used in the preceding proof with $\xi = 2$ yield

(4.2) $$\sigma_n^2(x) \leq \|K\|_{\infty}^2(nh^{2\nu+2})^{-1}\int_x^{x+h}(\bar{f}/u).$$

Remark 4.1. If $u \equiv 1$ a.e. on $\{t|f_j(t) > 0 \text{ for some } j \geq 1\}$, then (4.2) is strengthened to $nh^{2\nu+2}\|\sigma_n^2\|_{\infty} \leq \|K\|_{\infty}^2$, since then $\int_t^{t+h}\bar{f} \leq 1\ \forall\ t \in R$.

Lemma 2. If

(A_1): $$h^{-1}\int_x^{x+h}|\frac{\bar{f}}{u}(y) - \frac{\bar{f}}{u}(x)|dy = o(1)$$

then

(4.3) $$(nh)^{-1}\Sigma\ P_jY_j^2 = \|K\|_2^2(\bar{f}/u) + o(1).$$

Remark 4.2. (A_1) is implied if x is a rt-Lebesgue point of $(u)^{-1}$, $\bar{f}(x)$ is bounded in n, $\Delta_0(x) = o(1)$, and either $\sup_{x\leq t<x+h}(u(t))^{-1}$ or $\sup_{x\leq t<x+h}\bar{f}(t)$ is bounded in n. Obviously, (A_1) holds if $(\bar{f}/u)$, as a sequence in n, is asymptotically rt-equicontinuous at x.

Proof of Lemma 2. Since the lhs of (4.3) is $h^{-1}\int K^2((y-x)/h)(\bar{f}/u)dy$, and $\|K\|_2^2 \doteq \int K^2 = h^{-1}\int K^2((y-x)/h)dy$, to prove (4.3) it suffices to show that

(4.4) $$h^{-1}\int K^2(\frac{y-x}{h})|\frac{\bar{f}}{u}(y) - \frac{\bar{f}}{u}(x)|dy = o(1).$$

But since K is bounded and vanishes off $(0,1)$, (4.4) follows from (A_1). ■

Lemma 3. If for an integer $i \in [0,\nu]$,

$$(A_2^{(i)}): h^{2i-1}n^{-1}\Sigma_{j=1}^{n}\{(\int_x^{x+h}|f_j^{(i)}|dy)(\int_{x'}^{x'+h}|f_j^{(i)}|dy)\} = o(1),$$

and in case i is positive,

$(A_0^{(i)})^+$: for each $j \geq 1$, $t = x,x'$ and $0 < y < 1$,

$$f_j(t+hy) = \Sigma_0^{i-1}\frac{(hy)^{\ell}}{\ell!}f_j^{(\ell)}(t) + \frac{1}{(i-1)!}\int_t^{t+hy}(t+hy-z)^{i-1}f_j^{(i)}(z)dz$$

holds, then

(4.5) $$(nh)^{-1}\Sigma\{(P_jY_j(x))(P_jY_j(x'))\} = o(1).$$

Proof. Since K is bounded and vanishes off $(0,1)$, the lhs of (4.5) is bounded above, in its absolute value, by $||K||_\infty^2(nh)^{-1}\Sigma\{(\int_x^{x+h}f_j)(\int_{x'}^{x'+h}f_j)\} = o(1)$ by $(A_2^{(0)})$.

Now consider the case $i \geq 1$. By the transformation theorem, $(A_0^{(i)})^+$ and the orthogonality properties of K, the lhs of (4.5) is

(4.6) $$hn^{-1}\Sigma\{(\int K(y)f_j(x+hy)dy)(\int K(y)f_j(x'+hy)dy) = hn^{-1}\Sigma\gamma_{nj}(x)\gamma_{nj}(x'),$$

where $(i-1)!\gamma_{nj}(t) = \int K(y)\int_t^{t+hy}(t+hy-z)^{i-1}f_j^{(i)}(z)dzdy$. Since K vanishes off $(0,1)$, $(i-1)!|\gamma_{nj}(t)| \leq ||K||_\infty h^{i-1}\int_t^{t+h}|f_j^{(i)}|$. Consequently, since $||K||_\infty < \infty$, the rhs of (4.6) is $o(1)$ by $(A_2^{(i)})$. ■

Corollary 4. If (4.5) holds, then with $h \le |x-x'|$

$$nh^{2\nu+1}\sigma_n(x,x') = o(1); \tag{4.7}$$

and if (4.3) and (4.5) with $x' = x$ hold, then

$$nh^{2\nu+1}\sigma_n^2 - \|K\|_2^2(\bar{f}/u) = o(1). \tag{4.8}$$

Proof. Since K vanishes off $(0,1)$, with $h \le |x-x'|$, $Y_j(x)Y_j(x') \equiv 0$. Hence, by independence of $Y_1,\ldots,Y_n$ and by (1.2), $(nh^{2\nu+1})^{-1}\sigma_n(x,x')$ is minus the lhs of (4.5).

Again by independence of $Y_1,\ldots,Y_n$ and by (1.2), $nh^{2\nu+1}\sigma_n^2 = (nh)^{-1}\Sigma(P_jY_j^2 - P_j^2Y_j)$. Therefore, (4.8) follows from (4.3) and (4.5) with $x' = x$. ■

As an immediate consequence of the preceding corollary, if $\liminf \bar{f} > 0$ and if (4.8) holds, then

$$nh^{2\nu+1}\sigma_n^2 \sim \|K\|_2^2\bar{f}/u . \tag{4.9}$$

Remark 4.3. If (A_1) holds (which is assumed indirectly for (4.9)), then the simple result in (4.2) gives a rate for σ_n^2 equal to the exact rate obtained in (4.9).

Theorem 3. With $x_1,\ldots,x_m$ in R, suppose (4.7) for pairs (x_i,x_j), $i \ne j$, $i,j = 1,\ldots,m$, and the hypotheses for (4.9) for each x_i, $i = 1,\ldots,m$, hold. If for each $t = x_1,\ldots,x_m$

$$(\bar{f}(t)/u(t))^{-3/2}h^{-1}\int_t^{t+h}(\bar{f}/u^2) = o((nh)^{\frac{1}{2}}), \tag{4.10}$$

then

$$\underset{\sim}{Z} \doteq (Z_n(x_1),\ldots,Z_n(x_m)) \xrightarrow{\mathcal{L}} \mathcal{N}(\underset{\sim}{0},\underset{\sim}{I}) \tag{4.11}$$

where $\sigma_n Z_n = C_n \doteq (\hat{g}_n - \underset{\sim}{P}_n \hat{g}_n)$, $\underset{\sim}{I}$ is a $m \times m$ identity matrix and $\underset{\sim}{0} = (0,\ldots,0)$ is a $1 \times m$ matrix.

__Proof__. Let $Y_j^* = (nh^{\nu+1})^{-1}(Y_j - P_j Y_j)/\sigma_n$. Then $Z_n = \Sigma Y_j^*$. Let $\underset{\sim}{0} \neq \underset{\sim}{c} \in R^m$ with coordinates $c_1,\ldots,c_m$, and let $L_j = \Sigma_1^m c_i Y_j^*(x_i)$. Then, since $\hat{g}_n = (nh^{\nu+1})^{-1}\Sigma Y_j$, $\underset{\sim}{c}\,\underset{\sim}{Z} \doteq \Sigma_1^m c_i Z_n(x_i) = \Sigma L_j$ is the sum of n independent random variables L_j centered at expectations. Therefore, by Berry-Esseen theorem (see Loève (1963), p. 288), with $\eta_n^2 = \mathrm{var}(\underset{\sim}{c}\,\underset{\sim}{Z})$,

$$(4.12) \qquad |P[(\underset{\sim}{c}\,\underset{\sim}{Z} \leq \xi\, \eta_n] - \Phi(\xi)| \leq C\, \Sigma\, P_j |L_j|^3/\eta_n^3,$$

where C is the Berry-Esseen constant.

Recall that $\sigma_n^2 = \mathrm{var}(\hat{g}_n)$ and $\sigma_n(x_i,x_j) = \mathrm{cov}(\hat{g}_n(x_i),\hat{g}_n(x_j))$. Therefore

$$(4.13) \qquad \eta_n^2 \doteq \mathrm{var}(\Sigma_1^m c_i Z_n(x_i)) = \underset{\sim}{c}\,\underset{\sim}{c} + \Sigma_{\substack{i,j=1\\ i\neq j}}^m (c_i c_j \sigma_n(x_i,x_j)/\sigma_n(x_i)\sigma_n(x_j))$$
$$= \underset{\sim}{c}\,\underset{\sim}{c} + o(1)$$

where the second step follows from (4.7) and the hypotheses of (4.9). Moreover, (4.1) with $\xi = 3$ and (4.9) applied at the second step below yield, for each $t = x_1,\ldots,x_m$,

$$(4.14) \qquad \Sigma\, P_j |Y_j^*(t)|^3 = (nh^{\nu+1})^{-3}\Sigma\, P_j |Y_j(t) - P_j Y_j(t)|^3/\sigma_n^3(t)$$
$$\leq \sim (nh)^{-\frac{1}{2}}(2\|K\|_\infty/\|K\|_2)^3 \text{ (lhs in (4.10))} = o(1) \text{ by (4.10)}.$$

Since, by m-1 uses of c_r-inequality, $|L_j|^3 \leq 2^{2(m-1)}\Sigma_1^m |c_i|^3 |Y_j^*(x_i)|^3$, by (4.13) and (4.14), the rhs in (4.12) is $o(1)$. Consequently, since by (4.13) $\eta_n^2 \to \underset{\sim}{c}\,\underset{\sim}{c}$, (4.15) $$\underset{\sim}{c}\,\underset{\sim}{Z} \overset{\mathcal{L}}{\to} \mathcal{N}(0,\underset{\sim}{c}\,\underset{\sim}{c}) \ .$$

Since (4.15) holds for every $\underset{\sim}{c} \neq \underset{\sim}{0}$ in R^m, by a well known theorem, (see Billingsley (1968), p. 49), our desired conclusion follows.■

<u>Remark 4.4</u>. From (4.12)-(4.14) with $m = 1$, $c_1 = 1$ and $x_1 = x$, a rate of closeness of the distribution of $Z_n = C_n/\sigma_n$ to that of $\mathcal{N}(0,1)$ is given by

$$\|P[Z_n(x) \le \cdot] - \Phi(\cdot)\|_\infty \lesssim O((nh)^{-\frac{1}{2}}(\text{lhs of (4.10) with } t = x)) .$$

<u>Remark 4.5</u>. Inequality (4.10), with t as indicated there, is used in the proof of Theorem 3 only to show that $\Sigma\, P_j|Y_j^*(t)|^3 = o(1)$, (cf. (4.14)). If u_h is given by (3.2), then $|Y_j(t)| \le \|K\|_\infty/u_h(t)$. Therefore, the definitions of Y_j^* given in the preceding proof, and of $\sigma_n^2(t)$ followed by the fact that $P_j|Y_j^*|^3 \le 2\|K\|_\infty(nh^{\nu+1}u_h\sigma_n)^{-1}P_j|Y_j^*|^2$, lead to

$$\Sigma\, P_j|Y_j^*(t)|^3 \le 2\|K\|_\infty(nh^{\nu+1}u_h(t)\sigma_n(t))^{-1} \sim 2\|K\|_\infty(\|K\|_2^2 nh\bar{f}(t)u_h^2(t)/u(t))^{-\frac{1}{2}}$$

where the last relation follows from (4.9) which is assumed in Theorem 3. Hence (4.10) in Theorem 3 can be replaced by $((\bar{f}(t)u_h^2(t)/u(t))^{-1} = o(nh)$. In case $h^{-1}\int_t^{t+h}(\bar{f}/u^2) \sim \bar{f}(t)/u^2(t)$ at $t = x_1,\ldots,x_m$, then obviously (4.10) can be replaced by $(u(t)\bar{f}(t))^{-1} = o(nh)$.

<u>Remark 4.6</u>. If $\lim \bar{f}(x_i) = \alpha_i > 0$ and (4.8) holds at $x_1,\ldots,x_m$, then (4.11) implies

(4.16) $$\lim \underset{\sim}{P}_n[(nh^{2\nu+1})^{\frac{1}{2}}C_n(x_i) \le t_i,\ 1 \le i \le m] = \prod_1^m \Phi\left(\frac{t_i}{\|K\|_2(\alpha_i/u(x_i))^{\frac{1}{2}}}\right) .$$

For the i.i.d. case with $\nu = 0$ and $u \equiv 1$, (4.16) is quite similar to the univariate version of the result obtained in Theorem 3.5 of Cacoullos (1966).

In view of the identity $E_n = B_n + C_n$ and (4.16), certain interesting results about the asymptotic distribution of $(E_n(x_1),\ldots,E_n(x_m))$ can easily be obtained from (2.8) and (2.9) both at $x_1,\ldots,x_m$.

1.5 Mean Square and Integrated Mean Square Consistencies with the Exact Rates.

Define the mean square error (MSE_n) and the integrated mean square error ($IMSE_n$) of the estimators $\hat{g}_n$ by

(5.0) $$MSE_n = \underset{\sim}{P}_n(\hat{g}_n - g_n)^2 \quad \text{and} \quad IMSE_n = \int_a^\infty MSE_n dt$$

respectively, where $a \geq -\infty$ is fixed. Obviously,

(5.1) $$MSE_n = B_n^2 + \sigma_n^2 .$$

This section is divided into two parts. The first one deals with properties of MSE_n and the other deals with those of $IMSE_n$. We obtain, among other results, rates and the exact rates for MSE_n and $IMSE_n$.

In view of (5.1), various results concerning MSE_n can be obtained from those of B_n and σ_n contained in Sections 2 and 4 respectively. We describe some of them as follows. By (4.2), if $(nh^{2\nu+2})^{-1}\sup_{t\in D}\int_t^{t+h}(\bar{f}/u)\,dy = o(1)$, then sufficient and (somewhat) necessary conditions for $\|MSE_n\|_D = o(1)$ can be obtained from Remark 2.2 and Corollary 1. Regarding rates of convergence, we have for $r > \nu$, Theorem 4. If (2.8) holds, then

$$\text{(5.2)} \quad MSE_n \le (|k|_{r-1} h^{r-\nu-1} \int_x^{x+h} |\bar{f}^{(r)}|)^2 + \|K\|_\infty^2 (nh^{2\nu+2})^{-1} \int_x^{x+h} (\bar{f}/u);$$

and if (2.13) with $D = \{x\}$, and (4.9) hold , then

$$\text{(5.3)} \qquad MSE_n \sim (k_r h^{r-\nu} \bar{f}^{(r)})^2 + (nh^{2\nu+1})^{-1} \|K\|_2^2 (\bar{f}/u).$$

Proof. Inequalities (2.8) and (4.2) combined with (5.1) yield (5.2). Since $a_n \sim b_n > 0$ and $c_n \sim d_n > 0$ imply $a_n + c_n \sim b_n + d_n$, (5.3) is an immediate consequence of its hypothesis. ■

Remark 5.1. Suppose K is bounded. If for some $0 < p$, $q \le 1$, $\sup_{t\in D}\{(\int_t^{t+h} |\bar{f}^{(r)}|^{1/p}) \vee (\int_t^{t+h} (\bar{f}/u)^{1/q}\}$ is bounded in n, then (5.2) followed by use of Hölder inequality gives

$$\text{(5.4)} \qquad \|MSE_n\|_D = O(h^{2(r-\nu-p)} + (nh^{2\nu+1+q})^{-1}) = O(n^{-2s(r-\nu-p)})$$

where $s^{-1} = 2r + 1 - 2p + q$, and the second equation follows by taking h proportional to n^{-s}, a choice of h balancing the two terms in the middle of (5.4).

The result in (5.4) specialized to $\nu = 0$, $f_n \equiv f$, $u \equiv 1$ and $D = R$ improves the corresponding result obtained in Theorem 2 of Schwartz (1967). Assuming f is continuous, of bounded variation and $x^j f^{(r-j)} \in L_2$ for each $j = 0,1,\ldots,r$, he exhibits an estimator of f by orthogonal series method, and shows that MSE_n of his estimator is $O(n^{-(r-2)/r})$ uniformly on R. This rate is much slower (especially when r is not large) than $O(n^{-(2r-1)/(2r+1)})$ obtained in (5.4) with $p = \frac{1}{2}$, $q = 1$ and $\nu = 0$, which is guaranteed in this case simply by the assumption that $\sup_{t\in R} \int_t^{t+\epsilon} |f^{(r)}|^2 < \infty$ for some $\epsilon > 0$. Moreover, he requires $r > 2$ instead of $r > 0$.

Remark 5.2. If $\sup_{t\in D} h^{-1}\{(\int_t^{t+h}|\bar{f}^{(r)}|) \vee (\int_t^{t+h}(\bar{f}/u))\}$ is bounded in n, then taking h proportional to $n^{-1/(1+2r)}$ in (5.2), we get

$$(5.4)' \qquad \|MSE_n\|_D = O(n^{-2(r-\nu)/(1+2r)})$$

improving the rate in (5.4) (with the excess in the rate of the order n^{-c} where $c = 2s\{(r-\nu)q + (4\nu+1)p\}/(1+2r))$. For the case $D = \{x\}$, $f_n \equiv f$ and $\nu = 0,1$, Yu (1970), (Section 2 of his appendix); and for the case $D = \{x\}$, $f_n \equiv f$, $u \equiv 1$ and $\nu = 0$, Parzen (1962), (Section 4), and Wahba (1971), (Theorem 2), obtain the rate in (5.4)' for their estimators. Yu makes a little stronger assumption that $f^{(r)}$ and f/u are bounded on $[x,x+h]$; while Parzen and Wahba make still stronger assumption (adding others) that, respectively, $f^{(r)}$ is continuous and is in L_2.

An optimal h, in the sense of minimizing the asymptotic expression for MSE_n in (5.3) is given by

$$(5.5) \qquad h^{1+2r} = n^{-1}(2\nu+1)\|K\|_2^2\bar{f}/(2(r-\nu)k_r^2(\bar{f}^{(r)})^2u) \ .$$

Thus approximations of the optimal h could be based on suitable guesses or estimates of the magnitude of $\bar{f}/(\bar{f}^{(r)})^2$.

Using h given by (5.5), (5.3) becomes

$$(5.3)' \qquad MSE_n \sim c_{r,\nu}\{(k_r\bar{f}^{(r)})^{1+2\nu}(n^{-1}\|K\|_2^2\bar{f}/u)^{r-\nu}\}^{2/(1+2r)} ,$$

where

$$c_{r,\nu} = ((2\nu+1)(2r-2\nu)^{-1})^{2(r-\nu)/(1+2r)} + (2(r-\nu)(2\nu+1)^{-1})^{(2\nu+1)/(1+2r)} .$$

Relations (5.3), (5.5) and (5.3)' specialized to the case $f_n \equiv f$,

$u \equiv 1$ and $\nu = 0$ coincide (up to the factors k_r and $\|K\|_2$) with (4.12), (4.15) and (4.16), respectively, of Parzen (1962).

In the remainder of this section we derive certain properties of $\text{IMSE}_n \doteq \int_a^\infty \text{MSE}_n(t)\,dt$.

Lemma 4. For each $n \geq 1$,

$$(5.6) \qquad \int_a^\infty \sigma_n^2 \leq (nh^{2\nu+1})^{-1}\|K\|_2^2 \int_a^\infty(\bar{f}/u).$$

Proof. Integrating both sides of the inequality in (4.0) and then making use of the Tonelli theorem at the second step below we get

$$(nh^{2\nu+1})\int_a^\infty \sigma_n^2 dt \leq h^{-1}\int_a^\infty\int(K^2(y-t)/h)/u(y))\bar{f}(y)\,dydt$$

$$= \int(\bar{f}(y)/u(y))\int_a^\infty h^{-1}K^2((y-t)/h)\,dtdy$$

$$\leq \|K\|_2^2 \int_a^\infty(\bar{f}/u). \blacksquare$$

Lemma 5. Suppose $r = \nu$. If $B_n = o(1)$ a.e. on (a,∞), $\bigvee_n |g_n| \in L_2(a,\infty)$ and, for the case $\nu > 0$, $(A_0^{(\nu)})$ holds on (a,∞), then

$$(5.7) \qquad \int_a^\infty B_n^2 = o(1)\ .$$

Proof. Consider first the case $\nu = 0$. Let $s(t) = \bigvee_n \int|K(y)|\bar{f}(t+hy)\,dy$. Using Tonelli theorem and Schwarz inequality at the second step below, we get

$$\int_a^\infty s^2(t)\,dt \leq \int_a^\infty\int\int|K(y)K(\omega)|(\bigvee_n \bar{f}(t+hy))(\bigvee_n \bar{f}(t+h\omega))\,dyd\omega dt$$

$$\leq \int\int|K(y)K(\omega)|(\int_a^\infty \bigvee_n \bar{f}^2(t+hy)\,dt)^{\frac{1}{2}}(\int_a^\infty \bigvee_n \bar{f}^2(t+h\omega)\,dt)^{\frac{1}{2}}dyd\omega$$

$$\leq \int\int|K(y)K(\omega)|\int_a^\infty \bigvee_n \bar{f}^2(t)\,dtdyd\omega < \infty$$

since $K \in L_1$ by its definition, and $\bigvee_n \bar{f}(t) \in L_2(a,\infty)$ by hypothesis. Since by (1.5) and by c_r-inequality $|B_n(t)|^2 = (\int K(y)\bar{f}(t+hy)dt - \bar{f}(t))^2 \leq 2(s^2(t) + \bigvee_n \bar{f}^2(t))$ and since by hypothesis $B_n = o(1)$ a.e. on (a,∞), the desired conclusion for the case $\nu = 0$ follows by dominated convergence theorem.

Now consider the case $\nu \geq 1$. Since K vanishes off $(0,1)$, (2.4) followed by (2.3) gives, $\forall\, x \in (a,\infty)$,

$$(5.8) \qquad |B_n(x)| \leq h^{-1}\int y^{\nu-1}|K(y)|\int_x^{x+hy}|g_n(t)|dtdy + |g_n(x)|$$

$$\leq \int y^{\nu}|K(y)|\int_0^1|g_n(x+h\omega y)|d\omega dy + |g_n(x)|$$

where the last inequality follows by use of the transformation theorem. Since $y^{\nu}K(y) \in L_1$ by the definition of K, and since $\bigvee_n |g_n| \in L_2(a,\infty)$, by the technique used to prove $s^2 \in L_2(a,\infty)$, it can be shown that the $\sup_n$ of the extreme rhs of (5.8) is in $L_2(a,\infty)$. Thus the proof is complete by dominated convergence theorem. ■

<u>Lemma 6</u>. If for $r > \nu$, (1.7) holds a.e. on (a,∞), then

$$(5.9) \qquad \int_a^{\infty} B_n^2 \leq |k|_r^2 \; h^{2(r-\nu)}\int_a^{\infty}(\bar{f}^{(r)})^2.$$

<u>Proof</u>. By (1.7) we have, after use of the transformation theorem, $(r-1)!\,h^{\nu}B_n(x) = \int K(y)\int_0^{hy} z^{r-1}\bar{f}^{(r)}(x+hy-z)dzdy$ for almost all $x \in (a,\infty)$. Therefore, using Tonelli theorem, and Schwarz inequality at the second step below, we have

$$h^{2\nu}\int_a^\infty B_n^2 dx \le ((r-1)!)^{-2}\int_a^\infty \Big(\iiint_0^{hy_1}\int_0^{hy_2} |K(y_1)K(y_2)|(z_1z_2)^{r-1}|\bar f^{(r)}(z+hy_1-z_1)$$

$$\bar f^{(r)}(x+hy_2-z_2)|dz_2dz_1dy_2dy_1\Big)dx$$

$$\le ((r-1)!)^{-2}\iint |K(y_1)K(y_2)|\int_0^{hy_1}\int_0^{hy_2}(z_1z_2)^{r-1}\Big(\int_a^\infty |\bar f^{(r)}(x+hy_1-z_1)|^2dx\Big)^{\frac12}$$

$$\Big(\int_a^\infty |\bar f^{(r)}(x+hy_2-z_2)|^2dx\Big)^{\frac12}dz_2dz_1dy_2dy_1$$

$$\le h^{2r}|k|_r^2\int_a^\infty |\bar f^{(r)}|^2 \ . \blacksquare$$

<u>Lemma 7</u>. For $r > \nu$, suppose (2.8) holds on (a,∞), and both (2.9) and $\Delta_r = o(1)$ hold a.e. on (a,∞). If $\bigvee_n |\bar f^{(r)}|^2 \in L_1(a,\infty)$, then

$$\int_a^\infty \big|h^{-2(r-\nu)}B_n^2 - k_r^2|\bar f^{(r)}|^2\big| = o(1). \tag{5.10}$$

<u>Proof</u>. By Schwarz inequality, the square of the lhs in (5.10) is bounded above by $I_1\cdot I_2$, where

$$I_1 = \int_a^\infty |h^{-(r-\nu)}B_n - k_r\bar f^{(r)}|^2 \quad\text{and}\quad I_2 = \int_a^\infty |h^{-(r-\nu)}B_n + k_r\bar f^{(r)}|^2.$$

Since (2.8) holds on (a,∞), by transformation theorem and the Schwarz inequality, we get on (a,∞),

$$(|k|_{r-1}h^{r-\nu})^{-2}B_n^2(t) \le \Big(\int_0^1 |\bar f^{(r)}(t+h\omega)|d\omega\Big)^2 \tag{5.11}$$

$$\le \bigvee_n \int_0^1 |\bar f^{(r)}(t+h\omega)|^2 d\omega \ .$$

Since $\bigvee_n |\bar f^{(r)}|^2 \in L_1(a,\infty)$, so is the extreme rhs of (5.11). By c_r-inequality the integrands in I_1 and I_2 are bounded in n by an $L_1(a,\infty)$-function. Hence, since (2.9) and $\Delta_r = o(1)$ both hold a.e. on (a,∞), by dominated convergence theorem, $I_1\cdot I_2 = o(1)$. $\blacksquare$

Lemma 8. If (4.8) holds a.e. on (a,∞) and $(\vee_n \bar{f}/u) \in L_1(a,\infty)$, then

(5.12) $$\int_a^\infty \left| nh^{2\nu+1}\sigma_n^2 - \|K\|_2^2(\bar{f}/u)\right| = o(1).$$

Proof. By (4.0), $\forall\, t \in (a,\infty)$,

(5.13) $$nh^{2\nu+1}\sigma_n^2(t) \le (nh)^{-1}\Sigma_1^n P_j Y_j^2(t) = h^{-1}\int_t^{t+h} K^2((y-t)/h)(\bar{f}/u)dy$$
$$\le \int K^2(\omega)\vee_n (\bar{f}(t+h\omega)/u(t+h\omega))d\omega .$$

Since $(\vee_n \bar{f}/u) \in L_1(a,\infty)$, by an application of Tonelli theorem, we see that the extreme rhs of (5.13) is in $L_1(a,\infty)$. Hence, the integrand in (5.12) is bounded in n by a $L_1(a,\infty)$-function, and, since (4.8) holds a.e. on (a,∞), (5.12) follows by dominated convergence theorem. ∎

As an immediate corollary to Lemmas 7 and 8, we have

Corollary 5. If $k_r \ne 0$, $\liminf \int_a^\infty |\bar{f}^{(r)}|^2 > 0$ and (5.10) holds, then

(5.14) $$h^{-2(r-\nu)}\int_a^\infty B_n^2 \sim k_r^2\int|\bar{f}^{(r)}|^2;$$

and if $\liminf \int_a^\infty(\bar{f}/u) > 0$, and (5.12) holds, then

(5.15) $$nh^{2\nu+1}\int_a^\infty \sigma_n^2 \sim \|K\|_2^2 \int_a^\infty(\bar{f}/u) .$$

We will use (5.14) and (5.15) to prove (5.17) below.

In view of (5.1), various results on $IMSE_n$ can be obtained from those on $\int_a^\infty \sigma_n^2$ and $\int_a^\infty B_n^2$, e.g., if $\int_a^\infty \sigma_n^2 = o(1)$ (by (5.6) it is sufficient that $(nh^{2\nu+1})^{-1}\|K\|_2^2\int_a^\infty(\bar{f}/u) = o(1)$), then (5.7) implies $IMSE_n = o(1)$. Regarding rates of convergence, we have for $r > \nu$, Theorem 5. If (1.7) holds a.e. on (a,∞), then

$$(5.16)\quad IMSE_n \le |k|_r^2\, h^{2(r-\nu)} \int_a^\infty |\bar{f}^{(r)}|^2 + \|K\|_2^2 (nh^{2\nu+1})^{-1} \int_a^\infty (\bar{f}/u)\,;$$

and if (5.14) and (5.15) hold, then

$$(5.17)\quad IMSE_n \sim \text{rhs of (5.16) with } |k|_r \text{ replaced by } |k_r|.$$

Proof. Equation (5.1) followed by (5.6) and (5.9) yields (5.16). By (5.1),(5.17) is an immediate consequence of its hypotheses. ■

It may be recalled that a sufficient condition for (1.7) at a point is that $(A_0^{(r)})$ holds at that point. Thus a simple assumption gives (via (5.16)) a rate for $IMSE_n$ quite similar to the exact rate obtained in (5.17).

Since $|k_r| \le \|K\|_2$ by Schwarz inequality, (5.16) with $h = n^{-1/(1+2r)}$ yields

$$(5.18)\quad IMSE_n \le (\|K\|_2 h^{r-\nu})^2 \int_a^\infty \{|\bar{f}^{(r)}|^2 + (\bar{f}/u)\}\ .$$

Remark 5.3. The result in (5.18) specialized to the case $u \equiv 1$, $f_n \equiv f$, $\nu = 0$ and $a = -\infty$ improves the result in (3.6) of Schwartz (1967) who exhibits an estimator of f by orthogonal series method. Assuming $t^j f^{(r-j)}(t)$, $j = 0,1,\ldots,r$, are in L_2, he shows that $IMSE_n$ of his estimator is $O(n^{-(r-1)/r})$. This rate is significantly weaker (especially when r is not large) than our rate $O(n^{-2r/(1+2r)})$, which is guaranteed in this case if we only assume that $f^{(r)} \in L_2$. Moreover, he restricts $r > 1$, while we assume $r > 0$.

An optimal choice of h as a function of n and independent of the point at which g_n is to be estimated can be obtained by considering a global measure of how good $\hat{g}_n$ is as an estimator of g_n. The integrated mean square error is a standard measure of this type. The global optimal h, as the minimizer of the asymptotic expression in (5.17) for $IMSE_n$ is given by

$$(5.5)' \quad h^{1+2r} = (n^{-1}(1+2\nu)\|K\|_2^2\int_a^\infty(\bar{f}/u)/(2(r-\nu)k_r^2\int_a^\infty(\bar{f}^{(r)})^2),$$

and, hence, could be approximated by some suitable guess or estimate of the magnitude of the ratio $\int_a^\infty(\bar{f}/u)/\int_a^\infty(\bar{f}^{(r)})^2$. Using h given by (5.5)', the asymptotically minimum possible value of $IMSE_n$ can be obtained by (5.17).

1.6 Estimation of Mixed Partial Derivatives of the Average of Multivariate μ-Densities.

Let $X_1,\ldots,X_n$ be independent m-variate random variables with $X_j \sim P_j \ll \mu$, where P_j's and μ are over R^m, and μ is absolutely continuous wrt Lebesgue measure. Unless stated otherwise, throughout this section, the product Π is over $1,\ldots,m$. With t in R^m, and u , a fixed determination of $d\mu/dt$, let $f_j(t) = (u(t))^{-1}\lim_{\epsilon_i \downarrow 0, i=1,\ldots,m}(\Pi\epsilon_i^{-1})\int_{t_1}^{t_1+\epsilon_1}\cdots\int_{t_m}^{t_m+\epsilon_m} dP_j$ if limit exists $\forall\ j \geq 1$ and $u(t) > 0$, and 0 otherwise. For v and t in R^m with elements of v non-negative integers, and for $|v| = \Sigma_1^m v_j$, let $f_j^{(v)}(t) = \partial^{|v|}f_j(t)/(\Pi\partial t_i^{v_i})$. For a fixed vector $\nu = (\nu_1,\ldots,\nu_m)$, $\nu_j \geq 0$ integers, we consider estimation of $\bar{f}^{(\nu)} = n^{-1}\Sigma_1^n f_j^{(\nu)}$.

Let $h = (h_1, \ldots, h_m)$ be such that $0 < h_i \doteq h_{i,n} \leq 1$ and $h_i \downarrow 0$ as $n \uparrow \infty$. With $r = (r_1, \ldots, r_m)$, $r_i \geq \nu_i$ integers, let $\mathcal{K}_{\nu_i}^{r_i}$ be defined as in Section 1. For fixed K_i in $\mathcal{K}_{\nu_i}^{r_i}$, let

$$(6.0) \quad \hat{\bar{f}}^{(\nu)}(x) = n^{-1}\Sigma_{j=1}^{n}\{\{\Pi(h_i^{\nu_i+1})^{-1}K_i(\frac{X_{ij}-x_i}{h_i})\}[u(X_j) > 0]/u(X_j)\}$$

where $X_{1j}, \ldots, X_{mj}$ are coordinates of X_j. $\hat{\bar{f}}^{(\nu)}$ is our proposed estimator of $\bar{f}^{(\nu)}$. Taking expectation of $\hat{\bar{f}}^{(\nu)}(x)$ wrt $\underset{\sim}{P}_n = P_1 \times \ldots \times P_n$, and then making use of the transformation theorem, we get

$$(6.1) \qquad \underset{\sim}{P}_n \hat{\bar{f}}^{(\nu)}(x) = \int(\Pi h_i^{-\nu_i} K_i(y_i))\bar{f}(x+h\cdot y)dy$$

where $h \cdot y = (h_1 y_1, \ldots, h_m y_m)$.

For z and t in R^m, let $(z)_i(t) = (t_1, \ldots, t_i, z_{i+1}, \ldots, z_m)$. With $1 \leq \ell \leq m$ and with the first ℓ-elements of r non-negative integers, we introduce

$(A_0^{(r)})_\ell$: For $\forall\, y \in (0,1)^m$ and for each $i \leq \ell$, $\bar{f}(x+h\cdot y)$ has r_i-th order Taylor expansion in $h_i y_i$ about x_i with integral form of the remainder, while other components of $(x + h\cdot y)$ are held fixed. (Such expansion in the univariate case is given in $(A_0^{(r)})$ in Section 1.)

Suppose ℓ $(0 \leq \ell \leq m)$ elements of ν are positive. Without loss of generality, let these be $\nu_1, \ldots, \nu_\ell$. Suppose $(A_0^{(\nu)})_\ell$ holds. Using Taylor formula, we expand $\bar{f}(x+h\cdot y)$ (appearing in (6.1)) in $h_1 y_1$ about x_1 with integral form of the remainder at the ν_1-th term, we perform the integration on the rhs of (6.1) wrt y_1 and use the orthogonality properties of K_1. Then using

Taylor formula, we expand $\bar{f}^{(\nu_1,0,\ldots,0)}((x+h\cdot y)_1(t))$ (appearing in the resultant) in h_2y_2 about x_2 with integral form of the remainder at the ν_2-th term, we perform the integration wrt y_2 and use the orthogonality properties of K_2; then we do the similar operations wrt (x_3,ν_3,K_3) with $\bar{f}^{(\nu_1,\nu_2,0,\ldots,0)}((x+h\cdot y)_2(t))$ (appearing in the resultant), and so on until such operations wrt (x_ℓ,ν_ℓ,K_ℓ) are completed. We finally get

$$(6.2)\quad \underset{\sim}{P}_n\hat{\bar{f}}^{(\nu)}(x) = \int(\Pi h_i^{-\nu_i}K_i(y_i))\int J_y^\nu(t)\bar{f}^{(\nu)}((x+h\cdot y)_\ell(t))d_\ell t dy$$

where $J_y^\nu(t) = \Pi\{[x_i \le t_i < x_i + h_iy_i](x_i+h_iy_i-t_1)^{\nu_i-1}((\nu_i-1)!)^{-1}[\nu_i>0]+[\nu_i = 0]\}$, $d_\ell t = \Pi_1^\ell dt_i$ and the second integral on the rhs of (6.2) is ℓ-tuple. Since $\int z^{\nu_i}K_i(z)dz = \nu_i!$, $\int(\Pi h_i^{-\nu_i}K_i(y_i))\int J_y^\nu(t)d_\ell t dy = 1$. Consequently, using the transformation $x_i + h_iy_i = t_i$, $\ell+1 \le i \le m$, and the facts that $J_y^\nu(t) \le \Pi_1^\ell((h_iy_i)^{\nu_i-1}/(\nu_i-1)!)$ and $K_i \equiv 0$ off $(0,1)$, all at the second step below, we get, with $B_n = \underset{\sim}{P}_n\hat{\bar{f}}^{(\nu)} - \bar{f}^{(\nu)}$,

$$(6.3)\quad |B_n(x)| \le \int(\Pi h_i^{-\nu_i}|K_i(y_i)|)\int J_y^\nu(t)|\bar{f}^{(\nu)}((x+h\cdot y)_\ell(t))-\bar{f}^{(\nu)}(x)|d_\ell t dy$$

$$\le (\Pi_1^\ell|{}_ik|_{\nu_i-1})(\Pi_{\ell+1}^m\|K_i\|_\infty)\Delta_\nu(x),$$

where $\Delta_\nu(x) = \int(\Pi h_i^{-1}[x_i \le t_i < x_i + h_i])|\bar{f}^{(\nu)}]_x^t|dt$, and $|{}_ik|_j = \int z^j|K_i(z)|dz/j!$. Thus, if $\|K_i\|_\infty < \infty$ for $\ell+1 \le i \le m$ (bounded kernels $K_i \in \mathcal{K}_{\nu_i}^{r_i}$, $r_i \ge \nu_i$, can always be exhibited, see Section 1 and Remark 3.1), and $(A_0^{(\nu)})_\ell$ holds and $\Delta_\nu(x) = o(1)$, then

$$(6.4)\quad B_n(x) = o(1).$$

Now suppose for $r_i > \nu_i$, $(A_0^{(r)})_m$ holds. An analysis similar to that given for (6.2) (this time use r_1-th, r_2-th, ..., r_m-th order coordinatewise Taylor expansion with integral form of the remainder) gives

$$(6.5)\quad \underset{\sim}{P}_n \hat{\bar{f}}^{(\nu)}(x) = \bar{f}^{(\nu)}(x) + \int (\Pi h_i^{-\nu_i} K_i(y_i)) \int J_y^r(t) \bar{f}^{(r)}(t)\,dt\,dy .$$

Since K_i's vanish off $(0,1)$, $J_y^{(r)}(t) \le \Pi((h_i y_i)^{r_i - 1}/(r_i - 1)!)$ and $\int (\Pi h_i^{-r_i} K_i(y_i)) \int J_y^{(r)}(t)\,dt\,dy = \Pi({}_i k_{r_i})$ where ${}_i k_j = \int z^j K_i(z)\,dz/j!$, it follows from (6.5) that

$$(6.6)\quad |B_n(x)| \le (\Pi h_i^{r_i - \nu_i - 1} |{}_i k|_{r_i - 1}) \int (\Pi[x_i \le t_i < x_i + h_i]) \bar{f}^{(r)}(t)\,dt$$

and

$$(6.7)\quad |(\Pi h_i^{\nu_i - r_i} B_n(x) - (\Pi_i k_{r_i}) \bar{f}^{(r)}(x)| \le (\Pi |{}_i k|_{r_i - 1}) \Delta_r(x) .$$

Results obtained in (6.3), (6.6) and (6.7) for $m = 1$ coincide with (2.6), (2.8) and (2.9), respectively.

Since $X_1, \ldots, X_n$ are independent, the inequality $\operatorname{var} X \le EX^2$ followed by the transformation theorem gives

$$(6.8)\quad \sigma_n^2(x) \doteq \operatorname{var} \hat{\bar{f}}^{(\nu)}(x) \le M^2 (n \Pi h_i^{2(\nu_i + 1)})^{-1} \int (\Pi[x_i \le t_i < x_i + h_i])(\bar{f}/u)\,dt$$

where $M \doteq \Pi \|K_i\|_\infty$. Since $MSE_n \doteq \underset{\sim}{P}_n |\hat{\bar{f}}^{(\nu)} - \bar{f}^{(\nu)}|^2 = B_n^2 + \sigma_n^2$, rates of convergence for MSE_n can be obtained from (6.4), (6.6) and (6.8).

If, corresponding to u and h here, u_h is defined analogous to (3.2), then by Theorem 2 of Hoeffding (1963), $\forall\ \eta > 0$

$$\underset{\sim}{P}_n[|\hat{\bar{f}}^{(\nu)} - \underset{\sim}{P}_n \hat{\bar{f}}^{(\nu)}| > \eta] \le 2\exp\{-\tfrac{n}{2}((\Pi h_i^{\nu_i + 1})\eta\, u_h/M)^2\}.$$

Thus by Borel-Cantelli lemma $(\Pi h_i^{\nu_i + 1}) |\hat{\bar{f}}^{(\nu)} - \underset{\sim}{P}_n \hat{\bar{f}}^{(\nu)}| = O((n^{-1}\log n)^{\frac{1}{2}}(M/u_h))$ a.s., and rates for strong consistency of $\hat{\bar{f}}^{(\nu)}(x)$ can be obtained from (6.4) and (6.6), since $|\hat{\bar{f}}^{(\nu)} - \bar{f}^{(\nu)}| \leq |B_n| + |\hat{\bar{f}}^{(\nu)} - \underset{\sim}{P}_n \hat{\bar{f}}^{(\nu)}|$.

Though we have verified some of the asymptotic properties of $\hat{\bar{f}}^{(\nu)}$, it is not our intent to encounter and verify all the properties of $\hat{\bar{f}}^{(\nu)}$ that we have already studied in the univariate case. However, regarding some of these, it can be verified that under the assumptions analogous to those given for (4.7), (4.8), (4.11), (5.2), (5.3), (5.16) and (5.17), results analogous to these also hold good in the multivariate case here (analogoue of (4.8) is $n(\Pi h_i^{2\nu_i+1})\sigma_n^2 - (\Pi\|K_i\|_2^2)(\bar{f}/u) = o(1)$ and that of (5.16) is $IMSE_n \leq (\Pi|{}_ik|_{r_i}^2 h_i^{2(r_i-\nu_i)})\int_{a_1}^{\infty} \cdots \int_{a_m}^{\infty} |\bar{f}^{(r)}|^2 dt +$ $(\Pi\|K_i\|_2^2 h_i^{-2\nu_i-1})n^{-1}\int_{a_1}^{\infty} \cdots \int_{a_m}^{\infty} (\bar{f}/u)dt$, etc.). Conditions (somewhat) necessary for asymptotic unbiasedness, (and also for strong or mean square consistency), uniform on any subset of R^m, are analogous to those given for the same in the univariate case (cf. Remark 2.2 and Corollary 1).

CHAPTER 2

CONVERGENCE RATES IN SEQUENCE-COMPOUND SQUARED ERROR LOSS ESTIMATION OF CERTAIN UNBOUNDED FUNCTIONALS IN EXPONENTIAL FAMILIES

2.0 Introduction.

Let Ω be a parameter space indexing a family of probability measures $\mathcal{P} = \{P_\omega | \omega \in \Omega\}$ on a sample space $\mathfrak{X}$. With an observation on a random variable $X \sim P_\omega$, let the component problem be squared error loss estimation (SELE) of real $\theta(\omega)$.

Suppose this component problem occurs repeatedly and independently. Then, after n such occurrences, we have an unknown vector $\underset{\sim}{\omega} = (\omega_1, \ldots, \omega_n) \in \Omega^n$ and a corresponding vector of independent random variables, $\underset{\sim}{X} = (X_1, \ldots, X_n)$ with $X_j \sim P_j \doteq P_{\omega_j}$. With θ_j abbreviating $\theta(\omega_j)$, we consider estimation of each component of $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_n)$ with loss taken to be the average of the squared-error losses in the individual components.

We call $\underset{\sim}{\varphi} = (\varphi_1, \ldots, \varphi_n)$ a sequence compound estimator (henceforth, compound estimator or simply estimator) if φ_j is $(X_1, \ldots, X_j)$-measurable. Let G_i be the empiric distribution function of the first i components of $\underset{\sim}{\omega}$, and $R(\cdot)$ be the Bayes envelope for the component problem. With a $\delta > 0$, we say $\underset{\sim}{\varphi}$ achieves a rate δ (at $\underset{\sim}{\theta}$) if the modified regret of $\underset{\sim}{\varphi}$, defined by

$$(0.1) \qquad D_n(\underset{\sim}{\omega}, \underset{\sim}{\varphi}) = n^{-1} \Sigma_1^n \underset{\sim}{P}_j (\varphi_j - \theta_j)^2 - R(G_n)$$

is $O(n^{-\delta})$ as $n \to \infty$, where $\underset{\sim}{P}_j = P_1 \times \ldots \times P_j$. We now describe the main results briefly as follows.

In Section 1, we use the method of Gilliland (1968), (Section 2), to obtain an explicit bound for $|D_n(\underset{\sim}{\omega}, \underset{\sim}{\varphi})|$.

In Section 2, we introduce some further assumptions and notations. For the results in Sections 3-6, $\mathcal{X} = R$, $\mathcal{P}$ is an exponential family wrt μ, a σ-finite measure dominated by Lebesgue measure on R and Ω is the natural parameter space.

Using the technique developed by Gilliland (1966), (Chapter III), we exhibit, in Section 3, a divided difference estimator for $\underset{\sim}{\omega}$ with a rate 1/5.

Based on estimators (introduced in Chapter 1) of derivatives of the average of μ-densities, we exhibit, in Section 4, kernel estimators of $\underset{\sim}{\theta}$ for (integer) $r > 1$ when $\theta(\omega) = \omega$, and for (integer) $r > 0$ when $\theta(\omega) = e^{\omega}$ or ω^{-1}. These estimators are shown to have rates $(r-1)/(1+2r)$, $r/(1+2r)$ or $(r-)/(1+2r)$ in their respective cases of ω, e^{ω} or ω^{-1}.

In Section 5, we show that, when θ is an identity map and $\underset{\sim}{\omega}$ has identical components, rates with the divided difference and the kernel estimators are near, but cannot be more than, 2/5 and $2(r-1)/(1+2r)$, respectively.

A comparison between the divided difference and the kernel estimators, when θ is identity, is made in Section 6. Because of the reason stated there, the latter one is preferable to the former one.

2.1 A Bound for the Modified Regret.

In this section we will prove two simple but useful lemmas. Special forms of both have been studied, among others, by Gilliland (1968) and Susarla (1970). Lemma 1 is essentially due to Gilliland (1968), and Lemma 2 is a consequence of inequalities (8.8) and (8.11) of Hannan (1957), and of Lemma 1.

With μ some σ-finite measure dominating $P_j \; \forall \; j = 1,\ldots,n$, let f_j be a determination of $dP_j/d\mu$. Let $m_i \geq \max_{1\leq j\leq i} f_j$ and $N_i \geq \max_{1\leq j\leq i}|\theta_j|$ be such that m_i and N_i are non-decreasing. Recall that θ_j abbreviates $\theta(\omega_j)$. As the Bayes response against G_i in the component problem, we take the version of conditional expectation

$$(1.0) \qquad \psi_i = \frac{\Sigma_1^i \theta_j f_j}{\Sigma_1^i f_j} [\Sigma_1^i f_j > 0] .$$

Thus $|\psi_i| \leq N_i$. For the purpose of this section only, take ψ_0 arbitrary real valued function on R, and for $j \geq 1$, define $\Delta_j = \psi_j - \psi_{j-1}$.

Lemma 1. With ψ_0 taking values in $[-N_n, N_n]$,

$$\Sigma_1^n P_i|\Delta_i(X_i)| \leq 2N_n(1 + \log n)\mu(m_n) .$$

Proof. Abbreviate, throughout this proof, N_n by N. From (1.0) it follows that, for $1 \leq i \leq n$,

$$\Delta_i = \frac{(\theta_i - \psi_{i-1}) f_i}{\Sigma_1^i f_j} \quad \text{a.e. } P_i .$$

Consequently, since $|\theta_i - \psi_{i-1}| \leq 2N$ for $\forall \; 1 \leq i \leq n$, (1.1) $$\Sigma_1^n P_i|\Delta_i(X_i)| \le 2N\,\mu\{m_n\Sigma_1^n \frac{(f_i/m_n)^2}{\Sigma_1^i(f_j/m_n)}\}\,.$$

Since by Lemma 2.1 of Gilliland (1968), $\Sigma_1^n a_i^2(\Sigma_1^i a_j)^{-1} \le \Sigma_1^n i^{-1}$ for all $0 \le a_i \le 1$, $1 \le i \le n$ and $n \ge 1$, the rhs of (1.1) is bounded above by $2N(\Sigma_1^n i^{-1})\mu(m_n) \le 2N(1 + \log n)\mu(m_n)$. ■

<u>Lemma 2</u>. For any estimator $\underset{\sim}{\varphi} = (\varphi_1,\ldots,\varphi_n)$ with φ_1 and, for $i = 2,\ldots,n$, φ_i taking values in $[-N_n,N_n]$ and $[-N_i,N_i]$ respectively,

$$|D_n(\underset{\sim}{\omega},\underset{\sim}{\varphi})| \le 4n^{-1}\Sigma_2^n N_j \underset{\sim}{P}_j|\varphi_j(X_j) - \psi_{j-1}(X_j)| + 8n^{-1}N_n^2(1 + \log n)\mu(m_n).$$

<u>Proof</u>. Unless stated otherwise, sums in this proof are taken from 1 to n. Let the argument X_j in various summands below in this proof be abbreviated by omission. Inequalities (8.8) and (8.11) of Hannan (1957) specialized to the SELE problem here yield

(1.2) $$\Sigma P_j|\psi_j - \theta_j|^2 \le nR(G_n) \le \Sigma P_j|\psi_{j-1} - \theta_j|^2.$$

The identity $b^2 - c^2 = (b-c)(b+c)$ followed by (0.1) and (1.2) gives

$$\Sigma \underset{\sim}{P}_j((\varphi_j - \psi_{j-1})(\varphi_j + \psi_{j-1} - 2\theta_j)) \le nD_n(\underset{\sim}{\omega},\underset{\sim}{\varphi})$$

(1.3) $$\le \Sigma \underset{\sim}{P}_j((\varphi_j-\psi_j)(\varphi_j+\psi_j-2\theta_j))$$

$$= \Sigma \underset{\sim}{P}_j((\varphi_j-\psi_{j-1}-\Delta_j)(\varphi_j+\psi_j-2\theta_j)).$$

Since ψ_0 is arbitrary, we can (and do) take $\psi_0 = \varphi_1$. Then, since, for $j \ge 2$, φ_j, ψ_j, ψ_{j-1} and θ_j are in $[-N_j,N_j]$, and

$\max_{1\le j\le n}|\varphi_j + \psi_j - 2\theta_j| \le 4N_n$, from (1.3),

$-4\Sigma_2^n N_j P_{\tilde{j}} |\varphi_j - \psi_{j-1}| \le nD_n(\underset{\sim}{\omega},\underset{\sim}{\varphi}) \le 4(\Sigma_2^n N_j P_{\tilde{j}} |\varphi_j - \psi_{j-1}| + N_n \Sigma P_j |\Delta_j|)$.

The last inequalities and Lemma 1 now complete the proof. ■

2.2 Some Assumptions and Notations.

For the remainder of this chapter, we take $\mathfrak{X} = R$, the real line, and assume $\mathcal{P} \ll \mu$, where μ is a σ-finite measure dominated by Lebesgue measure on R. With u , a fixed determination of $d\mu/dx$, we assume the existence of an $a \ge -\infty$ such that

(2.0) $$u(x) > 0 \quad \text{iff} \quad x > a.$$

Furthermore, we take

(2.1) $$\Omega = \{\omega \in R | (C(\omega))^{-1} \doteq \int e^{\omega x} d\mu(x) < \infty\};$$

and, for $\omega \in \Omega$,

(2.2) $$f_\omega(x) = C(\omega)e^{\omega x} \quad \text{for} \quad x > a,$$

(and zero otherwise), as a fixed density of P_ω wrt μ. (Thus, with f_j abbreviating f_{ω_j}, uf_j is a Lebesgue density of X_j). Let $\alpha_i \le \min_{1\le j\le i}\omega_j$ and $\beta_i \ge \max_{1\le j\le i}\omega_j$ be in Ω for each $1 \le i \le n$, and $\alpha_i\downarrow$ and $\beta_i\uparrow$. We also take

(2.3) $$m_i = \sup\{f_\omega | \omega \in [\alpha_i,\beta_i]\} \quad \text{and} \quad N_i = \sup\{|\theta(\omega)| \, | \omega \in [\alpha_i,\beta_i]\}.$$

For $\tau > 0$ and $x > a$, define

(2.4) $u_\tau(x)$ = Lebesgue-inf of the restriction to $[x,x+\tau)$ of u.

The conclusion of Lemma 2 will be used in obtaining certain rates for various estimators to be introduced in later sections. Since the upper bound in the lemma does not depend on the first component (with values in $[-N_n, N_n]$) of the estimator $\underset{\sim}{\varphi}$ there, without any further indication, the such component of each of the estimators (yet to be introduced), is taken to be arbitrary with values in $[-N_1, N_1]$.

Our work in each of the next two sections is comprised of mainly two steps: First to exhibit an appropriate estimator φ_{i+1} of ψ_i and then to obtain a suitable bound for $N_{i+1}\underset{\sim}{P}_{i+1}|\varphi_{i+1}(X_{i+1}) - \psi_i(X_{i+1})|$ for each $i = 1,\ldots,n-1$. Using this and Lemma 2, we will obtain a bound for $|D_n(\underset{\sim}{\omega},\underset{\sim}{\varphi})|$ uniformly in $\underset{\sim}{\omega} \in \times_1^n[\alpha_i, \beta_i]$.

Let $0 < h_n \le h_{n-1} \le \ldots \le h_1 \le 1$. Unless stated otherwise, we, hereafter, fix i with $1 \le i \le n-1$ and drop the subscripts in m_i, α_i, β_i, N_i, h_i and X_{i+1}. For $a_j \in R$, let $\bar{a} = i^{-1}\Sigma_1^i a_j$.

Note that $\log C(\omega) \doteq -\log \int e^{\omega\cdot} d\mu(\cdot)$ is concave on $[\alpha,\beta]$ and, hence, so is $\log f_\omega(x) = \omega x + \log C(\omega)$ for each x. Thus, $\inf_{\alpha\le\omega\le\beta} f_\omega = f_\alpha \wedge f_\beta$. Hence, $\forall\ \gamma \ge 0$

$$(2.5) \qquad q_\gamma \doteq (m_{i+1}/(f_\alpha \wedge f_\beta)^{\gamma/2}) \ge f_{i+1}/(\bar{f})^{\gamma/2}.$$

For a real valued function g on R and for numbers $b < c$, abbreviate the retraction of g to $[b,c]$ by $(g)_{b,c}$. Unless stated otherwise, all the limits (of functions depending on i) are taken as $i \uparrow \infty$ (hence, necessarily as $n \uparrow \infty$).

2.3 A Divided Difference Estimator of $\underset{\sim}{\omega}$ with a Rate 1/5.

In this section we consider the case when θ is the identity map. Since $f_j(x) = C(\omega_j)\exp(\omega_j x)$, by (1.0), $\psi_i = (\log \bar{f})^{(1)}[\bar{f} > 0]$. Motivated by this expression, the compound estimator $\hat{\underset{\sim}{\psi}}$ to be introduced here will be based on a divided difference estimator of $\log \bar{f}$. The main idea behind the construction of this kind of estimator is developed by Gilliland (1966), (Chapter 3), in sequence-compound SELE of means in the family of normal densities. Our technique to be introduced here in defining $\hat{\underset{\sim}{\psi}}$ is, however, a little different than those of Gilliland (1966), (Chapter 3), Susarla (1970), (Section 1.2), and Hannan and Macky (1971); and does not require the continuity of u for $\hat{\underset{\sim}{\psi}}$ to have a rate. The method used here to get rid of the continuity requirement of u is partly due to Yu (1970), (Section 2 of the appendix), where he exhibits kernel estimators of a density function and its derivative.

Define a real valued functional Q on the space of all real valued non-negative functions t on R by

(3.0) $$Q(t)(x) = h^{-1}(\log \frac{t(x+h)}{t(x)})[t(x+h) + t(x) > 0] .$$

Let $\eta = e^{2hN}$ and, for $j = 1,\ldots,i$, let $\delta_j(y) = \int_y^{y+h} f_j$ and $\hat{\delta}_j(y) = [y \le X_j < y+h]/u(X_j)$. Note that $\hat{\delta}_j$ is well defined with probability one, and is an unbiased estimator of δ_j.

The compound estimator $\hat{\underset{\sim}{\psi}}$, which we propose for $\underset{\sim}{\omega}$, has $(i+1)$st component

(3.1) $$\hat{\psi}_{i+1}(X) = (Q(\bar{\hat{\delta}})(X))_{\alpha,\beta}.$$

Abbreviate $Q(\bar{\delta})(x)$ and $Q(\bar{\hat{\delta}})(x)$ by $Q(x)$ and $\hat{Q}(x)$ respectively.

For $x > a$, define

$u^*(x)$ = Lebesgue-sup of the restriction to $[x, x+2h)$ of u.

Denote u_{2h} by u_*. In Lemma 3 below and in its proof, Q, $\hat{Q}$, u_*, u^*, m and $\bar{f}$ all are evaluated at a fixed point $x > a$.

Lemma 3. $\forall\ \gamma > 0$

$$(3.2)\qquad \underset{\sim}{P}_i(|Q - \hat{Q}| \wedge 2N)^\gamma \le k_0(\gamma)(ih^3\bar{f}\ u_*^2/u^*)^{-\gamma/2}$$

where $k_0(\gamma) = \gamma\Gamma(\gamma/2)(16\eta^3(1 + \eta^2)/3k^+)^{\gamma/2}$ with $k = 1 - h\eta u^* m$.

Proof. The lhs of (3.2) is

$$(3.3)\qquad \int_0^{2N} \underset{\sim}{P}_n[|Q - \hat{Q}| > v]d(v^\gamma) = \int_0^{2N}(p_1(v) + p_2(v))d(v^\gamma),$$

where $p_1(v) = \underset{\sim}{P}_i[(\hat{Q} - Q) > v]$ and $p_2(v) = \underset{\sim}{P}_i[(Q - \hat{Q}) > v]$. Our method of the proof here involves obtaining an appropriate upper bound for $p_1(v) + p_2(v)$ with $0 < v < 2N$.

Fix v in $(0,2N)$ until stated otherwise. For $j = 1,\ldots,i$, let $Y_j = \hat{\delta}_j(x+h) - R\ e^{hv}\hat{\delta}_j(x)$, where $R = \bar{\delta}(x+h)/\bar{\delta}(x)$. Let $\nu_j = P_jY_j$ and $\sigma^2 = i\ \mathrm{var}(\bar{Y})$. We will first obtain (3.7) below by obtaining suitable upper bounds for $\bar{\nu}$ and σ^2. Notice that $\nu_j = \delta_j(x+h) - R\ e^{hv}\delta_j(x)$. Hence $\bar{\nu} = (1 - e^{hv})\bar{\delta}(x+h)$, and we get

$$(3.4)\qquad -\eta\bar{\delta}(x+h) \le \bar{\nu} \le -hv\bar{\delta}(x+h)\ .$$

By independence of $Y_1,\ldots,Y_i$ and by c_r-inequality (Loève (1963), p. 155) we have

$$(3.5)\qquad i\sigma^2 \le \Sigma_1^i\ P_jY_j^2 \le 2\Sigma_1^i(P_j\hat{\delta}_j^2(x+h) + R^2e^{2hv}P_j\hat{\delta}_j^2(x)).$$

Since $v < 2N$, $R = \bar{\delta}(x+h)/\bar{\delta}(x)$ and, for $y = x, x+h$, $P_j\hat{\delta}_j^2(y) \le \delta_j(y)/u_*$, by (3.5) we get

$$\sigma^2 \le 2(1 + R\eta^2)\bar{\delta}(x+h)/u_*$$
$$= 2((\bar{\delta}(x+h))^{-1} + \eta^2(\bar{\delta}(x))^{-1})\bar{\delta}^2(x+h))/u_*.$$

Now, since, for $1 \le j \le i$, $\omega_j \in [-N,N]$,

$$(3.6) \qquad h\eta^{-1} \le \frac{\delta_j(y)}{f_j(x)} = \int_y^{y+h} e^{\omega_j(t-x)} dt \le h\eta \quad \text{for } y = x, x+h.$$

Therefore, weakening the final upper bound obtained above for σ^2 by the first inequality in (3.6) we get $u_*\bar{f}\sigma^2 \le 2(1+\eta^2)\eta h^{-1}\bar{\delta}^2(x+h)$. This last inequality and (3.4) give

$$(3.7) \qquad \frac{(-\bar{\nu})^2}{\sigma^2} \ge \frac{h^3 v^2 \bar{f} u_*}{2(1+\eta^2)\eta} .$$

Next we will obtain (3.10) below by obtaining appropriate lower bounds for σ^2, $\bar{\nu}$, ν_j and $-Y_j$. By independence of $Y_1,\ldots,Y_i$ and by the facts that $v > 0$, $P_j(\hat{\delta}_j(\cdot)) > 0$, and $\hat{\delta}_j(x+h)\hat{\delta}_j(x) = 0$ with probability one, we get

$$(3.8) \qquad \sigma^2 \ge i^{-1}\Sigma_1^i(\text{var}(\hat{\delta}_j(x+h)) + R^2\text{var}(\hat{\delta}_j(x))) .$$

Now the definition of u^* and the second inequality in (3.6) yield, for $y = x, x+h$,

$$\text{Var}(\hat{\delta}_j(y)) = \int_y^{y+h}(f_j/u) - \delta_j^2(y)$$
$$\ge (\delta_j(y)(1 - u^*\delta_j(y))^+/u^*)$$
$$\ge (\delta_j(y)(1 - h\eta u^* f_j)^+/u^*) \ge k^+\delta_j(y)/u^*,$$

where k is as given in the lemma, and the last inequality follows from the definition of m given in (2.3). Consequently, from (3.8) we get

$$(3.9) \qquad u^{*}\sigma^{2} \ge (\bar{\delta}(x+h) + R^{2}\bar{\delta}(x))k^{+} = (1+R)\bar{\delta}(x+h)k^{+}.$$

Next observe that $-R\,e^{hv}\hat{\delta}_j(x) \le Y_j \le \hat{\delta}_j(x+h)$. Therefore, since for $y = x, x+h$, $\hat{\delta}_j(y) \le 1/u_*$ with probability one, $Y_j \le 1/u_*$ and $-\nu_j \le R\eta/u_*$. These upper bounds for Y_j and $-\nu_j$ together with (3.4) and (3.9) yield $(Y_j-\nu_j)(-\bar{\nu}/\sigma^2) \le \{(1+\eta R)\eta u^{*}/(k^{+}(1+R)u_*)\}$ $\le \eta^2 u^{*}(k^{+}u_*)^{-1}$. Hence

$$(3.10) \qquad Y_j - \nu_j \le \frac{\eta^2 u^{*}}{k^{+}u_*}\left(-\frac{\sigma^2}{\bar{\nu}}\right).$$

We will use (3.7) and (3.10) to obtain a suitable upper bound for $p_1(v)$. Note that the event in $p_1(v)$ is $[\bar{Y} > 0]$. Therefore, (3.10) and the Bernstein inequality stated in (2.13) of Hoeffding (1963) give

$$(3.11) \qquad p_1(v) = \underset{\sim}{P}_i[\bar{Y} - \bar{\nu} > -\bar{\nu}] \le \exp\left\{-\frac{i(-\bar{\nu})^2}{\sigma^2 2\left(1 + \frac{\eta^2 u^{*}}{3k^{+}u_*}\right)}\right\}$$

$$\le \exp\left\{-\frac{3ik^{+}h^3 v^2 \bar{f} u_*^2}{16\eta^3(1+\eta^2)u^{*}}\right\}$$

where the last inequality follows by (3.7) and by the fact that $(1 + \eta^2 u^{*}/(3k^{+}u_*)) \le 4(3k^{+}u_*)^{-1}\eta^2 u^{*}$, since $\eta \ge 1$, $k^{+} \le 1$ and $u^{*} \ge u_*$.

By interchanging x, x+h in the definition of Y_j's and by applying the techniques used for bounding $p_1(v)$, we see that $p_2(v)$ is also bounded above by the extreme rhs in (3.11).

Now bounding above the integrand on the rhs of (3.3) by the upper bound just obtained for $p_1(v) + p_2(v)$ and then performing the integration there after extending the range of integration from $(0,2N)$ to $(0,\infty)$ we get the desired conclusion. ■

<u>Lemma 4</u>.

$$\sup_{t>a} |(Q(\bar{\delta}) - \frac{\bar{f}^{(1)}}{\bar{f}})(t)| \le 4(N\eta)^2 h .$$

<u>Proof</u>. Since, for $1 \le j \le i$, $\omega_j \in [-N,N]$, for each integer $\nu \ge 0$ and $\forall\, t \in [\cdot, \cdot + 2h]$ we have

$$(3.12) \qquad \frac{|f_j^{(\nu)}(t)|}{f_j(\cdot)} = |\omega_j^\nu| e^{\omega_j(t-.)} \le N^\nu \eta ,$$

and

$$(3.13) \qquad \frac{f_j(t)}{f_j(\cdot)} = e^{\omega_j(t-\cdot)} \ge \eta^{-1} .$$

For the purpose of this proof only, let $g_j = \omega_j^{-1} f_j$. Since $\bar{\delta}(t) = \bar{g}(t+h) - \bar{g}(t)$, by Cauchy-mean value theorem, see Graves (1956), p. 81, for some ϵ in $(0,1)$

$$(3.14) \qquad \frac{\bar{\delta}(t+h)}{\bar{\delta}(t)} = \frac{\bar{g}^{(1)}(t + h + \epsilon h)}{\bar{g}^{(1)}(t + \epsilon h)} = \frac{\bar{f}(t + h + \epsilon h)}{\bar{f}(t + \epsilon h)} .$$

Therefore, by (3.14) and by mean value theorem, $Q(\bar{\delta})(t) = h^{-1}\log(\bar{f}(t + h + \epsilon h)/\bar{f}(t + \epsilon h)) = (\log \bar{f}(t'))^{(1)}_{t'=t+\gamma h}$ for some $\gamma \in (0,2)$. Making another use of mean value theorem at the third step below, we thus have, for some $\gamma', \gamma'' \in (0,\gamma h)$

$$|Q(\bar{\delta})(t) - (\frac{\bar{f}^{(1)}}{\bar{f}})(t)| = |(\frac{\bar{f}^{(1)}}{\bar{f}})(t+\gamma h) - (\frac{\bar{f}^{(1)}}{\bar{f}})(t)|$$

$$\le \frac{1}{\bar{f}(t+\gamma h)}(|\bar{f}^{(1)}(t+\gamma h) - \bar{f}^{(1)}(t)| + |(\frac{\bar{f}^{(1)}}{\bar{f}})(t)||\bar{f}(t+\gamma h) - \bar{f}(t)|)$$

(3.15)
$$= \frac{\gamma h}{\bar{f}(t+\gamma h)}(|\bar{f}^{(2)}(t+\gamma')| + |(\frac{\bar{f}^{(1)}}{\bar{f}})(t)||\bar{f}^{(1)}(t+\gamma'')|)$$

$$\le 4h(N\eta)^2$$

where the last inequality follows by applying (3.12) for $\nu = 2,1$, (3.13) and the fact that $|\bar{f}^{(1)}/\bar{f}| \le N$ and $\gamma < 2$.

Since the rhs of (3.15) is independent of t, the proof of the lemma is complete. ■

Observe that $e^{\alpha h} \le (\delta_j(x+h)/\delta_j(x)) \le e^{\beta h}$ for each $1 \le j \le i$. Therefore, Q is in $[\alpha,\beta]$. Since $\psi_i = \bar{f}^{(1)}/\bar{f}$ and $N = |\alpha| \vee |\beta|$, by (3.1) and Lemma 4 we get

(3.16)
$$|\psi_i - \hat{\psi}_{i+1}| \le |Q - (\hat{Q})_{\alpha,\beta}| + |Q - \psi_i| \le (|Q - \hat{Q}| \wedge 2N) + 4h(N\eta)^2.$$

Therefore, (3.16) followed by c_r-inequality (see Loève (1963), p. 155), Lemma 3 and (2.5) leads to

<u>Lemma 5</u>. $\forall\ \gamma \ge 0$

$$\underset{\sim}{P}_{i+1}|\psi_i(X) - \hat{\psi}_{i+1}(X)|^\gamma \le k_0'(\gamma)\{(ih^3)^{-\gamma/2}\mu((u^*/u_*^2)^{\gamma/2}q_\gamma) + (hN^2)^\gamma\}$$

where $k_0'(\gamma) = 2^{(\gamma-1)^+}\{(4\eta^2)^\gamma \vee k_0(\gamma)$ with k in $k_0(\gamma)$ replaced by $\inf_{x>a}(1 - h\eta u^*(x)m(x))\}$.

For the remainder of this section, let $c_0, c_1, \ldots$ denote <u>absolute</u> constants, and let

$$h \doteq h_i = c_0 i^{-1/5}.$$

We will now state and prove our main result of this section. Numbers $k_1, k_2, \ldots$ below are finite and independent of n.

<u>Theorem 1</u>. If $\forall\ i = 1,\ldots,n,$

(A1.0) $$h \eta u^* m \le 1 - c_1,$$

and if for a $\delta \in [0,1]$ $\exists$ a $\gamma \in [\delta,1]$ and k_1 and $k_2 \ni$ with $\xi^{-1} = 2 + \gamma$, $\forall\ i = 1,\ldots,n,$

(A1.1) $$\mu((u^*/u_*^2)^{\gamma/2} q_\gamma) \le k_1 N^{\gamma-1} h^{(\delta-\gamma)(1-\xi)},$$

and

(A1.2) $$N \le k_2 h^{\xi(\delta-\gamma)},$$

then $\exists$ a $k_3 \ni |D_n(\underset{\sim}{\omega},\underset{\sim}{\hat{\gamma}})| \le k_3 h_n^\delta$, uniformly in $\underset{\sim}{\omega} \in \times_1^n[\alpha_i,\beta_i]$.

<u>Remark 3.1</u>. Assumptions (A1.0), (A1.1) and (A1.2) together imply the existence of a $k_4 \ni \forall\ i = 1,\ldots,n,$

(3.17) $$\mu(m) \le k_4 N^{-2} h^{(8\delta-11\gamma)/6}.$$

To prove (3.17) we proceed as follows. Since $\eta \ge 1$, (A1.0) implies $\|um\|_\infty < h^{-1}$. Consequently, $f_\alpha \wedge f_\beta < (hu)^{-1}$. Therefore, since $m_{i+1} \ge m$ and since $(u^*/u_*^2) \ge (u)^{-1}$, from (2.5) the μ-integrand in (A1.1) is no less than $mh^{\gamma/2}$. Hence, by (A1.1)

$$(3.18)\qquad N^2\mu(m) \le k_1 N^{1+\gamma}\, h^{-(\gamma/2)+(\delta-\gamma)(1-\xi)}$$
$$\le k_4 h^{2(\delta-\gamma)(1-\xi)-(\gamma/2)},$$

where the second inequality follows from (A1.2) and the fact that $1+\gamma = (1-\xi)/\xi$. Since $1-\xi = ((1+\gamma)/(2+\gamma)) \le 2/3 \ \forall\ \gamma \in [0,1]$, (3.17) follows from (3.18). We will use (3.17) in the proof of the theorem here, and in a comparison (which we will make later) of the hypotheses of Theorem 1 with those of Theorem 2 in Section 4.

Remark 3.2. From (2.5) via the definition of m, and from the fact that $u^* \ge u \ge u_*$, the lhs of (A1.1) is bounded below by

$$(3.19)\quad \mu\Big(\frac{f_\beta}{(uf_\alpha)^{\gamma/2}}\Big) = \frac{C(\beta)}{C^{\gamma/2}(\alpha)} \int_a^\infty u^{1-\gamma/2} \exp((\beta - \gamma\alpha/2\,)x)\,dx.$$

Note that if $\beta - (\gamma\alpha/2) > 0$, then the rhs integral in (3.19) is not finite, unless $u(x) \to 0$ (as $x \to \infty$) at least as fast as $e^{-\ell x}$ for some $\ell > (2\beta - \gamma\alpha)/(2-\gamma)$, which holds, since $0 \le \gamma \le 1$, $\alpha < \beta$, α and $-\beta \downarrow$ in i, only if $N\ (= |\beta| \vee |\alpha|) = O(1)$. Thus, in such situations (A1.1) holds only if $N = O(1)$. It will be shown in the example following the proof of the theorem that, for the case $u(x) = (2\pi)^{-\frac{1}{2}} e^{-x^2/2} [-\infty < x < \infty]$, (A1.1) holds only if $N \uparrow \infty$ not faster than $(\log i)^{\frac{1}{2}}$. From these examples we conjecture here that for (A1.1) it is perhaps necessary that, for some τ_1, τ_2 non-negative, $N = O(1) + \tau_1 (\log i)^{\tau_2}$, whatever be the form of u. If this is true, then of course (A1.1) also implies (A1.2).

Proof of Theorem 1. By (A1.2), $\eta\ (= e^{2hN}) \downarrow 1$, and hence by (A1.0), k_0' in Lemma 5 is bounded in i. Now fix γ and δ satisfying the hypothesis of the theorem. The trivial bound

$|\hat{\psi}_{i+1} - \psi_i| \le 2N$ yields $\underset{\sim}{P}_{i+1}|\hat{\psi}_{i+1}(X) - \psi_i(X)| \le 2N^{1-\gamma}\underset{\sim}{P}_{i+1}|\hat{\psi}_{i+1}(X) - \psi_i(X)|^{\gamma}$. Therefore, since $h = i^{-1/5}$, Lemma 5 gives k_5 and $k_6 \ni \forall\ i = 1,\dots,n$

$$\text{(3.20)}\quad N_{i+1}\underset{\sim}{P}_{i+1}|\hat{\psi}_{i+1}(X) - \psi_i(X)| \le k_5 N_{i+1} N^{1-\gamma} i^{-\gamma/5}(\mu((u^*/u_*^2)^{\gamma/2} q_\gamma) + N^{2\gamma})$$
$$\le k_6 i^{-\delta/5},$$

where the second inequality follows from (A1.1) and (A1.2).

Since X abbreviates X_{i+1} and (3.20) holds for each $1 \le i \le n-1$, $n^{-1}\Sigma_2^n N_j \underset{\sim}{P}_j|\psi_{j-1}(X_j) - \hat{\psi}_{j-1}(X_j)| \le k_6 n^{-1}\Sigma_1^{n-1} j^{-\gamma/5}$. Thus, the first term on the rhs of the inequality in Lemma 2 with $\underset{\sim}{\varphi}$ there replaced by $\underset{\sim}{\hat{\psi}}$ is bounded above by $(k_3/2)h_n^{\delta}$ uniformly in $\underset{\sim}{\omega} \in X_1^n[\alpha_i,\beta_i]$, and so is the second term there by (3.17), since k_4 is independent of i and n. ■

Now we will show how the conditions of Theorem 1 reduce to a single condition on N when the family of densities involved is normal.

<u>Example $N(\omega,1)$</u>. Let $u(x) = (2\pi)^{-\frac{1}{2}} e^{-x^2/2}[-\infty < x < \infty]$. Then $a = -\infty$, $C(\omega) = e^{-\omega^2/2}$ and $\Omega = R$. Let $-\alpha = \beta = N > 0$. We will show that all the assumptions of Theorem 1, with $\gamma = 1$ and any fixed $\delta \in [0,1]$, are satisfied iff

$$\text{(3.21)}\qquad N \doteq N_i = O(1) + \left(\frac{1-\delta}{15}\log i\right)^{\frac{1}{2}}.$$

We will first prove the 'if' part. Clearly (A1.2) holds. Considering the upper and lower bounds for the ratio $u(t)/u(x)$ for $x \le t < x + 2h$, we get $u^*(x) \le u(x)e^{2h|x|}$ and $u_*(x) \ge u(x)e^{-2h(|x|+h)}$. Therefore,

$$(3.22)\quad u^*(x)f_\omega(x) \le e^{2h|x|}u(x)f_\omega(x) = (2\pi)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}((|x|-\omega \operatorname{sgn} x)^2-4h|x|)\}$$

$$\le \exp\{2h(h+\omega \operatorname{sgn} x)\}.$$

By (3.22) $u^*m \doteq u^*\sup_{|\omega|\le N} f_\omega \le \exp(2h^2+2hN)$. Therefore, since $h = c_0 i^{-1/5}$ and hence hN is bounded uniformly wrt c_0 in a neighborhood of zero, by a suitable choice of c_0, (A1.0) holds.

For this paragraph only, let N abbreviate N_{i+1} (instead of N_i). Now observe that

$$m_{i+1}(x) \doteq \sup_{|\omega|\le N} f_\omega(x) \le \exp(x^2/2)[|x| \le N] + \exp(N|x|-N^2/2)[|x|>N].$$

Therefore, since $f_\alpha(x) \wedge f_\beta(x) \ge \exp(-N|x| - N^2/2)$, by (2.5)

$$(3.23)\quad q_1(x) \le e^{\frac{1}{4}N^2+\frac{1}{2}N|x|}(e^{\frac{1}{2}x^2}[|x| \le N] + e^{-\frac{1}{2}N^2+N|x|}[|x| > N]).$$

Moreover, using the bounds obtained above for u_* and u^* we get

$$(3.24)\quad \left(\frac{u^*}{u_*^2}(x)\right)^{\frac{1}{2}} \le (2\pi)^{\frac{1}{4}}\exp(\tfrac{1}{4}x^2 + 3h|x| + 2h^2)\ .$$

Thus by (3.23) and (3.24),

$$\mu\{((u^*/u_*^2)^{\frac{1}{2}}q_1)[|x| \le N]\} \le 2N\exp(N^2+3hN+2h^2)$$

and

$$\mu\{((u^*/u_*^2)^{\frac{1}{2}}q_1)[|x| > N]\} \le (2\pi)^{-\frac{1}{4}}\exp(-\frac{N^2}{4}+2h^2)\int\exp(-\tfrac{1}{4}x^2 + 3(h+\frac{N}{2})|x|)dx$$

$$\le c_2\exp(2N^2+9hN)\ .$$

Consequently, since hN is bounded, we get

$$(3.25)\quad \mu((u^*/u_*^2)^{\frac{1}{2}}q_1) \le c_3 e^{2N^2}\ .$$

Thus, by (3.21) and (3.25), (A1.1) holds with $\gamma = 1$.

Conversely, by Remark 2.2 we note that the lhs of (A1.1) is bounded below by the rhs of (3.19). Therefore, since $\beta = -\alpha = N$, with $\gamma = 1$,

$$(3.25)' \quad \mu((u^*/u_*^2)^{\frac{1}{2}}q_1) \geq (2\pi)^{-\frac{1}{2}}\exp(-\frac{N^2}{4})\int\exp(-\frac{x^2}{4}+\frac{3Nx}{2})dx$$
$$= (8\pi)^{\frac{1}{2}}e^{2N^2}.$$

Thus (A1.1) with $\gamma = 1$ and any $\delta \in [0,1]$ holds only if $e^{2N^2} \leq (8\pi)^{-\frac{1}{2}}k_1h^{2(\delta-1)/3}$, or only if (3.21) holds. ■

The following corollary, which is a consequence of Theorem 1, asserts that for certain families of densities, $D_n(\underset{\sim}{\omega},\underset{\sim}{\hat{\psi}}) = O(h_n)$ uniformly in $\underset{\sim}{\omega} \in \Omega^n$. It also shows how the condition (A1.1) of the theorem is simplified in fixed N case. Recall from (2.4) and the definition following (3.1), that $u_*(\doteq u_{2h})$ and u^* depend on i, $1 \leq i \leq n-1$.

Corollary 1. Let α and β be constants wrt i. If μ is such that (A1.0) holds, and for a $\delta \in [0,1]$, with $w(\alpha,\beta) = \alpha - (\delta\beta/2)$,

$$(3.26) \quad \mu\{(\exp(xw(\alpha,\beta))[x \leq 0] + \exp(xw(\beta,\alpha))[x > 0])(\frac{u^*}{u_*^2})^{\delta/2}\} < \infty$$

for $i = 1$, (e.g., take any δ in $[0,1]$ with $\delta < 2\beta/\alpha$, and $u(x) = x^{\tau-1}[x > 0]$, $\tau \geq 1$ or $\Sigma_0^\infty(j+1)[j \leq x < j+1]$), then $D_n(\underset{\sim}{\omega},\underset{\sim}{\hat{\psi}}) = O(h_n^\delta)$ uniformly in $\underset{\sim}{\omega} \in [\alpha,\beta]^n$.

Proof. Since $N \equiv |\alpha| \vee |\beta|$ is constant wrt i, (A1.2) holds with $\gamma = \delta$. $C(\omega)$ is clearly bounded away from 0 and ∞ on $[\alpha,\beta]$. Hence, since $m \doteq \sup_{|\omega|\leq N} f_\omega$, $q_\delta \doteq (m_{i+1}/(f_\alpha \wedge f_\beta)^{\delta/2}) \leq c_4(u_*^2/u^*)^{\delta/2}$ times the μ-integrand in (3.26). Thus, since $u^*\downarrow$

and $u_*\uparrow$ in i, (A1.1) with $\gamma = \delta$ holds by (3.26). ■

As a final comment to this section, we state here that the rate 1/5, that is shown to be achieved, under certain conditions, by $\hat{\underset{\sim}{\psi}}$, is perhaps much slower than that $\hat{\underset{\sim}{\psi}}$ could actually achieve. This will be supplemented in Section 5, by showing that $D_n(\underset{\sim}{\omega},\hat{\underset{\sim}{\psi}}) = O(n^{-(2/5)+})$ $\forall\, \underset{\sim}{\omega} \in \Omega^n$ with identical components.

2.4 Kernel Estimators with Rates Near ½ when $\theta(\omega) = \omega$, e^{ω} or ω^{-1}.

In this section we consider the situations where $\theta(\omega)$ is ω, e^{ω} or ω^{-1}. In each of these cases we will exhibit, for each $\epsilon > 0$, a class of compound estimators with rates $(1-\epsilon)/2$ uniformly in $\underset{\sim}{\omega} \in \times_1^n[\alpha_i,\beta_i]$.

The classes of compound estimators to be exhibited in this section are based on types of kernel functions introduced by Johns and Van Ryzin (1972) in Empirical Bayes Linear loss two-action problems in exponential families. Thus in this section we will have two sets of assumptions; the one, which was not needed in Section 2.3, involves the kernel functions defining the classes of estimators, and the other involves the family of densities.

Recall from the latter part of Section 2 that the dependence of h,α,β,N,m and q_γ on i is abbreviated by omission, where i is fixed with $1 \le i \le n-1$. For $\nu = 0,1$ and integer $r > \nu$, let $\mathcal{K}_\nu^r$ be defined as in Section 1 of Chapter 1. As in Chapter 1, with a fixed $K_\nu \in \mathcal{K}_\nu^r$, define

$$(4.0) \qquad \hat{f}^{(\nu)}(\cdot) = (ih^{\nu+1})^{-1}\Sigma_1^i\{(K_\nu((X_j - \cdot)/h)/u(X_j))[u(X_j) > 0]\} .$$

Estimation of ψ_i (and hence exhibition of compound estimators) in this section, involves estimation of one or both of the functions $\bar{f}$ and $\bar{f}^{(1)}$. It has been seen in Chapter 1 that $\hat{\bar{f}}^{(\nu)}$, as an estimator of $\bar{f}^{(\nu)}$, has various asymptotic properties. We will make , according to our need, applications of one or both of the functions $\hat{\bar{f}}^{(0)}$ and $\hat{\bar{f}}^{(1)}$ in defining our compound estimators here.

The reason we have taken here $r > \nu$ (instead of $r \geq \nu$, as is taken in Chapter 1) is that $\mathcal{K}_\nu^r$, for any integer $r > \nu$, is non-empty and $\bar{f}$ here is (infinitely) differentiable. Moreover, we assume here that $K_\nu \in L_2(0,1)$, (e.g., K_ν could be the ν-th element of the dual basis for the subspace of $L_2(0,1)$ with basis $\{1, y, \ldots, y^{r-1}\}$). Denote $\int y^j |K_\nu(y)| dy/j!$ by $|k|_{j,\nu}$.

Let $s = |\alpha| \vee |\beta|$. (In case θ is identity map, $s = N$.) Since $\max_{1\leq j\leq i}|\omega_j| \leq s$, by mean value theorem, $\int_x^{x+h} |\bar{f}^{(r)}| dy \leq hs^r e^{hs}\bar{f}$. Hence, (2.8) of Chapter 1 gives

$$(4.1) \qquad |\underset{\sim}{P}_i \hat{\bar{f}}^{(\nu)} - \bar{f}^{(\nu)}| \leq |k|_{r-1,\nu} h^{r-\nu} s^r e^{hs} \bar{f} \ .$$

Moreover, since (Lebesgue) $\text{ess-sup}_{0\leq t\leq 1}(\bar{f}/u)(\cdot + ht) \leq e^{hs}(\bar{f}/u_h)(\cdot)$, the inequality in (4.0) of Chapter 1 followed by the equation in the proof of Lemma 1 there, gives

$$(4.2) \qquad \text{var}(\hat{\bar{f}}^{(\nu)}) \leq \bar{f} e^{hs}(ih^{2\nu+1}u_h)^{-1} \|K_\nu\|_2^2 \ .$$

As in Remark 5.2, and in Inequality (5.18), both of Chapter 1, a choice of h, that balances the two terms $h^{r-\nu}$ and $(ih^{2\nu+1})^{-\frac{1}{2}}$ appearing in the bounds for the bias in (4.1) and the standard deviation in (4.2) of the estimator $\hat{\bar{f}}^{(\nu)}$, is

$$(4.3) \qquad h = i^{-1/(1+2r)} .$$

This choice of h has been adopted by various authors, (e.g., Susarla (1970, Theorem 2; Yu (1970), Theorems 1.1 and 2.1; Johns and Van Ryzin (1972), Theorems 3 and 4), working on certain problems utilizing kernel estimators of a density or of its derivative. We too adopt (4.3) throughout this section and in Theorem 6 of the next section.

For $0 < \gamma \le 2$, let $M_{\gamma,\nu}$ be the γ-th mean error of $\hat{\bar{f}}^{(\nu)}$, i.e., $M_{\gamma,\nu} = \underset{\sim}{P}_i|\hat{\bar{f}}^{(\nu)} - \bar{f}^{(\nu)}|^{\gamma}$. Then, since $M_{2,\nu}$ = (lhs of (4.1))^2 + lhs of (4.2), Liapounov's inequality followed by c_r-inequality and (4.3) yields

$$(4.4) \qquad M_{\gamma,\nu} \le c_\nu(\gamma) h^{\gamma(r-\nu)}\{(s^r\bar{f})^{\gamma} + (\bar{f}/u_h)^{\gamma/2}\}$$

where

$$c_\nu(\gamma) = (|k|_{r-1,\nu}e^{hs})^{\gamma} \vee (e^{hs}\|K_\nu\|_2^2)^{\gamma/2} .$$

Hereafter, we abbreviate $\hat{\bar{f}}^{(\nu)}$ by $\hat{\bar{f}}$ and $M_{\gamma,0}$ by M_γ. Recall that X abbreviates X_{i+1}. Inequality (4.4) will be used in obtaining an upper bound for $\underset{\sim}{P}_{i+1}|\varphi_{i+1}(X) - \psi_i(X)|$ where φ_j's are yet to be defined. Let $c_1, c_2, \ldots$ denote absolute constants. We now discuss the three cases separately.

Case ω. We consider here the case when θ is the identity map. Since $f_j(\cdot) = C(\omega_j)\exp(\omega_j\cdot)$, (1.0) specialized to the case $\theta(\omega_j) = \omega_j$ yields

$$(4.5) \qquad \psi_i = (\bar{f}^{(1)}/\bar{f}).$$

Since $\mathcal{K}_\nu^r$ is non-empty for any $r > \nu$, and $\bar{f}$ here is (infinitely) differentiable, we restrict, throughout our discussion of this case, r in (4.0) to be at least 2. Define a compound estimator

$\hat{\underset{\sim}{\psi}}$ with its (i+1)st component, $\hat{\psi}_{i+1}(X)$, given by

(4.6) $$\hat{\psi}_{i+1} = (\hat{\bar{f}}^{(1)}/\hat{\bar{f}})_{\alpha,\beta},$$

where $\hat{\bar{f}}^{(\nu)}$ is given by (4.0) with $h = i^{-1/(1+2r)}$. Define $H = H_i$ by

(4.7) $$H = h^{r-1} \doteq i^{-(r-1)/(1+2r)}.$$

Recall that $N = s = |\alpha| \vee |\beta|$ and each of α,β and N hides subscript i. The following lemma which plays the central role in proving Theorem 2 below is a consequence of (4.4) and Lemma A.2 of the appendix.

Lemma 6. $\forall$ $p > 0$ and $0 < \gamma \le p \wedge 2$,

(4.8) $$\underset{\sim}{P}_{i+1}|\psi_i(X) - \hat{\psi}_{i+1}(X)|^p \le B(p)N^{p-\gamma}H^\gamma(N^{r\gamma} + \mu(q_\gamma/u_h^{\gamma/2}))$$

where $B(p) = 2^{p+(\gamma-1)^+}(1 + (hN)^\gamma(1+2^\gamma))\max_{\nu=0,1}C_\nu(\gamma)$.

Proof. Fix $0 < \gamma \le p \wedge 2$. Since ψ_i and $\hat{\psi}_{i+1}$ are in $[\alpha,\beta]$ and $N = |\alpha| \vee |\beta|$, by (4.5) and (4.6), $|\psi_i - \hat{\psi}_{i+1}| \le 2N$. Consequently, Lemma A.2 of the Appendix and the definitions of $M_{\gamma,\nu}$ yield

(4.9) $$\underset{\sim}{P}_i|\psi_i - \hat{\psi}_i|^p \le 2^{p+(\gamma-1)^+}N^{p-\gamma}(\bar{f})^{-\gamma}(M_{\gamma,1} + (1+2^\gamma)N^\gamma M_{\gamma,0}).$$

Since $s = N$, by (4.3) and (4.4), $h^{(\nu-1)\gamma}M_{\gamma,\nu}$ is bounded above by $C_\nu(\gamma)(H\bar{f})^\gamma(N^{r\gamma} + (\bar{f}u_h)^{-\gamma/2})$. Consequently, the rhs of (4.9) is bounded above by $B(p)N^{p-\gamma}H^\gamma(N^{r\gamma} + (\bar{f}u_h)^{-\gamma/2})$, where $B(p)$ is as given in the lemma.

Since X abbreviates X_{i+1}, taking expectation wrt P_{i+1} on both sides of the inequality just obtained we get the desired conclusion from the definition of q_γ given in (2.5). ■

Lemma 6 with $p = 1$ will be used to prove our main result below. The numbers $b_0, b_1, \ldots$ below are finite and independent of n.

Theorem 2. Recall from (4.7) that $H = h^{r-1} = i^{-(r-1)/(1+2r)}$. If for a $\delta \in [0,1]$ $\exists$ a $b_0 \ni \forall\ i = 1,\ldots,n$,

(A2.0) $$\mu(m) \le b_0 i H^\delta / (N^2(1 + \log i)),$$

and if $\exists$ a $\gamma \in [\delta,1]$ and b_1 and $b_2 \ni$ with $\xi^{-1} = 2 + \gamma(r-1)$,

(A2.1) $$\mu(q_\gamma / u_h^{\gamma/2}) \le b_1 N^{\gamma-1} H^{(\delta-\gamma)(1-\xi)} \quad \forall\ i = 1,\ldots,n-1,$$

and

(A2.2) $$N \le b_2 H^{\xi(\delta-\gamma)} \qquad \forall\ i = 1,\ldots,n,$$

then $\exists$ a $b_3 \ni |D_n(\underset{\sim}{\omega},\underset{\sim}{\hat{v}})| \le b_3 H_n^\delta$ uniformly in $\underset{\sim}{\omega} \in \times_1^n[\alpha_i,\beta_i]$.

Remark 4.1. For $r = 2$, (A2.2) is equivalent to (A1.2), while (A2.1) is implied by (A1.1). Moreover, since the rhs in (A2.0) is no less than $b_0 i^{1-\delta/2}/(N^2(1 + \log i))$, by Remark 2.1 via (3.17) there, for each $r \ge 2$ (A2.0) is implied by (A1.0), (A1.1) and (A1.2) together. Thus assumptions of Theorem 1 are stronger than those of Theorem 2,at least for $r = 2$.

Remark 4.2. By (2.5) via the definition of m, the lhs of (A2.1) is no less than

(4.10) $$\mu\Big(\frac{f_\beta}{(uf_\alpha)^{\gamma/2}}\Big) = \frac{C(\beta)}{C^{\gamma/2}(\alpha)} \int_a^\infty u^{1-\gamma/2} e^{(\beta-\gamma\alpha/2)x} dx .$$

Equation (4.10) is the same as (3.19). Hence the comments in Remark 2.2, regarding possible necessary conditions for the finiteness of the integral on the rhs of (4.10) (and hence for (A2.1)), remain valid here too.

<u>Proof of Theorem 2</u>. Since $h^{r-1} \doteq H$ and $(\gamma-\delta)\xi < (r-1)^{-1}$, by (A2.2) $hN \downarrow 0$. Therefore, since $s = N$, $c_\nu(\gamma)$ in (4.4) is bounded in i, and so is $B(1)$ in Lemma 6. Consequently, Lemma 6 with $p = 1$ gives a b_4 such that

$$(4.11)\quad N_{i+1}\underset{\sim}{P}_{i+1}|\psi_i(X) - \hat{\psi}_{i+1}(X)| \le b_4 N_{i+1} N^{1-\gamma} H^\gamma (N^{r\gamma} + \mu(q_\gamma/u_h^{\gamma/2}))$$
$$\le b_5 i^{-\delta(r-1)/(1+2r)}$$

where, remembering $\xi^{-1} \doteq 2 + \gamma(r-1)$, the last inequality follows from (A2.1) and (A2.2).

Since (4.11) holds for each $1 \le i \le n-1$, and X there abbreviates X_{i+1}, $n^{-1}\Sigma_2^n N_i \underset{\sim}{P}_i |\psi_{i-1}(X_i) - \hat{\psi}_i(X_i)| \le b_5 n^{-1}\Sigma_1^{n-1} i^{-\delta(r-1)/(1+2r)}$. Thus, the first term on the rhs of the inequality in Lemma 2, with $\underset{\sim}{\varphi}$ there replaced by $\hat{\underset{\sim}{\psi}}$, is no more than $(b_3/2)H_n^\delta$ uniformly in $\underset{\sim}{\omega} \in X_1^n[\alpha_i,\beta_i]$, and so is the second term there by (A2.0), since b_0 is independent of i and n. ■

The hypotheses of Theorem 2 are satisfied for many exponential families. In Example $N(\omega,1)$, introduced in Section 2, we will show that all the assumptions of Theorem 2, with $-\alpha = \beta = N>0$, $\gamma = 1$ and any fixed $\delta \in [0,1]$ are satisfied iff

$$(4.12)\qquad N \doteq N_i = O(1) + \left(\frac{r(r-1)(1-\delta)}{2(1+r)(1+2r)} \log i\right)^{\frac{1}{2}}.$$

Note that for the case $r = 2$, (4.12) is the same as (3.21).

Since the lhs of (3.25) is bounded below by $\mu(q_1/u_h^{\frac{1}{2}})$, we have from there

(4.13) $$\mu(q_1/u_h^{\frac{1}{2}}) \le c_1 e^{2N^2}.$$

Thus, if (4.12) holds, then, with $\gamma = 1$, (A2.2) holds; from (4.13), (A2.1) holds; and, from the fact that $m(x) \doteq \sup_{|\omega|\le N} f_\omega(x) \le e^{N|x|}$ implies $\mu(m) \le e^{N^2/2}$, (A2.0) holds.

On the other hand, we have noted in Remark 3.2 that the lhs of (A2.1) is no less than the rhs in (4.10). Therefore, with $\gamma = 1$

(4.14) $$\mu(q_1/u_h^{\frac{1}{2}}) \ge (2\pi)^{-\frac{1}{4}} e^{-N^2/4} \int e^{-x^2/4 + \frac{3Nx}{2}} dx = (8\pi)^{\frac{1}{4}} e^{2N^2}.$$

Hence, (A2.1) with $\gamma = 1$ holds only if (4.12) holds. ∎

Remark 4.3. Theorem 2, specialized to the above Example $N(\omega,1)$, improves the univariate version of the result in Theorem 3 of Susarla (1970). We have shown, through a simpler and shorter proof, the existence of less restrictive kernel estimators with rates $(r-1)\delta/(1+2r)$ where δ is given by (4.12). Our rates are strictly higher than the rates $(r-1)/(4+2r)$ shown to be achieved by his kernel estimators in the bounded N-case, provided, in our unbounded N-case, N satisfies (4.12) with some $1 > \delta > (1+2r)/(4+2r)$. Note that the number of restrictions on the kernel functions increases as r increases.

The following corollary shows how the conditions of Theorem 2 are simplified greatly in fixed N case. From (2.4) and (4.3), remember that u_h depends on i with $1 \le i \le n-1$.

Corollary 2. Let α and β be constants wrt i. If for a $\delta \in [0,1]$, and with $w(\alpha,\beta) = \alpha - (\delta\beta/2)$,

(4.15) $$\mu\{(\exp(xw(\alpha,\beta))[x \le 0] + \exp(xw(\beta,\alpha))[x > 0])/u_h^{\delta/2}\} < \infty$$

for $i = 1$, (e.g., take any δ in $[0,1]$ with $\delta < 2\beta/\alpha$, and $u(x) = x^{\tau-1}[x > 0]$, $\tau \geq 1$ or $\Sigma_0^\infty(i+1)[i \leq x < i+1]$), then $D_n(\underset{\sim}{\omega},\hat{\underset{\sim}{\psi}}) = O(H_n^\delta)$ uniformly in $\underset{\sim}{\omega} \in [\alpha,\beta]^n$.

Proof. Since $N \equiv |\alpha| \vee |\beta|$ is constant wrt i, (A2.2) holds for $\gamma = \delta$. $C(\omega)$ is clearly bounded away from 0 and ∞ on $[\alpha,\beta]$. Therefore, $m(x) \doteq \sup_{|\omega|\leq N} f_\omega(x) \leq c_2(\exp(\beta x)[x > 0] + \exp(\alpha x)[x \leq 0])$, hence (A2.0) holds since α and β are in Ω. Moreover, since $(f_\beta \wedge f_\alpha)(x) = C(\omega)(\exp(\alpha x)[x > 0] + \exp(\beta x)[x \leq 0])$, $q_\delta \doteq m_{i+1}/(f_\alpha \wedge f_\beta)^{\delta/2} \leq b_6 u_h^{\delta/2}$ times the μ-integrand in (4.15). Thus since $u_h \uparrow$ in i, (A2.1) holds for $\gamma = \delta$ by (4.15). ■

In the next section we will show that $\hat{\underset{\sim}{\psi}}$ perhaps could achieve rates much higher than those indicated in Theorem 2. Specifically, we will show that, when $\underset{\sim}{\omega}$ has identical components, $c_3 H_n^2 \leq D_n(\underset{\sim}{\omega},\hat{\underset{\sim}{\psi}}) \leq c_4 H_n^{2-}$.

From (4.3) recall that $h = i^{-1/(1+2r)}$. For the remainder of this section, redefine H by

(4.7)' $$H = h^r \doteq i^{-r/(1+2r)}.$$

Case e^ω. We consider here the case when $\theta(\omega) = e^\omega$. Only for the purpose of our discussion here, let ' on a function indicate its translation to the right by 1, e.g., $\bar{f}'(x) = \bar{f}(x+1)$. Since $f_j(\cdot) = C(\omega_j)\exp(\omega_j\cdot)$, specialization of (1.0) to the case $\theta(\omega_j) = \exp(\omega_j)$ gives

(4.16) $$\psi_i = \frac{\bar{f}'}{\bar{f}}.$$

Taking r in (4.0) at least 1, define a compound estimator $\hat{\underset{\sim}{\psi}}$

with its (i+1)st component given by

(4.17) $$\hat{\psi}_{i+1} = (\hat{\bar{f}}'/\hat{\bar{f}})_{e^\alpha, e^\beta}.$$

Recall that $s \doteq |\alpha| \vee |\beta|$ and $N \doteq e^\alpha \vee e^\beta = e^\beta$, and each of α, β, s and N hides subscript i.

Now using Lemma A.2 of the appendix, and (4.4) with $\nu = 0$, we prove

<u>Lemma 7</u>. $\forall\ p > 0$ and $0 < \gamma \le p \wedge 2$,

$$P_{i+1}(|\psi_i(X) - \hat{\psi}_{i+1}(X)|^p) \le B(p)H^\gamma N^p\{s^{r\gamma} + N^{-\gamma/2}\mu\{q_\gamma((u_h)^{-\gamma/2} + (u_h')^{-\gamma/2})\}\},$$

where $B(p) = 2^{p+(\gamma-1)^+}(2 + 2^\gamma)c_0(\gamma)$.

<u>Proof</u>. Since $\omega_j \in [\alpha, \beta]$ for all $1 \le j \le i$, by (1.0) $e^\alpha \le \psi \le e^\beta$. Thus (4.16) and (4.17) followed by the fact that $N = e^\beta$ give $|\psi_i - \hat{\psi}_{i+1}| \le 2N$. Therefore, since $(\bar{f}'/\bar{f}) \le N$, it follows from Lemma A.2 of the appendix, that

(4.18) $$P_i|\psi_i - \hat{\psi}_{i+1}|^p \le 2^{p+(\gamma-1)^+}(\bar{f})^{-\gamma}N^{p-\gamma}\{M_\gamma' + (1+2^\gamma)N^\gamma M_\gamma\}.$$

Since $\bar{f}' \le N\bar{f}$, from (4.4) with $\nu = 0$, $M_\gamma' \le c_0(\gamma)H^\gamma\{(Ns^r\bar{f})^\gamma + (N\bar{f}/u_h')^{\gamma/2}\}$. Consequently, by (4.4) with $\nu = 0$,

$$\text{rhs of (4.18)} \le B(p)N^pH^\gamma\{s^{r\gamma} + (N\bar{f})^{-\gamma/2}((u_h)^{-\gamma/2} + (u_h')^{-\gamma/2})\},$$

where $B(p)$ is as given in the lemma. Now (4.18) followed by the preceding inequality and the definition of q_γ in (2.5) yield the desired conclusion.■

Let the numbers $b_0, b_1, \ldots$ below be finite and independent of n. Now we will obtain the main result for the case under study.

Theorem 3. Recall from (4.7)' that $H = h^r = i^{-r/(1+2r)}$.
If for a $\delta \in [0,1]$ $\exists$ a $b_0 \ni \forall\ i = 1,\ldots,n$,

(A3.0) $$\mu(m) \leq b_0 iH^{\delta}/N^2(1 + \log i),$$

and if $\exists$ a $\gamma \in [\delta,1]$ and b_1 and $b_2 \ni \forall\ i = 1,\ldots,n-1$,

(A3.1) $$\mu\{q_\gamma((u_h)^{-\gamma/2} + (u_h')^{-\gamma/2})\} \leq b_1 N_{i+1}^{-2+\gamma/2} H^{\delta-\gamma}$$

and

(A3.2) $$s^{r\gamma} \leq b_2 N_{i+1}^{-2} H^{\delta-\gamma},$$

then $\exists$ a $b_3 \ni |D_n(\underset{\sim}{\omega},\underset{\sim}{\hat{\psi}})| \leq b_3 H_n^{\delta}$ uniformly in $\underset{\sim}{\omega} \in X_1^n[\alpha_i,\beta_i]$.

Remark 4.4. If $\Omega \leq (-\infty,0]$, then taking $\beta \equiv 0$ (implying $N \doteq e^{\beta} \equiv 1$) (A3.2) becomes $|\alpha|^{r\gamma} \leq b_2 H^{\delta-\gamma}$. In general, (A3.2) holds if $|\alpha| \vee |\beta| \uparrow \infty$ at rates not faster than $b_4 \log i$ for some $2b_4 < r/(1+2r)$. Keeping the difference in H's in two Cases ω and e^{ω} in mind, we see that (A3.0) implies (A2.0). Finally, since the lhs of (A3.1) is bounded below by the lhs of (A2.1), comments, quite similar to those given in Remark 3.2 regarding possible necessary conditions on α and β for (A3.1), can be stated here too.

Proof of Theorem 3. Since $h = i^{-1/1+2r}$, $H = h^r$ and $N_{i+1} \geq N_1$, (A3.2) implies $hs \to 0$. Hence $c(\gamma)$ in (4.4) is bounded in i, and so is B(1) in Lemma 7. Consequently, since $N\ (\doteq N_i) \leq N_{i+1}$, Lemma 7 with $p = 1$, and the hypotheses (A3.1) and (A3.2) yield a b_5 such that $\forall\ i = 1,\ldots,n-1$,

(4.19) $$N_{i+1} \underset{\sim}{P}_{i+1} |\psi_i(X) - \hat{\psi}_{i+1}(X)| \leq b_5 i^{-\delta r/(1+2r)}.$$

In view of (4.19) and (A3.0), the remainder of the proof follows by arguments identical to those given in the second paragraph of the proof of Theorem 2. ■

Remark 4.5. Theorem 3 here improves Theorem 6 of Samuel (1965). Restricting ω's to a bounded interval of Ω, she exhibits estimators $\underset{\sim}{\varphi}^*$ and shows that, under certain conditions, $\forall\ \varepsilon > 0$ $D_n(\underset{\sim}{\omega},\underset{\sim}{\varphi}^*) < \varepsilon\ \forall\ n \geq$ some $n_0(\omega,\varepsilon) < \infty$. We do not require her continuity assumption on u; and her other hypotheses always imply ours, (this may readily be seen through Corollary 3 below).

By analyses analogous to those made earlier in Example $N(\omega,1)$, it can be verified that the hypotheses of Theorem 3, with $\gamma = 1$, $\delta \in [0,1]$ and $-\alpha = \beta > 0$ (so that $s = \beta$ and $N = \exp(\beta) \geq 1$) are satisfied iff

$$\beta = O(1) + \left(\frac{(1-\delta)r}{2(1+2r)} \log i\right)^{\frac{1}{2}}.$$

The hypotheses of Theorem 3 reduce to a rather simple one in the fixed N case, as we can see in the following corollary.

Corollary 3. Let α and β be constants wrt i. If for a $\delta \in [0,1]$, (4.15) with $u_h^{\delta/2}$ there replaced by $u_h^{\delta/2} + (u_h')^{\delta/2}$ holds, (e.g., take examples mentioned in Corollary 2), then $D_n(\underset{\sim}{\omega},\underset{\sim}{\hat{\psi}}) = O(H_n^{\delta})$ uniformly in $\underset{\sim}{\omega} \in [\alpha,\beta]^n$.

Proof. The proof is identical to the one given for Corollary 3. ■

In the next section we will point out that, in certain cases, $D_n(\underset{\sim}{\omega},\underset{\sim}{\hat{\psi}})$ are $O(H_n^{2-})$. Thus one may expect $\underset{\sim}{\hat{\psi}}$ to achieve rates much higher than those indicated in Theorem 3.

Case ω^{-1}. We now consider the situation where the component problem is SELE of $\theta(\omega) = \omega^{-1}$.

One of the important examples, where such estimation arises, is sequence compound SELE of scale λ in $\Gamma(\lambda,\tau)$-family: $(\Gamma(\tau))^{-1}x^{\tau-1}\lambda^{-\tau}e^{-x/\lambda}[x,\tau > 0]$. This of course includes the case of sequence compound SELE of σ^2 in $N(0,\sigma^2)$-family: $(2\pi\sigma^2)^{-\frac{1}{2}}\exp(-x^2/\sigma^2)[-\infty < x < \infty]$, since x^2 is sufficient for σ^2.

Throughout the study of this case, we <u>assume</u> that $\beta_n < 0$. (Thus $\beta < 0 \ \forall \ i = 1,\ldots,n)$. Since $f_j(x) = C(\omega_j)e^{\omega_j x}$ and, for $j = 1,\ldots,i$, $\omega_j < 0$, $\omega_j^{-1}f_j(x) = -\int_x^\infty f_j$. Thus specialization of (1.0) to $\theta_j \doteq \theta(\omega_j) = \omega_j^{-1}$ gives

$$(4.20) \qquad \psi_i(x) = -(\int_x^\infty \bar{f})/\bar{f}(x).$$

Since $\alpha \le \omega_j \le \beta < 0 \ \forall \ j = 1,\ldots,i$, by (1.0) $\beta^{-1} \le \psi_i \le \alpha^{-1}$.

For the remainder of this section, let $\hat{\bar{f}}$ be given by (4.0) <u>with</u> $r > 0$, and let L denote $\beta^{-1}\log H$. Motivated by (4.20), our proposed compound estimator has (i+1)st component

$$(4.21) \qquad \hat{\psi}_{i+1}(X) = (-(\int_X^{X+L} \hat{\bar{f}})/\hat{\bar{f}}(X))_{\beta^{-1},\alpha^{-1}}.$$

From (4.3) and (4.7)' recall that $H \doteq h^r \doteq i^{-r/(1+2r)}$. Also, note that $s \doteq |\alpha| \vee |\beta| = |\alpha|$ and $N \doteq \sup\{|\omega^{-1}| \,|\, \alpha \le \omega \le \beta\} = |\beta|^{-1}$.

<u>Lemma 8</u>. $\forall \ p > 0$ and $0 < \gamma \le p \wedge 1$,

$$(4.22) \qquad \underset{\sim}{P}_{i+1}|\psi_i(X) - \hat{\psi}_{i+1}(X)|^p \le B(p)H^\gamma|\beta|^{-p}\{|\alpha|^{r\gamma} + 1 + (|\log H|^{\gamma/2} + 1)\mu(q_\gamma/u_{h+L}^{\gamma/2})\}$$

where $B(p) = 2^p \cdot 3(c_0(\gamma) \vee 1)$ with $c_0(\gamma)$ given by (4.4).

Proof. Fix $0 < \gamma \le p \wedge 1$. For $j = 1,\ldots,i$, $\int_{x+L}^{\infty} f_j = -\omega_j^{-1} e^{L\omega_j} f_j(x) \le |\beta|^{-1} H f_j(x)$, since $L\beta = \log H$ and $\alpha \le \omega_j \le \beta < 0$. Therefore,

$$\int_{x+L}^{\infty} \bar{f} \le |\beta|^{-1} H\, \bar{f}(x) \ . \tag{4.23}$$

As in Case e^{ω}, abbreviate $M_{1,0}$, introduced preceding (4.4), to M_1. Now (4.4), the inequality $u_{h+L}(x) \le u_h(t)\ \forall\, x \le t < x+L$, and Schwarz inequality give

$$\begin{aligned} \underset{\sim}{P}_i \int_x^{x+L} |\bar{f} = \hat{\bar{f}}| &= \int_x^{x+L} M_1 \le c_0(1) H\{\int_x^{x+L} (|\alpha|^r \bar{f} + (\bar{f}/u_h)^{\frac{1}{2}})\} \\ &\le c_0(1) H\{|\alpha|^r \int_x^{x+L} \bar{f} + (L(\int_x^{x+L} \bar{f})/u_{h+L}(x)))^{\frac{1}{2}}\} \\ &\le c_0(1) H\{|\alpha^r \beta^{-1}| \bar{f}(x) + (L\bar{f}(x)/(|\beta| u_{h+L}(x)))^{\frac{1}{2}}\} \end{aligned} \tag{4.24}$$

since $\int_x^{\infty} \bar{f} \le |\beta|^{-1} \bar{f}(x)$. Liapunov's inequality, (4.23), (4.24) and c_r-inequality (Loève (1963), p. 155) give

$$\begin{aligned} \underset{\sim}{P}_i \Big|\int_x^{\infty} \bar{f} - \int_x^{x+L} \hat{\bar{f}}\Big|^{\gamma} &\le \{\int_{x+L}^{\infty} \bar{f} + \underset{\sim}{P}_i \int_x^{x+L} |\bar{f} - \hat{\bar{f}}|\}^{\gamma} \\ &\le (c_0(\gamma) \vee 1)(|\beta^{-1}|H)^{\gamma}\{(|\alpha|^{r\gamma} + 1)(\bar{f}(x))^{\gamma} \\ &\quad + (|\log H| \bar{f}(x)/u_{h+L}(x))^{\gamma/2}\} \end{aligned} \tag{4.25}$$

since $c_0^{\gamma}(1) = c_0(\gamma)$ and $L\beta = \log H$.

Since $\beta^{-1} \le \psi_i \le \alpha^{-1} < 0$, by (4.21) $|\psi_i - \hat{\psi}_{i+1}| \le |\beta|^{-1}$. Therefore, by (4.20), (4.21) and Lemma A.2 of the appendix,

$$\begin{aligned} \underset{\sim}{P}_i |\psi_i(x) - \hat{\psi}_{i+1}(x)|^p &\le 2^p |\beta|^{\gamma-p} (\bar{f}(x))^{-\gamma}\{\text{lhs of } (4.25) \\ &\quad + 2|\beta|^{-\gamma} M_{\gamma}(x)\} \ . \end{aligned} \tag{4.26}$$

But, since (4.4) followed by the inequality $u_h \le u_{h+L}$ gives $M_\gamma \le c_0(\gamma)H^\gamma\{(|\alpha|^r\bar{f})^\gamma + (\bar{f}/u_{h+L})^{\gamma/2}\}$, by (4.25),

$$\text{(4.27)} \quad \text{rhs of (4.26)} \le B(p)H^\gamma|\beta|^{-p}\{|\alpha|^{r\gamma} + 1 + (\bar{f}(x)u_{h+L}(x))^{-\gamma/2}(|\log H|^{\gamma/2} + 1)\}$$

where $B(p)$ is as given in the lemma. Since $X \sim P_{i+1}$, (4.26) followed by (4.27) and the definition of q_γ in (2.5) leads to (4.22). ■

We will use Lemma 8 with $p = 1$ in order to prove our main result below. Numbers $b_0, b_1, \ldots$ below are finite and independent of n.

<u>Theorem 4</u>. Recall that $H \doteq h^r = i^{-r/(1+2r)}$. If for $\delta \in [0,1]$ and $\zeta > 0$ ∃ a $b_0 \ni \forall\ i = 1,\ldots,n$,

$$\text{(A4.0)} \qquad \mu(m) \le b_0 iH^\delta|\log H|^\zeta|\beta|^2/(1 + \log i),$$

and if ∃ $\gamma \in [\delta,1]$ and b_1 and $b_2 \ni$

$$\text{(A4.1)} \qquad |\beta|^2\mu(q_\gamma/u_{h+L}^{\gamma/2}) \le b_1H^{\delta-\gamma}|\log H|^{\zeta-\gamma/2} \quad \forall\ i = 1,\ldots,n-1,$$

and

$$\text{(A4.2)} \qquad |\beta|^{-2}|\alpha|^{r\gamma} \le b_2H^{\delta-\gamma}|\log H|^\zeta \qquad \forall\ i = 1,\ldots,n$$

then ∃ a $b_3 \ni |D_n(\underset{\sim}{\omega},\underset{\sim}{\hat{\psi}})| \le b_3H_n^\delta|\log H_n|^\zeta$ uniformly in $\underset{\sim}{\omega} \in \times_1^n[\alpha_j,\beta_j]$.

<u>Proof</u>. Since $0 > \beta\uparrow$, by (A4.2), $h|\alpha| \doteq H^{1/r}|\alpha| \to 0$. Therefore, since $|s| = |\alpha|$, by (4.4), $B(1)$ in Lemma 8 is bounded in i. Consequently, since β abbreviates β_i, N_i here is

$|\beta_i|^{-1}$, $H = i^{-r/(1+2r)}$ and (A4.2) holds $\forall\ i = 1,\ldots,n$, Lemma 8 with $p = 1$ followed by (A4.2) and (A4.1) give a $b_4 \ni$ $\forall\ i = 1,\ldots,n-1$,

$$(4.28)\quad N_{i+1}\underset{\sim}{P}_{i+1}|\psi_i(X) - \hat{\psi}_{i+1}(X)| \le b_4 H^{\delta}|\log H|^{\zeta}$$
$$\le b_4 i^{-r\delta/(1+2r)}|\log H_n|^{\zeta}.$$

In view of (4.28) and (A4.0), the remainder of the proof follows by arguments identical to those used in the second paragraph of the proof of Theorem 2. ∎

Assumption (A4.1) of the theorem is the most stringent one. Comments, regarding a possible necessary condition for this, are the same as those contained in Remark 4.2.

Corollary 4. If α and β are constants wrt i, and for $\delta \in [0,1]$ and $\zeta > 0$ $\exists$ a $\gamma \in [\delta,1]$ and a $b_5 \ni \forall\ i = 1,\ldots,n-1$

$$(4.29)\quad (\text{lhs of (4.15) with } \delta \text{ and } u_h \text{ replaced by } \gamma \text{ and } u_{h+L}) \le b_5 H^{\delta-\gamma}|\log H|^{\zeta-\gamma/2},$$

then $D_n(\underset{\sim}{\omega},\underset{\sim}{\hat{\psi}}) = O(H_n^{\delta}|\log H_n|^{\zeta})$ uniformly in $\underset{\sim}{\omega} \in [\alpha,\beta]^n$.

Proof. The proof is analogous to that of Corollary 2. ∎

Example. For $\tau > 0$ fixed, let $u(x) = (\Gamma(\tau))^{-1}x^{\tau-1}[x > 0]$. Moreover, let α and β be constants wrt i. Then, since by c_r-inequality,

$$(4.30)\quad (q_{\gamma}/u_{h+L}^{\gamma/2}) \le \{\Gamma(\tau)(x^{\tau-1}[\tau \ge 1] + (x^{1-\tau} + (h+L)^{1-\tau})[0 < \tau < 1])\}^{\gamma/2}$$
$$\exp((\beta - \gamma\alpha/2)x)[x > 0],$$

(4.29), with any $0 < \delta \le 1 \ni \delta < 2\beta/\alpha$, $\zeta = (\delta/2) + (1-\tau)[0 < \tau < 1]$ and $\gamma = \delta$, is satisfied uniformly in $\underset{\sim}{\omega} \in [\alpha,\beta]^n$.

Remark 4.6. Section 2.1 of Susarla (1970) deals with sequence compound SELE in the example just mentioned. His condition on the parameter space implies $2\beta < \alpha < \beta$, and his assumption (0.8) together with his hypothesis $\tau > 2$ restricts τ to be in $\{3,4,\ldots,r+1\} \cup \{t|t \geq r+2\}$. Moreover, his presentations are rather complicated and proofs of lemmas are lengthy. His estimator, which depends also on certain other auxiliary random variables independent of $X_1,\ldots,X_n$, achieves (a rather weaker) rate $r/2(1+r)$ uniform in $\underset{\sim}{\omega} \in [\alpha,\beta]^n$. Note that this example with $\tau > 1/2$ does not cover the case of sequence compound problems where the component is SELE (based on X^2) of σ^2 in $N(0,\sigma^2)$-family.

2.5 Rates Near the Best Possible Rates with the Divided Difference and the Kernel Estimators of $\underset{\sim}{\omega}$ with Identical Components.

This section deals with only the case when $\underset{\sim}{\omega}$ has identical components. We will show that, when θ is identity, rates with the divided difference and the kernel estimators are arbitrarily close to, but cannot be more than, 2/5 and 1, respectively. We will also indicate that, for Case e^{ω}, kernel estimators achieve rates near 1.

Throughout this section, let $\underset{\sim}{\omega} = (\omega,\ldots,\omega) \in \Omega^n$, and let α and β be constants wrt i such that $\alpha \leq \omega < \beta$. Let f abbreviate f_{ω}. It may be noted at the outset that the conclusions of Lemmas 5, 6 and 7 remain valid if q_{γ} there is replaced by

$$\tilde{q}_{\gamma} \doteq f_{i+1}/(\bar{f})^{\gamma/2}.$$

Theorems 5 and 6 below are proved for the case θ is identity, and for the lower bounds there we assume:

(5.0) $\exists$ an $\varepsilon > 0$ and a finite $\ell > a$ $\ni$

Lebesgue-sup of the restriction to $(\ell,\ell+\varepsilon)$ of u is finite,

and

Lebesgue-inf of the restriction to $(\ell,\ell+\varepsilon)$ of u is positive.

With ε and ℓ in (5.0), we have

(5.1) $$0 < \mu(\ell < x < \ell + \varepsilon) < \infty,$$

and, since $f(t) \doteq C(\omega)\exp(\omega t)$,

(5.2) $$0 < \inf_{\ell<t<\ell+\varepsilon} f(t) \le \sup_{\ell<t<\ell+\varepsilon} f(t) < \infty .$$

Since θ is identity in both of the theorems and α_i and β_i are constants wrt i, $N_i \doteq |\alpha_i| \vee |\beta_i| \equiv A$ (say). Moreover, since $\omega_j \equiv \omega$, $\psi_j \equiv \omega$ and $R(G_n) \equiv 0$. Therefore, for every compound estimator $\underset{\sim}{\hat\psi} = (\hat\psi_1,\ldots,\hat\psi_n)$ with values in $[-A,A]^n$, (0.1) gives

(5.3) $$D_n(\underset{\sim}{\omega},\underset{\sim}{\hat\psi}) = n^{-1}\Sigma_{1\sim j}^{n} P_j(\hat\psi_j(X_j) - \omega)^2$$
$$\le 4n^{-1}A^2 + (2A)^{2-\gamma}n^{-1}\Sigma_{2\sim j}^{n} P_j|\hat\psi_j(X_j) - \omega|^{\gamma}.$$

Throughout the remainder of this section, $c_0, c_1, \ldots$ will denote finite positive constants.

<u>Theorem 5</u>. For each fixed $i = 1,\ldots,n-1$, let h ($\doteq h_i \doteq c_0 i^{-1/5}$), u_* and u^* be defined as in Section 3, (cf. preceding Theorem 1 as well as following (3.1). Also note that u_* and u^* depend on h, and hence, on i). Let $\underset{\sim}{\hat\psi}$ be the divided difference estimator introduced in Section 3. If for $i = 1$

(5.4) $$\sup_{x>a} u^*(x) f(x) < \infty$$

and for a $\gamma \in [0,2)$ and $i = 1$

(5.5) $$\mu((u^*/u_*^2)^{\gamma/2} f^{1-\gamma/2}) < \infty,$$

then $\exists$ a $c_1 \ni$

(5.6) $$D_n(\underset{\sim}{\omega}, \underset{\sim}{\hat{\psi}}) \le c_1 h_n^{\gamma} .$$

On the other hand, if (5.0) holds, then

(5.7) $$D_n(\underset{\sim}{\omega}, \underset{\sim}{\hat{\psi}}) \ge c_2 h_h^2 \quad \forall \text{ sufficiently large } n.$$

<u>Proof</u>. Throughout this proof, fix i with $1 \le i \le n-1$, and abbreviate X_{i+1} to X.

Let γ be given by (5.5). By (5.4), $k_0'(\gamma)$ in Lemma 5 is bounded in i (by choosing c_0 suitably, if necessary). Since the conclusion of that lemma holds even if q_γ there is replaced by $\tilde{q}_\gamma$, which here becomes $f^{1-\gamma/2}$, by (5.5), $i^{\gamma/5} \underset{\sim}{P}_{i+1} |\hat{\psi}_{i+1}(X) - \omega|^{\gamma}$ is bounded in i. This conclusion and the inequality in (5.3) give (5.6).

To prove (5.7) we proceed as follows. Recall that $\beta > \omega$. Since by (3.1), $\hat{\psi}_{i+1}(X) = (Q(\bar{\delta})(X))_{\alpha,\beta}$,

(5.8) $$\underset{\sim}{P}_{i+1} |\hat{\psi}_{i+1}(X) - \omega| \ge \int_0^{\beta-\omega} \underset{\sim}{P}_{i+1}[\hat{\psi}_{i+1}(X) - \omega > v]dv \ge$$

$$P_{i+1}\{[\ell < X < \ell + \epsilon/2] \int_0^{\beta-\omega} \underset{\sim}{P}_i [Q(\bar{\delta})(X) > v + \omega]dv\},$$

where (here and throughout this proof) ℓ and ϵ are given by (5.0).

Fix $X \in (\ell, \ell + \epsilon/2)$ and $v \in (0, \beta-\omega)$ until stated otherwise. From Section 3, (following (3.0)), for $j = 1,\ldots,i$, $\hat{\delta}_j(\cdot) = [\cdot \le X_j < \cdot + h]/u(X_j)$. As in the second paragraph in the proof of Lemma 3, for $j = 1,\ldots,i$, let $Y_j = \hat{\delta}_j(X + h) - e^{h(v+\omega)}\hat{\delta}_j(X)$. Since $X_1,\ldots,X_i$ are i.i.d. so are $Y_1,\ldots,Y_i$. The definition of Q in (3.0) and the Berry-Esseen theorem (Loève (1963), p. 288) lead to

$$(5.9)\qquad \underset{\sim}{P}_i[Q(\bar{\delta})(X) > v + \omega] = \underset{\sim}{P}_i[\Sigma_1^i Y_j > 0]$$
$$\ge \Phi\left(\frac{i^{\frac{1}{2}}P_1Y_1}{\sigma_1}\right) - c_3\,\frac{i^{-\frac{1}{2}}P_1|Y_1 - P_1Y_1|^3}{\sigma_1^3}$$

where σ^2 = var of Y_1 and Φ is the distribution function of $N(0,1)$.

Inequalities in the remainder of this proof are valid only $\forall$ sufficiently large i. Since $h \downarrow 0$, take $h \le \epsilon/4$, where ϵ is as in (5.0). Let $\delta_1(x) = \int_x^{x+h} f(t)dt$. Then, since $(\delta_1(\cdot+h)/\delta_1(\cdot)) = e^{h\omega}$,

$$(5.10)\qquad P_1Y_1 = \delta_1(X+h) - e^{h(v+\omega)}\delta_1(X) = \delta_1(X)e^{h\omega}(1-e^{hv})$$
$$\ge -hv\delta_1(X)e^{h\omega} \ge -c_4h^2v,$$

where the last inequality follows from (5.2). Moreover, since $\hat{\delta}_1(X+h)\hat{\delta}_1(X) = 0$ with probability one and $P_1\hat{\delta}_1(\cdot) > 0$, $\sigma_1^2 \ge e^{2h\omega}\text{var } \hat{\delta}_1(X)$. Thus,

$$e^{-2h\omega}\sigma_1^2 \ge \text{var}(\hat{\delta}_1(X)) = \int_X^{X+h}(f/u) - \delta_1^2(X)$$
$$\ge \delta_1(X)(1 - u^*(X)\delta_1(X))/u^*(X) \ge c_5h$$

since, by (5.0) and (5.2), $\sup_{\ell<t<\ell+\epsilon/2} h^{-1}u^*(t)\delta_1(t) < \infty$ and $\inf_{\ell<t<\ell+\epsilon/2} h^{-1}(\delta_1(t)/u^*(t)) > 0$. Consequently, by (5.10),

$$(5.11) \qquad \frac{P_1Y_1}{\sigma_1} \geq -c_6vh^{3/2},$$

and, since by (5.0), $P_1|Y_1 - P_1Y_1|^3 \leq (\text{constant})\sigma_1^2$,

$$(5.12) \qquad \frac{P_1|Y_1 - P_1Y_1|^3}{\sigma_1^3} \leq c_7h^{-\frac{1}{2}}.$$

Now weakening the integrand on the extreme rhs of (5.8) by (5.9), (5.11) and (5.12) and then making the transformation $c_6v(ih^3)^{\frac{1}{2}} = t$ we get, after recognizing that X has μ-density f satisfying (5.2),

$$(5.13) \quad \underset{\sim}{P}_{i+1}|\hat{\psi}_{i+1}(X) - \omega| \geq \mu(\ell < X < \ell + \epsilon/2)$$
$$\{c_8(ih^3)^{-\frac{1}{2}}\int_0^{(\beta-\omega)c_6(ih^3)^{\frac{1}{2}}} \Phi(-t)dt - c_9(ih)^{-\frac{1}{2}}\}.$$

Since $h = c_0 i^{-1/5}$, the integral in (5.13) converges to $\int_0^\infty \Phi(-t)dt$ as $i \to \infty$, and hence by (5.1), $i^{1/5}$ times the lhs of (5.13) is bounded below by a positive quantity for all large i. Therefore, since $P_{i+1}|\hat{\psi}_{i+1}(X) - \omega|^2 \geq P_{i+1}^2|\hat{\psi}_{i+1}(X) - \omega|$, (5.7) follows from the equality in (5.3). ■

<u>Theorem 6</u>. Let $\underset{\sim}{\hat{\psi}}$ be the kernel estimator introduced under Case ω in Section 4. (See (4.6). Also recall that $\underset{\sim}{\hat{\psi}}$ is defined for each integer $r > 1$.) As in (4.3) and (4.7), take $h = h_i \doteq i^{-1/(1+2r)}$ and $H = H_i = h^{r-1}$. If for a $\gamma \in [0,2)$ and $i = 1$,

$$(5.14) \qquad \mu(f^{1-\gamma/2}/u_h^{\gamma/2}) < \infty$$

then $\exists$ c_{10} $\ni$

(5.15) $$D_n(\underset{\sim}{\omega},\underset{\sim}{\hat{\psi}}) \le c_{10}H_n^{\gamma};$$

and if the kernel functions K_0 and K_1 defining $\underset{\sim}{\hat{\psi}}$ are bounded, and (5.0) holds, then

(5.16) $$D_n(\underset{\sim}{\omega},\underset{\sim}{\hat{\psi}}) \ge c_{11}H_n^2 \quad \forall \text{ sufficiently large } n.$$

<u>Proof</u>. Fix i with $1 \le i \le n-1$ and abbreviate X_{i+1} by X.

In view of Lemma 6, (5.15) follows by arguments identical to those given for the corresponding part of Theorem 5.

Now we prove (5.16). Recall that $\beta > \omega$. Since $\underset{\sim}{\hat{\psi}} = (\hat{\bar{f}}^{(1)}/\hat{\bar{f}})_{\alpha,\beta}$,

(5.17) $$\underset{\sim}{P}_{i+1}|\hat{\psi}_{i+1}(X) - \omega| \ge \int_0^{\beta-\omega}\underset{\sim}{P}_{i+1}[\hat{\psi}_{i+1}(X) - \omega > v]dv \ge$$

$$P_{i+1}\{[\ell < X < \ell + \epsilon/2]\int_0^{\beta-\omega}\underset{\sim}{P}_i[\hat{\bar{f}}^{(1)} - \omega\bar{f} > v|\hat{\bar{f}}|]dv\},$$

where the argument X in $\hat{\bar{f}}$ and $\hat{\bar{f}}^{(1)}$ is abbreviated by omission, and ℓ and ϵ are given by (5.0).

Fix $X \in (\ell, \ell + \epsilon/2)$ and $v \in (0, \beta-\omega)$ until stated otherwise. For $1 \le j \le i$, let

(5.18) $$u(X_j)T_j = \{(-h^{-1}K_1 + \omega K_0 + v|K_0|)\,(\frac{X_j - X}{h})\}[u(X_j) > 0]$$

where K_0 and K_1 are the kernels used in the definition of $\underset{\sim}{\hat{\psi}}$.

Since $X_1,\ldots,X_i$ are i.i.d., so are $T_1,\ldots,T_i$. The definitions of $\hat{\bar{f}}^{(j)}$, $j = 0,1$, given in (4.0) and Berry-Esseen theorem give

$$(5.19)\quad \underset{\sim}{P}_i[\hat{\bar{f}}^{(1)} - \omega\hat{\bar{f}} > v|\hat{\bar{f}}|] \geq \underset{\sim}{P}_i[\Sigma_1^i T_j < 0]$$
$$\geq \Phi(-\frac{i^{\frac{1}{2}} P_1 T_1}{\sigma_1}) - c_{13}\frac{i^{-\frac{1}{2}} P_1|T_1 - P_1 T_1|^3}{\sigma_1^3}$$

where $\sigma_1^2 = \text{var } T_1$.

Inequalities in the remainder of this proof are obtained only ∀ sufficiently large i. Since $h \downarrow 0$, we take $h \leq \varepsilon/2$, where ε is as in (5.0). Let Z_1, Z_2 and Z_3 denote, respectively, the first, second and the third term in the expression for T_1. Then the transformation theorem followed by r-th order Taylor expansion with integral form of the remainder and the orthogonality properties of $K_j \in \mathcal{K}_j^r$ for $j = 0,1$ gives $P_1 Z_1 + hf^{(1)}(X) = \int_0^1 K_1(y)\int_x^{x+hy}(X + hy-t)^{r-1}f^{(r)}(t)dtdy/(r-1)!$. Thus, since K_1 is bounded, by (5.2), $P_1 Z_1 \leq -hf^{(1)}(X) + \text{const.}h^r$. By similar arguments, $P_1 Z_2 \leq h\omega f(X) + \text{const.}h^{r+1}$. Therefore, since by (5.2), $PZ_3 = v\int_X^{X+h}|K_0(\frac{t-X}{h})|f(t)dt \leq hv$ const. and since $f^{(1)} = \omega f$,

$$(5.20)\qquad P_1 T_1 \leq c_{14}h(h^r + h^{r-1} + v).$$

Next observe that

$$(5.21)\qquad \sigma_1^2 \geq \sigma^2(Z_1) + \underset{j\neq j'}{\Sigma_1^3}\,\text{cov}(Z_j, Z_{j'}).$$

Since $h \downarrow 0$ and K_0 and K_1 are bounded, writing the exact expression for $\text{cov}(Z_1,Z_2)$, we see, after making use of the transformation theorem and (5.0) and (5.2), that $|\text{cov}(Z_1,Z_2)|$ is bounded in i. The same conclusion holds for $\text{cov}(Z_1,Z_3)$, $\text{cov}(Z_2,Z_3)$ and $(P_1 Z_1)^2$. Therefore, there exists a finite constant

ξ (could be negative) such that

(5.22) $\sigma_1^2 \geq P_1 Z_1^2 + \xi = h^{-1}\int K_1^2(t)((f/u)(X + ht))dt + \xi$.

Consequently, by (5.0) and (5.2),

(5.23) $$h\sigma_1^2 \geq c_{15},$$

and hence, by (5.20)

(5.24) $$\frac{P_1 T_1}{\sigma_1} \leq c_{16} h^{3/2}(h^r + h^{r-1} + v) .$$

Moreover, since by (5.0), $hP_1|T_1 - P_1 T_1|^3 \leq$ const. σ_1^2, by (5.23)

(5.25) $$\frac{P_1|T_1 - P_1 T_1|^3}{\sigma_1^3} \leq c_{17} h^{-\frac{1}{2}} .$$

Now weakening the integrand on the extreme rhs of (5.17) by (5.19), (5.24) and (5.25), and then, doing the analysis exactly similar to that given (following (5.12)) in the proof of Theorem 5, we get the desired conclusion. ■

For Case e^{ω} in Section 4 we have taken $H_i = h_i^r$ with $h_i = i^{-1/(1+2r)}$. In this case, if for a $\gamma \in [0,2)$ and $i = 1$

(5.26) $$\mu\{f^{1-\gamma/2}((u_h)^{-\gamma/2} + (u_h')^{-\gamma/2})\} < \infty,$$

where $u_h'(\cdot)$ in (5.26) stands for $u_h(\cdot+1)$, then using a proof similar to that used for (5.6) and making an application of Lemma 7 with $p = \gamma$, it follows that $\hat{\underset{\sim}{\psi}}$ given by (4.17) satisfies $D_n(\underset{\sim}{\omega},\hat{\underset{\sim}{\psi}}) = O(H_n^{\gamma})$ as $n \uparrow \infty$.

<u>Remark 5.1</u>. In view of the definitions of u_*, u^* and u_h, each of (5.5) and (5.26) implies (5.14). Densities satisfying

(5.5) and (5.26) $\forall\ \gamma \in [0,2)$ exist, e.g., take $u(x) = (2\pi)^{-\frac{1}{2}} e^{-x^2/2}[-\infty < x < \infty]$, $(\Gamma(\tau))^{-1} x^{\tau-1}[x > 0]$, $\tau \geq 1$ or $\Sigma_0^\infty (j+1)[j < x \leq j+1]$. Thus situations exist where the lower and upper bounds in each of Theorems 5 and 6 are considerably tight.

2.6 The Divided Difference Versus the Kernel Estimators.

The divided difference estimator introduced in Section 3 and the kernel estimator introduced under Case ω, in Section 4 are compound estimators of the same vector $\underset{\sim}{\omega} = (\omega_1, \ldots, \omega_n) \in \Omega^n$. Therefore, it is rather natural to make a comparison between them. Denote them here by $\underset{\sim}{\hat{\psi}}$ and $\underset{\sim}{\hat{\psi}}_K$ respectively. Recall that $\underset{\sim}{\hat{\psi}}_K$ is defined for each integer $r > 1$.

Under certain conditions, Theorems 2 and 5 show that $\underset{\sim}{\hat{\psi}}_K$ with $r > 6$ is better than $\underset{\sim}{\hat{\psi}}$ in the sense that, $\forall$ large n,

$$\sup_{\underset{\sim}{\omega}} |D_n(\underset{\sim}{\omega}, \underset{\sim}{\hat{\psi}}_K)| \leq c_0 n^{-(r-1)/(1+2r)} \leq c_1 n^{-2/5} \leq \sup_{\underset{\sim}{\omega}} |D_n(\underset{\sim}{\omega}, \underset{\sim}{\hat{\psi}})|$$

where c_0 and c_1 are some finite positive constants. By Theorems 2 and 6, $\underset{\sim}{\hat{\psi}}_K$ with $r > 6$ is better than $\underset{\sim}{\hat{\psi}}_K$ with $r = 2$ in the same sense.

Results obtained in Theorems 2 and 6 for $\underset{\sim}{\hat{\psi}}_K$ with $r = 2$ coincide, respectively, with those obtained in Theorems 1 and 5 for $\underset{\sim}{\hat{\psi}}$. However, as we have noted in Remarks 4.1 and 5.1, conditions for latter ones are stronger than those for former ones. Hence, $\underset{\sim}{\hat{\psi}}_K$, even with $r = 2$, could be preferable to $\underset{\sim}{\hat{\psi}}$. Nevertheless, $\underset{\sim}{\hat{\psi}}$ is a more natural estimator compared to $\underset{\sim}{\hat{\psi}}_K$.

Estimators $\hat{\underset{\sim}{\psi}}$ and $\hat{\underset{\sim}{\psi}}_K$ are somewhat (but not completely) similar to $\underline{\psi}^{**}$ and $o\hat{\underline{\psi}}$ respectively, prescribed by Susarla (1970), (Chapter 1), for the case $u(x) = (2\pi)^{-\frac{1}{2}}\exp(-x^2/2)[-\infty < x < \infty]$ and $-\alpha_i = \beta_i \equiv c_2$, a finite positive number. Results of Theorems 1 and 5 for $\hat{\underset{\sim}{\psi}}$, specialized to the above case and (in Theorem 5) $\omega = 0$, coincide with those obtained by Susarla for $\underline{\psi}^{**}$. However, in order to make $o\hat{\underline{\psi}}$ (which is, in comparison of $\hat{\underset{\sim}{\psi}}_K$, rather complicated to exhibit) better than $\underline{\psi}^{**}$ in the sense described above he requires $r > 12$.

APPENDIX

APPENDIX

Here we prove two useful lemmas; one concerning the weighted empiricals based on independent random variables and the other concerning the difference of two random ratios.

A.1. On Glivenko-Cantelli Theorem for the Weighted Empiricals Based on Independent Random Variables.

Let $X_1,\ldots,X_n$ be independent real valued random variables, and, for $w \in [0,1]$, let $F_j(x) = wP[X_j < x] + (1-w)P[X_j \le x]$ and $Y_j(x) = w[X_j < x] + (1-w)[X_j \le x]$. Furthermore, with $c_1,\ldots,c_n$ non-negative numbers such that $\Sigma_1^n c_j^2 = 1$, let

$$H_n = \Sigma_1^n c_j F_j, \quad H_n^* = \Sigma_1^n c_j Y_j$$

and

$$D_n^+ = \sup_{x,w} \max_{N \le n} (H_N^*(x) - H_N(x)).$$

A special case of the result in Remark A.1 (following the proof of Lemma A.1 below) is used in the proof of Theorem 2(b) of Chapter 1.

<u>Lemma A.1</u>. With $c = \Sigma_1^n c_j$, $\forall\, M \ge 1$,

(1) $$P[D_n^+ \ge M] < 2c\, M \exp(-2(M^2 - 1)) .$$

<u>Proof</u>. Let $\Delta = \max_{N \le n}(H_N^* - H_N)$. The remark following (2.17) of Hoeffding (1963), p. 17, and Theorem 2 therein, applied to random variables $c_j Y_j$ with $w = 1$ yield

(2) $P[\Delta(x-) \geq \eta] \leq \exp(-2\eta^2) \quad \forall\, x \in R$ and $\forall\, \eta > 0$.

Fix (temporarily) $0 < \gamma < M$ and partition R into k intervals with endpoints $-\infty = x < x_1 < \ldots < x_k = \infty$ such that $H_n(x_{j-1}, x_j) \leq \gamma$ for $j = 1, \ldots, k$. Since $0 \leq H_n(\cdot) \leq c$, we can (and do) take $k < c\gamma^{-1} + 1$. Since $H_N(x_{j-1}, x_j) \leq H_n(x_{j-1}, x_j) \leq \gamma$ for $N \leq n$, using the monotonicity of H_N and H_N^*, we get

(3) $$\sup_{x_{j-1}<x<x_j} \Delta(x) \leq \max_{N\leq n}(H_N^*(x_j-) - H_N(x_{j-1}+))$$
$$\leq \Delta(x_j-) + \gamma$$

since $\Delta(x_k-) = 0$. The rhs of (3) is independent of w.

Now observe that $\Delta(x) \leq \Delta(x+) \vee \Delta(x-) \leq \sup_{x\in S}\sup_w \Delta(x)$, where S is any dense subset of R. Therefore, $D_n^+ \doteq \sup_{x,w}\Delta(x) \leq \sup_w \max_{1\leq j\leq k} \sup_{x_{j-1}<x<x_j} \Delta(x)$, and from (3) and (2) we have

(4) $$P[D_n^+ \geq M] \leq P(\bigcup_1^{k-1} [\Delta(x_j-) \geq M - \gamma])$$
$$< c\gamma^{-1}\exp(-2(M-\gamma)^2).$$

Since the lhs of (4) is independent of γ, substituting γ on the rhs of (4) by $\gamma_0 = M(1 - (1 - M^{-2})^{\frac{1}{2}})$ and noting that $\gamma_0^{-1} \leq 2M$, we get the desired conclusion. ■

<u>Remark A.1</u>. If D_n^- is defined by interchanging H_N^* and H_N in D_n^+, then $P[D_n^- \geq M] <$ rhs of (1). This follows from Lemma A.1 since $D_n^-(\underset{\sim}{X}_n) = D_n^+(-\underset{\sim}{X}_n)$ where $\underset{\sim}{X}_n = (X_1, \ldots, X_n)$. Thus, with $D_n = \sup_{x,w}\max_{N\leq n}|H_N^*(x) - H_N(x)| (= D_n^+ \vee D_n^-)$, $P[D_n \geq M] < 2$(rhs of (1)).

With $c_1 = \ldots = c_n = n^{-\frac{1}{2}}$ and $M = (1 + \log n)^{\frac{1}{2}}$, it follows from (1) and Remark A.1 that

$$(5)\quad P[\sup_{x,w}\max_{N\le n}|\Sigma_1^N(Y_j(x) - F_j(x))| \ge n^{\frac{1}{2}}(1 + \log n)^{\frac{1}{2}}] < 4n^{-3/2}(1 + \log n)^{\frac{1}{2}}.$$

Thus, by Borel-Cantelli lemma,

$$\sup_{x,w}\max_{N\le n}|\Sigma_1^N(Y_j(x) - F_j(x))| = O((n \log n)^{\frac{1}{2}})$$

with probability one.

A.2. A Bound for the γ-th Mean of the Bounded Difference of Two Random Ratios.

We apply Lemma A.2 below in the proof of Lemmas 6, 7 and 8 of Chapter 2, in order to obtain certain suitable bound for the p-th mean distance between the compound and Bayes estimators there.

Lemma A.2. Let y,z and L be in R with $z \ne 0$ and $L > 0$. If Y and Z are two real valued random variables, then $\forall\ \gamma > 0$

$$(1)\quad E(|\tfrac{y}{z} - \tfrac{Y}{Z}| \wedge L)^\gamma \le 2^{\gamma+(\gamma-1)^+}|z|^{-\gamma}\{E|y-Y|^\gamma + (|\tfrac{y}{z}|^\gamma + 2^{-(\gamma-1)^+}L^\gamma)E|z-Z|^\gamma\}.$$

Proof. Since $[2|z-Z| \le |z|] \le [2|Z| \ge |z|]$, the lhs of (1) is exceeded by

$$(2)\qquad E(|\tfrac{y}{z} - \tfrac{Y}{Z}|^\gamma[2|Z| \ge |z|]) + L^\gamma E[2|z-Z| \ge |z|]\ .$$

Now by Markov-inequality, the second term in (2) is no more than $(2L)^\gamma|z|^{-\gamma}E|z-Z|^\gamma$. By triangle inequality with intermediate term y/Z, and by c_r-inequality (Loève (1963), p. 155), the first term in (2) is bounded above by $2^{\gamma+(\gamma-1)^+}|z|^{-\gamma}(E|y-Y|^\gamma + |\tfrac{y}{z}|^\gamma E|z-Z|^\gamma)$. Putting these results together we conclude (1). ■

Remark A.2. The proof given above also proves: If y,z,Y,Z and L are real valued random variables, and $L > 0$ with probability one, then $\forall\ \gamma > 0$

$$(3) \quad E\left(\left|\frac{y}{z} - \frac{Y}{Z}\right| \wedge L\right)^{\gamma} \leq 2^{\gamma+(\gamma-1)^{+}} E\{|z|^{-\gamma}(|y-Y|^{\gamma} + (|\frac{Y}{Z}|^{\gamma} + 2^{-(\gamma-1)^{+}} L^{\gamma})|z - Z|^{\gamma})\}.$$

Thus (1) becomes a special case of (3).

BIBLIOGRAPHY

BIBLIOGRAPHY

BHATTACHARYA, P.K. (1967). Estimation of a probability density function and its derivatives. Sankhyā Ser. A **29** 373-382.

BILLINGSLEY, PATRICK (1968). Convergence of Probability Measures. John Wiley & Sons, Inc., New York.

CACOULLOS, THEOPHILOS (1966). Estimation of a multivariate density. Ann. Inst. Statist. Math. **18** 179-189.

CENCOV, N.N. (1962). Evaluation of an unknown distribution density from observations. Soviet Math. **3** 1559-1562.

GILLILAND, DENNIS C. (1966). Approximation to Bayes risk in sequences of non-finite decision problems. RM-162, Department of Statistics and Probability, Michigan State University.

GILLILAND, DENNIS C. (1968). Sequential compound estimation. Ann. Math. Statist. **39** 1890-1905.

HANNAN, JAMES F. and ROBBINS, HERBERT (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. Ann. Math. Statist. **26** 37-51.

HANNAN, JAMES F. (1956). The dynamic statistical decision problem when the component problem involves a finite number, m, of distributions (Abstract). Ann. Math. Statist. **27** 212.

HANNAN, JAMES (1957). Approximation to Bayes risk in repeated play. Contributions to the Theory of Games **3** 97-139. Ann. Math. Studies No. 39, Princeton Univ. Press.

HANNAN, J.F. and VAN RYZIN, J.R. (1965). Rate of convergence in the compound decision problem for two completely specified distributions. Ann. Math. Statist. **36** 1743-1752.

HANNAN, JAMES and MACKY, DAVID W. (1971). Empirical Bayes squared error loss estimation of unbounded functionals in exponential families. RM-290, Department of Statistics and Probability, Michigan State University.

HEWITT, EDWIN and STROMBERG, KARL (1965). Real and Abstract Analysis. Springer-Verlag New York, Inc.

HOEFFDING, WASSILY (1963). Probability inequalities for sums of bounded random variables. J. Amer. Stat. Assoc. 58 13-30.

JOHNS, M.V., Jr. (1967). Two-action compound decision problems. Proc. Fifth Berkeley Symp. Math. Statist. Prob. 1, University of California Press.

JOHNS, M.V., Jr. and VAN RYZIN, J. (1972). Convergence rates for empirical Bayes two-action problems II. Continuous case. Ann. Math. Statist. 43 934-947.

KIEFER, J. (1961). On large deviations of the empiric d.f. of vector chance variables and a law of iterated logarithm. Pacific J. Math. 11 649-660.

KRONMAL, R. and TARTER, M. (1968). The estimation of probability densities and cumulatives by Fourier series methods. J. Amer. Statist. Assoc. 38 482-493.

LOÈVE, MICHEL (1963). Probability Theory (3rd ed.). Van Nostrand, Princeton.

NADARAYA, É.A. (1965). On non-parametric estimates of density function and regression curves. Theor. Prob. Appl. 10 186-190.

NATANSON, I.P. (1955). Theory of Functions of a Real Variable. Frederick Ungar Publishing Co., New York.

OATEN, ALLAN (1972). Approximation to Bayes risk in compound decision problems. Ann. Math. Statist. 43 1164-1184.

PARZEN, EMANUEL (1962). On the estimation of probability density and mode. Ann. Math. Statist. 33 1065-1076.

RAO, B.L.S. PRAKASA (1969). Estimation of a unimodal density. Sankhyā Ser. A. 31 23-36.

ROSENBLATT, MURRAY (1956). Remarks on some non parametric estimates of a density function. Ann. Math. Statist. 27 832-837.

SAMUEL, ESTER (1963). Asymptotic solutions of the sequential compound decision problem. Ann. Math. Statist. 34 1079-1094.

SAMUEL, ESTER (1965). Sequential compound estimators. Ann. Math. Statist. 36 879-889.

SCHWARTZ, STUART C. (1967). Estimation of a probability density by an orthogonal series. Ann. Math. Statist. 38 1261-1265.

SCHUSTER, EUGENE F. (1969). Estimation of a probability density function and its derivatives. Ann. Math. Statist. **40** 1187-1195.

SUSARLA, V. (1970). Rates of convergence in sequence-compound squared-distance loss estimation and two-action problems. RM-262, Department of Statistics and Probability, Michigan State University.

VAN RYZIN, J. (1966). The compound decision problem with $m \times n$ finite loss matrix. Ann. Math. Statist. **37** 412-424.

VAN RYZIN, J. (1970). On a histogram method of density estimation. Univ. of Wisconsin, Department of Statistics T.R. No. 226.

VAN VLECK, F.S. (1973). A remark concerning absolutely continuous functions. Amer. Math. Monthly **80** 286-287.

WAHBA, GRACE (1971). A polynomial algorithm for density estimation. Ann. Math. Statist. **42** 1870-1886.

WATSON, G.S. and LEADBETTER, M.R. (1963). On the estimation of the probability density I. Ann. Math. Statist. **34** 480-491.

WATSON, G.S. (1969). Density estimation by orthogonal series. Ann. Math. Statist. **40** 1496-1498.

WEGMAN, E.J. (1969). Maximum likelihood estimation of a unimodal density function. Inst. Statist. Mimeo Ser. No. 608, University of North Carolina at Chapel Hill.

WEISS, L. and WOLFOWITZ, J. (1967). Estimation of a density at a point. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **7** 327-335.

YU, BENITO (1970). Rates of convergence in empirical Bayes two-action and estimation problems and in extended sequence-compound estimation problems. RM-279, Department of Statistics and Probability, Michigan State University.