RATES OF CONVERGENCE IN SEQUENCE-COMPOUND SQUARED-DISTANCE LOSS ESTIMATION AND TWO-ACTION PROBLEMS

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

VYAGHRESWARUDU SUSARLA

1970

This is to certify that the

thesis entitled

RATES OF CONVERGENCE IN SEQUENCE-COMPOUND SQUARED-DISTANCE LOSS ESTIMATION AND TWO-ACTION PROBLEMS

presented by

Vyaghreswarudu Susarla

has been accepted towards fulfillment of the requirements for

Ph.D. degree in Statistics and Probability

James Hannan

Major professor

Date August 12, 1970

O-169

ABSTRACT

RATES OF CONVERGENCE IN SEQUENCE-COMPOUND SQUARED-DISTANCE LOSS ESTIMATION AND TWO-ACTION PROBLEMS

By

Vyaghreswarudu Susarla

We consider a sequence of repetitions of a statistical decision problem which has the structure of one of the statistical decision problems described below. These statistical decision problems will be referred to later on as component problems.

When the family of distributions $\mathcal{P}$ is, (1) the family of m-variate normal distributions with covariance matrix I and mean θ in $\Theta = [|\theta| \le \alpha]$, the problem is to estimate θ with squared-distance loss, (2) the family of $\Gamma(\alpha)$ distributions with scale parameter θ in $\Theta = [a,b]$ where $0 < a < b < \infty$, the problem is to estimate θ with squared-distance loss and (3) same as (2) except that the problem is a linear loss two-action problem. For any distribution G on Θ, let $R(G)$ denote the Bayes risk in the component problem.

$\underline{X} = \{\underset{\sim}{X}_n\}$ is a sequence of independent random variables with distributions $\{P_{\theta_n}\}$ in $\underset{n}{\times}\mathcal{P}$. Let G_n be the empiric distribution of $\theta_1,\ldots,\theta_n$. Let s be a positive integer and γ be in $(0,1)$. All the orders stated here are uniform in the parameter sequences $\underline{\theta}$ in $\underset{n}{\times}\Theta$.

When the component problem is described by (1), we exhibit procedures ψ^{**}_n, $\hat{\psi}_n$ and ${}_o\hat{\psi}_n$, which are functions of $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$,

such that $D_n(\underline{\theta},\underline{\psi}^{**}) = n^{-1}\Sigma_1^n E|\psi_j^{**} - \theta_j|^2 - R(G_n)$, $D_n(\underline{\theta},\underline{\hat{\psi}})$ and $D_n(\underline{\theta},{}_o\underline{\hat{\psi}})$ are $O(n^{-1/(m+4)})$, $O(n^{-(2-1)\gamma/(2s+m)(1+\gamma)})$ and $O(n^{-(s-1)/2(s+m+1)})$ respectively. Whenever $m \geq 5$ and $(s-1)\gamma(m+4) \geq 2(2s+m)(1+\gamma)$, $\underline{\hat{\psi}}$ is better than $\underline{\psi}^{**}$ in the sense that $\sup\{D_n(\underline{\theta},\underline{\psi}^{**})|\underline{\theta}\}$ converges to zero at a faster rate than $\sup\{D_n(\underline{\theta},\underline{\psi}^{**})|\underline{\theta}\}$ does. Similar comparison has been given between $\underline{\psi}^{**}$ and ${}_o\underline{\hat{\psi}}$. The results stated above for $\underline{\psi}^{**}$ and $\underline{\hat{\psi}}$ have been extended to the case when the covariance matrix I is replaced by $\sigma^2 I$ (σ^2 unknown) and the means θ_n lie in lower dimensional subspaces having the same dimension.

When the component problem is given by (2), we exhibit a procedure ψ_n^* such that $D_n(\underline{\theta},\underline{\psi}^*) = O(n^{-s/2(s+1)})$ when a,b and α satisfy certain conditions. For the same set of conditions on a,b and α, when the component problem is described by (3) with loss function L, we define a procedure $\hat{\psi}_n$ such that $n^{-1}\Sigma_1^n E\, L(\theta_j,\psi_j) - R(G_n) = O(n^{-s/2(s+1)})$.

RATES OF CONVERGENCE IN SEQUENCE-COMPOUND SQUARED-DISTANCE LOSS ESTIMATION AND TWO-ACTION PROBLEMS

By

Vyaghreswarudu Susarla

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

1970

G - 65572
1 - 22 - 71

TO MY PARENTS

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to Professor J.F. Hannan for introducing me to compound decision theory and for suggesting the problems treated in the thesis. His comments aided greatly in improving and simplifying most of the results of the thesis.

I wish to thank Professors D.C. Gilliland and J.S. Huang for going through the thesis and pointing out some misprints. I wish to thank Mr. T. O'Bryan for suggesting changes in the phrasology. Special thanks are due to Mrs. Noralee Barnes for her excellent typing and cheerful attitude in the preparation of the manuscript.

I am grateful to the Department of Statistics and Probability, Michigan State University and the National Science Foundation for the financial support during my stay at Michigan State University.

TABLE OF CONTENTS

INTRODUCTION

In Chapter I, $\mathcal{P} = \{P_\theta\}$ is the family of m-variate normal distributions with covariance matrix I and mean θ in $\Theta = [|\theta| \le \alpha]$ and the component problem is squared-distance loss estimation of θ. In Chapter II, $\mathcal{P}$ is the family of $\Gamma(\alpha)$ distributions with scale parameter θ in $\Theta = [a,b]$ where $0 < a < b < \infty$ and the component problem is either squared-distance loss estimation or a linear loss two-action problem. For any distribution G on Θ, let ψ_G and $R(G)$ denote the Bayes estimate and the Bayes risk in the component problem.

The sequence-compound problem consists of a sequence of repetitions of the component problem with the loss taken to be the average of the component losses. $\underline{X} = \{\underset{\sim}{X}_n\}$ is a sequence of independent random variables with distributions $\{P_{\theta_n}\}$ in $\times_n \mathcal{P}$ and the nth component decision ξ_n depends only on $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$. With G_n denoting the empiric distribution of $\theta_1,\ldots,\theta_n$, let

$$(0.1) \qquad D_n(\underline{\theta},\underline{\xi}) = \frac{1}{n}\sum_{j=1}^{n} E[L(\theta_j,\xi_j)] - R(G_n).$$

$D_n(\underline{\theta},\underline{\xi})$ is known as the modified regret of $\underline{\xi}$.

Since the work reported here is a continuation of Gilliland (1966, 1968) and Johns (1967), we describe some of the main results contained in these references. All the orders stated below are uniform in the parameter sequences concerned. For the purpose of this introduction only, abbreviate $O(n^{-a})$ to order $-a$.

When $\mathcal{P}$ is the family of univariate normal distributions with variance unity and mean θ in $[-\alpha,+\alpha]$ and the component problem is squared-distance loss estimation, Gilliland (1966) exhibited a procedure whose modified regret is order $-1/5$. When $\mathcal{P}$ is a certain family of discrete distributions and the component problem is the linear loss two-action problem, Johns (1967) exhibited a procedure whose modified regret is order $-1/2$. When $\mathcal{P}$ is a certain discrete exponential family and the component problem is squared-distance loss estimation, Gilliland (1968) exhibited two procedures whose modified regrets are order $-1/2$.

Now we briefly describe the main results obtained in this work. In Chapter I, the Bayes estimate against G_{n-1} is

$$\psi_{n-1}(\underset{\sim}{X}_n) = \underset{\sim}{X}_n + \frac{\bar{q}}{\bar{p}}$$

with $\bar{p}$ denoting the mixed density $\int p_\theta dG_{n-1}$, $\bar{q}$ denoting the matrix of partial derivatives of $\bar{p}$ and indication of the evaluation of both at $\underset{\sim}{X}_n$ abbreviated by omission.

In section §1.2, we define ψ_n^{**} based on a divided difference estimate of $\bar{q}/\bar{p}$ whose D_n is order $-(m+4)^{-1}$. This generalizes the result of Gilliland (1966) for $m = 1$ case.

In section §1.3, for each positive integer s and γ in $(0,1)$, we define ψ_n based on kernel estimators for $\bar{p}$ and $\bar{q}$ analogous to Johns and Van Ryzin (1967) estimates of $\int p_\theta dG$ and its derivative in empirical Bayes two-action problem in exponential families and show $D_n(\underset{\sim}{\theta},\underset{\sim}{\psi})$ is order $-(s-1)\gamma/(2s+m)(1+\gamma)$. For each integer $s > 1$, we exhibit ${}_0\psi_n$, specializing $\underset{\sim}{\psi}$ but for the latter's retraction to $[\beta,\infty)$, whose D_n is order $-(s-1)/2(s+m+1)$.

In section §1.5, we show that $D_n(\underline{0},\underline{\psi}^{**}) \geq c\, n^{-2/m+4}$ where c is a constant depending on α. Hence, whenever $m \geq 5$ and s and γ are such that $(s-1)\gamma(m+4) > 2(2s+m)(1+\gamma)$, $\underline{\hat{\psi}}$ is better than $\underline{\psi}^{**}$ in the sense that $\sup\{D_n(\underline{\theta},\underline{\hat{\psi}})|\underline{\theta}\}$ converges to zero at a faster rate than $\sup\{D_n(\underline{\theta},\underline{\psi}^{**})|\underline{\theta}\}$. A similar comparison is made between $\underline{\psi}^{**}$ and $\underline{{}_o\hat{\psi}}$.

Section §1.6 extends the main results of sections §1.2 and §1.3 to the case when the covariance matrix I is replaced by $\sigma^2 I$ (σ^2 unknown) under the additional assumption that the means lie in lower dimensional subspaces having the same dimension.

In Chapter II, as already indicated earlier, $\mathcal{P}$ is the family of $\Gamma(\alpha)$ distributions with sclae parameter θ in $\Theta = [a,b]$. In section §2.1, the component problem is squared-distance loss estimation. For each positive integer s, we define ψ_n^* based on kernel estimates for two densities and show that $D_n(\underline{\theta},\underline{\psi}^*)$ is order $-s/2(s+1)$ whenever a,b and α satisfy certain conditions. In section §2.2, the component problem is linear loss two-action. For each positive integer s, we define $\hat{\psi}_n$ based on kernel estimates for two densities and show that $D_n(\underline{\theta},\underline{\hat{\psi}})$ is order $-s/2(s+1)$ whenever a,b and α satisfy the conditions imposed on them in section §2.1.

Throughout this work, we let Φ and ϕ denote the standard normal distribution and its density respectively. We suppress the arguments of functions whenever it is convenient not to exhibit them. Indulging in the abuse of notation, we let sets denote their own indicator functions and, infrequently, are forced to let the value of a function denote the function. For any measure μ, we let $\mu[f]$ or μf denote $\int f d\mu$.

CHAPTER I

RATES IN THE ESTIMATION PROBLEM FOR A FAMILY OF m-VARIATE NORMAL DISTRIBUTIONS

§1.0 Introduction and Notation.

For fixed $\alpha < \infty$ and for fixed positive integer m, let $\mathcal{P} = \{P_\theta \mid |\theta| \le \alpha\}$ be the family of distributions with P_θ denoting the m-variate normal law with mean θ and covariance $\sigma^2 I$, where I is the $m \times m$ identity matrix and $\sigma^2 > 0$.

We consider the following estimation problem which will be called the component problem hereafter. Based on an observation of a random vector $\underset{\sim}{X}$ whose distribution P_θ belongs to $\mathcal{P}$, the problem is to estimate θ with squared-distance loss.

For any distribution G on the m-sphere of radius α, let ψ_G and $R(G)$ denote the Bayes estimate and the Bayes risk versus G in the above estimation problem. Since the problem considered here is the squared-distance loss estimation problem, ψ_G is given by the conditional expectation of θ given $\underset{\sim}{X}$. If p_θ denotes the usual density of P_θ wrt Lebesgue measure on $(R^m, \mathcal{B}^m)$, then the conditional expectation of θ given $\underset{\sim}{X}$ is $G[\theta p_\theta]/G[p_\theta]$ which, can be expressed as $\underset{\sim}{X} + \sigma^2 q_G$ where q_G is the vector of partial derivatives of $\log G[p_\theta]$ wrt the various coordinates of $\underset{\sim}{X}$. Hence,

$$\psi_G = \underset{\sim}{X} + \sigma^2 q_G \ . \tag{0.1}$$

We consider a sequence of component problems as described above. That is, let $\{\underset{\sim}{X}_n\}$ be a sequence of independent random variables with $\underset{\sim}{X}_n$ distributed as P_{θ_n} belonging to $\mathcal{P}$ and the problem is to estimate every component of $\{\theta_n\}$ with loss taken as the average of squared-distance losses in individual components. For each n, let the product measure $\times_{i=1}^{n} P_i$, where

P_i is an abbreviation for P_{θ_i}, be denoted by $\underline{P}_n$. Let $\underline{\xi} = \{\xi_n\}$ be a sequence-compound procedure (abbreviated to procedure hereafter). For any parameter sequence $\underline{\theta} = \{\theta_n\}$ and for any non randomized procedure $\underline{\xi} = \{\xi_n\}$, define

$$D_n(\underline{\theta},\underline{\xi}) = n^{-1} \sum_{j=1}^{n} \underline{P}_j[|\xi_j - \theta_j|^2] - R(G_n) \tag{0.2}$$

where G_n is the empiric distribution of $\theta_1,\ldots,\theta_n$. $D_n(\underline{\theta},\underline{\xi})$ is called the modified regret of the procedure $\underline{\xi}$.

The orders stated in the results of sections §1.1, §1.2, §1.3 and §1.4 are uniform in all parameter sequences $\underline{\theta}$ in $\times_n [|\theta_n| \le \alpha]$ and the order stated in section §1.6 is uniform in all parameter sequences $\underline{\theta}$ belonging to $\times_n ([|\theta_n| \le \alpha] \cap \mathcal{R}_n)$, where, for each n, $\mathcal{R}_n$ is a d $(d < m)$-dimensional subspace of R^m. To reduce the complexity of the statements of various results in this chapter, the range of the parameter sequences will not be exhibited, but is understood to be as in the preceeding sentence. Henceforth, we use these conventions.

In section §1.1, we get an upper bound for $|D_n(\underline{\theta},\underline{\xi})|$ under the assumption that $\underline{\xi}$ is in $\times_n [-\alpha,+\alpha]^m$ and a useful lemma, both results holding for each σ^2. In section §1.2, we exhibit a procedure $\underline{\psi}^{**}$ for which $D_n(\underline{\theta},\underline{\psi}^{**}) = O(n^{-\frac{1}{m+4}})$ when $\sigma^2 = 1$. In section §1.3, for each $\gamma > 0$, we exhibit a procedure $\hat{\underline{\psi}}$ for which $D_n(\underline{\theta},\hat{\underline{\psi}}) = O(n^{-(\frac{1}{4}-\gamma)})$ again for $\sigma^2 = 1$. In section §1.4, for each positive integer s, we exhibit a procedure $\underline{{}_o\hat{\psi}}$ for which $D_n(\underline{\theta},\underline{{}_o\hat{\psi}}) = O(n^{-(s-1)/2(m+s+1)})$ for $\sigma^2 = 1$. Section §1.5 shows that

$D_n(\underline{0}, \psi^{**}) \geq c\, n^{-\frac{2}{m+4}}$ for all n, where $\underline{0} = \{0\}$ and c is a positive constant. Section §1.6 has two subsections. These subsections extend respectively the main results of sections §1.2 and §1.3 to the case when σ^2 is unknown and when, for each n, θ_n lies in $\mathcal{R}_n$ intersected with m-sphere of radius α.

Let μ denote the Lebesgue measure on $(R^m, \mathcal{B}^m)$. For any two points u,v in R^m with coordinates $u_1,\ldots,u_m$, $v_1,\ldots,v_m$ respectively, let $|u|^2 = \sum_{i=1}^{m} u_i^2$, $\|u\| = \sum_{i=1}^{m} |u_i|$ and $(u,v) = \sum_{i=1}^{m} u_i v_i$. The inequalities $|u| \leq \|u\| \leq \sqrt{m}\,|u|$ will be used without further comment. Also, a vector in R^m will be denoted by $<\ >$ with the general coordinate of the vector exhibited inside the brackets.

Let p_n be an abbreviation for p_{θ_n}, the density of P_{θ_n}. For each n, let ψ_{G_n} be abbreviated by ψ_n. Then, specializing (0.1),

$$\psi_n = \underset{\sim}{X} + \sigma^2 \bar{q}_n \tag{0.3}$$

where $\bar{q}_n$ is the vector of partial derivatives of the function $\log \sum_{j=1}^{n} p_j$ wrt the coordinates of $\underset{\sim}{X}$.

§1.1 A Bound for the Modified Regret $D_n(\underline{\theta},\underline{\xi})$.

We state and prove two lemmas which are higher dimensional generalizations of proposition 1 and corollary 1 of Chapter I of Gilliland (1966) for the case of the family of normal distributions $\mathcal{P}$.

Lemma 1. $P_n[|\psi_n - \psi_{n-1}|] \le 2\alpha\, e^{4\sigma^{-2}\alpha^2}\, n^{-1}$ for $n > 1$.

Proof. From $\psi_n = G_n[\theta p_\theta]/G_n[p_\theta]$, the triangle inequality and Jensen's inequality, respectively, it follows that

$$
\begin{aligned}
|\psi_n - \psi_{n-1}| &= p_n\Big(\sum_{j=1}^{n} p_j\Big)^{-1}\Big(\sum_{j=1}^{n-1} p_j\Big)^{-1}\Big|\sum_{j=1}^{n-1}(\theta_j - \theta_n)p_j\Big| \\
&\le 2\alpha\, p_n\Big(\sum_{j=1}^{n} p_j\Big)^{-1} \le 2\alpha\, n^{-2} p_n \sum_{j=1}^{n} p_j^{-1}.
\end{aligned}
\tag{1.1}
$$

Since $p_n p_j^{-1} = \exp \sigma^{-2}(\theta_n - \theta_j,\ x - (\theta_n + \theta_j)2^{-1})$,

$$P_n[p_n p_j^{-1}] = \exp \sigma^{-2}|\theta_n - \theta_j|^2 \le \exp \sigma^{-2}4\alpha^2$$

which, when substituted in (1.1), completes the proof.

Lemma 2. If the procedure $\underline{\xi}$ is in $\times_n [-\alpha,+\alpha]^m$, then, for each $\sigma^2 > 0$,

$$|D_n(\underline{\theta},\underline{\xi})| \le 4\alpha\, n^{-1}\sum_{j=1}^{n} \underline{P}_j[\|\xi_j - \psi_{j-1}\|] + O(n^{-1}\log n)$$

where ψ_o is an arbitrary decision rule taking values in $[-\alpha,+\alpha]^m$.

Proof. Inequalities (8.8) and (8.11) of Hannan (1957) when specialized to the squared-distance loss estimation problem here give the inequality

$$(1.2) \qquad n^{-1}\sum_{j=1}^{n} \underline{P}_j[|\psi_j - \theta_j|^2] \le R(G_n) \le n^{-1}\sum_{j=1}^{n} \underline{P}_j[|\psi_{j-1} - \theta_j|^2].$$

By bounding the term $R(G_n)$ appearing in the definition (0.2) of $D_n(\underline{\theta},\underline{\xi})$ above and below by using (1.2), we obtain, by using the equality $|a|^2 - |b|^2 = (a + b,\ a-b)$ for a,b in R^m, the double inequality

$$(1.3) \qquad n^{-1}\sum_{j=1}^{n} \underline{P}_j[(\xi_j + \psi_{j-1} - 2\theta_j,\ \xi_j - \psi_{j-1})] \le D_n(\underline{\theta},\underline{\xi})$$

$$\le n^{-1}\sum_{j=1}^{n} \underline{P}_j[(\xi_j + \psi_j - 2\theta_j,\ \xi_j - \psi_j)].$$

Since, by assumption ξ_j is in $[-\alpha,+\alpha]^m$, $|\theta_j| \le \alpha$ and ψ_j, ψ_{j-1}, being the Bayes estimates respectively against G_j, G_{j-1} whose supports lie in m-sphere of radius α, are in m-sphere of radius α, we obtain that the moduli of the ℓth coordinates of $\xi_j + \psi_j - 2\theta_j$ and $\xi_j + \psi_{j-1} - 2\theta_j$ are at most 4α. Therefore, we obtain from (1.3) that

$$(1.4) \quad -4\alpha n^{-1}\sum_{j=1}^{n} \underline{P}_j[\|\xi_j - \psi_{j-1}\|] \le D_n(\underline{\theta},\underline{\xi}) \le 4\alpha n^{-1}\sum_{j=1}^{n} \underline{P}_j[\|\xi_j - \psi_j\|].$$

The triangle inequality applied to the rhs of (1.4) and Lemma 1 will complete the proof of the lemma.

We use Lemma 2 to obtain rates of convergence of the modified regret of certain sequence-compound procedures to be defined in later sections.

§1.2 A Rate of Convergence for $D_n(\underline{\theta},\underline{\psi}^{**})$ with ψ^{**} Based on a Divided Difference Estimator for the Derivative of the log of a Density.

Some notation, which is similar to that of Gilliland (1966) for $m = 1$ case, is required to define $\underline{\psi}^{**}$. The notation to be given below is for each n and will be used also in section §1.6.1. We abbreviate by omission the dependency on n of the functions to be defined below.

Let $\bar{F}$ denote the average of the distributions of $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_{n-1}$. For each x in R^m with coordinates $x_1,\ldots,x_m$, let $\square = \times_{j=1}^{m} I_j$ where $I_j = [x_j,x_j+h]$ for $j = 1,\ldots,m$, and for $\ell = 1,\ldots,m$, $\square_\ell = \times_{j=1}^{m} I_j'$ where $I_j' = I_j$ for $j \neq \ell$ and $I_\ell' = I_\ell + k = [x_\ell + k, x_\ell + k + h]$. Let $0 < k \leq h$.

For any distribution F on R^m, let $t(F)$ denote the vector valued function $\langle k^{-1}\log (F\square_\ell/F\square)\rangle$ from R^m to R^m where $F\square$ and $F\square_\ell$ represent the measures of $\square$ and $\square_\ell$ for $\ell = 1,\ldots,m$ respectively under F and any undefined ratios are taken to be 1. We abbreviate $t(\bar{F})$ frequently by t hereafter.

Let the function $t(F^*)$, where F^* is the empiric distribution of $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_{n-1}$, be denoted by t^*. Let X abbreviate $\underset{\sim}{X}_n$ and $X_1,\ldots,X_m$ denote the coordinates of X. Let

$$(2.1) \qquad \psi^{**} = tr'(X + \sigma^2 t^*(X)), \quad \psi^* = tr(X + \sigma^2 t^*(X))$$

where tr' and tr stand for the coordinatewise retraction to the intervals $[-\alpha,+\alpha]$ and $[-\alpha - k - h, \alpha + k + h]$ respectively.

With ψ abbreviating ψ_{n-1}, we have, since $|\psi| \le \alpha$ and $\psi^{**} = tr'\psi^{*}$, $\|\psi^{**} - \psi\| \le \|\psi^{*} - \psi\|$. Therefore, by the triangle inequality

$$(2.2)\quad \underline{P}_n[\|\psi^{**} - \psi\|] \le \underline{P}_n[\|\psi^{*} - (X + \sigma^2 t(X))\|] + P_n[\|X + \sigma^2 t(X) - \psi\|].$$

Lemma 3. For all x in R^m,

(1) $x + \sigma^2 t(\bar{F})(x) \in [-\alpha - \frac{k}{2} - h, \alpha]^m$,

(2) $\bar{F}\square_\ell \ge \bar{p}(\frac{h}{\sigma})^m \exp - \frac{(k+h)}{\sigma^2}(\|x\| + \frac{mh+k}{2})$ where $\bar{p}$ is the density of $\bar{F}$ at x and

(3) $\bar{F}\square'_\ell \le \bar{F}\square_\ell \frac{k+h}{h} \exp \frac{k+h}{\sigma^2}(|x_\ell| + \alpha + k + h)$ for $\ell = 1,\ldots,m$ where $\square'_\ell = \times_{j=1}^{m} I''_j$ with $I''_j = I_j$ for $j \ne \ell$ and $I''_\ell = [x_\ell, x_\ell + k + h]$.

Proof. In this proof, let F_j denote the distribution of $\underset{\sim}{X}_j$ and $\theta_{j1},\ldots,\theta_{jm}$ denote the coordinates of θ_j.

Proof of (1). Let ℓ be in $\{1,\ldots,m\}$. Since the coordinates of $\underset{\sim}{X}_j$ are independent, we can express $F_j\square$ and $F_j\square_\ell$ as the products of univariate normal probabilities. Therefore, by cancelling out the common terms in these products, we obtain that

$$\frac{F_j\square_\ell}{F_j\square} = \frac{\Phi(\sigma^{-1}(x_\ell - \theta_{j\ell} + k + h)) - \Phi(\sigma^{-1}(x_\ell - \theta_{j\ell} + k))}{\Phi(\sigma^{-1}(x_\ell - \theta_{j\ell} + h)) - \Phi(\sigma^{-1}(x_\ell - \theta_{j\ell}))}.$$

Applying Cauchy's mean value theorem (Graves (1946), p. 81) to the rhs of this equality over $(0,\sigma^{-1}h)$ with the function in the denominator to be taken as $\Phi(\sigma^{-1}(x_\ell - \theta_{j\ell}))$ while that in the numerator to be taken as $\Phi(\sigma^{-1}(x_\ell - \theta_{j\ell} + k))$, we obtain, by using $a^2 - b^2 = (a + b)(a - b)$ for a,b in R^1, the existence of ω in $(0,1)$ such that

$$\frac{F_{j\square\ell}}{F_{j\square}} = \exp - \frac{k}{\sigma^2}\,(x_\ell - \theta_{j\ell} + \frac{k}{2} + \omega h).$$

Hence, since $|\theta_{j\ell}| \le \alpha$,

$$\exp - \frac{k}{\sigma^2}\,(x_\ell + \alpha + \frac{k}{2} + h) \le \frac{F_{j\square\ell}}{F_{j\square}} \le \exp \frac{k}{\sigma^2}\,(\alpha - x_\ell).$$

Since these bounds for $F_{j\square\ell}/F_{j\square}$ are independent of j, they also bound $\bar{F}_{\square\ell}/\bar{F}_{\square}$. These inequalities are equivalent to (1) in view of the definition of $t(\bar{F})$. Since ℓ is arbitrary the proof of (1) is complete.

Proof of (2). We temporarily abbreviate $\mu[S\phi]$ by $\Phi(S)$ for any S in $\mathcal{B}^m$. Then, $F_{j\square\ell} = \Phi(\sigma^{-1}(I_\ell - \theta_{j\ell} + k))\prod_{i\ne\ell} \Phi(\sigma^{-1}(I_i - \theta_{ji}))$. Hence, applying the mean value theorem to $\Phi(\sigma^{-1}(I_i - \theta_{ji}))$ for $i \ne \ell$ and to $\Phi(\sigma^{-1}(I_\ell - \theta_{j\ell} + k))$, we obtain the existence of $\langle\omega_i\rangle$ in $(0,1)^m$ such that

$$F_{j\square\ell} = (\frac{h}{\sigma})^m \prod_{i=1}^{m} \phi(\frac{x_i - \theta_{ji} + \omega_i h + \delta_{i\ell} k}{\sigma})$$

where $\delta_{i\ell} = [i = \ell]$. Hence, since $-\log(\phi(u)/\phi(v)) = (u-v)(u+v)/2$, we obtain that

$$-\sigma^2 \log(\frac{F_{j\square\ell}}{(\frac{h}{\sigma})^m p_j}) = \sum_{i=1}^{m} (\omega_i h + \delta_{i\ell} k)(x_i - \theta_{ji} + \frac{\omega_i h + \delta_{i\ell} k}{2}).$$

Hence, since the functions of ω_i appearing on the rhs of this equality, being convex, attain their maxima at $\omega_i = 0$ or 1, we obtain that the rhs of the last equality is exceeded by

$$\sum_{i=1}^{m} (\delta_{i\ell} k(x_i - \theta_{ji} + \delta_{i\ell} k/2)) \vee ((h + \delta_{i\ell} k)(x_i - \theta_{ji} + \frac{h + \delta_{i\ell} k}{2}))$$

$$\le \sum_{i=1}^{m} (h + \delta_{i\ell}k)(|x_i| + |\theta_{ji}| + \frac{h + \delta_{i\ell}k}{2})$$

$$\le (k + h)(\|x\| + \sqrt{m}\,\alpha + \frac{mh + k}{2}).$$

Since this bound for $-\sigma^2\log(F_{j\Box_\ell}/(h/\sigma)^m p_j)$ is independent of j, it also bounds $-\sigma^2\log(\bar{F}\Box_\ell/(h/\sigma)^m \bar{p})$. Since this inequality is equivalent to (2), the proof of (2) is complete.

Proof of (3). Using the notation $\Phi(S)$ for S in $\mathcal{B}^m$ introduced in the proof of (2), we have
$F_{j\Box_\ell} = \Phi(\sigma^{-1}(I_\ell - \theta_{j\ell} + k)) \prod_{i\ne\ell} \Phi(\sigma^{-1}(I_i - \theta_{ji}))$ and $F_{j\Box'_\ell} =$
$\Phi(\sigma^{-1}(I''_\ell - \theta_{j\ell})) \prod_{i\ne\ell} \Phi(\sigma^{-1}(I_i - \theta_{ji}))$. Hence, applying the mean value theorem to $\Phi(\sigma^{-1}(I_\ell - \theta_{j\ell} + k))$ and $\Phi(\sigma^{-1}(I''_\ell - \theta_{j\ell}))$, we obtain the existence of ω, ω' in $(0,1)$ such that

$$\frac{F_{j\Box'_\ell}}{F_{j\Box_\ell}} = \frac{k + h}{h}\,\frac{\phi(\sigma^{-1}(x_\ell - \theta_{j\ell} + \omega'(k+h))}{\phi(\sigma^{-1}(x_\ell - \theta_{j\ell} + k + \omega h))}\,.$$

Hence, since $\log(\phi(u)/\phi(v)) = (v-u)(v+u)/2$, we obtain from the above equality that

$$\sigma^2\log\frac{h}{k+h}\frac{F_{j\Box'_\ell}}{F_{j\Box_\ell}} = ((1 - \omega')k + (\omega - \omega')h)(x_\ell - \theta_{j\ell} + \frac{\omega'+1}{2}k + \frac{\omega+\omega'}{2}h).$$

Hence, since $0 < \omega, \omega' < 1$, we obtain from the above equality that

$$\sigma^2\log\frac{h}{k+h}\frac{F_{j\Box'_\ell}}{F_{j\Box_\ell}} \le (k+h)(|x_\ell| + \alpha + k+h).$$

Since this bound is independent of j, it also bounds $\sigma^2\log(h\,\bar{F}\Box'_\ell/(k+h)\bar{F}\Box_\ell)$. This inequality is equivalent to (3). Hence the proof of (3) is completed.

Now we bound the integrals on the rhs of (2.2). The method of bounding the first integral is essentially a generalization of that given in Chapter III of Gilliland (1966). We get a simpler method of bounding this integral because of the definition of ψ^{**} in (2.1). This definition of ψ^{**} differs from that of a similar function introduced by Gilliland (1966). The method of bounding the second integral of the rhs of (2.2) differs from that of Gilliland (1966). Let $c_1, c_2, \ldots$ denote finite functions of σ^2. Let

$$K = \{k \mid 0 < k < (5 + \sigma^{-2}(2\alpha + 1))^{-1}\}.$$

Lemma 4. If k is in K, then

$$\underline{P}_n[\|\psi^* - (X + \sigma^2 t(X))\|] \le c_1\left(\frac{k+h}{nk^2h^{m+1}}\right)^{\frac{1}{2}} + c_2\left(\frac{1}{nh^m}\right)^{\frac{1}{2}}.$$

Proof. Since the lhs is the sum of $\underline{P}_n$-integrals of the moduli of the coordinates of $\psi^* - X - \sigma^2 t(X)$, the lemma will be proved by showing that these integrals are bounded by rhs/m.

Let the dependency of t on X be suppressed and $X_1, \ldots, X_m$ denote the coordinates of X. We abbreviate in this proof the ℓth coordinates of ψ^* and t by omission. Let α' denote $2(\alpha + k + h)$.

Since ψ^*, by definition (2.1), is the retraction of $X_\ell + \sigma^2 t^*$ to $[-\alpha - k - h, \alpha + k + h]$ and since $X_\ell + \sigma^2 t$, by (1) of Lemma 3, is in $[-\alpha - \frac{k}{2} - h, \alpha]$, it follows that $|\psi^* - X_\ell - \sigma^2 t| \le \alpha'$ and $|\psi^* - X_\ell - \sigma^2 t| \le \sigma^2 |t^* - t|$.

Therefore,

$$\underset{\sim}{P}_{n-1}[|\psi^* - X_\ell - \sigma^2 t|] \le \int_0^{\alpha'} \underset{\sim}{P}_{n-1}[|\psi^* - X_\ell - \sigma^2 t| > u]du$$

$$\le \int_0^{\alpha'} \underset{\sim}{P}_{n-1}[\sigma^2|t^* - t| > u]du \tag{2.3}$$

$$= \int_0^{\alpha'} \underset{\sim}{P}_{n-1}[\sigma^2(t^* - t) > u]du$$

$$+ \int_{-\alpha'}^{0} \underset{\sim}{P}_{n-1}[\sigma^2(t^* - t) < u]du.$$

The main part of the proof bounds $\underset{\sim}{P}_{n-1}[\sigma^2(t^* - t) > u]$ for $0 \le u \le \alpha'$ and $\underset{\sim}{P}_{n-1}[\sigma^2(t^* - t) < u]$ for $-\alpha' \le u < 0$ by using the Berry-Esseen theorem. The rest of the proof shows that the P_n-integral of m times the bound for the rhs of (2.3) is exceeded by the bound in the lemma.

Let X be fixed until otherwise stated. Let $\tilde{\delta}_j = [\underset{\sim}{X}_j \in \square_\ell]$, $\delta_j = [\underset{\sim}{X}_j \in \square]$ and

$$Y_j(u) = \tilde{\delta}_j - \delta_j e^{k(t+u/\sigma^2)} \quad \text{for} \quad |u| \le \alpha'. \tag{2.4}$$

Let the dependency of Y_j on u be suppressed hereafter. Let $\beta^2 = \mathrm{Var}(\Sigma Y_j)$ and $L = \beta^{-3} \Sigma P_j|Y_j - P_jY_j|^3$ where Σ stands for summation over j from 1 to $n-1$.

Sublemma. For $|u| \le \alpha'$,

$$|\underset{\sim}{P}_{n-1}[\Sigma Y_j \ge 0] - \Phi(\beta^{-1}\Sigma P_jY_j)| \le \frac{2}{c_4} \frac{e^{k\sigma^{-2}(\alpha+\alpha'+|X_\ell|)}}{(n-1)^{\frac{1}{2}}(\bar{F}\square_\ell)^{\frac{1}{2}}}.$$

Proof. With $\mathbf{B}$ denoting the Berry-Esseen constant, the Berry-Esseen theorem (Loeve (1963), p. 288) implies that $|\underset{\sim}{P}_{n-1}[\Sigma Y_j \ge 0] - \Phi(\beta^{-1}\Sigma P_jY_j)|$ is exceeded by BL. Hence, we complete the proof of the sublemma by showing that L is exceeded by B^{-1} times the bound of the sublemma. In order to get a bound on L, we first get a lower bound on β^2.

By applying L1.A (see Appendix) to the Y_j, we obtain a lower bound for β^2. We observe that Y_j defined by (2.4) takes three values; namely

$$0, 1 \text{ and } -e^{k(t+u/\sigma^2)} \tag{2.5}$$

with probabilities $1 - F_j\Box_\ell - F_j\Box$, $F_j\Box_\ell$ and $F_j\Box$ respectively where F_j is the distribution of $\underset{\sim}{X}_j$.

Therefore, it follows by L1.A that

$$\begin{aligned}\mathrm{Var}(Y_j) \geq (1 - F_j\Box_\ell - F_j\Box)(F_j\Box(1 - F_j\Box) + \\ e^{2k(t+u/\sigma^2)} F_j\Box_\ell(1 - F_j\Box_\ell)).\end{aligned}$$

Weakening this inequality by dropping the second term of the above inequality, we obtain that $\mathrm{Var}(Y_j) \geq$ $(1 - F_j\Box_\ell - F_j\Box)F_j\Box_\ell(1-F_j\Box_\ell)$. Hence, denoting by c_3^2 the infimum

$$\inf \{(1 - \Phi]_{-h/2\sigma}^{h/2\sigma})^m (1 - (\Phi]_{-h/2\sigma}^{h/2\sigma})^{m-1}\ \Phi]_{-(k+h)/2\sigma}^{(k+h)/2\sigma}) \mid h \leq k \in K\},$$

we obtain that

$$\mathrm{Var}(Y_j) \geq c_3^2\, F_j\Box_\ell. \tag{2.6}$$

Therefore, since $\beta^2 = \Sigma\, \mathrm{Var}(Y_j)$,

$$\beta^2 \geq c_3^2\ (n - 1)\ \overline{F}\Box_\ell. \tag{2.7}$$

Since $0 \le u \le \alpha'$ and since $\sigma^2 t \le \alpha - X_\ell$ by (1) of Lemma 3, the maximum of the values of $|Y_j|$ in (2.5) is exceeded by

$$e^{k\sigma^{-2}(\alpha+\alpha'+|X_\ell|)} . \tag{2.8}$$

Therefore, the standardized range bound for L, with the help of (2.7) and (2.8), gives that

$$L \le \frac{2\ e^{k\sigma^{-2}(\alpha+\alpha'+|X_\ell|)}}{c_3\ (n-1)^{\frac{1}{2}}(\bar{F}\square_\ell)^{\frac{1}{2}}} .$$

In view of the remarks of the first paragraph of the sublemma, we obtain the result of the sublemma by taking $c_3 = c_4 B$.

Before proceeding further, we note that it follows from (2.5) and (2.8) that EY_j^2 is exceeded by

$$e^{2k\sigma^{-2}(\alpha+\alpha'+|X_\ell|)} F_j\square_\ell'$$

where $\square_\ell'$ is defined in (3) of Lemma 3. Hence, since $\beta^2 = \Sigma\, \mathrm{Var}(Y_j) \le \Sigma\, EY_j^2$,

$$\beta^2 \le (n-1)\ e^{2k\sigma^{-2}(\alpha+\alpha'+|X_\ell|)}\ \bar{F}\square_\ell' . \tag{2.9}$$

Now we proceed with the main part of the proof of the lemma. Let $0 \le u \le \alpha'$. Then the definitions of t^* and Y_j imply that $[\sigma^2(t^* - t) > u] \le [\Sigma Y_j \ge 0]$. Hence, by the sublemma, it follows that $\underline{P}_{n-1}[\sigma^2(t^* - t) > u]$ is exceeded by

$$\Phi(\beta^{-1}\Sigma\, P_j Y_j) + \text{bound in the sublemma.} \tag{2.10}$$

Since $e^{kt} = \bar{F}\square_\ell / \bar{F}\square$ by the definition of t,

$\Sigma P_jY_j = (n-1)\bar{F}\square_\ell(1 - \exp k\sigma^{-2}u) \le -(n-1)k\sigma^{-2}\bar{F}\square_\ell u$. Therefore, by using the upper bound for β in (2.9), we obtain that $\Phi(\beta^{-1}\Sigma P_jY_j) \le \Phi(-((n-1)h^m k^2)^{\frac{1}{2}} f u)$ where f is the positive solution of the equation

$$\text{(2.11)} \qquad \sigma^4 h^m \bar{F}\square'_\ell e^{2k\sigma^{-2}(\alpha+\alpha'+|X_\ell|)} f^2 = (\bar{F}\square_\ell)^2.$$

Therefore, since (2.10) is a bound for $\underline{P}_{n-1}[\sigma^2(t^* - t) > u]$, we have, for $0 \le u \le \alpha'$,

$$\text{(2.12)} \quad \underline{P}_{n-1}[\sigma^2(t^*-t) > u] \le \Phi(-((n-1)h^m k^2)^{\frac{1}{2}} f\, u) + \text{bound in the sublemma.}$$

Now we consider bounding the probability $\underline{P}_{n-1}[\sigma^2(t^*-t) < u]$ for $-\alpha' \le u < 0$. The definitions of t^* and Y_j imply that $[\sigma^2(t^* - t) < u] \le [\Sigma Y_j \le 0] = [\Sigma - Y_j \ge 0]$. Since the sublemma continues to hold when $\tilde{\delta}_j$ and δ_j in the definition of Y_j are replaced by $-\tilde{\delta}_j$ and $-\delta_j$ respectively, we obtain, by applying the sublemma to $\underline{P}_{n-1}[\Sigma - Y_j \ge 0]$, that $\underline{P}_{n-1}[\sigma^2(t^* - t) < u]$ is at most

$$\text{(2.13)} \qquad \Phi(-\beta^{-1}\Sigma P_jY_j) + \text{bound in the sublemma.}$$

Again since $\Sigma P_jY_j = (n-1)\bar{F}\square_\ell(1 - \exp k\sigma^{-2}u) \ge -2^{-1}(n-1)k\sigma^{-2}\bar{F}\square_\ell u$ where the inequality follows since $k\sigma^{-2}\alpha' < 1$ by the hypothesis on k, we obtain, by using the upper bound β^2 in (2.9) and the definition of f in (2.11), that $\Phi(-\beta^{-1}\Sigma P_jY_j)$ is exceeded by $\Phi(2^{-1}((n-1)k^2h^m)^{\frac{1}{2}} f\, u)$. Therefore

$$\underline{P}_{n-1}[\sigma^2(t^* - t) < u] \le \Phi(\tfrac{1}{2}(n-1)k^2h^m)^{\frac{1}{2}} f\, u) + \text{bound in the sublemma.}$$

Integrating this inequality wrt u on $[-\alpha',0)$ and the inequality (2.12) wrt u on $[0,\alpha']$, then bounding their first terms by using the inequality $\int_0^{\alpha'}\Phi(-au)du \le (2\pi)^{-\frac{1}{2}}a^{-1}$ for any $a > 0$, we obtain, by using the inequality (2.3), that

$$P_{n-1}[|\psi^* - X_\ell - \sigma^2 t|] \le \frac{3}{\sqrt{2\pi}}\frac{1}{((n-1)k^2h^m)^{\frac{1}{2}}f} + 2\alpha' \text{ (bound in the sublemma)}.$$

Hence we complete the proof of the lemma by showing below that the P_n-integrals of $m(h(k+h)^{-1})^{\frac{1}{2}}f^{-1}$ and $m(nh^m)^{\frac{1}{2}}$ (bound in the sublemma) are uniformly bounded in n.

By definition of f in (2.11), we have

$$f^{-1} = \sigma^2 \left(\frac{\bar{F}\Box'_\ell}{\bar{F}\Box_\ell}\right)^{\frac{1}{2}}\left(\frac{h^m}{\bar{F}\Box_\ell}\right)^{\frac{1}{2}} e^{k\sigma^{-2}(\alpha+\alpha'+|X_\ell|)}.$$

By bounding above $(\bar{F}\Box'_\ell/\bar{F}\Box_\ell)^{\frac{1}{2}}$ by using (3) of Lemma 3 and by bounding below $\bar{F}\Box_\ell/h^m$ by using (2) of Lemma 3, we get an upper bound for $(h(k+h)^{-1})^{\frac{1}{2}}f^{-1}$. Weakening this upper bound for $(h(k+h)^{-1})^{\frac{1}{2}}f^{-1}$ by using the fact that $0 < h \le k < 1/5$, we obtain that

$$\left(\frac{h}{k+h}\right)^{\frac{1}{2}}\frac{1}{f} \le c_5 \frac{e^{\frac{1}{2\sigma^2}(3|X_\ell|+\|X\|)}}{p^{-\frac{1}{2}}}$$

for some c_5. Since $(2\pi\sigma^2)^m p_n^2 \le \exp(-\sigma^{-2}((|X| - \alpha)^+)^2)$ and $(2\pi\sigma^2)^{m-2}p \ge \exp - (\sigma^{-2}(\alpha + |X|)^2)$, we obtain that the above upper bound for $(h(k+h)^{-1})^{\frac{1}{2}}f^{-1}$ is uniformly bounded in n and P_n-integrable. Now by using (2) of Lemma 3, and the inequality $0 < h \le k < 1/5$, we obtain that $(nh^m)^{\frac{1}{2}}$ (bound in the sublemma) is exceeded by

$$c_6 \frac{e^{\sigma^{-2}(|X_\ell| + \frac{\|X\|}{2})}}{p^{-\frac{1}{2}}}$$

for some c_6. Again, since $(2\pi\sigma^2)^m p_n^2 \le \exp - (\sigma^{-2}((|X| - \alpha)^+)^2)$ and $(2\pi\sigma^2)^m \bar{p}^{-2} \ge \exp - (\sigma^{-2}(\alpha + |X|)^2)$, we obtain that the P_n-integral of the above upper bound for $h^{m/2}$ (bound in the sublemma) is uniformly bounded in n. This completes the proof of the lemma.

The next lemma is a slight generalization of a particular case of Cauchy's mean value theorem (Graves (1946), p. 81).

Lemma 5. For each $j = 1,\dots,n-1$, $i = 1,\dots,m$, let the functions f_{ji}, g_{ji} be real valued, continuous on $[a_i,b_i]$ and differentiable on (a_i,b_i) and let the derivative of g_{ji} be finite and positive. Then there exist c_1 in $(a_1,b_1),\dots,c_m$ in (a_m,b_m) such that

$$\frac{\Sigma\, \pi\, f_{ji}]_{a_i}^{b_i}}{\Sigma\, \pi\, g_{ji}]_{a_i}^{b_i}} = \frac{\Sigma\, \pi\, f'_{ji}(c_i)}{\Sigma\, \pi\, g'_{ji}(c_i)}$$

where Σ stands for the summation over j from 1 through $n-1$, π stands for product over i from 1 through m and prime over any function denotes its derivative.

Proof. Define the functions ξ_1 and η_1 on $[a_1,b_1]$ as follows.

$$\xi_1(x) = \Sigma\, f_{j1}(x) \prod_{i=2}^{m} f_{ji}]_{a_i}^{b_i}$$

and

$$\eta_1(x) = \Sigma\, g_{j1}(x) \prod_{i=2}^{m} g_{ji}]_{a_i}^{b_i}$$

for x in $[a_1,b_1]$.

With these definitions, we obtain that

$$\frac{\Sigma\,\pi\, f_{ji}]_{a_i}^{b_i}}{\Sigma\,\pi\, g_{ji}]_{a_i}^{b_i}} = \frac{\xi_1]_{a_1}^{b_1}}{\eta_1]_{a_1}^{b_1}} \,. \tag{2.14}$$

Since f_{j1} and g_{j1} are continuous on $[a_1,b_1]$ and differentiable on (a_1,b_1) for all j, so are ξ_1 and η_1. Moreover, since the derivative of g_{ji} is finite and positive for all j and i by assumption, so is the derivative of η_1. Hence, applying Cauchy's mean value theorem to the rhs of (2.14), we obtain that there exists c_1 in (a_1,b_1) such that

$$\frac{\Sigma\,\pi\, f_{ji}]_{a_i}^{b_i}}{\Sigma\,\pi\, g_{ji}]_{a_i}^{b_i}} = \frac{\xi_1'(c_1)}{\eta_1'(c_1)} \,. \tag{2.15}$$

Now, we define ξ_2 and η_2 on $[a_2,b_2]$ as follows.

$$\xi_2(x) = \Sigma\, f_{j1}'(c_1) f_{j2}(x) \prod_{i=3}^{m} f_{ji}]_{a_i}^{b_i}$$

and

$$\eta_2(x) = \Sigma\, g_{j1}'(c_1) g_{j2}(x) \prod_{i=3}^{m} g_{ji}]_{a_i}^{b_i}$$

for x in $[a_2,b_2]$. Then it follows that the ratio $\xi_1'(c_1)/\eta_1'(c_1)$ is identically the ratio $\xi_2]_{a_2}^{b_2}/\eta_2]_{a_2}^{b_2}$. Again, ξ_2 and η_2 are continuous on $[a_2,b_2]$ and differentiable on (a_2,b_2) since f_{j2},g_{j2} are continuous on $[a_2,b_2]$ and differentiable on (a_2,b_2) for all j. Also, since the derivative of g_{ji} is finite and positive for all j and i, the derivative of

η_2 is finite and positive. Therefore, again using Cauchy's mean value theorem, the definitions of ξ_2 and η_2 and (2.15), we obtain the existence of c_2 in (a_2, b_2) such that

$$\frac{\Sigma\, \pi\, f_{ji}]_{a_i}^{b_i}}{\Sigma\, \pi\, g_{ji}]_{a_i}^{b_i}} = \frac{\xi_2'(c_2)}{\eta_2'(c_2)} \quad . \tag{2.16}$$

Iterating the above procedure of obtaining (2.16) from (2.15) $(m-2)$ times, we obtain the result of the lemma.

We apply this lemma to prove the following lemma.

Lemma 6. $|\ell$th coordinate of $X + \sigma^2 t - \psi| \le k(1 + \frac{\alpha^2}{\sigma^2}) + h(1 + m\frac{\alpha^2}{\sigma^2})$ for $\ell = 1, \ldots, m$.

Proof. Let the dependency of t on X be suppressed and abbreviate the indication of the ℓth coordinates of ψ and t by omission.

Let H abbreviate $h^{-m}\bar{F}\square$ and e_ℓ denote the unit vector in the ℓth direction. Since $t = k^{-1}$ $(\log H(X + k\, e_\ell) - \log H(X))$, by the mean value theorem, there exists ϵ in $(0,1)$ such that

$$t = \frac{\partial \log H}{\partial X_\ell}(X + \epsilon k\, e_\ell). \tag{2.17}$$

Since $\psi - X_\ell = \sigma^2\, \partial \log \bar{p}/\partial X_\ell$, the above equality together with the triangle inequality implies that

$$|X_\ell + \sigma^2 t - \psi| \le \sigma^2(|I_1| + |I_2|) \tag{2.18}$$

where

$$I_1 = \frac{\partial \log \bar{p}}{\partial X_\ell}\Big]_X^{X+\epsilon k\, e_\ell} \tag{2.19}$$

and

$$I_2 = \frac{\partial \log H}{\partial X_\ell}(X + \epsilon k\, e_\ell) - \frac{\partial \log \bar{p}}{\partial X_\ell}(X + \epsilon k\, e_\ell). \tag{2.20}$$

By the mean value theorem, $I_1 = \epsilon k(\partial^2 \log \bar{p}/\partial X_\ell^2)(X + \epsilon^* k\, e_\ell)$ for some ϵ^* in $(0,\epsilon)$. With $\theta_{j1},\ldots,\theta_{jm}$ denoting the coordinates of θ_j, we have

$$\sigma^2\left(1 + \sigma^2 \frac{\partial^2 \log \bar{p}}{\partial X_\ell^2}\right) = \frac{\Sigma (X_\ell - \theta_{j\ell})^2 p_j}{\Sigma\, p_j} - \left(\frac{\Sigma (X_\ell - \theta_{j\ell}) p_j}{\Sigma\, p_j}\right)^2.$$

The rhs of this equality can be recognized as the conditional variance of the ℓth coordinate of $X - \theta$ given X when the pair (θ,X) has the joint distribution resulting from G_{n-1} on θ and P_θ on X for given θ. Hence, since the support of G_{n-1} is in the m-sphere of radius α, we obtain that

$$\sigma^2 \left|\frac{\partial^2 \log \bar{p}}{\partial X_\ell^2}\right| \le 1 + \frac{\alpha^2}{\sigma^2}. \tag{2.21}$$

Hence

$$\sigma^2 |I_1| \le k\left(1 + \frac{\alpha^2}{\sigma^2}\right). \tag{2.22}$$

We complete the proof of the lemma by showing that $\sigma^2|I_2| \le h(1 + m\alpha^2\sigma^{-2})$ with the help of Lemma 5.

The definition of H gives

$$(n-1)h^m H = \Sigma\, F_j \,\square \tag{2.23}$$

where, since the coordinates of $\underset{\sim}{X}_j$ are independent,

$$F_j \,\square = \pi\, \Phi\Big]_{(x_i-\theta_{ji})/\sigma}^{(x_i-\theta_{ji}+h)/\sigma}. \tag{2.24}$$

Therefore,

$$(2.25)\qquad \sigma(n-1)h^m \frac{\partial H}{\partial X_\ell} = \Sigma\ \phi]_{(X_\ell-\theta_{j\ell})/\sigma}^{(X_\ell-\theta_{j\ell}+h)/\sigma} \prod_{i\neq\ell} \Phi]_{(X_i-\theta_{ji})/\sigma}^{(X_i-\theta_{ji}+h)/\sigma} .$$

Now we apply Lemma 5 to the ratio $(\partial H/\partial X_\ell)/H$ obtained by using (2.23), (2.24) and (2.25) with the following identification. For all $j = 1,\ldots,n-1$, $f_{ji} = g_{ji} = \Phi(\sigma^{-1}(y-\theta_{ji}))$ for $i \neq \ell$, $f_{j\ell} = \phi(\sigma^{-1}(y-\theta_{j\ell}))$, $g_{j\ell} = \Phi(\sigma^{-1}(y-\theta_{j\ell}))$ and $(a_i,b_i) = (\sigma^{-1}X_i, \sigma^{-1}(X_i + h))$ for all i. Then there exists a δ in $(0,1)^m$ such that

$$\frac{\partial \log H}{\partial X_\ell} = \frac{\partial \log \bar{p}}{\partial X_\ell}(X + h\delta).$$

By subtracting $\partial \log \bar{p}/\partial X_\ell$ and then applying the mean value theorem to this function of h, we obtain the existence of h' in $(0,h)$ such that

$$(2.26)\qquad \frac{\partial \log H}{\partial X_\ell} - \frac{\partial \log \bar{p}}{\partial X_\ell} = h \sum_{i=1}^{m} \delta_i \frac{\partial^2 \log \bar{p}}{\partial X_i \partial X_\ell}(X + h'\delta).$$

For $i \neq \ell$, we obtain directly that

$$\sigma^4 \frac{\partial^2 \log \bar{p}}{\partial X_i \partial X_\ell} = \frac{\Sigma(\theta_{j\ell} - X_\ell)(\theta_{ji} - X_i)p_j}{\Sigma\, p_j} - \frac{\Sigma(\theta_{j\ell} - X_\ell)p_j}{\Sigma\, p_j} \cdot \frac{\Sigma(\theta_{ji} - X_i)p_i}{\Sigma\, p_j} .$$

The rhs of this equality can be recognized as the i,ℓth element in the covariance matrix of $\theta - X$ conditional on X when the joint distribution of (θ,X) results from G_{n-1} on θ and P_θ on X for given θ. Hence, since the support of G_{n-1} lies in m-sphere of radius α, it follows by Schwarz's inequality that

$$\sigma^4 \left|\frac{\partial^2 \log \bar{p}}{\partial X_i \partial X_\ell}\right| \leq \alpha^2 \quad \text{for} \quad i \neq \ell.$$

This inequality, together with (2.21) and (2.26), implies that

$$\sigma^2 \left| \frac{\partial \log H}{\partial X_\ell} - \frac{\partial \log \bar{p}}{\partial X_\ell} \right| \le h(1 + m \frac{\alpha^2}{\sigma^2}).$$

Thus, by (2.20), $|I_2| \le h(1 + m\alpha^2\sigma^{-2})$ and the proof of the lemma is complete.

Before stating a theorem as a Corollary to Lemmas 2, 4 and 6, we make a remark on the proof of Lemma 6.

Remark 1. The method of proof of the lemma differs much from that of Gilliland (1966) for $m = 1$ case. He has never used the fact that the conditional variances and covariances are uniformly bounded by explicit functions of α^2. Moreover, the constants multiplying k and h in the result of the lemma are specific functions of α while those of Gilliland are complicated integrals. A proof similar to the proof obtained by particularizing our proof to $m = 1$ is simpler than that of Gilliland.

In the rest of the section, we let h and k depend on n. We assume in the theorem to be stated below that $\sigma^2 = 1$. The choices of h and k given in the following theorem are optimal for the convergence to 0 of the expression obtained by adding the right hand sides of Lemmas 4 and 6.

Theorem 1. If $h = n^{-\frac{1}{m+4}}$, $k = a\, n^{-\frac{1}{m+4}}$ for a in $[1,\infty)$ and ψ^{**} is defined by (2.1), then

$$\underline{P}_n[\|\psi^{**} - \psi\|] = O(n^{-\frac{1}{m+4}})$$

and

$$D_n(\underline{\theta},\underline{\psi}^{**}) = O(n^{-\frac{1}{m+4}}).$$

Proof. The first result is a direct consequence of (2.2), Lemmas 4 and 6 and the definitions of h and k. Since, by definition $\underline{\psi}^{**}$ is in $\times_n [-\alpha,+\alpha]^m$, the second result follows from the first result and Lemma 2 with $\sigma^2 = 1$.

§1.3 Rates Near $O(n^{-\frac{1}{2}})$ for $D_n(\underline{\theta},\underline{\hat{\psi}})$ with $\underline{\hat{\psi}}$ Based on Kernel Estimators for a Density and its Derivative

In this section, for each positive integer s and γ in $(0,1)$, we exhibit a procedure $\underline{\hat{\psi}}$ belonging to a class of procedures whose modified regret $D_n(\underline{\theta},\underline{\hat{\psi}})$ is $O(n^{-(s-1)\gamma/(2s+m)(1+\gamma)})$. The definition of $\underline{\hat{\psi}}$ depends on kernel estimators for a density and its derivative. These kernel estimators are similar to those defined by Johns and Van Ryzin (1967) for estimating the unconditional density and its derivative in the empirical Bayes two-action problem in exponential families.

For $\ell = 0,1,\ldots,m$, let K_ℓ be bounded with $\mu[\|u\|^s K_\ell] = s!c_{\ell s} < \infty$ and for all nonnegative integers $t_1,\ldots,t_m$,

$$(3.1)\qquad \mu[\prod_{j=1}^{m} u_j^{t_j} K_0] = 1 \text{ or } 0 \text{ as } \Sigma\, t_j = 0 \text{ or in } \{1,\ldots,s-1\}$$

and, for $1 \le \ell \le m$, $u_\ell K_\ell$ satisfies (3.1) with s replaced by $s-1$.

As a result of these conditions on $K_0,\ldots,K_m$ and their intent, if f is a function on R^m with partials of order s uniformly bounded by M, then the substitution of the sth order Taylor expansion with Lagrange's form of the remainder shows

$$(3.2)\qquad |\mu[f\,K_0] - f(0)| \le M\,c_{0s}$$

and if, in addition, all partials of f not involving the ℓth variable vanish at 0, (3.3) $$|\mu[f\, K_\ell] - f_\ell(0)| \le M\, c_{\ell s}$$

where f_ℓ stands for the first partial of f wrt the ℓth variable.

The notation to be introduced below is defined for each n. We abbreviate by omission the dependency on n of the functions to be defined below. We let Σ denote summation over j from 1 to $n-1$. Let ϵ, δ be positive. As in section §1.2, let X abbreviate $\underset{\sim}{X}_n$. Define

(3.4) $$\hat{p}_j(X) = \epsilon^{-m} K_0(\epsilon^{-1}(\underset{\sim}{X}_j - X)),\ (n-1)\bar{\hat{p}} = \Sigma\, \hat{p}_j$$

and $\bar{\hat{q}} = \langle \bar{\hat{q}}_\ell \rangle$ where

(3.5) $$(n-1)\bar{\hat{q}}_\ell = \Sigma\, \hat{q}_{\ell j} \quad \text{with} \quad \delta^{m+1}\hat{q}_{\ell j}(X) = \frac{1}{2} K_\ell(I_\ell \delta^{-1}(\underset{\sim}{X}_j - X)) - K_\ell(\delta^{-1}(\underset{\sim}{X}_j - X))$$

where I_ℓ is the $m \times m$ identity matrix reduced by 1/2 in the ℓth diagonal element.

Now we state and prove some lemmas which will be useful in obtaining a rate of convergence for the modified regret of a certain procedure $\underset{\sim}{\hat{t}}$ to be defined in the latter part of the section. Let $c_1, c_2, \ldots$ denote finite functions of σ^2. In the following lemmas, $\bar{p}$, the average of the densities of $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_{n-1}$ and $\bar{q}$, the vector of partial derivatives of $\bar{p}$ are evaluated at X. We do not require the condition that $|\theta_n| \le \alpha$ to prove lemmas 7 and 8.

<u>Lemma 7</u>. $\underset{\sim}{P}_{n-1}[|\bar{\hat{p}} - \bar{p}|] \le c_1\, (\epsilon^s + ((n-1)\epsilon^m)^{-\frac{1}{2}})$.

Proof. Since $\mu[p_j\ \varepsilon^{-m}K_0(\varepsilon^{-1}(\cdot - X))] = \mu[p_j(X + \varepsilon\cdot)K_0]$, its absolute difference from $p_j(X)$, by the uniform boundedness of partials of order s of p_j and (3.2), is at most $c_2\ \varepsilon^s$. Hence

$$(3.6) \qquad |\underline{P}_{n-1}[\bar{\hat{p}}] - \bar{p}| \le c_2\ \varepsilon^s\ .$$

Let $V_X(\bar{\hat{p}})$ denote the conditional variance of $\bar{\hat{p}}$ given X. Since

$$\mu[p_j\varepsilon^{-2m}K_0^2(\varepsilon^{-1}(\cdot - X))] = \varepsilon^{-m}\ \mu[p_j(X + \varepsilon\cdot)K_0^2] \le \varepsilon^{-m}(2\pi\sigma^2)^{-m/2}\mu[K_0^2]$$

and $\mu[K_0^2] < \infty$,

$$(3.7) \qquad V_X(\bar{\hat{p}}) \le c_3\ ((n-1)\varepsilon^m)^{-1}\ .$$

Since for any random variable R, $E|R| \le |ER| + \mathrm{Var}^{\frac{1}{2}}(R)$, (3.6) and (3.7) will yield the bound in the lemma with $c_1 = c_2 \vee c_3$.

Since $\sigma^2\|\bar{q}\|/\bar{p} \le \sqrt{m}\ \alpha + \|X\|$ and since $|\theta_n| \le \alpha$ implies that $P_n[\|X\|]$ is uniformly bounded, the following corollary is a direct consequence of Lemma 7.

Corollary 1. $\quad \underline{P}_n[\|\bar{q}\|\,|(\bar{\hat{p}}/\bar{p}) - 1|] \le c_4(\varepsilon^s + ((n-1)\varepsilon^m)^{-\frac{1}{2}})$.

Lemma 8. $\quad \underline{P}_n[\|\bar{\hat{q}} - \bar{q}\|] \le c_5(\delta^{s-1} + ((n-1)\delta^{m+2})^{-\frac{1}{2}})$.

Proof. In this proof, we abbreviate by omission the indication of the ℓth coordinates of $\bar{\hat{q}}$ and $\bar{q}$. Since, by two usages of the transformation theorem,

$$\mu[p_j\ \hat{q}_{\ell j}] = \delta^{-1}\mu[K_\ell(p_j(X + I_\ell^{-1}\delta\cdot) - p_j(X + \delta\cdot))],$$

its absolute difference from the partial derivative of p_j wrt the ℓth coordinate, by the uniform boundedness of partials of order s of p_j and (3.3), is at most $c_6 \delta^{s-1}$. Hence,

(3.8) $$|\underline{P}_{n-1}[\bar{\hat{q}}] - \bar{q}| \le c_6 \delta^{s-1} .$$

Let $V_X(\bar{\hat{q}})$ denote the conditional variance of $\bar{\hat{q}}$ given X. By the inequality $(a+b)^2 \le 2(a^2+b^2)$ for a,b in R^1 and the transformations as above, we have

$$\mu[p_j(\hat{q}_{\ell j})^2] \le (\delta^{m+2})^{-1}\mu[K_\ell^2(\tfrac{1}{2} p_j(X + I_\ell \delta\cdot) + 2p_j(X + \delta\cdot))].$$

Hence, since $\mu[K_\ell^2] < \infty$,

(3.9) $$V_X(\bar{\hat{q}}) \le c_7 \, ((n-1)\delta^{m+2})^{-1} .$$

Since for any random variable R, $E|R| \le |ER| + \mathrm{Var}^{\frac{1}{2}}(R)$, inequalities (3.8) and (3.9) yield the bound in the lemma $c_5 = c_6 \vee c_7$.

Lemma 9. For any a in $(0,1)$, there exists a finite function of σ^2, c_8, such that

$$P_n[\bar{p} < \beta] \le c_8 \, \beta^a .$$

Proof. With $\sigma Z = X - \theta_n$ and therefore $\sigma^{-1}|X - \theta_j| \le |Z| + 2\sigma^{-1}\alpha$,

(3.10) $$p_j(X) = c \exp - \frac{|X-\theta_j|^2}{2\sigma^2} \ge c \exp - \tfrac{1}{2}(|Z| + 2\sigma^{-1}\alpha)^2$$

with $c = (2\pi\sigma^2)^{-m/2}$. Let M be the minimum value of $|Z|$ for which rhs of (3.10) $\le \beta$. Since, for all t, $P[|Z|^2/2 > t] \le e^{-bt}(1-b)^{-m/2}$ for b in $(0,1)$, we get from

(3.10) that

$$\beta^{-a} P_n[\bar{p} < \beta] \le \beta^{-a} P_n[|Z| > M] \le c^{-a} e^{-\frac{a}{2}(M+2\sigma^{-1}\alpha)^2 - bM^2/2} (1-b)^{-m/2}$$

which is bounded in M for $b > a$.

Corollary 2. For any a in (0,1), there exists a function of σ^2, c_9, such that

$$P_n[\frac{\|\bar{q}\|}{\bar{p}}[\bar{p} < \beta]] \le c_9 \beta^a .$$

Proof. Since $\sigma^2\|\bar{q}\|/\bar{p} \le \sqrt{m}\,\alpha + \|X\|$ and, therefore, has all moments, Hölders inequality yields, for any $r > 1$, the bound

$$P_n^{\frac{r-1}{r}}[(\frac{\|\bar{q}\|}{\bar{p}})^{\frac{r}{r-1}}] P_n^{\frac{1}{r}}[\bar{p} < \beta]$$

for the lhs of the corollary. By Lemma 9, $P_n^{\frac{1}{r}}[\bar{p} < \beta] \le c_8^{\frac{1}{r}} \beta^{\frac{b}{r}}$ for b in (a,1). Choosing r such that $a\,r = b$, we get the result of the corollary.

Henceforth, we take δ to minimize the bound in Lemma 8. That is,

$$\delta^{2s+m} = (n-1)^{-1} . \tag{3.11}$$

We also choose ε to be such that

$$\delta^{\frac{m+2}{2}} \le \varepsilon \le \delta^{\frac{s-1}{s}} . \tag{3.12}$$

Let β be defined by

$$\beta^{1+\gamma} = \delta^{s-1} \qquad \text{for any } \gamma \text{ in } (0,1).$$

With these choices for ε, δ and β, we define Ψ as follows. Let

$$\hat{\psi} = \mathrm{tr}'(X + \sigma^2 \frac{\hat{\bar{q}}}{\hat{\bar{p}}'}) \tag{3.13}$$

where tr' stands for retraction to $[-\alpha,+\alpha]^m$ and for y in R^1, let $y' = y \vee \beta$.

In the following lemma, ψ is evaluated at X.

Lemma 10. For each positive integer s and γ in $(0,1)$, there exists c_{10} such that if $\delta^{2s+m} = (n-1)^{-1}$, $\delta^{\frac{m+2}{m}} \le \epsilon \le \delta^{\frac{s-1}{2}}$ and $\beta^{1+\gamma} = \delta^{s-1}$, then

$$\underline{P}_n[\|\hat{\psi} - \psi\|] \le c_{10}(n-1)^{-\frac{s-1}{2s+m}\frac{\gamma}{1+\gamma}} \quad \text{for each } n > 1 .$$

Proof. Since ψ lies in the m-sphere of radius α and $\hat{\psi}$ is the retraction of $X + \sigma^2 \hat{\bar{q}}/\hat{\bar{p}}'$ to $[-\alpha,+\alpha]^m$, we have by using the inequality $\hat{\bar{p}}' \ge \beta$

$$\sigma^{-2}\|\hat{\psi} - \psi\| \le \|\frac{\hat{\bar{q}}}{\hat{\bar{p}}'} - \frac{\bar{q}}{\bar{p}}\| \le \frac{1}{\beta}\|\hat{\bar{q}} - \frac{\bar{q}}{\bar{p}}\hat{\bar{p}}'\| \le \frac{1}{\beta}\{\|\hat{\bar{q}} - \bar{q}\| + \frac{\|\bar{q}\|}{\bar{p}}|\bar{p} - \hat{\bar{p}}'|\}.$$

Since $|\bar{p} - \hat{\bar{p}}'| \le |\bar{p} - \hat{\bar{p}}| + \beta[\bar{p} < \beta]$, the result of the lemma follows from the above inequality, Lemma 8, Lemma 7 and Corollary 2 and the hypothesis on ϵ, δ and β.

Now we state the main result of this section.

Theorem 2. If $\sigma^2 = 1$, the hypothesis of Lemma 10 is satisfied and $\hat{\underline{\psi}}$ is defined by (3.13), then

$$D_n(\underline{\theta},\hat{\underline{\psi}}) = O(n^{-\frac{s-1}{2s+m}\frac{\gamma}{1+\gamma}}) .$$

Proof. Since $\hat{\underline{\psi}}$, by definition (3.13), lies in $\times_n [-\alpha,+\alpha]^m$, the theorem is a consequence of Lemma 2 with $\sigma^2 = 1$ and Lemma 10.

§1.4 <u>Rates Near</u> $O(n^{-\frac{1}{2}})$ <u>for</u> $D_n(\underline{\theta}, {}_o\hat{\psi})$ <u>where</u> ${}_o\hat{\psi}$, <u>a Particular</u> $\hat{\psi}$

Let $\sigma^2 = 1$ and let $s > 1$ be a fixed integer throughout this section. Letting ${}_o\hat{\psi}$ denote a specialization, less a retraction to $[\beta, \infty)$, of the $\hat{\psi}$ of section §1.3, with certain additional assumptions on the kernels, we show that $D_n(\underline{\theta}, {}_o\hat{\psi}) = O(n^{-(s-1)/2(m+s+1)})$.

We specialize $\bar{p}$ and $\bar{q}$ (defined by (3.4) and (3.5) respectively) by setting $\varepsilon = \delta$ and denote their common value by h. Let

$$ {}_o\hat{\psi} = \operatorname{tr}'(X + \frac{\bar{\hat{q}}}{\bar{\hat{p}}}) \tag{4.1} $$

where tr' (as in previous sections §1.2 and §1.3) stands for retraction to the cube $[-\alpha, +\alpha]^m$ and any undefined ratios are taken to be zero. Let $hZ_j = \underset{\sim}{X}_j - X$, $v = h(u + \bar{q}_\ell/\bar{p})$ and $Y_j(u) = \langle Y_{\ell j}(u)\rangle$ with

$$ Y_{\ell j}(u) = (\tfrac{1}{2} K_\ell \circ I_\ell - K_\ell - v\, K_0) \circ Z_j = h^{m+1}\hat{q}_{\ell j} - h^m v\, \hat{p}_j . \tag{4.2} $$

In the following lemma, ψ will be evaluated at X. Let $c_1, c_2, \ldots$ denote constants.

<u>Lemma 11</u>. If $K_0, \ldots, K_m$ are bounded with $\mu[\|u\|^s K_\ell] = c_{\ell s} < \infty$, K_0 satisfies (3.1) and $u_1K_1, \ldots, u_mK_m$ satisfy condition (3.1) with s replaced by s-1 and are such that for $|u| \le 2\alpha$, $h \le s^{-1}\alpha$,

$$ c_1\, e^{-c_2|X|} \le \frac{\operatorname{Var}(Y_{\ell j}),\ \operatorname{Var}(K_0 \circ Z_j)}{h^m \phi(|X|)} \le c_3\, e^{c_4|X|}, \tag{4.3} $$

then $\underline{P}_n[\|{}_o\hat{\psi} - \psi\|] \le c_5(((n-1)h^{2s+m})^{\frac{1}{2}} + \dfrac{1}{((n-1)h^{m+2})^{\frac{1}{2}}})$.

Proof. Let the indication of the ℓth coordinate of ${}_o\hat{\psi}$ and ψ be abbreviated by omission. Since ${}_o\hat{\psi}$ lies in $[-\alpha,+\alpha]$ and since ψ lies in $[-\alpha,+\alpha]$, it follows that $|{}_o\hat{\psi} - \psi| \le 2\alpha$ and $|{}_o\hat{\psi} - \psi| \le |D|$ where

$$D = \frac{\hat{\bar{q}}_\ell}{\hat{\bar{p}}} - \frac{\bar{q}_\ell}{\bar{p}}. \tag{4.4}$$

Therefore

$$\underline{P}_{n-1}[|{}_o\hat{\psi} - \psi|] \le \int_0^{2\alpha} \underline{P}_{n-1}[|D| > u]du = \int_0^{2\alpha} \underline{P}_{n-1}[D > u]du + \int_{-2\alpha}^{0} \underline{P}_{n-1}[D < -u]du. \tag{4.5}$$

The main part of the proof bounds the integrands of the rhs of this inequality by using the Berry-Esseen theorem and (4.3). The rest of the proof shows that the P_n-integral of a bound for the rhs of (4.5) is at most the bound in the lemma.

With $\beta^2 = \mathrm{Var}(\Sigma\, Y_{\ell j})$ and $L = \beta^{-3}\Sigma\, P_j|Y_{\ell j} - P_j Y_{\ell j}|^3$, the standardized range bound for L, together with lhs inequality of (4.3), the inequality

$$|v| \le h(3\alpha + |X_\ell|) \tag{4.6}$$

and the fact that $K_0,\ldots,K_m$ are bounded, implies that

$$L \le \frac{c_7(1 + h(3\alpha + |X_\ell|))}{c_1((n-1)h^m)^{\frac{1}{2}}\phi^{\frac{1}{2}}(|X|)e^{-c_2|X|/2}} \quad \text{for} \quad |u| \le 2\alpha . \tag{4.7}$$

Let $0 \le u \le 2\alpha$. Then the definitions of D in (4.4), $Y_{\ell j}$ in (4.2) imply that $[D > u] \le [\Sigma\, Y_{\ell j} > 0] + [\hat{\bar{p}} < 0]$. The Berry-Esseen theorem (Loève (1963), p. 288) and the triangle inequality imply that $\underline{P}_{n-1}[\Sigma Y_{\ell j} > 0]$ is at most

$$(4.8)\qquad \Phi(-(n-1)h^{m+1}\beta^{-1}\bar{p}\,u) + |\Phi(-(n-1)h^{m+1}\beta^{-1}\bar{p}\,u) - \Phi(\beta^{-1}\Sigma\, P_jY_j)| + B\,L.$$

Since rhs inequality of (4.3) implies that $\beta^2 \le c_3h^m\phi(|X|)e^{c_4|X|}$, the first term in (4.8) can be bounded above by replacing β by this upper bound for β. Also, by the equality

$$(4.9)\qquad (n-1)h^{m+1}\bar{p}\,u + \Sigma\, P_jY_{\ell j} = (n-1)h^{m+1}(h^{-1}v(\bar{p} - \underline{P}_{n-1}[\hat{p}]) + \underline{P}_{n-1}[\hat{q}] - \bar{q}),$$

the lhs inequality in (4.3), the bounds (3.6), (3.8) and the inequality (4.6) imply that the second term in (4.8) is at most

$$(4.10)\qquad \frac{c_8((n-1)h^{2s+m})^{\frac{1}{2}}(1 + h(3\alpha + |X_\ell|))}{c_1^{\frac{1}{2}}\,\phi^{\frac{1}{2}}(|X|)\,e^{-c_2|X|/2}}\,.$$

Hence, with f defined as the positive solution of the equation

$$(4.11)\qquad c_3e^{c_4|X|}\phi(|X|)f^2 = \bar{p}^{-2},$$

we obtain that

$$(4.12)\, \underline{P}_{n-1}[\Sigma Y_{\ell j} > 0] \le \Phi(-((n-1)h^{m+2})^{\frac{1}{2}}f\,u) + (4.10) + B \text{ rhs of } (4.7)$$

Now we consider $-2\alpha \le u < 0$. The definitions of D in (4.4) and $Y_{\ell j}$ in (4.2) imply that $[D < u] \le [\Sigma\, Y_{\ell j} < 0] + [\bar{\hat{p}} \le 0]$. The Berry-Esseen theorem and the triangle inequality imply that $\underline{P}_{n-1}[\Sigma Y_{\ell j} < 0]$ is at most

$$(4.13)\quad \Phi((n-1)h^{m+1}\beta^{-1}\bar{p}\,u) + |\Phi((n-1)h^{m+1}\beta^{-1}\bar{p}\,u) - \Phi(-\beta^{-1}\Sigma P_jY_{\ell j})| + BL.$$

Since the rhs inequality of (4.3) implies that $\beta^2 \le c_3 h^m \phi(|X|) e^{c_4|X|}$, the first term in (4.13) is bounded by $\Phi(((n-1)h^{m+2})^{\frac{1}{2}} f\ u)$ where f is the positive solution of (4.13). The lhs inequality of (4.3), the equality (4.9) and the bounds (3.6), (3.8) and (4.6) imply that the second term of (4.13) is at most (4.10). Therefore,

$$(4.14)\quad \underline{P}_{n-1}[\Sigma Y_{\ell j}<0] \le \Phi(((n-1)h^{m+2})^{\frac{1}{2}} f\ u) + (4.10) + B \text{ rhs of } (4.7).$$

Integrating (4.12) wrt u over $[0, 2\alpha]$ and (4.14) wrt u over $[-2\alpha,0)$, then bounding their first terms by using the inequality $\int_0^{2\alpha} \Phi(-At)dt \le A^{-1}$ for $A > 0$, we obtain (since the corresponding Berry-Esseen, followed by normal tail bound, treatment of $\int_0^{2\alpha} \underline{P}_{n-1}[\bar{p} \le 0]du$ contributes no more than $1+\alpha^2/s$ times the rest) that $\underline{P}_{n-1}[|{}_0\hat{\psi}-\psi|]$ is at most

$$\left\{\frac{1}{((n-1)h^{m+2})^{\frac{1}{2}}}\frac{1}{f} + 4\alpha[(4.10) + B \text{ rhs of } (4.7)]\right\}\left(2 + \frac{\alpha^2}{s}\right).$$

Hence we complete the proof of the lemma by showing that the P_n-integrals of f^{-1} and $(1 + h(3\alpha + |X_\ell|))e^{c_2|X|/2}\phi^{-\frac{1}{2}}(|X|)$ are uniformly bounded.

Since $(2\pi)^m p_n^2 \le \exp -((|X| - \alpha)^+)^2$ and $(2\pi)^{m} p^{-2} \ge \exp -(\alpha+|X|)^2$, we obtain from the definition of f in (4.11) that $p_n f^{-1}$ is at most

$$\frac{c_8\ \phi((|X| - \alpha)^+)\phi^{\frac{1}{2}}(|X|)e^{c_4|X|/2}}{\phi(|X| + \alpha)}$$

which is μ-integrable. Again by using the upper bound p_n, we can show that the P_n-integral of $(1 + h(3\alpha + |X_\ell|))e^{c_2|X|/2}\phi^{\frac{1}{2}}(|X|)$ is uniformly bounded. This ends the proof of the lemma.

Now we state the main result of the section.

<u>Theorem 3</u>. If the kernel functions $K_0, \dots, K_m$ satisfy the conditions of Lemma 11, $h = a\, n^{-1/m+s+1}$ where $0 < a \le s^{-1}$ and ${}_{o}\hat{\underline{\theta}}$ is defined by (4.1), then

$$D_n(\underline{\theta}, {}_{o}\hat{\underline{\theta}}) = O(n^{-(s-1)/2(s+m+1)}).$$

<u>Proof</u>. Since ${}_{o}\hat{\underline{\theta}}$ lies in $\times_n [-\alpha,+\alpha]^m$, the result of the theorem is a direct consequence of Lemma 11, the hypothesis on h and Lemma 2.

Now we exhibit kernel functions $K_0, \dots, K_m$ satisfying the conditions of Lemma 11. We develop these kernels in $m = 2$ case for the sake of simplicity of the notation.

Let $[c_{ij}]$ be an $\infty \times \infty$ matrix whose ijth element is c_{ij}. For each pair of positive integers i,j, let $\neg^{i,j}$ be the indicator function of the south-west quadrant of (i,j) intersected with the north-east quadrant of $(0,0)$. We will determine $[a_{ij}]$, $[b_{ij1}]$ and $[b_{ij2}]$ with only finitely many entries different from zero such that

$$(4.15) \qquad K_0 = \sum_{i,j} a_{ij} \neg^{i,j}, \quad K_1 = \sum_{i,j} b_{ij1} \neg^{i,j} \quad \text{and} \quad K_2 = \sum_{i,j} b_{ij2} \neg^{i,j}$$

satisfy the conditions of Lemma 11.

For any two positive integers S, T, let $[a_{ij}]_{S,T}$ denote the modification of $[a_{ij}]$ obtained by replacing a_{ij} by zero if $i > S$ or $j > T$. We note that for any two sets of distinct non-negative integers, $k_1, \dots, k_S$ and $\ell_1, \dots, \ell_T$, the vectors

$$(4.16) \qquad [i^{k_1} j^{\ell_1}]_{S,T}, \dots, [i^{k_S} j^{\ell_T}]_{S,T} \quad \text{are a basis for } R^{ST}.$$

(For $\Sigma c_{rt}[i^{k_r} j^{\ell_t}] = [0]$ iff $\Sigma c_{rt} x^{k_r} y^{\ell_t} = 0$ has the roots $\{1,\ldots,S\} \times \{1,\ldots,T\}$, which by iterative application of Descarte's rule of signs requires the c_{rt} to vanish.) We use this fact to show that certain norms are different from zero and to show that certain coefficients are zero. The kernel conditions (3.1) on K_0 and K_1 specialize to the following requirements on inner products,

$$([a_{ij}], [i^{\ell_1} j^{\ell_2}]) = \begin{matrix} 1 & \ell_1 = \ell_2 = 1 \\ 0 & 1 \le \ell_1, \ell_2,\ 3 \le \ell_1 + \ell_2 \le s+1 \end{matrix}$$

and

$$([b_{ij1}], [i^{\ell_1} j^{\ell_2}]) = \begin{matrix} 2 & \ell_1 = 2,\ \ell_2 = 1 \\ 0 & 4 \le \ell_1 + \ell_2 \le s+1 \end{matrix} .$$

We choose $[a_{ij}]$ for simplicity to be the

(4.17) projection of $[ij]_{s,s}$ on $\perp \{[i^{\ell_1} j^{\ell_2}]_{s,s} | 1 \le \ell_1, \ell_2, 3 \le \ell_1 + \ell_2 \le s+1\}$ divided by its squared norm,

and in order to satisfy the variance requirements (4.3), we take $[b_{ij1}]$ to be

(4.18) projection of $[i^2 j]_{s,s}$ on $\perp \{[i^{\ell_1} j^{\ell_2}]_{s,s} | (\ell_1, \ell_2) \ne (2,1), 1 \le \ell_1 \le s, 1 \le \ell_2 \le s\}$ divided by its squared norm.

The squared norms are non-zero by the aforenoted linear independence for $(S,T) = (s,s)$. Moreover, $b_{sj1} \ne 0$ for some j in $\{1,\ldots,s\}$ for, otherwise $[b_{ij1}]$ defined in (4.18) will lie in $R^{(s-1)s}$ and is orthogonal to a basis in $R^{(s-1)s}$, hence is 0. Let $M = \mathrm{Max}\{j | b_{sj} \ne 0\}$. Interchanging i and j, we get a solution for $[b_{ij2}]$ such that K_2 satisfies the kernel conditions cultminating in (3.1).

With A denoting a bound of K_0, K_1 and K_2,

(4.19) $\quad \mathrm{Var}(Y_{\ell j}) \le A^2(\frac{3}{2} + v)^2 \; \underline{P}_{n-1}[\underset{\sim}{X}_j \in (X, X + sh) \times (X, X + sh)]$.

By the mean value theorem, the probability on the rhs of this inequality is $s^2h^2p_j(X + \xi sh)$ for some ξ in the unit square. Hence, factoring out $h^2\phi(|X|)$, the restriction $h \le s^{-1}\alpha$ and the inequality (4.6) show that the rhs of (4.19) is bounded by the rhs of (4.3) for suitable c_3 and c_4.

Now we observe that $Y_{\ell j}$ defined by (4.2) takes finite number of values including zero and $2^{-1}b_{sM}$. The probability that it takes the value zero is $\underline{P}_{n-1}[\underset{\sim}{X}_j - X \notin (0, sh) \times (0, sh)]$ and that it takes $2^{-1}b_{sM}$ is $\underline{P}_{n-1}[\underset{\sim}{X}_j - X \in (2(s-1)h, 2sh) \times ((M-1)h, Mh)]$. Therefore by L1.A of the Appendix, we obtain that

(4.20) $\quad \mathrm{Var}(Y_{1j}) \ge c_9 \; \underline{P}_{n-1}[\underset{\sim}{X}_j - X \in (2(s-1)h,\ 2sh) \times ((M-1)h,\ Mh)]$.

By the mean value theorem, the probability on the rhs of this inequality is $h^2p_j(X + \xi h)$ for some ξ in $(2(s-1), 2s) \times (M-1, M)$. Hence, factoring out $h^2\phi(|X|)$, the restriction $h \le s^{-1}\alpha$ shows that the rhs of (4.20) is bounded below by (4.3) for suitable c_1 and c_2 when $\ell = 1$. Similarly that $\mathrm{Var}(Y_{2j})$ is bounded by lhs of (4.3) can be similarly proved.

By following the argument given above, we can show that $\mathrm{Var}(K_0 \circ Z_j)$ also satisfies inequality (4.3).

§1.5 A Lower Bound for $D_n(\underline{0},\underline{\psi}^{**})$.

In this section, we use the notation of section §1.2 specialized to the $\sigma^2 = 1$ case. Let $c_1, c_2, \ldots$ denote absolute constants. With

$$\beta^2 = (n-1)k^2h^m, \tag{5.1}$$

by using the Berry-Esseen theorem and Lemma 1 of the Appendix, we show that $D_n(\underline{0},\underline{\psi}^{**}) \geq c_1^2\,\beta^{-2}$ under certain conditions on β.

Theorem 4. If $\beta(\frac{k}{2} + h) \to a < \infty$, $\beta \to \infty$ and $\underline{\psi}^{**}$ is defined by (2.1), then

$$D_n(\underline{0},\underline{\psi}^{**}) \geq c_1^2\,\beta^{-2}.$$

Proof. Let the first coordinate of ψ_n^{**} be abbreviated by ψ^{**} and let the indication of the first coordinate of t^* be abbreviated by omission. As in section §1.2, let X, with coordinates $X_1,\ldots,X_m$, abbreviate $\underset{\sim}{X}_n$. Our method of proof is to show that $\underline{P}_n[[X_1 > \alpha]|\psi^{**}|]$ exceeds the square-root of the bound of the theorem. This completes the proof of the theorem since $D_n(\underline{0},\underline{\psi}^{**}) = n^{-1}\sum_{j=1}^{n}\underline{P}_j[|\psi_j^{**}|^2]$ and $\underline{P}_n[|\psi_n^{**}|^2] \geq \underline{P}_n^2[|\psi_n^{**}|] \geq \underline{P}_n^2[[X_1 > \alpha]|\psi^{**}|]$.

Since, by definition, $|\psi^{**}| \leq \alpha$ and since $[|\psi^{**}| > u] = [|X_1 + t^*| > u]$ for $u < \alpha$, we obtain by Fubini's theorem that

$$\underline{P}_n[|\psi^{**}|] = \int_0^\alpha \underline{P}_n[|X_1 + t^*| > u]du \geq P_n[[X_1 > \alpha]\int_0^\alpha \underline{P}_{n-1}[X_1 + t^* > u]du]. \tag{5.2}$$

Let X in $(\alpha,\infty) \times R^{m-1}$ and u be in $(0,\alpha)$ fixed until otherwise stated. As in section §1.2, let

$\tilde{\delta}_j = [\underset{\sim}{X}_j \in \sqcap_1]$, $\delta_j = [\underset{\sim}{X}_j \in \sqcap]$ and

$$(5.3) \qquad Y_j = \tilde{\delta}_j - \delta_j \, e^{k(u-X_1)} .$$

With this definition of Y_j, we obtain that

$$(5.4) \qquad [X_1 + t^* > u] = [\Sigma Y_j \geq 0]$$

where Σ, as in sections §1.2, §1.3 and §1.4, denotes summation over j from 1 to n-1. Note that $X_1 > \alpha$ implies that $[\Sigma Y_j \geq 0, \Sigma \tilde{\delta}_j = 0, \Sigma \delta_j = 0] \subseteq [X_1 + t^* > u]$ for $u < \alpha$.

Since $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_{n-1}$ are i.i.d., so are $Y_1, \ldots, Y_{n-1}$. Hence, with B denoting the Berry-Esseen constant, the Berry-Esseen theorem and (5.4) give that

$$(5.5) \qquad \underset{\sim}{P}_{n-1}[X_1 + t^* > u] \geq \Phi\left(\frac{(n-1)^{\frac{1}{2}} P_1 Y_1}{\text{s.d.} Y_1}\right) - B(n-1)^{-\frac{1}{2}} \frac{P_1 |Y_1 - P_1 Y_1|^3}{(\text{s.d.} Y_1)^3} .$$

The definition of Y_1 gives that $P_1 Y_1 = F_1 \sqcap_1 - e^{k(u-X_1)} F_1 \sqcap$. Hence, since the alternative expression for $F_j \sqcap_\ell / F_j \sqcap$ in the proof of (2) of Lemma 3 when specialized to the case $\sigma^2 = j = \ell = 1$ gives that $F_1 \sqcap_1 / F_1 \sqcap = \exp - k(X_1 + \frac{k}{2} + \omega h)$ for some ω in $(0,1)$, we obtain that

$$(5.6) \qquad P_1 Y_1 = F_1 \sqcap_1 e^{k(u-X_1)} \left(e^{-k(u + \frac{k}{2} + \omega h)} - 1\right)$$

$$\geq -k \, F_1 \sqcap_1 (u + \frac{k}{2} + h) \quad \text{for} \quad k < (\alpha + 4)^{-1}$$

where the inequality follows from the inequalities $u < \alpha < X_1$ and $e^{-\lambda} - 1 \geq -\lambda$.

Applying L1.A (see Appendix) to the random variable Y_1, we obtain, since Y_1 takes value 1 with probability $F_1 \sqcap_1$, that Var (Y_1) is at least $(1 - F_1\square_1 - F_1\square)F_1\square_1(1 - F_1\square_1)$. Hence, since $(1 - F_1\square_1 - F_1\square)$ is bounded away from zero for $h < (\alpha + 4)^{-1}$, we obtain that for some $c_2 > 0$

$$(5.7) \qquad \text{Var } (Y_1) \geq c_2^2 F_1\square_1 \qquad \text{for} \quad h < (\alpha + 4)^{-1}.$$

Using (5.6), (5.7) and the definition of β in (5.1), we obtain that, for $k < (\alpha + 4)^{-1}$,

$$(5.8) \quad (n-1)^{\frac{1}{2}} \frac{P_1 Y_1}{s.d.Y_1} \geq - \frac{\beta}{c_2^{\frac{1}{2}}} (u + \frac{k}{2} + h)f \quad \text{with} \quad h^{m/2} f = (F_1\square_1)^{\frac{1}{2}}.$$

The standardized range bound for $P_1|Y_1 - P_1Y_1|^3/(s.d.Y_1)^3$, with the help of the inequality range of $Y_1 \leq 1 + e^{(u-X_1)k} \leq 2$ since $u < \alpha < X_1$, (5.7) and the definition of β, gives that, for $h < (\alpha + 4)^{-1}$,

$$(5.9) \qquad \frac{P_1|Y_1 - P_1Y_1|^3}{(n-1)^{\frac{1}{2}}(s.d.Y_1)^3} \leq \frac{2k}{\beta \; c_2^{\frac{1}{2}} f} .$$

Integrating the inequality obtained by weakening (5.5) with the help of (5.8) and (5.9) wrt u over $(0,\alpha)$, then using the transformation $\beta(u + \frac{k}{2} + h)f = c_2^{\frac{1}{2}}v$ in the first integral, we obtain that

$$\beta \int_0^\alpha P_{n-1}[X_1 + t^* > u]du \geq \frac{c_2^{\frac{1}{2}}}{f} \int_{(\frac{k}{2}+h)f/c_2^{\frac{1}{2}}}^{(\alpha+\frac{k}{2}+h)/c_2^{\frac{1}{2}}} \Phi(-v)dv - \frac{2k\alpha}{c_1^{\frac{1}{2}} f} \quad \text{for} \quad k < (\alpha+4)^{-1}.$$

In view of (5.2), we complete the proof by showing that the P_n-integral of the first term of the rhs of this inequality on $[X_1 > \alpha]$ converges to a positive constant while that of the second term on $[X_1 > \alpha]$ converges to zero.

Since f, defined in (5.8), converges to $p_1^{\frac{1}{2}}$ and since specialization of (2) of Lemma 3 to the case of $\sigma^2 = \ell = 1$ and $n = 2$ gives that f^{-1} is exceeded by $p_1^{-\frac{1}{2}} \exp(\|X\| + \frac{m+1}{2})$ which is P_n-integrable on $[X_1 > \alpha]$, it follows by dominated convergence theorem and the hypothesis on β that

$$P_n[[X_1 > \alpha] \int_{\beta(\frac{k}{2}+h)f/c_2^{\frac{1}{2}}}^{\beta(\alpha+\frac{k}{2}+h)/c_2^{\frac{1}{2}}} \Phi(-v)dv] \to P_1[[X_1 > \alpha] \int_{a(c_2 p_1)^{-\frac{1}{2}}}^{\infty} \Phi(-v)dv] > 0$$

and

$$\frac{2k\alpha}{c_2^{\frac{1}{2}}} P_n[[X_1 > \alpha]f] \to 0 .$$

The proof of the theorem is complete.

Now we make a remark concerning the procedures $\underline{\psi}^{**}$, $\underline{\Psi}$ and $_o\underline{\Psi}$ defined in sections §1.2, §1.3 and §1.4 respectively.

Remark 4. For the choice of h and k given in Theorem 1 of section §1.2, we obtain by the theorem proved above that

$$D_n(\underline{0},\underline{\psi}^{**}) \geq c\, n^{-\frac{2}{m+4}}$$

for some $c > 0$.

For any $\gamma > 0$, Theorem 2 of section §1.3 shows that we can define a procedure $\underline{\Psi}$ such that $D_n(\underline{\theta},\underline{\Psi}) = O(n^{-(\frac{1}{4}-\gamma)})$. Hence, since $\gamma > 1/36$ implies that $\frac{1}{4} - \gamma \geq \frac{2}{m+4}$ for $m \geq 5$, it follows that the procedure $\underline{\Psi}$ is better than $\underline{\psi}^{**}$ in the sense that

$$\sup D_n(\underline{\theta},\underline{\Psi}) \leq c_1\, n^{-(\frac{1}{4}-\gamma)} \leq c_2\, n^{-\frac{2}{m+4}} \leq \sup D_n(\underline{\theta},\underline{\psi}^{**})$$

where the sup is taken over all parameter sequences.

For any positive integer s, Theorem 3 of section §1.4 shows that we can define a procedure $\underline{{}_o\hat{\psi}}$ such that $D_n(\underline{\theta}, \underline{{}_o\hat{\psi}}) = O(n^{-s/2(s+3)})$. Hence, if $ms \geq 5m + 8$, the procedure $\underline{{}_o\hat{\psi}}$ is better than $\underline{\psi}^{**}$ in the sense described above.

§1.6 Extension of Results in Sections §1.2 and §1.3 to Constrained Mean Vectors and Unknown Covariance Matrix

Let Y be a d-variate normal with mean ω and covariance matrix $\sigma^2 I$. If ω is assumed to lie in a lower dimensional subspace $\mathcal{R}$, say of dimension $m < d$, then the square of the projection of Y onto the subspace orthogonal to $\mathcal{R}$ has expectation $\sigma^2(d-m)$ and variance $2\sigma^4(d-m)$. In this section, this fact has been used to extend the results of sections §1.2 and §1.3.

Let $\{\underset{\sim}{Y}_n\}$ be a sequence of independent random variables with $\underset{\sim}{Y}_n$ distributed as d-variate normal with unknown covariance matrix $\sigma^2 I$ and mean ω_n belonging to an m-dimensional subspace $\mathcal{R}_n$ of R^d intersected with the d-sphere of radius α. While stating the results of the present section in section §1.0, we interchanged m and d in order to make proper references to sections §1.2 and §1.3.

Let B_n be an orthogonal matrix whose first m columns generate $\mathcal{R}_n$. Let $\underset{\sim}{X}_n$ and θ_n denote the vectors formed by the first m coordinates of $B_n' \underset{\sim}{Y}_n$ and $B_n' \omega_n$ respectively where B_n' is the transpose of B_n. Let $(m-d)Z_n$ denote the square of the projection of $\underset{\sim}{Y}_n$ onto the subspace which is orthogonal to $\mathcal{R}_n$. Let E stand for expectation wrt the joint distribution of $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n, Z_1,\ldots,Z_n$.

This section is divided into two subsections. In the first subsection, with the help of the procedure $\underline{\psi}^{**}$ defined in (2.1), we exhibit a procedure $\underline{T}^{**}$ for which $D_n(\underline{\theta},\underline{T}^{**}) = O(n^{-1/(m+4)})$ for each σ^2. In the second subsection, for each positive integer s and each γ in $(0,1)$, with the help of the procedure $\underline{\hat{\psi}}$ defined by (3.13), we exhibit a $\underline{\hat{T}}$, for which

$D_n(\underline{\theta},\underline{\hat{T}}) = O(n^{-(s-1)\gamma/(2s+m)(1+\gamma)})$ for each σ^2. Let Σ denote summation over i from 1 to n.

§1.6.1 Definition of $\underline{T}^{**}$ and a Rate of Convergence for $D_n(\underline{\theta},\underline{\psi}^{**})$

In this subsection, we use the notation of section §1.2. We require the following notation for each n, but, as in earlier sections, we suppress the dependency on n of the functions to be defined below.

Define $\underline{T}^{**} = \{T^{**}\}$ as follows,

$$T^{**} = tr'(X + \frac{\Sigma Z_i}{n} tr_\lambda t^*) \tag{6.1}$$

where tr' (as in section §1.2) and tr_λ stand for retractions to $[-\alpha,+\alpha]^m$ and $\overset{m}{\underset{\ell=1}{\times}} [-\lambda^{-1}(|X_\ell| + \alpha + k + h), \lambda^{-1}(|X_\ell| + \alpha + k + h)]$ respectively.

Let T be the modification of T^{**} obtained by replacing $n^{-1}\Sigma Z_i$ in the definition of T^{**} by σ^2. Let T^* be the modification of T obtained by replacing tr' in the definition of T by retraction to the cube $[-\alpha',+\alpha']^m$ where $\alpha' = \alpha + k + h$. Let $c_1,c_2,\ldots$ denote finite functions of σ^2.

Lemma 12. $E\|T^* - T\| \le \dfrac{c_1}{n^{\frac{1}{2}}\lambda}$.

Proof. Since the distance between two points retracted in R^m to the same cube is at most the distance between the points and since $\lambda\|tr_\lambda t^*\| \le \|X\| + m\alpha'$, we obtain that $\lambda\|T^* - T\| \le (\|X\| + m\alpha')|n^{-1}\Sigma Z_i - \sigma^2|$. Since $(d-m)Z_1/\sigma^2, \ldots, (d-m)Z_n/\sigma^2$ are i.i.d. χ^2 - random variables with d-m degrees of freedom, application of Schwarz inequality to the rhs of the last inequality and the fact that $E[(\|X\| + m\alpha')^2]$ is bounded by a finite function of σ^2 completes the proof of the lemma.

Theorem 5. If $h = n^{-1/m+4}$, $k = a\, n^{-1/m+4}$ for a in $[1,\infty)$, $\lambda^{2(m+4)} = n^{-(m+2)}$ and $\underline{T}^{**}$ is defined by (6.1), then

$$D_n(\underline{\theta},\underline{T}^{**}) = O(n^{-\frac{1}{m+4}}) \quad \text{for each } \sigma^2.$$

Proof. Let σ^2 be fixed. In the proof, we consider only those n for which $\lambda < \sigma^2$.

Since $|\psi| \le \alpha$ and since $T = tr'T^*$, it follows that $\|T - \psi\| \le \|T^* - \psi\|$ and hence $\|T^{**} - \psi\| \le \|T^{**} - T\| + \|T^* - \psi\|$. If the ℓth coordinate of t^* (its negative) $> \lambda(|X_\ell| + \alpha')$, then, since $\lambda < \sigma^2$, T^* and ψ^* defined by (2.1) turn out to equal α' (its negative). Hence, $T^* = \psi^*$. Therefore the last inequality, together with Lemmas 12, 4 and 6 and the definitions of λ, h and k, implies that $E\|T^* - \psi\| = O(n^{-1/m+4})$. Since $\underline{T}^{**}$, by definition (6.1), takes values in $X_n[-\alpha,+\alpha]^m$, Lemma 2 and this order relation give the result of the theorem.

§1.6.2 Definition of $\hat{\underline{T}}$ and a Rate of Convergence of $D_n(\underline{\theta},\hat{\underline{T}})$

In this subsection, we use the notation of section §1.3. We require the following notation for each n, but as in previous sections, we suppress the dependency on n of the functions to be defined below.

Define $\underline{\hat{T}} = \{\hat{T}\}$ as follows,

$$\hat{T} = tr'(X + \frac{\Sigma Z_i}{n} tr_\lambda(\frac{\hat{q}}{\bar{\hat{p}}}))$$

where tr' (as in section §1.3) and tr_λ stand for retractions to $[-\alpha,+\alpha]^m$ and $\times_{\ell=1}^{m} [-\lambda^{-1}(|X_\ell| + \alpha), \lambda^{-1}(|X_\ell| + \alpha)]$ and $\bar{\hat{p}}' = \hat{p} \vee \beta$ ((3.13)). Let T be the modification of $\hat{T}$ obtained by replacing $n^{-1}\Sigma Z_i$ in the definition of $\hat{T}$ by σ^2.

As a consequence of replacing T^* by $\hat{T}$, T of subsection 1.6.2 by T of this subsection and α' by α in the proof of Lemma 12, we obtain the following lemma.

Lemma 13. $E\|\hat{T} - T\| \le \frac{c_1}{n^{\frac{1}{2}}\lambda}$.

Now we state and prove the main result of the subsection.

Theorem 6. If the hypothesis of Lemma 10 is satisfied, $\underline{\hat{T}}$ is defined by (1.6.2) and $\lambda = n^{\frac{s-1}{2s+m}\frac{\gamma}{1+\gamma} - \frac{1}{2}}$, then

$$D_n(\underline{\theta},\underline{\hat{T}}) = O(n^{-\frac{s-1}{2s+m}\frac{\gamma}{1+\gamma}}) \quad \text{for each } \sigma^2.$$

Proof. Let σ^2 be fixed. In the proof, we consider only those n for which $\lambda < \sigma^2$.

If the ℓth coordinate of T (its negative) $> \lambda(|X_\ell| + \alpha)$, then, since $\lambda < \sigma^2$, T and $\hat{\psi}$ defined by (3.13) turn out to equal α (its negative). Hence $T = \hat{\psi}$. Therefore the inequality $\|\hat{T} - \psi\| \le \|\hat{T} - T\| + \|\hat{\psi} - \psi\|$, together with Lemmas 13 and 10 and the hypothesis of the theorem, gives that $E\|\hat{T} - \psi\| = O(n^{-\frac{s-1}{2s+m}\frac{\gamma}{1+\gamma}})$. Since, by definition, $\hat{T}$ is in $[-\alpha,+\alpha]^m$, this order relation and Lemma 2 complete the proof of the theorem.

CHAPTER II

RATES IN THE ESTIMATION AND TWO-ACTION PROBLEMS FOR A FAMILY OF SCALE PARAMETER $\Gamma(\alpha)$ DISTRIBUTIONS

§2.0 Introduction and Notation

For $0 < a < b < 2a < \infty$ and $\alpha > 2$, let $\mathcal{P} = \{P_\theta | \theta \in [a,b]\}$ be the family of distributions with P_θ representing the $\Gamma(\alpha)$ distribution with scale parameter θ. Let s be a positive integer.

Let $\{X_n\}$ be a sequence of independent random variables with X_n distributed as P_{θ_n} belonging to $\mathcal{P}$. Let $\underline{X}_n = (X_1,\ldots,X_n)$, $\underline{\theta} = \{\theta_n\}$ and G_n be the empiric distribution of $\theta_1,\ldots,\theta_n$.

In section §2.1, we consider a sequence of estimation problems each having the structure of the following component estimation problem. Based on an observable random variable X whose distribution P_θ belongs to $\mathcal{P}$, the problem is to estimate θ with squared-error loss. Let $R(G_n)$ denote the Bayes risk against G_n in the estimation problem just described. Let $\underline{\phi} = \{\phi_n\}$ be a randomized sequence-compound procedure (abbreviated to randomized procedure hereafter). That is, for each n, ϕ_n is a randomized function of $\underline{X}_n$. For any such $\underline{\phi}$, $\underline{\theta}$ in $\times_n [a,b]$, let

$$D_n(\underline{\theta},\underline{\phi}) = n^{-1} \sum_{j=1}^{n} E|\phi_j - \theta_j|^2 - R(G_n) \tag{0.1}$$

where E stands for expectation wrt the joint distribution of all the random variables involved. In section §2.1, we exhibit a randomized procedure $\underline{\psi}^* = \{\psi_n^*\}$ such that $D_n(\underline{\theta},\underline{\psi}^{**}) = O(n^{-s/2(s+1)})$ uniformly in all parameter sequences $\underline{\theta}$ in $\times_n [a,b]$.

In section §2.2, we consider a sequence of two-action problems each having the structure of the following component two-action problem. Based on an observable random variable X whose distribution P_θ belongs to $\mathcal{P}$, the problem is to choose one of two possible actions a_1 and a_2 when the loss functions corresponding to a_1 and a_2 are $L(a_1,\theta) = (\theta-c)^+$ and $L(a_2,\theta) = (\theta-c)^-$ for some c in (a,b). Let $R(G_n)$ denote the Bayes risk against G_n in the two-action problem described above. Then, in section §2.2, we exhibit a randomized procedure $\underline{\hat{\psi}} = \{\hat{\psi}_n\}$ such that the absolute value of $D_n(\underline{\theta},\hat{\psi})$ defined by

$$D_n(\underline{\theta},\underline{\hat{\psi}}) = n^{-1} \sum_{j=1}^{n} EL(\hat{\psi}_j,\theta_j) - R(G_n) \tag{0.2}$$

is $O(n^{-s/2(s+1)})$ uniformly in all parameter sequences $\underline{\theta}$ in $\times_n [a,b]$.

The orders stated in the results of both sections §2.1 and §2.2 are uniform in all parameter sequences $\underline{\theta}$ in $\times_n [a,b]$. Hence, in order to reduce the complexity of the statements of the results in this chapter, the range of the parameter sequences will not be exhibited, but is understood to be $\times_n [a,b]$.

We introduce some notation which is common to both sections §2.1 and §2.2. Let $\{\lambda_n\}$ be a sequence of i.i.d. random variables with the density of λ_1 as $(\alpha-1)\lambda_1^{\alpha-2}[0 < \lambda_1 < 1]$ wrt Lebesgue measure μ on $((0,\infty), \mathcal{B} \cap (0,\infty))$. Furthermore, we assume that $\{\lambda_n\}$ is independent of $\{X_n\}$. Define, for each n, $Y_n = \lambda_n X_n$. Then, Y_n has $\Gamma(\alpha-1)$ distribution with scale parameter θ_n. We let Σ and Σ' denote summations over j from 1 to n-1 and from 1 to s respectively.

Now we introduce some notation which is similar to that introduced in section §1.4. Since the Vandermonde determinant involved does not vanish, there exists a unique vector $d = (d_1,\ldots,d_s)$ in R^s such that $d_s \neq 0$ and

$$\Sigma' d_i(i^{\ell} - (i-1)^{\ell}) = \begin{matrix} 1 & \text{for } \ell = 1 \\ 0 & \text{for } \ell = 2,\ldots,s. \end{matrix} \tag{0.3}$$

For any $h > 0$ and any real valued function g on $(0,\infty)$, define

$$\Delta g(u) = h^{-1}(g(u+h) - g(u)) \qquad \text{for} \quad u > 0.$$

With $\bar{F}$ and $\bar{H}$ denoting the averages of the distributions of $X_1,\ldots,X_{n-1}$ and $Y_1,\ldots,Y_{n-1}$ respectively and with X abbreviating X_n, let

$$\bar{\eta} = \Sigma' d_i \Delta \bar{F}(X + (i-1)h) \tag{0.4}$$

and

$$\bar{\xi} = \Sigma' d_i \Delta \bar{H}(X + (i-1)h). \tag{0.5}$$

With F^* and H^* denoting the empiric distributions of $X_1,\ldots,X_{n-1}$ and $Y_1,\ldots,Y_{n-1}$, let

$$\eta^* = \Sigma' d_i \Delta F^*(X + (i-1)h) \tag{0.6}$$

and

$$\xi^* = \Sigma' d_i \Delta H^*(X + (i-1)h). \tag{0.7}$$

Let p_θ and q_θ denote the densities of $\Gamma(\alpha)$ distribution with scale parameter θ and $\Gamma(\alpha-1)$ distribution with scale parameter θ respectively. Let $\bar{p}$ and $\bar{q}$ denote the densities of $\bar{F}$ and $\bar{H}$ respectively. With $p_\theta^{(s)}$ and $q_\theta^{(s)}$ denoting the sth order derivatives of p_θ and q_θ respectively, we assume throughout this chapter that α is $\ni$

$$(0.8)\qquad \sup\{|p_\theta^{(s)}|, |q_\theta^{(s)}| \mid a \le \theta \le b\} < \infty.$$

Under this assumption (0.8), it follows from the condition on d in (0.3) and (3.2) of Chapter I that

$$(0.9)\qquad |\bar{\eta} - \bar{p}| \le k_6 h^s$$

and

$$(0.10)\qquad |\bar{\xi} - \bar{q}| \le k_7 h^s$$

where k_6 and k_7 are constants.

§2.1 Estimation Problem. Rates of Convergence for $D_n(\underline{\theta},\underline{\psi}^*)$ with ψ^* Based on Kernel Estimators for a Density

In this section, under certain conditions on Θ, we show that, for each positive integer s, the modified regret of the procedure $\underline{\psi}^*$ (to be defined below by (1.2)) is $O(n^{-s/2(s+1)})$ when the component problem involved is the estimation problem described in section §2.0. The method of proving this rate of convergence is similar to that of Theorem 3 of Chapter I.

Let ψ denote the Bayes estimate against G_{n-1} in the component estimation problem described in section §2.0. Then ψ can be expressed as

$$\psi = \frac{u}{\alpha-1}\,\frac{\bar{q}(u)}{\bar{p}(u)} \quad \text{for} \quad u > 0. \tag{1.1}$$

Define the procedure $\underline{\psi}^*$ as follows. Let

$$\psi^* = \mathrm{tr}\left(\frac{X}{\alpha-1}\,\frac{\xi^*}{\eta^*}\right) \tag{1.2}$$

where tr stands for retraction to $[a,b]$. Any undefined ratios are taken to be zero.

Let $K_1, K_2, \ldots$ denote constants in this section. Let E stand for the expectation wrt the joint distribution of the random variables involved unless otherwise specified. In the following lemma, $\bar{q}$, $\bar{p}$ are evaluated at X.

Lemma 1. If $\alpha > 2$, $b < 2a$, (0.8) is satisfied and h is in $\mathcal{H} = \{h \,|\, 0 < h((\alpha-1)^{\alpha-1} \vee (\alpha-2)^{\alpha-2}) < e^{\alpha-2}\Gamma(\alpha-1)a\,r\,s^{-1}\}$ for some r in $(0,\frac{1}{2})$, then

$$E[|\psi^* - \psi(X)|] \le K_1((nh)^{-\frac{1}{2}} + (nh^{2s+1})^{\frac{1}{2}}).$$

Proof. We have by the definition of conditional expectation

(1.3) $$E[|\psi^* - \psi(X)|] = E[E[|\psi^* - \psi(X)|X]]$$

where $E[\cdot|X]$ stands for the conditional expectation operation given X.

Since ψ^*, by definition (1.2), is the retraction of $X\xi^*/(\alpha-1)\eta^*$ to $[a,b]$ and since $\psi(X) = X\bar{q}/(\alpha-1)\bar{p}$, the Bayes estimate against G_{n-1} whose support lies in $[a,b]$, is in $[a,b]$, we have $|\psi^* - \psi| \le b-a$ and $|\psi^* - \psi| \le X(\alpha-1)^{-1}|D|$ where

(1.4) $$D = \frac{\xi^*}{\eta^*} - \frac{\bar{q}}{\bar{p}}.$$

We then have

(1.5) $$\begin{aligned} E[|\psi^* - \psi(X)||X] &\le \int_0^{b-a} P[|\psi^* - \psi| > u]du \\ &\le \int_0^{b-a} P[|D| > (\alpha-1)X^{-1}u]du \\ &= \int_0^{b-a} P[D > (\alpha-1)X^{-1}u] + \int_{a-b}^{0} P[D < (\alpha-1)X^{-1}u]du \end{aligned}$$

where P stands for the joint probability measure of $X_1,\ldots,X_{n-1}$ and $Y_1,\ldots,Y_{n-1}$.

The main part of the proof bounds $P[D > (\alpha-1)X^{-1}u]$ for $0 \le u \le b-a$ and $P[D < (\alpha-1)X^{-1}u]$ for $a-b \le u < 0$ by using the Berry-Esseen theorem. The rest of the proof shows that the expectation of a bound for the rhs of (1.5) is exceeded by the bound in the lemma.

Let $X > 0$ be fixed until otherwise stated. Let

$$(1.6)\quad Z_j(u) = \Sigma' d_i([X + (i-1)h < Y_j < X + ih] - ((\alpha-1)X^{-1}u + \frac{\bar{q}}{\bar{p}})[X + (i-1)h < X_j < X + ih]) \quad \text{for } |u| \le b-a.$$

Let the dependency of Z_j on u be suppressed. Let $\beta^2 = \mathrm{Var}(\Sigma Z_j)$ and $L = \beta^{-3}\Sigma E|Z_j - EZ_j|^3$. Now we prove the following sublemma.

Sublemma. For any $|u| \le b-a$ and any constant k_2 such that $|d_i| \le k_2$ for $i = 1,\dots,s$,

$$L \le \frac{k_2(1 + 2(\alpha-1)X^{-1}b)}{k_4((n-1)h)^{\frac{1}{2}}(\phi\, \bar{H}(X+(s-1)h))^{\frac{1}{2}}}\, .$$

Proof. In order to obtain the result of the sublemma, we need a lower bound for β^2. This bound will be obtained by applying L1.A (see Appendix) to the Z_j.

Since $Y_j = \lambda_j X_j$ and since the distribution of λ_j is supported on $(0,1)$, $P[Y_j \ge X_j] = 0$ and hence Z_j defined by (1.6) takes $2^{-1}s(s+5) + 1$ values; namely,

$$(1.7)\quad \begin{array}{l} 0,\ d_i - d_j((\alpha-1)X^{-1}u + \frac{\bar{q}}{\bar{p}}) \quad \text{for } 1 \le i \le j \le s, \\ d_i \quad \text{for } i = 1,\dots,s \text{ and } -d_i((\alpha-1)X^{-1}u + \frac{\bar{q}}{\bar{p}}) \quad \text{for } i = 1,\dots,s. \end{array}$$

with nonzero probability.

The probability that Z_j takes the value zero in (1.7) is given by $P[X_j \notin (X,X + sh),\ Y_j \notin (X,X + sh)]$. Since

$e^{-u}u^m \le e^{-m}m^m$ for $m > 0$ and since $\alpha > 2$ and $\theta_j > 2$ by assumption, we have

$$P[X_j \in (X, X + sh)] = \frac{1}{\Gamma(\alpha)\theta_j} \int_X^{X+sh} (\frac{u}{\theta_j})^{\alpha-1} e^{-\frac{u}{\theta_j}} du$$

$$\le \frac{e^{-(\alpha-2)}}{\Gamma(\alpha-1)a} ((\alpha-1)^{\alpha-1} \vee (\alpha-2)^{\alpha-2})sh,$$

(1.8)

$$P[Y_j \in (X, X + sh)] = \frac{1}{\Gamma(\alpha-1)\theta_j} \int_X^{X+sh} (\frac{u}{\theta_j})^{\alpha-2} e^{-\frac{u}{\theta_j}} du$$

$$\le \frac{e^{-(\alpha-2)}}{\Gamma(\alpha-1)a} ((\alpha-1)^{\alpha-1} \vee (\alpha-2)^{\alpha-2})sh.$$

Therefore, it follows by the hypothesis on h that $P[X_j \in (X, X + sh)]$ and $P[Y_j \in (X, X + sh)]$ are exceeded by r. Hence, since $P(A \cap B) \ge P(A) + P(B) - 1$ for any two events A, B, we obtain that $P[X_j \notin (X, X + sh), Y_j \notin (X, X + sh)] > 1-2r$. Hence, since Z_j takes the value d_s with probability $P[X + (s-1)h < Y_j < X + sh \le X_j]$, we obtain by L1.A. that

$$\text{(1.9)} \qquad Var(Z_j) \ge k_3^2(P[X + (s-1)h < Y_j < X + sh \le X_j])$$

where $k_3^2 = d_s^2(1-2r) \inf\{1 - P[u + (s-1)h < Y_j < u + sh] | u > 0, h \in \mathscr{H}\}$. We observe that $k_3 \ne 0$ since $d_s \ne 0$ and $2r < 1$. Hence, since $[X + (s-1)h < Y_j < X + sh, X_j - Y_j \ge h] \subseteq [X + (s-1)h < Y_j < X + sh \le X_j]$ and since $X_j - Y_j$ and Y_j are independent, we obtain that

$$Var(Y_j) \ge k_3^2 \inf\{P[X_j - Y_j \ge h] | h \in \mathscr{H}\} P[X + (s-1)h < Y_j < X + sh].$$

Therefore,

$$\text{(1.10)} \qquad \beta^2 \ge k_4^2(n-1)h \, \phi \, \bar{H}(X + (s-1)h)$$

where $\qquad k_4^2 = k_3^2 \inf\{P[X_j - Y_j \ge h] | h \in \mathscr{H}\}.$

Since $|u| \le b-a$ and since $X\bar{q}/(\alpha-1)\bar{p} \le b$, the maximum of the moduli of the values of $(-Z_j)$ in (1.7) is at most

$$(1.11) \qquad k_2(1 + 2(\alpha-1)X^{-1}b)$$

where k_2 is the constant stated in the sublemma.

Therefore, the standardized range bound for L, together with the help of (1.10) and (1.11), gives the result of the sublemma.

Proceeding with the proof of the lemma, we obtain an upper bound for β^2. The definition of Z_j in (1.6) and (1.11) imply that

$$E\, Z_j^2 \le k_2^2\, (1 + 2(\alpha-1)X^{-1}b)^2 (F_j]_X^{X+sh} + H_j]_X^{X+sh})$$

where F_j and H_j are the distributions of X_j and Y_j respectively. Therefore, since $\beta^2 = \Sigma\, Var(Z_j) \le \Sigma\, EZ_j^2$, we obtain that

$$(1.12) \qquad \beta^2 \le k_2^2(1 + 2(\alpha-1)X^{-1}b)^2(n-1)(\bar{F}]_X^{X+sh} + \bar{H}]_X^{X+sh}).$$

Let $0 \le u \le b-a$. Then the definitions of D in (1.4) and Z_j in (1.6) imply that $[D > (\alpha-1)X^{-1}u] \le [\Sigma Z_j > 0] + [\eta^* \le 0]$. Hence, with $b(L)$ denoting the bound in the sublemma and B denoting the Berry-Esseen constant, by the Berry-Esseen theorem, the sublemma and the triangle inequality, we obtain that $P[\Sigma\, Z_j > 0]$ is exceeded by

$$(1.13) \quad \Phi(-\beta^{-1}(n-1)h(\alpha-1)X^{-1}\bar{p}\, u) + |\Phi(-\beta^{-1}(n-1)h(\alpha-1)X^{-1}\bar{p}\, u) - \Phi(\beta^{-1}\Sigma\, EZ_j)| + B\, b(L).$$

By using the upper bound for β^2 in (1.12), we obtain that

$$(1.14) \qquad \beta^{-1}(n-1)h(\alpha-1)X^{-1}\bar{p} \geq ((n-1)h)^{\frac{1}{2}}f$$

where f is the positive solution of the equation

$$(1.15) \qquad k_2^2X^2(1 + 2(\alpha-1)X^{-1}b)^2(\bar{F}]_X^{X+sh} + \bar{H}]_X^{X+sh})f^2 = h(\alpha-1)^2\bar{p}^2.$$

Since $\Sigma\, EZ_j + (n-1)h(\alpha-1)X^{-1}\bar{p}\, u = (n-1)h((\alpha-1)X^{-1}u + \frac{\bar{q}}{\bar{p}})(\bar{p} - \bar{\eta}) + \bar{\xi} - \bar{q})$ for all $|u| \leq b-a$ and $X\bar{q}/(\alpha-1)\bar{p} \leq b$, it follows from (0.9) and (0.10) that

$$(1.16) \qquad |\Sigma\, EZ_j + (n-1)h(\alpha-1)X^{-1}\bar{p}\, u| \leq (n-1)h^{s+1}(2K_6b(\alpha-1)X^{-1} + K_7) \quad \text{for all } |u| \leq b-a.$$

Therefore, it follows by the mean value theorem and the lower bound for β^2 in (1.10) that the second term in (1.13) is exceeded by

$$(1.17) \qquad \frac{((n-1)h^{2s+1})^{\frac{1}{2}}(2k_6(\alpha-1)X^{-1} + K_7)}{k_4(\phi\, \bar{H}(X + (s-1)h))^{\frac{1}{2}}}.$$

Hence it follows from (1.13) and (1.14) that

$$(1.18) \qquad P[\Sigma\, Z_j > 0] \leq \Phi(-((n-1)h)^{\frac{1}{2}}f\, u) + (1.17) + b\, b(L) \quad \text{for } 0 \leq u \leq b-a.$$

Let $a-b \leq u < 0$. Then the definitions of D in (1.4) and Z_j in (1.6) imply that $[D < (\alpha-1)X^{-1}u] \leq [\Sigma(-Z_j) > 0] + [\eta^* \leq 0]$. Hence, since the sublemma continues to hold if d is replaced by $-d$, we have by the Berry-Esseen theorem, the triangle inequality and the sublemma that $P[\Sigma(-Z_j) > 0]$ is exceeded by

$$(1.19)\quad \Phi(\beta^{-1}(n-1)h(\alpha-1)X^{-1}\bar{p}\ u) + |\Phi(\beta^{-1}(n-1)h(\alpha-1)X^{-1}u) - \Phi(-\beta^{-1}\Sigma\ EZ_j)| + B\ b(L).$$

By using the upper bound for β^2 in (1.12), we obtain that the first term of (1.19) is bounded by $\Phi(((n-1)h)^{\frac{1}{2}} f\ u)$ where f is the positive solution of (1.15). By using (1.16) and the lower bound for β^2 in (1.10), we obtain that the second term of (1.19) is exceeded by (1.17). Hence, it follows from (1.19) that

$$P[D < (\alpha-1)X^{-1}u] \le \Phi(((n-1)h)^{\frac{1}{2}} f\ u) + (1.17) + B\ b(L)$$
$$\text{for}\quad a-b \le u < 0.$$

Integrating this inequality wrt u over $[a-b,0)$ and the inequality (1.18) wrt u over $[0,b-a]$, then bounding their first terms by using the inequality $\int_0^{b-a} \Phi(-Au)du \le (2\pi)^{-\frac{1}{2}} A^{-1}$ for $A > 0$, we obtain from (1.5) that

$$E[|\psi^* - \psi(X)||X] \le \frac{2}{((n-1)h)^{\frac{1}{2}} f} + 2(b-a)\ (1.17) + 2(b-a)B\ b(L).$$

In view of this inequality, (1.17) and the bound in the sublemma, we continue the proof of the lemma by showing that f^{-1} and $(1 + X^{-1})(\blacktriangle\ \bar{H}(X + (s-1)h))^{-\frac{1}{2}}$ are uniformly bounded and P_n-integrable.

Since $\bar{F}]_X^{X+sh} / \bar{H}]_X^{X+sh} = \bar{p}(X + \epsilon sh)/\bar{q}(X + \epsilon sh)$ for some ϵ in (0,1) by Cauchy's mean value theorem, $X\bar{q}/(\alpha-1)\bar{p} \ge a$, and since f is defined as the positive solution of (1.15), we obtain that f^{-1} is exceeded by

$$(1.20) \qquad \frac{k_2\, X\,(1 + 2(\alpha-1)X^{-1}b)\,(\bar{H}]_X^{X+sh})^{\frac{1}{2}}((1+((\alpha-1)a)^{-1}(X+\epsilon sh))^{\frac{1}{2}}}{h^{\frac{1}{2}}\bar{p}}$$

Since $\bar{H}]_X^{X+sh} = sh\,\bar{q}(X + \delta sh)$ for some $0 < \delta < 1, (\alpha-1)^{-1}$ $\bar{q}(X + \epsilon sh)/\bar{p}(X) \le b(X + s)^{\alpha-2}/X^{\alpha-1}$ and $b^{-\alpha}\, e^{-X/a} \le \Gamma(\alpha)X^{1-\alpha}p_j \le$ $a^{-\alpha}\, e^{-X/b}$, the condition $b < 2a$ implies that the expectation of the upper bound (1.20) for f^{-1} is uniformly bounded.

Since, by the mean value theorem, $\blacktriangle\, \bar{H}(X + (s-1)h) =$ $\bar{q}(X + \epsilon(s-1)h)$ for some ϵ in $(0,1)$ and since $\Gamma(\alpha-1)b^{\alpha-1}\bar{q} \ge X^{\alpha-2}e^{-X/a}$ and $\Gamma(\alpha)p_n \le X^{\alpha-1}a^{-\alpha}e^{-X/b}$, the conditions $b < 2a$ and $\alpha > 2$ imply that the expectation of $X^{-1}(\blacktriangle\, \bar{H}(X + (s-1)h))^{\frac{1}{2}}$ is uniformly bounded.

The same method of bounding $\underline{P}_n[\eta^* \le 0]$ completes the proof.

Now we state and prove the main result of the section. This result is a consequence of Lemma 1, Theorem 2.1 and (2.5) of Gilliland (1968).

<u>Theorem 1</u>. If $\alpha > 2$, $b < 2a$, (0.8) is satisfied, $h = \gamma n^{-1/s+1}$ with $0 < \gamma((\alpha-1)^{\alpha-1} \vee (\alpha-2)^{\alpha-2}) < e^{\alpha-2}\,\Gamma(\alpha-1)a\; r\; s^{-1}$ for some r in $(0,\frac{1}{2})$ and $\underline{\psi}^*$ is defined by (1.2), then

$$D_n(\underline{\theta},\underline{\psi}^*) = O(n^{-s(2(s+1)}).$$

<u>Proof</u>. Since $p_\theta(u)$ is exceeded by $(\Gamma(\alpha))^{-1}a^{-\alpha}u^{\alpha-1}e^{-u/b}$ uniformly in all θ belonging to $[a,b]$ and $\mu[u^{\alpha-1}e^{-u/b}] < \infty$, it follows by Theorem 2.1 of Gilliland (1968) that

$$(1.\ \ ) \qquad n^{-1}\sum_{j=1}^{n} E[|\psi_j(X_j) - \psi_{j-1}(X_j)|] = O(n^{-1}\log n)$$

where $O(n^{-1}\log n)$ is uniform in all parameter sequences $\underline{\theta}$ in $\underset{n}{X}[a,b]$.

Since the inequality (2.5) of Gilliland (1968) continues to hold when the ϕ_i mentioned there are randomized procedures and since $\underline{\psi}^*$, by definition (1.2), takes values in $[a,b]$, it follows by (1.19) that

$$|D_n(\underline{\theta},\underline{\psi}^*)| \le 4b\, n^{-1} \sum_{j=1}^{n} E[|\psi_j^* - \psi_{j-1}(X_j)|] + O(n^{-1}\log n).$$

Hence the result of the theorem follows from Lemma 1 and the definition of h in the statement of the theorem.

§2.2 Two-action Problem. Rates of Convergence for $D_n(\underline{\theta},\underline{\Psi})$ with Ψ Based on Kernel Estimators for a Density

In this section, under certain conditions on $\mathcal{P}$, we show that, for each positive integer s, the modified regret of the procedure $\underline{\Psi}$ (defined below by (2.5) and (2.6)) is $O(n^{-s/2(s+1)})$ when the component problem involved is the two action problem described in section §2.0. The method of proving this rate of convergence is similar to that of Johns (1967).

In this section, we make it a convention that the value of any decision function is the probability of taking action a_1. Define, for each n,

$$\gamma_n = (\theta_n - c)p_n. \tag{2.1}$$

If $R(G_n)$ denotes the component two-action problem described in section §2.0, then

$$R(G_n) = \inf_\delta \mu[\delta\, n^{-1}\sum_{j=1}^n \gamma_j] + n^{-1}\sum_{j=1}^n (\theta_j - c)^-.$$

Hence, with m_n defined by

$$m_n = \sum_{j=1}^n \gamma_j, \tag{2.2}$$

$$n\,R(G_n) = -\mu[m_n^-] + \sum_{j=1}^n (\theta_j - c)^- . \tag{2.3}$$

With E denoting the expectation operation, for any randomized procedure $\underline{\Psi} = \{\Psi_n\}$, the risk of using Ψ_n to decide about θ_n is given by $(\theta_n - c)E[\Psi_n] + (\theta_n - c)^-$ and hence the average risk of using $\Psi_1,\ldots,\Psi_n$ to decide about $\theta_1,\ldots,\theta_n$ respectively is given by

$$\frac{1}{n}\sum_{j=1}^{n}\mu[\gamma_j(u)E[\psi_j|X_j = u]] + \frac{1}{n}\sum_{j=1}^{n}(\theta_j - c)^-$$

where $E[\psi_j|X_j = u]$ is a conditional expectation of ψ_j given $X_j = u$. Hence, it follows from (0.2) and (2.3) that

(2.4) $$n\, D_n(\underline{\theta},\underline{\psi}) = \mu[\gamma_1\psi_1] + \mu[\sum_{j=2}^{n}\gamma_j(u)E[\psi_j|X_j = u] + m_n^-(u)].$$

Let $h > 0$ be a function of n. We define $\underline{\psi} = \{\psi_n\}$ as follows. Let

(2.5) $$\psi_1 = 1$$

and, for $n > 1$,

(2.6) $$\psi_n = [X\xi^*/\alpha\text{-}1 < c\eta^*]$$

where η^* and ξ^* are defined by (0.6) and (0.7) respectively and X is an abbreviation for X_n.

Let, for $n > 1$,

(2.7) $$S_{n-1} = (n-1)\left(\frac{u\xi^*}{\alpha-1} - c\eta^*\right) \qquad \text{for} \quad u > 0$$

where ξ^* and η^* are evaluated at $Y_1,\ldots,Y_{n-1}, u$ and $X_1,\ldots,X_{n-1}, u$ respectively.

(2.8) $$m_{n-1}^* = E[S_{n-1}] \qquad \text{for} \quad u > 0$$

and

(2.9) $$\beta_{n-1}^2 = \mathrm{Var}(S_{n-1}) \qquad \text{for} \quad u > 0.$$

Lemma 2. If $\alpha > 2$, $b < 2a$, (0.8) is satisfied and h is in $\mathcal{K} = \{h|0 < h((\alpha-1)^{\alpha-1} \vee (\alpha-2)^{\alpha-2}) < e^{\alpha-2}\,\Gamma(\alpha-1)a\,r\,s^{-1}\}$ for some $0 < r < \frac{1}{2}$, then

$$|\mu[\gamma_n(\Phi(-\frac{m^*_{n-1}}{\beta_{n-1}}) - \Phi(-\frac{m_{n-1}}{\beta_{n-1}}))]| \le K_1((n-1)h^{2s+1})^{\frac{1}{2}} \quad \text{for} \quad n > 1.$$

Proof. By the mean value theorem and the inequality $\sqrt{2\pi}\,\phi \le 1$, we obtain that

$$(2.10) \qquad |\Phi(-\frac{m^*_{n-1}}{\beta_{n-1}}) - \Phi(-\frac{m_{n-1}}{\beta_{n-1}})| \le \frac{|m^*_{n-1} - m_{n-1}|}{\sqrt{2\pi}\,\beta_{n-1}}.$$

Since (0.8) is satisfied by the hypothesis, it follows from (0.9), (0.10) and the definitions of m^*_{n-1} in (2.8) and m_{n-1} in (2.2) that

$$(2.11) \qquad |m^*_{n-1} - m_{n-1}| \le (n-1)h^s(\frac{u}{\alpha-1}\, k_6 + c\, k_7).$$

Now we get a lower bound for β^2_{n-1}. Let

$$(2.12) \quad h\, Z_j(u) = \Sigma' d_i(u(\alpha-1)^{-1}[u + (i-1)h < Y_j < u + ih]$$
$$- c[u + (i-1)h < X_j < u + ih]).$$

Then, since $P[Y_j \ge X_j] = 0$, $h\, Z_j$ defined above takes $2^{-1}s(s+5) + 1$ values; namely,

$$(2.13) \quad 0,\ d_i u(\alpha-1)^{-1} - d_j c \quad \text{for} \quad 1 \le i \le j \le s,$$
$$d_i u(\alpha-1)^{-1} \quad \text{for} \quad i = 1,\dots,s \quad \text{and} \quad d_i c \quad \text{for} \quad i = 1,\dots,s.$$

with nonzero probability. The probability that $h\, Z_j$ takes the value zero in (2.13) is given by $P[X_j \notin (u, u + sh),\ Y_j \notin (u, u + sh)]$. Then, it follows by the hypothesis on h, (1.8) and the inequality $P(A \cap B) \ge P(A) + P(B) - 1$ for any two events A and B that this probability is at least $1-2r > 0$. Hence, since $h\,Z_j$ takes the value d_s with probability $P[u + (s-1)h < Y_j < u + sh \le X_j]$, we have from L1.A. (see Appendix) that

$$(2.14)\qquad \operatorname{Var}(h\,Z_j) \ge (1-2r)d_s^2\, P[u + (s-1)h < Y_j < u + sh \le X_j]\,(1 - P[u + (s-1)h < Y_j < u + sh \le X_j]).$$

Hence, since $\inf\{1 - P[u + (s-1)h < Y_j < u + sh \le X_j] \mid u > 0,\ \theta_j \in [a,b],\ h \in \mathcal{H}\} > 0$, by using the argument given to obtain (1.10) from (1.9), we obtain that

$$(2.15)\qquad h\,\beta_{n-1}^2 \ge k_2^2(n-1)\, \spadesuit\, \bar{H}(u + (s-1)h).$$

Therefore, we have from (2.10) and (2.11) that

$$(2.16)\qquad \left|\Phi\left(-\frac{m^*_{n-1}}{\beta_{n-1}}\right) - \Phi\left(-\frac{m_{n-1}}{\beta_{n-1}}\right)\right| \le \frac{((n-1)h^{2s+1})^{\frac{1}{2}}\left(k_6\frac{u}{\alpha-1} + k_7 c\right)}{k_2\,(\spadesuit\, \bar{H}(u + (s-1)h))^{\frac{1}{2}}}.$$

We have

$$(2.17)\qquad a^{\alpha}\,\Gamma(\alpha)\,|\gamma_n| \le (b + c)\, u^{\alpha-1} e^{-u/b}.$$

By the mean value theorem,

$$(2.18)\qquad \spadesuit\, \bar{H}(u + (s-1)h) = \bar{q}(u + \epsilon h) \quad \text{for some } \epsilon \text{ in } (s-1,s).$$

Hence, since

$$(2.19)\qquad b^{\alpha-1}\Gamma(\alpha-1)\bar{q}(u) \ge u^{\alpha-2} e^{-u/a},$$

it follows by the hypothesis on h,

$$(2.20)\qquad b^{\alpha-1}\Gamma(\alpha)\, \spadesuit\, \bar{H}(u + (s-1)h) \ge u^{\alpha-2} e^{-1-u/a}.$$

Since $b < 2a$ the result follows from the inequalities (2.16), (2.17) and (2.20).

Let

$$L_{n-1} = \beta_{n-1}^{-3} \Sigma E|Z_j - EZ_j|^3 \tag{2.21}$$

where Z_j is defined by (2.12).

Lemma 3. If the hypothesis of Lemma 2 is satisfied, then

$$\mu[|\gamma_n|L_{n-1}] \le K_3((n-1)h)^{-\frac{1}{2}} \quad \text{for} \quad n > 1.$$

Proof. The standardized range bound for L_{n-1}, together with (2.15) and the fact that the maximum of the moduli of the values of $h\,Z_j$ defined in (2.12) is at most

$$\max\{|d_1|,\ldots,|d_s|\}(u(\alpha-1)^{-1} + c), \tag{2.22}$$

implies that

$$L_{n-1} \le \frac{\max\{|d_1|,\ldots,|d_s|\}(u(\alpha-1)^{-1} + c)}{k_2((n-1)h)^{\frac{1}{2}}(\phi\,\bar{H}(u + (s-1)h))^{\frac{1}{2}}}.$$

Since $b < 2a$ and $\alpha > 2$ implies that the μ-integral of the rhs of the inequality obtained by weakening this inequality for L_{n-1} by using (2.20) is uniformly bounded, the proof of the lemma is complete.

Below, we get an upper bound for β_{n-1}^2. We have by the definition of $h\,Z_j$ in (2.12) and (2.22) that

$$h^2\,EZ_j^2 \le (\max\{|d_1|,\ldots,|d_s|\})^2(u(\alpha-1)^{-1} + c)^2(F_j]_u^{u+sh} + H_j]_u^{u+sh})$$

where F_j and H_j are the distribution functions of X_j and

Y_j respectively. Therefore, since $\beta_{n-1}^2 \le \Sigma\, EZ_j^2$, we have

$$(2.23)\quad h^2\beta_{n-1}^2 \le (\max\{|d_1|,\ldots,|d_s|\})^2(u(\alpha-1)^{-1}+c)^2(n-1)(\bar{F}]_u^{u+sh} + \bar{H}]_u^{u+sh}).$$

Lemma 4. $\mu[\beta_{n-1}] \le k_4((n-1)h^{-1})^{\frac{1}{2}}$ for $n > 1$.

Proof. By (2.23) and the inequality $(a+b)^{\frac{1}{2}} \le a^{\frac{1}{2}} + b^{\frac{1}{2}}$ for $a,b > 0$, we have

$$((n-1)h^{-1})^{-\frac{1}{2}}\beta_{n-1} \le \max\{|d_1|,\ldots,|d_s|\}((h^{-1}\bar{F}]_u^{u+sh})^{\frac{1}{2}} + (h^{-1}\bar{H}]_u^{u+sh})^{\frac{1}{2}}).$$

By the mean value theorem $h^{-1}\bar{F}]_u^{u+sh} = s\,\bar{p}(X+\epsilon sh)$ and $h^{-1}\bar{H}]_u^{u+sh} = s\,\bar{q}(X+\delta sh)$ for some $0 < \epsilon,\delta < 1$. Hence, since $\Gamma(\alpha)a^{\alpha}\bar{p} \le u^{\alpha-1}e^{-u/b}$, $\Gamma(\alpha-1)a^{\alpha-1}\,\bar{q} \le u^{\alpha-2}e^{-u/b}$ and $u^{\alpha-2}e^{-u/b}$ is μ-integrable, the result of the lemma follows.

The proof of the following theorem depends on Lemmas 2,3,4 and part of the method of proof of Theorem 1 of Johns (1967).

Theorem 2. For each positive integer s, if (0.8) is satisfied, $\alpha > 2$, $b < 2a$, $h = \gamma\, n^{-1/s+1}$ where $\gamma((\alpha-1)^{\alpha-1} \vee (\alpha-2)^{\alpha-2}) < e^{\alpha-2}\,\Gamma(\alpha-1)a\,r\,s^{-1}$ for some $0 < r < \frac{1}{2}$ and $\underline{\hat{\psi}}$ is defined by (2.5) and (2.6), then

$$D_n(\underline{\theta},\underline{\hat{\psi}}) = O(n^{-s/2(s+1)}).$$

Proof. By (2.4) and the definition of $\underline{\hat{\psi}}$ in (2.5) and (2.6), we have

$$(2.24)\quad n|D_n(\underline{\theta},\underline{\hat{\psi}})| \le \mu[|\gamma_1|] + |\mu[\sum_{j=2}^{n}\gamma_j(u)E[\hat{\psi}_j|X_j = u] + m_n^-(u)]|.$$

To start with, we consider bounding the integrand of the second term of the rhs of (2.24) on the set $[m_n > 0]$. Afterwards we consider the case when $m_n \le 0$. So, let $m_n \ge 0$ until otherwise stated. Since $m_n \ge 0$, we have by the definition of Ψ_j in (2.6) and S_{j-1} in (2.7) for $j > 1$ that

$$\sum_{j=2}^{n} \gamma_j(u) E[\Psi_j | X_j = u] + m_n^-(u) = \sum_{j=2}^{n} \gamma_j(u) P[S_{j-1} < 0] \tag{2.25}$$

where P stands for the joint probability measure of all the random variables involved.

By the triangle inequality,

$$\Big|\sum_{j=2}^{n} \gamma_j(u) P[S_{j-1} < 0]\Big| \le |D_1| + |D_2| + |D_3| \tag{2.26}$$

where

$$D_1 = \sum_{j=2}^{n} \gamma_j \Big(P[S_{j-1} < 0] - \Phi\Big(-\frac{m_{j-1}^*}{\beta_{j-1}}\Big)\Big), \tag{2.27}$$

$$D_2 = \sum_{j=2}^{n} \gamma_j \Big(\Phi\Big(-\frac{m_{j-1}^*}{\beta_{j-1}}\Big) - \Phi\Big(-\frac{m_{j-1}}{s_{j-1}}\Big)\Big) \tag{2.28}$$

and

$$D_3 = \sum_{j=2}^{n} \gamma_j \, \Phi\Big(-\frac{m_{j-1}}{\beta_{j-1}}\Big). \tag{2.29}$$

With B denoting the Berry-Esseen constant, the Berry-Esseen theorem (Loève (1963), p. 288) gives

$$|D_1| \le B \sum_{j=2}^{n} |\gamma_j| L_{j-1}.$$

Therefore, by Lemma 3,

$$\mu[[m_n > 0]|D_1|] \le k_3 \sum_{j=2}^{n} ((j-1)h)^{-\frac{1}{2}}. \tag{2.30}$$

By Lemma 2, we have

$$\mu[[m_n > 0]|D_2|] \le k_1 \sum_{j=2}^{n} ((j-1)h^{2s+1})^{\frac{1}{2}}. \tag{2.31}$$

Replacing α_j and s_j by γ_j and β_j in (2.6) through (2.13) of Theorem 1 of Johns (1967), we obtain that

$$|D_3| \le \phi(0) \sum_{j=2}^{n} \frac{\gamma_j^2}{\beta_{j-1}} + \sum_{j=2}^{n} \frac{|\gamma_j|}{j^2} + (\beta_n + \beta_1)(A_1\Phi(A_1) + 2\phi(0)) \tag{2.32}$$

where $\max_{x>0} x\Phi(-x) = A_1\Phi(A_1)$.

The lower bound for β_{j-1}^2 in (2.15), the lower bound for $\phi\bar{H}(u + (s-1)h)$ in (2.20) and the upper bound for γ_j in (2.17), together with the conditions $b < 2a$ and $\alpha > 2$, imply that $\mu[\gamma_j^2/\beta_{j-1}] \le k_5((j-1)h^{-1})^{-\frac{1}{2}}$. Hence, since (2.17) implies that $\mu[|\gamma_j|]$ is uniformly bounded, it follows from (2.32) and Lemma 4 that

$$\mu[[m_n > 0]|D_3|] \le k_8\Big(\sum_{j=2}^{n} ((j-1)h^{-1})^{-\frac{1}{2}} + \sum_{j=2}^{n} \frac{1}{j^2} + (nh^{-1})^{\frac{1}{2}} + 1\Big).$$

Hence (2.25) to (2.32) imply that

$$\begin{aligned}
&\Big|\mu[[m_n > 0]\Big(\sum_{j=2}^{n} \gamma_j(u)E[\Phi|X_j = u] + m_n^-(u)\Big)]\Big| \\
&\qquad = \Big|\mu[[m_n > 0] \sum_{j=2}^{n} \gamma_j(u)P[S_{j-1} < 0]]\Big| \\
&\le k_9\Big(\sum_{j=2}^{n} ((j-1)h)^{-\frac{1}{2}} + \sum_{j=2}^{n} ((j-1)h^{2s+1})^{\frac{1}{2}} \\
&\qquad + \sum_{j=2}^{n} ((j-1)h^{-1})^{-\frac{1}{2}} + \sum_{j=2}^{n} \frac{1}{j^2} + \left(\frac{n}{h}\right)^{\frac{1}{2}} + 1\Big).
\end{aligned} \tag{2.33}$$

Now we consider bounding the integrand of the second term of the rhs of (2.24) on $[m_n < 0]$. For u in $[m_n \le 0]$, we have by the definition of Ψ_j in (2.6) and S_{j-1} in (2.7) for $j > 1$, we obtain that

$$\sum_{j=2}^{n} \gamma_j(u)E[\Psi_j|X_j = u] + m_n^-(u) = -\gamma_1(u) - \sum_{j=2}^{n} \gamma_j(u)P[S_{j-1} \geq 0].$$

Following the same argument we gave to bound $\Sigma_2^n \gamma_j(u)P[S_{j-1} < 0]$ by rhs of (2.33), we obtain that

$$|\mu[[m_n < 0](\sum_{j=2}^{n} \gamma_j(u)E[\Psi_j|X_j = u] + m_n^-(u))]| \leq \text{rhs of (2.33)} + \mu[|\gamma_1|].$$

Since (2.17) implies that $\mu[|\gamma_1|]$ is uniformly bounded, this inequality and (2.33), together with (2.24) and the hypothesis concerning h, imply the result of the theorem.

APPENDIX

APPENDIX

We apply the following lemma for obtaining lower bounds for certain variances in Lemmas 4 and 11 of Chapter I and Lemmas 1, 2 and 3 of Chapter II. The inequality in the following lemma is trivially true when $p_0 = 1$.

Lemma 1.A. Let $p_0 < 1, p_1, \ldots, p_i, \ldots$ be a probability distribution on $\{0,1,\ldots,i,\ldots\}$ and let Z be the r.v. $Z(i) = z_i$ for specified $z_0 = 0, z_1, \ldots, z_i, \ldots$ with $\Sigma z_i^2 p_i < \infty$. Let q_i abbreviate $1 - p_i$ and let $I(\lambda) = \Sigma_1^\infty p_i(1 - \lambda q_i)^{-1}$. Then $I \uparrow$ from q_0 to $\#\{i \geq 1 | p_i > 0\}$ as $\lambda \uparrow$ from 0 to 1 and

$$\mathrm{Var}(Z) \geq \lambda_1 \Sigma z_i^2 p_i q_i$$

with λ_1 the unique root of $I(\lambda) = 1$. Since $I(p_0) \leq 1$, $\lambda_1 \geq p_0$.

<u>Proof</u>. $I \uparrow$ since each summand $p_i(1 - \lambda q_i)^{-1}$ with $p_i > 0 \uparrow$. Since equality holds in the inequality when $\lambda_1 = 1$, we consider below the case $\lambda_1 < 1$.

To prove the inequality when $\lambda_1 < 1$, let $\psi(z) = \mathrm{Var}(Z) - \lambda_1 \Sigma z_i^2 p_i q_i$ for $z = (z_1, z_2, \ldots)$. Denoting the first and second partials wrt z_j by ψ_j and ψ_{jj} respectively,

$$\psi_j(z) = 2p_j\{(1-\lambda_1 q_j)z_j - \Sigma z_i p_i\} \quad , \quad \psi_{jj}(z) = 2(1-\lambda_1)p_j q_j .$$

For j with $p_j > 0$, ψ is, therefore, minimal wrt z_j variation iff $z_j = (1-\lambda_1 q_j)^{-1} \Sigma z_i p_i$. These conditions are satisfied iff, for some constant c, $z_j = c(1-\lambda_1 q_j)^{-1}$ for j with $p_j > 0$. For such z,

$$\psi(z) = c^2\{\Sigma\, p_i(1-\lambda_1 q_i)^{-2} - 1 - \lambda_1\, \Sigma\, p_i q_i(1-\lambda_1 q_i)^{-2}\} = 0$$

which yields the nonnegativity of ψ asserted by the lemma.

BIBLIOGRAPHY

BIBLIOGRAPHY

Clemmer, Bennie A. and Krtuchkoff, Richard G. (1968). The use of empirical Bayes estimators in a linear regression model. Biometrika 55, 525-534.

Fox, Richard (1968). Contributions to compound decision theory and empirical squared error loss estimation. RM-214, Department of Statistics and Probability, Michigan State University.

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

Gilliland, Dennis (1968). Sequential compound estimation. Ann. Math. Statist. 39, 1890-1904.

Graves, Lawrence M. (1956). The theory of functions of real variables, 2nd Edition. Macmillan.

Hannan, James F. (1957). Approximation to Bayes risk in repeated play. Contributions to the Theory of Games, 3, 97-139. Ann. Math. Studies No. 39, Princeton University Press.

Hannan, J. (1964). Mathematical Reviews 27, 828.

Johns, M.V., Jr. (1967). Two-action compound decision problems. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 463-478. University of California Press.

Johns, M.V., Jr. and Van Ryzin, J. (1967). Convergence rates for empirical Bayes two-action problems II. Continuous case. Technical Report No. 132, Department of Statistics, Stanford University.

Loève, Michel (1963). Probability Theory, 3rd Edition. Van Nostrand.

Martz, Harry F., Jr. and Krutchkoff, Richard G. (1969). Empirical Bayes estimators in a multiple linear regression model. Biometrika 56, 367-374.

Miyasawa, Koichi (1961). An empirical Bayes estimator of the mean of a normal distribution. Bull. Inst. Internat. Statist. 38, 181-188.

Pi-Erh, Lin (1968). Estimation of a multivariate density and its partial derivatives, with empirical Bayes applications. Ph.D. Thesis, Columbia University.

Rutherford, John R. (1965). Some parametric empirical Bayes techniques. Ph.D. Thesis, Virginia Polytechnic Institute.