RATES OF CONVERGENCE IN EMPIRICAL BAYES TWO - ACTION AND ESTIMATION PROBLEMS AND IN EXTENDED SEQUENCE - COMPOUND ESTIMATION PROBLEMS

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

BENITO ONG YU

1970

This is to certify that the

thesis entitled

RATES OF CONVERGENCE
IN EMPIRICAL BAYES TWO-ACTION AND ESTIMATION PROBLEMS
AND IN EXTENDED SEQUENCE-COMPOUND ESTIMATION PROBLEMS

presented by

Benito Ong Yu

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics and Probability

James Hannan
Major professor

Date 11/30/70

ABSTRACT

RATES OF CONVERGENCE
IN EMPIRICAL BAYES TWO-ACTION AND ESTIMATION PROBLEMS
AND IN EXTENDED SEQUENCE-COMPOUND ESTIMATION PROBLEMS

By

Benito Ong Yu

Throughout, our component problems concern exponential families of distributions of x conditional on the parameter θ.

In Part I we consider exponential families determined by a measure with Lebesgue density h, where $h(x) > 0$ if and only if $x > a$, and assume the parameter θ has a distribution G. Based on a sequence of observations $x_1, x_2, \ldots, x_n$, iid according to the marginal distribution of x, estimates of the posterior mean are used to define estimates for the Bayes test in the linear loss two-action problem. Rates of convergence of the excess risk are obtained under certain integrability conditions. The scale parameter exponential and the location parameter Normal densities are given as examples where the finiteness of certain moments of G is sufficient for these integrability conditions.

These results, proved under weaker hypotheses than those of Johns and Van Ryzin (1967), are obtained under the assumption $h^{(r)}$ exists for some $r \geq 2$. Analogous results are also obtained without any differentiability assumption on h.

In the squared error loss estimation problem, a truncation of the previous estimates for the posterior mean are used to estimate

θ. By a different method of proof, rates of convergence of the excess risk are established.

It is shown that the excess risk of the linear loss two-action problem is exceeded by the squared root of that of the estimation problem and, consequently, certain improved rates in the location parameter Normal two-action problem can be obtained as a corollary to those obtained in the estimation problem.

In Part II we consider certain discrete exponential and the location parameter Normal families, and assume that the parameter θ is bounded. Based on all past observations $x_1, x_2, \ldots, x_n$, with the x_i conditional on θ_i being independently distributed according to P_{θ_i}, squared error loss estimation of θ_n is considered with the aim that the average risk across the first n problems approach the extended Bayes envelope $R^k(G_n^k)$ evaluated at G_n^k, the empirical distribution function of the k-vectors $(\theta_1, \ldots, \theta_k)$, $(\theta_2, \ldots, \theta_{k+1}), \ldots, (\theta_{n-k+1}, \ldots, \theta_n)$.

Swain (1965) obtained rates of $O(n^{-\frac{1}{4}} \log^k n)$ and $o(1)$ for the discrete exponential and the Normal families, respectively. Gilliland (1966 and 1968) considered the unextended $(k = 1)$ versions of these problems and obtained improved rates of $O(n^{-\frac{1}{2}})$ and $O(n^{-1/5})$, respectively. In Chapters 3 and 4, the same order of improved rates, namely, $O(n^{-\frac{1}{2}})$ and $O(n^{-\frac{1}{k+4}})$, are obtained in these families, respectively.

RATES OF CONVERGENCE
IN EMPIRICAL BAYES TWO-ACTION AND ESTIMATION PROBLEMS
AND IN EXTENDED SEQUENCE-COMPOUND ESTIMATION PROBLEMS

By

Benito Ong Yu

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

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to Professor James F. Hannan for the patience he accorded me in the preparation of this thesis. His guidance and comments greatly improved and simplified the results in this work. To Professor Dennis C. Gilliland, I owe my thanks for his encouraging and helpful suggestions in reviewing the second part of this thesis.

This work was made possible through the financial support provided by the Department of Statistics and Probability and the National Science Foundation.

TABLE OF CONTENTS

PART I

EMPIRICAL BAYES IN EXPONENTIAL FAMILIES

INTRODUCTION

Johns and Van Ryzin (1967) studied the empirical Bayes two-action problem in the exponential family. They used kernel estimates for the marginal density f and its derivative g to define tests ϕ_n, and showed, in their Theorem 3, that under certain conditions, including (C) and (D) of Theorem 1.1, the risk $R_n(\phi_n, G)$ converges to the Bayes risk R^*. Furthermore, a rate was obtained. They gave the scale exponential and the Normal densities as examples where the existence of certain moments of the prior G is sufficient for the conditions (C) and (D).

Lin (1968) considered the multivariate estimation problem with squared error loss. A multivariate version of Theorem 2.1 was considered.

Chapter 1 considers the same empirical Bayes two-action problem that Johns and Van Ryzin studied. Theorem 1.1 improves upon their Theorem 3 by deleting assumption (B) in §1.4 and by relaxing (A). The scale exponential and the Normal densities are given to show that in each case their moment assumptions on G can be relaxed.

Chapter 2 considers the squared error loss estimation problem. Using a truncation different from that of Lin, Theorem 2.1 establishes a certain rate of convergence. Lemma 2.4 shows that for certain natural tests derivable from estimates the excess risk in the two-action problem is bounded by the square root of the corresponding excess risk in the estimation problem. Corollary 2.3 utilizes this fact to obtain better rates for the Normal two-action problem (Corollary 1.2) from those obtained in the Normal estimation problem (Corollary 2.2). The improved rates are exactly those corresponding to priors not having finite $(3+\sqrt{89})/10$ - th absolute moment.

Notational Conventions.

Sets and their corresponding indicator functions will be used interchangeably. The same symbols will be used to denote distribution functions and their induced Lebesgue-Stieltje measures. For any measure μ, the μ-integral of Y will be denoted by μY, $\mu[Y]$ or $\mu\{Y\}$. Dependence on arguments will be suppressed for simplicity and dummy variables of integration will not be displayed except for emphasis.

CHAPTER 1

LINEAR LOSS TWO-ACTION PROBLEM

1.1. Introduction.

Let us consider the following hypotheses testing problem. Let $\theta \sim G$. We test

$$H_1 : \theta \le c \quad \text{against} \quad H_2 : \theta > c$$

based on an observation X, with $X|\theta$ being distributed according to some F_θ with Lebesgue density f_θ. Let A_1 and A_2 respectively denote the actions of deciding on H_1 and H_2, and

$$L_1(\theta) \ge 0, \; L_2(\theta) \ge 0$$

denote the losses of A_1 and A_2 when θ is the true parameter.

Let P denote the p-measure on (X,θ). A randomized test ϕ in the Bayes problem above incurs a risk given below by

$$R(\phi,G) = P\{\phi L_1 + (1-\phi)L_2\}. \tag{1.1}$$

Let R^* or $R^*(G)$ denote the Bayes risk versus G. (We tacitly assume that $P_X(L_1 - L_2)$ is well-defined. This will be the case for the application of the theory to the two-action problem in exponential families with linear losses.)

Since a test is Bayes if and only if it minimizes the expected loss given x,

$$\phi_G(x) = [P_X(L_1 - L_2) \le 0] \tag{1.2}$$

is Bayes versus G. Johns (1957) considered the linear losses

$$L_1(\theta) = (\theta - c)^+ , \quad L_2(\theta) = (\theta - c)^-, \tag{1.3}$$

and intended, as a consequence, that $P_X(L_1 - L_2)$ be expressible in terms of the posterior mean; that is,

$$P_X(L_1 - L_2) = P_X(\theta - c) . \tag{1.4}$$

Hereafter, unless stated otherwise, we will assume that L_1 and L_2 are as defined in (1.3).

We remark that, although the losses in (1.3) are unbounded, the Bayes risk $R^*(G)$ may be uniformly bounded on the class of all priors; for example, let $X \sim N(\theta,1)$ and consider the natural test $\phi'(X) = [X \le c]$. Taking conditional expectation given θ, $P_\theta\{\phi' L_1 + (1-\phi')L_2\} = |\theta - c|\Phi(-|\theta - c|)$ is less than $(2\pi)^{-\frac{1}{2}}$ by the Normal tail bound (Feller (1962), p. 166). Therefore, the Bayes risk in the Normal two-action problem is less than $(2\pi)^{-\frac{1}{2}}$ whatever be G.

1.2. The Empirical Bayes Problem.

In this chapter we shall consider the case when a sequence of past observations $X_1, X_2, \ldots, X_n$ is available, with each of the X's i.i.d. according to the marginal distribution of x. At the $(n+1)^{st}$ problem, the decision rule ϕ_n is allowed to depend on all the past observations as well as the $(n+1)^{st}$. Hence, ϕ_n is a measurable function of $X_1, X_2, \ldots, X_n$ and $X = X_{n+1}$. With P extended to denote the product measure on (X, θ),

$X_1, X_2, \ldots, X_n$, we can express the risk of ϕ_n by

$$(1.5) \qquad R_n(\phi_n, G) = P\{\phi_n L_1 + (1-\phi_n) L_2\} .$$

We note that since $P_{\underline{X}_n, X}\{g(\theta)\} = P_X\{g(\theta)\}$ for any function $g(\theta)$, it follows that ϕ_G continues to be Bayes in the empirical Bayes problem. This motivates the use of the excess risk (regret)

$$(1.6) \qquad R_n - R^* = R_n(\phi_n, G) - R^*$$

as a measure of goodness of a test ϕ_n. Restricting G to those with finite Bayes risk, the excess risk satisfies

$$(1.7) \qquad 0 \le R_n - R^* = P\{(\phi_n - \phi_G)(P_X \theta - c)\} .$$

Note that the integrand $(\phi_n - \phi_G)(P_X \theta - c)$ is non-negative since ϕ_G continues to be Bayes.

1.3. Exponential Families.

Let h be a non-negative measurable function defined on the real line, and

$$\Omega = \{-\infty < \theta < \infty : \int e^{-\theta x} h \, dx < \infty\} .$$

For each θ in the natural parameter space Ω, let

$$(1.8) \qquad f_\theta(x) = \beta(\theta) \, h(x) \, e^{-\theta x} , \text{ where } \frac{1}{\beta(\theta)} = \int e^{-\theta x} h(x) \, dx.$$

The following lemma, due to Professor J. Hannan, yields a choice $h_{\mathscr{L}}$ of h such that on the set of x for which $h_{\mathscr{L}}$ is positive, the function

$$(1.9) \qquad J(x) = \int \beta(\theta) \, e^{-\theta x} \, dG(\theta)$$

is infinitely differentiable and its derivatives can be computed by repeated differentiation under the integral sign.

__Lemma 1.1__. Let $\mathscr{G} = \{G : G$ is a distribution on $\Omega\}$ and $C_G = \{x : J(x) < \infty\}$, for each $G \in \mathscr{G}$. Then there exists a determination $h_{\mathscr{G}}$ within the Lebesgue equivalence class of h (independent of $G \in \mathscr{G}$), for which $[h_{\mathscr{G}} > 0] \subseteq \text{int}(C_G)$, whatever be G.

__Proof__. The fact that hJ is a density implies that $[h > 0] \leq C_G$ a.e. for each $G \in \mathscr{G}$. The closed convex set $\bar{C}_{\mathscr{G}} = \cap\{\bar{C}_G : G \in \mathscr{G}\}$ is also the countable intersection $\cap\{\bar{C}_{G_r} :$ rational $r \notin \bar{C}_{\mathscr{G}}\}$ where $\bar{C}_{G_r}$ is any one of the $\bar{C}_G$ that excludes r. The above considerations, together with the fact that a countable union of null sets is null, imply that $[h > 0] \leq \bar{C}_{\mathscr{G}}$ a.e. and, therefore, also $[h > 0] \leq \text{int}(\bar{C}_{\mathscr{G}})$ a.e. Hence, by defining $h_{\mathscr{G}} = 0$ off $\text{int}(\bar{C}_{\mathscr{G}})$ and $h_{\mathscr{G}} = h$ on $\text{int}(\bar{C}_{\mathscr{G}})$, it follows that $[h > 0] \subseteq \text{int}(\bar{C}_{\mathscr{G}}) \subseteq \text{int}(\bar{C}_G) \subseteq \text{int}(C_G)$, whatever be G.

__Remark__.

Since J is well known to be infinitely differentiable on $\text{int}(C_G)$ and its derivatives can be computed by repeated differentiation under the integral sign, it follows that the same hold true on the subset $[h_{\mathscr{G}} > 0]$. Therefore, with

$$f = \int f_\theta \, dG(\theta) \tag{1.10}$$

denoting the marginal density, the existence of $h_{\mathscr{G}}^{(r)}$ on $[h_{\mathscr{G}} > 0]$ will imply the existence of $f^{(r)}$ via the Leibniz's rule of differentiation for the product $f = J\, h_{\mathscr{G}}$. We shall make use of this fact immediately after the following summary.

1.4. Summary and Some Useful Results.

Johns and Van Ryzin (1967) considered the two-action empirical Bayes problem in exponential families with densities (1.8) under the additional assumption that there is an $a \geq -\infty$ such that

$$(1.11) \qquad h(x) > 0 \quad \text{if and only if} \quad x > a.$$

For each integer $r \geq 2$, they exhibited procedures ϕ_n such that under the assumptions:

(A) $\quad h^{(r)}$ exists and is continuous for $x > a$

and

(B) $\quad G|\theta|^r < \infty$,

together with the conditions (C) and (D) of Theorem 1.1, the regret can be shown to converge to zero at a rate no worse than $n^{-\gamma}$, where $\gamma = (r-1)\delta/(2r+1)$ and $0 \leq \delta \leq 2$. Moreover, they gave the Normal $(-\theta,1)$ and the scale exponential families as examples where conditions (C) and (D) hold for some $0 \leq \delta \leq 1$ whenever the prior G has certain moments finite.

We shall show in Theorem 1.1 that only the existence of $h^{(r)}$ together with (C) and (D) are required for the regret convergence of $O(n^{-\gamma})$. The Normal and the scale exponential examples will be discussed in Corollaries 1.1 and 1.2; and we will show that in each case their moment assumptions can be relaxed.

We will further show in Theorem 1.2 that analysis similar to that in Theorem 1.1 can be carried out in exponential families (1.8) where h is not assumed to have any derivatives.

In the remainder of Part 1, $\mathcal{G}$ is assumed to be the class of priors G for which the Bayes risk is finite, and only exponential families as defined in (1.8) and (1.11) will be considered; moreover, since $[x \le a]$ is a P-null set, all statements are assumed to be quantified by $x > a$ unless stated otherwise.

We note that since $[x > a]$ is an open set, the h in (1.11) is already its own $h_{\mathcal{G}}$ determination. By the remark follow- Lemma 1.1, the existence of $h^{(r)}$ implies the existence of $f^{(r)}$. This improves upon Lemmas 2, 3 and 4 of Johns and Van Ryzin in that their respective moment assumptions $G|\theta| < \infty$, $G|\theta|^r < \infty$ and $G|\log \theta|^r < \infty$ are deleted.

For the exponential family in (1.8) and (1.11),

$$(1.12) \qquad P_X(\theta) = -\frac{J^{(1)}}{J} \qquad (\text{for } x > a).$$

Hence, the quantity $P_X(L_1 - L_2) = P_X(\theta - c)$ and, therefore, also the Bayes test ϕ_G in (1.2), are well defined without any assumption on G. In addition, if $h^{(1)}$ exists then, with

$$(1.13) \qquad v = \frac{h^{(1)}}{h}, \quad g = f^{(1)} \quad \text{and} \quad \alpha = f \quad P_X(\theta - c),$$

we have

$$(1.14) \qquad P_X(\theta) = v - \frac{g}{f} \quad \text{and} \quad \alpha = (v-c)f - g .$$

We note that the Bayes test in (1.2) becomes

$$(1.15) \qquad \phi_G(x) = [\alpha(x) \le 0] .$$

When a sequence of i.i.d. observations $X_1,\ldots,X_n$ and X is available, it is the special form of ϕ_G in (1.15) that we will exploit in defining reasonable extimates ϕ_n by estimating the density f and its derivative g by the kernel method so successfully employed by Johns and Van Ryzin.

To conclude this section, we state and prove Lemma 1 of Johns and Van Ryzin (1967) as a consequence of (1.7).

Lemma 1.2. Let α_n be any measurable function of $X_1,\ldots,X_n$ and X. Then the excess risk of

$$\phi_n = [\alpha_n \le 0] \tag{1.16}$$

satisfies

$$0 \le R_n - R^* \le \int_a^\infty |\alpha| P_X[|\alpha_n - \alpha| \ge |\alpha|]dx. \tag{1.17}$$

Proof. From (1.7) and (1.13),

$$0 \le R_n - R^* = \int_a^\infty |\alpha| P_X| \phi_n - \phi_G |dx . \tag{1.18}$$

The result follows from (1.18) since $|\phi_n - \phi_G| \le [|\alpha_n - \alpha| \ge |\alpha|]$.

1.5. Main Result and Examples.

In view of (1.14) and (1.17), the excess risk $R_n - R^*$ can be made small if f and g can be adequately estimated. The appendix provides kernel estimates f_n and g_n for which the bias terms $P_X f_n - f$ and $P_X g_n - g$ are small. These estimates will be used in the obvious way to define α_n and ϕ_n in (1.19).

Theorem 1.1 below is an improvement of Theorem 3 of Johns and Van Ryzin (1967) in that their assumptions $G|\theta|^r < \infty$ and

$h^{(r)}$ is continuous are deleted. Their proof is reproduced below for completeness.

For each integer $r \geq 2$, let

(1.19) $$\phi_n = [\alpha_n \leq 0] \text{ , where } \alpha_n = (v-c)f_n - g_n$$

with

$$f_n(x) = n^{-1} \sum_{j=1}^{n} W_j^0(\Delta) \text{ , } W_j^0(\Delta) = \Delta^{-1} K_0((X_j - x)/\Delta)$$

and

$$g_n(x) = (n\Delta)^{-1} \sum_{j=1}^{n} (W_j^1(2\Delta) - W_j^1(\Delta)), \; W_j^1(\Delta) = \Delta^{-1} K_1((X_j - x)/\Delta)$$

being the type of kernel estimates of f and g given in (A.8) of the appendix. We note that $r \geq 2$ is required in (A.1).

<u>Theorem 1.1</u>. Let ϕ_n be as in (1.19) with $\Delta = n^{-1/(2r+1)}$. If $h^{(r)}$ exists (for $x > a$), and if there is some $\epsilon > 0$ such that

(C) $$\int_a^\infty |\alpha|^{1-\delta}(1 + |v|)^\delta (q_\epsilon^{(0)})^{\delta/2} \, dx < \infty, \; q_\epsilon^{(0)}(x) = \sup_{0<u<1} f(x + \epsilon u)$$

(D) $$\int_a^\infty |\alpha|^{1-\delta}(1 + |v|)^\delta (q_\epsilon^{(r)})^\delta \, dx < \infty, \; q_\epsilon^{(r)}(x) = \sup_{0<u<1} |f^{(r)}(x + \epsilon u)|$$

then,

$$0 \leq R_n(\phi_n, G) - R^* = O(n^{-\gamma}), \text{ where } \gamma = \frac{r-1}{2r+1}\delta .$$

<u>Proof</u>. Lemma 1.2, followed by the Markov inequality, yields

(1.20) $$0 \leq R_n(\phi_n, G) - R^* \leq \int_a^\infty |\alpha|^{1-\delta} P_X|\alpha_n - \alpha|^\delta dx .$$

Since (1.14), (1.19) together with the C_r-inequality (Loève p. 155) imply

$$|\alpha_n - \alpha|^\delta \le C_\delta \{|v - c|^\delta |f_n - f|^\delta + |g_n - g|^\delta\} ,$$

we have, by (1.20),

$$0 \le R_n(\phi_n,G) - R^* \le C_\delta \{A + B\}$$

where

$$A = \int_a^\infty |\alpha|^{1-\delta} |v - c|^\delta P_X|f_n - f|^\delta dx$$

and

$$B = \int_a^\infty |\alpha|^{1-\delta} P_X|g_n - g|^\delta dx .$$

Thus, the rate at which the regret converges to zero is no worse than that of $\max(A,B)$. Let us first consider A. For $\delta > 0$, the C_r-inequality yields

$$|f_n - f|^\delta \le C_\delta \{|f_n - P_X f_n|^\delta + |P_X f_n - f|^\delta\} \tag{1.21}$$

and for $0 < \delta < 2$, Holder's inequality yields

$$P_X|f_n - P_X f_n|^\delta \le (\mathrm{Var}_X f_n)^{\delta/2} .$$

Since the above inequality trivially holds for $\delta = 0$ and 2, it follows from (1.21) that

$$P_X|f_n - f|^\delta \le C_\delta \{(\mathrm{Var}_X f_n)^{\delta/2} + |P_X f_n - f|^\delta\} . \tag{1.22}$$

Thus by (A.9) and (A.10) of the appendix,

$$P_X|f_n - f|^\delta \le \text{const} \times \{[(n\Delta)^{-1} q_\epsilon^{(0)}]^{\delta/2} + [\Delta^r q_\epsilon^{(r)}]^\delta\}$$

so that by (C), (D), and the choice $\Delta = n^{-1/(2r+1)}$, one has

$$A = O((n\Delta)^{-\delta/2}) + O(\Delta^{r\delta}) = O(n^{-r\delta/(2r+1)}) .$$

Similarly, for $0 \le \delta \le 2$,

$$P_X|g_n - g|^\delta \le C_\delta \{(\mathrm{Var}_X g_n)^{\delta/2} + |P_X g_n - g|^\delta\}$$

$$= \text{const} \times \{[(n\Delta^3)^{-1} q_\varepsilon^{(0)}]^{\delta/2} + [\Delta^{r-1} q_\varepsilon^{(r)}]^\delta\}$$

so that by (C) and (D),

$$B = O((n\Delta^3)^{-\delta/2}) + O(\Delta^{(r-1)\delta}) = O(n^{-\gamma}) .$$

The proof is completed by this weaker rate of B.

For the remainder of this section, the scale exponential and the location Normal families will be given as examples to illustrate how conditions (C) and (D) relate to the moments of G.

<u>Example 1</u>. (Scale Exponential)

Consider the exponential density in (1.8) with $h = [x > 0]$ and $\beta(\theta) = \theta$; i.e., for each $\theta > 0$

$$f_\theta(x) = \begin{cases} \theta\, e^{-\theta x} & , \quad x > 0 \\ 0 & , \quad \text{otherwise.} \end{cases} \tag{1.23}$$

The density f satisfies the following facts:

(1.24a) f_θ is monotonically decreasing, and so is f.

(1.24b) Since $h^{(r)} = 0$ for $x > 0$, $f^{(r)}$ exists (for $x > 0$) by Lemma 1.1; moreover, $v = 0$ so that conditions (C) and (D) simplify.

(1.24c) $|f^{(r)}| = \int \theta^r f_\theta dG(\theta)$ is monotonically decreasing

and, therefore,

$$q_\varepsilon^{(r)} = |f^{(r)}| . \tag{1.24d}$$

Corollary 1.1 is an improvement over Corollary 3.1 of Johns-Van Ryzin (1967). They proved the same result under the assumptions $G\theta^{r+1} < \infty$ and (1.26) below.

Corollary 1.1. For the scale exponential in (1.23), the hypothesis of Theorem 1.1 holds for each $0 \le \delta \le 1$ if

(1.25) $$G[\theta^r] < \infty ,$$

(1.26) $$G[\theta^{-\eta}] < \infty, \text{ where } \eta = (1+t)\delta/(2-\delta) \text{ for some } t > 0.$$

Proof. Since $v = 0$, condition (D) simplifies and is implied by the integrability of α and $q_\epsilon^{(r)}$, subsequently illustrated. By Tonelli's theorem (Royden (1965), p. 234),

$$\int|\alpha|dx \le \iint|\theta - c|f_\theta dG\,dx = G|\theta - c|.$$

By (1.24c) and (1.24d),

$$\int q_\epsilon^{(r)}dx = \iint \theta^r f_\theta dG\,dx = G[\theta^r] .$$

Hence, we have shown that $G[\theta^r] < \infty$ is sufficient for condition (D).

Let us next verify condition (C). Since α is bounded by $G|\theta(\theta - c)|$, $v = 0$, and $q_\epsilon^{(0)} = |f| \le G[\theta]$, it follows that, under (1.25), condition (C) is implied by

(1.27) $$\int_1^\infty |\alpha|^{1-\delta} f^{\delta/2}dx < \infty .$$

Since $\theta \le e^\theta$, $|f^{(1)}(x)| \le f(x - 1)$ for $x > 1$; consequently, $|\alpha(x)| = |cf + f^{(1)}| \le (c+1)f(x-1)$ for $x > 1$. Thus, by the Holder inequality,

(1.28) $$\int_1^\infty |\alpha|^{1-\delta} f^{\delta/2} dx \le (c+1)^{1-\delta}(1/t)^{\delta/2}\{P[1+x]^\eta\}^{1-\delta/2}.$$

The proof is completed by the equality $P\{x^{\eta}\} = G\{\theta^{-\eta}\}\Gamma(1+\eta)$.

<u>**Remark**</u>. Corollary 1.1 shows that procedures ϕ_n exist, for which the regret convergence rate can be arbitrarily close to $n^{-\frac{1}{2}}$ provided $\delta = 1$ and r is sufficiently large, i.e., G has finite (-1)- as well as arbitrarily high moments.

<u>Example 2</u>. (Normal $(-\theta,1)$).

Consider the exponential family in (1.8) with $h(x) = e^{-x^2/2}$ and $\beta(\theta) = (2\pi)^{-\frac{1}{2}} e^{-\theta^2/2}$; that is, for each $-\infty < \theta < \infty$,

$$f_\theta(x) = (2\pi)^{-\frac{1}{2}} e^{-(\theta+x)^2/2}, \quad \text{where} \quad -\infty < x < \infty .$$

We have shown earlier (§1.3) that for this family the Bayes risk $R^*(G) < (2\pi)^{-\frac{1}{2}}$ whatever be G.

Since the function $e^{-y^2/2} + e^{-(y+\varepsilon)^2/2}$ is symmetric with respect to $y = -\varepsilon/2$, and has a unique minimum there with value $2e^{-\varepsilon^2/8}$, it follows that

$$(1.29) \qquad \begin{aligned} f_\theta(x+t) &\le f_\theta(x) + f_\theta(x+\varepsilon), \\ q_\varepsilon^{(0)}(x) &\le f(x) + f(x+\varepsilon), \end{aligned} \qquad \text{for } 0 \le t \le \varepsilon \le \sqrt{8 \log 2} .$$

By repeated differentiation under the integral sign,

$$f^{(r)}(x) = (-1)^r \int H_r(x+\theta) f_\theta(x) dG(\theta) ,$$

where H_r is the r-th Hermite polynomial. Thus, for $\varepsilon \le \sqrt{8 \log 2}$,

$$(1.30) \qquad \begin{aligned} |f^{(r)}(x)| &\le \sum_0^r |a_j| \int |x+\theta|^j f_\theta(x)\, dG(\theta) , \\ q_\varepsilon^{(r)}(x) &\le \sum_0^r |a_j| C_j \int (|x+\theta|^j + \varepsilon^j)(f_\theta(x) + f_\theta(x+\varepsilon)) dG(\theta), \end{aligned}$$

where the second inequality follows from the first via (1.29) and the C_r-inequality. Lastly,

$$f_\theta \le (2\pi)^{-\frac{1}{2}} \quad , \quad f \le (2\pi)^{-\frac{1}{2}}$$

(1.31) $$|\alpha| \le \int |\theta - c| f_\theta \, dG(\theta) \le (2\pi)^{-\frac{1}{2}} G|\theta - c| .$$

$$q_\epsilon^{(0)} \le (2\pi)^{-\frac{1}{2}} , \ q_\epsilon^{(r)} \text{ is bounded.}$$

<u>Remark</u>. Corollary 1.2 below is an improvement of Corollary 3.2 of Johns-Van Ryzin. They proved the corollary under the stronger assumption $G|\theta|^{1+(3+t)\delta/(2-\delta)} < \infty$, and $G|\theta|^r < \infty$.

<u>Corollary 1.2</u>. Consider the Normal $(-\theta,1)$ family. For each $0 \le \delta \le 1$, if

(1.32) $$G|\theta|^{1+(2+t)\delta/(2-\delta)} < \infty \quad \text{for some} \quad t > 0 ,$$

then the hypothesis of Theorem 1.1 holds for each $r \ge 2$.

<u>Proof</u>. Condition (D) is implied by the integrability of α and $|x|\, q_\epsilon^{(r)}$, since $1 + |v| = 1 + |x|$ is bounded by $2|x|$ for $|x| > 1$. By (1.31), if $G|\theta| < \infty$ then

(1.33) $$\int |\alpha| dx \le G|\theta - c| < \infty .$$

Denote by b_j the constant $(2\pi)^{-\frac{1}{2}} \int |z|^j e^{-z^2/2} dz$. Since $P_\theta |x + \theta|^j = b_j$, it follows, by the triangle inequality, that $P_\theta[|x|\,|x + \theta|^j] \le P_\theta[(|x+\theta| + |\theta|)|x + \theta|^j] = b_{j+1} + |\theta| b_j$. Hence, $G|\theta| < \infty$ implies $P(|x|\ |x + \theta|^j) < \infty$ for each j and therefore $|x|\ q_\epsilon^{(r)}$ is integrable by (1.30). This completes the verification of (D) under $G|\theta| < \infty$.

Let us next consider condition (C) for $\delta = 0$, $\delta = 1$, and $0 < \delta < 1$.

Case 1 ($\delta = 0$). (1.33) proves this case.

Case 2 ($\delta = 1$). Since $|v| = |x|$ and $q_\varepsilon^{(0)} \le (2\pi)^{-\frac{1}{2}}$, we need only to verify the integrability of $[|x| > 1]\,|x|(q_\varepsilon^{(0)})^{\frac{1}{2}}$. By Holder's inequality,

$$\int_{|x|>1} |x|(q_\varepsilon^{(0)})^{\frac{1}{2}}dx \le \left(\frac{2}{t}\right)^{\frac{1}{2}} \{\int |x|^{3+t} q_\varepsilon^{(0)} dx\}^{\frac{1}{2}} ,$$

where the last integral is bounded by

$$\int |x|^{3+t}(f(x) + f(x+\varepsilon))dx$$

$$\le P[|x+\theta| + |\theta|]^{3+t} + P[|x+\theta| + |\theta+\varepsilon|]^{3+t},$$

via (1.29) and the triangle inequality. Again by the fact that $(x+\theta)$ given θ is standard Normal, $G|\theta|^{3+t} < \infty$ implies Case 2.

Case 3 ($0 < \delta < 1$). Let $0 \le \xi \le 1$, $0 < t$. With $0 < 1/p = \delta/2 < 1$, $0 < 1/q = \frac{2-\delta}{2} < 1$, $X = |x|^{(1-\xi)\delta}(q_\varepsilon^{(0)})^{\delta/2}$ and $Y = |x|^{\xi\delta}|\alpha|^{1-\delta}$ in the Holder inequality, it follows that (C) is implied by the integrability of X^p and Y^q. By (1.29),

$$\int X^p dx \le P\,x^{2(1-\xi)} + 2\,P|x-\varepsilon|^{2(1-\xi)} ,$$

so that $G|\theta|^{2(1-\xi)} < \infty$ implies the integrability of X^p.

If $G|\theta| < \infty$, then α is bounded by (1.31), and Y is bounded on $|x| \le 1$. Therefore, the integrability of Y^q is implied by that of $[|x| > 1]Y^q$. By Holder's inequality,

$$\int [|x| > 1]Y^q dx \le \left(\frac{2}{t}\right)^{\delta/(2-\delta)} \times \{\int |x|^u |\alpha| dx\}^{2(1-\delta)/(2-\delta)},$$

where $u = \frac{1}{2}(1 + t + 2\xi)\delta/(1-\delta)$. By Tonelli's theorem, $\int|x|^u|\alpha|dx \le \int|\theta - c|P_\theta|x|^u dG(\theta)$. Since $(x+\theta)$ given θ is standard Normal, $P_\theta|x|^u$ is bounded by

$$P_\theta|x|^u \le C_u \times \{P_\theta|x + \theta|^u + |\theta|^u\} ,$$

by the C_r-inequality. Thus, $G|\theta|^{1+u} < \infty$ implies that Y^q is integrable. Balancing between $1+u$ and $2(1-\xi)$, we get $\max(1+u, 2(1-\xi))$ is minimized when $1-2\xi = \delta(2+t)/(2-\delta)$, so that $2(1-\xi) = 1+\delta(2+t)/(2-\delta)$. Therefore, (1.32) implies Case 3.

Remark. Corollary 1.2 shows that for the Normal $(-\theta,1)$ family there exist procedures for which the regret convergence to zero is of a rate no worse than $n^{-\gamma}$, provided that the prior has finite $1+\delta(2+t)/(2-\delta)$th absolute moment , where $0 \le \delta \le 1$. In the case where $\delta = 1$ and r is sufficiently large, a rate close to $n^{-\frac{1}{2}}$ can be achieved provided the prior G has 3+ absolute moments. However, for $\delta = 0$, the finiteness of the first moment of G guarantees only the boundedness of the excess risk. This lack of rate will be removed in Corollary 2.4.

1.6. Result Without Differentiability of h.

In Section 1.5 we discussed the exponential family in (1.11) and (1.8). We took advantage of the existence of $h^{(r)}$ and obtained the result in Theorem 1.1. In this section we shall not assume h to have any derivative. We recall from (1.9) the definition

$$J(x) = \int e^{-\theta x} \beta(\theta)dG(\theta) . \tag{1.34}$$

It was shown in Lemma 1.1 that J is infinitely differentiable on

$[h_{\ell} > 0]$ and, therefore, also on $[x > a]$. Since $f = Jh$, it follows from (1.12), (1.13) that

(1.35) $$\alpha = -(J^{(1)} + cJ)h.$$

In view of the method of attack exhibited in Sections 1.4 and 1.5 , we shall estimate ϕ_G through J and $J^{(1)}$.

For each $r \geq 2$, let

(1.36) $$J_n(x) = n^{-1} \sum_{j=1}^{n} W_j^0(\Delta)/h(X_j)$$
$$J_n'(x) = (n\Delta)^{-1} \sum_{j=1}^{n} \{W_j^1(2\Delta) - W_j^1(\Delta)\}/h(X_j)$$

where W_j^0 and W_j^1 are as defined in (A.8) of the appendix. Let

(1.37) $$\phi_n = [\alpha_n \leq 0], \text{ where } \alpha_n = -(J_n' + cJ_n)h .$$

<u>Theorem 1.2</u>. Let ϕ_n be as in (1.37). Consider the exponential family in (1.11) and (1.8). For each $0 \leq \delta \leq 2$, if there exists some $\epsilon > 0$ such that

(C') $$\int_a^\infty |\alpha|^{1-\delta}(T_\epsilon^{\frac{1}{2}} h)^\delta \, dx < \infty \ , \ T_\epsilon(x) = \sup_{0<u<1} \frac{J(x+\epsilon u)}{h(x+\epsilon u)}$$

(D') $$\int_a^\infty |\alpha|^{1-\delta}(S_\epsilon^{(r)} h)^\delta \, dx < \infty \ , \ S_\epsilon^{(r)}(x) = \sup_{0<u<1} |J^{(r)}(x+\epsilon u)| ,$$

then, with $\Delta = n^{-1/(2r+1)}$, we have

(1.38) $$0 \leq R_n(\phi_n, G) - R^* = O(n^{-\gamma}), \text{ where } \gamma = (r-1)\delta/(2r+1).$$

<u>Proof</u>. From (1.20), it follows by the C_r-inequality that the excess risk is bounded above by $C_\delta \times (A + B)$ where

$$A = |c|^{\delta} \int_a^{\infty} |\alpha|^{1-\delta} P_X|(J_n - J)h|^{\delta}\, dx$$

and

$$B = \int_a^{\infty} |\alpha|^{1-\delta} P_X|(J_n' - J^{(1)})h|^{\delta}\, dx .$$

With J_n and J replacing f_n and f in (1.22) we have,

$$P_X|J_n - J|^{\delta} \le C_{\delta}\{(\mathrm{Var}_X J_n)^{\delta/2} + |P_X J_n - J|^{\delta}\}, \text{ for } 0 \le \delta \le 2.$$

Under (C') and (D') and Lemmas A.5 and A.6 of the appendix and the choice $\Delta = n^{-1/(2r+1)}$, then

$$A = O((n\Delta)^{-\delta/2}) + O(\Delta^{r\delta}) = O(n^{-r\delta/(2r+1)}) .$$

Similarly, for $0 \le \delta \le 2$,

$$P_X|J_n' - J^{(1)}|^{\delta} \le C_{\delta}\{(\mathrm{Var}_X J_n')^{\delta/2} + |P_X J_n' - J^{(1)}|^{\delta}\}.$$

Invoking (C'), (D') and Lemmas A.5 and A.6 of the appendix,

$$B = O((n\Delta^3)^{-\delta/2}) + O(\Delta^{(r-1)\delta}) = O(n^{-\gamma}) .$$

The proof is completed by the weaker rate of B.

Example 3. Consider the exponential family with

(1.39) $\quad h = [0 < x \le 1] + 2[1 < x < \infty]$.

Then $\Omega = (0,\infty)$ and $\beta(\theta) = \theta/(1 + e^{-\theta})$. We note that

(1.40a) $\quad h$ is non-decreasing while J is strictly decreasing.

(1.40b) $\quad |J^{(r)}| = \int \theta^r \beta(\theta) e^{-\theta x} dG(\theta)$.

(1.40c) $\quad S_{\epsilon}^{(r)} = |J^{(r)}|$ and $T_{\epsilon} = \frac{J}{h}$.

Corollary 1.3. Consider the exponential family with h in (1.39). The hypothesis of Theorem 1.2 holds provided (1.25) and (1.26) hold.

Proof. The proof of Corollary 1.1 works with

(1.41) $$q_\epsilon^{(r)},\ q_\epsilon^{(0)},\ f,\ |\alpha| \le G[\theta|\theta - c|] \text{ and } f \le G[\theta]$$

respectively replaced by

(1.42) $$s_\epsilon^{(r)},\ T_\epsilon,\ J,\ |\alpha| \le 2G[\theta|\theta - c|] \text{ and } f \le 2G[\theta].$$

CHAPTER 2

SQUARED ERROR LOSS ESTIMATION PROBLEM

2.1. Introduction.

Suppose θ is distributed according to some prior G, and one is to estimate θ based on an observation X with $X|\theta$ distributed according to the exponential family given in (1.8) and (1.11); that is, for some $a \geq -\infty$,

$$f_\theta(x) = \beta(\theta)\, h(x)\, e^{-\theta x} , \tag{2.1}$$

where

$$h > 0 \quad \text{if and only if} \quad x > a . \tag{2.2}$$

Let P denote the joint p-measure on (X,θ) as in Chapter 1. Let the loss function be the squared error loss. The risk of an estimate ϕ is then given by $R(\phi,G) = P(\phi - \theta)^2$ with Bayes risk

$$R^*(G) = \inf_\phi P(\phi - \theta)^2 . \tag{2.3}$$

We note that R and R^* denote different quantities in Chapter 1.

In order that the problem not be totally uninteresting, we restrict G to those with finite Bayes risk. We note that the Bayes risk $R^*(G)$ can be uniformly bounded in G. For example, let $X \sim N(\theta,1)$. Then the natural estimate $\phi'(X) = X$

has risk $P(\phi' - \theta)^2 = 1$. Therefore, $R^*(G) \le 1$ whatever be G.

Extend P to denote the product p-measure on (X,θ), $X_1, X_2, \ldots,$ and X_n. Let ψ_n be any measurable function of $X_1,\ldots,X_n$ and X. The risk of ψ_n is then given by

$$R_n(\psi_n, G) = P(\psi_n - \theta)^2 .$$

Let ψ_G be a Bayes estimate versus G. If $\psi_n - \psi_G \in L_2(P)$ then $P(\psi_n - \psi_G)(\psi_G - \theta) = 0$, and the excess risk satisfies

$$(2.4) \qquad 0 \le R_n(\psi_n, G) - R^*(G) = P(\psi_n - \psi_G)^2.$$

We recall the following definitions from Chapter 1.

$$(2.5) \qquad v = \frac{h^{(1)}}{h}, \quad f = G(f_\theta), \quad g = f^{(1)}, \quad \text{and} \quad J(x) = \int e^{-\theta x}\beta(\theta)\,dG(\theta).$$

It is well known that a Bayes estimate under squared error loss is the posterior mean $P_X\theta$. Hence, by (1.12), the Bayes estimate ψ_G is well defined without any assumption on the prior G. Furthermore, (1.14) remains valid with $P_X\theta$ replaced by ψ_G, i.e.,

$$(2.6) \qquad \psi_G = v - \frac{g}{f} .$$

In view of (2.4), it is now a matter of estimating ψ_G by estimating the density f and its derivative g.

2.2. Estimation of $\psi_G = P_X\theta$.

We shall exploit the expression in (2.6) in estimating ψ_G when a sequence of observations $X_1,\ldots,X_n$, i.i.d. according to the common density f, is available.

Let f_n and g_n respectively be any estimates of f and g. Let $\eta > 0$. Truncate f_n away from 0 by

$$f'_n = f_n \vee \eta \, , \tag{2.7}$$

and define

$$\psi_n = v - \frac{g_n}{f'_n} \, . \tag{2.8}$$

<u>Lemma 2.1</u>. For each $\eta > 0$, the estimate ψ_n in (2.8) satisfies

$$P(\psi_n - \psi_G)^2 \le 3(\eta^{-2} A + \eta^{-2} B + C) \, , \tag{2.9}$$

where $A = P(g_n - g)^2$, $B = P(g/f)^2(f_n - f)^2$, and $C = P(g/f)^2[f < \eta]$.

<u>Proof</u>. From (2.7) and (2.8), simple algebraic manipulation followed by the triangle inequality will yield

$$\eta|\psi_n - \psi_G| = \eta\left|\frac{g_n}{f'_n} - \frac{g}{f}\right| \le \left|g_n - \frac{g}{f} f'_n\right| \tag{2.10}$$

$$\le |g_n - g| + \left|\frac{g}{f}\right| \, |f - f'_n| \, .$$

Since $|f - f'_n| \le \eta[f < \eta] + |f - f_n|$, the proof follows from (2.10) and the inequality $(a + b + c)^2 \le 3(a^2 + b^2 + c^2)$.

Lemma 2.1 shows that for any estimate ψ_n of the form in (2.8), the regret can be bounded in terms of A, B and C in (2.9). The first two terms, namely A and B, involve $P_X(f_n - f)^2$ and $P_X(g_n - g)^2$. The appendix gives kernel estimates f_n and g_n, for which these quantities are small. Therefore, hereafter, we shall consider f_n and g_n to be the kernel estimates given in (A.8) and that ψ_n in (2.8) is to be defined in terms of these estimates.

2.3. Summary.

Theorem 2.1 below is a 1-dimensional specialization of a result considered by Lin (1968). The scale exponential and the Normal densities again will serve as examples to show that the existence of certain moments of G is sufficient for the hypothesis of Theorem 2.1. In Corollary 2.4, better rates are obtained for the Normal two-action problem from those obtained in the Normal estimation problem.

2.4 Main Results and Examples.

Theorem 2.1. Let ψ_n be of the form in (2.8) with f_n and g_n being kernel estimates of f and g as given in (A.8) of the appendix. If $h^{(r)}$ exists and if for some $0 < \epsilon$

(2.13) $$P\{(1 + (g/f)^2)q_\epsilon^{(0)}\} < \infty ,$$

(2.14) $$P\{(1 + (g/f)^2)(q_\epsilon^{(r)})^2\} < \infty ,$$

and if $\delta \geq 0$ such that

(2.15) $$P\{(g/f)^2[f < \eta]\} \leq c_1 \eta^\delta ,$$

then, with $\Delta = n^{-1/(2r+1)}$ and $\eta = n^{-\frac{1}{2+\delta}\frac{2(r-1)}{2r+1}}$

(2.16) $$0 \leq R_n(\psi_n, G) - R^* = O(n^{-\gamma'}) ,$$

where $\gamma' = \frac{\delta}{2+\delta}\frac{2(r-1)}{2r+1}$.

Proof. Let A, B and C be as in (2.9). With $\Delta = n^{-1/(2r+1)}$, Lemmas A.3 and A.4 of the appendix followed by (2.13) and (2.14) will yield

$$A \le c_2' \times (n\Delta^3)^{-1} + c_2'' \times \Delta^{2(r-1)} \le c_2 \times n^{-2(r-1)/(2r+1)} ,$$

and

$$B \le c_3' \times (n\Delta)^{-1} + c_3'' \times \Delta^{2r} \le c_3 \times n^{-2r/(2r+1)} ,$$

with the rate on A being the smaller of the two. The choice $\eta = n^{-\frac{1}{2+\delta}\frac{2(r-1)}{2r+1}}$ balances the rates of C and $\eta^{-2}A$ to $n^{-\gamma'}$. The proof is completed by Lemma 2.1.

Example 1 (Scale exponential).

Consider the scale exponential with Lebesgue densities given by (1.23), i.e.,

$$(2.17) \qquad f_\theta(x) = \begin{cases} \theta\, e^{-\theta x} , & \text{for } x > 0 \\ 0 , & \text{otherwise.} \end{cases}$$

Consider the extreme case where G is degenerate at $\theta = 1$ with all moments finite. The quantity $C = P(g/f)^2[f < \eta]$ in (2.9) can be computed to be exactly η. This motivates the bound in the following lemma.

Lemma 2.2. For the scale exponential in (2.17) if $0 < \eta \le f(1)$, then for each $p > 1$ and $1/p + 1/q = 1$,

$$(2.18) \quad P(g/f)^2\,[f < \eta] \le (\Gamma(1+2q))^{1/q}(\eta/(2p - 1))^{1/p} .$$

Proof. The inequality $(g/f)^2 = (-P_X\theta)^2 \le P_X(\theta^2)$ followed by Holder's inequality yields

$$P(g/f)^2\,[f < \eta] \le P(\theta x)^2\, x^{-2}[f < \eta]$$

(2.19) $$\le (PP_\theta(\theta x)^{2q})^{1/q}(P\,x^{-2p}[f < \eta])^{1/p}$$

$$= (\Gamma(1+2q))^{1/q}(P\,x^{-2p}[f < \eta])^{1/p}\ ,$$

where the last equality follows from the fact that conditioned on θ, θx is standard scale exponential. For $0 < \eta \le f(1)$, $[f < \eta] \le [x > 1]$ so that

$$P\,x^{-2p}[f < \eta] \le \eta \int_1^\infty x^{-2p}\,dx = \eta/(2p-1)\ .$$

This completes the proof.

Lemma 2.2 shows that (2.15) holds with $\delta = 1/p$ and $c_1 = \Gamma^{1/q}(1+2q)/(2p-1)^{1/p}$ without any assumption on the prior G. For priors with densities $\frac{1}{a}\,\theta^{a-1}[0 < \theta < 1]$, $a > 0$, it can be shown that $f(x) = x^{-(1+a)}\int_0^x z^a e^{-z} dz \sim x^{-(1+a)}\Gamma(1+a)$ and $|g(x)| \sim x^{-(2+a)}\Gamma(2+a)$ as $x \to \infty$. Hence, $(g/f)^2 \sim (1+a)^2 x^{-2}$ as $x \to \infty$, and $C \le c_1 \eta^{(2+a)/(1+a)}$. Here we see that the bound on C deteriorates as a, the number of finite moments of θ^{-1}, increases.

<u>Corollary 2.1</u>. For the scale exponential family in (2.17), the hypothesis of Theorem 2.1 holds for each $r \ge 2$ and $\delta < 1$, provided

(2.20) $$G\,\theta^{r+\frac{1}{2}} < \infty\ .$$

<u>Proof</u>. Since $(g/f)^2 \le P_X(\theta^2)$ and $q_\epsilon^{(r)} = G(\theta^{r+1} e^{-\theta x})$, it suffices to note that with $\theta_i \sim G_i = G$, $i = 1,2$,

$$P[(1+P_X(\theta^2))G(\theta\ e^{-\theta x})] = G_1\ G\ F_\theta[(1+\theta^2)\theta_1\ e^{-\theta_1 x}]$$
$$= G_1\ G[(1+\theta^2)\theta_1\ \theta/(\theta+\theta_1)]$$
$$\le G(\theta)(1+G(\theta^2)) ,$$

and furthermore, by the Arithmetic-Mean-Geometric-Mean inequality (Beckenback-Bellman (1961), p. 54),

$$P(1+P_X(\theta^2))G^2(\theta^{r+1}\ e^{-\theta x})$$
$$= G_1\ G_2\ GF_\theta[(1+\theta^2)\theta_1^{r+1}\ e^{-\theta_1 x}\ \theta_2^{r+1}\ e^{-\theta_2 x}]$$
$$= G_1\ G_2\ \theta_1^{r+1}\ \theta_2^{r+1}\ G[(1+\theta^2)\theta/(\theta+\theta_1+\theta_2)]$$
$$\le \tfrac{1}{2}\ G_1\ G_2\ \theta_1^{r+\frac{1}{2}}\ \theta_2^{r+\frac{1}{2}}\ G[(1+\theta^2)\theta^{\frac{1}{2}}]$$
$$= \tfrac{1}{2}\ G^2(\theta^{r+\frac{1}{2}})G[\theta^{\frac{1}{2}}(1+\theta^2)] .$$

The proof is complete.

Example 2 (Normal).

Let us consider the Normal $(-\theta,1)$ family with Lebesgue densities

(2.21) $$f_\theta(x) = (2\pi)^{-\frac{1}{2}}\ e^{-(x+\theta)^2/2} .$$

For each $0 \le u$ and $1 \le v$, we note that

(2.22) $$b_u = P_\theta|x + \theta|^u$$

is the finite constant $\int |z|^u\ e^{-z^2/2}dz/(2\pi)^{\frac{1}{2}}$ and by Jensen's inequality,

(2.23) $$|g/f|^v = |P_X(x + \theta)|^v \le P_X|x + \theta|^v .$$

Remark. Consider again the extreme case when G degenerates at $\theta = 0$. Then the quantity $C = P(g/f)^2[f < \eta] \sim 2L\eta$ as $\eta \to 0$

with $L^2 = -\log(2\pi\,\eta^2) = o(\eta^{-t})$ for any $t > 0$.

Proof. Since $f(L) = \eta$ and $[f < \eta] = [|x| > L]$, we have $C = 2\int_L^\infty x^2 f\,dx$ which, upon integration by parts, yields $C = 2L\eta + 2P[x > L]$. By the Normal tail bound (Feller (1962), p. 166), it follows that $2Lf(L) + 2f(L)(\frac{1}{L} - \frac{1}{L^3}) < C < 2Lf(L) + 2f(L)/L$. Consequently, $C \sim 2L\eta$. The proof is completed by the fact that $L = o(n^{-t})$ for any $t > 0$.

The above remark motivates the bound in the next lemma.

Lemma 2.3. Consider the Normal $(-\theta,1)$ in (2.21). For each $0 \le \delta < 1$, (2.15) holds if

$$(2.24) \qquad G|\theta|^{(1+t)\delta/(1-\delta)} < \infty \quad \text{for some } t > 0.$$

Proof. By the Holder inequality,

$$P(g/f)^2[f < \eta] \le I^{1/p}\ II^{1/q},$$

where

$$I = P|g/f|^{2p} \le PP_X|x+\theta|^{2p} = PP_\theta|x+\theta|^{2p} = b_{2p}$$

by (2.23) and (2.22), and

$$II = P[f < \eta] \le \eta^s \int f^{1-s}dx, \text{ for any } 0 \le s < 1.$$

Since the density f is bounded by $(2\pi)^{-\frac{1}{2}}$, the integrability of f^{1-s} is implied by that of $[|x| > 1]f^{1-s}$. Temporarily, let $\nu(s) = (1+a)s/(1-s)$ for each $a > 0$. The Holder inequality followed by the C_r-inequality yields

$$\int [|x| > 1]f^{1-s}dx \le (2/a)^s\, P^{1-s}(|x|^\nu)$$
$$\le (2/a)^s\, \{(b_\nu + P|\theta|^\nu)C_\nu\}^{1-s}.$$

Hence, $G|\theta|^{\nu(s)} < \infty$ implies that f^{1-s} is integrable and, therefore, (2.15) holds with the rate s/q. Since (2.24) implies that there exists some $0 < a < t$ for which $G|\theta|^{\nu(\delta+)} < \infty$ with $\delta+ > \delta$, the proof above shows that (2.15) holds with rate $\delta+/q$. The proof is completed by the choice $\delta+/q = \delta$. Such a choice is possible since $1 < q$ is a free parameter.

Corollary 2.2. Consider the Normal $(-\theta,1)$ family. For each $0 \le \delta < 1$, if (2.24) holds, then the hypothesis of Theorem 2.1 holds for any $r \ge 2$.

Proof. Conditions (2.13) and (2.14) are satisfied because $q_\varepsilon^{(0)}$ and $q_\varepsilon^{(r)}$ are bounded functions, and $(g/f)^2$ is P-integrable by (2.23). The proof is completed by Lemma 2.3.

Remark. For δ close to 1, Corollary 2.2 shows that a rate of $O(n^{-\gamma'})$, with γ' arbitrarily close to 1/3, can be attained, provided $G|\theta|^m < \infty$ for sufficiently large m. On the other hand, for δ close to zero, lower convergence rates are attained. This last result is completely absent in the two-action problem (Cf. the remark following Corollary 1.2). We shall presently remedy the situation by obtaining better rates in the Normal two-action problem as a corollary of the estimation problem.

Let ψ_n be the estimate prescribed in Theorem 2.1. Consider the test $\phi_n' = [\psi_n - c \le 0]$ in the two-action problem in Theorem 1.1.

Lemma 2.4. $P\{(P_X\theta - c)(\phi_n' - \phi_G)\} \le P^{\frac{1}{2}}(P_X\theta - \psi_n)^2$. Consequently, the excess risk of ϕ_n' in the two-action problem is bounded by the square root of the excess risk of ψ_n in the estimation problem.

Proof. Since

$$(P_X\theta - c)(\phi_n' - \phi_G) = \begin{cases} P_X\theta - c & \text{if} \quad \psi_n \le c < P_X\theta \\ c - P_X\theta & \text{if} \quad P_X\theta \le c < \psi_n \ , \end{cases}$$

it follows that

(2.25) $P\{(P_X\theta - c)(\phi_n' - \phi_G)\} \le P|P_X\theta - \psi_n| \le P^{\frac{1}{2}}(P_X\theta - \psi_n)^2$,

where the second inequality follows by the Liapounov inequality. The proof is completed by (1.7) and (2.3).

We note that (2.25) is a statement about the excess risk in the two-action problem being bounded by the L_1-norm of $\psi - \psi_n$, which in turn is bounded by the L_2-norm of $\psi - \psi_n$.

Applying Lemma 2.4 to the Normal two-action problem, a rate of $O(n^{-\frac{1}{2}\gamma'})$ is possible provided G has finite $((1+t)\delta/(1-\delta))$-th moment. If we let m denote the number of finite moments of G and v the obtained rate, we have the parametric equation in δ

(2.26) $m = (1+t)\dfrac{\delta}{1-\delta}$, $v = q\dfrac{\delta}{2+\delta}$, where $q = \dfrac{r-1}{2r+1}$.

Similarly, we obtain the parametric equation

(2.27) $$m = 1 + (2+t)\frac{\delta}{2-\delta} \ , \ v = q\delta \ ,$$

from Corollary 1.2. The two parametric equations have a solution at $m = .3(1-t) + [.09(1-t)^2 + .80(1+t)]^{\frac{1}{2}}$. For $t = 0$, $m = m_0 = .3 + (.89)^{\frac{1}{2}}$. Therefore, for priors not having finite m_0-th moment, $\gamma < \frac{1}{2}\gamma'$. We have thus proved the following corollary.

<u>Corollary 2.3</u>. Consider the Normal two-action problem. Let ψ_n be as in Theorem 2.1, and $\phi_n' = [\psi_n - c \le 0]$. Then the excess risk in the two-action problem satisfies

$$0 \le R_n(\phi_n', G) - R^* = O(n^{-\gamma'/2}) \ ,$$

provided (2.24) is satisfied.

PART II

EXTENDED SEQUENCE-COMPOUND ESTIMATION

INTRODUCTION

Let $\underline{\theta} = (\theta_1,\ldots,\theta_n,\ldots)$ be a sequence of parameters. Let G_n denote the empirical distribution of $\theta_1,\theta_2,\ldots,\theta_n$. The usual standard in compound decision problems is $R(G_n)$, the Bayes envelope of the component problem evaluated at G_n.

Let $k \geq 1$. Let G_n^k denote the empirical distribution of the k-vectors $\underline{\theta}_k^k = (\theta_1,\ldots,\theta_k)$, $\underline{\theta}_{k+1}^k = (\theta_2,\ldots,\theta_{k+1}),\ldots,$ $\underline{\theta}_n^k = (\theta_{n-k+1},\ldots,\theta_n)$. Gilliland and Hannan (1969) considered the following extended game. Player I picks $\underline{\omega}_k = (\omega_1,\ldots,\omega_k) \in \Omega^k$ and Player II, after observing $\underline{X}_k \sim P_{\omega_1} \times\ldots\times P_{\omega_k}$, picks an action $a \in \mathcal{A}$ according to some randomized decision rule $\varphi(\underline{x}_k)$. With $L(\omega,a)$ denoting the loss, the risk Player II incurs is given by

$$R^k(\underline{\omega}_k,\varphi) = \iint L(\omega_k,a)\varphi(\underline{x}_k)(da)\ d(P_{\omega_1} \times\ldots\times P_{\omega_k}) .$$

The Bayes risk versus a p-measure G on Ω^k is

$$R^k(G,\varphi) = \int R^k(\cdot,\ \varphi)dG .$$

Swain (1965) used $R^k(G_n^k)$ as standards for compound problems, and called the resulting versions the extended compound decision problems. He considered squared error loss estimation problems in the discrete exponential and the Normal families and obtained rates of $O(n^{-\frac{1}{4}} \log^k n)$ and $o(1)$, respectively, uniformly in $\underline{\theta}$. Samuel (1965) and Gilliland (1966 and 1968) considered the unextended $(k = 1)$ versions of these same problems with Gilliland obtaining the improved rates of $O(n^{-\frac{1}{2}})$ and $O(n^{-1/5})$, respectively.

It is the purpose of this work to re-instate the k in Gilliland's results.

Chapter 3 considers the discrete exponential families. Lemma 3.2, a corollary of a theorem of Bikelis (1966), is used in (3.31) to bound certain probabilities involving k-dependent random variables. Without Lemma 3.2, the knowledge of a lower bound for the variances r_ℓ^2 in (3.28) seems to be necessary. Theorem 3.2, an improvement of Theorem 3.5 of Gilliland (1968), gives a rate of $O(n^{-\frac{1}{2}})$, uniformly in $\underline{\theta}$, for the estimates $\underline{\phi}^*$ that subsume those of Gilliland's.

Chapter 4 considers the Normal family. Here there is much in common with the estimation problem in the k-multivariate Normal considered by Susarla (1970). Most of the results in his §1.2 are applicable to our extended problem. Theorem 4.1 gives a rate of $O(n^{-1/(k+4)})$, uniformly in $\underline{\theta}$. We note that the rate deteriorates as k increases.

CHAPTER 3

ESTIMATION IN DISCRETE EXPONENTIAL FAMILIES UNDER SQUARED ERROR LOSS

3.1 Introduction.

We shall consider a sequence of statistical decision problems each of which is structurally identical to the component problem described below.

A component problem consists of a family of probability measures $\{P_\theta : \theta \in \Omega\}$ on a measurable space $(\mathcal{X},\mathcal{B})$, a measurable space $(\mathcal{A},\mathcal{C})$, and a loss function $0 \le L$ defined on $\Omega \times \mathcal{A}$. A randomized decision rule $\phi \in \Phi$ is a function defined on $\mathcal{X} \times \mathcal{C}$ such that for each $x \in \mathcal{X}$, $\phi(x,\cdot)$ is a probability measure on $\mathcal{C}$, and for each $C \in \mathcal{C}$, $\phi(\cdot,C)$ is $\mathcal{B}$-measurable. The risk of a procedure ϕ is defined by

$$R(\theta,\phi) = \iint L(\theta,A)\phi(x,dA)P_\theta(dx) \quad . \tag{3.1}$$

A sequence (non-Bayes) compound problem is one in which the decision rule ϕ_n for the n-th problem is allowed to depend on all past observations $\underline{x}_n = (x_1,x_2,\ldots,x_n)$ and the loss is taken to be the average of the component losses. We require that $\phi_n(\underline{x}_n,\cdot)$ be a probability measure on $\mathcal{C}$, for each $\underline{x}_n$; and that $\phi_n(\cdot,C)$ be $\mathcal{B}^n$-measurable, for each $C \in \mathcal{C}$.

Let $\underline{\phi} = (\phi_1, \phi_2, \ldots)$ be a procedure in a sequence-compound problem. The average risk of using $\underline{\phi}$ against $\underline{\theta}$ in the first n problems is given by

$$(3.2) \qquad R_n(\underline{\theta}, \underline{\phi}) = n^{-1} \sum_{i=1}^{n} \iint L(\theta_i, A)\phi_i(\underline{x}_i, dA)\underline{P}_i(d\underline{x}_i) ,$$

where $\underline{P}_i$ denotes the product measure $P_{\theta_1} \times P_{\theta_2} \times \ldots \times P_{\theta_i}$.

A compound procedure $\underline{\phi}$ is simple if $\phi_i(\cdot, C)$ is x_i-measurable for each $C \in \mathcal{C}$. If, in addition, all ϕ_i are identical, say $\phi_i = \phi$, it is simple symmetric. For every simple symmetric procedure $\underline{\phi}$ and any $\underline{\theta}$,

$$R_n(\underline{\theta}, \underline{\phi}) = n^{-1} \sum_{i=1}^{n} R(\theta_i, \phi) = \int R(\cdot, \phi)dG_n ,$$

where G_n denotes the empirical distribution of the first n θ's; i.e.,

$$(3.3) \qquad G_n \text{ puts mass } 1/n \text{ on each of } \theta_1, \theta_2, \ldots, \theta_n .$$

With $R(G, \phi)$ denoting $\int R(\cdot, \phi)dG$ and

$$(3.4) \qquad R(G) = \inf\{R(G, \phi) : \phi \in \Phi\}$$

denoting the Bayes risk versus the distribution G, it is obvious that for any simple symmetric procedure $\underline{\phi} = (\phi, \phi, \ldots)$

$$(3.5) \qquad R_n(\underline{\theta}, \underline{\phi}) = R(G_n, \phi) \geq R(G_n).$$

This motivates the use of the modified regret

$$(3.6) \qquad D_n(\underline{\theta}, \underline{\phi}) = R_n(\underline{\theta}, \underline{\phi}) - R(G_n)$$

as a measure of goodness for compound procedures.

Swain (1965) considered the following extended version of $R(G_n)$.

Let $k \geq 1$ be an integer. Let $\underline{\theta} \in \Omega^\infty$ and G_n^k be the k-th order empirical distribution of the first n θ's which puts equal mass $1/(n-k+1)$ on each of the k-vectors:

$$\underline{\theta}_k^k = (\theta_1, \theta_2, \ldots, \theta_k) ,$$

$$\underline{\theta}_{k+1}^k = (\theta_2, \theta_3, \ldots, \theta_{k+1}) ,$$

$$\ldots ,$$

$$\underline{\theta}_i^k = (\theta_{i-k+1}, \ldots, \theta_i) ,$$

$$\ldots ,$$

$$\underline{\theta}_n^k = (\theta_{n-k+1}, \ldots, \theta_n) .$$

Correspondingly, an extension of a simple symmetric procedure is a k-simple symmetric procedure $\underline{\phi}^k$ for which $\phi_i^k(\cdot, C)$ is $\underline{x}_i^k$-measurable for each $C \in \mathcal{C}$, $\phi_i^k(\underline{x}_i^k, \cdot)$ is a p-measure on $\mathcal{C}$ and all ϕ_i^k are identical to some ϕ^k. The risk of any k-simple symmetric procedure against $\underline{\theta} \in \Omega^\infty$ in the first n problems, not counting the first $k-1$, is given by

$$R_n(\underline{\theta}, \underline{\phi}^k) = (n-k+1)^{-1} \sum_{i=k}^{n} R^k(\underline{\theta}_i^k, \phi^k) \tag{3.7}$$

$$= R^k(G_n^k, \phi^k) ,$$

where

$$R^k(\underline{\theta}_i^k, \phi^k) = \iint L(\theta_i, A)\phi^k(\underline{x}_i^k, dA)\underline{P}_i^k(d\underline{x}_i^k) , \tag{3.8}$$

$$\underline{P}_i^k = \prod_{i-k+1}^{i} P_{\theta_j} \quad \text{and} \quad R^k(G_n^k, \phi^k) = \int R^k(\underline{\omega}^k, \phi^k) dG_n^k(\underline{\omega}^k).$$

It follows from (3.7) that for any k-simple symmetric procedure $\underset{\sim}{\phi}^k = (\phi^k, \phi^k, \ldots)$,

$$(3.9) \qquad R_n(\underline{\theta}, \underset{\sim}{\phi}^k) = R^k(G_n^k, \phi^k) \geq R^k(G_n^k) \ ,$$

where

$$(3.10) \qquad R^k(G_n^k) = \inf_{\phi^k} R^k(G_n^k, \phi^k) \ .$$

Swain (1965) used the k-th order Bayes envelopes $R^k(\cdot)$ in (3.10), or effectively

$$(3.11) \qquad D_n^k(\underline{\theta}, \underset{\sim}{\phi}) = R_n(\underline{\theta}, \underset{\sim}{\phi}) - R^k(G_n^k) \ ,$$

as standards in defining goodness of compound procedures $\underline{\phi}$, and called the resulting problem the extended compound decision problem.

Gilliland and Hannan (1969), in an improvement of a result of Swain, showed that for each $1 \leq k \leq n$ and $\underline{\theta}$,

$$(3.12) \qquad (n-k)\ R^{k+1}(G_n^{k+1}) \leq (n-k+1)\ R^k(G_n^k) \ .$$

In special cases, $\overline{\lim_{n\to\infty}} \{R^{k+1}(G_n^{k+1}) - R^k(G_n^k)\} < 0$, so that R^{k+1} is truly asymptotically more stringent than R^k.

Swain exhibited procedures, for the discrete exponential and the Normal families, that attained regret convergence of rates no worse than $O(n^{-\frac{1}{4}} \log^k n)$ and $o(1)$ respectively. Gilliland (1968) considered the $(k=1)$ unextended versions of these problems and was able to exhibit procedures that possessed regret convergence of rates no worse than $O(n^{-\frac{1}{2}})$ and $O(n^{-1/5})$ for the discrete exponential and the Normal families, respectively.

It is the purpose of the remainder of this thesis to reinstate the k in Gilliland's results and to show that the same improved rates of $O(n^{-\frac{1}{2}})$ and $O(n^{-1/(k+4)})$ hold. In the course of doing so, several of Gilliland's lemmas and theorems will be extended and, in some cases, strengthened.

3.2. A Bound for the Modified Regret D_n^k.

It is well known that under squared error loss, the posterior mean is Bayes. With respect to G_n^k, a version of the posterior mean of the k-th component of $\underline{\theta}_k$ is given by

$$(3.13) \qquad \psi_n^k(\underline{y}) = [\rho_n > 0] \sum_{j=k}^{n} \theta_j \, \pi_j / \rho_n$$

where $p_j = p_{\theta_j}$, $\pi_j = \prod_{\ell=1}^{k} p_{j-k+\ell}(y_\ell)$ and $\rho_n = \sum_{j=k}^{n} \pi_j$.

Under squared error loss, a non-randomized estimate $\underline{\phi}$ has a modified regret

$$D_n^k(\underline{\theta},\underline{\phi}) = (n-k+1)^{-1} \sum_{i=k}^{n} \underline{P}_i(\phi_i - \theta_i)^2 - R^k(G_n^k) ,$$

where $\underline{P}_i = \prod_1^i P_{\theta_j}$. Thus, by Theorem 2 of Gilliland and Hannan (1969) (i.e.,

$$\sum_{i=k}^{n} R^k(\underline{\theta}_i^k,\psi_i^k) \le (n-k+1)\, R^k(G_n^k) \le \sum_{i=k}^{n} R^k(\underline{\theta}_i^k,\psi_{i-1}^k) ,$$

where ψ_{k-1}^k is arbitrary), one can show that D_n^k is bounded above and below by

$$(3.14) \qquad (n-k+1)^{-1} \sum_{k}^{n} \underline{P}_i((\phi_i - \psi_i^k)(\phi_i + \psi_i^k - 2\theta_i))$$

and

$$(3.15)\quad (n-k+1)^{-1}\sum_k^n \underline{P}_i((\phi_i - \psi_i^k)(\phi_i + \psi_i^k - 2\theta_i))$$
$$+ (n-k+1)^{-1}\sum_k^n \underline{P}_i^k((\psi_i^k - \psi_{i-1}^k)(\psi_i^k + \psi_{i-1}^k - 2\theta_i)) ,$$

respectively, where the argument of ψ_i^k, ψ_{i-1}^k is $\underline{x}_i^k$.

If we assume $\Omega = \mathcal{A} = [-a,a]$, then the bounds (3.14) and (3.15) yield the following bound on the modified regret:

$$(3.16)\quad |D_n^k(\underline{\theta},\underline{\phi})| \le 4a\ (n-k+1)^{-1}\sum_k^n \underline{P}_i\{|\phi_i - \psi_i^k| + |\Delta_i|\} ,$$

where $\Delta_i = \psi_i^k - \psi_{i-1}^k$ for $i \ge k$. Let us show that following extended version of Theorem 2.1 of Gilliland (1968).

<u>Theorem 3.1</u>. Let $\Omega = [-a,a]$. For each P_θ, let p_θ be its Radon-Nikodym derivative with respect to some σ-finite measure μ. If $M = \sup\{p_\theta : \theta \in \Omega\}$ is μ-integrable, then

$$(3.17)\quad (n-k+1)^{-1}\sum_k^n \underline{P}_i^k|\Delta_i| = O(n^{-1}\log n) \text{ uniformly in } \underline{\theta} .$$

<u>Proof</u>. From the form of ψ_i^k in (3.13), it is easily verified by simple algebra that

$$|\Delta_i| \le 2a\,[\rho_{i-1} > 0]\pi_i/\rho_i + a[\rho_{i-1} = 0,\ \rho_i > 0]$$

from which

$$(3.18)\quad \sum_k^n \underline{P}_i^k\ |\Delta_i| \le 2a\int \sum_k^n [(\pi_i/\underline{M})^2/(\rho_i/\underline{M})]\underline{M}\ d\mu^k$$
$$+ a\int \sum_k^n [\rho_{i-1} = 0,\ \rho_i > 0](\pi_i/\underline{M})\underline{M}\ d\mu^k ,$$

where $\underline{M} = \prod_{\ell=1}^k M(y_\ell)$. The first term on rhs of (3.18), according to Lemma 3.1 below, is bounded by

$$2a \int \Big(\sum_{i=k}^{n} 1/i\Big)\underline{M}\, d\mu^k = 0(\log n) \int \underline{M}\, d\mu^k ,$$

and the second term is bounded by $a \int \underline{M}\, d\mu^k$. But since $\int \underline{M}\, d\mu^k = (\int M\, d\mu)^k < \infty$, the result follows.

We state without proof Lemma 2.1 of Gilliland (1968).

<u>Lemma 3.1</u>. For all $0 \le a_i \le 1$, $k \le i \le n$,

$$S_n = \sum_{i=k}^{n} a_i^2 \Big/ \sum_{j=k}^{i} a_j \le \sum_{i=k}^{n} 1/i .$$

Combining (3.16) and (3.17), we have

<u>Corollary 3.1</u>. If $\Omega = \mathcal{A} = [-a,a]$ and the hypothesis of Theorem 3.1 is satisfied, then

$$(3.19) \qquad |D_n^k(\underline{\theta},\underline{\phi})| \le 4a\ (n-k+1)^{-1} \sum_{k}^{n} \underline{P}_i|\phi_i - \psi_i^k| + 0(n^{-1} \log n)$$

uniformly in $\underline{\theta}$, for any compound procedure $\underline{\phi}$.

3.3 <u>Estimation in Discrete Exponential Families Under Squared Error Loss</u>.

Consider the family of probability measures on the non-negative integers having densities

$$(3.20) \qquad p_\theta(x) = \theta^x\, h(\theta)\, g(x) \ , \ x = 0,1,2,\ldots,$$

with respect to counting measure μ, where $g > 0$, and let

$$(A\ 1) \qquad \Omega = \mathcal{A} = [0,a] \ , \quad 0 < a < \infty .$$

For this family, the Bayes estimate in (3.13) takes the form

$$(3.21) \qquad \psi_i^k(\underline{y}) = [\rho_i > 0](g/\tilde{g})(\tilde{\rho}_i/\rho_i) \ ,$$

where for each $\underline{y} = (y_1,\ldots,y_k)$,

$$g = g(y_k) \ , \quad \tilde{g} = g(1 + y_k) \ ,$$

$$\tilde{\pi}_j = \prod_{\ell=1}^{k-1} p_{j-k+\ell}(y_\ell) p_j(1 + y_k) \quad \text{and} \quad \tilde{p}_i = \sum_{j=k}^{i} \tilde{\pi}_j \ .$$

In view of (3.21), when a sequence of past observations is available, a natural estimate for $\psi_i^k(\underline{X}_i^k)$ is

$$(3.22) \qquad \phi_i^*(\underline{X}_i^k) = \{[S > 0]((g/\tilde{g})(\tilde{S} + v_1)/(S + v_2))\} \wedge a \ , \ 2k \le i,$$

where $\tilde{f}(\underline{y}) = f(y_1,\ldots,y_{k-1},1 + y_k)$ for any f,

$$S = \sum_{k}^{i-k} \delta_j \ , \quad \delta_j = \delta_j(\underline{X}_i^k) = \begin{cases} 1 & \text{if } \underline{X}_j^k = \underline{X}_i^k \\ 0 & \text{if } \underline{X}_j^k \neq \underline{X}_i^k \end{cases} \ , \ \tilde{S} = \sum_{k}^{i-k} \tilde{\delta}_j \ ,$$

and $0 \le v_1, v_2 \le k$. We note that ϕ_i^* depends on the last k observations $\underline{X}_i^k$ taken as a k-vector, and is essentially a ratio between the number of times the k-vectors $\underline{X}_j^k$ equals $(X_{i-k+1},\ldots,X_{i-1},1+X_i)$ and the number of times $\underline{X}_j^k$ equals $\underline{X}_i^k$, except for the perturbations v_1 and v_2 in the numerator and denominator. It will be shown that these perturbations are negligible by comparing ϕ_i^* to the unattainable procedure

$$(3.23) \qquad \phi_i'(\underline{X}_i^k) = \{[S > 0](g/\tilde{g})(\tilde{S} + \tilde{S}')/(S + S')\} \wedge a \ ,$$

where ratios 0/0 are taken to be 0, $S' = \sum_{i-k+1}^{i} \delta_j'$, $\delta_j'(\underline{X}_i^k) = \delta((X_{j-k+1},\ldots,X_{i-k},X_{i-k+1}',\ldots,X_j'), \underline{X}_i^k)$ with X_j' independently distributed according to P_{θ_j} and independent of $\underline{X}_j$.

It will also be shown that ϕ_i' possesses a certain rate of the regret convergence. To be more specific, we will show that under suitable conditions, with E_i denoting the product measure on $(\underline{X}_i', \underline{X}_{i-k})$,

$$(n-k+1)^{-1}\sum_k^n \underline{P}_i^k E_i|\phi_i' - \psi_i^k| = O(n^{-\frac{1}{2}}) \quad \text{uniformly in } \underline{\theta}$$

(Proposition 3.1) and

$$(n-k+1)^{-1}\sum_k^n \underline{P}_i^k E_i|\phi_i^* - \phi_i'| = O(n^{-\frac{1}{2}}) \quad \text{uniformly in } \underline{\theta}$$

(Proposition 3.2), so that, by the triangle inequality,

$$(n-k+1)^{-1}\sum_k^n \underline{P}_i|\phi_i^* - \psi_i^k| = O(n^{-\frac{1}{2}}) \quad \text{uniformly in } \underline{\theta}$$

(Theorem 3.2).

<u>A Useful Result of Bikelis (1966)</u>.

Let Y_i, $i = 1,2,\ldots,n$ be a sequence of independent random variables that possess finite $2 + \delta$ $(0 < \delta \le 1)$ moments. Let $\bar{F}_n$ denote the distribution function of the normalized sum $S_n = \sum_{i=1}^n (Y_i - EY_i)/s_n$, where $s_n^2 = \sum_1^n \operatorname{Var} Y_i$. There exists a universal constant c such that $|\bar{F}_n(x) - \Phi(x)| \le c\, L_{2+\delta,n}/(1+|x|^{2+\delta})$, where $L_{2+\delta,n}$ is the Liapounov quotient $\sum_1^n E|Y_i - EY_i|^{2+\delta}/s_n^{2+\delta}$ and $\Phi(x) = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^x e^{-t^2/2}\,dt$.

The lemma below is an immediate corollary of the Bikelis theorem. We will use the lemma in bounding the error term in the Normal approximation.

<u>Lemma 3.2</u>. Let Y_i, $i = 1,\ldots,n$ be a sequence of independent bounded random variables with $|Y_i - EY_i| \le B < \infty$ for each i. Then

$$|\bar{F}_n(x/s_n) - \Phi(x/s_n)| \le c\, 2^{1+\delta}B^\delta/(s_n + |x|)^\delta \; . \tag{3.24}$$

<u>Proof</u>. By the Bikelis theorem, we have

$$|\bar{F}_n(x) - \Phi(x)| \le c\, L_{2+\delta,n}/(1 + |x|^{2+\delta}) \; ,$$

where $L_{2+\delta,n} \le B^\delta/s_n^\delta$ and $1 + |x|^{2+\delta} \ge (1+|x|)^{2+\delta}/2^{1+\delta}$, by the C_r-inequality. Hence,

$$|\bar{F}_n(x/s_n) - \Phi(x/s_n)| \le c\ 2^{1+\delta}\ B^\delta\ s_n^2/(s_n + |x|)^{2+\delta}$$

$$\le c\ 2^{1+\delta}\ B^\delta/(s_n + |x|)^\delta.$$

The proof is completed.

Henceforth until (3.34), we will let $\underline{X}_i^k = \underline{x}_i^k$ be fixed and abbreviate $\psi_i^k(\underline{x}_i^k)$ and $\phi_i'(\underline{x}_i^k)$ to ψ and ϕ', respectively.

Let E abbreviate E_i. Since $0 \le \phi', \psi \le a$ by (A1), it follows that

$$E|\phi' - \psi| = \int_0^a E[\phi' - \psi \ge u]du + \int_{-\psi}^0 E[\phi' - \psi \le u]du \ . \tag{3.25}$$

We shall next place bounds on the two integrands by the use of Lemma 3.2.

For each i and $|u| \le a$, put

$$q = (\tilde{g}/g)(\psi + u) \ ,$$

$$Y_j = \begin{cases} \tilde{\delta}_j - q\ \delta_j, & \text{for } k \le j \le i-k, \\ \tilde{\delta}_j' - q\ \delta_j', & \text{for } i-k < j \end{cases} \tag{3.26}$$

$$w = (\tilde{g}/g)\ \rho_i/k \ .$$

Since $\sum_k^i EY_j = \sum_k^i (\tilde{\pi}_j - q\ \pi_j) = -k\ w\ u$,

$$[\phi' - \psi \ge u] \le [\sum_k^i Y_j \ge 0]$$

$$= [\sum_k^i (Y_j - EY_j) \ge k\ w\ u] \ , \tag{3.27}$$

$$\le \Sigma'[\Sigma''(Y_{\ell+dk} - EY_{\ell+dk}) \ge w\ u]$$

where Σ' denotes summation over ℓ from 1 to k and Σ'' denotes summation over d for which $k \le \ell + dk \le i$. For each $1 \le \ell \le k$ and $i \ge k$, we let

$$r_\ell^2 = \Sigma'' \operatorname{Var} Y_{\ell+dk}$$

(3.28)
$$r^2 = \sum_k^i \operatorname{Var} Y_j \quad .$$

Then, with $c' = 2^{1+\delta}c$, it follows from (3.27), (3.24) and the fact that $|Y_j - EY_j| \le 1+q$,

(3.29) $$E[\phi' - \psi \ge u] \le \Sigma'\{\Phi(-w\ u/r_\ell) + c'(1+q)^\delta/(w\ u)^\delta\} ,$$

where $1+q$ playes the role of B in Lemma 3.2. We shall next bound the terms on the rhs of (3.29) by a quantity not involving the index ℓ. Let $Q = 2a\ \tilde{g}/g$ and $T^2 = Q(1+Q)$, then, for each $|u| \le a$, (A1) yields

(3.30a) $$q \le Q ,$$

(3.30b) $$\tilde{\pi}_j \le \tfrac{1}{2} Q\ \pi_j ,$$

(3.30c) $$r_\ell^2 \le r^2 \le \sum_k^i (\tilde{\pi}_j + q^2 \pi_j) \le T^2\ \rho_i \ .$$

Thus, (3.29) yields

(3.31) $$E[\phi' - \psi \ge u] \le k\{\Phi(-w\ u/(T\ \rho_i^{\frac{1}{2}})) + c'(1+Q)^\delta/(w\ u)^\delta\} .$$

Upon setting $\delta = \frac{1}{2}$ and integrating u between 0 and a, we obtain,,via the inequality $\int_0^a \Phi(-bt)dt \le b^{-1}(2\pi)^{-\frac{1}{2}}$,

(3.32) $$\int_0^a E[\phi' - \psi \ge u]du \le B_1\{T\rho_i^{\frac{1}{2}}/w + (1+Q)^{\frac{1}{2}}/w^{\frac{1}{2}}\} ,$$

where B_1 is a constant independent of i, $\underline{x}_i^k$ and $\underline{\theta}$. We shall next bound the second term in the rhs of (3.25).

Since $[\phi' - \psi \le u] \le [\sum_k^i - Y_j \ge 0]$, the arguments in (3.27) through (3.29) hold with Y_j replaced by $-Y_j$; consequently, by (3.30) and the arguments leading to (3.32), it follows that

$$(3.33) \qquad \int_{-\psi_i}^{a} E[\phi_i' - \psi_i \le u]du \le B_2\{T\rho_i^{\frac{1}{2}}/w + (1+Q)^{\frac{1}{2}}/w^{\frac{1}{2}}\} ,$$

where B_2 is a constant independent of i, $\underline{x}_i^k$ and $\underline{\theta}$.

Combining (3.25), (3.32), (3.33) and (3.30), we have

$$(3.34) \qquad E|\phi_i' - \psi_i| \le B_3\{T\,\rho_i^{\frac{1}{2}}/w + (1+Q)^{\frac{1}{2}}/w^{\frac{1}{2}}\}$$

$$\le B_4[(1 + g/\tilde{g})/\rho_i]^{\frac{1}{2}} .$$

Before we prove the next proposition, we quote Lemma 3.1 of Gilliland (1968), i.e.,

$$(3.35) \qquad \sum_k^n a_i(\sum_k^i a_j)^{-\frac{1}{2}} \le 2(\sum_k^n a_i)^{\frac{1}{2}}, \text{ for all } a_i \ge 0,\ k \le i \le n .$$

Let p_a denote $p_{\theta=a}$.

<u>Proposition 3.1</u>. Under the assumptions

$$(A1) \qquad \mathcal{A} = \Omega = [0,a] , \quad 0 < a < \infty ,$$

$$(A2) \qquad \sum_x p_a^{\frac{1}{2}} < \infty$$

and

$$(A3) \qquad \sum_x [(g/\tilde{g})p_a]^{\frac{1}{2}} < \infty ,$$

$$(3.36) \quad (n-k+1)^{-1} \sum_k^n \underline{P}_i^k E|\phi_i' - \psi_i^k| = O(n^{-\frac{1}{2}}) \text{ uniformly in } \underline{\theta} .$$

<u>Proof</u>. Let $M = \sup\{p_\theta : \theta \in \Omega\}$. Since $h(\theta)$ is a decreasing function, it follows from (A.1) that

(3.37) $\qquad M \le p_a(x)\, h(0)/h(a)$ and M is μ-integrable.

From (3.34) and (3.35), we have

$$\sum_{k}^{n} \underline{P}_i^k E|\phi_i' - \psi_i^k| \le 2B_4 \sum_{\underline{y}_k} [(1 + g/\tilde{g})\rho_n]^{\frac{1}{2}} \tag{3.38}$$
$$\le 2B_4 \sum_{\underline{y}_k} [(1 + g/\tilde{g})(\rho_n/\underline{M})]^{\frac{1}{2}}\, \underline{M}^{\frac{1}{2}} ,$$

where $\underline{M} = \prod_{\ell=1}^{k} M(y_\ell)$ is bounded by $\prod_{\ell=1}^{k} p_a(y_\ell)[h(0)/h(a)]^k$ via (3.37), and $\rho_n/\underline{M}$ is bounded trivially by n-k+1. Hence,

$$\sum_{k}^{n} \underline{P}_i^k E|\phi_i' - \psi_i^k| \le B_5(n-k+1)^{\frac{1}{2}} \sum_{\underline{y}_k} (1 + g/\tilde{g})^{\frac{1}{2}}[\prod_1^k p_a(y_\ell)]^{\frac{1}{2}}$$
$$= B_5(n-k+1)^{\frac{1}{2}}[\sum_x p_a^{\frac{1}{2}}(x)]^{k-1} \sum_x [(1 + g/\tilde{g})p_a]^{\frac{1}{2}} .$$

The result follows from (A2) and (A3).

Remark. Proposition 3.2 of Gilliland (1968) proved (3.36) under the stronger assumption $(A1^+)$, (A2) and (A3) with k=1. The Bikelis bound on the error term in the Normal approximation enabled us to weaken the assumption $(A1^+)$ to (A1). Gilliland used the Berry-Esseen bound to prove his Proposition 3.2.

Bounds for $D_n^k(\underline{\theta},\underline{\phi}^*)$.

Proposition 3.1 shows that $\underline{\phi}'$ and $\underline{\psi}^k$ are not more than $O(n^{-\frac{1}{2}})$ apart in a Cesaro sense. In view of (3.19), it remains to be shown that $\underline{\phi}^*$ and $\underline{\phi}'$ are close in order to show that $\underline{\phi}^*$ and $\underline{\psi}^k$ are not far apart.

Henceforth until (3.41), let $\underline{X}_i^k = \underline{x}_i^k$ be fixed and abbreviate $\phi_i^*(\underline{x}_i^k)$ and $\phi_i'(\underline{x}_i^k)$ by ϕ^* and ϕ', respectively.

Lemma 3.3. $|\phi^* - \phi'| \le k(a + g/\tilde{g})[S > 0]/S$.

Proof. Let $I = [(g/\tilde{g})(\tilde{S}/(k+S)) < a]$. On $[S > 0]I$, $|\phi^* - \phi'| \le (g/\tilde{g})|(\tilde{S} + k)/S - \tilde{S}/(k + S)| \le k(a + g/\tilde{g})/S$. Since $|\phi^* - \phi'| = 0$ on $\{[S > 0]I\}^c$, the result follows.

Remark. Lemma 3.3 is an analogue of (3.28) of Gilliland (1968). The truncation of ϕ' in (3.23) results in the better bound in Lemma 3.3.

Lemma 3.4.

$$\underline{P}_{i-k}([S > 0]/S) < (k+2)/\rho_i \ . \tag{3.39}$$

Proof. If $S > 0$, then the inequality $S+k+1 \le S(k+2)$ implies $[S > 0]/S \le (k+2)/(S+k+1) \le (k+2)/(S + \sum_{i-k+1}^{i} \delta_j' + 1)$. By the convexity of $1/(1+z)$, Hoeffding's Theorem 3 (1956) applies to yield

$$\underline{P}_{i-k}(S + \sum_{i-k+1}^{i} \delta_j' + 1)^{-1} \le \sum_{j=0}^{i} \binom{i}{j}\rho^{i}(1 - \rho)^{i-j}/(1+j) \ ,$$

where $\rho = \rho_i/(i-k+1)$. The rhs of the last inequality is bounded by

$$\sum_{j=0}^{i} \binom{1+i}{1+j}\rho^{1+j}(1 - \rho)^{i-j}/((1+i)\rho)$$

$$\le (1 - (1-\rho)^{1+i})/((1+i)\rho) \le 1/((1+i)\rho) \ .$$

Since $(1+i)\rho = (1+i)\rho_i/(i-k+1) > \rho_i$, the result follows.

Lemma 3.4 with k specialized to 1 improves upon Lemma 3.3 of Gilliland (1968).

The next lemma is suggested by the proof of Lemma 3.5 of Gilliland (1968).

Lemma 3.5. Under (A1),

$$(3.40)\qquad \sum_{k}^{n} \pi_i \, \underline{P}_{i-k}([S > 0]/S) < b(n-k+1)^{\frac{1}{2}}\Big(\prod_{\ell=1}^{k} p_a(x_{i-k+\ell})\Big)^{\frac{1}{2}}$$

where $b = 2(k+2)^{\frac{1}{2}}(h(0)/h(a))^{k/2}$.

Proof. Since $[S > 0]/S \le 1$, $[S > 0]/S \le ([S > 0]/S)^{\frac{1}{2}}$. Consequently, Jensen's inequality applies to give

$$\underline{P}_{i-k}([S > 0]/S) \le (\underline{P}_{i-k}[S > 0]/S)^{\frac{1}{2}} < ((k+2)/\rho_i)^{\frac{1}{2}}$$

where the last inequality follows from (3.39). Thus, by (3.35),

$$(3.41)\qquad \sum_{k}^{n} \pi_i \, \underline{P}_{i-k}([S > 0]/S) < 2(k+2)^{\frac{1}{2}} \rho_n^{\frac{1}{2}} .$$

Under (A1), (3.37) holds. Hence it follows from (3.41) that

$$\begin{aligned}\sum_{k}^{n} \pi_i \underline{P}_{i-k}([S>0]/S) &< 2(k+2)^{\frac{1}{2}}(h(0)/h(a))^{k/2}(\rho_n/\underline{M})^{\frac{1}{2}} \Big(\prod_{\ell=1}^{k} p_a(x_{i-k+\ell})\Big)^{\frac{1}{2}} \\ &\le b(n-k+1)^{\frac{1}{2}}\Big(\prod_{\ell=1}^{k} p_a(x_{i-k+\ell})\Big)^{\frac{1}{2}} .\end{aligned}$$

The proof if completed.

Proposition 3.2. If the family of distributions satisfies the assumptions

(A1) $\qquad \Omega = \mathcal{A} = [0,a]\ ,\quad 0 < a < \infty\ ,$

(A2) $\qquad \sum_{x} p_a^{\frac{1}{2}} < \infty\ ,$

and

(A3') $\qquad \sum_{x} (g/\tilde{g})\, p_a^{\frac{1}{2}} < \infty\ ,$

then

(3.42) $\quad (n-k+1)^{-1}\sum_{k}^{n}\underline{P}_i^k E|\phi_i' - \phi_i^*| = O(n^{-\frac{1}{2}})$ uniformly in $\underline{\theta}$.

Proof. From Lemma 3.3,

(3.43) $$\sum_{k}^{n}\underline{P}_i^k E|\phi_i' - \phi_i^*| \le k\sum_{k}^{n}\underline{P}_i((a + g/\tilde{g})[S > 0]/S) .$$

Via the equality $\underline{P}_i((a + g/\tilde{g})[S > 0]/S) = \sum_{\underline{y}_k}(a + g/\tilde{g})\pi_i\underline{P}_{i-k}([S > 0]/S)$, (3.43) and (3.40) yield

(3.44) $$\sum_{k}^{n}\underline{P}_i^k E|\phi_i' - \phi_i^*| < b(n-k+1)^{\frac{1}{2}}\sum_{\underline{y}_k}(a + g/\tilde{g})\Big(\prod_{\ell=1}^{k} p_a(y_\ell)\Big)^{\frac{1}{2}} .$$

Since

$$\sum_{\underline{y}_k}(a + g/\tilde{g})\Big(\prod_{1}^{k} p_a\Big)^{\frac{1}{2}} = \Big(\sum_{x} p_a\Big)^{(k-1)/2}\sum_{x}(a + g/\tilde{g})p_a^{\frac{1}{2}} ,$$

the proof is completed by (A2) and (A3').

Theorem 3.2. Under (A1), (A2) and (A3'),

(3.45) $$|D_n^k(\underline{\theta}, \underline{\phi}^*)| = O(n^{-\frac{1}{2}}) \quad \text{uniformly in } \underline{\theta} .$$

Proof. Under (A1), (A2), (A3') and (A3), Corollary 3.1 together with (3.36) and (3.42) implies (3.45). Since (A2) and (A3') imply (A3) via the Cauchy-Schwarz inequality

$$\sum((g/\tilde{g})p_a)^{\frac{1}{2}} \le \Big(\sum(g/\tilde{g})p_a^{\frac{1}{2}}\Big)^{\frac{1}{2}}\Big(\sum p_a^{\frac{1}{2}}\Big)^{\frac{1}{2}} ,$$

the result follows.

Remark. Theorem 3.5 of Gilliland (1968) proved (3.45) under the stronger assumption ($A2^+$) together with (A1) and (A3'). The procedure $\underline{\phi}^*$ in (3.45) extends and includes that of $\underline{\phi}^*$ and $\underline{\phi}^{**}$ in Gilliland. For examples of distribution satisfying these assumptions see Gilliland (1968).

CHAPTER 4

SQUARED ERROR LOSS ESTIMATION IN THE NORMAL FAMILY

4.1 Introduction.

Consider the Normal $(\theta,1)$ family

$$(4.1) \qquad p_\theta(x) = (2\pi)^{-\frac{1}{2}} e^{-(x-\theta)^2/2}, \quad -\infty < x < \infty$$

with $|\theta| \leq a$. The Bayes estimate in (3.13) takes the form

$$(4.2) \qquad \psi_n^k(\underline{y}) = y + u_n, \text{ for each } \underline{y} = (y_1,\ldots,y_k) \in R^k,$$

where $y = y_k$, $u_n = \frac{\partial}{\partial y} \log(\sum_k^n \pi_j)$. In view of (3.19), let us consider estimating ψ_n^k. The method of estimation is contained in §1.2 of Susarla (1970).

Let $\underline{P}_j^k$ denote the product measure on $\underline{X}_j^k$ and $\bar{Q} = (n-k+1)^{-1} \sum_k^n \underline{P}_j^k$. For each $\underline{x} = \underline{x}_k$ in R^k, let

$$\square = \mathop{\times}_{\ell=1}^{k} I_\ell, \text{ where } I_\ell = [x_\ell, x_\ell + \epsilon] \text{ for } \ell = 1,\ldots,k,$$

and

$$\square_k = \mathop{\times}_{\ell=1}^{k} I'_\ell, \text{ where } I'_\ell = I_\ell \text{ for } \ell \neq k \text{ and } I'_k = [x_k+\eta, x_k+\eta+\epsilon].$$

and

$$\square'_k = \mathop{\times}_{\ell=1}^{k} I''_\ell, \text{ where } I''_\ell = I_\ell \text{ for } \ell \neq k \text{ and } I''_k = [x_k, x_k+\eta+\epsilon].$$

For any distribution F on R^k let $t(F)(\underline{x})$ denote the function $\eta^{-1} \log(F\,\square_k/F\,\square)$ where $F\,\square$ and $F\,\square_k$ represent the measures of $\square$ and $\square_k$ under F and undefined ratios are taken to be 1. We abbreviate $t(\bar{Q})(\underline{x}) = \eta^{-1}\log(\bar{Q}\,\square_k/\bar{Q}\,\square)$ by $t(\underline{x})$.

Let Q^* be the k-order empiric distribution of $X_1,\ldots,X_n$ and abbreviate $t(Q^*)(\underline{x}) = \eta^{-1}\log(Q^*\,\square_k/Q^*\square)$ by $t^*(\underline{x})$.

Let X abbreviate X_{n+k}, $\underline{X}$ abbreviate $\underline{X}^k_{n+k}$ and

$$(4.3)\qquad \psi^*_{n+k} = \mathrm{tr}(X + t^*(\underline{X}))\ ,\ \psi^{**}_{n+k} = \mathrm{tr}'(X + t^*(\underline{X}))$$

where tr and tr' stand for retraction to the intervals $[-(a+\eta+\varepsilon),\ a + \eta + \varepsilon]$ and $[-a,a]$ respectively.

With ψ abbreviating ψ^k_n and suppressing the subscripts in ψ^*_{n+k} and ψ^{**}_{n+k}, we have $|\psi| \le a$, $\psi^{**} = \mathrm{tr}'\,\psi^*$ and therefore $|\psi^{**} - \psi| \le |\psi^* - \psi|$. Consequently, by the triangle inequality,

$$(4.4)\qquad \underline{P}_{n+k}|\psi^{**} - \psi| \le \underline{P}_{n+k}|\psi^* - (X + t)| + \underline{P}^k_{n+k}|X + t - \psi|\ .$$

We state without proof Lemma 3 of Susarla with $\sigma^2 = 1$, and $\bar{F} = \bar{Q}$.

<u>Lemma 4.1</u> (Susarla). For each $\underline{x}$ in R^k

(1) $x + t(\bar{Q})(\underline{x}) \in [-a - \frac{\eta}{2} - \varepsilon,\ a]$

(2) $\bar{Q}\,\square_k \ge \bar{\pi}\ \varepsilon^k \exp\{-\ (\eta+\varepsilon)(\|\underline{x}\| + \frac{k\varepsilon + \eta}{2})\}$,

(3) $\bar{Q}\,\square'_k \le \bar{Q}\,\square_k\ \frac{\eta+\varepsilon}{\varepsilon}\ \exp\{(\eta + \varepsilon)(|x\ | + a + \eta + \varepsilon)\}$

where $x = x_k$, $\bar{\pi} = \sum_k^n \pi_j/(n-k+1)$ and $\|\underline{x}\| = \sum_{\ell=1}^k |x_\ell|$.

4.2 Bounding $\underline{P}_{n+k}|\psi^* - (X + t)|$

Fix $\underline{X} = \underline{x}$ until (4.10). Since $x + t$, by (1) of Lemma 4.1, is in $[-a - \frac{\eta}{2} - \varepsilon, a]$ it follows from the definition of ψ^* that $|\psi^* - (X + t)|$ is bounded by the quantity $a' = 2a + \frac{3}{2}\eta + 2\varepsilon$, and at the same time bounded by $|t^* - t|$. Therefore, for each $\underline{x}$ in R^k,

$$(4.5) \qquad \underline{P}_n|\psi^* - (x+t)| \le \int_0^{a'} A\, du + \int_{-a'}^{0} B\, du$$

where $A = \underline{P}_n[t^* - t > u]$ and $B = \underline{P}_n[t^* - t < u]$. We shall first bound A and B by the Bikelis theorem.

Put

$$\tilde{\delta}_i = [\underline{X}_i^k \in \square_k] \ , \quad \delta_i = [\underline{X}_i^k \in \square] \ ,$$

$$(4.6) \qquad Y_i(u) = \tilde{\delta}_i - \delta_i\, e^{\eta(t+u)} \ , \quad \text{for } |u| \le a' \ , \ k \le i \ .$$

$$r^2 = \sum_k^n \mathrm{Var}\, Y_i$$

Let $w = (n-k+1)\, \bar{Q}\, \square_k\, \eta/k$, $R = e^{\eta(|x|+a+a')}$, Σ denote summation over i from k to n, Σ' denote summation over ℓ from 1 to k and, Σ'' denote summation over d for which $k \le \ell + dk \le n$, for each ℓ.

Lemma 4.2. For some constant c_1,

$$A \le k\, \Phi(-wu/r) + c_1\, R^{\frac{1}{2}}/|wu|^{\frac{1}{2}} \ , \ 0 \le u \ ,$$

and, for $\eta\, a' \le 1$,

$$B \le k\, \Phi(wu/2r) + c_1\, R^{\frac{1}{2}}/|wu/2|^{\frac{1}{2}} \ , \quad -a' \le u \le 0 \ .$$

<u>Proof</u>. Note that $A = \underline{P}_n[\Sigma Y_i \geq 0] \leq \Sigma' \underline{P}_n[\Sigma''(Y_{\ell+dk} - \underline{P}_n Y_{\ell+dk}) \geq -\Sigma \underline{P}_n Y_i/k]$ and, similarly, $B = \underline{P}_n[\Sigma(-Y_i) \geq 0] \leq \Sigma' \underline{P}_n[\Sigma'' - (Y_{\ell+dk} - \underline{P}_n Y_{\ell+dk}) \geq \Sigma \underline{P}_n Y_i/k]$. By (4.6),

$$\Sigma \underline{P}_n Y_i = (n-k+1)\, \bar{Q}\, \Box_k (1 - e^{\eta u}) . \tag{4.7}$$

For $0 \leq u$, $1 - e^{\eta u} \leq -\eta u$. For $-a' \leq u \leq 0$, $\eta a' \leq 1$ implies $1 - e^{\eta u} > -\frac{1}{2}\eta u$. Thus, by (4.7),

$$\begin{aligned} A &\leq \Sigma' \underline{P}_n[\Sigma''(Y_{\ell+dk} - \underline{P}_n Y_{\ell+dk}) \geq wu] , \quad 0 \leq u \\ B &\leq \Sigma' \underline{P}_n[\Sigma'' - (Y_{\ell+dk} - \underline{P}_n Y_{\ell+dk}) \geq -wu/2], \quad -a' \leq u \leq 0 . \end{aligned} \tag{4.8}$$

Since $|Y_i - EY_i| \leq 2R$, we have, by Lemma 3.2

$$\begin{aligned} A &\leq \Sigma'\{\Phi(-wu/r_\ell) + c\, R^{\frac{1}{2}}/|wu|^{\frac{1}{2}}\} , \quad 0 \leq u \\ B &\leq \Sigma'\{\Phi(wu/2r_\ell) + c\, R^{\frac{1}{2}}/|wu/2|^{\frac{1}{2}}\} , \quad -a' \leq u \leq 0 \end{aligned} \tag{4.9}$$

where $r_\ell^2 = \mathrm{Var}\, \Sigma'' Y_{\ell+dk}$. The proof is completed by the bound $r_\ell^2 \leq \Sigma \,\mathrm{Var}\, Y_i = r^2$.

We note that

$$r^2 \leq \Sigma \underline{P}_n Y_i^2 \leq (n-k+1)\, R^2\, \bar{Q}\, \Box_k' . \tag{4.10}$$

With (4.10), we prove an analogue of Lemma 4 of Susarla.

<u>Lemma 4.3</u>. For $0 < \epsilon \leq \eta \leq 1/(6 + 2a)$,

$$\underline{P}_{n+k}|\psi^* - (X + t)| \leq B_1 (n- k+1)^{-\frac{1}{2}}\{(\frac{\eta + \epsilon}{\eta^2 \epsilon\, k+1})^{\frac{1}{2}} + (\frac{1}{\eta\epsilon k})^{\frac{1}{2}}\} , \tag{4.11}$$

where B_1 is independent of n and $\underline{\theta}$.

Proof. Since $\int_0^\infty \Phi(-bt)dt \le (2\pi)^{-\frac{1}{2}}/b$, for $b > 0$, it follows from (4.5) and Lemma 4.2 that, for $\eta a' \le 1$,

$$(4.12) \qquad \underline{P}_n|\psi^* - (x+t)| \le c_2 \frac{r}{w} + c_3 \frac{R^{\frac{1}{2}}}{w^{\frac{1}{2}}} .$$

By (4.10) and the definitions of w and R, the above inequality yields

$$(4.13) \quad \underline{P}_n|\psi^* - (x+t)| \le B_2(n-k+1)^{-\frac{1}{2}}\{(\frac{\eta + \epsilon}{\eta^2 \epsilon^{k+1}})^{\frac{1}{2}} C^{\frac{1}{2}}D^{\frac{1}{2}}R + (\frac{1}{\eta\epsilon^k})^{\frac{1}{2}}D^{\frac{1}{2}}R^{\frac{1}{2}}\},$$

where $C = \frac{\epsilon}{\eta+\epsilon}\frac{\bar{Q}\,\Box_k'}{\bar{Q}\,\Box_k}$ and $D = \frac{\epsilon^k}{\bar{Q}\,\Box_k}$. By (2) and (3) of Lemma 4.1, $C \le \exp\{(\eta+\epsilon)(|x| + a + \eta + \epsilon)\}$ and $D \le (\bar{\pi})^{-1}\exp(\eta+\epsilon)(\|\underline{x}\| + \frac{k\epsilon+\eta}{2})$. Hence, it follows from (4.13), the definition of R and $0 \le \epsilon \le \eta \le 1/(6 + 2a)$ that

$$(4.14) \quad \underline{P}_n|\psi^* - (x + t)| \le B_3(n-k+1)^{-\frac{1}{2}}\{(\frac{\eta + \epsilon}{\eta^2 \epsilon^{k+1}})^{\frac{1}{2}} + (\frac{1}{\eta\epsilon^k})^{\frac{1}{2}}\} \times$$

$$\times \exp\{(2|x| + \|\underline{x}\|)\eta\}(\bar{\pi})^{-\frac{1}{2}} .$$

To complete the proof, we shall show that the $\underline{P}^k_{n+k}$-integral of the function $g = \exp\{(2|x| + \|\underline{x}\|)\}(\bar{\pi})^{-\frac{1}{2}}$ is uniformly bounded in n. Let $c = (2\pi)$. Since $c^{\frac{1}{2}}p_\theta(y) \le \exp\{-[(|y| - a)^+]^2/2\}$ and $c^{\frac{1}{2}}p_\theta(y) \ge \exp\{-[|y| + a]^2/2\}$, we have $(\bar{\pi})^{-\frac{1}{2}} \le$ $c^{k/4}\exp[\Sigma'(|x_\ell| + a)^2/4]$, and $\pi_{n+k} \le c^{-k/2}\exp[-\Sigma'[(|x_\ell| - a)^+]^2/2\}$. Consequently, the $\underline{P}^k_{n+k}$-integral of g is exceeded by the constant

$$\int c^{-k/4}\exp\{(2|x| + \|\underline{x}\|) + \Sigma'(|x_\ell| + a)^2/4 - \Sigma'[(|x_\ell| - a)^+]^2/2\}d\underline{x} .$$

The proof is completed.

We state without proof a special case of Lemma 6 of Susarla.

Lemma 4.4. $|x + t - \psi| \le \eta(1 + a^2) + \epsilon(1 + ka^2)$.

The next lemma, suggested by Professor Gilliland, is an analogue of Theorem 3.1.

Lemma 4.5. Consider the Normal $(\theta,1)$ family in (4.1). For any $1 \le b$, $b + k \le n$

$$\underline{P}_n^k|\psi_n^k - \psi_{n-b}^k| = O(n^{-1}) \tag{4.15}$$

uniformly in $\underline{\theta}$.

Proof. Let $1 \le k$, $k \le n-b$. Since for each fixed $\underline{x}_n^k$

$$|\psi_n^k - \psi_{n-b}^k| \le 2a \frac{\sum_{n-b+1}^{n} \pi_j}{\sum_k^n \pi_j}$$

and, by Jensen's inequality, $1/\sum_k^n \pi_j \le (n-k+1)^{-2} \sum_k^n \pi_j^{-1}$, we have $|\psi_n^k - \psi_{n-b}^k| \le 2a(n-k+1)^{-2} \sum_{n-b+1}^{n} \pi_j \sum_k^n \pi_i^{-1}$. But for any $\underline{x} \in R^k$, $\pi_j \pi_i^{-1}(\underline{x}) \le e^{2a\|\underline{x}\|}$, therefore,

$$\underline{P}_n|\psi_n^k(\underline{x}) - \psi_{n-b}^k(\underline{x})| \le 2ab(n-k+1)^{-1} \int e^{2a\|\underline{x}\|} \pi_n \, d\underline{x} \, .$$

By the monotone likelihood ratio property of the Normals, $P_\theta e^{2a|x|} \le 2e^{a^2/2} \int_0^\infty e^{2ax} p_a(x)dx = c(a)$ is a finite constant. Consequently, $\int e^{2a\|\underline{x}\|} \pi_n \, d\underline{x} \le c^k(a)$, uniformly in n; therefore, the result follows.

With Lemma 4.5, it follows from (3.19), via the triangle inequality, that for the Normal family in (4.1)

(4.16) $$|D_n^k(\underset{\sim}{\theta},\underset{\sim}{\phi})| \le 4a(n-k+1)^{-1} \sum_k^n \underline{P}_i|\phi_i - \psi_{i-k}^k| + O(n^{-1}\log n),$$

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

Theorem 4.1. With $\epsilon = n^{-1/(k+4)}$ and $\eta = b\epsilon$ for $1 < b$, then

(4.17) $$\underline{P}_{n+k}|\psi^{**} - \psi_n^k| = O(n^{-1/(k+4)})$$

and

(4.18) $$D_n^k(\underset{\sim}{\theta},\underset{\sim}{\psi}^{**}) = O(n^{-1/(k+4)}).$$

Proof. Lemmas 4.3 and 4.4 imply (4.17), via (4.4). The result follows from (4.16) and (4.17).

REFERENCES

REFERENCES

Beckenbach, E. and Bellman, R. (1961). An Introduction to Inequalities. Random House.

Bikelis, A. (1966). Estimates of remainder term in the central limit theorem. Litovsk. Mat. Sb. **6**, 323-346.

Feller, William (1962). An Introduction to Probability Theory and Its Applications, Vol. 1, 2nd ed. John Wiley & Sons.

Ferguson, T.S. (1967). Mathematical Statistics. Academic Press.

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.

Gilliland, D.C. and Hannan, J.F. (1969). On an extended compound decision problem. Ann. Math. Statist. **40**, 1536-1541.

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.

Hewitt, E. and Stromberg, K. (1965). Real and Abstract Analysis. Springer-Verlag New York.

Hoeffding, Wassily (1956). On the distribution of the number of successes in independent trials. Ann. Math. Statist. **27**, 713-731.

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.

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

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

Royden, H.L. (1963). Real Analysis. Macmillan.

Samuel, E. (1965). Sequential compound estimators. Ann. Math. Statist. **36** 879-889.

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

Swain, Donald D. (1965). Bounds and rates of convergence for the extended compound estimation problem in the sequence case. Tech. Report No. 81, Department of Statistics, Stanford.

APPENDIX

APPENDIX

SOME KERNEL ESTIMATES OF DENSITIES AND THEIR DERIVATIVES

Estimation of Lebesgue density f and its derivative $g = f^{(1)}$ will be discussed in Section 1. Estimation of a density J with respect to $d\mu = h\,dx$ and its derivative $J^{(1)}$ will be discussed in Section 2. Estimates for the above quantities are based on the kernel method that Johns and Van Ryzin (1967) used.

We shall first discuss briefly the existence of some of the kernels. Let r be an integer ≥ 2 and let K_0 and K_1 be $L_2(0,1)$ functions vanishing off $(0,1)$ with $\int |u^r K_j|\,du = r!\, c_{jr}$, $j = 0,1$ such that

$$\text{(A.1)} \qquad \int u^t K_0(u)\,du = \begin{matrix} 1 & \text{if} & t = 0 \\ 0 & \text{if} & 0 < t \leq r-1 \end{matrix}$$

and uK_1 satisfies (A.1) with r replaced by $r-1$. For example, K_0 and K_1 can be the first two elements of the dual basis for the subspace of $L_2(0,1)$ with basis $\{1,u,\ldots,u^{r-1}\}$.

As the intended result of these conditions on K_0 and K_1, if S has its r^{th} derivative bounded by M on $(0,1)$, then substitution of the r^{th} order Taylor expansion with Lagrange's remainder shows

$$\text{(A.2)} \qquad \left|\int S K_0\,du - S(0)\right| \leq M c_{0r}$$

and, if in addition $S(0) = 0$,

(A.3) $$\left|\int S\, K_1\, du - S^{(1)}(0)\right| \le M\, c_{1r}\ .$$

Let $X_1, X_2, \ldots$ be a sequence of random variables i.i.d. according to some Lebesgue density f. Let E denote the product measure on $X_1, X_2, \ldots, X_n$.

1. Lebesgue Density

In this section kernel estimates f_n and g_n for f and $g = f^{(1)}$, respectively, will be discussed. Johns and Van Ryzin (1967) proposed these estimates and it appears that they showed (A.9) below under the extra assumption that $f^{(r)}$ is continuous for $x > a$. The bounds on the bias terms in (A.9) improve as the number of derivatives of f increases.

Lemma A.1. (Approximation of f and g). For each x and each $\Delta > 0$, let

(A.4) $$\begin{aligned} \bar{f}(x) &= \int K_0(u)\ f(x+\Delta u)\,du \\ \bar{g}(x) &= \int \Delta^{-1}\ f\Big]_{x+\Delta u}^{x+2\Delta u} K_1(u)\,du\ . \end{aligned}$$

If $f^{(r)}$ exists on $[x, x+2\Delta]$, then

(A.5) $$|\bar{f} - f| \le \Delta^r\, q_\Delta^{(r)}\ c_{0r}$$

(A.6) $$|\bar{g} - g| \le \Delta^{r-1}(q_\Delta^{(r)} + 2^r\, q_{2\Delta}^{(r)})\,c_{1r}\ ,$$

where

(A.7) $$q_\Delta^{(r)}(x) = \sup\{|f^{(r)}(x+\Delta u)| : 0 < u < 1\}\ .$$

Proof. Since $f^{(r)}(x+\Delta\cdot)$ is bounded by $\Delta^r q_\Delta^{(r)}(x)$, (A.5) follows from (A.2). With $S(u) = f\Big]_{x+\Delta u}^{x+2\Delta u}$ in (A.3), the fact $S(0) = 0$

together with $|S^{(r)}| \le \Delta^r(q_\Delta^{(r)} + 2^r\, q_{2\Delta}^{(r)})$ implies (A.6).

Lemma A.2. (Unbiased estimation of $\bar{f}$ and $\bar{g}$). For each x and $\Delta > 0$, let

$$\text{(A.8)} \qquad f_n(x) = n^{-1}\sum_{j=1}^{n} W_j^0(\Delta) \quad \text{and} \quad g_n(x) = n^{-1}\sum_{j=1}^{n} \Delta^{-1}(W_j^1(2\Delta) - W_j^1(\Delta))$$

where $W_j^0(\Delta) = \Delta^{-1} K_0((X_j - x)/\Delta)$ and $W_j^1(\Delta) = \Delta^{-1} K_1((X_j - x)/\Delta)$. Then $f_n(x)$ and $g_n(x)$ are unbiased for $\bar{f}(x)$ and $\bar{g}(x)$, respectively.

Proof. Since the X_j are i.i.d., the proof follows readily from (A.8) and the transformation theorem.

Combining Lemmas A.1 and A.2, we have

Lemma A.3. (Johns and Van Ryzin). Let $\Delta > 0$. If $f^{(r)}$ exists on $[x, x + 2\Delta]$, then

$$\text{(A.9)} \qquad \begin{aligned} |E\, f_n(x) - f(x)| &\le \Delta^r\, q_\Delta^{(r)}(x)\, c_{0r}\,, \\ |E\, g_n(x) - g(x)| &\le \Delta^{r-1}(q_\Delta^{(r)}(x) + 2^r\, q_{2\Delta}^{(r)}(x))c_{1r}\,. \end{aligned}$$

Lemma A.4. (Johns and Van Ryzin). Under the hypothesis of Lemma A.3,

$$\text{(A.10)} \qquad \begin{aligned} \mathrm{Var}\, f_n(x) &\le (n\Delta)^{-1}\, q_\Delta^{(0)}(x)\|K_0\|_2^2\,, \\ \mathrm{Var}\, g_n(x) &\le 3(n\Delta^3)^{-1}\, q_{2\Delta}^{(0)}(x)\|K_1\|_2^2\,, \end{aligned}$$

where Var denotes the variance taken with respect to the measure E, and $\|\cdot\|_2$ denotes the L_2-norm with respect to Lebesgue.

Proof. Since the X_j are i.i.d., the inequality $\operatorname{Var} X \le E(X^2)$ followed by the transformation theorem, and with the C_r-inequality applied at the proper place, yields (A.10).

2. Density with Respect to $d\mu = h\,dx$.

Let f be a Lebesgue density of the form $f = hJ$, where $h > 0$ if and only if $x > a$. Then J is a density with respect to $d\mu = h\,dx$. The estimation of J and its derivative $J^{(1)}$ will be discussed next.

Let $\Delta > 0$. For each x, let

$$\text{(A.11)}\qquad \begin{aligned} J_n(x) &= n^{-1} \sum_{j=1}^{n} W_j^0(\Delta)/h(X_j) , \\ J_n'(x) &= n^{-1} \sum_{j=1}^{n} \Delta^{-1}(W_j^1(2\Delta) - W_j^1(\Delta))/h(X_j) . \end{aligned}$$

Lemma A.5. If $J^{(r)}$ exists on $[x, x + 2\Delta]$, then

$$\text{(A.12)}\qquad \begin{aligned} |E\, J_n - J| &\le \Delta^r S_\Delta^{(r)} c_{0r} , \\ |E\, J_n' - J^{(1)}| &\le \Delta^{r-1} (S_\Delta^{(r)} + 2^r S_{2\Delta}^{(r)}) c_{1r} , \end{aligned}$$

where $S_\Delta^{(r)}(x) = \sup\{|J^{(r)}(x+\Delta u)| : 0 < u < 1\}$.

Proof. The proof is the same as that of Lemma A.3 with W_j^0, W_j^1 and $q_\Delta^{(r)}$ replaced by $W_j^0/h(X_j)$, $W_j^1/h(X_j)$ and $S_\Delta^{(r)}$, respectively.

Lemma A.6. Under the hypothesis of Lemma A.5,

$$\text{(A.13)}\qquad \begin{aligned} \operatorname{Var} J_n &\le (n\Delta)^{-1}\, T_\Delta\, \|K_0\|_2^2 \\ \operatorname{Var} J_n' &\le 3(n\Delta^3)^{-1}\, T_\Delta\, \|K_1\|_2^2 , \end{aligned}$$

where $T_\Delta(x) = \sup \{\frac{J(x+\Delta u)}{h(x+\Delta u)} : 0 < u < 1\}$.

<u>Proof</u>. The proof is the same as that of Lemma A.4 with W_j^0, W_j^1 and $q_\Delta^{(0)}$ replaced by $W_j^0/h(X_j)$, $W_j^1/h(X_j)$ and T_Δ, respectively.