ABSTRACT

CONTRIBUTIONS TO COMPOUND DECISION THEORY

AND EMPIRICAL BAYES SQUARED ERROR LOSS ESTIMATION

By

Richard John Fox

Let $\underline{x}_n = (x_1, x_2, \ldots, x_n)$ be a set of n independent random variables where for $i = 1, 2, \ldots, n$, $x_i \sim P_{\theta_i}$ and $\theta_i \in \Omega$, a real interval. Consider the n testing problems corresponding to $\underline{\theta}_n = (\theta_1, \ldots, \theta_n)$, each having the structure of the following component problem, H: $\theta \le a$, K: $\theta > a$, a in the interior of Ω and letting L denote loss, $L(H,\theta) = (\theta - a)^+$, $L(K,\theta) = (\theta - a)^-$. A compound procedure $\underline{\psi}_n(\underline{x}_n)$ is a sequence $(\psi_1(\underline{x}_n), \ldots, \psi_n(\underline{x}_n))$ of $\underline{x}_n$-measurable test functions where $\psi_i(\underline{x}_n)$ is to be used for testing θ_i. The compound risk of $\underline{\psi}_n$ is the average of the individual risks, $R_n(\underline{\theta}, \underline{\psi}_n) = n^{-1} \Sigma_{i=1}^n R(\underline{\theta}\ , \psi_i)$, where $\underline{\theta} = (\theta_1, \theta_2, \ldots) \in \Omega^\infty$. Define the modified regret of $\underline{\psi}_n$, denoted by $D_n(\underline{\theta}, \underline{\psi}_n)$, by $D_n(\underline{\theta}, \underline{\psi}_n) = R_n(\underline{\theta}, \underline{\psi}_n) - R(G_n)$, where $R(G_n)$ is the Bayes risk versus G_n, the empiric distribution of $\underline{\theta}_n$, in the component problem. For both discrete and continuous exponential families, compound testing procedures are presented whose modified regrets converge to zero as $n \to \infty$.

Consider the problem of estimating G_n, based on $\underline{x}_n$. Let the class of distributions be the uniform on the interval $(0, \theta)$, $\theta > 0$, family. An estimator is presented whose Lévy

distance from G_n, for a certain class of $\underline{\theta}$'s, almost surely converges to zero as $n \to \infty$. Estimators are presented possessing this property for all $\underline{\theta}$'s, when the family is the uniform on the interval $[\theta,\theta+1)$, $\theta \in (-\infty,\infty)$, distributions. For these same two families, it is shown that if the θ_i's are i.i.d. $\sim G$, the same estimators converge in Lévy metric to G.

Let x have distribution function F_θ, $\theta \in \Omega$, a subset of the reals, where θ is a random variable possessing distribution function G. Let $x_1,x_2,\ldots$ be a sequence of random variables i.i.d. according to the marginal distribution on x. Based on $x_1,x_2,\ldots,x_n$, we estimate the conditional mean of θ given x and show that the risk, assuming squared error loss, of using this estimate of θ converges to Bayes risk for three different families of distributions, namely the two uniform families previously discussed and a certain family of Gamma distributions. No assumptions are made concerning G in the uniform $[\theta,\theta+1)$ case and in the other two cases we assume $\int \theta^2 dG(\theta) < \infty$.

Consider the estimation problem discussed immediately above when θ indexes an exponential family on the non-negative integers. Assuming a bounded parameter space, sufficient conditions are presented for obtaining a rate of $n^{-\frac{1}{2}}$ of convergence to Bayes risk. Finally, this same problem is considered in the context of a bivariate exponential family where one component of the two-dimensional parameter indexing the family is to be estimated. An estimator is displayed whose risk, under a set of assumptions, converges to Bayes risk.

CONTRIBUTIONS TO COMPOUND DECISION

THEORY AND EMPIRICAL BAYES

SQUARED ERROR LOSS ESTIMATION

By

Richard John Fox

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

1968

To my parents

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to Professor J. Hannan for his patience and guidance throughout the preparation of this manuscript. His suggestions and comments were of great value in obtaining the results of this thesis. I also wish to thank Professor D.C. Gilliland for his help in the final review.

Finally, I wish to express my gratitude to the National Science Foundation which provided the financial support for this investigation through Grant GP 7362.

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION	1
II.	COMPOUND TESTING IN EXPONENTIAL FAMILIES	4
	2.1 General Remarks	4
	2.2 Discrete Case	8
	2.3 Continuous Case	20
III.	ESTIMATING THE EMPIRICAL DISTRIBUTION OF A PARAMETER SEQUENCE IN COMPOUND DECISION PROBLEMS	27
	3.1 Introduction	27
	3.2 Uniform $(0,\theta)$ Case	29
	3.3 Uniform $[\theta,\theta+1)$ Case	33
	3.4 Estimating the Prior Distribution	37
IV.	SOME EMPIRICAL BAYES SOLUTIONS	42
	4.1 Introduction	42
	4.2 Uniform $(0,\theta)$ Case	43
	4.3 Uniform $[\theta,\theta+1)$ Case	52
	4.4 Estimation of a Location Parameter in Certain Gamma Distributions	54
V.	EMPIRICAL BAYES ESTIMATION IN EXPONENTIAL FAMILIES	62
	5.1 A Rate for the Discrete Case	62
	5.2 Estimation in the Presence of a Nuisance Parameter	70
	BIBLIOGRAPHY	79

CHAPTER I

INTRODUCTION

The problems considered in this thesis fall into the following three categories: Compound Decision Theory, Estimation of Distribution Functions and Empirical Bayes Estimation. In Chapter II, we consider a compound decision problem, where the component problem is a test on the parameter of an exponential family. For both discrete and continuous cases, compound testing procedures which possess a certain desirable asymptotic property are displayed.

Consider a compound decision problem where the underlying family of distributions is either uniform on the interval $(0,\theta)$, $\theta \in (0,\infty)$ or uniform on the interval $[\theta,\theta+1)$, $\theta \in (-\infty,\infty)$. Let G_n denote the empiric distribution of the n parameters corresponding to n observations. In Chapter III, for both families, estimators are presented whose Lévy distances from G_n almost surely converge to zero as $n \to \infty$. For these same two families, we obtain as corollaries that if the θ's are i.i.d. random variables with common distribution function G, these same estimators converge in Lévy metric to G. Robbins (1964) presents a minimum distance technique in a general context for obtaining Lévy-convergent estimators of the prior distribution function G.

In Chapter IV we deal with the Empirical Bayes Quadratic Loss Estimation Problem, which is treated for certain exponential families by Robbins (1964). Suppose x has a distribution depending on a random variable θ possessing distribution function G and that the value of θ is to be estimated. Further suppose that this problem has occurred n times in the past. For three families, namely the two families considered in Chapter III and a certain family of Gamma distributions, estimators of the conditional mean of θ given x, based on the past n observations, are presented whose risk converges to the Bayes risk versus G, i.e. the estimators are asymptotically optimal. The class of prior distributions G, for which this result holds varies with the family.

In Chapter V, two more empirical Bayes quadratic loss estimation results are presented. Macky (1966) displays an asymptotically optimal procedure for a family of exponential distributions on the non-negative integers, under the assumption that the prior distribution possesses a second moment. Under the assumption that the parameter space is bounded, we present sufficient conditions for obtaining a rate of convergence to Bayes risk for this procedure. Finally, asymptotically optimal estimators are presented for one component of a two-dimensional parameter indexing a bivariate exponential family.

We now make some remarks concerning notation. If A is an event, $[A]$ will sometimes be used to denote the indicator function of A. For any distribution function, say F, the letter F also represents the corresponding Lebesgue-Stieltjes measure and $F]_a^b = F(b) - F(a)$. We adopt the convention that distribution functions are right continuous.

For any measure μ and a function f, $\mu(f)$ will occasionally

be used to denote $\int f\, d\mu$. The abbreviation i.o. stands for infinitely often. For any function of a real variable, say g, g' and g'' denote its first and second derivatives. If A is a set, A' denotes its complement. Finally, Φ stands for the standard normal distribution function.

We now make some remarks concerning a certain type of three-point distribution which occurs frequently throughout this paper. Let the random variable x take on the three values $-v, 0, w$, where v and w are positive, with corresponding probabilities $q, 1-q-p, p$. By direct calculation, letting V denote variance,

$$V(x) = v^2 q(1-q) + 2v\, w\, q\, p + w^2 p(1-p). \tag{1.1}$$

The following lemma is due to Gilliland and Hannan (1968).

Lemma 1.1. If x is a random variable assuming the three values $-v, 0, w$, where both v and w are positive, with corresponding probabilities $q, 1-p-q, p$, then letting the range of x be denoted by $r = v + w$ and $\sigma^2 = V(x)$,

$$\frac{r^2}{\sigma^2} \le \frac{1}{q} + \frac{1}{p} .$$

CHAPTER II

COMPOUND TESTING IN EXPONENTIAL FAMILIES

1. General Remarks.

Let $(\mathfrak{X}, \mathcal{B})$ be the measurable space consisting of the Borel field on the real line and let $\mathcal{P} = \{P_\theta | \theta \in \Omega\}$, Ω being a real interval, be a family of probability measures on $(\mathfrak{X}, \mathcal{B})$. Suppose that $\mathcal{P} \ll \mu$ and that

$$\frac{dP_\theta}{d\mu} = p_\theta.$$

Consider the following test of hypothesis problem which we call the component problem. Based on an observation of a random variable x, distributed according to P_θ, $\theta \in \Omega$, we test:

$$H: \theta \leq a$$

versus

$$K: \theta > a,$$

where a is in the interior of Ω. The loss function L is defined as follows:

$$L(H, \theta) = (\theta - a)^+,$$

$$L(K, \theta) = (\theta - a)^-.$$

Let $\underline{x}_n = (x_1, x_2, \ldots, x_n)$ be a sequence of n independent random variables with x_i distributed according to P_{θ_i} and $\theta_i \in \Omega$ for

$i = 1,2,\ldots,n$. Abbreviate P_{θ_i} to P_i, p_{θ_i} to p_i and $\Pi_{i=1}^n P_i$ to $\underline{P}_n$. Let $\underline{\theta}_n = (\theta_1,\theta_2,\ldots,\theta_n)$. We consider the compound decision problem consisting of the n testing problems corresponding to $\underline{\theta}_n$, where the hypotheses are as in the component problem.

We adopt the convention that the value of a test function is the probability of accepting H. By a compound testing procedure is meant a sequence, $\underline{\psi}_n(\underline{x}_n) = (\psi_1(\underline{x}_n),\psi_2(\underline{x}_n),\ldots,\psi_n(\underline{x}_n))$ of $\underline{x}_n$-measurable test functions, where ψ_i is the test function for testing the hypothesis concerning θ_i. Also, for any sequence $\underline{\theta}$, let $\underline{P}$ denote the corresponding product measure on $(\mathfrak{X}^\infty,\mathcal{B}^\infty)$.

For any sequence $\underline{\theta}$, we define the compound risk of $\underline{\psi}_n$, denoted by $R_n(\underline{\theta},\underline{\psi}_n)$, as follows:

$$(2.1)\quad R_n(\underline{\theta},\underline{\psi}_n) = n^{-1}\sum_{i=1}^{n}[\underline{P}_n(\psi_i)(\theta_i-a)^+ + (1-\underline{P}_n(\psi_i))(\theta_i-a)^-],$$

i.e., $R_n(\underline{\theta},\underline{\psi}_n)$ is the average of the risks for the n problems.

Let G_n be the empiric distribution function of $\underline{\theta}_n$. If we restrict $\underline{\psi}_n$ to simple symmetric compound procedures, i.e., $\psi_i(\underline{x}_n) = \psi(x_i)$ where ψ is some test function for the component problem, then the minimum of $R_n(\underline{\theta},\underline{\psi}_n)$ over such procedures is $R(G_n)$, the Bayes risk against G_n in the component problem. Define the modified regret of a procedure $\underline{\psi}_n$, denoted by $D_n(\underline{\theta},\underline{\psi}_n)$, as follows:

$$(2.2)\qquad D_n(\underline{\theta},\underline{\psi}_n) = R_n(\underline{\theta},\underline{\psi}_n) - R(G_n).$$

Hence, a procedure whose modified regret tends to zero, asymptotically

is as good in terms of average risk as the best simple symmetric procedure.

Let ϕ_{G_n} be the Bayes test versus G_n in the component problem which chooses H if and only if $P_x(\theta) < a$, where $P_x(\theta)$ is the conditional expectation of θ given x when the joint distribution on the pair (θ, x) results from G_n on θ and P_θ on x. Thus,

$$\phi_{G_n}(x) = [P_x(\theta) < a]. \tag{2.3}$$

Let $\underline{\phi}_{G_n}$ be the simple symmetric compound procedure which uses ϕ_{G_n} at each stage.

We now derive a useful expression for the modified regret of a compound procedure $\underline{\psi}_n$. Since $R(G_n) = R_n(\underline{\theta}, \underline{\phi}_{G_n})$, by (2.1) and (2.2),

$$D_n(\underline{\theta}, \underline{\psi}_n) = n^{-1} \sum_{i=1}^{n} (\theta_i - a) \underline{P}_n(\psi_i - \phi_{G_n}(x_i)). \tag{2.4}$$

For each i in the right hand side of (2.4), interchange the order of integration so that $\underline{P}_n$ becomes $P_i(\Pi_{j \neq i}^{n} P_j)$. Then, converting the P_i-integrals to μ-integrals with the variable of integration changed from x_i to x, replacing $\Pi_{j \neq i}^{n} P_j$ by $\underline{P}_n$ and interchanging summation and integration, we obtain from (2.4)

$$D_n(\underline{\theta}, \underline{\psi}_n) = \mu(I_n),$$

(2.5)

$$I_n(x) = n^{-1} \sum_{i=1}^{n} (\theta_i - a) p_i(x) [\underline{P}_n(\psi_i(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n)) - \phi_{G_n}(x)].$$

In this chapter μ will either be Lebesgue measure or counting measure on the non-negative integers. $\mathcal{P}$ will be an exponential family specified by the following density with respect to μ:

$$p_\theta(x) = \theta^x C(\theta) m(x), \tag{2.6}$$

where $\theta \in \Omega$, Ω being an interval subset of the non-negative reals and m is a positive function. In both the discrete and continuous cases we shall exhibit compound testing procedures whose modified regrets tend to zero as n increases.

For any real-valued function, say g, on the $\mathcal{P}$-support of $\mathfrak{X}$, define the linear functional T by

$$T(g(x)) = \frac{g(x+1)}{m(x+1)} - a\,\frac{g(x)}{m(x)} . \tag{2.7}$$

Let

$$|T|(x) = (m(x+1))^{-1} + a(m(x))^{-1}. \tag{2.8}$$

Throughout the remainder of this chapter we will occasionally abbreviate expressions involving functions of x by omitting the display of the argument x. Define $\bar{p} = n^{-1}\,\Sigma_{i=1}^n p_i$. Since by (2.7) and the definition of $\bar{p}$,

$$T(\bar{p}(x)) = n^{-1} \sum_{i=1}^{n} (\theta_i - a)\theta_i^x\, C(\theta_i) \tag{2.9}$$

and since $p_\theta(x)$ being defined by (2.6) implies that $P_x(\theta) < a$ iff $T(\bar{p}(x)) < 0$, by (2.3), (2.10) $$\phi_{G_n} = [T(\bar{p}) < 0].$$

Let F^* be the empiric distribution function of $\underline{x}_n = (x_1, x_2, \ldots, x_n)$ and let F_i^* be the empiric distribution function of $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ multiplied by $(n-1)/n$, i.e. $F_i^*(x) = n^{-1} \Sigma_{j \neq i}^n [x_j \leq x]$.

2. Discrete Case.

In this case μ is counting measure on the non-negative integers and the family considered is specified by density (2.6). Our results will be obtained under the following two assumptions:

(A1) $$\Omega = [0, \beta], \ \beta < \infty.$$

(A2) $$\Omega = [\alpha, \beta], \ 0 < \alpha < \beta < \infty.$$

Johns (1966) considers a sequential compound testing problem, i.e., only the first i observations can be used at the i^{th} problem. He considers $\mathcal{P}$ to be a class of distributions having some common discrete support. In particular, the family dealt with in this section is considered by Johns.

Under (A1) and the further assumption that β is in the interior of the natural parameter space, $\{\theta | \theta > 0, \mu(\theta^x m(x)) < \infty\}$, Johns displays a procedure whose modified regret is of order $n^{-\frac{1}{2}}$ uniformly in $\underline{\theta}$. The statistic used in the testing procedure involves artificial randomization. Under (A2) and the same assumption on β, Johns points out that randomization is not necessary.

In the following, a non-randomized compound testing procedure is

ea

me

Al

su

pr

Equ

def

(2.

mod

to

(2.

By

(2.

Also

is given whose modified regret under (A1) converges to zero for each $\underline{\theta}$ and under (A2) converges to zero uniformly in $\underline{\theta}$. The method of proof differs very much from the technique of Johns. Also, an example is given which shows that unless further assumptions are made, no rate of convergence can be found for this procedure under (A1) or (A2).

Define

$$f_i^*(x) = \frac{dF_i^*}{d\mu} = n^{-1} \sum_{j \neq i}^{n} [x_j = x],$$

$$f^*(x) = \frac{dF^*}{d\mu} = n^{-1} \sum_{j=1}^{n} [x_j = x].$$

Equation (2.10) motivates the testing procedure $\underline{\psi}_n^* = (\psi_1^*, \ldots, \psi_n^*)$ defined by

$$\text{(2.11)} \qquad \psi_i^*(\underline{x}_n) = [T(f_i^*(x_i)) < 0], \; i = 1,2,\ldots,n.$$

We now proceed to show that this testing procedure has modified regret converging to zero as $n \to \infty$. It is convenient to introduce the random variables $Y_j(x)$ defined by

$$\text{(2.12)} \qquad Y_j(x) = T([x_j = x]).$$

By (1.1),

$$\text{(2.13)} \quad V(Y_j(x)) = \frac{a^2 p_j(x)(1-p_j(x))}{m^2(x)} + \frac{2ap_j(x)p_j(x+1)}{m(x)m(x+1)} + \frac{p_j(x+1)(1-p_j(x+1))}{m^2(x+1)} .$$

Also,

$$P_j(Y_j(x)) = (\theta_j - a)\theta_j^x \, C(\theta_j). \tag{2.14}$$

We note that by the linearity of T, for any x,

$$T(f_i^*(x)) = n^{-1} \sum_{j \neq i}^{n} Y_j(x) \tag{2.15}$$

and

$$T(f^*(x)) = n^{-1} \sum_{j=1}^{n} Y_j(x) = \overline{Y(x)}. \tag{2.16}$$

Lemma 2.1. $T(f_i^*) - T(\overline{p}) \to 0$ in $\underline{P}$-measure uniformly in i and $\underline{\theta}$ for each x.

Proof: Let x be fixed. By (2.15), (2.16) and the definition of $|T|$ in (2.8),

$$|T(f_i^*) - T(f^*)| = n^{-1}|Y_i| \leq n^{-1}|T|.$$

By (2.9) and (2.14), $T(\overline{p}) = \underline{P}_n(\overline{Y})$. Thus, by (2.16), $|T(f^*) - T(\overline{p})| = |\overline{Y} - \underline{P}_n(\overline{Y})|$ which converges to zero in $\underline{P}$-measure uniformly in $\underline{\theta}$ by the Tchebichev Inequality. Hence, by the triangle inequality $T(f_i^*) - T(\overline{p}) \to 0$ in $\underline{P}$-measure uniformly in $\underline{\theta}$ and i.

Corollary 2.1. For each x, there exist two sequences $\{\delta_n\}$ and $\{\epsilon_n\}$, both positive and decreasing to zero, such that for all $\underline{\theta}$ and $i = 1,2,\ldots,n$, if $|T(\overline{p})| > \delta_n$, then $\underline{P}([T(f_i^*) < 0] = [T(\overline{p}) < 0]) \geq 1 - \epsilon_n$.

Proof: Fix x and let $\{\delta'_n\}$ and $\{\epsilon'_n\}$ be two sequences of positive reals decreasing to zero. Define n'_j, $j = 1,2,\ldots$ such th

i

Le

De

j

n_j

ar

P(

co

a

le

th

Le

po

an

nu

fo

Pr

whe

that $n \geq n'_j$ implies $\underline{P}(|T(f_i^*) - T(\bar{p})| \geq \delta'_j) \leq \epsilon'_j$ for all $\underline{\theta}$ and $i = 1,2,\ldots,n$. The existence of n'_j for each j is guaranteed by Lemma 2.1. Define $n_1 = 1$ and for $j \geq 2$ let $n_j = (n_{j-1} \vee n'_j) + 1$. Define the sequence $\{\epsilon_n\}$ by $\epsilon_n = 1$ for $n_1 \leq n < n_2$ and for $j \geq 2$, $\epsilon_n = \epsilon'_j$ for $n_j \leq n < n_{j+1}$. Define $\{\delta_n\}$ by $\delta_n = \delta'_j$ for $n_j \leq n < n_{j+1}$, for $j = 1,2,\ldots$. The sequences $\{\delta_n\}$ and $\{\epsilon_n\}$ are positive, decreasing to zero and satisfy the condition: $\underline{P}(|T(f_i^*) - T(\bar{p})| \geq \delta_n) \leq \epsilon_n$ for all $\underline{\theta}$ and $i = 1,2,\ldots,n$ which completes the proof.

We now state and prove a lemma which follows immediately from a well known theorem of probability theory. Applications of this lemma to the variables $Y_j(x)$ yield results which are useful in this development.

<u>Lemma 2.2</u>. If z_j, $j = 1,2,\ldots$ are independent random variables possessing finite ranges r_j such that $r_j \leq r < \infty$ for $j = 1,2,\ldots$ and if $s_n^2 = \Sigma_{j=1}^n V(z_j) \to \infty$ as $n \to \infty$, then for any pair of real numbers $b < d$, $P\{b \leq \Sigma_{j=1}^n z_j \leq d\} \to 0$ as $n \to \infty$. Further, if for $j = 1,2,\ldots$, $V(z_j) \geq \delta^2 > 0$, $P\{b \leq \Sigma_{j=1}^n z_j \leq d\} = O(n^{-\frac{1}{2}})$.

<u>Proof</u>: By the Berry-Esseen Theorem, page 288 of Loève (1963),

$$P\{b \leq \sum_{j=1}^{n} z_j \leq d\} \leq \Phi(d^*) - \Phi(b^*) + 2c\ r\ s_n^{-1},$$

where

$$s_n d^* = d - \Sigma_{j=1}^n E(z_j),\ s_n b^* = b - \Sigma_{j=1}^n E(z_j)$$

ar

s

co

ar

Le

P(

nu

un

Pr

If

wh

He

|T

fr

and

$\underline{\theta}$.

the

and c is the Berry-Esseen constant. Since $\Phi(d^*) - \Phi(b^*) \leq s_n^{-1}(2\pi)^{-\frac{1}{2}}(d-b)$, the proof of the first result of the lemma is complete. If $\mathrm{Var}\,(z_j) \geq \delta^2 > 0$ for all j, then $s_n \geq n^{\frac{1}{2}}\delta$ and the proof is complete.

Lemma 2.3. If $\theta_j \nrightarrow 0$ as $j \to \infty$, for each x, $\underline{P}\{b \leq \Sigma_{j=1}^{n} Y_j(x) \leq d\} \to 0$ as $n \to \infty$ where $b < d$ are two real numbers. Further, under (A2), $\underline{P}\{b \leq \Sigma_{j=1}^{n} Y_j(x) \leq d\} = O(n^{-\frac{1}{2}})$ uniformly in $\underline{\theta}$ for each x.

Proof: If $\theta_j \nrightarrow 0$ there exists $\eta > 0$ such that $\theta_j \geq \eta$ i.o. If $\theta_j \geq \eta$, by (2.13)

$$V(Y_j(x)) \geq \frac{a \;\; d_\eta(x)\, d_\eta(x+1)}{m(x)\, m(x+1)}$$

where for $x = 0,1,2,\ldots$

$$d_\eta(x) = \inf_{\theta \in [\eta,\beta]} p_\theta(x) > 0.$$

Hence, $s_n^2 = \Sigma_{j=1}^{n} V(Y_j(x)) \to \infty$. Also, the range of $Y_j(x)$ is $|T|(x)$ for all j. Thus, the first result of the lemma follows from the first part of Lemma 2.2.

Under (A2), for all j

$$V(Y_j(x)) \geq \frac{a \;\; d_{\alpha'}(x)\, d_\alpha(x+1)}{m(x)\, m(x+1)}$$

and the right hand side of this last inequality is independent of $\underline{\theta}$. Hence, the second result of the lemma follows directly from the second part of Lemma 2.2.

(

S[illegible]

Cr[illegible]

an[illegible]

We now define

$$(2.17)\quad V_n(x) = \sup_{1\le i\le n} \underline{P}_n\{T(f_i^*(x)) < 0\} - \inf_{1\le i\le n} \underline{P}_n\{T(f_i^*(x)) < 0\}.$$

<u>Lemma 2.4</u>. If $\theta_j \not\to 0$ as $j \to \infty$, $V_n(x) \to 0$ as $n \to \infty$ for each x. Also, under (A2), $V_n(x) = O(n^{-\frac{1}{2}})$ uniformly in $\underline{\theta}$ for each x.

<u>Proof</u>: Let x be fixed. By (2.15) $T(f_i^*) = \bar{Y} - n^{-1}Y_i$. Hence, by (2.12), for $i = 1,2,\ldots,n$

$$(2.18)\quad \underline{P}_n\{\bar{Y} < -a(nm)^{-1}\} \le \underline{P}_n\{T(f_i^*) < 0\} \le \underline{P}_n\{\bar{Y} < (nm(x+1))^{-1}\}.$$

By the definition of $V_n(x)$ in (2.17) and (2.18),

$$(2.19)\qquad V_n(x) \le \underline{P}_n\{-am^{-1} \le \sum_{j=1}^{n} Y_j \le (m(x+1))^{-1}\}.$$

By Lemma 2.3, $V_n(x) \to 0$ for each x. Under (A2), by the second part of Lemma 2.3, $V_n(x) = O(n^{-\frac{1}{2}})$ uniformly in $\underline{\theta}$ for each x.

<u>Lemma 2.5</u>. If $\theta_j \to 0$ as $j \to \infty$, then

$$\inf_{1\le i\le n} \underline{P}_n(T(f_i^*(0)) < 0) \to 1.$$

<u>Proof</u>: By the lower bound of (2.18),

$$(2.20)\qquad \inf_{1\le i\le n} \underline{P}_n(T(f_i^*(0)) < 0) \ge \underline{P}_n\{\overline{Y(0)} < -a(nm(0))^{-1}\}.$$

Since $\theta_j \to 0$, by (2.14) $\underline{P}_n(\overline{Y(0)}) \to -a(m(0))^{-1} < 0$. By Kolmogorov's Criterion, page 238 of Loève (1963), $\overline{Y(0)} - \underline{P}_n(\overline{Y(0)}) \to 0$ a.s. $\underline{P}$ and the result is immediate from (2.20).

We note that since $C(\theta) = (\Sigma_{x=0}^{\infty} \theta^x m(x))^{-1}$,

(2.21) $$C(0) = (m(0))^{-1} = \sup\{C(\theta) \mid \theta \in [0,\beta]\}.$$

It follows from (2.6) and (2.21), with I_n defined in (2.5), that for $x = 0,1,2,\ldots$

(2.22) $$|I_n(x)| \le C(0)\beta^{x+1}m(x).$$

<u>Proposition 2.1</u>. Under (A1), for each $\underline{\theta}$, $D_n(\underline{\theta},\underline{\psi}_n^*) \to 0$ as $n \to \infty$ where $\underline{\psi}_n^*$ is defined by (2.11).

<u>Proof</u>: Case 1: $\theta_j \to 0$ as $j \to \infty$. Let $x \ge 1$. Recalling the definition of I_n in (2.5), we have

$$|I_n(x)| \le \beta\, n^{-1} \sum_{i=1}^{n} p_i(x)$$

and since $\theta_j \to 0$, the right hand side converges to zero. Let $x = 0$. Since $\theta_j \to 0$, by (2.9) and (2.10) $\phi_{G_n}(0) = 1$ for n sufficiently large. By Lemma 2.5,

$$\inf_{1\le i\le n} \underline{P}_n\{T(f_i^*(0) < 0\} \to 1$$

so that $|I_n(0)| \le n^{-1}\beta\, \Sigma_{i=1}^n |\underline{P}_n\{T(f_i^*(0)) < 0\} - \phi_{G_n}(0)| \to 0$. Thus, since the right hand side of (2.22) is μ-integrable, by the Dominated Convergence Theorem, $\mu(I_n) \to 0$ which by (2.5) completes the proof of Case 1.

Case 2: $\theta_j \not\to 0$ as $j \to \infty$. Let $x = 0,1,\ldots$ be fixed. By Corollary 2.1, there exist two sequences of positive reals depending on x, say $\{\delta_n\}$ and $\{\epsilon_n\}$, both decreasing to zero such that $|T(\overline{p})| > \delta_n$ implies that for $i = 1,2,\ldots,n$

(

w

Pa

re

ger

Lem

$$\underline{P}_n\{[T(f_i^*) < 0] = [T(\bar{p}) < 0]\} \geq 1 - \varepsilon_n.$$

Hence, if $|T(\bar{p})| > \delta_n$, since $p_\theta(x) \leq 1$ for all x and θ,

$$(2.23) \qquad |I_n| \leq n^{-1}\beta \sum_{i=1}^{n} |\underline{P}_n\{T(f_i^*) < 0\} - [T(\bar{p}) < 0]| \leq \beta\, \varepsilon_n.$$

If $|T(\bar{p})| \leq \delta_n$, by (2.9) and the triangle inequality

$$(2.24) \qquad |I_n| \leq m\, \delta_n(1 + q_n) + \beta\, V_n$$

where

$$q_n(x) = \frac{1}{2}\left(\sup_{1\leq i\leq n} \underline{P}_n\{T(f_i^*(x)) < 0\} + \inf_{1\leq i\leq n} \underline{P}_n\{T(f_i^*(x)) < 0\}\right)$$

and V_n is defined by (2.17). Adding the bounds in (2.23) and (2.24) and replacing $1 + q_n$ by 2 we have

$$(2.25) \qquad |I_n| \leq 2m\, \delta_n + \beta(\varepsilon_n + V_n).$$

The right hand side of (2.25) converges to zero by Lemma 2.4 and the construction of $\{\delta_n\}$ and $\{\varepsilon_n\}$. Hence, since the bound of (2.22) is μ-integrable, by the Dominated Convergence Theorem $\mu(I_n) \to 0$ which by (2.5) completes the proof.

In order to prove uniform in $\underline{\theta}$ convergence under (A2), we will need the following lemma which appears as Lemma 5c of Parzen (1959). For a proof of this lemma, Parzen refers the reader to another text where the result appears in a much more general context.

Lemma 2.6. Let $(\mathfrak{X}, \mathcal{A}, \mu)$ be a measure space. Let $f_{n,\lambda}$,

$n = 1,2,\ldots$, $\lambda \in \Lambda$, Λ being some set of possible values of λ, be real-valued Borel measurable functions defined on $(\mathfrak{X},\mathcal{A})$ such that for each n and λ, $|f_{n,\lambda}| \le g$ with $\mu(g) < \infty$. If $f_{n,\lambda}(x) \to 0$ uniformly in $\lambda \in \Lambda$ for each x, as $n \to \infty$, then $\mu(f_{n,\lambda}) \to 0$ uniformly in λ as $n \to \infty$.

Proof: Without loss of generality, assume $0 \le f_{n,\lambda}$ for all n and λ. For $\epsilon > 0$, $\mu(f_{n,\lambda}) = A_n + B_n + C_n$ where

$$A_n = \mu([g \le \epsilon]f_{n,\lambda}),$$

$$B_n = \mu([g > \epsilon][f_{n,\lambda} \le \epsilon^2]f_{n,\lambda})$$

and

$$C_n = \mu([g > \epsilon][f_{n,\lambda} > \epsilon^2]f_{n,\lambda}).$$

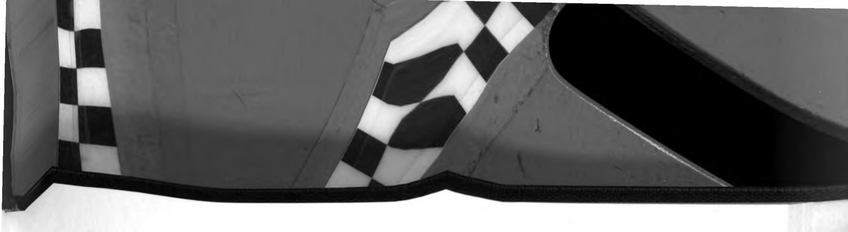
Let ν be the finite measure such that $d\nu/d\mu = g$ and note that $A_n \le \mu([g \le \epsilon]g) = \nu([g \le \epsilon])$, which can be made arbitrarily small by choice of ϵ, since $\nu([g \le 0]) = 0$. Also $B_n \le \epsilon^2\mu([g > \epsilon]) \le \epsilon\mu(g)$ which can be made arbitrarily small by choice of ϵ. Note that by the uniform in λ convergence of $f_{n,\lambda}$ to zero and the finiteness of ν, for any $\delta > 0$,

$$\sup_{\lambda} \nu([f_{n,\lambda} > \delta]) \to 0.$$

Hence, since $C_n \le \nu([f_{n,\lambda} > \epsilon^2])$, as $n \to \infty$

$$\sup_{\lambda} C_n \to 0,$$

which completes the proof.



Proposition 2.2. Under (A2), $D_n(\underline{\theta}, \underline{\psi}_n^*) \to 0$ uniformly in $\underline{\theta}$ as $n \to \infty$ where $\underline{\psi}_n^*$ is defined by 2.11.

Proof: The proof is the same as that of Case 2 of Proposition 2.1 except for the following modifications. Since the sequences $\{\delta_n\}$ and $\{\epsilon_n\}$ are independent of $\underline{\theta}$ for each x and by Lemma 2.4, $V_n(x) = O(n^{-\frac{1}{2}})$ uniformly in $\underline{\theta}$ for each x, the bound on $|I_n|$ of (2.25) converges to zero uniformly in $\underline{\theta}$ for each x. Applying Lemma 2.6 instead of the Dominated Convergence Theorem completes the proof.

Consider the compound testing procedure defined by

$$\psi_i(\underline{x}_n) = [T(f^*(x_i)) < 0], \tag{2.26}$$

$i = 1,2,\ldots,n$. With $\underline{\psi}_n$ defined by (2.26) and $\underline{\psi}_n^*$ by (2.11), since $T(f^*(x_i)) = T(f_i^*(x_i)) - a/n\ m(x_i)$ we obtain by (2.1),

$$|R_n(\underline{\theta}, \underline{\psi}_n) - R_n(\underline{\theta}, \underline{\psi}_n^*)| \le \left|n^{-1} \sum_{i=1}^{n} (\theta_i - a)\underline{P}_n\{0 \le T(f_i^*(x_i)) < a/n\ m(x_i)\}\right|. \tag{2.27}$$

Dealing with the right hand side of (2.27) as we did to obtain (2.5), since $|\theta - a| \le \beta$,

$$|R_n(\underline{\theta}, \underline{\psi}_n) - R_n(\underline{\theta}, \underline{\psi}_n^*)| \le \beta\mu(n^{-1} \sum_{i=1}^{n} p_i(x)\underline{P}_n\{0 \le T(f_i^*(x)) < a/n\ m(x)\}). \tag{2.28}$$

By the definitions of $Y_j(x)$, $|T|(x)$ ((2.12) and (2.8)) and (2.15), for each x

I

e

(

t

c

Fo

to

con

Dom

Theo

unif

Proof

by the

$$(2.29) \quad n^{-1}\sum_{i=1}^{n} p_i(x)\underline{P}_n\{0 \le T(f_i^*(x)) < a/n\ m(x)\} \le \underline{P}_n\{-a/m(x) \le \sum_{j=1}^{n} Y_j(x) < |T|(x)\}.$$

<u>Theorem 2.1</u>. Under (A1), with $\underline{\psi}_n$ defined by (2.26), $D_n(\underline{\theta},\underline{\psi}_n) \to 0$ for each $\underline{\theta}$.

<u>Proof</u>: By Proposition 2.1, $D_n(\underline{\theta},\underline{\psi}_n^*) \to 0$ for each $\underline{\theta}$. Hence, by the triangle inequality, it suffices to show that for each $\underline{\theta}$, $|R_n(\underline{\theta},\underline{\psi}_n) - R_n(\underline{\theta},\underline{\psi}_n^*)| \to 0$.

Case 1. $\theta_j \not\to 0$ as $j \to \infty$.

By Lemma 2.3, the right hand side of (2.29) converges to zero for each x. Hence, since the integrand of the right hand side of (2.28) is bounded by the μ-integrable function $C(0)\beta^x m(x)$, by the Dominated Convergence Theorem the right hand side of (2.28) converges to zero which completes the proof of Case 1.

Case 2. $\theta_j \to 0$ as $j \to \infty$.

For $x \ge 1$, the integrand of the right hand side of (2.28) converges to zero, since $p_j(x) \to 0$ as $j \to \infty$. For $x = 0$, this integrand converges to zero by Lemma 2.5. Again an application of the Dominated Convergence Theorem completes the proof.

<u>Theorem 2.2</u>. Under (A2), with $\underline{\psi}_n$ defined by (2.26), $D_n(\underline{\theta},\underline{\psi}_n) \to 0$ uniformly in $\underline{\theta}$.

<u>Proof</u>: By Proposition 2.2 $D_n(\underline{\theta},\underline{\psi}_n^*) \to 0$ uniformly in $\underline{\theta}$. Hence, by the triangle inequality it suffices to show that

(

It

has

ove

Proce

domin

slowly

$|R_n(\underline{\theta},\underline{\psi}_n) - R_n(\underline{\theta},\underline{\psi}_n^*)| \to 0$ uniformly in θ. Under (A2) the right hand side of (2.29) converges to zero uniformly in $\underline{\theta}$ by Lemma 2.3. Hence, since the integrand of the right hand side of (2.28) is bounded by the μ-integrable function $C(0)\beta^x m(x)$, the right hand side of (2.28) converges to zero uniformly in $\underline{\theta}$ by Lemma 2.6 and the proof is complete.

Example 2.1. The following is a slight modification of Example 3.1 of Gilliland (1966) which shows that no rate of convergence to zero can be found for the modified regret of the compound testing proceudre defined by (2.26) under (A1) or (A2). Let $m(0) = 1$, $m(x) = x^{-3}$ for $x = 1,3,5,\ldots$ and $m(x) = r(x)$ for $x = 2,4,5,\ldots$ with $r(x) \le x^{-3}$ and $r(x)$ strictly decreasing. The parameter space for this family is $[0,1]$. Let $0 < a < 1$ and $\underline{\theta} = (1,1,\ldots)$. Since $R(G_n) = 0$, with $\underline{\psi}_n$ defined by (2.26)

$$D_n(\underline{\theta},\underline{\psi}_n) = n^{-1}\Sigma_{i=1}^n(1-a)\underline{P}_n\{T(f^*(x_i)) < 0\}.$$

It then follows that the right hand side of this last equality has a lower bound of $(1-a)\Sigma_{x=0}^{\infty}p_1(x)(1-p_1(x+1))^{n-1}$. Summing over $x = 1,3,5,\ldots$, since $p_1(x) = C(1)m(x)$, we have

$$D_n(\underline{\theta},\underline{\psi}_n) \ge (1-a)C(1)\sum_{x=0}^{\infty}(2x+1)^{-3}(1-C(1)r(2x+2))^{n-1}.$$

Proceeding exactly as Gilliland, we see that the modified regret dominates a positive null sequence which decreases arbitrarily slowly by choice of $r(x)$.

(

We

(A

Let

By

0 <

Henc

We e

For c

upon

will

real v

(

(2.30)

for i =

3. <u>Continuous Case</u>.

In this section $\mathcal{G}$ is characterized by density (2.6) with respect to Lebesgue measure μ. We recall that m is positive and further assume that:

(A1) $\qquad\qquad$ m is continuous.

We also assume:

(A2) $$\Omega = [\alpha,\beta],\ 0 < \alpha < \beta < \infty.$$

Let F_θ be the distribution function corresponding to p_θ. By (A1), $F'_\theta = p_\theta$.

Abbreviate F_{θ_i} to F_i. Define $\overline{F} = n^{-1}\Sigma_{i=1}^{n} F_i$. Let $0 < h(n) \le 1$ be a sequence of positive reals decreasing to zero. Henceforth, we omit the display of the dependence of h on n. We estimate $\overline{p}$ for the i^{th} problem by a divided difference of F_i^*. For convenience we use central differences, but it is obvious upon analyzing the development that other differencing procedures will work just as well. Let g be a real-valued function of a real variable, x. Define the linear functional $\blacktriangle$ by

$$\blacktriangle\, g(x) = (2h)^{-1} g]_{x-h}^{x+h}.$$

Consider the procedure $\underline{\psi}_n^*$ defined by

(2.30) $$\psi_i^*(\underline{x}_n) = [T(\blacktriangle\, F_i^*(x_i)) < 0],$$

for $i = 1,2,\ldots,n$. For g a $\underline{P}$-square integrable function of

$\underline{x}$, let $||g||$ denote the $L_2(\underline{P})$ norm of g. We note that by (A2) for all $\theta \in \Omega$,

$$C(\theta) \le \overline{C} = \{\mu(([x > 0]\alpha^x + [x \le 0]\beta^x)m(x))\}^{-1}, \tag{2.31}$$

$$C(\theta) \ge \underline{C} = \{\mu(([x > 0]\beta^x + [x \le 0]\alpha^x)m(x))\}^{-1}. \tag{2.32}$$

Lemma 2.7. If $nh^2 \to \infty$, then for each x, $||T(\phi\, F^*(x)) - T(\phi\, \overline{F}(x))|| \to 0$ uniformly in $\underline{\theta}$.

Proof: By the linearity of ϕ and T and the triangle inequality,

$$||T(\phi\, F^*(x)) - T(\phi\, \overline{F}(x))|| \le (m(x+1))^{-1}||\phi(F^*-\overline{F})(x+1)||$$
$$+ a(m(x))^{-1}||\phi(F^*-\overline{F})(x)||.$$

For any y, $||\phi(F^*-\overline{F})y||^2$ is the product of $(2h)^{-2}$ and the variance of the average of n independent Bernoulli random variables. Hence, $||\phi(F^*-\overline{F})y||^2 \le (16nh^2)^{-1}$ which completes the proof.

Lemma 2.8. Under (A1) and (A2), for each fixed x, $|T(\phi\, \overline{F}(x)) - T(\overline{p}(x))| \to 0$ uniformly in $\underline{\theta}$.

Proof: By the linearity of T and the triangle inequality, for each x

$$|T(\phi\, \overline{F}(x)) - T(\overline{p}(x))| \le (m(x+1))^{-1}|\phi\, \overline{F}(x+1) - \overline{p}(x+1)|$$
$$+ a(m(x))^{-1}|\phi\, \overline{F}(x) - \overline{p}(x)|.$$

For any y, since by the linearity of ϕ, $\phi\, \overline{F}(y) = n^{-1}\Sigma_{i=1}^{n}\phi\, F_i(y)$, We n

unif

(A1)

Unde

ineq

$|\alpha^{\delta}_{m}$

$\delta \to ($

to z

Lemm

$\|T(\phi$

Proo

inequ

$\underline{\theta}$. S

is co

(2.33

Let

and not

$$|\phi\,\overline{F}(y) - \overline{p}(y)| \le n^{-1}\sum_{i=1}^{n} |\phi\, F_i(y) - p_i(y)|.$$

We now bound the summands in the immediately preceding expression uniformly in θ and show that this bound converges to zero. By (A1) and the Mean-Value Theorem, $\phi\, F_\theta(y) = p_\theta(y + \delta)$, $|\delta| \le h$. Also

$$|p_\theta(y + \delta) - p_\theta(y)| \le \theta^y C(\theta)|\theta^\delta m(y + \delta) - m(y)|.$$

Under (A2), for all $\theta \in \Omega$, the right hand side of this last inequality is bounded by $(\beta^y \vee \alpha^y)\,\overline{C}\,(|\beta^\delta m(y+\delta) - m(y)| \vee |\alpha^\delta m(y+\delta) - m(y)|)$, where $\overline{C}$ is defined by (2.31). Since $h \to 0$, $\delta \to 0$ and hence, by the continuity of m, this bound converges to zero which completes the proof.

<u>Lemma 2.9</u>. Under (A1) and (A2), if $nh^2 \to \infty$, for each x, $\|T(\phi\, F_i^*(x)) - T(\overline{p}(x))\| \to 0$ uniformly in $\underline{\theta}$ and i.

<u>Proof</u>: Let x be fixed. By Lemmas 2.7 and 2.8 and the triangle inequality for L_2-norm, $\|T(\phi\, F^*(x)) - T(\overline{p}(x))\| \to 0$ uniformly in $\underline{\theta}$. Since $\|T(\phi\, F_i^*(x)) - T(\phi\, F^*(x))\| \le (2hn)^{-1}|T|(x)$, the proof is complete by another application of the triangle inequality.

We now define $M_n(x)$, the continuous analog of $V_n(x)$, by

(2.33) $$M_n(x) = \sup_{1\le i\le n} \underline{P}_n\{T(\phi\, F_i^*(x)) < 0\} - \inf_{1\le i\le n} \{\underline{P}_n(T(\phi\, F_i^*(x)) < 0\}.$$

Let

$$r(x) = \inf\{m(y)|\ |y-x| \le 1\},$$

and note that $r > 0$ by (A1). Then, by (2.32), for all y such t

(

w

Le

an

un

Pr

wh

V(

s_n^2

(2.

the

(2.

wher

and

$\Phi(d^*$

hand

Lemma

that $|y-x| \le 1$ and all $\theta \in \Omega$,

$$p_\theta(y) \ge A(x) \tag{2.34}$$

where

$$A(x) = \underline{C}\ r(x) \exp\{-(|\log \beta| \vee |\log \alpha|)(|x|+1)\}.$$

Lemma 2.10. Under (A1) and (A2), if $nh^2 \to \infty$ then for each x and any two real numbers $b < d$, $\underline{P}\{h^{-1}b \le nT(\phi\, F^*(x)) \le h^{-1}d\} \to 0$ uniformly in $\underline{\theta}$.

Proof: Let x be fixed and note that $nT(\phi\, F^*(x)) = \Sigma_{j=1}^n w_j$ where $w_j = (2h)^{-1}T([x-h < x_j \le x+h])$. By (1.1), $V(w_j) \ge 2a(2h)^{-2}(m(x)m(x+1))^{-1} F_j]_{x-h}^{x+h} F_j]_{x+1-h}^{x+1+h}$. Thus, with $s_n^2 = \Sigma_{j=1}^n V(w_j)$, by (2.34),

$$s_n^2 \ge \frac{2\ anA(x)A(x+1)}{m(x)m(x+1)}. \tag{2.35}$$

By the Berry-Esseen Theorem, page 288 of Loève (1963) and the fact that the range of $w_j = (2h)^{-1}|T|(x)$ for $j = 1,2,\dots,n$,

$$\underline{P}_n\{h^{-1}b \le \sum_{j=1}^n w_j \le h^{-1}d\} \le \Phi(d^*) - \Phi(b^*) + c(s_n h)^{-1}|T|(x), \tag{2.36}$$

where $s_n b^* = h^{-1}b - \Sigma_{j=1}^n P_j(w_j)$ and $s_n d^* = h^{-1}d - \Sigma_{j=1}^n P_j(w_j)$ and c is the Berry-Esseen constant. Since $\Phi(d^*) - \Phi(b^*) \le (s_n h)^{-1}(d-b)$ and $nh^2 \to \infty$, by (2.35) the right hand side of (2.36) converges to zero which completes the proof.

Lemma 2.11. Under (A1) and (A2), if $nh^2 \to \infty$, for each x,

$M_n(x) \to 0$ uniformly in $\underline{\theta}$.

<u>Proof</u>: Let x be fixed. For $i = 1,2,\ldots,n$,

$$\underline{P}_n\{nT(\spadesuit\, F^*(x)) < -a(2hm(x))^{-1}\} \le \underline{P}_n\{T(\spadesuit\, F_i^*(x)) < 0\} \le \underline{P}_n\{nT(\spadesuit\, F^*(x)) < (2hm(x+1))^{-1}\}.$$

Hence, by the definition of $M_n(x)$ in (2.33), $M_n(x)$ is bounded by the difference between the upper and lower bounds of the above inequality. This difference can be expressed as $\underline{P}_n\{-a(2hm(x))^{-1} \le nT(\spadesuit\, F^*(x)) < (2hm(x+1))^{-1}\}$ which converges to zero uniformly in $\underline{\theta}$ by Lemma 2.10, which completes the proof.

We now note that by the bound on $C(\theta)$ of (2.31) and (A2)

$$p_\theta(x) \le v(x) \tag{2.37}$$

where $v(x) = \overline{C}(\beta^x[x > 0] + \alpha^x[x \le 0])m(x)$.

<u>Proposition 2.3</u>. With $\underline{\psi}_n^*$ defined by (2.30), under (A1) and (A2), if $nh^2 \to \infty$, $D_n(\underline{\theta},\underline{\psi}_n^*) \to 0$ uniformly in $\underline{\theta}$.

<u>Proof</u>: Let x be fixed and recall the definition of I_n in (2.5). By the bound on p_θ of (2.37), bounding $|\theta-a|$ by β, we have

$$|I_n(x)| \le \beta v(x)n^{-1}\sum_{i=1}^{n}|\underline{P}_n\{T(\spadesuit\, F_i^*(x)) < 0\} - [T(\overline{p}(x)) < 0]|. \tag{2.38}$$

By Lemma 2.9 and the Tchebichev Inequality, $T(\spadesuit\, F_i^*(x)) - T(\overline{p}(x)) \to 0$ in $\underline{P}$-measure uniformly in $\underline{\theta}$ and i. Using the same construction as in Corollary 2.1 following Lemma 2.1, there exist two sequences $\{\delta_n\}$ and $\{\epsilon_n\}$ of positive reals both decreasing to zero such that

|

fo

ri

si

tr

of

we

fo

fr

th

(2.

The

if

Pro

(2.4

Proce

$P_\theta(x)$

$|T(\bar{p}(x))| > \delta_n$ implies $\underline{P}\{[T(\Phi\, F_i^*(x)) < 0] = [T(\bar{p}(x)) < 0]\} \geq 1-\varepsilon_n$ for $i = 1,2,\ldots,n$ and all $\underline{\theta}$. Hence, if $|T(\bar{p}(x))| > \delta_n$, the right hand side of (2.38) is bounded by $\beta v(x)\varepsilon_n$. If $|T(\bar{p}(x))| \leq \delta_n$, since by (2.9), $T(\bar{p}(x)) = n^{-1}\Sigma_{i=1}^n (\theta_i - a)\theta_i^x C(\theta_i)$, adding and subtracting $q_n(x)$ to the term $[\underline{P}_n(\psi_i^*(x_1,\ldots,x_{i-1},x,x_{i+1},\ldots,x_n)) - \phi_{G_n}(x)]$ of the expression defining I_n in (2.5), where

$$q_n(x) = \frac{1}{2}\left(\sup_{1\leq i\leq n} \underline{P}_n\{T(\Phi\, F_i^*(x)) < 0\} + \inf_{1\leq i\leq n} \underline{P}_n\{T(\Phi\, F_i^*(x)) < 0\}\right),$$

we have $|I_n(x)| \leq m(x)(\delta_n + q_n(x)\delta_n) + \beta v(x)M_n(x)$.

Since ε_n, δ_n and M_n converge to zero uniformly in $\underline{\theta}$ for each x and $|I_n| \leq \beta v$ which is μ-integrable, it follows from Lemma 2.6 that $\mu(|I_n|) \to 0$ uniformly in $\underline{\theta}$, which completes the proof.

Consider the compound testing procedure defined by

$$(2.39)\qquad \psi_i(\underline{x}_n) = [T(\Phi\, F^*(x_i)) < 0].$$

<u>Theorem 2.3</u>. Under (A1) and (A2), with $\underline{\psi}_n(\underline{x}_n)$ defined by (2.39), if $nh^2 \to \infty$, then $D_n(\underline{\theta},\underline{\psi}_n) \to 0$ uniformly in $\underline{\theta}$.

<u>Proof</u>: Since $T(\Phi\, F^*(x_i)) = T(\Phi\, F_i^*(x_i)) - a(2hnm(x_i))^{-1}$, by (2.1),

$$(2.40)\qquad |R_n(\underline{\theta},\underline{\psi}_n) - R_n(\underline{\theta},\underline{\psi}_n^*)| \leq \beta n^{-1}\sum_{i=1}^n \underline{P}_n\{0 \leq T(\Phi\, F_i^*(x_i)) < a(2hnm(x_i))^{-1}\}.$$

Proceeding as we did to obtain (2.5) and applying the bound on $p_\theta(x)$ in (2.37) we get that the right hand side of this inequality

is bounded by:

$$(2.41) \qquad \beta\mu\left(v(x)n^{-1}\sum_{i=1}^{n} \underline{P}_n\{0 \le T(F_i^*(x) \le a(2hnm(x))^{-1}\}\right).$$

Since for $i = 1,2,\ldots,n$,

$$\underline{P}_n\{0 \le T(\spadesuit\, F_i^*(x)) \le a(2hnm(x))^{-1}\} \le \underline{P}_n\{-a(2hm(x))^{-1}$$

$$\le nT(\spadesuit\, F^*(x)) \le (2h)^{-1}|T|(x)\},$$

the integrand of (2.41) converges to zero uniformly in $\underline{\theta}$ for each x by Lemma 2.10. Since this integrand is bounded above by $v(x)$ which is μ-integrable, the integral of (2.41) converges to zero uniformly in $\underline{\theta}$ by Lemma 2.6. It follows that the left hand side of (2.40) converges to zero uniformly in $\underline{\theta}$. Hence, by Proposition 2.3 and the triangle inequality, $D_n(\underline{\theta},\underline{\psi}_n) \to 0$ umiformly in $\underline{\theta}$.

<u>Remark</u>. Suppose we consider $\mathfrak{X}$ to be an interval neighborhood of $+\infty$. Redefine $\spadesuit$ to be a right divided difference, i.e., $\spadesuit\, g(x) = h^{-1}\, g]_x^{x+h}$ and let $r(x) = \inf\{r(y) \mid 1 \ge y-x \ge 0\}$. If one modifies the techniques of this section accordingly, Theorem 2.3 can be obtained in this more general context. Taking $\mathfrak{X}$ to be of this form includes the Gamma and Negative Exponential distributions as special cases.

CHAPTER III

ESTIMATING THE EMPIRICAL DISTRIBUTION FUNCTION OF A PARAMETER SEQUENCE IN COMPOUND DECISION PROBLEMS

1. Introduction.

In a statistical compound decision problem one is faced with a set of n problems all having the structure of a certain component problem. For example, in Chapter II the component problem is a test on the parameter of an exponential family. In such problems, procedures are desired whose compound risk (average risk over the n problems) converges to $R(G_n)$, the Bayes Risk versus G_n, the empirical distribution of the n parameters, in the component problem. It is evident that knowledge of G_n is useful in these problems.

Let F_θ be a distribution function for each $\theta \in \Omega$, some subset of the real line and f_θ be a corresponding density with respect to Lebesgue measure μ. Let $\underline{x} = (x_1, x_2, \ldots)$ be a sequence of independent random variables where x_i has distribution function F_{θ_i}, henceforth abbreviated to F_i, and $\theta_i \in \Omega$ for $i = 1, 2, \ldots$. Also abbreviate f_{θ_i} to f_i. Let $\underline{F} = \Pi_{i=1}^{\infty} F_i$ and G_n be the empiric distribution of the n parameters corresponding to $x_1, x_2, \ldots, x_n$.

Define the following functions

$$\bar{F} = G_n(F_\theta) = n^{-1} \sum_{i=1}^{n} F_i, \tag{3.1}$$

(3.

and

(3.

For

(3.4

disc

metr

foll

F_1

d

Loèv

dist

inter

whose

certa

of uni

In thi

conver

conside

accordi

(3.2) $$\bar{f} = G_n(f_\theta) = n^{-1}\sum_{i=1}^{n} f_i$$

and note that $\bar{f} = d\bar{F}/d\mu$. Let

(3.3) $$F^*(x) = n^{-1}\sum_{i=1}^{n}[x_i \le x].$$

For any real-valued function g of a real vaiable, say x, define

(3.4) $$\Delta g(x) = h^{-1}g]_x^{x+h}, \; h > 0.$$

We now make some remarks about the Lévy metric which is discussed on page 215 of Loève (1963). The Lévy metric is the metric on the space of all distribution functions defined by the following distance formula. For any two distribution functions F_1 and F_2, letting d denote distance,

$$d(F_1,F_2) = \inf\{\epsilon > 0 \mid \text{for all } x,\ F_1(x-\epsilon)-\epsilon \le F_2(x) \le F_1(x+\epsilon)+\epsilon\}.$$

Loève mentions that convergence in Lévy metric of a sequence of distribution functions is equivalent to complete convergence.

In section 2, we consider the family of uniform on the interval $(0,\theta)$ distributions, $\theta \in (0,\infty)$ and exhibit an estimator whose Lévy distance from G_n converges to zero a.s. $\underline{F}$ for a certain class of $\underline{\theta}$'s. In section 3, we deal with the family of uniform on the interval $[\theta,\theta+1)$ distributions, $\theta \in (-\infty,\infty)$. In this case, we find an estimator whose Lévy distance from G_n converges to zero a.s. $\underline{F}$ for every $\underline{\theta}$. In section 4, we again consider these two families and assume that the θ's are i.i.d. according to some distribution G. We then apply the results of

s

t

p

sh

mi

of

fu

fur

ass

the

for

cha

sec

cont

find

the e

game.

2. Un

$\theta \in \Omega =$

It then

sections 2 and 3 to the problem of estimating this prior distribution function G.

Robbins (1964) treats the general problem of estimating a prior distribution function G. Under certain conditions, he shows that if the estimate of G is chosen so that the resulting mixed distribution function is within ϵ_n ($\epsilon_n \downarrow 0$ as $n \to \infty$) of minimizing over the class of possible mixed distribution functions, the sup norm distance from the empiric distribution function of the observations $(x_1, x_2, \ldots, x_n)$, then, under certain assumptions, almost surely the estimator will converge to G on the continuity set of G. However, no explicit method is given for obtaining this estimator. The family of section 2 of this chapter is discussed in Robbins' Example 3 and the family of section 3 is a special case of his Theorem 2.

Deely and Kruse (1968) further assume that $F_\theta(x)$ is continuous in x for each θ. They then exhibit a method of finding an estimator satisfying Robbins' condition. Calculating the estimate involves finding an optimal strategy in a certain game.

2. <u>Uniform</u> $(0,\theta)$ <u>Case</u>.

We consider the following family of distributions. For $\theta \in \Omega = (0,\infty)$, let

$$f_\theta(x) = \theta^{-1}[0 < x < \theta].$$

It then follows that

w

F

a

or

from

defin

<u>Lemma</u>

$$F_\theta(x) = \begin{cases} 0 & x \le 0, \\ x\theta^{-1} & 0 < x < \theta, \\ 1 & x \ge \theta. \end{cases}$$

Thus, for any x, by the definitions of $\overline{F}$ and $\overline{f}$ in (3.1) and (3.2) respectively,

$$\overline{F}(x) = G_n(x) + x\overline{f}(x). \tag{3.5}$$

We estimate G_n at any point $x \ge 0$ by

$$G_n^*(x) = F^*(x) - x \phi F^*(x) \tag{3.6}$$

where F^* and ϕ are defined by (3.3) and (3.4) respectively. For each n form the following grid:

$$0 = x_{n0} < x_{n1} < \ldots < x_{nN} < x_{nN+1}$$

and let $\beta = x_{nN+1}$. For any $\varepsilon > 0$, define for all $x \ge 0$.

$$A_n(x) = \{\underline{x} | x \phi F^*(x) > (x-\varepsilon)\overline{f}(x-\varepsilon) + \overline{F}]_{x-\varepsilon}^{x} + \varepsilon/2\},$$

$$B_n(x) = \{\underline{x} | x \phi F^*(x) < (x+\varepsilon)\overline{f}(x+\varepsilon) - \overline{F}]_{x}^{x+\varepsilon} - \varepsilon/2\}.$$

Note that $\overline{F}(x)$ can be written as either $\overline{F}(x-\varepsilon) + \overline{F}]_{x-\varepsilon}^{x}$ or $\overline{F}(x+\varepsilon) - \overline{F}]_{x}^{x+\varepsilon}$. The following lemma then follows immediately from equation (3.5), definition (3.6) which defines G_n^* and the definitions of $A_n(x)$ and $B_n(x)$.

<u>Lemma 3.1</u>. For any $\varepsilon > 0$, for each $x \ge 0$

L

Pr

It

x

fol

Sin

bel

sub

$x \notin$

$A_n($

by T

$\underline{F}(A_n$

(3.7)

it fol

$x \notin F^*($

$$\{\underline{x} \mid G_n^*(x) < G_n(x-\epsilon) - \epsilon\} \subset \{\underline{x} \mid F^*(x) < \overline{F}(x) - \epsilon/2\} \cup A_n(x),$$

$$\{\underline{x} \mid G_n^*(x) > G_n(x+\epsilon) + \epsilon\} \subset \{\underline{x} \mid F^*(x) > \overline{F}(x) + \epsilon/2\} \cup B_n(x).$$

<u>Lemma 3.2</u>. If $h \to 0$ and $\Sigma_{n=1}^{\infty} N \exp\{-nh^2\epsilon^2/2\beta^2\} < \infty$, then

$$\underline{F}\{\bigcup_{j=0}^{N} (A_n(x_{nj}) \cup B_n(x_{nj})) \text{ i.o.}\} = 0.$$

<u>Proof</u>: Let $0 < x < \beta$ be fixed and note that

$$x \Delta F^*(x) = n^{-1} \sum_{i=1}^{n} x \Delta[x_i \le x].$$

It follows from the definition of f_i that the variables $x \Delta[x_i \le x]$ have expectations $x \Delta F_i(x) \le xf_i(x)$. It then follows that $x \Delta F^*(x)$ has expectation bounded above by $x\overline{f}(x)$. Since $\overline{f}$ is decreasing on $x > 0$, we bound $\overline{F}]_{x-\epsilon}^{x} = \int_{x-\epsilon}^{x} \overline{f}\, d\mu$ below by $\epsilon\, \overline{f}(x)$ if $x \ge \epsilon$ and $x\overline{f}(x)$ if $x < \epsilon$. Thus subtracting the upper bound, $x\overline{f}(x)$, on the expectation of $x \Delta F^*(x)$ from the right hand side of the inequality defining $A_n(x)$ yields a quantity which is bounded below by $\epsilon/2$. Hence, by Theorem 2 of Hoeffding (1963), since $x < \beta$, $\underline{F}(A_n(x)) \le \exp(-nh^2\epsilon^2/2\beta^2)$. Then, since $\underline{F}(A_n(0)) = 0$ for all n,

(3.7) $$\underline{F}(\bigcup_{j=0}^{N} A_n(x_{nj})) \le N \exp(-nh^2\epsilon^2/2\beta^2).$$

If $h \le \epsilon$, $f_i(x+\epsilon) > 0$ implies $\Delta F_i(x) = f_i(x+\epsilon)$ and it follows $x \Delta F_i(x) \ge xf_i(x+\epsilon)$. Hence, the expectation of $x \Delta F^*(x)$ is bounded below by $x\overline{f}(x+\epsilon)$. Hence, since

$\bar{F}]_x^{x+\epsilon} \geq \epsilon\bar{f}(x+\epsilon)$, subtracting $x\bar{f}(x+\epsilon)$ from the right hand side of the inequality defining $B_n(x)$ yields a quantity less than or equal to $-\epsilon/2$. Again applying Hoeffding's Theorem 2, since $x < \beta$ and $\underline{F}(B_n(0)) = 0$ for all n we obtain, for $h \leq \epsilon$, the bound of the right hand side of (3.7) for $\underline{F}(\cup_{j=0}^{N} B_n'(x_{nj}))$. Since the infinite series formed by summing this bound over n converges by assumption, by the Borel-Cantelli Lemma, the proof is complete.

Define the distribution function $\hat{G}_n$ by $\hat{G}_n(0-) = 0$, $\hat{G}_n(\beta) = 1$ and for $0 \leq x < \beta$

$$\hat{G}_n(x) = \max\{G_n^*(x_{nj}) \mid 0 \leq x_{nj} \leq x\}.$$

Note that $G_n^*(0) = 0$.

<u>Theorem 3.1</u>. If $\delta = \max\{x_{nj+1} - x_{nj} \mid 0 \leq j \leq N\} \to 0$, $h \to 0$, $G_n(\beta) \to 1$ and for all $\epsilon > 0$

$$\sum_{n=1}^{\infty} N \exp(-nh^2\epsilon^2/2\beta^2) < \infty,$$

then $d(\hat{G}_n, G_n) \to 0$ a.s. $\underline{F}$.

<u>Proof</u>: Let $\epsilon > 0$ be arbitrary. By the extension of the Glivenko-Cantelli Theorem to non-identically distributed independent random variables, see Theorem 4.1 of Wolfowitz (1953), a.s. $\underline{F}$, $\bar{F} - F^* \to 0$ uniformly in x. It then follows from Lemmas 3.1 and 3.2 that

$$(3.8)\quad \underline{F}\{\bigcup_{j=0}^{N}(G_n(x_{nj} - \epsilon) - \epsilon \leq G_n^*(x_{nj}) \leq G_n(x_{nj} + \epsilon) + \epsilon)'\ \text{i.o.}\} = 0.$$

3

f

w

3

θ

It

(3.

By (

(3.1(

point

$[F^{*}(x),$

Since for $0 \le x < \beta$, $\hat{G}_n(x) = \hat{G}_n(x')$ where x' is the largest x_{nj} which is not larger than x,

$$\bigcup_{0\le x<\beta} \{\underline{x} \mid \hat{G}_n(x) > G_n(x+\epsilon) + \epsilon\} \subset \bigcup_{j=0}^{N} \{\underline{x} \mid G_n^*(x_{nj}) > G_n(x_{nj} + \epsilon) + \epsilon\}$$

and if $\delta \le \epsilon$,

$$\bigcup_{0\le x<\beta} \{\underline{x} \mid \hat{G}_n(x) < G_n(x-2\epsilon) - \epsilon\} \subset \bigcup_{j=0}^{N} \{\underline{x} \mid G_n^*(x_{nj}) < G_n(x_{nj} - \epsilon) - \epsilon\}.$$

Since $\delta \to 0$ and $\hat{G}_n(\beta) \to 1$, by (3.8) and the fact that $\hat{G}_n = G_n$ for $x < 0$, a.s. $\underline{F}$, for n sufficiently large, $d(\hat{G}_n, G_n) \le 2\epsilon$, which completes the proof.

3. Uniform $[\theta,\theta+1)$ Case.

We now consider the following family of distributions. For $\theta \in \Omega = (-\infty,\infty)$,

$$f_\theta(x) = [\theta \le x < \theta + 1].$$

It then follows that for all x

$$\bar{f}(x) = G_n(x) - G_n(x-1). \tag{3.9}$$

By (3.9),

$$G_n(x) = \sum_{r=0}^{\infty} \bar{f}(x-r). \tag{3.10}$$

Since $F^*(x) \le G_n(x) \le F^*(x+1)$, we estimate G_n at a point x by $G_n^*(x)$ which is the truncation to the interval $[F^*(x), F^*(x+1)]$ of $\Sigma_{r=0}^{\infty} \Delta F^*(x-r)$, i.e.

W

r

(3

By

$\underline{F}(T$

(3.

(3.11) $$G_n^*(x) = \{((\sum_{r=0}^{\infty} \spadesuit\, F^*(x-r)) \vee F^*(x)) \wedge F^*(x+1)\}.$$

For convenience we assume that $h \le 1$.

Lemma 3.3. For any $\epsilon > 0$, if $h \le \epsilon$, then for all x,

$$\underline{F}(\{x \mid G_n(x-\epsilon) - \epsilon \le G_n^*(x) \le G_n(x+\epsilon) + \epsilon\}') \le 2\exp(-2nh^2\epsilon^2).$$

Proof: Since the truncation involved in the definition of G_n^* can only improve the estimator, it suffices to prove the lemma for the estimator T_n, defined by

$$T_n(x) = \sum_{r=0}^{\infty} \spadesuit\, F^*(x-r) = n^{-1} \sum_{i=1}^{n} \sum_{r=0}^{\infty} \spadesuit\, [x_i \le x-r].$$

By the definition of $\overline{F}$ in (3.1),

$$\underline{F}(T_n(x)) = h^{-1} \sum_{r=0}^{\infty} \overline{F}\Big]_{x-r}^{x+h-r}.$$

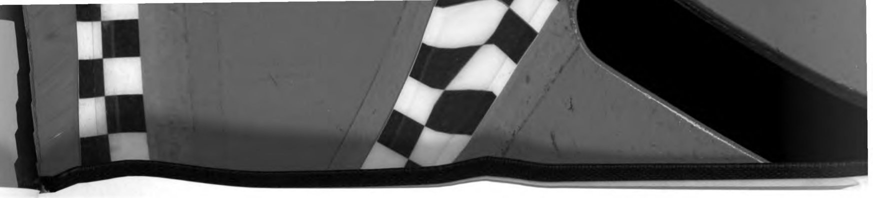
By (3.9),

$$\sum_{r=0}^{\infty} \overline{F}\Big]_{x-r}^{x+h-r} = \sum_{r=0}^{\infty} \left(\int_{x-r}^{x+h-r} (G_n(t) - G_n(t-1))\,dt\right).$$

Writing $\int_{x-r}^{x+h-r} G_n(t-1)\,dt$ as $\int_{x-r-1}^{x+h-r-1} G_n(t)\,dt$, we see that the right hand side is a telescopic series and we obtain

(3.12) $$\sum_{r=0}^{\infty} \overline{F}\Big]_{x-r}^{x+h-r} = \int_{x}^{x+h} G_n(t)\,dt.$$

By Hoeffding's (1963) Theorem 2, since by (3.12), $\underline{F}(T_n(x)) \ge G_n(x)$, $\underline{F}(T_n(x) < G_n(x-\epsilon) - \epsilon) \le \exp(-2nh^2\epsilon^2)$. Similarly, if $h \le \epsilon$, by (3.12), $\underline{F}(T_n(x)) \le G_n(x+\epsilon)$; applying Hoeffding's bounds again



$\underline{F}(T_n(x) > G_n(x+\varepsilon) + \varepsilon) \le \exp(-2nh^2\varepsilon^2)$, and the lemma is proved.

Let $\delta = N^{-1}$, N a positive integer depending on n and consider the following grid on the real line:

$$\ldots < -2\delta < -\delta < 0 < \delta < 2\delta < \ldots \quad .$$

Define the following distribution function.

(3.13) $\qquad \hat{G}_n(x) = \sup\{G_n^*(j\delta) \mid j\delta \le x,\ j = 0, \pm 1, \pm 2, \ldots\}.$

<u>Theorem 3.2</u>. If $\Sigma_{n=1}^{\infty} N \exp(-2nh^2\varepsilon^2) < \infty$ for any $\varepsilon > 0$, $N \to \infty$ and $h \to 0$, then for any $\underline{\theta}$, a.s. $\underline{F}$, $d(\hat{G}_n, G_n) \to 0$, where $\hat{G}_n$ is defined by (3.13).

<u>Proof</u>: Let $\varepsilon > 0$ be arbitrary. Let $h \le \varepsilon$ and $\delta \le \varepsilon$. Let J be the largest integer such that $F^*(J\delta + 1) < \varepsilon$. Define the following subset of the real line, letting $\mathcal{J} = \{j \mid F^*]_{j\delta}^{(j+1)\delta+1} \ge \varepsilon$, $j \ge J$, $j = 0, \pm 1, \ldots\}$,

$$A_n = \bigcup_{j \in \mathcal{J}} [j\delta, (j+1)\delta) \ .$$

Note that there are at most $L = (N+1)M$ grid points in A_n where M is the smallest integer greater than or equal to ε^{-1}. Also note that A_n may be empty. If $x < J\delta$, since $F^*(J\delta + 1) < \varepsilon$, $G_n(x) < \varepsilon$ and $\hat{G}_n(x) < \varepsilon$ and it follows trivially that $G_n(x-\varepsilon) - \varepsilon \le \hat{G}_n(x) \le G_n(x+\varepsilon) + \varepsilon$. For $m \ge J$, let $F^*]_{m\delta}^{(m+1)\delta+1} < \varepsilon$. Then for $x \in [m\delta, (m+1)\delta)$, since both $G_n(x)$ and $G_n(x)$ are in the interval $[F^*(m\delta), F^*((m+1)\delta + 1)]$, $G_n(x-\varepsilon) - \varepsilon \le \hat{G}_n(x) \le G_n(x+\varepsilon) + \varepsilon$.

Let $[m\delta, (m+1)\delta) \subset A_n$. For all x in this interval x

E

a

Le

Proo

be fi

$\hat{G}_n(x) = \hat{G}_n(m\delta)$. Thus,

$$\bigcup_{x\in A_n} \{\underline{x} | \hat{G}_n(x) > G_n(x+\epsilon) + \epsilon\} \subset \bigcup_{j\delta\in A_n} \{\underline{x} | G_n^*(j\delta) > G_n(j\delta+\epsilon) + \epsilon\}$$

and since $\delta \le \epsilon$,

$$\bigcup_{x\in A_n} \{\underline{x} | \hat{G}_n(x) < G_n(x-2\epsilon) - \epsilon\} \subset \bigcup_{j\delta\in A_n} \{\underline{x} | G_n^*(j\delta) < G_n(j\delta-\epsilon) - \epsilon\}.$$

The $\underline{F}$ measure of the union of the two right hand sides of the above inclusions, by Lemma 3.3 is less than or equal to $2L \exp\{-2nh^2\epsilon^2\}$. Since $\Sigma_{n=1}^{\infty} N \exp(-2nh^2\epsilon^2) < \infty$, by the Borel-Cantelli Lemma

$$\underline{F}(\bigcup_{x\in A_n} \{\underline{x} | G_n(x-2\epsilon) - \epsilon \le \hat{G}_n(x) \le G_n(x+\epsilon) + \epsilon\}' \text{ i.o.}) = 0.$$

It follows that a.s. $\underline{F}$, $d(\hat{G}_n, G_n) \le 2\epsilon$ for n sufficiently large, which completes the proof.

Remark: We have tacitly been assuming, in both sections 2 and 3, that $d(G_n, \hat{G}_n)$ is for each n a Borel-measurable function of $\underline{x}_n$, where the σ-field on the space of $\underline{x}_n$'s is the n-dimensional Borel sets. It will be shown that these measurability assumptions are satisfied in the proofs of Corollaries 3.1 and 3.2.

Lemma 3.4. For $a > 0$, $\alpha > 0$ and all c,

$$\sum_{n=1}^{\infty} n^c \exp\{-an^{\alpha}\} < \infty.$$

Proof: It suffices to prove the lemma for $c > 0$. Let $c > 0$ be fixed and let m, a positive integer, be such that $x \ge m$

P

n

r

br

Le

me

pro

Pro

$\underline{\theta}$,w

implies $(x+1)^c < x^{2c}$. Then

$$\sum_{n=m+1}^{\infty} n^c \exp(-an^{\alpha}) < \int_m^{\infty} x^{2c} \exp(-ax^{\alpha})dx,$$

and the integral on the right hand side can be shown to be finite by the change of variable: $y = x^{\alpha}$, which completes the proof.

<u>Remark</u>: Letting $N = n^c$, c a positive integer, $h = n^{-\alpha}$ and $\beta = n^{\gamma}$ with $\alpha,\gamma > 0$, and $\alpha + \gamma < \frac{1}{2}$, by Lemma 3.4 the series of the hypothesis of Theorem 3.1 converges. Letting $N = n^c$, c a positive integer and $h = n^{-\alpha}$, $0 < \alpha < \frac{1}{2}$, again by Lemma 3.4, the series of the hypothesis of Theorem 3.2 converges.

4. Estimating the Prior Distribution.

Let $(R,\mathcal{B})$ be the measurable space consisting of the real line and the Borel field. Let $(R^m,\mathcal{B}^m)$, where m is a positive integer or infinity, be the usual product space. We now drop the condition that F_{θ} is absolutely continuous with respect to μ for each $\theta \in \Omega$.

We refer the reader to page 137 of Loève (1963) for a brief discussion of regular conditional probability.

<u>Lemma 3.5</u>. If F_{θ}, $\theta \in \Omega$, is a regular conditional probability measure on $(R,\mathcal{B})$, then $\underline{F}$, $\underline{\theta} \in \Omega^{\infty}$, is a regular condition probability measure on $(R^{\infty},\mathcal{B}^{\infty})$.

<u>Proof</u>: Since $\underline{F}$ is a probability measure on $(R^{\infty},\mathcal{B}^{\infty})$ for each $\underline{\theta}$, we only need show that for each set $B \in \mathcal{B}^{\infty}$, $\underline{F}(B)$ is a

measurable function of $\underline{\theta}$. If B is a measurable cylinder set, i.e. $B = \Pi_{i=1}^{\infty} B_i$, $B_i \in \mathcal{B}$ for $i = 1,2,\ldots$, where only a finite number of B_i's are not equal to R, $\underline{F}(B)$ is a finite product of terms of the form $F_i(B_i)$ and hence is a measurable function of $\underline{\theta}$. Thus, since it is easily seen that the class of all subsets whose $\underline{F}$- measure is a measurable function of $\underline{\theta}$ is a σ-field and since the measurable cylinders are the generators of $\mathcal{B}^{\infty}$, the proof is complete.

Suppose θ_i, $i = 1,2,\ldots$, are i.i.d. according to some distribution G. Let $G^{\infty}(\underline{F})$ denote the marginal distribution on $\underline{x}$ of the joint distribution on pairs $(\underline{\theta},\underline{x})$ resulting from G^{∞} on $\underline{\theta}$ and $\underline{F}$ on $\underline{x}$.

<u>Theorem 3.3</u>. If θ_i are i.i.d. according to G and $\hat{G}_n$ is an estimator of G_n based on $(x_1,x_2,\ldots,x_n)$ such that $d(\hat{G}_n,G_n)$ is jointly measurable in $(\underline{\theta}_n,\underline{x}_n)$ for each n and if $G^{\infty}\{\underline{\theta} \mid d(\hat{G}_n,G_n) \to 0 \text{ a.s. } \underline{F}\} = 1$, and F_θ, $\theta \in \Omega$, is a regular conditional probability measure,then $d(\hat{G}_n,G) \to 0$ a.s. $G^{\infty}(\underline{F})$.

<u>Proof</u>: By the triangle inequality $d(\hat{G}_n,G) \leq d(\hat{G}_n,G_n) + d(G_n,G)$. By the Glivenko-Cantelli Theorem, page 20 of Loeve (1963), $d(G_n,G) \to 0$ a.s. G^{∞}. Let C be the set of pairs $(\underline{\theta},\underline{x})$ such that $d(\hat{G}_n,G_n) \to 0$. C is jointly measurable in $(\underline{\theta},\underline{x})$ and since by Lemma 3.5 $\underline{F}$ is regular, the measure of C is $G^{\infty}(\underline{F}(C))$. Since $\underline{F}(C) = 0$ a.s. G^{∞}, the proof is complete.

Let the σ-field on Ω be the restriction of the Borel Field to Ω.

Corollary 3.1. If F_θ corresponds to the uniform distribution on the interval $(0,\theta)$, $\theta \in \Omega = (0,\infty)$, if θ_i, $i = 1,2,\ldots$, are i.i.d. according to G, if $\hat{G}_n$ is defined as in Section 2 of this chapter and if the hypotheses of Theorem 3.1 are satisfied with $\beta \to \infty$ replacing $G_n(\beta) \to 1$, then $d(\hat{G}_n,G) \to 0$ a.s. $G^\infty(\underline{F})$.

Proof: Let B be a Borel set. $F_\theta(B) = \theta^{-1}\mu(B(0,\theta))$ which is a continuous function of $\theta > 0$ and hence F_θ, $\theta \in \Omega$, is a regular condition probability measure.

For each n, $\hat{G}_n(\underline{x}_n)$ assumes one of a finite set of values, each of which is a step function with discontinuity points restricted to the selected grid. The set of $\underline{x}_n$'s for which $\hat{G}_n$ assumes a particular value is a finite union of sets each of which corresponds to a particular set of values for the restriction of G_n^* to the selected grid, where G_n^* is defined by (3.6). Each set of this union is a finite intersection of sets which correspond to a specific value of G_n^* at each point of the grid. Hence, since it is easily seen that G_n^* is a Borel-measurable function of $\underline{x}_n$ for each x, the set of $\underline{x}_n$'s where $\hat{G}_n$ assumes a specific value is measurable. Consider the partition of $\Omega^n \times R^n$ into sets $J_1,J_2,\ldots,J_M$ where M is the number of possible values of $\hat{G}_n$ and each J_i, $i = 1,2,\ldots,M$, is the product of Ω^n and a set of $\underline{x}_n$'s on which $\hat{G}_n$ assumes one of the M possible values and hence is measurable. Then $d(\hat{G}_n,G_n) = \Sigma_{i=1}^M J_i d(\hat{G}_n,G_n)$. Since $d(\hat{G}_n,G_n)$ is continuous in $\underline{\theta}_n$ on each J_i, $J_i d(\hat{G}_n,G_n)$ is a

I

f

C

[θ

di

G^∞

Pr

a c

The

a m

c >

(3.1

where

and

ε is

(3.15)

where

seen th

$\underline{x}_n$ for

measurable function of pairs $(\underline{\theta}_n, \underline{x}_n)$ for $i = 1,2,\ldots,M$. Hence, $d(\hat{G}_n, G_n)$ is jointly measurable for each n.

Since the θ_i's are i.i.d. according to G, by the Glivenko-Cantelli Theorem, $\beta \to \infty$ implies $G_n(\beta) \to 1$ a.s. G^∞. Thus, by Theorem 3.1, $G^\infty\{\underline{\theta} \mid d(\hat{G}_n, G_n) \to 0 \text{ a.s. } \underline{F}\} = 1$ and it follows by Theorem 3.3 that $d(\hat{G}_n, G) \to 0$ a.s. $G^\infty(\underline{F})$.

Corollary 3.2. If F_θ corresponds to the uniform distribution on $[\theta, \theta+1), \theta \in (-\infty, \infty)$ and if $\hat{G}_n$ is defined by (3.13) and if the conditions of Theorem 3.2 are satisfied, then $d(\hat{G}_n, G) \to 0$ a.s. $G^\infty(\underline{F})$.

Proof: Let B be a Borel set. $F_\theta(B) = \mu(B[\theta, \theta+1))$ which is a continuous function of θ and hence F_θ is regular. By Theorems 3.2 and 3.3 it remains to be shown that $d(\hat{G}_n, G_n)$ is a measurable function of pairs $(\underline{\theta}_n, \underline{x}_n)$ for each n. For any $c > 0$,

$$(3.14) \quad \{d(\hat{G}_n, G_n) < c\} = \bigcup_{\underline{\theta}'_n, \epsilon} (\{d(G'_n, G_n) < \epsilon\} \cap \{d(\hat{G}_n, G'_n) < c-\epsilon\}),$$

where $\underline{\theta}'_n$ is an n-dimensional vector with rational components and G'_n is its corresponding empiric distribution function and ϵ is a positive rational. For any fixed $\underline{\theta}_n$ and any $c' \geq 0$

$$(3.15) \quad \{d(\hat{G}_n, G_n) \leq c'\} = \bigcup_r \{G_n(r-c') - c' \leq \hat{G}_n(r) \leq G_n(r+c') + c'\},$$

where r ranges over the set of rationals. Since it is easily seen that $G^*_n(x)$, defined in (3.11), is a measurable function of $\underline{x}_n$ for each x, it follows by the definition of $\hat{G}_n$ that $\hat{G}_n(x)$

is

d((

Als

mea

uni

two

cons

is measurable for each x. Hence, it follows from (3.15) that $d(\hat{G}_n, G_n)$ is a measurable function of $\underline{x}_n$ for each fixed $\underline{\theta}_n$. Also, for each $\underline{\theta}_n'$, $d(G_n', G_n)$ is continuous in $\underline{\theta}_n$ and hence measurable in $\underline{\theta}_n$. Thus, each of the sets of the countable union of the right hand side of (3.14) is an intersection of two measurable cylinders in the space of pairs $(\underline{\theta}_n, \underline{x}_n)$ and consequently is measurable and the proof is complete.

l

d:

St

x

Let

whi

(θ,

fro

whic

μ. I

rando

these

seque

the jo

CHAPTER IV

SOME EMPIRICAL BAYES SOLUTIONS

1. Introduction.

As previously mentioned a symbol, say F, representing a distribution function, also represents the corresponding Lebesgue-Stieltjes measure. For Ω a subset of the real line, let $\{F_\theta \mid \theta \in \Omega\}$ be a class of distributions for a random variable x possessing densities, f_θ, with respect to Lebesgue measure, μ. Let G be a distribution on Ω.

Define

$$K(x) = G(F_\theta(x))$$

which is the marginal distribution function of x of the pair (θ,x), where (θ,x) possesses the joint distribution resulting from G on θ and F_θ on x.

Let

$$k(x) = G(f_\theta(x))$$

which is a determination of the density of K with respect to μ. Let $(x_1,x_2,\ldots)$ be a sequence of i.i.d. according to K random variables and K^∞ be the product measure on the space of these sequences. Let P be the product measure on the space of sequences $(x_1,x_2,\ldots,(\theta,x))$, i.e. P is the product of K^∞ and the joint distribution of (θ,x).

The Bayes Estimator versus G in the problem of estimating θ based on observing x, under quadratic loss, is the conditional expectation of θ given x, $P_x(\theta)$. Denoting the Bayes risk versus G by R, we have $R = P(\phi(x) - \theta)^2$, where ϕ is a Bayes response.

Our objective is to find an estimator, say ϕ_n, based on $(x_1, x_2, \ldots, x_n)$, of ϕ, for which $R_n = P(\phi_n - \theta)^2 \to R$ as $n \to \infty$. If $P(\phi_n - \theta)^2 < \infty$ and $P(\phi - \theta)^2 < \infty$, then $P((\phi_n - \phi)(\phi - \theta)) = 0$ and it follows that

$$R_n - R = P(\phi_n - \phi)^2. \tag{4.1}$$

We shall also consider the technique of first estimating G based on $(x_1, x_2, \ldots, x_n)$, and then using the Bayes Estimator versus this estimate of G to estimate θ.

Throughout this chapter, + or - appearing as an affix on the lower limit of an integral means respectively to exclude and include the lower limit in the range of integration.

2. Uniform $(0,\theta)$ Case.

We consider the same family of distributions discussed in section 2 of Chapter III. For $\theta \in \Omega = (0,\infty)$,

$$f_\theta(x) = \theta^{-1}[0 < x < \theta]$$

and

$$F_\theta(x) = \begin{cases} 0 & x \le 0, \\ x\,\theta^{-1} & 0 < x < \theta, \\ 1 & x \ge \theta. \end{cases}$$

N

k

De

and

(4.5

i.e.

where

In this case we have the following:

$$(4.2)\qquad k(x) = G(f_\theta(x)) = [x > 0]\int_{x+}^{\infty} \theta^{-1} dG,$$

$$(4.3)\qquad K(x) = G(F_\theta(x)) = G(x) + x\,k(x).$$

Henceforth, we only consider $x > 0$. From the definition of f_θ and k,

$$(4.4)\qquad P_x(\theta) = \frac{G(\theta f_\theta(x))}{G(f_\theta(x))} = \frac{1-G(x)}{k(x)}\ .$$

Note that $k(x) = 0$ implies $G(x) = 1$ and $K(x) = 1$. If $k(x) > 0$, by (4.3),

$$P_x(\theta) = x + \frac{1-K(x)}{k(x)}\ .$$

Define

$$\psi = \frac{1 - K}{k}$$

and consider the following Bayes response,

$$(4.5)\qquad \phi(x) = \psi(x) + x.$$

We now make the following definitions:

$$K_n(x) = n^{-1} \sum_{i=1}^{n} [x_i \le x],$$

i.e. K_n is the empiric distribution of $(x_1, x_2, \ldots, x_n)$,

$$k_n(x) = h^{-1} K_n]_{x-h}^{x}$$

where h depends on n and is positive. Define

(4.6) $$\psi_n(x) = (a_n(x) \wedge (1 - K_n(x))/k_n(x))[x \geq h],$$

where undefined ratios are taken to be zero and a_n is a bounded non-negative function of $x > 0$ for each n and

(4.7) $$\phi_n(x) = x + \psi_n(x).$$

We now assume

(A1) $$G(\theta^2) < \infty.$$

By (A1) and Jensen's Inequality $P(\phi^2) < \infty$. Hence, since a_n is bounded for each n the conditions implying (4.1) are satisfied. Thus, we are interested in choosing a_n so that $P(\phi_n - \phi)^2 \to 0$.

Remark. If the failure rate of the marginal distribution of x is bounded away from zero, ψ is bounded, say by C. If we estimate ψ by $(1-K_n)/k_n$ truncated at C, under the proper conditions on h, at every x where $k(x)$ is positive and is the derivative of K, hence a.s. K, the estimator converges in K^∞ measure to $\psi(x)$. Then, since $P(\psi_n - \psi)^2 = P\,P_x(\psi_n - \psi)^2$, by twice applying the Bounded Convergence Theorem, $P(\psi_n - \psi)^2 = P(\phi_n - \phi)^2 \to 0$.

The following example illustrates a parametric class of prior distributions for which a rate of convergence to Bayes risk can be obtained for each member of the class with a non-truncated estimator.

Example 4.1. Suppose $G \ll \mu$ and

He

by

ar

V($\bar{x}$

of

such

Lemm

each

Proof

the d

k(x)

$k_n(x)$

as n

(4.8)

in K^∞

Loève

$$\frac{dG}{d\mu} = \lambda^2\theta\ e^{-\lambda\theta}[\theta > 0],$$

where $\lambda \in (0,\infty)$. Then for $x > 0$,

$$k(x) = \int_{x+}^{\infty}\theta^{-1}dG = \lambda^2\int_{x+}^{\infty} e^{-\lambda\theta}d\theta = \lambda\ e^{-\lambda x}.$$

Hence, $\psi(x) = e^{-\lambda x}/\lambda\ e^{-\lambda x} = \lambda^{-1}$. We estimate $\phi(x) = \lambda^{-1} + x$ by $\phi_n(x) = \bar{x} + x$ where $\bar{x} = n^{-1}\Sigma_{i=1}^{n} x_i$. Since $G(\theta^2) < \infty$ and $P(\bar{x}^2) < \infty$, (4.1) holds. Since $P(\bar{x}) = \lambda^{-1}$, $P(\phi_n - \phi)^2 = V(\bar{x}) = n^{-1}\lambda^{-2}$. Thus, for each G in this class, $R_n - R$ is of order n^{-1}.

We now proceed in the general problem of finding a_n such that $P(\phi_n - \phi)^2 \to 0$.

<u>Lemma 4.1</u>. Under (A1), if $nh^2 \to \infty$, $h \to 0$ and $a_n(x) \to \infty$ for each x, $P((\psi_n - \psi)^-)^2 \to 0$.

<u>Proof</u>: Let $x \in A = \{x | k(x) > 0,\ x \notin \mathcal{D}(G)\}$, where $\mathcal{D}(G)$ is the discontinuity set of G. Since k is continuous at x, $k(x) = K'(x)$. By the Tchebichev Inequality, since $nh^2 \to \infty$, $k_n(x) - h^{-1}K]_{x-h}^{x} \to 0$ in K^{∞}-measure. Since $h^{-1}K]_{x-h}^{x} \to k(x)$ as $n \to \infty$,

$$(4.8) \qquad k_n(x) \to k(x)$$

in K^{∞}-measure. By the Glivenko-Cantelli Theorem, page 20 of Loève (1963) and (4.8),

$$\frac{1-K_n(x)}{k_n(x)} \to \frac{1-K(x)}{k(x)} = \psi(x)$$

in K^∞-measure. Since $a_n(x) \to \infty$,

$$(4.9) \qquad (\psi_n(x) - \psi(x))^- \to 0.$$

Since

$$(4.10) \qquad 0 \le ((\psi_n - \psi)^-)^2 \le \psi^2,$$

by (4.9) and the Bounded Convergence Theorem

$$(4.11) \qquad P_x((\psi_n - \psi)^-)^2 = K^\infty((\psi_n - \psi)^-)^2 \to 0.$$

Since $P(A) = 1$ and under (A1), $P(\psi^2) < \infty$, by (4.10), (4.11) and the Dominated Convergence Theorem, $P\,P_x((\psi_n - \psi)^-)^2 = P((\psi_n - \psi)^-)^2 \to 0$ which completes the proof.

<u>Lemma 4.2</u>. If $x \ge h$ and $k(x) > 0$,

$$P_x((\psi_n - \psi)^+)^2 \le \frac{c(a_n^2 + ha_n)(1^+)}{(n\ h\ k)^{\frac{1}{2}}} + \frac{2(h^{-1}a_n^2 + a_n)}{n^{\frac{1}{2}}k}$$

where c is the Berry-Esseen constant and 1^+ denotes $(1-K(h))^{-\frac{1}{2}}$ which decreases to one as $h \to 0$.

<u>Proof</u>: Let $x \ge h$ be such that $k(x) > 0$. Since $P(\psi_n - \psi > v) = 0$ for $v \ge *$ where $* = (a_n - \psi)^+$, $P_x((\psi_n - \psi)^+)^2 = \int_0^* P_x(\psi_n - \psi > v)\,dv^2$. Let $0 < v < *$. $\psi_n - \psi > v$ iff $\overline{w} < 0$ where for $i = 1,2,\ldots,n$, $w_i = h^{-1}[x-h < x_i \le x](\psi+v) - [x_i > x]$. Since the x_i's are i.i.d., the w_i's are i.i.d. Since k is a decreasing function of $x \ge h$,

$$(4.12) \qquad P_x(w_1) = \phi(\psi + v) - (1-K) \geq vk,$$

where we define $\phi = h^{-1}K]_{x-h}^{x}$. By the Berry-Esseen Theorem, page 288 of Loève (1963), with $V(w_1) = \sigma^2$ and r denoting the range of w_1,

$$(4.13) \qquad P_x(\bar{w} < 0) \leq \Phi(z) + c\, n^{-\frac{1}{2}} r\sigma^{-1},$$

where $z = -n^{-\frac{1}{2}} P_x(w_1)\sigma^{-1}$ and c is the Berry-Esseen constant.

By (1.1), $h^{\frac{1}{2}}\sigma \geq \{(\psi+v)^2\phi(1-h\phi)\}^{\frac{1}{2}}$. Also, $hr = (\psi+v) + h$. Hence,

$$\int_0^{*} r\sigma^{-1} dv^2 \leq (h\phi(1-h\phi))^{-\frac{1}{2}} \int_0^{*} (\psi + v + h)(\psi + v)^{-1} dv^2.$$

Note that $(1-h\phi) \geq 1 - K(h)$ for all $x > 0$. Thus, bounding the integral in the right hand side of the above inequality, we obtain

$$(4.14) \qquad \int_0^{*} r\sigma^{-1} dv^2 \leq \{h\phi(1-K(h))\}^{-\frac{1}{2}}(*^2 + 2h*).$$

By a weakening of the tail bounds of the standard normal distribution function, page 166 of Feller (1957), the lower bound on $P_x(w_1)$ of (4.12) and the fact that $\sigma \leq r$, $\Phi(z) \leq (n^{\frac{1}{2}}vk)^{-1}r$. Hence, since $hr \leq a_n + h$ for $0 \leq v \leq *$,

$$(4.15) \qquad \int_0^{*} \Phi(z)\, dv^2 \leq \frac{2(h^{-1}a_n^2 + a_n)}{n^{\frac{1}{2}}\, k}.$$

Since $\phi \geq k$ for $x \geq h$ and $* \leq a_n$, by (4.13), (4.14) and (4.15), replacing $(1-K(h))^{-\frac{1}{2}}$ by (1^{+}), Re

fo

Le

Pro

the

seco

by in

Lemma

(4.16)

Now ex

$$\int_0^{*} P_x(\bar{w} < 0)\,dv^2 \le \frac{c(a_n^2 + 2ha_n)(1+)}{(n\ h\ k)^{\frac{1}{2}}} + \frac{2(h^{-1}a_n^2 + a_n)}{n^{\frac{1}{2}}k} ,$$

which completes the proof.

Let

$$\|a_n\|_\infty = \sup a_n ,$$

$$\|a_n\|_2 = (\mu(a_n^2))^{\frac{1}{2}} ,$$

$$\|a_n\|_1 = \mu(a_n) .$$

Recall that a_n is a bounded non-negative function on $x > 0$ for each n.

Lemma 4.3. With 1^+ as defined in Lemma 4.2,

$$P((\psi_n - \psi)^+)^2 \le \int_0^h \psi^2 dP + n^{-\frac{1}{2}}((1^+)ch^{-\frac{1}{2}}(\|a_n\|_\infty \|a_n\|_2 + 2h\|a_n\|_2) + 2(h^{-1}\|a_n\|_2^2 + \|a_n\|_1)) .$$

Proof: $P((\psi_n - \psi)^+)^2 = \int_0^h \psi^2 dP + \int_h^\infty P_x((\psi_n - \psi)^+)^2 dP$. Since the P-measure of the set where $k > 0$ is one, converting the second integral of the right hand side above to a μ-integral by introducing k in the integrand and applying the bound of Lemma 4.2, we obtain

$$(4.16) \quad \int_h^\infty P_x((\psi_n - \psi)^+)^2 dP \le n^{-\frac{1}{2}}\int_h^\infty ((1^+)ck^{\frac{1}{2}}h^{-\frac{1}{2}}(a_n^2+2ha_n)+2(h^{-1}a_n^2+a_n))\,d\mu .$$

Now extend the range of integration of this bound to $(0,\infty)$.

Then, since $a_n^2 \le \|a_n\|_\infty a_n$ and $\mu(k) = 1$, by the Schwarz Inequality the right hand side of (4.16) is bounded by

$$n^{-\frac{1}{2}}[(1^+)ch^{-\frac{1}{2}}(\|a_n\|_\infty \|a_n\|_2 + 2h\|a_n\|_2) + 2(h^{-1}\|a_n\|_2^2 + \|a_n\|_1)],$$

which completes the proof.

<u>Theorem 4.1</u>. Under (A1), if $a_n(x) \to \infty$ for each x, $h \to 0$ and

$$\|a_n\|_1 = o(n^{\frac{1}{2}}),$$

$$\|a_n\|_2 = o((nh^2)^{\frac{1}{4}}),$$

$$\|a_n\|_\infty = O((nh)^{\frac{1}{4}}),$$

then $P(\phi_n - \phi)^2 \to 0$ as $n \to \infty$.

<u>Proof</u>: Under (A1), $P(\psi^2) < \infty$ and it follows by the conditions of this theorem and Lemma 4.3 that $P((\psi_n - \psi)^+)^2 \to 0$. By Lemma 4.1, since $nh^2 \to \infty$, $P((\psi_n - \psi)^-)^2 \to 0$. Hence, $P(\phi_n - \phi)^2 = P(\psi_n - \psi)^2 \to 0$ as $n \to \infty$ and the theorem is proved.

We now consider the procedure of first estimating the prior G and then using the Bayes Estimator versus the estimate of G. As before let d denote the Lévy metric. Noting that K^∞ is the same measure on the space of <u>x</u>-sequences as $G^\infty(\underline{F})$, which is discussed immediately before Theorem 3.3, by Corollary 3.1 we can construct an estimator, $\hat{G}_n$, of G such that $d(\hat{G}_n, G) \to 0$ a.s. K^∞.

The following example shows that the risk of the Bayes Estimator versus a distribution function, converging in Lévy

metric to G, may not converge to Bayes risk.

Example 4.2. Let $0 < \beta < \infty$ be a continuity point of G with $G(\beta) = 1$ and $G(\theta) < 1$ for $\theta < \beta$. Let

$$\hat{G}_n(\theta) = \begin{cases} G(\theta) & \theta < b_n, \\ G(b_n) & b_n \le \theta < M_n, \\ 1 & \theta \ge M_n, \end{cases}$$

where $M_n \to \infty$, $b_n = \theta_n - c_n$ with $0 < c_n \le 1$ and $c_n \downarrow 0$ and θ_n satisfies the following conditions:

$$0 \le \theta_n \uparrow \beta,$$

$$G(\theta_n -) \le 1 - c_n.$$

Since $b_n \uparrow \beta$, $d(\hat{G}_n, G) \to 0$. For $x > 0$, let

$$\hat{\phi}_n(x) = \frac{1-\hat{G}_n(x)}{\int_{x+}^{\infty} \theta^{-1} d\hat{G}_n} [\int_{x+}^{\infty} \theta^{-1} d\hat{G}_n > 0],$$

which is a Bayes response versus $\hat{G}_n$. Let n be sufficiently large so that $b_n > 0$ and $M_n > \beta$. It follows that $K(\hat{\phi}_n^2) \ge \int [b_n, \beta) \hat{\phi}_n^2 dK = M_n^2 K([b_n, \beta))$. Since for $\theta_n \le \theta \le \beta$, $F_\theta([b_n, \beta)) \ge 1 - b_n \theta_n^{-1}$,

$$K([b_n, \beta)) = \int F_\theta([b_n, \beta)) dG \ge (1 - b_n \theta_n^{-1})(1 - G(\theta_n -)).$$

Thus, $K(\hat{\phi}_n^2) \ge M_n^2 (1 - b_n \theta_n^{-1}) c_n \ge \beta^{-1} (M_n c_n)^2$. Hence if $M_n c_n \to \infty$, since $G(\theta^2) < \infty$, by the triangle inequality for L_2- norm the ris

G(F

3.

s

risk of using $\hat{\phi}_n$ converges to infinity. However, since $G(\theta^2) < \infty$, the Bayes risk is finite.

3. <u>Uniform</u> $[\theta,\theta+1)$ <u>Case</u>.

We now consider the class of distributions considered in section 3 of Chapter III. For $\theta \in \Omega = (-\infty,\infty)$,

$$f_\theta(x) = [\theta \le x < \theta+1].$$

In this case, we have

$$k(x) = G(f_\theta(x)) = G(x) - G(x-1), \tag{4.17}$$

$$K(x) = G(F_\theta(x)) = G(x-1) + xk(x) - \int_{(x-1)+}^{x+} \theta dG, \tag{4.18}$$

where the affix + on the upper limit of the integral means to include the limit in the range of integration.

By the right continuity of G, k is right continuous and hence, for all x, k is the right hand derivative of K. Thus, we estimate $k(x)$ by

$$k_n(x) = h^{-1} K_n]_x^{x+h},$$

where $0 < h \le 1$ and h is allowed to depend on n and K_n is the empiric distribution function of $x_1, x_2, \ldots, x_n$ as in section 2 of this chapter. Note that central or left differences of K_n are good estimators of $k(x)$ when $K'(x) = k(x)$.

By equation (4.17), for all x,

$$G(x) = \sum_{j=0}^{\infty} k(x-j), \tag{4.19}$$

(4

Fo

(

(4.20) $$K(x) = \int_{-\infty}^{x} k(y)\,dy = \int_{x-1}^{x} G(y)\,dy.$$

For all x, define

(4.21) $$G_n^*(x) = \sum_{j=0}^{\infty} k_n(x-j).$$

Lemma 4.4. If $h \to 0$, for each x, $G_n^*(x) \to G(x)$ in K^∞-measure.

Proof: Let x be fixed. By (4.17),

$$\sum_{j=0}^{\infty} K\Big]_{x-j}^{x+h-j} = \sum_{j=0}^{\infty} \left(\int_{x-j}^{x-j+h} G(y)\,dy - \int_{x-j-1}^{x-j-1+h} G(y)\,dy\right).$$

Since the series on the right hand side is telescopic and $\int_{x-j}^{x+h-j} G(y)\,dy \to 0$ as $j \to \infty$, we have,

(4.22) $$\sum_{j=0}^{\infty} K\Big]_{x-j}^{x+h-j} = \int_{x}^{x+h} G(y)\,dy.$$

Since $G_n^*(x)$ is the average of n i.i.d. random variables, each distributed as $h^{-1}\Sigma_{j=0}^{\infty} [x-j < x_1 \le x-j+h]$ whose expectation is $h^{-1}\Sigma_{j=0}^{\infty} K]_{x-j}^{x-j+h}$ and since $nh^2 \to \infty$, by the Tchebichev Inequality, $G_n^*(x) - h^{-1}\Sigma_{j=0}^{\infty} K]_{x-j}^{x-j+h} \to 0$ in K^∞-measure. Hence, by (4.22) and the fact that the right continuity of G plus h converging to zero imply that $h^{-1}\int_x^{x+h} G(y)\,dy \to G(x)$, $G_n^*(x) \to G(x)$ in K^∞-measure, which completes the proof.

Since

$$P_x(\theta) = \frac{G(\theta f_\theta(x))}{G(f_\theta(x))},$$

by the equations describing the functions k and K, (4.17) and (4.18) respectively, Thus ,

(4.2

whe

Si

o

$$P_x(\theta) = \frac{G(x) + (x-1)k(x) - K(x)}{k(x)} .$$

Thus, we take as a Bayes response

$$\phi(x) = (x-1) + \psi(x), \tag{4.23}$$

where the function ψ is defined as follows:

$$\psi = \frac{G-K}{k} .$$

Since the conditional distribution of θ given x is concentrated on $(x-1,x]$, $0 \le \psi \le 1$. Define the function ψ_n by

$$\psi_n = \left(\frac{G_n^* - K_n}{k_n} \vee 0\right) \wedge 1 , \tag{4.24}$$

where $0/0$ is defined to be an arbitrary value in the interval $[0,1]$. We estimate $\phi(x)$ by

$$\phi_n(x) = (x-1) + \psi_n(x). \tag{4.25}$$

<u>Theorem 4.2</u>. With ϕ_n defined by (4.25), if $nh^2 \to \infty$ and $h \to 0$, $R_n \to R$ as $n \to \infty$.

<u>Proof</u>: Since the conditional distribution of θ given x is concentrated on $(x-1,x]$, $P(\phi - \theta) = P(P_x(\phi - \theta)^2) \le 1$ and $P(\phi_n - \theta) = P(P_x(\phi_n - \theta)^2) \le 1$. It follows that (4.1) holds and it is sufficient to show $P(\psi - \psi_n)^2 \to 0$. Let x be fixed. Since $nh^2 \to \infty$, using the Tchebichev and triangle inequalities as in the method used to obtain (4.8), in K^∞-measure,

$$k_n(x) \to k(x) .$$

By the Glivenko-Cantelli Theorem, page 20 of Loève (1963), a.s. K^∞,

$$K_n(x) \to K(x)$$

Hence, by Lemma 4.4 and the Slutsky Theorem, page 174 of Loève (1963), if $k(x) > 0$,

$$\psi_n(x) \to \psi(x)$$

in K^∞-measure. We note that $P(\{x|k(x) > 0\}) = 1$. Thus, by the Bounded Convergence Theorem, a.s. P, $P_x(\psi_n - \psi)^2 \to 0$. Again by the Bounded Convergence Theorem $P(P_x(\psi_n - \psi)^2) = P(\psi_n - \psi)^2 \to 0$, which completes the proof.

Note that the assumption $G(\theta^2) < \infty$ is not made in this case. This assumption is sufficient for $R < \infty$. However, in this case, $R \leq 1$ for any G as was shown in the proof of Theorem 4.2. The question arises as to whether or not the Bayes Estimator ϕ can have an infinite second moment if $R < \infty$. The following example shows that this is possible. Obviously, in this example, $G(\theta^2) = \infty$ since by Jensen's Inequality, $G(\theta^2) < \infty \Rightarrow P(\phi^2) < \infty$.

Example 4.3. Let G be concentrated on $I = \{1,2,3,\ldots\}$ such that the mass at $m \in I$ is $C\,m^{-3/2}$ where C is a normalizing constant. Since $\phi(x) = x-1 + \psi(x)$ and ψ is bounded above by 1 and below by 0, $P(\phi^2) = \infty$ iff $P(x^2) = \infty$. By (4.17)

T

$$P(x^2) = \int_1^\infty x^2 k(x)\,dx = \int_1^\infty x^2 (G(x) - G(x-1))\,dx.$$

Thus, by the definition of G,

$$P(x^2) \geq C \sum_{m=1}^{\infty} m^2 m^{-3/2} = \infty$$

and it follows that $P(\phi^2) = \infty$.

We now consider the technique of using the Bayes Estimator versus an estimate of the prior, G. By Corollary 3.2, we can construct an estimator, $\hat{G}_n$, of G such that $d(\hat{G}_n, G) \to 0$ a.s. K^∞. Let $\hat{G}_n$ be such an estimator and redefine

$$\phi_n(x) = \frac{\int_{(x-1)+}^{x+} \theta\, d\hat{G}_n}{\hat{G}_n\big]_{x-1}^{x}} \tag{4.26}$$

where $0/0$ is defined to be some value in the interval $[x-1, x]$. It then follows that for all x, $x-1 \leq \phi_n(x) \leq x$.

<u>Theorem 4.3</u>. With $\phi_n(x)$ defined by (4.26), $R_n - R \to 0$.

<u>Proof</u>: Let D be the discontinuity set of G and let $D + 1 = \{x+1 \mid x \in D\}$. Let $x \in A = \{x \mid x \notin D \cup (D + 1),\ k(x) > 0\}$ be fixed. Then, by the Helly-Bray Lemma, page 80 of Loève (1963), since $d(\hat{G}_n, G) \to 0$ a.s. K^∞, we have that a.s. K^∞,

$$\int_{(x-1)+}^{x+} \theta\, d\hat{G}_n \to \int_{x-1}^{x} \theta\, dG,$$

$$\hat{G}_n\big]_{x-1}^{x} \to G\big]_{x-1}^{x}.$$

Hence, $\phi_n(x) \to P_x(\theta)$ a.s. K^∞. Since $x-1 \leq P_x(\theta) \leq x$, by the Bounded Convergence Theorem, $P_x(\phi_n(x) - P_x(\theta))^2 = K^\infty(\phi_n(x) - P_x(\theta))^2 \to 0$.

He

P(

4.

fo

w

F

(

(

W

c

c

(

Re

(1

tr

of

to

Hence, since $P(A) = 1$, again by the Bounded Convergence Theorem, $P(\phi_n(x) - P_x(\theta))^2 \to 0$, which completes the proof.

4. Estimation of a Location Parameter in Certain Gamma Distributions.

Consider a family of distributions characterized by the following density with respect to Lebesgue measure, μ:

$$f_\theta(x) = \frac{(x-\theta)^{\alpha-1} e^{-(x-\theta)}}{\Gamma(\alpha)} [x \geq \theta]$$

with $\alpha \geq 1$, where $\theta \in \Omega = (-\infty, +\infty)$ and Γ represents the Gamma Function. Suppose that G is a distribution on Ω and assume

$$\text{(A1)} \qquad G(\theta^2) < +\infty.$$

In this case

$$\text{(4.27)} \qquad k(x) = G(f_\theta(x)) = \frac{1}{\Gamma(\alpha)} \int_{-\infty}^{x} (x-\theta)^{\alpha-1} e^{-(x-\theta)} dG.$$

We adopt the convention that the upper limit of an integral is included in the range of integration.

The Bayes Estimator in the problem of estimating θ based on an observation x, with quadratic loss, is

$$\text{(4.28)} \qquad \phi(x) = \frac{\int_{-\infty}^{x} \theta (x-\theta)^{\alpha-1} e^{-(x-\theta)} dG}{\Gamma(\alpha) k(x)} .$$

<u>Remark</u>. By part (ii) of the proposition of section 4 of Teicher (1961), a sufficient condition for identifiability of a class of translation parameter mixtures is that the characteristic function of the generating distribution function (take the location parameter to be zero) not be identically zero on a non-degenerate interval. In

this

trib

here

Lemm

ϕ(x

Pro

Inv

har

Th

pr

(4

We

(4

wh

and

bou

the

this case, the characteristic function of the generating distribution is $(1 - it)^{-\alpha}$. Hence, the family of mixtures considered here is identifiable.

Lemma 4.5. With ϕ defined in (4.28), if $k(x) > 0$, $\phi(x) = x - \alpha(k(x))^{-1}\int_{-\infty}^{x} e^{-(x-t)} dK(t)$.

Proof: By the definition of k in (4.27),

$$\alpha\int_{-\infty}^{x} e^{-(x-t)} dK(t) = (\Gamma(\alpha))^{-1}\alpha\int_{-\infty}^{x} e^{-(x-t)}\int_{-\infty}^{t}(t-\theta)^{\alpha-1}e^{-(t-\theta)} dG(\theta)\, dt.$$

Inverting the order of integration in the expression of the right hand side and performing the inner integration yields

$$(\Gamma(\alpha))^{-1}\int_{-\infty}^{x}(x-\theta)^{\alpha}e^{-(x-\theta)} dG(\theta) = \int_{-\infty}^{x}(x-\theta) f_{\theta}(x)\, dG(\theta).$$

This last expression is $xk(x) - k(x)\phi(x)$, which completes the proof.

Define

$$(4.29) \qquad \psi(x) = \frac{\int_{-\infty}^{x} e^{-(x-t)} dK(t)}{k(x)} .$$

We estimate $\psi(x)$ by

$$(4.30) \qquad \psi_n(x) = \frac{\int_{-\infty}^{x} e^{-(x-t)} dK_n(t)}{k_n(x)} \wedge a_n(x),$$

where K_n and k_n are defined as in section 3 of this chapter and h is positive and allowed to depend on n and a_n is a bounded non-negative function of x for each n. Also, define the estimator to be zero in the case of an undefined ratio. Let

(4.3

Note

(4.3

It

(4.

(4.

De

Le

Pr

i

S

w

Le

a
r

Pr

h

$$(4.31)\qquad \hat{\psi}(x) = \frac{\int_{-\infty}^{x} e^{-(x-t)}\,dK(t)}{h^{-1}K]_x^{x+h}}\ .$$

Note that by the definition of k, for any x and $\varepsilon > 0$,

$$(4.32)\qquad k(x+\varepsilon) \geq e^{-\varepsilon}k(x).$$

It then follows from (4.32) that

$$(4.33)\qquad h^{-1}K]_x^{x+h} \geq e^{-h}k(x).$$

Since $F_\theta(x^2) = \alpha(\alpha+1) + 2\theta\alpha + \theta^2$, by (A1),

$$(4.34)\qquad P(x^2) = G(F_\theta(x^2)) < \infty.$$

Define

$$A = \{x \mid k(x) > 0,\ k(x) = K'(x)\}.$$

<u>Lemma 4.6</u>. Under (A1), if $h \to 0$, $P(\hat{\psi} - \psi)^2 \to 0$.

<u>Proof</u>: Let $x \in A$. Then, since $h \to 0$, $\hat{\psi}(x) \to \psi(x)$. By (4.33), $|\hat{\psi}(x) - \psi(x)| \leq (e^h + 1)\psi(x)$. By (4.34) and the fact that (A1) implies that $P(\phi^2) < \infty$, since $\alpha\psi(x) = x - \phi(x)$, $P(\psi^2) < \infty$. Since $P(A) = 1$, by the Dominated Convergence Theorem, $P(\hat{\psi} - \psi)^2 \to 0$ which completes the proof.

<u>Lemma 4.7</u>. Under (A1), if $h \to 0$, $nh^2 \to \infty$ and for each x $a_n(x) \uparrow \infty$, $P((\psi_n - \hat{\psi})^-)^2 \to 0$.

<u>Proof</u>: Let $x \in A$ be fixed. By the Strong Law of Large Numbers, $\int_{-\infty}^{x} e^{-(x-t)}\,dK_n(t) \to \int_{-\infty}^{x} e^{-(x-t)}\,dK(t)$ a.s. K^∞. Since $nh^2 \to \infty$ and $h \to 0$, $k_n(x) \to k(x)$ in K^∞-measure by the method used to obtain

(4.8). Hence, since $a_n(x) \uparrow \infty$, $\psi_n(x) \to \psi(x)$ in K^∞-measure. Since $\hat{\psi}(x) \to \psi(x)$, $\psi_n(x) - \hat{\psi}(x) \to 0$ in K^∞-measure. By the bound on $\hat{\psi}(x)$ of (4.33), $(\{\psi_n(x) - \hat{\psi}(x)\}^-)^2 \le (e^h\psi(x))^2$. Hence, by the Bounded Convergence Theorem, $P_x(\{\psi_n(x) - \hat{\psi}(x)\}^-)^2 \to 0$. Since $P(A) = 1$ and under (A1), $P(\psi^2) < \infty$, by the Dominated Convergence Theorem $P((\psi_n - \hat{\psi})^-)^2 \to 0$ and the proof is complete.

<u>Lemma 4.8</u>. $P((\psi_n - \hat{\psi})^+)^2 \le 2e^h n^{-\frac{1}{2}}(c+1)\mu(a_n + h^{-1}a_n^2)$, where c is the Berry-Esseen constant.

<u>Proof</u>: Let $* = (a_n - \hat{\psi})^+ . P((\psi_n - \psi)^+)^2 = \int_0^* P_x(\psi_n - \hat{\psi} > t)dt^2$, since the integrand of the right hand side is zero for $t \ge *$. Fix $0 < t < *$ and suppose x is such that $k(x) > 0$. Then $[\psi_n - \hat{\psi} > t] = [\bar{w} > 0]$ where $w_i = [x_i \le x]e^{-(x-x_i)}$ $- (\hat{\psi} + t)h^{-1}[x < x_i \le x+h]$ for $i = 1,2,\ldots,n$. By (4.33),

$$P_x(\bar{w}) = P_x(w_1) = -th^{-1}K]_x^{x+h} \le -te^{-h}k(x). \tag{4.35}$$

Consider the following bounds on $\sigma^2 = V(w_1)$,

$$(th^{-1})^2K]_x^{x+h} \le ((\hat{\psi} + t)h^{-1})^2K]_x^{x+h} \le \sigma^2 \le (1 + h^{-1}a_n)^2. \tag{4.36}$$

By the Berry-Esseen Theorem, page 288 of Loève (1963) and (4.35), since the range of w_1 is bounded by $1 + h^{-1}a_n$,

$$P_x(\bar{w} > 0) \le \Phi(z) + c\, n^{-\frac{1}{2}}\sigma^{-1}(1 + h^{-1}a_n),$$

where $\sigma z = -n^{\frac{1}{2}}t\, e^{-h}k(x)$ and c is the Berry-Esseen constant. Applying a weakening of the bounds on the tails of the standard normal distribution of Feller (1957) and the bounds of (4.36),

we bound the right hand side of the above inequality by:

$$\frac{(1 + a_n h^{-1})e^h}{n^{\frac{1}{2}} t\ k(x)} + \frac{c(1 + a_n h^{-1})}{n^{\frac{1}{2}} t\ h^{-1} (K]_x^{x+h})^{\frac{1}{2}}} .$$

Hence, since $(K]_x^{x+h})^{\frac{1}{2}} \geq K]_x^{x+h}$ and by (4.33), $h^{-1} K]_x^{x+h} \geq e^{-h} k(x)$,

$$\int_0^* P_x(\bar{w} > 0)\,dt^2 \leq \frac{2a_n e^h (c+1)(1 + a_n h^{-1})}{n^{\frac{1}{2}} k(x)} .$$

Noting that $P(\{x | k(x) > 0\}) = 1$, converting the P-integral of the right hand side of the above inequality to a μ-integral we obtain, $P((\psi_n - \hat{\psi})^+)^2 \leq 2n^{-\frac{1}{2}} e^h (c+1)\mu(a_n + h^{-1} a_n^2)$, which completes the proof.

Define

$$\phi_n(x) = x - \alpha\psi_n(x) .$$

Theorem 4.4. If $\mu(a_n) = o(n^{\frac{1}{2}})$, $\mu(a_n^2) = o((nh^2)^{\frac{1}{2}})$, $h = o(1)$ and $a_n(x) \uparrow \infty$ for each x, then under (A1), $R_n - R \to 0$.

Proof: By (A1) and the fact that a_n is bounded for each n, the conditions implying (4.1) hold and thus it suffices to show that $P(\phi_n - \phi)^2 \to 0$. Since $a_n(x) \uparrow \infty$ for each x, $\mu(a_n^2) = o((nh^2)^{\frac{1}{2}})$ implies that $nh^2 \to \infty$. Hence, by Lemmas 4.7 and 4.8, $P(\psi_n - \hat{\psi})^2 \to 0$. By Lemma 4.6 and the triangle inequality for L_2-norm, $P(\psi_n - \psi)^2 \to 0$. Since for $x \in A$, $\phi_n - \phi = \alpha(\psi_n - \psi)$ and $P(A) = 1$, the proof is complete.

CHAPTER V

EMPIRICAL BAYES ESTIMATION IN EXPONENTIAL FAMILIES

1. A Rate for the Discrete Case.

Macky (1966) dealt with the Empirical Bayes Estimation Problem for the class of distributions considered in section 2 of Chapter II. The family is characterized by the following density with respect to μ, where μ is counting measure on the non-negative integers, $p_\theta(x) = \theta^x C(\theta) m(x)$, $m(x) > 0$ for $x = 0,1,2,\ldots$ and $\theta \in \Omega \subset (0,\infty)$. Let G be a distribution on Ω. As in Chapter III, let K be the marginal distribution of x of the pair (θ,x). Let $x_1, x_2, \ldots$ be a sequence of i.i.d. according to K random variables and let P be the product measure on the space of sequences $(x_1,x_2,\ldots,(\theta,x))$, i.e. P is the product of K^∞ and the joint measure of (θ,x).

Define $k(x) = G(p_\theta(x))$ which is the density of K with respect to μ. The Bayes estimator in the problem of estimating θ based on the observation x, assuming quadratic loss is:

$$\phi(x) = \frac{G(\theta^{x+1}C(\theta))}{G(\theta^x C(\theta))} = T(x)\,\frac{k(x+1)}{k(x)} ,$$

where $T(x) = m(x)/m(x+1)$.

Macky (1966) took Ω to be the natural parameter space for the family, i.e. $\{\theta | \mu(\theta^x m(x)) < \infty,\ \theta > 0\}$, and assumed $G(\theta^2) < \infty$. He then exhibited a procedure for estimating θ, based

risk

(A1)

whe

cer

whe

an

Ag

a

a

1

E

t

f

S

based on $(x_1,x_2,\ldots,x_n,x)$, whose risk converges to the Bayes risk versus G as $n \to \infty$.

In this section, we assume

$$\text{(A1)} \qquad \Omega = (0,\beta],\ \beta < \infty,$$

where $(0,\beta]$ is a subset of the natural parameter space. Under certain other assumptions, we show that $P(\phi_n - \phi)^2 = O(n^{-\frac{1}{2}})$, where

$$\phi_n(x) = T(x)\,\frac{k_n(x+1)}{k_n(x)} \wedge \beta$$

and

$$k_n(x) = n^{-1}\sum_{i=1}^{n}[x_i = x].$$

Again undefined ratios are taken to be zero. Let $R = P(\phi - \theta)^2$ and $R_n = P(\phi_n - \theta)^2$ as in Chapter III. Since $\beta < \infty$, $\phi \le \beta$ and $G(\theta^2) \le \beta^2$ and it follows that

$$R_n - R = P(\phi_n - \phi)^2.$$

__Lemma 5.1__. ϕ is an increasing function of x.

__Proof__: Define the measure G^* on Ω by $dG^*/dG = C(\theta)$. Note that G^* is a finite measure possessing all moments and that for $x = 0,1,2,\ldots,$

$$\phi(x) = \frac{G^*(\theta^{x+1})}{G^*(\theta^x)}\ .$$

Since $G^*(\theta^r)$, $r \ge 0$ is a log convex function of r, see 9.3b, pag

for

the

rate

in (

for

(A2

(A3

Le

P

f

o

R

th

(A

page 156 of Loève (1963),

$$(G^*(\theta^{x+1}))^2 \le G^*(\theta^x)G^*(\theta^{x+2}),$$

for $x = 0,1,2,\ldots$, which completes the proof.

We now make the following assumptions which are part of the group of assumptions made by Gilliland (1966) to obtain a rate of convergence to zero of the modified regret, discussed in Chapter II, in the sequential compound estimation problem for this family of distributions.

(A2) $$\sum_{x=0}^{\infty} p_\beta^{\frac{1}{2}}(x) < \infty.$$

(A3) $$\sum_{x=0}^{\infty} (T(x)p_\beta(x))^{\frac{1}{2}} < \infty.$$

Lemma 5.2. Under (A1), (A2) and (A3),

$$\sum_{x=0}^{\infty} k^{\frac{1}{2}}(x) < \infty,$$

$$\sum_{x=0}^{\infty} (T(x)k(x))^{\frac{1}{2}} < \infty.$$

Proof: The proof follows directly from the fact that under (A1) for $\theta \in \Omega$, $p_\theta(x) \le (m(0))^{-1}\beta^x m(x)$ which is a constant multiple of $p_\beta(x)$.

Remark. Gilliland (1966) mentions that β in the interior of the natural parameter space is sufficient for (A2) and (A2) and (A3) hold in the Poisson case.

Theo

P(¢

Proo

Noti

hand

and

w_i =

(5.1

Abbr

$\bar{k}$ =

Esse

(5.2

wher

(5.3

c i

r =

(5.4

Sinc

equi

Theorem 5.1. Under (A1), (A2) and (A3), uniformly in G,
$P(\phi_n - \phi)^2 = O(n^{-\frac{1}{2}})$.

Proof: $P(\phi_n - \phi)^2 \le \beta\, P(P_x(|\phi_n - \phi|))$ and

$$P_x(|\phi_n - \phi|) = \int_0^\infty P_x(\phi_n - \phi > v)\,dv + \int_0^\infty P_x(\phi_n - \phi < -v)\,dv.$$

Noting that the integrand of the first integral on the right hand side immediately above is zero for $v \ge *$ where $* = (\beta - \phi)^+$ and that for $0 < v < *$, $[\phi_n - \phi > v] = [\,\bar{w} > 0]$ where $w_i = T(x)[x_i = x+1] - (\phi + v)[x_i = x]$ for $i = 1,2,\ldots,n$, we have

$$\int_0^\infty P_x(\phi_n - \phi > v)\,dv = \int_0^* P_x(\bar{w} > 0)\,dv. \tag{5.1}$$

Abbreviate $T(x)$ and $k(x)$ by omission of x and let $\tilde{k} = k(x+1)$. Let $\sigma^2 = V(w_1)$ and let $0 < v < *$. By the Berry-Esseen Theorem, page 288 of Loève (1963), since $K(w_1) = -vk$,

$$P_x(\bar{w} > 0) \le \Phi(z_1) + cn^{-\frac{1}{2}}r\,\sigma^{-1} \tag{5.2}$$

where

$$\sigma z_1 = -n^{\frac{1}{2}}vk, \tag{5.3}$$

c is the Berry-Esseen constant and r is the range of w, i.e. $r = T + \phi + v$. By Lemma (1.1),

$$r\sigma^{-1} \le (\tilde{k}^{-1} + k^{-1})^{\frac{1}{2}}. \tag{5.4}$$

Since $\sigma^2 \le K(w_1^2) = T^2\tilde{k} + (\phi + v)^2 k$ and by (A1), $\phi \le \beta$ or equivalently $\tilde{k} \le T^{-1}\beta k$, for $0 < v < *$,

(5.5) $$\sigma^2 \le (T\beta + \beta^2)k.$$

Recalling the definition of z_1 in (5.3), replacing σ by the upper bound of (5.5) and then extending the range of integration from $*$ to ∞, we obtain

(5.6) $$\int_0^* \Phi(z_1)\,dv \le (2\pi nk)^{-\frac{1}{2}}(T\beta + \beta^2)^{\frac{1}{2}}.$$

Hence by (5.1), (5.2), (5.4) and (5.6), since $* \le \beta$,

(5.7) $$\int_0^\infty P_x(\phi_n - \phi > v)\,dv \le (2\pi nk)^{-\frac{1}{2}}(T\beta + \beta^2)^{\frac{1}{2}} + \beta c n^{-\frac{1}{2}}(\tilde{k}^{-1} + k^{-1})^{\frac{1}{2}}.$$

Letting $u_i = T[x_i = x+1] - (\phi - v)[x_i = x]$ for $i = 1,2,\dots,n$, noting that $P_x(\phi_n - \phi < -v) = 0$ for $v \ge \phi$ and that for $0 < v < \phi$, $[\phi_n - \phi < -v] \le [\bar{u} \le 0]$, we have for $0 < v < \phi$,

(5.8) $$P_x(\phi_n - \phi < -v) \le P_x(\bar{u} \le 0).$$

Let $0 < v < \phi$ and $s^2 = V(u_1)$. By (1.1), $s^2 \le T^2\tilde{k}(1 - \tilde{k}) + 2T(\phi - v)k\tilde{k} + (\phi - v)^2 k(1 - k)$. Since $\tilde{k} \le k\,\phi\,T^{-1}$, letting $v = 0$, we bound the right hand side of this last inequality by $k(\phi T(1 - \tilde{k}) + 2\phi^2 k + \phi^2(1 - k))$. It then follows that for all $0 < v < \phi$,

(5.9) $$s^2 \le k(2\phi^2 + \phi T).$$

By the Berry-Esseen Theorem again, since $K(u_1) = v\,k$ and by

Lemma 1.1, $rs^{-1} \le (\tilde{k}^{-1} + k^{-1})^{\frac{1}{2}}$ where r is the range of u_1,

$$(5.10) \qquad P_x(\bar{u} \le 0) \le \Phi(z_2) + c\, n^{-\frac{1}{2}}(\tilde{k}^{-1} + k^{-1})^{\frac{1}{2}},$$

where $sz_2 = -n^{\frac{1}{2}}vk$. By the same method used to obtain (5.6), using the upper bound of (5.9) for s, we obtain

$$(5.11) \qquad \int_0^{\phi} \Phi(z_2)\,dv \le (2\pi nk)^{-\frac{1}{2}}(2\phi^2 + \phi T)^{\frac{1}{2}}.$$

By (5.8), (5.10) and (5.11)

$$(5.12) \qquad \int_0^{\phi} P_x(\phi_n - \phi < -v)\,dv \le (2\pi nk)^{-\frac{1}{2}}(2\phi^2 + \phi T)^{\frac{1}{2}} + \phi c n^{-\frac{1}{2}}(\tilde{k}^{-1} + k^{-1})^{\frac{1}{2}}.$$

Combining (5.7) and (5.12), replacing ϕ by β, since the square root of a sum of positive quantities is no greater than the sum of the individual square roots we have,

$$(5.13) \qquad P_x(|\phi_n - \phi|) \le n^{-\frac{1}{2}}(\beta^{\frac{1}{2}}(2\pi)^{-\frac{1}{2}}A + c\beta\, D)$$

$$A = k^{-\frac{1}{2}}(2T^{\frac{1}{2}} + \beta^{\frac{1}{2}}(1 + \sqrt{2})),$$

$$D = \tilde{k}^{-\frac{1}{2}} + k^{-\frac{1}{2}}.$$

By Lemma 5.1, $k\,\tilde{k}^{-1} \le T\phi(0)$. Thus,

$$(5.14) \qquad P(D) \le \mu((T\phi(0))^{\frac{1}{2}}k^{\frac{1}{2}} + k^{\frac{1}{2}}).$$

By Lemma 5.2, the right hand side of (5.14) is finite. Also, by Lemma 5.2 $P(A) = \mu\{k^{\frac{1}{2}}(2T^{\frac{1}{2}} + \beta^{\frac{1}{2}}(1 + \sqrt{2}))\} < \infty$, so the P-

integral of the right hand side of (5.13) is $O(n^{-\frac{1}{2}})$ and the proof is complete.

The following example indicates that merely assuming that the parameter space is bounded, i.e. (A1), is not sufficient for obtaining a rate of convergence for the estimator ϕ_n.

<u>Example 5.1</u>. We again consider the family due to Gilliland (1966) that appeared in Example 2.1. Let $\Omega = (0,1]$, $m(0) = 1$, $m(x) = x^{-3}$ for $x = 1,3,5,\ldots$ and $m(x) = a(x) \le x^{-3}$ for $x = 2,4,\ldots$. Let G concentrate all mass at $\theta = 1$. Then making use of the fact that undefined ratios are zero,

$$R_n - R = P(\phi_n - \phi)^2 \ge \sum_{x=0}^{\infty} p_1(x)(1 - p_1(x+1))^n.$$

Since $p_1(x) = C(1)m(x)$,

$$R_n - R \ge \sum_{x=0}^{\infty} C(1)(2x+1)^{-3}(1 - C(1)a(2x+2))^n.$$

Proceeding exactly as Gilliland we see that $R_n - R$ dominates a positive null sequence which decreases arbitrarily slowly by choice of $a(x)$. This example also indicates that bounding the parameter space away from zero as well as above will not be sufficient for obtaining a rate. Also, if undefined ratios were defined otherwise, a slight modification of this example would yield the same result.

Tucker (1963) demonstrates a method of estimating G in the Poisson Case, i.e. $p_\theta(x) = \theta^x e^{-\theta}/x!$. He exhibits a distribution function concentrated on $(0,\infty)$, based on observations integral of the right hand side of (5.13) is $O(n^{-1/2})$ and the proof is complete.

The following example indicates that merely assuming that the parameter space is bounded, i.e. (A1), is not sufficient for obtaining a rate of convergence for the estimator ϕ_n.

Example 5.1. We again consider the family due to Gilliland (1966) that appeared in Example [illegible]. Here $\Omega = [0,1]$, $m(0) = 1$, $m(x) = x^{-3}$ for $x = 1,3,5,\ldots$, and $m(x) = a(x) \leq x^{-3}$ for $x = 2,4,\ldots$. Let G concentrate all mass at $\theta = 1$. Then making use of the fact that undefined ratios are zero,

$$R_n - R = \int (\phi_n - \theta)^2 \, dG \geq \sum_{x=0}^{\infty} p_1(x)(1 - p_1(x+1))^n .$$

Since $p_1(x) = C(1)m(x)$

$$R_n - R \geq \sum_{x=0}^{\infty} C(1)(2x+1)^{-3}(1 - C(1)a(2x+2))^n .$$

Proceeding exactly as Gilliland we see that $R_n - R$ dominates a positive null sequence which decreases arbitrarily slowly by choice of $a(x)$. This example also indicates that bounding the parameter space away from zero as well as above will not be sufficient for obtaining a rate. Also, if undefined ratios were defined otherwise, a slight modification of this example would yield the same result.

Tucker (1963) demonstrates a method of estimating G in the Poisson Case, i.e. $p_\theta(x) = \theta^x e^{-\theta}/x!$. He exhibits a distribution function concentrated on $(0,\infty)$, based on observations

$x_1, x_2, \ldots, x_n$ which are i.i.d. according to K, which converges a.s. K^∞ to G, on the continuity set of G, as $n \to \infty$.

In order to apply the exact same estimation procedure to the general exponential family on the non-negative integers, the following condition, say (A), would have to be satisfied:

(A) there exists $r > 0$ such that for $|z| < r$,

$$\sum_{x=0}^{\infty} z^x (x!)^{-1} (\int_0^\infty \theta^x dQ) < \infty$$

where

$$\frac{dQ}{dG} = \frac{C(\theta)}{k(0)} m(0).$$

Let us assume, as in the previous theorem, that G is concentrated on $(0,\beta]$, $\beta < \infty$. It then follows that condition (A) is satisfied with $r = \beta^{-1}$. Let $\hat{G}_n$ be the Tucker estimate of G, modified to meet this more general context. Since $C(\theta)\theta^x < m^{-1}(x)$, by the Helly-Bray Theorem, page 182 of Loève (1963), $\hat{G}_n(\theta^x C(\theta)) \to G(\theta^x C(\theta))$ a.s. K^∞ for each x. Thus, for each fixed x, a.s. K^∞

$$\hat{\phi}_n(x) \to \frac{G(\theta^{x+1}C(\theta))}{G(\theta^x C(\theta))} = \phi(x),$$

where we define

$$\hat{\phi}_n(x) = \frac{\hat{G}_n(\theta^{x+1}C(\theta))}{\hat{G}_n(\theta^x C(\theta))} \wedge \beta.$$

It then follows that $\hat{\phi}_n \to \phi$ a.s. P and by the Bounded Convergence Theorem, $P(\hat{\phi}_n - \phi)^2 \to 0$, so that the risk of using the Bayes Estimator versus $\hat{G}_n$ converges to the Bayes Risk versus G.

2. Estimation in the Presence of a Nuisance Parameter.

Consider a bivariate random variable with a discrete and a continuous component. Let the distribution depend on a two-dimensional parameter with one component pertaining to the discrete variable and the other to the continuous variable. In this section an empirical Bayes estimation procedure is given for the problem of estimating the discrete component of the parameter, under the assumption of quadratic loss.

Let $z = (x,y) \in \{0,1,2,\ldots\} \times (-\infty,+\infty)$. Let

$$p_\eta(z) = C_1(\theta)C_2(\xi)\theta^x e^{\xi y} m(x)r(y), \tag{5.15}$$

where $\eta = (\theta,\xi) \in \Omega$, Ω being the natural parameter space, i.e. $\Omega = \{(\theta,\xi) | \theta > 0, \sum_{x=0}^{\infty}\theta^x m(x) < \infty, \int e^{\xi y} r(y)\,dy < \infty\}$ and m and r are positive, be the density with respect to $\mu = \mu_1 \times \mu_2$, where μ_1 is counting measure on the non-negative integers and μ_2 is Lebesgue measure. Let $C(\eta) = C_1(\theta)C_2(\xi)$.

Let G be a distribution on Ω such that

$$G(\xi^2) < \infty, \tag{A1}$$

$$G(\theta^2) < \infty. \tag{A2}$$

Let P be the usual product measure on the space of sequences $(z_1,z_2,\ldots,(z,\theta))$ where z_i, $i = 1,2,\ldots$, are i.i.d. with density

$$k(z) = G(p_\eta(z)) = r(y)m(x)G(C(\eta)\theta^x e^{\xi y}) \tag{5.16}$$

with respect to μ and (z,θ) has the usual joint distribution.

The Bayes Estimator in the problem of estimating θ based on the observation z, under quadratic loss, is

$$(5.17)\qquad \phi(z) = \frac{G(\theta p_\eta(z))}{G(p_\eta(z))} = T(x)\,\frac{k(x+1,y)}{k(x,y)},$$

where

$$T(x) = \frac{m(x)}{m(x+1)}.$$

We now assume that r is twice differentiable and that

$$(A3)\qquad r, |r'|, |r''|$$

are bounded functions. By (A3) and Theorem 9, page 52 of Lehmann (1959), for each fixed x, $k(x,y)$ is continuous and possesses at least its first two derivatives with respect to y. Thus for each fixed x, for all y, letting $F_x(y) = \int_{-\infty}^{y} k(x,t)\,dt$,

$$\frac{dF_x(y)}{dy} = F_x'(y) = k(x,y).$$

Thus, by (5.17),

$$(5.18)\qquad \phi(z) = T(x)\,\frac{F_{x+1}'(y)}{F_x'(y)}.$$

Let the operator $\mathfrak{h}$ be defined as in Chapter II by

$$\mathfrak{h}\, g(y) = (2h)^{-1} g\Big]_{y-h}^{y+h},$$

where $1 \geq h > 0$ and g is any real-valued function of a real variable y. Define

$$(5.19) \qquad t_n(z) = T(x)\left(\frac{\blacklozenge F^*_{x+1}(y)}{\blacklozenge F^*_x(y)} \wedge a_n \right) ,$$

where a_n is a non-negative constant and

$$F^*_x(y) = n^{-1} \sum_{i=1}^{n} [x_i = x, y_i \le y].$$

We consider undefined ratios to be zero. Let N_n be a sequence of non-negative integers increasing to ∞.

Define

$$A_n = \{z \mid 0 \le x \le N_n, \ -N_n \le y \le N_n\},$$

$$A_n^+ = \{z \mid 0 \le x \le N_n + 1, \ -N_n - 1 \le y \le N_n + 1\},$$

$$D_n = \{\eta \mid -N_n \le \log \theta \le N_n, \ -N_n \le \xi \le N_n\},$$

$$r_n = \inf \{r(y) \mid -N_n - 1 \le y \le N_n + 1\},$$

$$m_n = \inf \{m(x) \mid 0 \le x \le N_n + 1\}.$$

Note that $m_n > 0$ and that $r_n > 0$ since r is positive and continuous.

Macky (1966) shows that with

$$B(y) = \sup_{\Omega_\xi} C_2(\xi) e^{\xi y},$$

where $\Omega_\xi = \{\xi \mid \int e^{\xi y} r(y)\,dy < \infty\}$, $B(y) < \infty$ for all y and that

$$\sup_{z \in A_n^+} B(y) = \max\,(B(-N_n - 1), B(N_n + 1)).$$

It then follows that for $z \in A_n^+$,

$$k(z) \le B_n, \tag{5.20}$$

where B_n is the product of the max $(B(-N_n-1),\ B(N_n+1))$ and the bound on r guaranteed by (A3). Since $k(z) \ge r(y)m(x)\int_{D_n} C(\eta)\theta^x e^{\xi y}dG$, for $z \in A_n^+$,

$$k(z) \ge c_n, \tag{5.21}$$

where $c_n = r_n m_n d_n \exp(-2(N_n + 1)^2)$ and $d_n = \int_{D_n} C(\eta)dG$ which converges to $\int C(\eta)dG > 0$ as $n \to \infty$.

Define

$$\psi(z) = \frac{\phi F_{x+1}(y)}{\phi F_x(y)} . \tag{5.22}$$

Lemma 5.3. For $z \in A_n$, $\psi(z) \le B_n c_n^{-1}$.

Proof: The proof follows immediately from the bounds, (5.20) and (5.21), on $k(z)$ for $z \in A_n^+$ and the definition of ψ in (5.22).

Define

$$\phi_n(z) = t_n(z)[z \in A_n], \tag{5.24}$$

$$\hat{t}_n(z) = \frac{\phi F^*_{x+1}(y)}{\phi F^*_x(y)} , \tag{5.25}$$

where t_n is defined by (5.19). As usual let $R = P(\phi - \theta)^2$, the Bayes risk and $R_n = P(\phi_n - \theta)^2$. By (A2) and the fact that ϕ_n is bounded for each n, we have that $R_n - R = P(\phi_n - \phi)^2$.

<u>Lemma 5.4</u>. For $z \in A_n$, $b > 0$,

$$P_z\{(\hat{t}_n - \psi)^* > b\} \le \exp\left(-\frac{8nh^2b^2c_n^4}{((1+b)c_n + B_n)^2}\right),$$

where $*$ denotes $+$ or $-$.

<u>Proof</u>: $[\hat{t}_n - \psi > b] = [\bar{v} > 0]$ where for $i = 1,2,\ldots,n$,

$v_i = [x_i = x+1, y-h < y_i \le y+h] - (\psi+b)[x_i = x, y-h < y_i \le y+h]$.

Hence, since $P_z(\bar{v}) = -b\, F_x]_{y-h}^{y+h}$ and the range of v_1 is $1 + (\psi+b)$, applying Theorem 2 of Hoeffding (1963) to $P_z(\bar{v} > 0)$,

$$(5.26) \qquad P_z(\hat{t}_n - \psi > b) \le \exp\left(\frac{-2n(b\, F_x]_{y-h}^{y+h})^2}{(1 + \psi + b)^2}\right).$$

Also, $[\hat{t}_n - \psi < -b] \le [\bar{w} \ge 0]$ where for $i = 1,2,\ldots,n$,

$w_i = (\psi - b)[x_i = x, y-h < y_i \le y+h] - [x_i = x+1, y-h < y_i \le y+h]$.

Since $\bar{w}$ has the same expectation as $\bar{v}$ and the range of w_1 is smaller than the range of v_1, the bound of the right hand side of (5.26) applies also to $P_z(\hat{t}_n - \psi < -b)$. In the right hand side of (5.26), we replace $F_x]_{y-h}^{y+h}$ by the lower bound, $2h\, c_n$, given in (5.21) and ψ by the upper bound, $c_n^{-1}B_n$, given in Lemma 5.3 and obtain the bound of this lemma.

Letting $k''(x,y)$ denote the second partial with respect to y,

$$(5.27) \qquad k''(x,y) = m(x)\{r(y)G^*(\xi^2 e^{\xi y}) + 2r'(y)G^*(\xi e^{\xi y}) + r''(y)G^*(e^{\xi y})\},$$

where

$$\frac{dG^*}{dG} = C(\eta)\theta^x.$$

Converting the G^*-integrals in the right hand side of (5.27) back to G-integrals by introducing $C(\eta)\theta^x$ into the integrands as a multiplicative factor, noting that $C_1(\theta)\theta^x m(x) < 1$ for all x and θ and recalling the definition of B_n, we see that for $z \in A_n^+$, by (A1), (A3) and the triangle inequality the absolute value of the right hand side of (5.27) is bounded by $L\ B_n$, where L is a constant depending on $G(|\xi|)$, $G(\xi^2)$ and the bounds on r, $|r'|$ and $|r''|$. Hence for $z \in A_n^+$,

(5.28) $$|k''(x,y)| \le L\ B_n.$$

Lemma 5.5. For $z \in A_n$,

$$\left|\psi - \frac{k(x+1,y)}{k(x,y)}\right| \le \frac{h^2 L\ B_n^2}{3\ c_n^2} .$$

Proof: For fixed x, since $\phi\ F_x(y) = (2h)^{-1}\int_{y-h}^{y+h} k(x,t)\,dt$, expanding $k(x,t)$ by a Taylor Series about y yields

$$\phi\ F_x(y) = (2h)^{-1}\int_{y-h}^{y+h} (k(x,y) + (t-y)k'(x,y) + 2^{-1}(t-y)^2 k''(x,s))\,dt,$$

where s depends on t and $y-h \le s \le y+h$. Distributing the integration we see that the right hand side of the above equation equals $k(x,y) + (2h)^{-1}\int_{y-h}^{y+h} 2^{-1}(t-y)^2 k''(x,s)\,dt$. By (5.28), for $z \in \{(x,y)|x = 0,1,2,\ldots,N_n+1,\ -N_n \le y \le N_n\}$, the second term of

where

$$\frac{dG^*}{dG} = C(\eta)\theta^*$$

Converting the G^*-integrals in the right hand side of (5.27) back to G-integrals by introducing $C(\eta)\theta^*$ into the integrands as a multiplicative factor, noting that $C_1(\theta)\theta^* m(x) < 1$ for all x and θ and recalling the definition of B_n we see that for $z \in A_n^+$, by (A1), (A3) and the Schwarz inequality the absolute value of the right hand side of (5.27) is bounded by $L B_n$, where L is a constant depending on $G(|\xi|)$, $G(\xi^2)$ and the bounds on r, $|r'|$ and $|v''|$. Hence for $z \in A_n^+$,

(5.28) $$|k''(x,y)| \le L B_n .$$

Lemma 5.5. For $z \in A_n$,

$$\left| \phi - \frac{k(x+1,y)}{k(x,y)} \right| \le \frac{h^2 L B_n^2}{3 c_n^2}$$

Proof: For fixed x, since $\phi T_x(y) = (2h)^{-1}\int_{y-h}^{y+h} k(x,t)dt$, expanding $k(x,t)$ by a Taylor Series about y yields

$$\phi T_x(y) = (2h)^{-1}\int_{y-h}^{y+h}(k(x,y) + (t-y)k'(x,y) + 2^{-1}(t-y)^2 k''(x,s))dt,$$

where s depends on t and $y-h \le s \le y+h$. Distributing the integration we see that the right hand side of the above equation equals $k(x,y) + (2h)^{-1}\int_{y-h}^{y+h} 2^{-1}(t-y)^2 k''(x,s)dt$. By (5.28), for $z \in \{(x,y) | x = 0,1,2,\ldots,N_n+1, -N_n \le y \le N_n\}$, the second term of this sum is bounded in absolute value by $h^2 L B_n/6$. Hence for $z \in A_n$,

(5.29) $|\Delta F_{x+1}(y)k(x,y) - k(x+1,y)\Delta F_x(y)| \le (k(x,y)+k(x+1,y))h^2 B_n L/6.$

Since $|\psi - (k(x,y))^{-1}k(x+1,y)|$ equals the product of the left hand side of (5.29) and $(\Delta F_x(y)k(x,y))^{-1}$ and $\Delta F_x(y) = k(x,a)$ where $y - h \le a \le y + h$, replacing k by the upper bound of (5.20) in the right hand side of (5.29) and then using the bound of (5.21) to obtain a lower bound of c_n^2 for $(k(x,a)k(x,y))^{-1}$, we see that for $z \in A_n$,

$$\left| \psi - \frac{k(x+1,y)}{k(x,y)} \right| \le \frac{h^2 B_n^2 L}{3\, c_n^2},$$

which completes the proof.

We choose the sequences a_n, h and N_n so that the following conditions are satisfied:

(i) $$\frac{B_n}{a_n c_n} \to 0,$$

(ii) $$\frac{a_n}{n^{\frac{1}{2}} h c_n} \to 0,$$

(iii) $$\frac{hB_n}{c_n} \to 0.$$

Note that since d_n converges to a positive constant, these conditions are implied by the set of conditions formed by replacing c_n by $c_n d_n^{-1}$, which is a quantity independent of G. Hence, it is possible to choose a_n, h and N_n independent of G.

We also assume

$$\text{(A4)} \qquad P(T^2) < \infty.$$

Theorem 5.2. If conditions (i), (ii) and (iii) are satisfied, then under (A1)-(A4), $P(\phi_n - \phi)^2 \to 0$ as $n \to \infty$, where ϕ and ϕ_n are defined by (5.17) and (5.24) respectively.

Proof: $\int (\phi_n - \phi)^2 dP = \int_{A_n} (\phi_n - \phi)^2 dP + \int_{A_n'} \phi^2 dP$. (A2) implies $P(\phi^2) < \infty$, so $\int_{A_n'} \phi^2 dP \to 0$, since N_n increases to infinity. Thus, by (5.24), it remains to be shown that $\int_{A_n} (t_n - \phi)^2 dP \to 0$. Let $z \in A_n$.

$$P_z(t_n - T\psi)^2 = T^2 \int_0^\infty P_z\{(\hat{t}_n \wedge a_n) - \psi)^2 > b\} db.$$

By Lemma 5.3 and (i), for n sufficiently large for $z \in A_n$, $\psi(z) \leq a_n$, so that the range of integration of the above right hand side can be reduced to $(0, a_n^2)$. It also follows that for large n we can remove the truncation of $\hat{t}_n$ at a_n and apply Lemma 5.4 to obtain the following asymptotic bound on the integrand of the right hand side of the above equation:

$$2 \exp\left(\frac{-8nh^2 b\, c_n^4}{((1+\sqrt{b})c_n + B_n)^2} \right).$$

Replacing $(1+\sqrt{b})$ by $(1+a_n)$ and integrating the resulting expression over the range $(0,\infty)$, we obtain

$$\text{(5.30)} \qquad P_z(t_n - T\psi)^2 \leq \frac{T^2((1+a_n)c_n + B_n)^2}{4\, nh^2 c_n^4}.$$

By Lemma 5.5, since $z \in A_n$,

$$(5.31) \qquad P_z(T\psi - \phi)^2 \le T^2 \left(\frac{h^2 L B_n^2}{3 c_n^2} \right)^2 .$$

By the Minkowski Inequality,

$$(5.32) \qquad P_z(t_n - \phi)^2 \le \{ (P_z(t_n - T\psi)^2)^{\frac{1}{2}} + (P_z(T\psi - \phi)^2)^{\frac{1}{2}} \}^2 .$$

By (i), (ii), (iii) and (A4) the P-integral over the set A_n of the right hand sides of (5.30) and (5.31) converge to zero as $n \to \infty$. Thus, by the Schwarz Inequality the P-integral over A_n of the right hand side of (5.32) converges to zero and it follows that $\int_{A_n} (t_n - \phi)^2 dP \to 0$ which completes the proof.

<u>Example 5.2</u>. This example points out that Theorem 5.2 applies to the Poisson-Normal $(\xi, 1)$ case. In this case $r(y) = e^{-\frac{1}{2}y^2}$ so (A3) is satisfied. Let G be such that (A2) is satisfied. $T(x) = x+1$ and $P(x+1)^2 = G(P_\theta(x+1)^2)$ where P_θ denotes the Poisson distribution with parameter θ. Since $P_\theta(x+1)^2 = \theta^2+3\theta+1$ and (A2) holds, (A4) holds. Also in this case,

$$r_n = \exp\,(-(N_n+1)^2/2),$$

$$m_n = ((N_n + 1)!)^{-1}$$

and

$$B_n = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{(N_n+1)^2}{2} \right).$$

BIBLIOGRAPHY

BIBLIOGRAPHY

Deely, J.J. and Kruse, R.L. (1968). Construction of Sequences Estimating the Mixing Distribution. Ann. Math. Statist., 39, 286-288.

Feller, W. (1957). An Introduction to Probability Theory and its Applications Volume I (2nd ed.), Wiley, New York.

Gilliland, D. (1966). Approximation to Bayes Risk in Sequences of Non-Finite Decision Problems. Tech. Report No. 10, Department of Statistics and Probability, Michigan State University.

Gilliland, D. and Hannan, J. (1968). The Role of Normal Approximation in Compound Decision Theory Problems. Research Memorandum RM-218, Department of Statistics and Probability, Michigan State University.

Hoeffding, W. (1963). Probability Inequalities for Sums of Bounded Random Variables. J. Amer. Statist. Assoc., 58, 13-30.

Johns, M.V. (1966). Two-Action Compound Decision Problems. Tech. Report No. 87, Department of Statistics, Stanford.

Lehmann, E.L. (1959). Testing Statistical Hypotheses. Wiley, New York.

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

Macky, D. (1966). Empirical Bayes Estimation in an Exponential Family. Research Memorandum RM-176, Department of Statistics and Probability, Michigan State University.

Parzen, E. (1959). Convergence of Families of Sequences of Random Variables. California University Publications in Statistics, 2, 23-53.

Robbins, H. (1964). The Empirical Bayes Approach to Statistical Decision Problems. Ann. Math. Statist., 35, 1-20.

Teicher, H. (1961). Identifiability of Mixtures. Ann. Math. Statist., 32, 244-248.

Tucker, H. (1963). An Estimate of the Compounding Distribution of a Compound Poisson Distribution. Theory of Probability and its Applications, 8, 195-200.

Wolfowitz, J. (1953). Estimation by the Minimum Distance Method. Ann. Inst. Math. Statist., V, 9-23.