APPROXIMATION TO BAYES RISK IN COMPOUND DECISION PROBLEMS

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
ALLAN OATEN
1969

This is to certify that the

thesis entitled

APPROXIMATION TO BAYES RISK IN
COMPOUND DECISION PROBLEMS

presented by

Allan Oaten

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics and Probability

James Hannan
Major professor

Date August 5, 1969

O-169

ABSTRACT

APPROXIMATION TO BAYES RISK IN
COMPOUND DECISION PROBLEMS

By

Allan Oaten

The set version of the compound decision problem involves simultaneous consideration of N statistical decision problems, called the component problems, with identical generic structure: state space Ω, action space A, sample space $\mathfrak{X}$ and non-negative loss function L defined on $\Omega \times A \times \mathfrak{X}$. With $\underline{x} = (x_1, x_2, \ldots, x_N)$ distributed according to $\prod_{i=1}^{N} P_{\theta_i} = P_{\underline{\theta}}$, a compound procedure is a vector, $\underline{\varphi} = (\varphi_1, \ldots, \varphi_N)$ such that, for each i, $\varphi_i: \mathfrak{X}^N \to A$.

The conditional risk, given $\underline{x}$, of the procedure $\underline{\varphi}$ is $W(\underline{\theta}, \underline{\varphi}, \underline{x}) = N^{-1} \sum_{r=1}^{N} L(\theta_r, \varphi_r(\underline{x}), x_r)$, the unconditional risk is $R(\underline{\theta}, \underline{\varphi}) = \int W(\underline{\theta}, \underline{\varphi}) dP_{\underline{\theta}}$, and the modified regret is $D(\underline{\theta}, \underline{\varphi}) = R(\underline{\theta}, \underline{\varphi}) - R(G_N)$ where G_N is the empirical distribution of $\theta_1, \theta_2, \ldots, \theta_N$ and $R(\cdot)$ is the Bayes envelope in the component problem.

For F a distribution on Ω, Φ_F is the set of procedures in the component problem which are ε-Bayes against F. Let $\hat{G}$ be an estimator of G_N - i.e. for each $\underline{x} \in \mathfrak{X}^N$, $\hat{G}(\underline{x})$ is a distribution on Ω and let $\Phi_{\hat{G}}$ be the set of compound procedures $\underline{\varphi}$ such that, for each $\underline{x}$, there is an element, $\varphi^o(\underline{x})$, of $\Phi_{\hat{G}(\underline{x})}$ such that $\varphi_r(\underline{x}) = \varphi^o(\underline{x})(x_r)$ for each r. Thus, to use a procedure in $\Phi_{\hat{G}}$ one first estimates G_N by $\hat{G}(\underline{x})$ and then plays ε-Bayes against $\hat{G}(\underline{x})$ in each component problem.

We consider two subsets of $\Phi_{\hat{G}}$, the "half-space" procedures and the "equivariant uniformly ϵ-Bayes" procedures. For the $m \times n$ problem (i.e. Ω has m elements, A has n) we establish the uniform almost sure convergence of $W(\underline{\theta},\underline{\varphi},\underline{x})$ to $R(G_N)$ for half-space procedures if $\hat{G}$ is "uniformly strongly consistent"; and if $\hat{G}$ is "uniformly consistent" we establish $D(\underline{\theta},\underline{\varphi}) < o(1) + \epsilon$ uniformly as $N \to \infty$ for both types. For the $m \times \infty$ problem, we again establish $D(\underline{\theta},\underline{\varphi}) < o(1) + \epsilon$ for the equivariant procedures.

We also consider the problem when Ω is infinite. We consider a class of procedures that differs from $\Phi_{\hat{G}}$ in that the corresponding component procedures are defined in the elements of a finite partition, $\mathcal{V}$, of $\mathfrak{X}$, rather than $\mathfrak{X}$ itself. Assuming the existence of "good" estimators $\hat{G}$, we establish results similar to, but slightly weaker than, those for finite state spaces, provided $\mathcal{V}$ is a "good approximation" to $\mathfrak{X}$ in respect of both the distributions P_ω and the loss function. We give conditions under which this latter requirement holds, and show that these conditions are satisfied by wide classes of problems.

APPROXIMATION TO BAYES RISK IN COMPOUND DECISION PROBLEMS

By

Allan Oaten

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

1969

TO MY FAMILY

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to Professor James Hannan, who introduced me to the Compound Decision Problem and guided me through this research. His suggestions have been the source of many of the results presented here and have improved virtually all results, either in substance or in form, almost beyond recognition. His comments helped me avoid many wrong turnings.

In addition, among many others who helped, I would like to thank Professor Dennis Gilliland for some useful discussions and for reading and commenting on a difficult rough draft; Mrs. Noralee Barnes, who typed both the rough and the final versions of this thesis, and whose speed, accuracy and cheerfulness leave me, at times, unnerved; my wife and son for their heroic patience and endurance; and Messrs. S. Guthery and F. Ruiz for several helpful discussions.

Finally I would like to thank the Department of Statistics and Probability at Michigan State University for the generous support, financial and otherwise, that I have enjoyed over the period of my graduate studies.

TABLE OF CONTENTS

0. INTRODUCTION AND NOTATION

In this thesis we consider a set of statistical decision problems, each having the structure of what is called the component problem: a state space Ω indexing a family of probability distributions $\{P_\omega, \omega \in \Omega\}$ over a σ-field $\mathcal{B}$ of a sample space $\mathfrak{X}$; a measurable action space $(A,\mathcal{A})$; and a loss function $L \geq 0$ defined on $\Omega \times A \times \mathfrak{X}$, which is $\mathcal{A}$-measurable for each ω and x. A (randomized) decision function Ψ has domain $\mathfrak{X} \times \mathcal{A}$ and is such that $\Psi(x)(\cdot)$ is a probability measure on $\mathcal{A}$ for each fixed x. If the state is ω, the conditional risk of Ψ given x is

$$(0.1) \qquad L(\omega,\Psi(x),x) = \int_A L(\omega,a,x)\Psi(x)(da);$$

and if $L(\omega,\Psi(x),x)$ is $\mathcal{B}$-measurable, the unconditional risk is

$$(0.2) \qquad R(\omega,\Psi) = \int L(\omega,\Psi(x),x)\,dP_\omega(x).$$

The notation of (0.1) will be extended to $L(\omega,\lambda,x) = \int_A L(\omega,a,x)\lambda(da)$ for any signed measure λ for which the right side exists; in this context, "a" will stand for the probability measure degenerate at a, so that, for example, $L(\omega,a-b,x) = L(\omega,a,x) - L(\omega,b,x)$.

For any $\underline{\theta} = (\theta_1,\theta_2,\ldots) \in \Omega^\infty$ we write, for the moment, $\underline{\theta}_N = (\theta_1,\theta_2,\ldots,\theta_N)$; then in the compound problem, $\underline{x}_N = (x_1,x_2,\ldots,x_N) \sim \prod_{r=1}^{N} P_{\theta_r} = P_{\underline{\theta}_N}$, and the choice of an action for the r^{th} component problem is allowed to depend on $\underline{x}_N$; this distinguishes the "set" compound problem from the "sequence" problem in which the action at the r^{th} stage can depend on $\underline{x}_r = (x_1,\ldots,x_r)$ only. Formally, a compound procedure of the type we consider is an array, $\underline{\varphi} = \{(\varphi_1^N,\varphi_2^N,\ldots,\varphi_N^N),\ N = 1,2,\ldots\}$ such that, for each N and $r \le N$, φ_r^N is defined on $\mathcal{X}^N \times \mathcal{A}$, with $\varphi_r^N(\underline{x}_N)$ being, for each $\underline{x}_N$, the probability measure on $\mathcal{A}$ according to which an action is chosen for the r^{th} problem. Since N is fixed in most of what follows (although our main concern is with the asymptotic properties of certain procedures), we will omit it henceforth and simply write $\underline{\varphi} = (\varphi_1,\varphi_2,\ldots,\varphi_N)$. For the same reason we shall write $\underline{\theta}$ for $\underline{\theta}_N$ and $\underline{x}$ for $\underline{x}_N$. If there are N problems, the conditional risk, given $\underline{x}$, of the procedure $\underline{\varphi}$ is

$$W(\underline{\theta},\underline{\varphi},\underline{x}) = N^{-1}\sum_{r=1}^{N} L(\theta_r,\varphi_r(\underline{x}),x_r) \tag{0.3}$$

and if $W(\underline{\theta},\underline{\varphi},\underline{x})$ is $\mathcal{B}^N$-measurable, the unconditional risk is

$$R(\underline{\theta},\underline{\varphi}) = \int W(\underline{\theta},\underline{\varphi},\underline{x})\,dP_{\underline{\theta}}(\underline{x}). \tag{0.4}$$

The above is the more or less standard setting for the compound decision problem. However to reduce somewhat the complexity of the notation, and to avoid the necessity of keeping track of a multiplicity of measures, we will adopt a slightly different point of view. Let P be any distribution on $\mathcal{X}^{\Omega}$ whose ω^{th} marginal is P_ω. Let $\underline{Y} = (Y_1,Y_2,\ldots)$ be a random matrix distributed according to P^{∞}, and suppose that, if the states are $\theta_1,\theta_2,\ldots,\theta_N$, only $(Y_1(\theta_1),Y_2(\theta_2),\ldots,Y_N(\theta_N))$ is observed, where $Y(\omega)$ is the ω^{th} coordinate of Y. We will write x for the observed coordinate of Y in the component problem and $\underline{x}$ for

$(x_1,x_2,\ldots,x_N) = (Y_1(\theta_1),Y_2(\theta_2),\ldots,Y_N(\theta_N))$. In addition we shall frequently omit the variable of integration. Consequently, in the notation to be used throughout, (0.2) and (0.4) become, respectively,

$$R(\omega,\Psi) = \int L(\omega,\Psi(x),x)\,dP(Y) = \int L(\omega,\Psi)\,dP \tag{0.5}$$

and

$$R(\underline{\theta},\underline{\varphi}) = \int W(\underline{\theta},\underline{\varphi},\underline{x})\,dP^{\infty}(\underline{Y}) = \int W(\underline{\theta},\underline{\varphi})\,dP^{\infty}. \tag{0.6}$$

As is becoming standard (cf. Gilliland, (1968) p. 1890), we say a compound procedure $\underline{\varphi}$ is simple if $\varphi_r(\cdot)(C)$ is a function of x_r for each r and C. If, in addition, the φ_r are identical, say $\varphi_r = \varphi$, we say $\underline{\varphi}$ is simple symmetric, with kernel φ. We shall, in general, identify simple symmetric procedures with their kernels, and write $R(\underline{\theta},\varphi)$ and $W(\underline{\theta},\varphi,\underline{x})$ for, respectively, the risk and the conditional risk given $\underline{x}$ of the simple symmetric procedure whose kernel is φ. For $\underline{\theta} \in \Omega^{\infty}$ and any simple symmetric procedure φ,

$$R(\underline{\theta},\varphi) = N^{-1} \sum_{r=1}^{N} R(\theta_r,\varphi).$$

From the right side it is clear that $R(\underline{\theta},\varphi)$ is the risk of the component procedure, φ, against G_N, the empirical distribution of $\theta_1,\theta_2,\ldots,\theta_N$. To emphasize this, we shall frequently write, when φ is simple symmetric,

$$\begin{aligned} R(\underline{\theta},\varphi) &= G_N[R(\omega,\varphi)] \\ &= R(G_N,\varphi) \geq R(G_N) \end{aligned} \tag{0.7}$$

where $R(\cdot)$ is the Bayes envelope for the component problem and we have used the convention of writing integrals in operator notation, i.e. $G[h(\omega)] = \int h(\omega)dG(\omega)$, a notation we shall continue to use for integrals on the space Ω only.

In (0.7), because of the nature of G_N, the $\mathcal{B}$-measurability of $L(\omega,\varphi(x),x)$ for each ω suffices to make $R(G_N,\underline{\varphi})$ meaningful. Future references to $R(G,\varphi)$, where G is a distribution on Ω, will be restricted to the class, Φ, of "measurable" component procedures - those for which $R(\omega,\varphi)$ exists and is a measurable function of ω. In particular, the Bayes envelope at G is given by $R(G) = \inf_{\Phi} R(G,\varphi)$.

For any simple procedure $\underline{\varphi}$, (cf. Gilliland (1968), p. 1890),

$$\sup_{\underline{\theta}} \{R(\underline{\theta},\underline{\varphi}) - R(G_N)\} \geq \sup_{\theta} \{N^{-1} \sum_{r=1}^{N} R(\theta,\varphi_i) - \inf_{\Psi} \int L(\theta,\Psi)dP\}$$

$$\geq \inf_{\varphi} \sup_{\theta} \{R(\theta,\varphi) - \inf_{\Psi} \int L(\theta,\Psi)dP\}.$$

The right side is positive, and is zero only when the component problem is trivial; hence, with modified regret defined by

$$D(\underline{\theta},\underline{\varphi}) = R(\underline{\theta},\underline{\varphi}) - R(G_N) \tag{0.8}$$

we have $D(\underline{\theta},\underline{\varphi}) \geq 0$ for all simple $\underline{\varphi}$; and if the component problem is non-trivial, there is a $\delta > 0$ such that $\sup_{\underline{\theta}} D(\underline{\theta},\underline{\varphi}) \geq \delta$ for all simple $\underline{\varphi}$.

However, it is possible, in some cases, to find non-simple procedures, $\underline{\varphi}$, for which $D(\underline{\theta},\underline{\varphi})$ converges in some sense to zero as $N \to \infty$. Robbins (1951) gives a heuristic argument for the existence of such procedures, and precedes it by an example in which the component problem is to distinguish between $N(1,1)$ and $N(-1,1)$. In the case where the component problem is to distinguish between two arbitrary distributions, Hannan and Robbins (1955) exhibit non-simple procedures, $\underline{\varphi}$, for which $D(\underline{\theta},\underline{\varphi}) < o(1)$ uniformly in $\underline{\theta}$ as $N \to \infty$ (Theorem 4); this result is obtained as a corollary to a theorem (Theorem 3) in which it is shown (in our notation) that to any $\epsilon > 0$ corresponds an $N(\epsilon)$ such that, for any $\underline{\theta} \in \Omega^{\infty}$,

$$P^{\infty}[W(\underline{\theta},\underline{\varphi},\underline{x}) - R(G_N) > \epsilon \text{ for some } N > N(\epsilon)] < \epsilon.$$

In addition, defining the "equivariant envelope" by

$$R^*(G_N) = \inf\{R(G_N,\underline{\varphi}): \underline{\varphi}(g\underline{x}) = g\,\underline{\varphi}(\underline{x}) \text{ for all } \underline{x} \in \mathfrak{X}^N,\ g \in \mathfrak{S}\}$$

where $\mathfrak{S}$ is the set of all permutations, g, of vectors of N coordinates (i.e. $g(r_1,r_2,\ldots,r_N) = (r_{g(1)},r_{g(2)},\ldots,r_{g(N)})$ where $g(1,2,\ldots,N) = (g(1),\ldots,g(N))$), they show that $|R^*(G_N) - R(G_N)| \le o(1)$ as $N \to \infty$. Hannan and Huang (1969) have generalized and strengthened this latter result by showing that, for a finite state space and under mild conditions on the loss function, $|R^*(G_N) - R(G_N)| \le O(N^{-\frac{1}{2}})$ as $N \to \infty$.

Hannan and Van Ryzin (1965), in the case considered by Hannan and Robbins, exhibit a function of the observation in the r^{th} problem which provides an unbiased estimate of the empirical distribution of the r^{th} state; they consider procedures which consist of playing Bayes, in each component problem, against the estimate of G_N given by the average of these estimates, and give sets of conditions, each stronger than the last, under which an upper bound for $D(\underline{\theta},\underline{\varphi})$ is $O(N^{-\frac{1}{2}})$, $o(N^{-\frac{1}{2}})$ and $O(N^{-1})$ respectively, each bound being uniform in $\underline{\theta}$. Van Ryzin (1966) considers the case in which the component problem consists of making one of n decisions based on an observation from one of m distributions; for procedures analogous to those of Hannan and Van Ryzin, he gives conditions which imply $D(\underline{\theta},\underline{\varphi}) \le 0(N^{-\frac{1}{2}})$ and further conditions implying $D(\underline{\theta},\underline{\varphi}) \le 0(N^{-1})$, both rates being uniform in $\underline{\theta}$.

Mention should also be made of the papers by Stein (1955) and James and Stein (1960) showing inadmissibility of the usual estimator of the mean of an N-variate normal distribution $(N > 2)$, with covariance matrix identity, under squared error loss. As is recognized in these papers, this is a compound problem whose r^{th} component problem is to estimate the mean of the r^{th} coordinate of the random vector; however the estimator which makes the usual estimator inadmissible does not seem to have been derived by an explicit compound procedure.

Other work on the compound problem includes that of Fox (1968, Chapter 2) who exhibits procedures, $\underline{\varphi}$, for which $D(\underline{\theta},\underline{\varphi}) \to 0$ for each $\underline{\theta}$ in the case where the component problem is a test on the parameter of an exponential family; that of Cogburn (1963 and 1967) who deals in particular with the case in which the component problem is to estimate the probability of success of a binomial distribution and who formulates a general approach from the point of view of the notion of stringent solutions; that of Suzuki (1968) who includes (among other things) some discussion of "ϵ-Bayes" simple symmetric procedures; and some expository work of Samuel (1967).

Throughout this thesis, ϵ denotes an arbitrary non-negative number, possibly depending on N, and we concern ourselves with the asymptotic properties of procedures which consist of playing ϵ-Bayes against an estimate $\hat{G}(\underline{x})$ of G_N. The main intuitive justification for these procedures is the following lemma, a slight generalization of earlier results (see, e.g., Hannan and Robbins (1955), Theorem 1), which we will refer to again, and which shows that a simple symmetric procedure which is Bayes against a distribution "close" to G_N will yield a risk close to the Bayes envelope at G_N.

Lemma 0 Let G and F be any distributions on Ω and let φ be any measurable component procedure. Then

$$R(G,\varphi) - R(G) \leq \sup_{\nu,\Psi\in\Phi} (G - F)[R(\omega,\nu) - R(\omega,\Psi)] + R(F,\varphi) - R(F)$$

where Φ is, as before, the class of measurable component procedures.

Proof. (1) $R(G,\varphi) - R(F,\varphi) = (G - F)[R(\omega,\varphi)]$

for any procedure φ. If φ_n is any procedure for which $R(G,\varphi_n) \leq R(G) + n^{-1}$, then

$$(2)\quad R(F) - R(G) \leq R(F,\varphi_n) - R(G,\varphi_n) + n^{-1}$$

$$= -(G - F)[R(\omega,\varphi_n)] + n^{-1}$$

Adding (1) and (2) and taking the supremum,

$$R(G,\varphi) - R(G) - R(F,\varphi) + R(F) \leq \sup_{\nu,\Psi\in\Phi} (G - F)[R(\omega,\nu) - R(\omega,\Psi)] + n^{-1}.$$

Since the left side is independent of n, the proof is complete.

If Ω is finite and $\int L(\omega,\varphi)dP_\omega \leq M < \infty$ for every ω and $\Psi \in \Phi$, and φ is ϵ-Bayes against F (i.e. $R(F,\varphi) \leq R(F) + \epsilon$) then

$$(0.9)\qquad R(G,\varphi) - R(G) \leq 2M|G - F|(\Omega) + \epsilon.$$

Hence if we play ϵ-Bayes against an estimator, $\hat{G}$, for which $|\hat{G}(\underline{x}) - G_N|(\Omega) \to 0$ in some sense as $N \to \infty$, we might expect to do reasonably well.

In Chapter 1 we consider the case in which Ω is finite; we obtain results of the form $\sup_{\underline{\theta}} D(\underline{\theta},\underline{\varphi}) \leq o(1) + \epsilon$ for a fairly wide class of "half-space" procedures, $\underline{\varphi}$, when the action space is finite and under rather weaker conditions on the loss function and the estimator $\hat{G}$ than are in Van Ryzin; this result is a corollary to Theorem 1, which is analogous to the result of Hannan and Robbins involving $W(\underline{\theta},\underline{\varphi},\underline{x})$. Later in Chapter I we obtain $\sup_{\underline{\theta}} D(\underline{\theta},\underline{\varphi}) \le o(1) + \varepsilon$ for equivariant procedures which play ε-Bayes, in a certain uniform sense, against $\hat{G}(\underline{x})$; and this result is used to achieve a similar result, under certain conditions, when the action space is infinite.

In Chapter 2 we consider the case when Ω is infinite. Assuming the existence of "good" estimators, we show that under certain conditions (mainly involving the total boundedness of $\mathcal{X}$ in an L_∞-norm for the loss function, and of Ω in an L_1-norm for the family of distributions) there exist, for arbitrarily small η, procedures $\underline{\varphi}$ for which $D(\underline{\theta},\underline{\varphi}) \le o(1) + \eta$ for each $\underline{\theta}$ and the result analogous to Theorem 3 of Hannan and Robbins (1955) holds. We conclude Chapter 2 by showing that the required conditions hold for a very large class of problems.

In addition to the notational conventions already described, we shall also identify sets with their indicator functions and write H^c for the complement of a set H.

CHAPTER I

FINITE STATE SPACES

§1.0 Definitions and Preliminaries

Throughout this chapter, $\Omega = \{1,2,\ldots,m\}$. In this case we have $P_\omega << \mu = \sum_{\omega=1}^{m} P_\omega$, and we define $f(\omega,\cdot) = \frac{dP_\omega}{d\mu}(\cdot)$. We assume throughout that $L(\omega,a,x)$ is $\mathcal{B}$-measurable for each fixed ω and a, and that

$$\text{(1.1)} \qquad \int L(\omega,a,x)dP \le M < \infty \quad \text{for all } \omega \text{ and } a.$$

We let $\mathcal{G}$ be the set of all distributions on Ω.

Definition 1. $\hat{G}$ is a uniformly consistent estimator (of G_N, the empirical distribution of $\theta_1,\theta_2,\ldots,\theta_N$) if there exists a function $N_1(\eta,\gamma)$, defined for all $\eta > 0$ and $\gamma > 0$ such that, for each $\underline{\theta} \in \Omega^\infty$,

$$\sup_{N > N_1(\eta,\gamma)} P^\infty[|\hat{G}(\underline{x}) - G_N|(\Omega) > \eta] < \gamma.$$

With the supremum inside the square brackets, $\hat{G}$ is uniformly strongly consistent. (An estimator $\hat{G}$ is really a sequence of functions $\hat{G}_1,\hat{G}_2,\ldots$, with $\hat{G}_N : \mathcal{X}^N \to \mathcal{G}$ being $\mathcal{B}^N$-measurable for each N. We shall not need to emphasize this formality, however.)

Definition 2. For each $F \in \mathcal{G}$, let Φ_F be the set of component procedures ϵ-Bayes against F. For $\hat{G}$ an estimator, let

$$\Phi_{\hat{G}} = \{\underline{\varphi} : \forall\, \underline{x}, \exists\, \varphi^o(\underline{x}) \in \Phi_{\hat{G}(\underline{x})} \text{ such that}, \forall\, r,\ \varphi_r(\underline{x}) = \varphi^o(\underline{x})(x_r)\}$$

Thus to use a procedure in $\Phi_{\hat{G}}$ one first estimates G_N by $\hat{G}(\underline{x})$ and then, using a simple symmetric procedure the choice of whose kernel is permitted to depend on $\underline{x}$, plays ϵ-Bayes against $\hat{G}(\underline{x})$ in each component problem. Henceforth, for each $\underline{x}$ and each $\underline{\varphi} \in \Phi_{\hat{G}}$, $\varphi^{o}(\underline{x})$ will denote the component procedure given by Definition 2.

It is with two subsets of $\Phi_{\hat{G}}$ that we will be concerned in this chapter, the first in sections 1.1, 1.3 and 1.4, and the second in sections 1.5, 1.6 and 1.7.

§1.1 Finite Action Spaces. Definitions of Half-Spaces and Half-Space Procedures.

Throughout this section (and also sections 1.3 and 1.6) $A = \{1,2,\ldots,n\}$. Let E^k be k-dimensional Euclidean space and let $Z: \mathcal{X} \to E^{mn}$ be given by

$$(1.2)\quad Z(\omega,a,x) = L(\omega,a,x)f(\omega,x) \qquad \omega = 1,2,\ldots,m;\ a = 1,2,\ldots,n.$$

We shall adopt, for Z, the same conventions as for L; i.e. $Z(\omega,\lambda,x) = \int Z(\omega,a,x)d\lambda(a)$ for any signed measure on $\mathcal{A}$ for which the right side exists.

Definition 3. A set $H \subset E^k$ is a half-space if, for some linear functional ℓ and some number p, either H or H^c is $\{y: \ell(y) < p\}$. Let $\mathcal{N}_s$ be the set of all intersections of s half-spaces, $\mathcal{N}_s^t$ the set of all unions of t members of $\mathcal{N}_s$, and $\mathcal{K}_s^t = Z^{-1}(\mathcal{N}_s^t)$.

For each $F \in \mathcal{J}$ we want to restrict attention to those members of Φ_F which take on only finitely many values, and for which the corresponding induced partition of $\mathcal{X}$ is a collection of regions each of which is an element of $\mathcal{K}_s^t$ for some t and s. Formally, $\Psi \in \Phi_F$ is an element of $\Phi_{F,s,t,v}$ if

$$\{\Psi(x) : x \in \mathfrak{X}\} \subset \{\nu_1, \nu_2, \ldots, \nu_v\}$$

where $\nu_1, \ldots, \nu_v$ are distinct measures on $\mathcal{Q}$ and, for each j, $\{x: \Psi(x) = \nu_j\} = Q_j$ for some (possibly empty) $Q_j \in \mathcal{K}_s^t$. Hence $\Psi(x)(a) = \sum_{j=1}^{v} Q_j(x)\nu_j(a)$ for each $x \in \mathfrak{X}$ and $a \in A$.

Definition 4. (Half-Space Procedures). Let $\hat{G}$ be an estimator. Then, with s,t,v all finite,

$$\Phi_{\hat{G},s,t,v} = \{\underline{\varphi} \in \Phi_{\hat{G}} : \text{for each } \underline{x},\ \varphi^o(\underline{x}) \in \Phi_{\hat{G}(\underline{x}),s,t,v}\},$$

where φ^o is as in Definition 2.

To use a procedure in $\Phi_{\hat{G},s,t,v}$, one first estimates G_N by $\hat{G}(\underline{x})$ and then, using a simple symmetric procedure whose kernel, the choice of which may depend on $\underline{x}$, is an element of $\Phi_{\hat{G}(\underline{x}),s,t,v}$, one plays ϵ-Bayes against $\hat{G}(\underline{x})$ in each component problem.

Most results obtained so far in the set version of the compound decision problem have been obtained only for special subsets of $\Phi_{\hat{G},s,t,v}$ (e.g. Hannan and Robbins (1955) and Van Ryzin (1966)) - usually the class $\Phi_{F,s,t,v}$ has been restricted to those procedures, Ψ, for which $\Psi(x)(a) = \sum_{B \subset A} Q_{FB}^o(x)\nu_B(a)$, where $x \in Q_{FB}^o$ if B is the set of Bayes acts against F when x is observed. The proof that $Q_{FB}^o \in \mathcal{K}_s^t$ for some t and s will be given in a more general context in Lemma 2, in section 1.5. Usually $\nu_B(a)$ is restricted to the values 0 and 1. In addition the form of the estimator, $\hat{G}$, is usually restricted, especially when rates have been obtained; the most common form is the "average of unbiased estimators" given by Hannan and Van Ryzin (1965) and mentioned in the Introduction.

§1.2 Unions of Intersections of Half-Spaces.

Recalling the definition of $\mathcal{K}_s^t$ in the previous section, and identifying sets and their indicator functions, we see that if $H \in \mathcal{K}_s^t$, then for some elements $\{H_j\}$ of $\mathcal{K}_s$, $H = \bigcup_1^t H_j = H_1 + H_1^c H_2 + H_1^c H_2^c H_3 + \ldots + H_1^c H_2^c \ldots H_{t-1}^c H_t$. Since $H_j \in \mathcal{K}_s$ implies that, for some half-spaces $\{H_{ji}\}$, $H_j = \bigcap_{i=1}^s H_{ji}$, so that $H_j^c = \bigcup_{i=1}^s H_{ji}^c = H_{j1}^c + H_{j1} H_{j2}^c + \ldots + H_{j1} H_{j2} \ldots H_{js-1} H_{js}^c$, we see that $H \in \mathcal{K}_s^t$ implies that H is the disjoint union of at most $\sum_{j=0}^{t-1} s^j$ members of $\mathcal{K}_{st}$. For future use we note that this implies:

(1.4) If $H_j \in \mathcal{K}_{s_j}^{t_j}$ for $j = 1,2,\ldots,J$, then $\bigcap_{j=1}^J H_j$ is the disjoint union of $\prod_{j=1}^J (\sum_{k=0}^{t_j - 1} s_j^k)$ members of $\mathcal{K}_q$, where $q = \sum_{j=1}^J s_j t_j$.

We now prove a lemma which is a slight generalization of the results of Ranga Rao (1962), and which we use in section 1.3.

Lemma 1. Let $(\mathfrak{X}, \mathcal{B}, P)$ be a probability space, and let P_N be the empirical distribution of N i.i.d. random variables $\sim P$. Let $h: \mathfrak{X} \to \mathcal{R}$ be P-integrable and let $g: \mathfrak{X} \to E^k$ be $\mathcal{B}$-measurable. Then, for any s and t,

$$P^\infty[\sup_{H \in \mathcal{K}_s^t} |\int g^{-1}(H)\, h\, d(P_N - P)| \to 0 \quad \text{as} \quad N \to \infty] = 1.$$

Proof. It is clearly sufficient to prove the lemma for $h \geq 0$; also, by the preceding discussion, and since g^{-1} preserves unions and intersections, it is sufficient to prove the theorem for $\mathcal{K}_s$. If $H = \prod_{j=1}^s H_j$, where the H_j are indicators of half-spaces, H can be written as a linear combination of open members of $\mathcal{K}_s$ by replacing any closed H_j by $1 - H_j^c$ in the product. Hence it is sufficient to prove the lemma for open members of $\mathcal{K}_s$.

For any Borel set $B \subset E^k$, let $\lambda(B) = \int g^{-1}(B)\, h\, dP$, and $\lambda_N(B) = \int g^{-1}(B) h\, dP_N$. Ranga Rao (1962, Lemma 7.3) shows the existence, for finite λ, of mutually orthogonal measures, $\lambda_r^{(i)}$, $r = 0,1,\ldots,k-1$, $i = 0,1,2,\ldots$ and $r = k$, $i = 0$ such that $\lambda = \sum_{r=0}^{k-1} \sum_{i=0}^{\infty} \lambda_r^{(i)} + \lambda_k^{(0)}$, and, for each r and i, there exists an r-dimensional subspace or a translate of such a subspace, $A_r^{(i)}$ say, such that $\lambda_r^{(i)}(E^k \sim A_r^{(i)}) = 0$, and $\lambda_r^{(i)}(A) = 0$ whenever A is a translate of a subspace of dimension less than r. For each Borel set $B \subset E^k$, let $\lambda_{Nr}^{(i)}(B) = \lambda_N(B \cap A_r^{(i)})$. Then Ranga Rao (Lemma 7.5) shows that if (i) λ_N converges weakly to λ, (ii) $\lambda_N(A_r^{(i)}) \to \lambda(A_r^{(i)})$ for all r and i, and (iii) $\lambda_{Nr}^{(i)}$ converges weakly to $\lambda_r^{(i)}$, then $\sup_{H \in \mathscr{H}^*_{st}} |\lambda_N(H) - \lambda(H)| \to 0$ as $N \to \infty$, where $\mathscr{H}^*_{st}$ are the open members of ${}^{st}\mathscr{H}_{st}$.

Thus our lemma is proved if we can show that (i), (ii) and (iii) hold almost surely $[P^\infty]$. However (ii) follows immediately from the strong law of large numbers and the fact that $\{A_r^{(i)}\}$ is a countable collection; and (i) and (iii) follow from the strong law together with the "sufficiency" part of Theorem 3.1 of Varadarajan (1958), which establishes that for any separable metric space S there is a sequence of functions $f_1, f_2, \ldots$ such that, for any finite measure λ on S, λ_N converges weakly to λ if $\int f_i\, d\lambda_N \to \int f_i\, d\lambda$ for each i. Hence the lemma is proved.

§1.3 Convergence of $D(\underline{\theta},\underline{\varphi})$ for Half-Space Procedures.

In this section we prove the main theorem for half-space procedures, establishing conditions for the uniform almost sure convergence of the conditional risk. This is followed by some remarks which attempt to point up some features of the proof which the statement of the theorem tends to obscure, by a Corollary concerning the unconditional risk, and some further remarks.

Theorem 1. Let $\hat{G}$ be a uniformly strongly consistent estimator of G_N. Then given $s < \infty$, $t < \infty$ and $v < \infty$, there exists a function $N(\eta,\gamma)$, defined for all $\eta > 0$ and $\gamma > 0$, such that, for all $\underline{\theta} \in \Omega^\infty$ and all $\underline{\varphi} \in \Phi_{\hat{G},s,t,v}$,

$$P^\infty[|W(\underline{\theta},\underline{\varphi},\underline{x}) - R(G_N)| > \epsilon + \eta \text{ for some } N > N(\eta,\gamma)] < \gamma.$$

Proof. Let $\Psi \in \Phi_{F,s,t,v}$, say $\Psi = \sum_{j=1}^{v} Q_j \nu_j$ where each Q_j is an element of $\mathcal{K}_s^t$, and $\nu_1,\ldots,\nu_v$ are distinct measures on $\mathcal{Q}$. Then

$$W(\underline{\theta},\Psi,\underline{x}) = N^{-1} \sum_{r=1}^{N} \sum_{j=1}^{v} Q_j(x_r) \sum_{a\in A} \nu_j(a) L(\theta_r,a,x_r)$$

$$= \sum_{\omega=1}^{m} N^{-1} \sum_{j=1}^{v} \sum_{a\in A} \nu_j(a) \sum_{\{r:\theta_r=\omega\}} Q_j(x_r) L(\theta_r,a,x_r).$$

Now $\sum_{\{r:\theta_r=\omega\}} Q_j(x_r)L(\theta_r,a,x_r) = N_\omega \int Q_j L(\omega,a)dP_{N_\omega}$ where P_{N_ω} is the empirical distribution of the $N_\omega = N_\omega(\underline{\theta}) = \sum_{r=1}^{N} [\theta_r=\omega]$ i.i.d. random variables $\{x_r:\ \theta_r=\omega\}$. Thus

$$W(\underline{\theta},\Psi,\underline{x}) = \sum_{\omega=1}^{m} \frac{N_\omega}{N} \sum_{j=1}^{v} \sum_{a\in A} \nu_j(a) \int Q_j L(\omega,a)dP_{N_\omega}.$$

Subtracting from each side its expectation:

$$W(\underline{\theta},\Psi,\underline{x}) - R(G_N,\Psi) = \sum_{\omega=1}^{m} \frac{N_\omega}{N} \sum_{j=1}^{v} \sum_{a\in A} \nu_j(a) \int Q_j L(\omega,a)d(P_{N_\omega} - P_\omega).$$

Finally, taking the supremum of the absolute value of the integrals and bounding out the $\nu_j(a)$ terms, we have, with

$$S(N,\underline{\theta},\omega,\underline{x}) = v \sum_{a\in A} \sup_{Q\in\mathcal{K}_s^t} \left|\int Q\, L(\omega,a)d(P_{N_\omega} - P_\omega)\right|,$$

(1.5) $$|W(\underline{\theta},\Psi,\underline{x}) - R(G_N,\Psi)| \le \sum_{\omega=1}^{m} \frac{N_\omega}{N} S(N,\underline{\theta},\omega,\underline{x}),$$

for any $\Psi \in \Phi_{F,s,t,v}$ and any $F \in \mathcal{F}$.

Let $\varphi^o(\underline{x})$ be the member of $\Phi_{\hat{G}(\underline{x}),s,t,v}$, guaranteed by Definition 4, for which $\varphi_r(\underline{x}) = \varphi^o(\underline{x})(x_r)$ for each r. Since A has n elements, and $\int L(\omega,a)dP \le M$ for each ω and a, $\int L(\omega,\Psi)dP \le Mn$ for any component procedure Ψ. Hence, from (0.9) and (1.5), we have, since $W(\underline{\theta},\underline{\varphi},\underline{x}) = W(\underline{\theta},\varphi^o(\underline{x}),\underline{x})$ for each $\underline{x}$,

$$(1.6)\quad |W(\underline{\theta},\underline{\varphi},\underline{x}) - R(G_N)| \le \sum_{\omega=1}^{m} \frac{N_\omega}{N} S(N,\underline{\theta},\omega,\underline{x}) + Mn|\hat{G}(\underline{x}) - G_N|(\Omega) + \epsilon.$$

We note that (1.6) does not depend on any properties of $\hat{G}$, nor on the measurability of the left side, points to which we shall return in the remarks to follow.

Applying Lemma 1 to the random variables $\{x_r: \theta_r = \omega\}$, there exists a function $k' = k'(\eta,\gamma)$, defined for all $\eta > 0$ and $\gamma > 0$, such that, for all ω, $\underline{\theta}$ and a,

$$P^\infty[\sup_{J(k',\omega,\underline{\theta})} \sup_{Q\in\mathcal{X}_s^t} |\int Q\, L(\omega,a)d(P_{N_\omega} - P_\omega)| > \eta] < \gamma$$

where $J(k',\omega,\underline{\theta}) = \{N: N_\omega(\underline{\theta}) > k'\}$.

Thus, with $h = k(\eta,\gamma) = k'(\eta(nv)^{-1},\gamma n^{-1})$, we have

$$P^\infty[\sup_{J(h,\omega,\underline{\theta})} S(N,\underline{\theta},\omega,\underline{x}) > \eta] < \gamma \text{ for all } \underline{\theta} \text{ and } \omega.$$

Thus, for any N', with $k = k(\eta m^{-1}, \gamma(2m)^{-1})$ and $H(k,\omega,\underline{\theta}) = \{N: N > N' \text{ and } N_\omega(\underline{\theta}) \le k\}$, we have

$$\begin{aligned}
&P^\infty[\sum_{\omega=1}^{m} \frac{N_\omega(\theta)}{N} S(N,\underline{\theta},\omega,\underline{x}) > \eta \text{ for some } N > N']\\
&\le \sum_{\omega=1}^{m} P^\infty[\frac{N_\omega(\theta)}{N} S(N,\underline{\theta},\omega,\underline{x}) > \eta m^{-1} \text{ for some } N > N']\\
&\le \sum_{\omega=1}^{m}\left\{P^\infty\left[\sup_{J(k,\omega,\underline{\theta})} S(N,\underline{\theta},\omega,\underline{x}) > \eta m^{-1}\right] + P^\infty\left[\frac{k}{N'} \max_{H(k,\omega,\underline{\theta})} S(N,\underline{\theta},\omega,\underline{x}) > \eta m^{-1}\right]\right\}\\
&< \gamma/2 + g(N')
\end{aligned}$$

where $g(N') \downarrow 0$ as $N' \uparrow \infty$, since $\max\limits_{H(k,\omega,\underline{\theta})} S(N,\underline{\theta},\omega,\underline{x})$ is finite valued, because $\{S(N,\underline{\theta},\omega,\underline{x}) : N \in H(k,\omega,\underline{\theta})\}$ has at most k elements.

Hence for N' sufficiently large, say $N_2(\eta,\gamma)$,

$$(1.7) \qquad P^{\infty}[\sum_{\omega=1}^{m} \frac{N_{\omega}}{N} S(N,\underline{\theta},\omega,\underline{x}) > \eta \text{ for some } N > N'] < \gamma.$$

Using now the condition on $\hat{G}$, let $N(\eta,\gamma) = \max\{N_1(\eta(2Mn)^{-1}, \gamma/2), N_2(\eta/2, \gamma/2)\}$ where N_1 is the function given by the uniform strong consistency of $\hat{G}$, as described in Definition 1. Then (1.6) and (1.7) together yield Theorem 1.

Remarks 1. Neither (1.6) nor (1.7) depends on the hypothesis concerning $\hat{G}$. Consequently this hypothesis could be omitted and the theorem restated as: given $s < \infty$, $t < \infty$ and $v < \infty$, there exists $N(\eta,\gamma)$ such that, for all $\underline{\theta}$ and all $\underline{\varphi} \in \Phi_{\hat{G},s,t,v}$

$$(1.8) \quad P^{\infty}[|W(\underline{\theta},\underline{\varphi},\underline{x}) - R(G_N)| - Mn|\hat{G}(\underline{x}) - G_N|(\Omega) > \eta+\epsilon \text{ for some } N>N(\eta,\gamma)] < \gamma.$$

2. It follows from (1.8) that if $\hat{G}$ is uniformly consistent (not necessarily strongly) and $\eta > 0$ then

$$(1.9) \qquad \sup_{\Phi_{\hat{G},s,t,v}} \sup_{\underline{\theta}} P^{\infty}[|W(\underline{\theta},\underline{\varphi}.\underline{x}) - R(G_N)| > \eta + \epsilon] \to 0 \quad \text{as} \quad N \to \infty.$$

It is (1.9) which yields the corollary to this theorem.

Corollary. If $\hat{G}$ is uniformly consistent then

$$\sup_{\Phi_{\hat{G},s,t,v}} \sup_{\underline{\theta}} D(\underline{\theta},\underline{\varphi}) < \epsilon + o(1) \quad \text{as} \quad N \to \infty.$$

Proof. Recalling our convention that $\underline{x} = (Y_1(\theta_1),\ldots,Y_N(\theta_N))$, let $C_N = \{\underline{Y}: W(\underline{\theta},\underline{\varphi},\underline{x}) > R(G_N) + \epsilon + \delta/2\}$, so that

$$R(\underline{\theta},\underline{\varphi}) \le R(G_N) + \epsilon + \delta/2 + \int C_N W(\underline{\theta},\underline{\varphi})dP^{\infty}.$$

Since $W(\underline{\theta},\underline{\varphi},\underline{x}) \le N^{-1}\sum_{r=1}^{N} \max_{\omega,a} L(\omega,a,Y_r(\omega)) = N^{-1}\sum_{r=1}^{N} V(Y_r)$, say,

$$\int C_N W(\underline{\theta},\underline{\varphi})dP^{\infty} \le N^{-1}\sum_{r=1}^{N}\int C_N V(Y_r)dP^{\infty}$$

$$< \delta/2 \quad \text{for } P(C_N) \text{ sufficiently small,}$$

because the $V(Y_r)$ are identically distributed and integrable.

Hence, from (1.9), $R(\underline{\theta},\underline{\varphi}) < R(G_N) + \epsilon + \delta$, for all $\underline{\theta}$ and $\underline{\varphi} \in \Phi_{\hat{G},s,t,v}$, for N sufficiently large.

Remarks 1. We repeat the observation in the Introduction that none of our results are affected if ϵ depends on N. In particular, if $\epsilon = o(1)$ as $N \to \infty$, then the conclusion of the corollary becomes $\sup_{\Phi_{\hat{G},s,t,v}} \sup_{\underline{\theta}} D(\underline{\theta},\underline{\varphi}) < o(1)$ as $N \to \infty$.

2. The class $\Phi_{\hat{G},s,t,v}$ may include procedures, $\underline{\varphi}$, which are not $\mathcal{B}^N$-measurable (though, for each $\underline{x}$, the component procedures $\varphi^o(\underline{x})$ must be $\mathcal{B}$-measurable). Such procedures are included in Theorem 1 and its corollary in the sense that, whether or not $W(\underline{\theta},\underline{\varphi},\underline{x})$ is $\mathcal{B}^N$-measurable, there is a measurable function, $W'(\underline{\theta},\underline{x})$ such that $W'(\underline{\theta},\underline{x}) \ge W(\underline{\theta},\underline{\varphi},\underline{x})$ for all $\underline{\theta}$ and $\underline{x}$, with W' having the properties asserted for W. This can be seen from the fact that both $S(N,\underline{\theta},\omega,\underline{x})$ and $|\hat{G}(\underline{x}) - G_N|(\Omega)$ are $\mathcal{B}^N$-measurable, and the comment following (1.6).

§1.4 Remarks on the Restrictions on Half-Space Procedures.

It is clear that the results of section 1.1 depend heavily on the use of Lemma 1. Indeed this is the reason for restricting the conclusions of Theorem 1 to the class $\Phi_{\hat{G},s,t,v}$.

The restrictions on the class $\Phi_{\hat{G},s,t,v}$ are two, both the result of restrictions on the classes $\Phi_{F,s,t,v}$ for $F \in \mathcal{J}$: that for

$\Psi \in \Phi_{F,x,t,v}$, (1) $\{\Psi(x): x \in \mathfrak{X}\}$ is finite and (2) $\{x: \Psi(x) = \nu_j\} \in \mathcal{K}_s^t$ for each j. Restriction (2) is clearly not essential: the application of Lemma 1 would not be affected if finitely many of the sets $\{\{x: \Psi(x) = \nu_j\}, 1 \le j \le v, \Psi \in \Phi_{F,x,t,v}, F \in \mathcal{B}\}$ were not elements of $\mathcal{K}_s^t$.

The crux of the proof of Theorem 1 is the uniform almost sure convergence of $S(N,\underline{\theta},\omega,\underline{x})$, which is obtained from Lemma 1 because of the structure of the family of functions $\bigcup_{F\in\mathcal{B}} \{\Psi(\cdot)(a): a \in A, \Psi \in \Phi_{F,s,t,v}\}$. However if, for each $F \in \mathcal{B}$, Φ_{F1} is a subclass of Φ_F and, for each $\omega \in \Omega$ and $a \in A$,

$$P^\infty[\sup|\int g(a)L(\omega,a)d(P_{N_\omega} - P_\omega)| \to 0 \quad \text{as} \quad N \to \infty] = 1$$

where the sup is taken over $g \in \bigcup_{F\in\mathcal{B}} \{\Psi(\cdot)(a): \Psi \in \Phi_{F1}\}$, then Theorem 1 would hold with $\Phi_{\hat{G},s,t,v}$ replaced by $\Phi_{\hat{G}1}(= \{\underline{\varphi} \in \Phi_{\hat{G}}: \varphi^o(\underline{x}) \in \Phi_{\hat{G}(\underline{x})1}$ for each $\underline{x}\})$. The families $\bigcup_{F\in\mathcal{B}} \{\Psi(\cdot)(a): \Psi \in \Phi_{F,s,t,v}\}$, $a \in A$, are by no means the only ones with this property; however they are of particular interest as a rather natural generalization of the standard situation, mentioned previously in section 1.1, in which, for each F, Φ_F is restricted to those procedures, Ψ, for which $\Psi(x)(a) = \sum_{B\subset A} Q_{FB}^o(x)\nu_B(a)$.

In fact, although ν_B usually depends only on B (the most common case, with $A = \{1,2,\ldots,n\}$, is to have ν_B degenerate at the "minimum" member of B; see, e.g., Hannan and Robbins (1955), Hannan and Van Ryzin (1965), Van Ryzin (1966)), Theorem 1 also applies to the case where ν_B is also permitted to depend on F; and the conclusions of Theorem 1 hold if ν_B is a measurable function of x (since $P^\infty[\sup_{Q\in\mathcal{K}_s^t} |\int \nu_B(a)Q\, L(\omega,a)d(P_{N_\omega} - P_\omega)| \to 0$ as $N_\omega \to \infty] = 1$ for each $a \in A$, $B \subset A$ and $\omega \in \Omega$, from Lemma 1).

However the methods of Theorem 1 fail if ν_B depends on both F and x. In this case (1.5) becomes

$$(1.10)\quad W(\underline{\theta},\Psi,\underline{x}) - R(G_N,\Psi) \le \sum_{\omega=1}^{m} \frac{N_\omega}{N} \sum_{B\subset A} \sum_{a\in A} \sup_{F\in\mathscr{F}} \left|\int Q^o_{FB}\nu_{FB}(a)L(\omega,a)d(P_{N_\omega} - P)\right|.$$

It is possible for the right side of (1.10) not to converge to zero; for there may exist, for a fixed B, containing at least 2 points, a set $C \in \mathcal{B}$, of non-atomic P_ω-measure, with card $C \le$ card $\mathfrak{F}$ where $\mathfrak{F} = \{F: C \subset Q^o_{FB}\}$. Let J map $\mathfrak{F}$ onto the finite subsets of C and let $\nu_{FBx}(a) = J(F)(x)$ for some $a \in B$ and all $F \in \mathfrak{F}$. Then, if points are measurable, $\int Q^o_{FB}\nu_{FB}(a)L(\omega,a)dP = \int J(F)L(\omega,a)dP_\omega = 0$ for any F, since $J(F)$ is finite. However for any empirical distribution P_{N_ω}, there is an F for which $P_{N_\omega}[x \in J(F)] = P_{N_\omega}[x \in C]$ and $C \subset Q^o_{FB}$. Hence

$$\sup_{F\in\mathfrak{F}} \int Q^o_{FB}\nu_{FB}(a)L(\omega,a)dP_{N_\omega} = \int C\ L(\omega,a)dP_{N_\omega} \to \int C\ L(\omega,a)dP_\omega \text{ a.s. } [P^\infty],$$

which may not be zero.

For example, let P_ω be the uniform distribution on $[0,1] \cup [\omega,\omega+1]$ for $\omega = 1,2,3$ with $L(\omega,1) = \omega = 4 - L(\omega,2)$. Let $C = [0,1]$. Then since $F(\omega,x) = \frac{1}{3}[0,1](x) + [\omega,\omega+1](x)$, we have

$$\mathfrak{F} = \{F: C \subset Q^o_{F\{1,2\}}\} = \{F: \sum_{\omega=1}^{3} F_\omega(L(\omega,1) - L(\omega,2))F(\omega,x) = 0\} = \{F: F_1 = F_3\},$$

if $\varepsilon = 0$, so that card C = card $\mathfrak{F}$. Then with J mapping $\mathfrak{F}$ onto the finite subsets of C, and $\nu_{F\{1,2\}x}(1) = J(F)(x)$, we have

$$\sup_{\mathscr{F}} \left|\int Q^o_{F\{1,2\}}\nu_{F\{1,2\}}(1)L(1,1)d(P_{N_1} - P_1)\right| = \sup_{\mathfrak{F}} \left|(P_{N_1} - P_1)(J(F))\right| = P_{N_1}[0,1].$$

Since $P_{N_1}[0,1] \to P_1[0,1] = \frac{1}{2}$, the equivalent of Lemma 1 does not hold in this case.

§1.5 Uniformly ϵ-Bayes Procedures. The sets Q^{ϵ}_{FB}. Equivariant Procedures.

In this section we define the procedures to be discussed in section 1.6 for finite action spaces, and in section 1.7 for infinite action spaces. We also prove, in a more general context, the claim of section 1.1, that the sets Q^{o}_{FB}, $F \in \mathscr{F}$, $B \subset A$, are elements of $\mathcal{K}^{t}_{s}$ for some t and s.

Definition 5. A measurable component procedure, φ, is uniformly ϵ-Bayes against a distribution $F \in \mathscr{F}$ if $\varphi(x)(B(F,x,\epsilon)) = 1$ for all x, where

$$B(F,x,\epsilon) = \{a\colon F[Z(\omega,a-b,x)] \le \epsilon/m \text{ for all } b \in A\}. \tag{1.11}$$

If φ is uniformly ϵ-Bayes against F, then

$$\begin{aligned} R(F,\varphi) &= F[\textstyle\int L(\omega,\varphi)\,dP] = F[\textstyle\int Z(\omega,\varphi(x),x)\,d\mu(x)] \\ &= \textstyle\int F[Z(\omega,\varphi(x),x)]\,d\mu(x) \le R(F) + \epsilon , \end{aligned} \tag{1.12}$$

since $F[Z(\omega,\varphi(x),x)] \le \min_{a\in A} F[Z(\omega,a,x)] + \epsilon/m$, and $\mu(\mathfrak{X}) = m$. The change in the order of integration is justified by the finiteness of Ω.

Hence a uniformly ϵ-Bayes procedure is ϵ-Bayes in the usual sense.

Lemma 2. Let $Q^{\epsilon}_{FB} = \{x\colon B(F,x,\epsilon) = B\}$. Then for each $F \in \mathscr{F}$, $B \subset A$ and $\epsilon \ge 0$, $Q^{\epsilon}_{FB} \in \mathcal{K}^{t}_{s}$ for $t = \sum_{k=0}^{r-1} n^{2k}$, and $s = n^2(1+r)$ where $r = n^n$.

Proof. Let $T_{Fba} = \{x\colon F[Z(\omega,b-a,x)] \le \epsilon/m\}$. Then T_{Fba} is the Z-inverse of a half space, so is an element of $\mathcal{K}_1$. Since

$$Q^{\epsilon}_{FB} = \{x: B \subset B(F,x,\epsilon)\} \cap \{x: (A\sim B) \cap B(F,x,\epsilon) = \phi\}$$

$$= \bigcap_{b\in B} \bigcap_{a\in A} T_{Fba} \quad \bigcap \quad \bigcap_{d\in A\sim B} \bigcup_{e\in A} T^{c}_{Fde} ,$$

and, from (1.4) and the fact that inverses of functions preserve unions and intersections, $\bigcap_{b\in B} \bigcap_{a\in A} T_{Fba} \in \mathcal{X}_q$ and $\bigcap_{d\in A\sim D} \bigcup_{e\in A} T^{c}_{Fde} \in \mathcal{X}^{r}_{q}$, where $q = n^2$ and $r = n^n$. Hence, by (1.4), $Q^{\epsilon}_{FB} \subset \mathcal{X}^{t}_{s}$ where $t = \sum_{k=0}^{r-1} q^k$ and $s = q(1+r)$. The proof is complete.

Since ϵ is fixed in our discussion, we shall abbreviate $B(F,x,\epsilon)$ and Q^{ϵ}_{FB} to $B(F,x)$ and Q_{FB} respectively, in future.

Definition 6. For each $F \in \mathcal{B}$, let Φ_{Fu} be the set of component procedures uniformly ϵ-Bayes against F; and for $\hat{G}$ an estimator, let $\Phi_{\hat{G}u} = \{\underline{\varphi}: \forall\, \underline{x}, \exists\; \varphi^{o}(\underline{x}) \in \Phi_{\hat{G}(\underline{x})u}$ such that, $\forall\, r,\ \varphi_r(\underline{x}) = \varphi^{o}(\underline{x})(x_r)\}$.

From (1.12), $\Phi_{\hat{G}u} \subset \Phi_{\hat{G}}$ for every $\hat{G}$.

Let $g(1,2,\ldots,N) = (g1,g2,\ldots,gN)$ be an arbitrary permutation of $(1,2,\ldots,N)$, and, for any vector $\underline{r} = (r_1,\ldots,r_N)$, let $g\underline{r} = (r_{g1},r_{g2},\ldots,r_{gN})$. Let $\mathfrak{S}$ be the set of permutations on $(1,2,\ldots,N)$.

Definition 7. A procedure $\underline{\varphi}$ is equivariant if, for each N, $\underline{x}$ and $g \in \mathfrak{S}$, $\underline{\varphi}(g\underline{x}) = g\underline{\varphi}(\underline{x})$, i.e., for each r, $\varphi_r(g\underline{x}) = \varphi_{gr}(\underline{x})$.

Definition 8. $\Phi^{*}_{\hat{G}}$ is the set of equivariant members of $\Phi_{\hat{G}u}$.

In sections 1.6 and 1.7 we establish asymptotic results for $D(\underline{\theta},\underline{\varphi})$ for the classes $\Phi^{*}_{\hat{G}}$ with $\hat{G}$ uniformly consistent.

§1.6 Invariant Estimates. Convergence of $D(\underline{\theta},\underline{\varphi})$ for Equivariant, Uniformly ϵ-Bayes Procedures. (Finite Action Spaces)

The main result of this section is the convergence of $D(\underline{\theta},\underline{\varphi})$ for procedures in the class $\Phi^{*}_{\hat{G}}$, to be proved in Theorem 2. In proving this result it will be convenient to make use of the invariance of the estimator $\hat{G}$, and we proceed now to show that no loss of generality occurs if this property is assumed.

Definition 9. An estimator $\hat{G}$ is invariant if, for each N, each $\underline{x}$ and each $g \in \mathfrak{S}$, $\hat{G}(\underline{x}) = \hat{G}(g\underline{x})$.

We note a particular class of procedures where the invariance of $\hat{G}$ implies the equivariance of the procedure. This occurs when the choice of the procedure to be used in the component problems depends on $\underline{x}$ only through $\hat{G}(\underline{x})$; i.e. if Ψ maps $\mathcal{J}$ into the class of component procedures and $\varphi_r(\underline{x}) = \Psi_{\hat{G}(\underline{x})}(x_r)$ for each $\underline{x}$ and r, then $\underline{\varphi}$ is equivariant if $\hat{G}$ is invariant, for $\varphi_r(g\underline{x}) = \Psi_{\hat{G}(g\underline{x})}(x_{gr}) = \Psi_{\hat{G}(\underline{x})}(x_{gr}) = \varphi_{gr}(\underline{x})$. We make use of this in Theorem 2.

Lemma 3. Let $\hat{G}$ be an estimator and let $\hat{\hat{G}}(\underline{x}) = E[\hat{G}(g\underline{x})]$, where E denotes expectation under the distribution with mass $\frac{1}{N!}$ at each $g \in \mathfrak{S}$. Then $\hat{\hat{G}}$ is invariant, $\Phi^*_{\hat{G}} \subset \Phi^*_{\hat{\hat{G}}}$, and $\hat{\hat{G}}$ is uniformly consistent if $\hat{G}$ is.

Proof. Clearly $\hat{\hat{G}}$ is invariant. Suppose $\underline{\varphi} \in \Phi^*_{\hat{G}}$. We shall show that $\underline{\varphi} \in \Phi^*_{\hat{\hat{G}}}$.

To show this, we need to show that $\varphi_r(\underline{x})(B(\hat{\hat{G}}(\underline{x}),x_r)) = 1$ for all $\underline{x}$ and r. Given $\underline{x}$, r and g, let $gk = r$. Then, since $\underline{\varphi} \in \Phi_{\hat{G}u}$, we have $\varphi_k(g\underline{x})(B(\hat{G}(g\underline{x}),x_{gk})) = 1$. Thus, since $\underline{\varphi}$ is equivariant, $\varphi_{gk}(\underline{x})(B(\hat{G}(g\underline{x}),x_{gk})) = 1$, so that $\varphi_r(\underline{x})(B(\hat{G}(g\underline{x}),x_r)) = 1$ for all g, $\underline{x}$ and r, i.e. $\varphi_r(\underline{x})[\bigcap_{g\in\mathfrak{S}} B(\hat{G}(g\underline{x}),x_r)] = 1$. Thus $\underline{\varphi} \in \Phi^*_{\hat{\hat{G}}}$ if $\bigcap_{g\in\mathfrak{S}} B(\hat{G}(g\underline{x}),x_r) \subset B(\hat{\hat{G}}(\underline{x}),x_r)$. For any $a \in \bigcap_{g\in\mathfrak{S}} B(\hat{G}(g\underline{x}),x_r)$, $\sum_{\omega=1}^{m} \hat{G}_\omega(g\underline{x})Z(\omega,a-b,x_r) \le \epsilon/m$ for all $b \in A$ and all $g \in \mathfrak{S}$. Hence $\frac{1}{N!}\sum_{g\in\mathfrak{S}}\sum_{\omega=1}^{m} \hat{G}_\omega(g\underline{x})Z(\omega,a-b,x_r) \le \epsilon/m$ for all $b \in \mathbf{A}$, so that $a \in B(\hat{\hat{G}}(\underline{x}),x_r)$ as required. Hence $\Phi^*_{\hat{G}} \subset \Phi^*_{\hat{\hat{G}}}$.

Suppose that $\hat{G}$ is uniformly consistent, and let α, β, γ and δ be arbitrary positive numbers. We write $|G|$ for $|G|(\Omega)$. Then for $N > N(\gamma,\delta)$ we have $P^{\infty}[|\hat{G}(\underline{Y}(\underline{\theta})) - G_N| > \gamma] < \delta$ for all $\underline{\theta}$, where $\underline{Y}(\underline{\theta}) = (Y_1(\theta_1),\ldots,Y_N(\theta_N))$. Hence, since $|\hat{G}-G_N| \le |\hat{G}| + |G_N| = 2$, $\int[|\hat{G}(\underline{Y}(\underline{\theta})) - G_N|]dP^{\infty}(\underline{Y}) < \gamma + 2\delta$ for all $\underline{\theta}$. Since the Y_i are i.i.d. and $G_N(g\underline{\theta}) = G_N(\underline{\theta})$ for all g and $\underline{\theta}$ (where $G_N(\underline{\theta})$ is the empirical distribution of $\theta_1,\theta_2,\ldots,\theta_N$) we have, by the transformation theorem, $\int|\hat{G}(g\underline{Y}(g\underline{\theta})) - G_N|dP^{\infty}(\underline{Y}) = \int|\hat{G}(\underline{Y}(g\underline{\theta})) - G_N|dP^{\infty}(\underline{Y}) < \gamma + 2\delta$ for all $\underline{\theta}$ and g. Hence

$$\int\Big|\frac{1}{N!}\sum_{g\in\mathfrak{S}}\hat{G}(g\underline{Y}(g\underline{\theta})) - G_N\Big|dP^{\infty}(Y) \le \frac{1}{N!}\sum_{g\in\mathfrak{S}}\int|\hat{G}(g\underline{Y}(g\underline{\theta})) - G_N|dP^{\infty}(\underline{Y})$$

$$< \gamma + 2\delta.$$

Thus, by the Markov inequality, $P^{\infty}[|\frac{1}{N!}\sum_{g\in\mathfrak{S}} G(g\underline{Y}(g\underline{\theta})) - G_N| > \alpha] < \frac{\gamma+2\delta}{\alpha}$. Since $g\underline{x} = g\underline{Y}(g\underline{\theta})$ for each $\underline{x}$, g and $\underline{\theta}$ we have that if $N > N(\gamma,\delta)$, with $\gamma + 2\delta < \alpha\beta$,

$$P^{\infty}[|\hat{\hat{G}}(\underline{x}) - G_N| > \alpha] < \beta \quad \text{for all } \underline{\theta}.$$

The lemma is proved.

We can now state the main result of this section.

<u>Theorem 2</u>. Let $\hat{G}$ be a uniformly consistent estimator. Then $\sup_{\underline{\varphi}\in\Phi_G^*}\sup_{\underline{\theta}} D(\underline{\theta},\underline{\varphi}) < o(1) + \epsilon$ as $N \to \infty$.

<u>Proof</u>. In view of Lemma 3, it suffices to prove the result for $\hat{G}$ invariant. Let $\underline{\varphi} \in \Phi_{\hat{G}}^*$ and let $W_r(\underline{\theta},\underline{\varphi},\underline{x}) = L(\theta_r,\varphi_r(\underline{x}),x_r)$. Then $W_{gr}(\underline{\theta},\underline{\varphi},\underline{x}) = L(\theta_{gr},\varphi_{gr}(\underline{x}),x_{gr}) = L(\theta_{gr},\varphi_r(g\underline{x}),x_{gr}) = W_r(g\underline{\theta},\underline{\varphi}\cdot g\underline{x})$.

Let E denote expectation under the distribution with mass $\frac{1}{N!}$ at each element of $\mathfrak{S}$. Then since $N^{-1}\sum_{r=1}^{N} h(r) = E[h(gN)]$ we have

$$R(\underline{\theta},\underline{\varphi}) = \int E[W_{gN}(\underline{\theta},\underline{\varphi},\underline{x})] \prod_{k=1}^{N} f(\theta_k,x_k)\,d\mu^N(\underline{x})$$

$$= E[\int W_N(g\underline{\theta},\underline{\varphi},g\underline{x}) \prod_{k=1}^{N} f(\theta_{gk},x_{gk})\,d\mu^N(\underline{x})]$$

$$= E[\int W_N(g\underline{\theta},\underline{\varphi},\underline{x}) \prod_{k=1}^{N} f(\theta_{gk},x_k)\,d\mu^N(\underline{x})]$$

by the transformation theorem, since μ^N is invariant under permutations of $\underline{x}$. Noting that $E[\theta_{gN} = \omega] = \frac{N_\omega}{N}$, we have

$$R(\underline{\theta},\underline{\varphi}) = \int E[L(\theta_{gN},\varphi_N(\underline{x}),x_N) \prod_{k=1}^{N} f(\theta_{gk},x_k)]\,d\mu^N(\underline{x})$$

$$= \int E[E[L(\theta_{gN},\varphi_N(\underline{x}),x_N) \prod_{k=1}^{N} f(\theta_{gk},x_k) \mid \theta_{gN} = \omega]]\,d\mu^N(\underline{x})$$

$$(1.13) \qquad = \int \sum_{\omega=1}^{m} \frac{N_\omega}{N} L(\omega,\varphi_N(\underline{x}),x_N) f(\omega,x_N) E[\prod_{k=1}^{N-1} f(\theta_{gk},x_k) \mid \theta_{gN} = \omega]\,d\mu^N(\underline{x}).$$

Denoting the integrand of (1.13) by $T(\varphi_N(\underline{x}),\underline{x})$ we have, with $\hat{B}(r) = B(\hat{G}(\underline{x}),x_r)$,

$$(1.14) \qquad R(\underline{\theta},\underline{\varphi}) \le \int \max_{a\in\hat{B}(N)} T(a,\underline{x})\,d\mu^N(\underline{x}) \quad \text{for all } \underline{\varphi} \in \Phi^*_{\hat{G}}.$$

Let $\underline{\zeta} \in \Phi_{\hat{G}u}$ be given by $\zeta_r(\underline{x})(a) = 1$ for $a = a_{\hat{B}(r)}(x_r)$, where $a_B(x)$ is the first maximizer, among elements of B, of $G_N[Z(\omega,a,x)]$. One might expect $\underline{\zeta}$ to do about as badly as possible against $\underline{\theta}$ since it "plays anti-Bayes" against G_N within the restrictions imposed by membership of $\Phi_{\hat{G}u}$. We shall show that this is, in fact, the case. Since $\hat{G}$ is invariant, $\underline{\zeta}$ is equivariant (see, e.g., the remarks following Definition 9). Hence

$$\int \max_{a\in\hat{B}(N)} T(a,\underline{x})\,d\mu^N(\underline{x}) - R(\underline{\theta},\underline{\zeta}) = \sum_{B\subset A} \int [B=\hat{B}(N)]\{\max_{a\in B} T(a,\underline{x}) - T(a_B(x_N),\underline{x})\}\,d\mu^N(\underline{x})$$

$$(1.15) \qquad \le \sum_{B\subset A} \int \max_{a\in B} T(a,\underline{x}) - T(a_B(x_N),\underline{x})\,d\mu^N(x),$$

since each integrand is positive. We show the right side is $O(N^{-\frac{1}{2}})$ as $N \to \infty$.

For each B, consider the problem obtained from the present problem by truncating the action space to B and using the loss function $L_B(\omega,a,x) = \sum_{b\in B} L(\omega,b,x) - L(\omega,a,x)$.

Replacing L by L_B in (1.13) and interchanging orders of summation, we see that the risk of an equivariant procedure, $\underline{\Psi}$, in this new game is

$$(1.16) \qquad R_B(\underline{\theta},\underline{\Psi}) = \int \sum_{b\in B} T(b,\underline{x}) - T(\Psi_N(\underline{x}),\underline{x})d\mu^N(\underline{x}).$$

In particular, the best equivariant procedure has risk

$$(1.17) \qquad R_B^*(G_N) = \int \sum_{b\in B} T(b,\underline{x}) - \max_{a\in B} T(a,\underline{x})d\mu^N(\underline{x})$$

and the best simple symmetric procedure has risk

$$(1.18) \qquad R_B(G_N) = \int \sum_{b\in B} T(b,\underline{x}) - T(a_B(x_N),\underline{x})d\mu^N(\underline{x})$$

since simple symmetric procedures are equivariant. We obtain (1.18) from (1.16) by taking $\underline{\Psi}$ to be the simple symmetric procedure whose kernel, in the r^{th} problem, is degenerate at the first minimizer in B of $G_N[L_B(\omega,a,x_r)f(\omega,x_r)] = G_N[f(\omega,x_r)\sum_{b\in B} L(\omega,b,x_r)] - G_N[Z(\omega,a,x_r)]$, i.e. at $a_B(x_r)$, the first maximizer of $G_N[Z(\omega,a,x_r)]$ in B.

Substituting into the left side of (1.15) the left of (1.14) and replacing the right of (1.15) by the difference of the left sides of (1.18) and (1.17), we obtain, for all $\underline{\varphi} \in \Phi_{\hat{G}}^*$,

$$(1.19) \qquad R(\underline{\theta},\underline{\varphi}) - R(\underline{\theta},\underline{\zeta}) \le \sum_{B\subset A} \{R_B(G_N) - R_B^*(G_N)\}.$$

Hannan and Huang (1969) have shown that each summand on the right of (1.19) is bounded by $O(N^{-\frac{1}{2}})$ uniformly in $\underline{\theta}$. Hence

$$(1.20)\quad \sup_{\varphi\in\Phi^*_{\hat G}} \sup_{\underline{\theta}} D(\underline{\theta},\underline{\varphi}) \le \sup_{\underline{\theta}}\{R(\underline{\theta},\underline{\zeta}) - R(G_N)\} + O(N^{-\frac{1}{2}}).$$

We now show that $\underline{\zeta}$ is a half-space procedure, so that, from the corollary to Theorem 1, $\sup_{\underline{\theta}}\{R(\underline{\theta},\underline{\zeta}) - R(G_N)\} < o(1) + \varepsilon$. We first note that if ζ^o is the function given by Definition 6 then $\{\zeta^o(\underline{x})(y)\colon y \in \mathfrak{X}\} \subset \{a_1,a_2,\ldots,a_n\}$ (where "a" denotes the measure degenerate at a) for each $\underline{x}$. It remains only to show that, for each $\underline{x}$, $\{y\colon \zeta^o(\underline{x})(y) = a\} \in \mathcal{K}^t_s$ for some t and s. But

$$\{y\colon \zeta^o(\underline{x})(y) = a\} = \bigcup_{\{B\subset A:\,a\in B\}} (Q_{\hat G(\underline{x})B} \cap Q_{G_N Ba})$$

where

$$\begin{aligned} Q_{G_N Ba} &= \{x\colon a_B(x) = a\} \\ &= \bigcap_{\substack{b\in B\\ b<a}} \{x\colon G_N[Z(\omega,b-a,x)] < 0\} \cap \bigcap_{\substack{b\in B\\ b>a}} \{x\colon G_N[Z(\omega,b-a,x)] \le 0\} \\ &\in \mathcal{K}_{n-1}. \end{aligned}$$

Since we already have $Q_{\hat G(\underline{x})B} \in \mathcal{K}^t_s$ from Lemma 2, an application of (1.4) yields the result.

Hence $\underline{\zeta} \in \Phi_{\hat G,s,t,n}$ for some bounded s and t, so we can apply the corollary to Theorem 1 in (1.20) to get

$\sup_{\Phi^*_{\hat G}} \sup_{\underline{\theta}} D(\underline{\theta},\underline{\varphi}) < o(1) + \varepsilon$ as $N \to \infty$, as required.

§1.7 Convergence of $D(\underline{\theta},\underline{\varphi})$ for Equivariant Uniformly ε-Bayes Procedures. (Totally Bounded Action Spaces)

In this section we replace the assumption that A is finite by

$$(1.21)\quad A \text{ is totally bounded in the metric } d(a,a') = \sup_{x,\omega} |L(\omega,a-a',x)|.$$

(Since, in fact, we deal only with uniformly ε-Bayes procedures, it is sufficient that some totally bounded subset of A contain

$\bigcup_{\substack{F\in\mathcal{G}\\ x\in\mathfrak{X}}} B(F,x)$.) We now state and prove the result analogous to Theorem 2.

<u>Theorem 3</u>. If (1.21) holds and $\hat{G}$ is uniformly consistent,

$$\sup_{\Phi^*_{\hat{G}}} \sup_{\underline{\theta}} D(\underline{\theta},\underline{\varphi}) < \varepsilon + o(1) \quad \text{as} \quad N \to \infty.$$

<u>Proof</u>. For each $\delta > 0$, let $D_\delta = \{a_1,a_2,\dots,a_k\}$, $k = k(\delta)$, be such that, for any $a \in A$, $d(a,a_j) < \delta$ for some $a_j \in D_\delta$, where d is the metric given in (1.21).

Let $\underline{\varphi} \in \Phi^*_{\hat{G}}$. Fix δ and let $\{A_j: j = 1,2,\dots,k\}$ be a partition of A such that, for each j, $d(a,a_j) < \delta$ for every $a \in A_j$.

Consider the reduced problem obtained by replacing A by D_δ. Let $\Phi_{\hat{G}u\delta}$ and $\Phi^*_{\hat{G}\delta}$ satisfy Definitions 6 and 8 (for "$\Phi_{\hat{G}u}$" and "$\Phi^*_{\hat{G}}$"), with "ε" replaced by "$\varepsilon+m\delta$"; and let $R_\delta(\cdot)$ be the Bayes envelope for this reduced game.

We observe that, for any $G \in \mathcal{G}$, we have

$$\min_{D_\delta} G[Z(\omega,a_j,x)] - \inf_{A} G[Z(\omega,a,x)] < \delta \max_{\omega} f(\omega,x) < \delta,$$

since $f(\omega,x) \le 1$ for all ω and x. Integrating this inequality with respect to μ, we have, since $\mu(\mathfrak{X}) = m$,

$$(1.22) \qquad R_\delta(G) - R(G) < m\delta \quad \text{for all} \quad G \in \mathcal{G}.$$

Let $\underline{\zeta} = \underline{\zeta}(\underline{\varphi})$ be the procedure in the reduced game given by $\zeta_r(\underline{x})(a_j) = \varphi_r(\underline{x})(A_j)$ for all r, $\underline{x}$ and j. We show that $\underline{\zeta} \in \Phi^*_{\hat{G}\delta}$.

First, since $\zeta_r(g\underline{x})(a_j) = \varphi_r(g\underline{x})(A_j) = \varphi_{gr}(\underline{x})(A_j) = \zeta_{gr}(\underline{x})(a_j)$, $\underline{\zeta}$ is equivariant. It remains to show that $\underline{\zeta} \in \Phi_{\hat{G}u\delta}$.

Let φ^o be the function, given by the definition of $\Phi_{\hat{G}u}$ (Definition 6), corresponding to $\underline{\varphi}$. For each $\underline{x} \in \mathfrak{X}^N$, $y \in \mathfrak{X}$ and $a_j \in D_\delta$, let $\zeta^o(\underline{x})(y)(a_j) = \varphi^o(\underline{x})(y)(A_j)$. Then for each $\underline{x}$ and r,

$\zeta_r(\underline{x}) = \zeta^o(\underline{x})(x_r)$. If $\zeta^o(\underline{x})(y)(a_j) > 0$ for some $\underline{x}$, y and j, then $\varphi^o(\underline{x})(y)(A_j) > 0$. Since $\varphi^o(\underline{x}) \in \Phi_{\hat{G}(\underline{x})u}$, there exists an $a \in A_j$ such that $\hat{G}(\underline{x})[Z(\omega,a-b,y)] \leq \epsilon/m$ for all $b \in A$. By definition of A_j this implies that $\hat{G}(\underline{x})[L(\omega,a_j-b,y)f(\omega,y) \leq \epsilon/m + \delta]$ for all $b \in A$ (and hence for all $b \in D_\delta$). Hence $\zeta^o(\underline{x})(y)(a_j) > 0$ implies $a_j \in B_\delta(\hat{G}(\underline{x}),y,\epsilon+m\delta)$ where $B_\delta(G,x,\epsilon)$ satisfies (1.11) when A is replaced by D_δ. Hence $\zeta^o(\underline{x}) \in \Phi_{\hat{G}(\underline{x})u\delta}$ for all $\underline{x}$, so that $\underline{\zeta} \in \Phi^*_{\hat{G}\delta}$.

By definition of $\underline{\zeta}$ we have, for all $\underline{\theta}$ and $\underline{x}$,

$$|W(\underline{\theta},\underline{\varphi},\underline{x}) - W(\underline{\theta},\underline{\zeta},\underline{x})| \leq N^{-1}\sum_{r=1}^{N}|W_r(\underline{\theta},\underline{\varphi},\underline{x}) - W_r(\underline{\theta},\underline{\zeta},\underline{x})|$$

$$\leq N^{-1}\sum_{r=1}^{N}\sum_{j=1}^{n}\left|\int_{A_j} L(\theta_r,a,x_r)\varphi_r(\underline{x},da) - L(\theta_r,a_j,x_r)\varphi_r(\underline{x},A_j)\right|$$

$$\leq N^{-1}\sum_{r=1}^{N}\sum_{j=1}^{n}\int_{A_j}|L(\theta_r,a,x_r) - L(\theta_r,a_j,x_r)|\varphi_r(\underline{x},da)$$

$$\leq N^{-1}\sum_{r=1}^{N}\sum_{j=1}^{n}\delta\,\varphi_r(\underline{x},A_j) = \delta.$$

Integrating this inequality we obtain,

$$R(\underline{\theta},\underline{\varphi}) - R(\underline{\theta},\underline{\zeta}) < \delta \quad \text{for all } \underline{\theta}. \tag{1.23}$$

Thus, for any $\delta > 0$,

$$\sup_{\Phi^*_G}\sup_{\underline{\theta}} D(\underline{\theta},\underline{\varphi}) \leq \sup_{\Phi^*_{\hat{G}}}\sup_{\underline{\theta}}\{|R(\underline{\theta},\underline{\varphi})-R(\underline{\theta},\underline{\zeta}(\underline{\varphi}))|+|R(\underline{\theta},\underline{\zeta}(\underline{\varphi})-R_\delta(G_N)| + |R_\delta(G_N) - R(G_N)|\}$$

$$\leq \delta + \sup_{\Phi^*_{\hat{G}\delta}}\sup_{\underline{\theta}}|R(\underline{\theta},\underline{\zeta})-R_\delta(G_N)| + m\delta, \text{ by (1.22) and (1.23).} \tag{1.24}$$

From Theorem 2 there is a function $N_\delta(\gamma) = N(\gamma,\delta)$ such that $N > N(\gamma,\delta)$ implies $\sup_{\Phi^*_{\hat{G}\delta}}\sup_{\underline{\theta}}|R(\underline{\theta},\underline{\zeta}) - R_\delta(G_N)| < \gamma + \epsilon + m\delta$.

Substituting this in (1.24), with $\delta(\gamma) = \frac{\gamma}{2(2m+1)}$, we have, for $N > N'(\gamma) = N(\frac{\gamma}{2}, \delta(\gamma))$, $\sup_{\Phi_G^*} \sup_{\underline{\theta}} D(\underline{\theta},\underline{\varphi}) < \gamma + \epsilon$ as required.

Remark. We again note that Theorems 2 and 3 continue to hold if ϵ depends on N.

CHAPTER II

INFINITE STATE SPACES

§2.0 Introduction.

When Ω is infinite, we face two problems not encountered earlier. Solutions to the problem of estimating the empirical distribution G_N from the observations $x_1, x_2, \ldots, x_N$ are not known in general. In what follows we simply assume the existence of appropriate estimators, and we will not discuss this question further except to mention the work of Fox (1968, Chapter III) in the case where the distribution P_ω is the uniform distribution on $[0,\omega]$ $(0 < \omega < \infty)$ and the case where P_ω is the uniform distribution on $[\omega,\omega+1]$ $(-\infty < \omega < \infty)$. Also, appropriate forms of Lemma 1 are not available because, among other things, of the partial failure of the Glivenko-Cantelli theorem in infinite dimensional spaces (see, e.g., Sazonov (1963)). The convergence for which Lemma 1 was used, however, could be expected if the sample space, $\mathfrak{X}$, were finite, since we would then be concerned with a supremum over a finite number of sets. However if $\mathfrak{X}$ is finite the problem of estimating G_N is virtually incapable of solution, since the distributions $\{P_\omega : \omega \in \Omega\}$ would not be linearly independent if Ω has more elements than $\mathfrak{X}$. One seems to need, then, an infinite sample space to allow the estimation and a finite sample space to ensure the convergence needed for the asymptotic optimality of the "Bayes against the estimate" procedures. It is these considerations that motivate this chapter.

In section 2.1 we define terms to be used in the following sections, and outline the basic approach. The main results of the chapter are in section 2.2, and this is followed, in section 2.3, by an attempt to show that, under reasonable conditions, constant terms appearing in the bounds in section 2.2 can be controlled.

§2.1 Finitely Based Decision Procedures.

Let $\mathcal{V}$ be a finite measurable partition of $\mathfrak{X}$, and for each $x \in \mathfrak{X}$ let x' be the member of $\mathcal{V}$ to which x belongs and $L(\omega,a,x')$ be the value of $L(\omega,a,y)$ at a fixed, but arbitrary, point $y \in x'$. As before, let $\mathcal{S}$ be the set of distributions on Ω.

Definition 10. For each $\omega \in \Omega$, let $P_{\omega\mathcal{V}}$ be the distribution on $\mathcal{V}$ induced by P_ω on $\mathcal{B}$. For the component game obtained by replacing $\mathfrak{X}$ by $\mathcal{V}$, P_ω by $P_{\omega\mathcal{V}}$ and $L(\omega,a,x)$ by $L(\omega,a,V)$, let $R_{\mathcal{V}}(\cdot)$ be the Bayes envelope, $R_{\mathcal{V}}(G,\varphi)$ the risk of a $\mathcal{V}$-measurable component procedure φ against $G \in \mathcal{S}$, $\Phi_{\mathcal{V}}$ the set of component procedures and, for each $F \in \mathcal{S}$, $\Phi_{F\mathcal{V}}$ the set of component procedures ε-Bayes against F. Component procedures available in the reduced game are also available in the original game in the sense that if $\varphi \in \Phi_{\mathcal{V}}$, the procedure Ψ in the original game, given by $\Psi(x) = \varphi(x')$ for every x, can be identified with φ. Since $\mathcal{V} \subset \mathcal{B}$, any $\mathcal{V}$-measurable procedure in the reduced game is $\mathcal{B}$-measurable in the original game. In the context of the original problem, a procedure $\varphi \in \Phi_{F\mathcal{V}}$ will be called "ε-$\mathcal{V}$-Bayes" against F.

For each F and $G \in \mathcal{S}$, let

$$(2.1) \qquad \lambda(F,G) = \sup_{\nu,\Psi\in\Phi_{\mathcal{V}}} \sum_{V\in\mathcal{V}} (G-F)[L(\omega,\nu(V)-\Psi(V),V)P_\omega(V)].$$

Then, from Lemma 0, for each F and $G \in \mathcal{S}$, and $\varphi \in \Phi_{\mathcal{V}}$,

$$(2.2) \qquad R_{\mathcal{V}}(G,\varphi) - R_{\mathcal{V}}(G) \le \lambda(F,G) + R_{\mathcal{V}}(F,\varphi) - R_{\mathcal{V}}(F)$$

where the interchange of the order of integration, for the term $\lambda(F,G)$, is justified by the finiteness of $\mathcal{V}$.

The main idea in what follows is that, if $\mathcal{V}$ is a "good" approximation to $\mathfrak{X}$ in the sense that both $|L(\omega,a,x) - L(\omega,a,x')|$ and $\sum_{V \in \mathcal{V}} \int V(f_\omega - \frac{P_\omega(V)}{\mu(V)})^+ d\mu$ are small for every ω, x, a and V (where $f_\omega = \frac{dP_\omega}{d\mu}$ for some measure μ), then $R_{\mathcal{V}}(\cdot)$ might be close to $R(\cdot)$. Then we might use $\hat{G}(x_1,\ldots,x_N)$ to estimate G_N, play ε-$\mathcal{V}$-Bayes against $\hat{G}(\underline{x})$ in each component problem, and use the finiteness of $\mathcal{V}$ to obtain the convergence which, in Theorem 1, came as a result of Lemma 1.

§2.2 Convergence Theorems for ε-$\mathcal{V}$-Bayes Compound Procedures.

In this section we give conditions under which the risk of an "ε-$\mathcal{V}$-Bayes against $\hat{G}$" procedure is close to $R_{\mathcal{V}}(G_N)$, and give a bound on the difference $R_{\mathcal{V}}(G_N) - R(G_N)$. These results are drawn together for a general theorem on the convergence to $R(G_N)$ of these procedures.

Definition 11. For $\hat{G}$ an estimator, let

$$\Phi_{\hat{G}\mathcal{V}} = \{\underline{\varphi}: \forall\, \underline{x}, \exists\, \varphi^o(\underline{x}) \in \Phi_{\hat{G}(\underline{x})\mathcal{V}} \text{ such that}, \forall\, r,\ \varphi_r(\underline{x}) = \varphi^o(\underline{x})(x_r)\}.$$

Theorem 4. Let Ω be totally bounded in the metric $d_1(\omega,\omega') = \sup\{|L(\omega,a,V) - L(\omega',a,V)|: a \in A,\ V \in \mathcal{V}\}$. Let

$$(2.4) \quad \delta(\mathcal{V}) = \sup\{|L(\omega,a,x) - L(\omega,a,y)| : \omega \in \Omega,\ a \in A,\ x' = y'\}$$

be finite and let $M < \infty$ be the uniform bound on L implied by these conditions.

Then with $\delta = \delta(\mathcal{V})$, there is a function $N(\eta,\gamma)$, defined for all $\eta > 0$ and $\gamma > 0$ such that, for all $\underline{\theta} \in \Omega^\infty$ and $\underline{\varphi} \in \Phi_{\hat{G}\mathcal{V}}$,

$$P^\infty[\bigcup_{N>N(\eta,\gamma)} \{W(\underline{\theta},\underline{\varphi},\underline{x}) - \lambda(\hat{G}(\underline{x}),G_N) > R_{\mathcal{V}}(G_N) + \epsilon + 5\delta + \eta\}] < \gamma.$$

Proof. Let $E_1,\ldots,E_k$ be a partition of Ω by sets of d_1-diameter $< \delta$, so that for each i, $\sup_a |L(\omega,a,x) - L(\omega',a,x')| < 2\delta$ whenever $\omega,\omega' \in E_i$. For each i, let ω_i be an arbitrary fixed element of E_i.

Let $\Psi \in \Phi_{\mathcal{V}F}$ for some F. Then

$$|W(\underline{\theta},\Psi,\underline{x}) - N^{-1}\sum_{r=1}^{N} L(\theta_r,\Psi(x_r'),x_r')| \le N^{-1}\sum_{r=1}^{N} |L(\theta_r,\Psi(x_r'),x_r) - L(\theta_r,\Psi(x_r'),x_r')| \le \delta.$$

Since the integral of the absolute value bounds the absolute value of the integral,

$$(2.5) \qquad |R(\underline{\theta},\Psi) - R_{\mathcal{V}}(\underline{\theta},\Psi)| \le \delta.$$

Also, for $\theta_r \in E_i$, we have

$$|L(\theta_r,\Psi(x_r'),x_r) - L(\omega_i,\Psi(x_r'),x_r')| \le \int |L(\theta_r,a,x_r) - L(\omega_i,a,x_r')|\Psi(x_r')(da) \le 2\delta.$$

Hence

$$(2.6) \qquad |W(\underline{\theta},\Psi,\underline{x}) - N^{-1}\sum_{i=1}^{m}\sum_{\{r:\theta_r\in E_i\}} L(\omega_i,\Psi(x_r'),x_r')| < 2\delta.$$

Let $N_i = N(\underline{\theta},i) = \sum_{r=1}^{N}[\theta_r \in E_i]$, $\bar{P}_{N_i} = N_i^{-1}\sum_{r=1}^{N}[\theta_r \in E_i]P_{\theta_r\mathcal{V}}$, and for each $D \subset \mathcal{V}$, let $P_{N_i}(D) = N_i^{-1}\sum_{r=1}^{N}[\theta_r \in E_i][x_r' \in D]$, so that $\bar{P}_{N_i}$ is the "average" distribution on $\mathcal{V}$ arising from the θ_r's in E_i, and P_{N_i} is the corresponding empirical distribution given by $\{x_r' \colon \theta_r \in E_i\}$.

Hence (2.6) becomes

$$(2.7) \qquad \left| W(\underline{\theta},\Psi,\underline{x}) - \sum_{i=1}^{m} \frac{N_i}{N} \int L(\omega_i,\Psi)\,dP_{N_i} \right| < 2\delta$$

which implies

$$\left| R(\underline{\theta},\Psi) - \sum_{i=1}^{m} \frac{N_i}{N} \int L(\omega_i,\Psi)\,d\overline{P}_{N_i} \right| < 2\delta,$$

so that, from (2.5), we have

$$(2.8) \qquad \left| R_V(\underline{\theta},\Psi) - \sum_{i=1}^{m} \frac{N_i}{N} \int L(\omega_i,\Psi)\,d\overline{P}_{N_i} \right| < 3\delta.$$

From (2.7) and (2.8) we have

$$(2.9) \quad W(\underline{\theta},\Psi,\underline{x}) \le 5\delta + \sum_{i=1}^{m} \frac{N_i}{N} \int L(\omega_i,\Psi)\,d(P_{N_i}-\overline{P}_{N_i}) + R_V(\underline{\theta},\Psi).$$

Let $\underline{\varphi} \in \Phi_{\hat{G}V}$ and let φ^o be the function guaranteed by Definition 11. Then since $W(\underline{\theta},\underline{\varphi},\underline{x}) = W(\underline{\theta},\varphi^o(\underline{x}),\underline{x})$ for each $\underline{x}$, we have, from (2.9) and using (2.2) and the definition of Φ_{FV} to bound $R_V(\underline{\theta},\varphi^o(\underline{x}))$,

$$(2.10) \quad W(\underline{\theta},\underline{\varphi},\underline{x}) \le 5\delta + \sum_{i=1}^{m} \frac{N_i}{N} S(N,\underline{\theta},E_i,\underline{x}) + R_V(G_N) + \epsilon + \lambda(\hat{G}(\underline{x}),G_N)$$

where $S(N,\underline{\theta},E_i,\underline{x}) = M \sum_{V\in\mathcal{V}_m} |P_{N_i} - \overline{P}_{N_i}|(V)$.

We now show that $\sum_{i=1}^{m} \frac{N_i}{N} S(N,\underline{\theta},i,\underline{x}) \to 0$ a.s. (P^∞) uniformly in $\underline{\theta}$ as $N \to \infty$, by considering the 4^{th} moment of $|P_{N_i} - \overline{P}_{N_i}|(V)$. We have

$$\int |(\tilde{P}_{N_i} - \overline{P}_{N_i})(V)|^4 dP^\infty = N_i^{-4}\Big\{ \sum_{\{r:\theta_r\in E_i\}} \int (V(x) - P_{\theta_r}(V))^4 dP_{\theta_r}(x)$$

$$+ 6 \sum_{\{r\neq s:\theta_r,\theta_s\in E_i\}} [\int (V(x)-P_{\theta_r}(V))^2 dP_{\theta_r}(x)][\int (V(x)-P_{\theta_s}(V))^2 dP_{\theta_s}(x)]\Big\}$$

$$\le N_i^{-4}\{N_i + 6N_i(N_i - 1)\} < 6N_i^{-2}$$

where we use the independence and zero expectation of the terms $V(x_r) - P_{\theta_r}(V)$, $r = 1,2,\ldots,N$.

Hence, by the Markov inequality, given $\eta > 0$,

$$P^{\infty}[\sup_{N>N'} S(N,\underline{\theta},i,\underline{x}) > \eta] \le \sum_{n=N'(\underline{\theta},i)}^{\infty} 6n^{-2}\eta^{-2} \to 0 \quad \text{as} \quad N'(\underline{\theta},i) \to \infty.$$

Hence there exists a function $k' = k'(\eta,\gamma)$, defined for all $\eta > 0$ and $\gamma > 0$, such that, for each i and each $\underline{\theta} \in \Omega^{\infty}$,

$$P^{\infty}[\sup_{N\in J(k',i,\underline{\theta})} S(N,\underline{\theta},i,\underline{x}) > \eta] < \gamma$$

where $J(k',i,\underline{\theta}) = \{N: N(\underline{\theta},i) > k'\}$.

Let $k = k'(\eta/m,\gamma/m)$. Then, with $H(k,i,\underline{\theta}) = \{N: N>N' \text{ and } N(\underline{\theta},i)<k\}$,

$$P^{\infty}[\sup_{N>N'} \sum_{i=1}^{m} \frac{N_i}{N} S(N,\underline{\theta},i,\underline{x})>\eta] \le \sum_{i=1}^{m} P^{\infty}[\sup_{N>N'} \frac{N_i}{N} S(N,\underline{\theta},i,\underline{x}) > \eta/m]$$

$$(2.11) \quad \le \sum_{i=1}^{m} (P^{\infty}[\sup_{N\in J(k,i,\underline{\theta})} S(N,\underline{\theta},i,\underline{x})>\eta/m] + P^{\infty}[\frac{k}{N'} \max_{N\in H(k,i,\underline{\theta})} S(N,\underline{\theta},i,\underline{x})>\eta/m])$$

$< \gamma/2 + g(N')$, where $g(N') \downarrow 0$ uniformly in $\underline{\theta}$ as $N \uparrow \infty$ since $S(N,\underline{\theta},i,\underline{x}) \le 2M < \infty$.

Theorem 4 now follows from (2.10) and (2.11).

We now deal with the term $R_{V}(G_N)$, by introducing a measure of the accuracy with which V "approximates" $\mathfrak{X}$.

Let $P_{\omega} << \mu$ for all ω, with $f_{\omega} = \frac{dP_{\omega}}{d\mu}$ and, for each x, $f_{\omega V}(x) = \frac{P_{\omega}(x')}{\mu(x')}$. Let

$$(2.12) \qquad \alpha_{\mu}(V) = \sup_{\omega} \alpha_{\mu\omega}(V) = \sup_{\omega} \int (f_{\omega} - f_{\omega V})^{+} d\mu.$$

It is clear that $\alpha_{\mu}(V)$ depends on μ. In fact, if μ is σ-finite, then $\alpha_{\nu}(V) \le \alpha_{\mu}(V)$ for any finite measure, ν, equivalent to μ and agreeing with μ on $\cup\{V \in V: \mu(V) < \infty\}$. This follows easily from the fact that $f_{\omega V}(x) = 0$ if $\mu(x') = \infty$.

<u>Definition 12</u>. For V a measurable partition of $\mathfrak{X}$, let $\alpha(V) = \inf \alpha_{\mu}(V)$, where the infimum is taken over the set of σ-finite

measures $\{\mu: P_\omega << \mu$ for all $\omega \in \Omega\}$.

Remark. We assume henceforth that Ω is separable in the metric $d(\omega,\omega') = \sup_{A \in \mathcal{B}} |P_\omega(A) - P_{\omega'}(A)|$; this implies domination of $\{P_\omega: \omega \in \Omega\}$ by a σ-finite measure.

Definition 13. For each $G \in \mathcal{J}$, let $R_1(G) = \inf R(G,\varphi)$, where the infimum is taken over all measurable procedures φ for which $G[\int L(\omega,\varphi) f_\omega d\mu] = \int G[L(\omega,\varphi) f_\omega] d\mu$.

Remark. The class of procedures for which this change of order is valid does not depend on μ, since

(i) μ can be taken to be equivalent to $\{P_\omega: \omega \in \Omega\}$ because of the separability of Ω under the metric d; and

(ii) if $\mu << \nu$ then $\frac{dP_\omega}{d\nu} = \frac{d\mu}{d\nu} f_\omega$, so that

$$\int G[\frac{dP_\omega}{d\nu} L(\omega,\varphi)] d\nu = \int \frac{d\mu}{d\nu} G[f_\omega L(\omega,\varphi)] d\nu = \int G[f_\omega L(\omega,\varphi)] d\mu.$$

Lemma 4. Let $L(\omega,a,x) \le M$ for all ω, a and x, and let $\delta = \delta(V)$ be given by (2.4). Then, with $\alpha = \alpha(V)$,

$$R_V(G) - R_1(G) \le M\alpha + \delta \tag{2.13}$$

for all $G \in \mathcal{J}$.

Proof. Let μ be any measure with $P_\omega << \mu$ for each ω, and let f_ω and $f_{\omega V}$ be as in (2.12). Let φ be any procedure for which $G[\int L(\omega,\varphi) f_\omega d\mu] = \int G[L(\omega,\varphi) f_\omega] d\mu$. Then

$$R(G,\varphi) \ge \int G[\{L(\omega,\varphi(x),x') - \delta\} f_\omega(x)] d\mu(x)$$
$$\ge \int G[L(\omega,\varphi(x),x') f_\omega(x)] d\mu(x) - \delta.$$

Since $L \le M$ and $f_\omega(x) \ge f_{\omega V}(x) - (f_{\omega V}(x) - f_\omega(x))^+$, we have

$$R(G,\varphi) + \delta \ge \int G[L(\omega,\varphi(x),x') f_{\omega V}(x)] d\mu(x) - MG[\int (f_{\omega V} - f_\omega)^+ d\mu]. \tag{2.14}$$

The first term on the right of (2.14) is bounded below by

$$(2.15)\quad \int \inf_{a\in A} G[L(\omega,a,x')f_{\omega\mathcal{V}}(x)]d\mu(x) \geq R_{\mathcal{V}}(G) - M \sum_{\mu(V)=\infty} G[P_{\infty}(V)].$$

To deal with the second term on the right of (2.14) we note that

$$\int (f_{\omega} - f_{\omega\mathcal{V}})^{+}d\mu - \int (f_{\omega\mathcal{V}} - f_{\omega})^{+}d\mu = \sum_{\mu(V)=\infty} \int V f_{\omega} d\mu + \sum_{\mu(V)<\infty} \int V(f_{\omega} - f_{\omega\mathcal{V}})d\mu$$

$$(2.16)\qquad = \sum_{\mu(V)=\infty} P_{\omega}(V).$$

Combining (2.12) and (2.16) we have, for all ω,

$$(2.17)\qquad \alpha_{\mu}(\mathcal{V}) - \sum_{\mu(V)=\infty} P_{\omega}(V) \geq \int (f_{\omega\mathcal{V}} - f_{\omega})^{+}d\mu.$$

Combining (2.14), (2.15) and (2.17) we have

$$R(G,\varphi) + \delta \geq R_{\mathcal{V}}(G) - M\alpha_{\mu}(\mathcal{V}).$$

Since φ and μ are arbitrary, the proof is complete.

<u>Corollary</u>. $R(G_N) \geq R\ (G_N) - M\alpha - \delta$ for all $\underline{\theta} \in \Omega^{\infty}$.

<u>Proof</u>. We have only to show that $R_1(G_N) - R(G_N)$. This, however, follows immediately from the fact that G_N is discrete, so the change in the order of integration in Definition 13 is valid for any measurable procedure φ.

We are now in a position to obtain a result, analogous to Theorem 1, for infinite state spaces.

<u>Theorem 5</u>. Let Ω and A be compact, L jointly continuous in ω and a for each fixed x, $P_{\omega}(V)$ continuous in ω for each $V \in \mathcal{V}$, and $\delta = \delta(\mathcal{V})$ and $\alpha = \alpha(\mathcal{V})$ as in (2.4) and Definition 12 respectively. Let M be the bound on L, and let either

(a) (Ω,d) be a metric space and $L^{*}(\hat{G}(\underline{x}),G_N) \to 0$ a.s. $[P^{\infty}]$

as $N \to \infty$ for each $\underline{\theta} \in \Omega^{\infty}$, where L^* is the Prohorov metric, or

(b) Ω be a subset of the real line and $L(\hat{G}(\underline{x}), G_N) \to 0$ a.s. $[P^{\infty}]$ as $N \to \infty$ for each $\underline{\theta} \in \Omega^{\infty}$, where L is the Lévy metric.

Then there exists a function $N(\eta,\gamma,\underline{\theta})$ such that for each $\underline{\varphi} \in \Phi_{\hat{G}V}$,

$$P^{\infty}[\sup_{N > N(\eta,\gamma,\underline{\theta})} W(\underline{\theta},\underline{\varphi},\underline{x}) - R(G_N) > \epsilon + M\alpha + 6\delta + \eta] < \gamma.$$

In addition, if the convergence of $\hat{G}(\underline{x})$ to G_N is uniform in $\underline{\theta}$ then $N(\eta,\gamma,\underline{\theta}) = N(\eta,\gamma)$.

Proof. For any ν and $\Psi \in \Phi_V$ and $V \in V$

$$|L(\omega,\nu(V) - \Psi(V),V)P_{\omega}(V) - L(\omega',\nu(V) - \Psi(V),V)P_{\omega'}(V)|$$

$$\le P_{\omega}(V)|L(\omega,\nu(V) - \Psi(V),V) - L(\omega',\nu(V) - \Psi(V),V)| + M|P_{\omega'}(V)-P_{\omega}(V)|$$

$$\le 2 \sup_a |L(\omega,a,V) - L(\omega',a,V)| + M|P_{\omega'}(V) - P_{\omega}(V)|. \tag{2.18}$$

Since Ω is compact, P_{ω} is uniformly continuous so that $|P_{\omega'}(V) - P_{\omega}(V)| \to 0$ uniformly in ω as $\omega' \to \omega$. We shall show that the same is true for $\sup_a |L(\omega,a,V) - L(\omega',a,V)|$.

Given $\rho > 0$ and $\omega \in \Omega$, for each $a \in A$ there exist open sets $U_a \subset \Omega$ and $W_a \subset A$ such that $(\omega,a) \in U_a \times W_a$, and $(\omega',a') \in U_a \times W_a \Rightarrow |L(\omega,a,V) - L(\omega',a',V)| < \rho/2$. Since A is compact, a finite covering $\{U_{a_i} \times W_{a_i}, i = 1,2,\ldots,n\}$ covers $\{\omega\} \times A$. Let $U = \bigcap_{i=1}^{n} U_{a_i}$. Then U is open, contains ω, and for any $\omega' \in U$ and $a \in A$ there is an i for which $\{(\omega,a),(\omega',a)\} \subset U_{a_i} \times W_{a_i}$, so that $|L(\omega,a,V)-L(\omega',a,V)| \le |L(\omega,a,V)-L(\omega,a_i,V)| + |L(\omega,a_i,V)-L(\omega',a,V)| < \rho$. Hence for each ω there is $\beta_{\omega}(\rho) > 0$ such that $d(\omega,\omega') < \beta_{\omega}(\rho)$ implies $\sup_a |L(\omega,a,V) - L(\omega',a,V)| < \rho$. An elementary argument, using the compactness of Ω, shows that $\inf_{\omega} \beta_{\omega}(\rho) = \beta(\rho) > 0$ for each $\rho > 0$. Hence $\sup_a |L(\omega,a,V) - L(\omega',a,V)| \to 0$ uniformly as $\omega' \to \omega$, for each V; and the above also suffices to show that Ω is compact (and hence totally bounded) in the metric $d_1(\omega,\omega') = \sup_{a,V} |L(\omega,a,V) - L(\omega',a,V)|$.

Thus both terms on the right of (2.18) tend uniformly to zero as $\omega \to \omega'$, so that, with $\alpha_h(\rho) = \sup\{h(\omega)-h(\omega'): d(\omega,\omega') < \rho\}$ and with $\mathcal{X} = \{L(\omega,\Psi(V)-\nu(V),V)P_\omega(V): V \in \mathcal{V};\ \Psi,\nu \in \Phi_{\mathcal{V}}\}$, we have $\sup_{h\in\mathcal{X}} \alpha_h(\rho) \to 0$ as $\rho \to 0$.

Hence, if (Ω,d) is a metric space, we can apply Lemma 7 in the Appendix, to get $\lambda(F,G) \to 0$ as $L^*(F,G) \to 0$; and if Ω is a subset of the real line, we can apply Lemma 8 (or 8') in the Appendix to get $\lambda(F,G) \to 0$ as $L(F,G) \to 0$.

Consequently, under either (a) or (b) we have, for each $\underline{\theta} \in \Omega^\infty$, $\lambda(\hat{G}(\underline{x}),G_N) \to 0$ a.s. $[P^\infty]$, with this convergence being uniform for $\underline{\theta} \in \Omega^\infty$ if the convergence of $\hat{G}(\underline{x})$ to G_N is uniform.

In addition, as has been shown, Ω is totally bounded in the metric $d_1(\omega,\omega') = \sup_{a,V} |L(\omega,a,V) - L(\omega',a,V)|$; thus the conclusion of Theorem 4 holds. This conclusion, with the conclusion of the corollary of Lemma 4 and the convergence of $\lambda(\hat{G}(\underline{x}),G_N)$, yields the required result.

<u>Corollary</u>. Under the conditions of Theorem 5, $\sup_{\Phi_{\hat{G}}} D(\underline{\theta},\underline{\varphi}) < o(1) + \varepsilon + M\alpha + 5\delta$, as $N \to \infty$, for each $\underline{\theta} \in \Omega^\infty$; and the convergence to 0 is uniform in $\underline{\theta}$ if the convergence of $\hat{G}(\underline{x})$ to G_N is uniform.

<u>Proof</u>. Given $\eta > 0$, by Theorem 5 we have

$$R(\underline{\theta},\underline{\varphi}) = \int W(\underline{\theta},\underline{\varphi},\underline{x})\,dP^\infty \le R(G_N) + \varepsilon + \eta + M\alpha + 6\delta$$

if $N > N(\eta/2,\ \eta/2M,\underline{\theta})$, since $W \le M$.

We remark, in concluding this section, that none of the results depend on the particular determination of $L(\omega,a,x')$. (The first condition for Theorem 4 does depend on this; however if d_1 were computed for a determination different from the one to be used, $L(\omega,a,x'')$ say, the first sentence of the proof would still hold and would imply $\sup_a |L(\omega,a,x) - L(\omega',a,x'')| < 4\delta$; and the conclusion of Theorem 4 would still hold with "5δ" replaced by "9δ".) Consequently any determination of $L(\omega,a,x')$ will suffice.

§2.3 Approximating the Sample Space by a Finite Partition.

The usefulness of Theorem 5 and its corollary will depend on the existence of estimators satisfying conditions (a) and (b) of Theorem 5, and on the availability of partitions, $\mathcal{V}$, for which $\alpha(\mathcal{V})$ and $\delta(\mathcal{V})$ are arbitrarily small. As has been said, we do not discuss the first of these problems in this thesis. The next two lemmas and the remarks which follow give a partial answer to the second.

Lemma 5. Let $P_\omega << \mu$ with $\frac{dP_\omega}{d\mu} = f_\omega$. Then, for each ω and $\alpha > 0$, there is a partition $\mathcal{V}_\omega$ such that $\alpha_{\omega\mu}(\mathcal{V}) \le \alpha$ whenever $\mathcal{V}$ is a sub-partition of $\mathcal{V}_\omega$.

Proof. Choose a_1 so that $P_\omega(f_\omega^{-1}[0,a_1)) < \alpha/3$. Let $a_o = 0$, and let $a_j = \frac{3a_{j-1}}{3-\alpha} = (\frac{3}{3-\alpha})^{j-1}a_1$ for $1 < j \le k$ where $k = \min\{j: P_\omega(f_\omega^{-1}[(\frac{3}{3-\alpha})^{j-1}a_1,\infty)) < \alpha/3\}$. Let $a_{k+1} = \infty$ and let $\mathcal{V}_\omega = \{V_0,V_1,\ldots,V_k\}$ where $V_j = f_\omega^{-1}[a_j,a_{j+1})$.

Let $\mathcal{V}$ be a sub-partition of $\mathcal{V}_\omega$. Then for $V \subset V_j$, $1 < j < k$, and $x \in V$, $(f_\omega(x) - \frac{P_\omega(V)}{\mu(V)})^+ \le f_\omega(x)(1 - \frac{a_{j-1}}{a_j}) \le f_\omega(x)\alpha/3$. Hence

$$\alpha_{\omega\mu}(\mathcal{V}) = \sum_{V\in\mathcal{V}} \int V(f_\omega - \frac{P_\omega(V)}{\mu(V)})^+ d\mu = \sum_{V\subset V_0\cup V_k} \int V f_\omega d\mu + \sum_{j=1}^{k} \sum_{V\subset V_\delta} \int V f_\omega \alpha/3 \, d\mu < \alpha.$$

This proves the lemma.

Lemma 6. If Ω is totally bounded in the metric $d(\omega,\omega') = \sup_A |P_\omega(A) - P_{\omega'}(A)|$, then for any $\alpha > 0$ there is a partition, $\mathcal{V}$, of $\mathfrak{X}$ for which $\alpha(\mathcal{V}) \le \alpha$.

Proof. Since Ω is totally bounded in d, it is separable so that, for some σ-finite μ, $P_\omega << \mu$ for all $\omega \in \Omega$. Let $f_\omega = \frac{dP_\omega}{d\mu}$. Then $d(\omega,\omega') = \frac{1}{2}\int |f_\omega - f_{\omega'}| d\mu$.

By Lemma 5 we can find, for each ω, a partition $\mathcal{V}_\omega$ such that $\alpha_{\omega\mu}(\mathcal{V}) \le \alpha/2$ for any sub-partition of $\mathcal{V}_\omega$. Let $U_1, U_2, \ldots, U_k$ be a covering of Ω be spheres of diameter $\le \alpha/8$, and let $\omega_i \in U_i$, $i = 1,\ldots,k$, be arbitrary. Then since, for any ω,ω' and any $\mathcal{V}$,

$$\sum_{\mathcal{V}} \int V \left| \left(f_\omega - \frac{P_\omega(V)}{\mu(V)}\right)^+ - \left(f_{\omega'} - \frac{P_{\omega'}(V)}{\mu(V)}\right)^+ \right| d\mu$$

$$\le \sum_{\mathcal{V}} \left\{ \int V |f_\omega - f'_\omega| d\mu + \int V \left| \frac{P_\omega(V)}{\mu(V)} - \frac{P_{\omega'}(V)}{\mu(V)} \right| d\mu \right\}$$

and since, for the second term on the right,

$$\int V \left| \frac{P_\omega(V)}{\mu(V)} - \frac{P_{\omega'}(V)}{\mu(V)} \right| d\mu \le |P_\omega(V) - P_{\omega'}(V)| \le \int V |f_\omega - f_{\omega'}| d\mu,$$

we have $\alpha_{\omega\mu}(\mathcal{V}) - \alpha_{\omega'\mu}(\mathcal{V}) \le 2\int |f_\omega - f_{\omega'}| d\mu = 4d(\omega,\omega')$.

Let $\mathcal{V}$ be any finite sub-partition of $\mathcal{V}_{\omega_1}, \mathcal{V}_{\omega_2}, \ldots, \mathcal{V}_{\omega_k}$. Then for any ω, $\alpha_{\omega\mu}(\mathcal{V}) \le \alpha_{\omega_i\mu}(\mathcal{V}) + 4d(\omega,\omega_i) \le \alpha$ for $\omega \in U_i$. Hence $\alpha_\mu(\mathcal{V}) = \sup_\omega \alpha_{\omega\mu}(\mathcal{V}) \le \alpha$. Since $\alpha(\mathcal{V}) \le \alpha_\mu(\mathcal{V})$, the lemma is proved.

Remarks 1. If the conditions of Lemma 6 hold and there is a partition, $\mathcal{V}_1$ say, for which $\delta(\mathcal{V}_1) \le \delta$, then any sub-partition, $\mathcal{V}_2$, of both $\mathcal{V}_1$ and the $\mathcal{V}$ of Lemma 6 will have $\alpha(\mathcal{V}_2) \le \alpha$ by Lemma 5 and $\delta(\mathcal{V}_2) \le \delta$ trivially. We do not discuss $\delta(\mathcal{V})$ further except to note that, obviously, $\delta(\mathcal{V}) = 0$ for all $\mathcal{V}$ if L is independent of x.

2. The condition of Lemma 6, that Ω be totally bounded, holds for many families of distributions. Scheffé's theorem, that $\int|f_\omega - f_{\omega'}|d\mu \to 0$ if, for every F with $\mu(F) < \infty$ and every $\eta > 0$, $\mu[F \cap \{x: |f_\omega(x) - f_{\omega'}(x)| > \eta\}] \to 0$, serves to establish total boundedness - often by using compactness - in many cases. We give some examples.

(a) Exponential families. Let T be a mapping from $\mathfrak{X}$ into E^k and let μ be a measure on $\mathfrak{X}$. Let $\Theta = \{\omega \in E^k: \int e^{\omega T(x)} d\mu < \infty\}$ where $\omega\, T(x)$ is an inner product. The class of densities $\{C(\omega)e^{\omega T(x)}: \omega \in \Theta\}$, where $C(\omega) = [\int e^{\omega T(x)} d\mu]^{-1}$, is the exponential family on $\mathfrak{X}$ generated by T and μ. It is well known that C is continuous on the interior of Θ; so, since $f_\omega(x) \to f_{\omega'}(x)$ as $\omega \to \omega'$ for all x and all ω' in the interior of Θ, we see, using Scheffé's Theorem, that any subset of a compact subset of the interior of Θ will be totally bounded in the metric

$d(\omega,\omega') = \sup\limits_{\mathcal{B}} |P_\omega(A) - P_{\omega'}(A)| = \frac{1}{2}\int|f_\omega - f_{\omega'}|d\mu.$

For example, in the one dimensional normal family, $T(x) = (x^2,x)$ and $\omega = (-\frac{1}{2\sigma^2}, \frac{\mu}{\sigma^2})$, for the distribution with mean μ and variance σ^2. Hence our requirement is satisfied if, for some positive numbers, a and b, $\sigma^2 \geq a$ and $|\mu| \leq b\sigma^2$.

(b) Translation parameter families. Let $\int p\, d\nu = 1$ where, for μ Lebesgue measure on E^k, $\nu << \mu$ and $\frac{d\nu}{d\mu}$ is bounded. Let $f_\omega(x) = p(x-\omega)$, $\omega \in E^k$. Then since $\int|f_\omega - f_{\omega'}|d\nu \to 0$ as $\omega \to \omega'$ (cf. Royden (1968), p. 91, Problem 17 (b)) our condition is satisfied if Ω is any bounded subset of E^k.

Remark. In both (a) and (b) above, we also have, for each $B \in \mathcal{B}$, $|P_\omega(B) - P_{\omega'}(B)| \to 0$ uniformly in $\omega \in \Omega$ as $\omega' \to \omega$, a result which, in Theorem 5, was obtained from the compactness of Ω

and was used in showing that $\lambda(F,G) \to 0$ as $L(F,G) \to 0$ or $L^*(F,G) \to 0$. The other requirement, that $\sup_{a,V} |L(\omega,a,V) - L(\omega',a,V)| \to 0$ uniformly in ω, still needs separate treatment however.

BIBLIOGRAPHY

BIBLIOGRAPHY

Cogburn, Robert (1965). On the estimation of a multivariate location parameter with squared error loss. Bernoulli (1723), Bayes (1763), Laplace (1813), Anniversary Vol., 24-29.

_______ (1967). Stringent solutions to statistical decision problems. Ann. Math. Statist. **38**, 447-464.

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

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

Hannan, James F. and Huang, J.S. (1969). Equivariant procedures in the compound decision problem with finite-state component problem. To be submitted for publication.

Hannan, James F. and Robbins, Herbert (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. Ann. Math. Statist. **26**, 37-51.

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

James, W. and Stein, Charles (1961). Estimation with quadratic loss. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 361-379. University of California Press.

Rao, R. Ranga (1962). Relations between weak and uniform convergence of measures with applications. Ann. Math. Statist. **33**, 659-680.

Robbins, Herbert (1951). Asymptotically sub-minimax solutions of compound decision problems. Proc. Second Berkeley Symp. Math. Statist. Prob., 131-148. University of California Press.

Royden, H.L. (1968). Real Analysis, 2nd Edition. Macmillan.

Samuel, E. (1967). The compound statistical decision problem. Sankhya **29**, 123-140.

Sazonov, V.V. (1963). On the Glivenko-Cantelli Theorem. Theory of Probability and Its Applications VIII, 282-290.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 197-206. University of California Press.

Suzuki, Giitiro (1968). A general class of non-parametric empirical Bayes procedures. Research Memorandum No. 19, The Institute of Statistical Mathematics.

Varadarajan, V.S. (1958). Weak convergence of measures on separable metric spaces. Sankhya **19**, 15-22.

APPENDIX

APPENDIX

We prove here lemmas which have been used in Chapter II but which were unsuitable for inclusion there because, except for their application, they have no particular connection with the Compound Decision Problem. These lemmas are concerned with the relations between certain metrics on sets of probability measures; although they were used in the proof of Theorem 5, their use may not have been necessary and they are included at least partly because of their general interest.

Before introducing the lemmas, we need some definitions.

<u>Definition 1</u>. Let h be any function on a metric space (Ω,d). The modulus of continuity of h is the function given by

$$\alpha_h(\varepsilon) = \sup\{|h(\omega) - h(\omega')| : d(\omega,\omega') < \varepsilon\}$$

for each $\varepsilon > 0$.

<u>Definition 2</u>. Let $\mathcal{P}$ be the space of probability distributions on a metric space Ω. The Prohorov metric on $\mathcal{P}$ is given by

$$L^*(F,G) = \inf\{\delta : F(A^\delta) + \delta \geq G(A) \text{ for all closed } A \subset \Omega\}$$

where

$$A^\delta = \{\omega : \text{for some } \omega' \in A,\ d(\omega,\omega') < \delta\}.$$

[We note $L^*(F,G) = L^*(G,F)$; for A^δ is open for each A, so $A^{\delta c}$ is closed; and $\omega \in A^{\delta c \delta} \Rightarrow d(\omega,\omega') < \delta$ for some $\omega' \in A^{\delta c} \Rightarrow \omega \in A^c$, so $A^{\delta c\delta} \subset A^c$. Hence, if $\delta < L^*(G,F)$, so that $G(A^\delta) + \delta < F(A)$ for some A, then $F(A^{\delta c\delta}) + \delta \le F(A^c) + \delta < G(A^{\delta c})$, i.e. $\delta < L^*(F,G)$. Thus $L^*(G,F) \le L^*(F,G)$.]

<u>Lemma 7</u>. Let $a \le h(\cdot) \le a + M$ be a real-valued function on a metric space Ω. Then if P and Q are distributions on Ω with $L^*(P,Q) = \rho$,

$$\left|\int h\,dP - \int h\,dQ\right| \le M\rho + \alpha_h(\rho).$$

<u>Proof</u>. Without loss of generality we take $a = 0$. Then

$$\int h\,dP - \int h\,dQ = \int_0^M x\,dPh^{-1} - \int_0^M x\,dQh^{-1}$$

$$(1) \qquad = \int_0^M Ph^{-1}[x > t] - Qh^{-1}[x > t]dt$$

Since $\omega \in \left(\overline{h^{-1}[t,M]}\right)^\rho$ (where the bar denotes closure) implies $h(\omega) \ge t - \alpha_h(\rho)$, we have

$$Ph^{-1}[x \ge t] < Q\left(\overline{h^{-1}[t,M]}\right)^\rho + \rho \le Qh^{-1}[t - \alpha_h(\rho),M] + \rho.$$

Thus

$$(1) \le \int_0^M Qh^{-1}[t - \alpha_h(\rho) \le x \le t] + \rho\ dt$$

$$= \rho M + Qh^{-1}[\int_0^M [x, x+\alpha_h(\rho)](t)dt] \quad \text{by Fubini's Theorem}$$

$$\le \rho M + \alpha_h(\rho).$$

The same argument, with P and Q interchanged, proves the lemma.

<u>Definition 3</u>. Let $\mathcal{S}$ be the space of probability distribution functions on the real line. The Levy metric on $\mathcal{S}$ is given by

$$L(F,G) = \inf\{\delta\colon F(x-\epsilon) - \epsilon < G(x) < F(x+\epsilon) + \epsilon \text{ for all real } x\}.$$

Lemma 8. Let a and b be real, and $c \le h(\cdot) \le c + M$ be a real valued function on $\Omega \subseteq [a,b]$. Let F and G be any distributions on Ω, and $\lambda > 2\rho = 2L(F,G)$. Then

$$\left|\int h\, dF - \int h\, dG\right| < M\left[\frac{b-a}{\lambda} + 1\right]\rho + \alpha_h(\lambda + \rho) + \alpha_h(\lambda + 2\rho),$$

where $[x]$ denotes the largest integer not greater than x.

Proof. Without loss of generality, we take $c = 0$, $M < \infty$ and $b-a < \infty$.

Choose σ so that $2\rho < \sigma < \lambda$ and $[\frac{b-a}{\lambda} + 1]\sigma - (b-a) = \delta > 0$, and let $x_j = a + j\sigma - \delta/2$ for $j = 0,1,2,\ldots,k = [\frac{b-a}{\lambda} + 1]$.

By definition of $L(F,G)$, we can find $x_0 = x_0' < x_1' <\ldots< x_k' = x_k$ such that, for each j, $|x_j - x_j'| \le \rho$ and $F(x_j'-) - \rho < G(x_j) < F(x_j') + \rho$, because $F(x_j - \rho) - \rho < G(x_j) < F(x_j + \rho) + \rho$ implies the existence of an $x_j' \in [x_j - \rho, x_j + \rho]$ for which either $G(x_j) - \rho < F(x_j') < G(x_j) + \rho$ or $F(x_j'-) < G(x_j) - \rho < G(x_j) + \rho < F(x_j')$.

For each j, let $y_j = \min\{x_j, x_j'\}$ and $z_j = \max\{x_j, x_j'\}$, and for $x \in (x_j, x_{j+1}]$ let $h_1(x) = \inf\{h(\omega): \omega \in [y_j, z_{j+1}] \cap \Omega\}$, with $h_1(x) = M$ if $[y_j, z_{j+1}] \cap \Omega = \phi$.

For $x \in (x_j', x_{j+1}')$, let $h_2(x) = h_1(y)$ for $y \in (x_j, x_{j+1}]$, and for each j let $h_2(x_j') = \max\{h_2(x_j'-), h_2(x_j'+)\}$.

Since $|x_j - x_{j-1}'| < \lambda + \rho$, $|x_j - x_{j+1}'| < \lambda + \rho$ and $|x_j' - x_{j+1}'| < \lambda + 2\rho$ for each j, $|h-h_1| \le \alpha_h(\lambda + \rho)$ and $|h-h_2| \le \alpha_h(\lambda + 2\rho)$.

We note that $F(x_i',x_j') < G(x_i,x_j] + 2\rho$ by the construction of $\{x_j'\}_{j=0}^k$.

Let $0 \le r < M$. If $h_2(x_j') \le r$ then $(x_{j-1}',x_{j+1}') \subset h_2^{-1}(-\infty,r]$ and $(x_{j-1},x_{j+1}] \subset h_1^{-1}(-\infty,r]$. Conversely $(x_{j-1},x_{j+1}] \subset h_1^{-1}(-\infty,r]$ implies $(x_{j-1}',x_j') \cup (x_j',x_{j+1}') \subset h_2^{-1}(-\infty,r]$ and this implies

$h_2(x'_j) \le r$. Hence $h_2^{-1}(-\infty,r]$ is the union of at most $k/2$ intervals of the form (x'_i,x'_j) and $h_1^{-1}(-\infty,r]$ is the union of the corresponding intervals $(x_i,x_j]$.

Hence, for all r,

$$F\, h_2^{-1}(-\infty,r] < G\, h_1^{-1}(-\infty,r] + k\rho .$$

Thus
$$\begin{aligned}\int_\Omega h_2 dF - \int_\Omega h_1 dG &= \int_0^M x\, dFh_2^{-1} - \int_0^M x\, dGh_1^{-1} \\ &= \int_0^M Gh_1^{-1}(-\infty,x] - Fh_2^{-1}(-\infty,x]dx \\ &\ge \int_0^M Gh_1^{-1}(-\infty,x] - [Gh_1^{-1}(-\infty,x] + k\rho]dx \\ &= -Mk\rho .\end{aligned}$$

Hence $\int h\, dF - \int h\, dG \ge -Mk\rho - \alpha_h(\lambda+\rho) - \alpha_h(\lambda+2\rho)$. The same argument with F and G interchanged proves the lemma.

Lemma 8' involves a special case of Lemma 8 in which a simpler proof leads to a somewhat stronger conclusion. It seems, though we have been unable to show this, that the proof could be used in the context of Lemma 8, with some modifications, to give an improved result; and also that it might be amenable to versions of Lemma 8 or Lemma 8' in higher dimensions.

Lemma 8'. Let h be a function on $[a,b]$, F and G any distributions on $[a,b]$ and $\lambda > \rho = L(F,G)$. Then

$$\left|\int h\, dF - \int h\, dG\right| \le \alpha_h(\lambda)\{3 + [\tfrac{b-a}{\lambda}]\rho\}$$

Proof. Choose σ such that $\rho < \sigma < \lambda$ and $[\frac{b-a}{\lambda} + 1]\sigma > b-a$, and let $x_j = b - (k+1-j)\sigma$, for $j = 0,1,\ldots,k+1 = [\frac{b-a}{\lambda} + 1]$.

Let $c_j = h\,\frac{x_{j-1}+x_j}{2}$ for $j = 1,2,\ldots,k+1$. Then $|h(x) - c_j| \le \alpha_h(\lambda)$ for each $x \in (x_{j-1},x_j]$, and $|c_j - c_{j-1}| \le \alpha_h(\lambda)$

for each j.

Let $D_j = F(x_j) - G(x_j)$ for each j.

Then

$$\int h \, d(F-G) \le \alpha_h(\lambda) + \sum_{j=1}^{k+1} c_j(D_j - D_{j-1})$$

$$= \alpha_h(\lambda) - c_1 D_0 + c_{k+1} D_{k+1} + \sum_{j=1}^{k} D_j(c_j - c_{j+1})$$

$$\le \alpha_h(\lambda) + \alpha_h(\lambda) \sum_{j=1}^{k} |D_j|.$$

But $|D_j| = [F(x_j)-G(x_j)] \vee [G(x_j)-F(x_j)] < [G(x_{j+1})+\rho-G(x_j)] \vee [G(x_j)-G(x_{j-1})+\rho]$

Hence $\sum_{j=1}^{k} |D_j| < \sum_{j=1} \{G(x_{j+1}) - G(x_j) + \rho + G(x_j) - G(x_{j-1})\}$

$$= k\rho + G(x_{k+1}) + G(x_k) - G(x_1) - G(x_0)$$

$$\le k\rho + 2.$$

Hence $\int h \, d(F-G) < \alpha_h(\lambda)(k\rho + 2)$

$$= [\tfrac{b-a}{\lambda}]\rho \, \alpha_h(\lambda) + 3\alpha_h(\lambda).$$

Again, reversing the roles of F and G proves the lemma.