$O(N^{-\frac{1}{2}})$ CONVERGENCE IN THE FINITE STATE RESTRICTED RISK COMPONENT SEQUENCE COMPOUND DECISION PROBLEM

Dissertation for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

STEPHEN BRUCE VARDEMAN

1975

This is to certify that the

thesis entitled

$O(N^{\frac{1}{2}})$ CONVERGENCE IN THE FINITE STATE RESTRICTED RISK COMPONENT SEQUENCE COMPOUND DECISION PROBLEM

presented by

Stephen Bruce Vardeman

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics and Probability

James Hannan
Major professor

Date 8/18/75

O-7639

ABSTRACT

$O(N^{\frac{1}{2}})$ CONVERGENCE IN THE FINITE STATE RESTRICTED RISK COMPONENT SEQUENCE COMPOUND DECISION PROBLEM

By

Stephen Bruce Vardeman

We consider a sequence of independent, structurally identical, finite state restricted risk component decision problems, where the choice of risk point in the αth problem is allowed to depend on observations from the previous problems, and the goal is to control the total risk incurred across the first N problems. k-extended standards for the risk of sequence compound procedures are introduced. In the most general situation in terms of allowable form of the component problem risk set and distributions of the observations, bounds are obtained for the risks of a family of procedures employing artificial randomization. Appropriate specification of a sequence of constants appearing in both the bounds and description of the procedures give total risk approximating the k-extended standard at a $O(N^{\frac{1}{2}})$ rate. It is noted that the formulation of the problem given includes a game theoretic situation in which the information about past states carried by the observations is perfect. Four procedures appropriate to such a situation are offered, each of which has risk approximating the k-extended standard at a $O(N^{\frac{1}{2}})$ rate. Finally, nondegeneracy conditions are imposed on the distributions of the observations and the

resulting statistical version of the problem is studied. A rate of weak convergence theorem of Bhattacharya ((1970). Rates of weak convergence for the multidimensional central limit theorem. *Theory of Probability and its Applications* **15**, 68-86.) and Mirahmedov ((1974). The rate of weak convergence in the multidimensional limit theorem. *Izv. Akad. Nauk UzSSR Ser. Fiz.-Mat. Nauk* **18** no. 2, 23-28, 92-93.) is applied to show that in a two state case, a natural procedure has risk approximating the usual unextended standard at a $O(N^{\frac{1}{2}})$ rate.

$O(N^{\frac{1}{2}})$ CONVERGENCE IN THE FINITE STATE RESTRICTED RISK COMPONENT SEQUENCE COMPOUND DECISION PROBLEM

By

Stephen Bruce Vardeman

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1975

ACKNOWLEDGEMENTS

I am truly grateful to Professor James Hannan for his guidance during my two years at Michigan State University. I feel fortunate to have been able to draw on his expert knowledge of mathematical statistics both in course work and in the preparation of this dissertation. Without his patience, encouragement and many suggestions this dissertation would never have come into existence.

I would also like to thank the other members of my guidance committee, Professors Dennis Gilliland, Hira Koul, Raoul LePage and Joel Shapiro. Conversations with Professors Gilliland and LePage concerning sections one through three were especially helpful.

Financial support provided by the Department of Statistics and Probability and the National Science Foundation made my graduate studies in Statistics possible. I wish to thank both of them.

I am grateful to Mrs. Noralee Burkhardt, who typed this dissertation. Her excellent and fast typing of the manuscript was done cheerfully and quite unselfishly.

Finally, I wish to thank my wife, Jo Ellen. The supportive and encouraging spirit that she has maintained over the past four years has been a major ingredient of my success as a graduate student.

TABLE OF CONTENTS

0. INTRODUCTION

Simultaneous consideration of a number of independent structurally identical decision problems with the goal of controlling the average or total risk incurred across the problems was first suggested by Robbins (1951). Robbins termed his original example involving N independent discriminations between normal $(1,1)$ and normal $(-1,1)$ distributions a compound decision problem. The procedure he proposed has total risk approximating N times the Bayes risk versus the normalized empirical distribution of states in the component testing problem, and finding procedures with similar total risk performance became the usual objective in compound decision theory.

Hannan (1956), (1957) in rather general finite state settings showed that the usual compound decision theoretic goal is achievable not only in situations in which all N of the problems are considered simultaneously, but also in situations where the independent structurally identical problems are faced serially. Such a modification of the compound setting has become known as a sequence compound decision problem. Hannan's procedures involve artificial randomization and are applicable in situations in which before making the αth decision one has available either the exact empiric distribution of states through the $(\alpha-1)$st problem or an estimate of the same. Van Ryzin (1966b) showed that in many finite state finite act statistical versions of the sequence compound problem, the extra randomization employed by Hannan is not necessary. In a sense, enough randomness enters through the estimation of empiric distributions of states. Van Ryzin's arguments are closely tied to his assumption of a finite action space.

Johns (1967) and Gilliland and Hannan (1969a) suggested standards of performance for compound procedures which are appropriate to sequence versions of compound problems and asymptotically more stringent than the usual standards. These standards, which take into account kth order empirical dependencies in the sequence of states have become known as k-extended standards. Work of Ballard (1974) and Ballard, Gilliland and Hannan (1974) shows that in Van Ryzin's finite state finite act statistical setting, generalizations of his non-randomized procedures have risk approximating these k-extended objectives.

In this thesis we treat a particularly tractable, yet quite general finite state sequence compound decision problem. The generality of the problem derives primarily from the fact that a risk structure rather than action space and loss structure is assumed for the component problem. In §1 this finite state restricted risk component sequence compound decision problem is described along with the unextended and k-extended standards for the problem. The estimation of kth order empirical distributions of states is also very briefly considered. Section 2 contains the description of procedures which are generalizations of the procedures involving randomization proposed originally by Hannan. We bound the total risk of the procedures and note that appropriate choice of arbitrary constants yields bounds approximating the k-extended standards at a $O(N^{\frac{1}{2}})$ rate. In §3 it is noted that the problem includes a game theoretic situation in which the component risk set is composed of the risk points available to player II, and before each repetition of the game, II is furnished with the empirical distribution of I's moves through the previous play. After noting a simple game theoretic decomposition of the k-extended standard, three procedures are provided in addition to the game theoretic specialization of the procedure from §2, which have risk approximating the k-extended standard at a $O(N^{\frac{1}{2}})$ rate. The basic technique employed in §3.3 and §3.4 has been used independently by Cover and Shenar (1974) in a less general situation. The results of §4 all concern the unextended version of the problem. Using a rate of weak convergence result of Bhattacharya (1970) - Mirahmedov (1974), we show that certain natural procedures in a two state case have risks approximating the usual standard at a $O(N^{\frac{1}{2}})$ rate. The procedures are related to those of Van Ryzin in that no extra randomization is employed. The appendix contains several results applied in the body of the thesis.

Several notational conventions should be mentioned. Vectors in R^m are considered to be column vectors, although to save space they occasionally appear as row vectors in the text. Euclidean vector norms are denoted as $\|\cdot\|$. $\vee$ and $\wedge$ stand for supremum and infimum respectively. Φ denotes the univariate standard normal distribution and Φ_m is the m-variate standard normal distribution function. $\mathcal{B}^m$ is the Borel sigma algebra on R^m. The term "range" of a matrix refers to its column space. lhs and rhs are used to abbreviate left hand side and right hand side respectively. For probabilities ν_1 and ν_2, $|\nu_1 - \nu_2|$ will denote the total variation of the signed measure $\nu_1 - \nu_2$. Displays are numbered consecutively in §1 through §4 with the displays in the appendix numbered separately.

1. THE k-EXTENDED FINITE STATE RESTRICTED RISK COMPONENT SEQUENCE COMPOUND DECISION PROBLEM

1.1. The component problem.

We consider a (component) decision problem with states $\theta \in \Theta = \{1,2,\ldots,m\}$ and risk set $S \subset [0,\infty)^m$. For each $\theta \in \Theta$ let P_θ be a probability on a measurable space $(\mathfrak{X},\mathcal{F})$. In all that follows we shall assume that S is bounded, $\|s\|_\infty \le B < \infty$ for each $s \in S$, where $\|\cdot\|_\infty$ denotes the supremum norm for m-vectors. Gilliland and Hannan (1974) use the term "restricted risk" to indicate that S may be a proper subset of the largest possible risk set for a given action space and loss function.

For w a finite signed measure on Θ and $s \in S$ we will let ws denote $\int s(\cdot)dw(\cdot)$. In the case where w is a probability ws is the Bayes risk of s against the prior distribution w. It will be convenient to identify each $\theta \in \Theta$ with the probability on Θ concentrated at θ. θs is then the θ coordinate of s. Further notational economy will be possible by agreeing to also identify finite signed measures on Θ with m-vectors, w corresponding to $(w(\{1\}), w(\{2\}),\ldots,w(\{m\})) \in R^m$.

1.2. The sequence compound problem.

A compound decision problem as introduced by Robbins (1951) involves N independent repetitions of the component problem described in section 1. We will consider a sequence version in which the choice of risk function in component α is allowed to depend upon independent, P_{θ_β} distributed observations for $\beta \le \alpha-1$, and the compound risk is taken as the sum of risks in components 1 through N.

More precisely, let k be a positive integer and $\underline{s} = (s_1, s_2, \ldots, s_N)$ be such that s_α is a $\mathcal{F}^{\alpha+k-2}$ measurable mapping into S. For $\underset{\sim}{\theta}_N = (\theta_{2-k}, \ldots, \theta_N)$ we suppose that $\underset{\sim}{X}_N = (X_{2-k}, \ldots, X_N)$ is distributed as $P_{\theta_{2-k}} \times \ldots \times P_{\theta_N}$. (The purpose of allowing indices $\alpha < 1$ in the case $k > 1$ here is to simplify later notation.) The compound risk of the sequence rule $\underline{s}$ is

$$(1) \qquad \sum_{\alpha=1}^{N} E\, \theta_\alpha s_\alpha(\underline{X}_{\alpha-1}) = \sum_{1}^{N} \int \theta_\alpha s(\cdot)\, dP_{\theta_{2-k}} \times \ldots \times P_{\theta_{\alpha-1}} .$$

When $\underline{s} = (s, s, \ldots, s)$ for $s \in S$ the risk (1) becomes

$$\sum_{1}^{N} \theta_\alpha s = \left(\sum_{1}^{N} \theta_\alpha\right) s = G_N^1 s ,$$

where as indicated earlier, we are identifying elements of Θ with probabilities on Θ and G_N^1 is the non-normalized empiric distribution of $\{\theta_1, \ldots, \theta_N\}$. With $\Psi^1(G_N^1) = \bigwedge_{s \in S} G_N^1 s$, Hannan (1957), (1956) first exhibited procedures in versions of the sequence-compound problem achieving $\Psi^1(G_N^1)$ asymptotically.

Let S^* be a class of $\mathcal{F}^{k-1}$ measurable mappings into S. We consider the sequence compound rules of the form $\underline{s} = (s^*, \ldots, s^*)$ for $s^* \in S^*$. For such $\underline{s}$ the risk (1) reduces to the functional

$$(2) \qquad \sum_{1}^{N} E\, \theta_\alpha s^*(X_{\alpha-k+1}, \ldots, X_{\alpha-1}) = \sum_{\alpha=1}^{N} \int \theta_\alpha s^*(\cdot)\, dP_{\theta_{\alpha-k+1}} \times \ldots \times P_{\theta_{\alpha-1}}$$

of G_N^k, the non-normalized empirical distribution on Θ^k of the vectors $\{(\theta_{2-k},\ldots,\theta_1),(\theta_{3-k},\ldots,\theta_2),\ldots,(\theta_{N-k+1},\ldots,\theta_N)\}$. Swain (1965), Johns (1967), and Gilliland and Hannan (1969) have termed

$$(3) \qquad \Psi^k(G_N^k) = \bigwedge_{s^*\in S^*} \sum_1^N E\, \theta_\alpha s^*(X_{\alpha-k+1},\ldots,X_{\alpha-1})$$

a k-extended simple envelope evaluated at G_N^k. Notice that in the case S^* is the class of all $\mathcal{F}^{k-1}$ measurable mappings into S, for fixed $\underline{\theta}_N$, $\Psi^k(G_N^k) \le \Psi^{k-1}(G_N^{k-1})$. That is, the extended envelopes are increasingly stringent.

The purpose of this work is to exhibit sequence compound rules which achieve risk $\Psi^k(G_N^k)$ asymptotically with rate.

1.3. Bayes rules in the component problem.

We will make the assumption that S is not only bounded, but also closed. For any $w \in R^m$, $\bigwedge_{s\in S} ws$ is then attained and we denote an infimizing s by $\sigma(w)$. It is a simple consequence of Corollary 1 of Brown and Purves (1973) that there is a Borel measurable determination of $\sigma(\cdot)$. In addition we may assume that $\sigma(\cdot)$ has the property that $\sigma(pw) = \sigma(w)$ for $p > 0$. (If not, we replace $\sigma(w)$ by $\sigma(w/\|w\|)$ for $w \neq 0$, where $\|\cdot\|$ is the usual Euclidean vector norm.) Notice that with this notation we have

$$\Psi^1(G_N^1) = \bigwedge_{s\in S} G_N^1 s = G_N^1\, \sigma(G_N^1) \ .$$

There is no essential loss of generality in the assumption that S is closed. If $\bar{S}$ denotes the closure of S in R^m, for any $\varepsilon > 0$ and any sequence compound rule $\underline{s}' = (s_1',s_2',\ldots,s_N')$ where s_α' is a $\mathcal{F}^{\alpha+k-2}$ measurable mapping into $\bar{S}$, there exists a rule $\underline{s} = (s_1, \ldots, s_N)$ such that s_α is a $\mathcal{F}^{\alpha+k-2}$ measurable mapping into S with $|\theta_\beta s_\alpha(\cdot) - \theta_\beta s'_\alpha(\cdot)| \leq 2^{-\alpha}\epsilon$ each $\theta_\beta \in \Theta$. Hence $|\sum_1^N E\, \theta_\alpha s_\alpha(\underline{X}_{\alpha-1}) - \sum_1^N E\, \theta_\alpha s'_\alpha(\underline{X}_{\alpha-1})| < \epsilon$ for all $\underline{\theta}_N$, and theorems concerning the risks of $\bar{S}$ valued rules have ϵ analogues for S valued rules.

1.4. <u>The Γ^k construct</u>.

In order to describe compound rules achieving risk $\Psi^k(G_N^k)$ asymptotically, we introduce a variant of Gilliland and Hannan's Γ^k decision problem. The Γ^k problem has finite state space Θ^k and risk set $\tilde{S} \subset [0,+\infty)^{m^k}$ where

$$\tilde{S} = \{\tilde{s} \in R^{m^k} \mid (\theta_1, \ldots, \theta_k)\tilde{s} = \int \theta_k s^*(\cdot) dP_{\theta_1} \times P_{\theta_2} \times \ldots \times P_{\theta_{k-1}} \text{ for some } s^* \in S^*\} .$$

The Γ^k problem inherits the property of bounded risk from the component problem. We shall use notational conventions for the Γ^k construct similar to those introduced for the component problem. That is, we will identify each $\underline{\theta} \in \Theta^k$ with the probability on Θ^k degenerate at $\underline{\theta}$, let m^k-vectors correspond to signed measures on Θ^k, and for v a signed measure on Θ^k, $\tilde{s} \in \tilde{S}$ denote $v\tilde{s} = \int \tilde{s}(\cdot) dv(\cdot)$. $\tilde{\sigma}(v)$ will denote a Borel measurable, positive homogeneous minimizer of $v\tilde{s}$. (That no essential generality will be lost by the assumption that $\tilde{S}$ is closed follows from a comment similar to that made in the previous paragraph.)

Using the Γ^k notation we have from (2) and (3)

$$\Psi^k(G_N^k) = \bigwedge_{s^* \in S^*} \sum_1^N \int \theta_\alpha s^*(\cdot)\, dP_{\theta_{\alpha-k+1}} \times \ldots \times P_{\theta_{\alpha-1}}$$

$$= \bigwedge_{\tilde{s} \in \tilde{S}} \sum_1^N (\theta_{\alpha-k+1}, \ldots, \theta_\alpha)\tilde{s}$$

$$= \bigwedge_{\tilde{s} \in \tilde{S}} G_N^k \tilde{s}$$

$$= G_N^k \tilde{\sigma}(G_N^k) \ .$$

We extend the domain of Ψ^k to all of R^{m^k} by defining

$$\Psi^k(v) = v\tilde{\sigma}(v) \quad \text{all} \quad v \in R^{m^k} \ . \tag{4}$$

It will be important to recover elements of S^* which give rise to values of the minimizer $\tilde{\sigma}(\cdot)$. Thus assume that $s^*(\cdot,\cdot)$ is a mapping from $R^{m^k} \times \mathfrak{X}^{k-1}$ into S with the property that for $v \in R^{m^k}$, $s^*(v,\cdot) \in S^*$ such that

$$\int \theta_k s^*(v,\cdot)\, dP_{\theta_1} \times \ldots \times P_{\theta_{k-1}} = (\theta_1, \ldots, \theta_k)\tilde{\sigma}(v) \quad \text{each} \quad \underline{\theta} \in \Theta^k . \tag{5}$$

Notice that in the case $k = 1$, the Γ^k construct is identical with the original component problem and we take $s^*(v) = \tilde{\sigma}(v) = \sigma(v)$.

1.5. <u>Assumption on the</u> P_θ <u>and estimation of empirics</u>.

We will assume that $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$ is a linearly independent family of measures. That is, for real numbers $a_1, \ldots, a_m$, $\sum_{\theta \in \Theta} a_\theta P_\theta$ is the zero signed measure only if each $a_\theta = 0$. Robbins (1964), Van Ryzin (1966a), and Ballard (1974) discuss the estimation of mixtures of a finite number of linearly independent distributions.

In the linearly independent situation there are R^m valued, bounded, $\mathcal{F}$ measurable mappings t with the property that $\int t(\cdot)\, dP_\theta$

is the m-vector with all zero entries except a 1 in the θ position. $t(X_\alpha)$ is then an unbiased estimate of the m-vector corresponding to θ_α. Ballard (1974) uses vectors of all possible products of coordinates of k such mappings t to construct R^{m^k} valued, bounded $\mathcal{F}^k$ measurable mappings $\tilde{t}$ with the property that $\int \tilde{t}(\cdot)dP_{\theta_1} \times \ldots \times P_{\theta_k}$ is the m^k-vector with all zero entries except a 1 in the $(\theta_1, \theta_2, \ldots, \theta_k)$ position. $\tilde{t}(X_{\alpha-k+1}, \ldots, X_\alpha)$ is then an unbiased estimate of the m^k-vector corresponding to $(\theta_{\alpha-k+1}, \ldots, \theta_\alpha)$, and $\sum_{j=1}^{\alpha} \tilde{t}(X_{j-k+1}, \ldots, X_j)$ is an unbiased estimate of the m^k-vector corresponding to G_α^k.

We will not assume a special product structure for our estimates but will assume only that $\tilde{t}$ is an $\mathcal{F}^k$ measurable mapping into R^{m^k} with the properties

$$\int \tilde{t}(\cdot)dP_{\theta_1} \times \ldots \times P_{\theta_k} = (\theta_1, \ldots, \theta_k) \in R^{m^k}, \tag{6}$$

and

$$\bigvee_{\underline{\theta} \in \Theta^k} \int (\|\tilde{t}(\cdot)\|_1)^2 dP_{\theta_1} \times P_{\theta_2} \times \ldots \times P_{\theta_k} \equiv \tau_k^2 < \infty, \tag{7}$$

where $\|\cdot\|_1$ is the usual ℓ_1 vector norm. (Ballard's product kernels provide examples of functions satisfying (6) and (7).) Let $\tilde{t}_\alpha$ denote $\tilde{t}(X_{\alpha-k+1}, \ldots, X_\alpha)$ and $\tilde{T}_\alpha$ denote $\sum_{j=1}^{\alpha} \tilde{t}_j$ for $\alpha \geq 1$, 0 otherwise.

2. A BOUND ON THE RISK OF A k-EXTENDED SEQUENCE COMPOUND PROCEDURE EMPLOYING ARTIFICIAL RANDOMIZATION

We introduce a generalization of a strategy in the sequence compound problem proposed by Hannan (1957), (1956) and bound its risk. When constants appearing in both the description of the procedure and the bound are appropriately chosen, the strategy is seen to achieve risk $\Psi^k(G_N^k) + O(N^{\frac{1}{2}})$ uniform in $\underline{\theta}_N$.

2.1. Two lemmas.

Forms of the two lemmas which follow appeared first in Hannan (1957) and variations of one or the other have since appeared in Samuel (1963), (1965), Swain (1965), Van Ryzin (1966), Gilliland (1969) and Gilliland and Hannan (1969a).

Lemma 1. Let $f_1, f_2, \ldots, f_N$ be real-valued functions on some set $\mathcal{D}$. Denote $F_\alpha = \sum_1^\alpha f_j$ and suppose that for each $1 \le \alpha \le N$, $\exists\, d_\alpha \in \mathcal{D}$ such that $F_\alpha(d_\alpha) = \bigwedge_{d \in \mathcal{D}} F_\alpha(d)$. Let d_0 be arbitrary. Then $\sum_{\alpha=1}^N f_\alpha(d_\alpha) \le F_N(d_N) \le \sum_{\alpha=1}^N f_\alpha(d_{\alpha-1})$.

Proof. $\sum_1^N f_\alpha(d_\alpha) = F_N(d_N) - \sum_1^{N-1} (F_\alpha(d_{\alpha+1}) - F_\alpha(d_\alpha))$. But for each α, $F_\alpha(d_{\alpha+1}) - F_\alpha(d_\alpha) \ge 0$. Also $\sum_1^N f_\alpha(d_{\alpha-1}) =$ $F_N(d_N) + \sum_1^N (F_\alpha(d_{\alpha-1}) - F_\alpha(d_\alpha))$. And for each α, $F_\alpha(d_{\alpha-1}) - F_\alpha(d_\alpha) \ge 0$. ∎

The application we will make of this lemma is to take $\mathcal{S} = \tilde{S}$, $f_\alpha(\tilde{s}) = v_\alpha \tilde{s}$ for $v_\alpha \in R^{m^k}$, let $V_\alpha = \sum_{\beta=1}^{\alpha} v_\beta$ and conclude

$$(8) \qquad \sum_{\alpha=1}^{N} v_\alpha \tilde{\sigma}(V_\alpha) \leq \Psi^k(V_N) \leq \sum_{\alpha=1}^{N} v_\alpha \tilde{\sigma}(V_{\alpha-1}) .$$

We will not prove the next lemma. Apart from slight notational differences, the proof of a similar lemma in section 3 of Gilliland (1969) applies in our case also. Gilliland's assumptions that v, v', z are R^∞ vectors may be altered to v, v', z being R^{m^k} vectors without change in the form of the proof, his B may be re-interpreted as our supremum norm on S, and his assumption that the co-ordinates of v, v' are non-negative may be dropped.

<u>Lemma 2</u>. Let Z be uniformly distributed on $[0,1]^{m^k}$ and let μ be the distribution of Z. For any v, v' belonging to R^{m^k}, any $\underline{\theta} \in \Theta^k$

$$|\mu \, \underline{\theta}(\tilde{\sigma}(v + z) - \tilde{\sigma}(v' + z))| \leq B\|v - v'\|_1,$$

where operator notation is used to indicate integration.

The lemma then gives

$$(9) \qquad \|\mu(\tilde{\sigma}(v + z) - \tilde{\sigma}(v' + z))\|_\infty \leq B\|v - v'\|_1 .$$

2.2. <u>Definition of the procedures $\hat{\underline{s}}$</u>.

Take $\{H_\alpha\}_{\alpha=1}^{\infty}$ to be a non-decreasing sequence of positive constants. Define $H_\alpha = 0$ for $\alpha \leq 0$ and denote $h_\alpha = H_\alpha - H_{\alpha-1}$. Let Z be a uniform $[0,1]^{m^k}$ random vector independent of X_α for each $2-k \leq \alpha \leq N$. We will consider the procedure $\hat{\underline{s}} = (\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_N)$ where

(10) $$\hat{s}_\alpha = s^*(\tilde{T}_{\alpha-k} + H_{\alpha-k}Z,\ (X_{\alpha-k+1},\ldots,X_{\alpha-1}))\ .$$

(In the α component, the proposed procedure uses an element of S^* corresponding to an element of $\tilde{S}$ which is Γ^k Bayes against a randomly perturbed estimate of $G^k_{\alpha-k}$.)

2.3. A bound for the risk of $\hat{s}$.

Theorem 1.

$$\sum_{\alpha=1}^{N} E\ \theta_\alpha \hat{s}_\alpha \le \Psi^k(G^k_N) + \frac{1}{2} B\ H_N m^k + B\ k\ \tau_k(1 + 2\tau_k \sum_{\alpha=1}^{N} \frac{1}{H_\alpha})\ ,$$

uniform in $\underline{\theta}_N$, where for each α, $E\ \theta_\alpha \hat{s}_\alpha$ is interpreted as an iterated integral, the first integration with respect to the distribution of $(X_{\alpha-k+1},\ldots,X_{\alpha-1})$ on $\mathcal{X}^{k-1}$, and the second with respect to the distribution of $(X_{2-k},\ldots,X_{\alpha-k},Z)$ on $\mathcal{X}^{\alpha-1} \times [0,1]^{m^k}$.

Proof. Use operator notation to indicate integration and the following notations for distributions. Let $\underline{P}$ denote the joint distribution of $\underline{X}_N$, $\underline{P}_\alpha = P_{\theta_{2-k}} \times\ldots\times P_{\theta_{\alpha-k}}$ the joint distribution of $\underline{X}_{\alpha-k}$, $\underline{Q}_\alpha = P_{\theta_{\alpha-k+1}} \times\ldots\times P_{\theta_{\alpha-1}}$ the joint distribution of $(X_{\alpha-k+1},\ldots,X_{\alpha-1})$, and μ the distribution of Z.

$$\sum_1^N E\ \theta_\alpha \hat{s}_\alpha = \sum_1^N \mu\ \underline{P}_\alpha \underline{Q}_\alpha \theta_\alpha s^*(\tilde{T}_{\alpha-k} + H_{\alpha-k}z,\ (X_{\alpha-k+1},\ldots,X_{\alpha-1}))$$

$$= \sum_1^N \mu\ \underline{P}_\alpha(\theta_{\alpha-k+1},\ldots,\theta_\alpha)\tilde{\sigma}(\tilde{T}_{\alpha-k} + H_{\alpha-k}z),$$

from (5). So

(11) $$\sum_1^N E\ \theta_\alpha \hat{s}_\alpha = \sum_1^N \mu\ \underline{P}(\theta_{\alpha-k+1},\ldots,\theta_\alpha)\tilde{\sigma}(\tilde{T}_{\alpha-k} + H_{\alpha-k}z)\ .$$

Recalling that $\tilde{t}_\alpha$ is unbiased for $(\theta_{\alpha-k+1},\ldots,\theta_\alpha)$ and independent of $\tilde{T}_{\alpha-k} + H_{\alpha-k}Z$, (11) gives

$$\sum_1^N E\,\theta_\alpha \hat{\theta}_\alpha$$

$$(12)\qquad = \sum_1^N \mu\,\underline{P}\,\tilde{\varepsilon}_\alpha \tilde{\sigma}(\tilde{T}_{\alpha-k} + H_{\alpha-k}z) = \sum_1^N \underline{P}\,\tilde{\varepsilon}_\alpha \mu(\tilde{\sigma}(\tilde{T}_{\alpha-k} + H_{\alpha-k}z) - \tilde{\sigma}(\tilde{T}_\alpha + H_\alpha z))$$

$$+ \sum_1^N \underline{P}\,\mu\tilde{\varepsilon}_\alpha \tilde{\sigma}(\tilde{T}_\alpha + H_\alpha z)\ .$$

Denote the first sum on the right of (12) by A and the second by C. We will set $U_\alpha = \tilde{T}_\alpha + H_\alpha z$, $\tilde{\sigma}(U_\alpha) = \tilde{\sigma}_\alpha$ and bound A and C separately.

First consider A. For $\alpha > k$

$$(13)\qquad |\tilde{\varepsilon}_\alpha \mu(\tilde{\sigma}_{\alpha-k} - \tilde{\sigma}_\alpha)| \le \|\tilde{\varepsilon}_\alpha\|_1 \|\mu(\tilde{\sigma}_{\alpha-k} - \tilde{\sigma}_\alpha)\|_\infty\ .$$

We write $\tilde{\sigma}_{\alpha-k} - \tilde{\sigma}_\alpha = \tilde{\sigma}\left(\frac{\tilde{T}_{\alpha-k}}{H_{\alpha-k}} + z\right) - \tilde{\sigma}\left(\frac{\tilde{T}_\alpha}{H_\alpha} + z\right)$ and (13) and (9) give

$$|\tilde{\varepsilon}_\alpha \mu(\tilde{\sigma}_{\alpha-k} - \tilde{\sigma}_\alpha)|$$

$$\le \|\tilde{\varepsilon}_\alpha\|_1 B \left\|\frac{\tilde{T}_\alpha}{H_\alpha} - \frac{\tilde{T}_{\alpha-k}}{H_{\alpha-k}}\right\|_1 = B\|\tilde{\varepsilon}_\alpha\|_1 \left\|\frac{\sum_{j=1}^{k} \tilde{\varepsilon}_{\alpha-k+j}}{H_\alpha} - \tilde{T}_{\alpha-k}\left(\frac{1}{H_{\alpha-k}} - \frac{1}{H_\alpha}\right)\right\|_1$$

$$\le B\|\tilde{\varepsilon}_\alpha\|_1 \left(\frac{1}{H_\alpha}\sum_{j=1}^{k} \|t_{\alpha-k+j}\|_1 + \left(\frac{1}{H_{\alpha-k}} - \frac{1}{H_\alpha}\right)\sum_{j=1}^{\alpha-k} \|\tilde{\varepsilon}_j\|_1\right)\ .$$

Hence

$$(14)\qquad \sum_{\alpha=k+1}^{N} \underline{P}|\tilde{\varepsilon}_\alpha \mu(\tilde{\sigma}_{\alpha-k} - \tilde{\sigma}_\alpha)|$$

$$\le B \sum_{\alpha=k+1}^{N} \underline{P}\left(\frac{1}{H_\alpha}\sum_{j=1}^{k} \|\tilde{\varepsilon}_\alpha\|_1 \|\tilde{\varepsilon}_{\alpha-k+j}\|_1 + \left(\frac{1}{H_{\alpha-k}} - \frac{1}{H_\alpha}\right)\sum_{j=1}^{\alpha-k} \|\tilde{\varepsilon}_\alpha\|_1 \|\tilde{\varepsilon}_j\|_1\right)\ .$$

The Schwarz inequality and (7) applied to (14) give

$$\sum_{\alpha=k+1}^{N} \underline{P}|\tilde{\epsilon}_\alpha \mu(\tilde{\sigma}_{\alpha-k} - \tilde{\sigma}_\alpha)| \leq B \sum_{\alpha=k+1}^{N} \left(\frac{k}{H_\alpha}\tau_k^2 + \left(\frac{1}{H_{\alpha-k}} - \frac{1}{H_\alpha}\right)(\alpha - k)\tau_k^2\right)$$

$$= B\tau_k^2 \left(2k \sum_{k+1}^{N} \frac{1}{H_\alpha} + \sum_{\alpha=1}^{k} \frac{\alpha}{H_\alpha} - \sum_{N-k}^{N} \frac{\alpha}{H_\alpha} \right) .$$

But

$$\sum_{\alpha=1}^{k} \underline{P}\, \mu|\tilde{\epsilon}_\alpha(\tilde{\sigma}_{\alpha-k} - \tilde{\sigma}_\alpha)| \leq \sum_{\alpha=1}^{k} \underline{P}\, \mu \|\tilde{\epsilon}_\alpha\|_1 \|\tilde{\sigma}_{\alpha-k} - \tilde{\sigma}_\alpha\|_\infty \leq B\, k\, \tau_k$$

by the moment inequality. So we have

$$A \leq Bk\tau_k + B\tau_k^2 \left(2k \sum_{k+1}^{N} \frac{1}{H_\alpha} + \sum_{\alpha=1}^{k} \frac{\alpha}{H_\alpha} - \sum_{N-k}^{N} \frac{\alpha}{H_\alpha} \right) \leq Bk\tau_k + 2Bk\tau_k^2 \sum_{1}^{N} \frac{1}{H_\alpha} .$$

Now bound C.

$$C = \underline{P}\, \mu \sum_{1}^{N} \tilde{\epsilon}_\alpha \tilde{\sigma}_\alpha \leq \underline{P}\, \mu \sum_{1}^{N} (\tilde{\epsilon}_\alpha + h_\alpha z)\tilde{\sigma}_\alpha \leq \underline{P}\, \mu\, \Psi^k(\tilde{T}_N + H_N z)$$

by (8). But by (4)

$$\begin{aligned}
\underline{P}\, \mu\, \Psi^k(\tilde{T}_N + H_N z) &= \underline{P}\, \mu(\tilde{T}_N + H_N z)\tilde{\sigma}(T_N + H_N z) \\
&\leq \underline{P}\, \mu(\tilde{T}_N + H_N z)\tilde{\sigma}(G_N^k) \\
&= \underline{P}(\tilde{T}_N + \tfrac{1}{2}H_N \underline{1})\tilde{\sigma}(G_N^k) \\
&= G_N^k \tilde{\sigma}(G_N^k) + \tfrac{1}{2}H_N \|\tilde{\sigma}(G_N^k)\|_1 \\
&\leq \Psi^k(G_N^k) + \tfrac{1}{2}H_N m^k B .
\end{aligned}$$

That is, $C \leq \Psi^k(G_N^k) + \frac{1}{2}H_N m^k B$ and combining the bounds, the theorem is proved. ∎

It is clear from the proof that the result of Theorem 1 is basically a Γ^k phenomenon, hence the iterated integral condition appears. Under conditions sufficient to allow a $\mathcal{B}^{m^k} \times \mathcal{F}^{k-1}$ measurable choice of $s^*(\cdot,\cdot)$ the special interpretation of expectation becomes unnecessary.

<u>Corollary 1</u>. With the choice $H_\alpha = \alpha^{\frac{1}{2}}$ each α,

$$\sum_{\alpha=1}^{N} E\, \theta_\alpha \hat{s}_\alpha = \Psi^k(G_N^k) + O(N^{\frac{1}{2}}) \tag{15}$$

uniform in $\underline{\theta}_N$.

Notice that the corollary shows that on an average, rather than total risk scale, with $H_\alpha = \alpha^{\frac{1}{2}}$, the risk incurred by the strategy $\underline{\hat{s}}$ is $\Psi^k(\frac{1}{N} G_N^k) + O(N^{-\frac{1}{2}})$ uniform in $\underline{\theta}_N$.

3. k-EXTENDED GAME THEORETIC RESULTS

The framework introduced in section 1 is quite flexible. Both decision theoretic and game theoretic problems are covered. In this section we consider a game theoretic setting, that is a situation where the information about past states is assumed to be perfect.

3.1. Specializations to a game theoretic setting.

We take $\mathfrak{X} = \Theta$, let $\mathcal{F}$ be the set of all subsets of Θ and suppose each P_θ to be degenerate at θ. S^* becomes the set of all functions from Θ^{k-1} into S and we may take $\mathfrak{E}(\underline{\theta}) = \underline{\theta} \in R^{m^k}$ where we are still identifying m^k-vectors with signed measures on Θ^k and elements of Θ^k with degenerate probabilities on Θ^k. Theorem 1 and Corollary 1 are in force in this situation so that specializations of the strategies $\hat{\underline{s}}$ provide asymptotic solutions of the k-extended game theoretic sequence compound problem.

In addition, a simple decomposition of the k-extended envelope is available in this setting that allows us to modify solutions of the unextended problem to produce solutions of the k-extended problem.

3.2. A decomposition of $\Psi^k(G_N^k)$ in the case $k > 1$.

For each $\underline{\theta} \in \Theta^{k-1}$ define $G_\alpha^k|\underline{\theta}$ to be the $\underline{\theta}$ section of G_α^k. That is, let $G_\alpha^k|\underline{\theta}$ be the measure on Θ defined by

$G^k_\alpha|\underline{\theta}(\{\theta\}) = G^k_\alpha(\{(\underline{\theta},\theta)\})$ for $\theta \in \Theta$.

Lemma 3. In the game theoretic context, for any $\underline{\theta}_N$

$$\Psi^k(G^k_N) = \sum_{\underline{\theta}\in\Theta^{k-1}} \Psi^1(G^k_N|\underline{\theta}) \; .$$

Proof. For $s^* \in S^*$

$$
\begin{aligned}
&E \sum_{\alpha=1}^{N} \theta_\alpha s^*(X_{\alpha-k+1},\ldots,X_{\alpha-1}) \\
&= \sum_{\alpha=1}^{N} \theta_\alpha s^*(\theta_{\alpha-k+1},\ldots,\theta_{\alpha-1}) = \sum_{\underline{\theta}\in\Theta^{k-1}} \sum_{\alpha \ni (\theta_{\alpha-k+1},\ldots,\theta_{\alpha-1})=\underline{\theta}} \theta_\alpha s^*(\underline{\theta}) \\
(16)\quad &= \sum_{\underline{\theta}\in\Theta^{k-1}} \left(\sum_{\alpha \ni (\theta_{\alpha-k+1},\ldots,\theta_{\alpha-1})=\underline{\theta}} \theta_\alpha \right) s^*(\underline{\theta}) = \sum_{\underline{\theta}\in\Theta^{k-1}} (G^k_N|\underline{\theta}) s^*(\underline{\theta}) \; .
\end{aligned}
$$

But (16) is minimal if $s^*(\underline{\theta}) = \sigma(G^k_N|\underline{\theta})$ for each $\underline{\theta} \in \Theta^{k-1}$.

Hence

$$\Psi^k(G^k_N) = \sum_{\underline{\theta}\in\Theta^{k-1}} (G^k_N|\underline{\theta})\,\sigma(G^k_N|\underline{\theta}) = \sum_{\underline{\theta}\in\Theta^{k-1}} \Psi^1(G^k_N|\underline{\theta}) \; . \blacksquare$$

The lemma suggests that given a strategy achieving the unextended ($k = 1$) envelope at some rate uniform in $\underline{\theta}_N$, and such that the risk function used at stage α depends on $\underline{\theta}_{\alpha-1}$ only through $G^1_{\alpha-1}$, it may be possible to achieve the k-extended envelope at the same rate by at stage α choosing the risk function according to $G^k_{\alpha-1}|(\theta_{\alpha-k+1},\ldots,\theta_{\alpha-1})$ rather than $G^1_{\alpha-1}$. Two examples of the use of this kind of technique follow.

3.3. A modification of Hannan's game theoretic strategy.

Hannan (1957) shows that for the case $k = 1$, the risk incurred in a game theoretic setting by the specialization of $\hat{\underline{s}}$

defined in (10) with $H_\alpha = (\frac{6\alpha}{m})^{\frac{1}{2}}$ achieves

$$-N^{\frac{1}{2}}(\tfrac{3}{2}\,m)^{\frac{1}{2}}B \le E\sum_{1}^{N}\theta_\alpha\hat{s}_\alpha - \Psi^1(G_N^1) \le N^{\frac{1}{2}}(6m)^{\frac{1}{2}}B\ . \tag{17}$$

If we modify Hannan's strategy by replacing $G^1_{\alpha-1}$ with $G^k_{\alpha-1}|(\theta_{\alpha-k+1},\ldots,\theta_{\alpha-1})$ and H_α with $H'_\alpha = (6\|G^k_{\alpha-1}|(\theta_{\alpha-k+1},\ldots,\theta_{\alpha-1})\|_1/m)^{\frac{1}{2}}$ we have $s_\alpha = \sigma(G^k_{\alpha-1}|(\theta_{\alpha-k+1},\ldots,\theta_{\alpha-1}) + H'_\alpha Z)$. Then

$$\sum_{\alpha=1}^{N}\theta_\alpha s_\alpha - \Psi^k(G_N^k)$$

$$= \sum_{\underline{\theta}\in\Theta^{k-1}}\left\{\sum_{\alpha\ \ni\ (\theta_{\alpha-k+1},\ldots,\theta_{\alpha-1})=\underline{\theta}}\theta_\alpha s_\alpha - \Psi^1(G_N^k|\underline{\theta})\right\}.$$

Denote the term in brackets by $A(\underline{\theta})$ for each $\underline{\theta}\in\Theta^{k-1}$, and the indices α for which $(\theta_{\alpha-k+1},\ldots,\theta_{\alpha-1}) = \underline{\theta}$ by $\alpha_1 < \alpha_2 < \ldots < \alpha_{N(\underline{\theta})}$ where $N(\underline{\theta}) = \|G_N^k|\underline{\theta}\|_1$.

$$A(\underline{\theta}) = \sum_{j=1}^{N(\underline{\theta})} E\ \theta_{\alpha_j}\,\sigma\left(G^k_{\alpha_j-1}|\underline{\theta} + (\frac{6(j-1)}{m})^{\frac{1}{2}}Z\right) - \Psi^1(G_N^k|\underline{\theta})\ .$$

The sequence $\{G^k_{\alpha_j}|\underline{\theta}\}$ is a sequence of non-normalized empirical distributions on Θ, with $G^k_{\alpha_j}|\underline{\theta} = \theta_{\alpha_j} + G^k_{\alpha_{(j-1)}}|\underline{\theta}$ and Hannan's result is applicable. So $-N^{\frac{1}{2}}(\underline{\theta})(\frac{3}{2}\,m)^{\frac{1}{2}}B \le A(\underline{\theta}) \le N^{\frac{1}{2}}(\underline{\theta})(6m)^{\frac{1}{2}}B$. Noting that $\sum_{\underline{\theta}} N(\underline{\theta}) = N$, an application of the Schwarz inequality yields

$$-N^{\frac{1}{2}}(\tfrac{3}{2}\,m^k)^{\frac{1}{2}}B \le E\sum_{\alpha=1}^{N}\theta_\alpha s_\alpha - \Psi^k(G_N^k) \le N^{\frac{1}{2}}(6m^k)^{\frac{1}{2}}B\ , \tag{18}$$

uniform in $\underline{\theta}_N$. Comparison of (17) and (18) shows the rate of convergence for the risk of the modified procedure to the extended envelope is the same as that for the original strategy. Indeed the bounds are $m^{(k-1)/2}$ times the original bounds.

3.4. A modification of Blackwell's game theoretic strategy.

Hannan (1957) states that an unextended game theoretic strategy proposed by Blackwell (1956) achieves risk $\Psi^1(G_N^1) + O(N^{\frac{1}{2}})$. We introduce this strategy and show that a natural modification achieves risk $\Psi^k(G_N^k) + O(N^{\frac{1}{2}})$.

For each $\alpha \geq 1$ we let ϕ_α denote the $m+1$ vector $(\theta_\alpha, \theta_\alpha s_\alpha)$ and $\bar{\phi}_\alpha = \frac{1}{\alpha}\sum_{\beta=1}^{\alpha} \phi_\beta = (\bar{\theta}_\alpha, \bar{r}_\alpha)$. With Δ the convex subset of R^{m+1} defined by

$$\Delta = \{(w,u) \in R^{m+1} \mid w \in R^m \text{ corresponds to a probability on } \Theta \text{ and } u \leq \Psi^1(w)\},$$

we let ρ_α be the Euclidean distance of $\bar{\phi}_\alpha$ from Δ. Arbitrarily set $\rho_0 = 0$. For each m dimensional probability vector w let $\gamma(w) = (w, \Psi^1(w))$ and let w_α be the probability vector minimizing

$$\|\bar{\phi}_\alpha - \gamma(w)\|^2 = \|\bar{\theta}_\alpha - w\|^2 + (\bar{r}_\alpha - \Psi^1(w))^2 .$$

Blackwell's strategy $\check{\underline{s}}$ is defined inductively,

$$\text{(19)} \quad \check{s}_\alpha = \begin{cases} \text{any } s \in S, \text{ if } \rho_{\alpha-1} = 0 \\ \text{any } s \in S \text{ which minimizes} \\ \bigvee_{\theta\in\Theta} (\theta(\bar{\theta}_{\alpha-1} - w_{\alpha-1}) + \theta s(\bar{r}_{\alpha-1} - \Psi^1(w_{\alpha-1})), \text{ if } \rho_{\alpha-1} \neq 0. \end{cases}$$

A proof of the following proposition is contained in the appendix.

Proposition. If S is convex, then

$$\sum_{\alpha=1}^{N} \theta_\alpha \check{s}_\alpha - \Psi^1(G_N^1) \leq N^{\frac{1}{2}}((2 + B^2)(1 + mB^2))^{\frac{1}{2}}$$

uniform in $\underline{\theta}_N$.

The convexity assumption on S appears in order to allow an application of the Minimax Theorem in the proof.

Abbreviate $\|G^k_\alpha|\underline{\theta}\|_1$ as $n_\alpha(\underline{\theta})$ for $\underline{\theta} \in \Theta^{k-1}$. With

$$\bar{\phi}^k_\alpha = \frac{1}{n_\alpha((\theta_{\alpha-k+2},\ldots,\theta_\alpha))} \sum_{\beta \ni (\theta_{\beta-k+1},\ldots,\theta_{\beta-1})=(\theta_{\alpha-k+2},\ldots,\theta_\alpha)} \phi_\beta$$

$$= (\bar{\theta}^k_\alpha, \bar{r}^k_\alpha),$$

where we interpret $\frac{0}{0} = 0$, let ρ^k_α be the Euclidean distance of $\bar{\phi}^k_\alpha$ from Δ and w^k_α be the m dimensional probability vector minimizing $\|\bar{\phi}^k_\alpha - \gamma(w)\|_2^2$. We consider a procedure $\underline{s}$ defined by

$$(20) \qquad s_\alpha = \begin{cases} \text{any } s \in S, \text{ if } \rho^k_{\alpha-1} = 0 \\ \text{any } s \in S \text{ which minimizes} \\ \bigvee_{\theta \in \Theta} (\theta(\bar{\theta}^k_{\alpha-1} - w^k_{\alpha-1}) + \theta s(\bar{r}^k_{\alpha-1} - \Psi^1(w^k_{\alpha-1})), \text{ if } \rho^k_{\alpha-1} \neq 0. \end{cases}$$

As before,

$$(21) \qquad \sum_{\alpha=1}^{N} \theta_\alpha s_\alpha - \Psi^k(G^k_N) = \sum_{\underline{\theta}\in\Theta^{k-1}} \left\{ \sum_{\alpha \ni (\theta_{\alpha-k+1},\ldots,\theta_{\alpha-1})=\underline{\theta}} \theta_\alpha s_\alpha - \Psi^1(G^k_N|\underline{\theta}) \right\} .$$

Denoting the term in brackets by $A(\underline{\theta})$ for each $\underline{\theta} \in \Theta^{k-1}$, and the indices α for which $(\theta_{\alpha-k+1},\ldots,\theta_{\alpha-1}) = \underline{\theta}$ by $\alpha_1 < \alpha_2 < \cdots < \alpha_{N(\underline{\theta})}$ with $N(\underline{\theta}) = n_N(\underline{\theta})$,

$$A(\underline{\theta}) = \sum_{j=1}^{N(\underline{\theta})} \theta_{\alpha_j} s_{\alpha_j} - \Psi^1\Big(\sum_{j=1}^{N(\underline{\theta})} \theta_{\alpha_j}\Big) .$$

With this notation $\bar{\phi}^k_{\alpha_j-1} = \frac{1}{j-1} \sum_{\ell=1}^{j-1} \phi_{\alpha_\ell}$, $\rho^k_{\alpha_j-1}$ is the Euclidean distance from Δ to $\frac{1}{j-1}\sum_{\ell=1}^{j-1}\phi_{\alpha_\ell}$, and $w^k_{\alpha_j-1}$ is the m dimensional probability vector which minimizes $\left\|\frac{1}{j-1}\sum_{\ell=1}^{j-1}\phi_{\alpha_\ell} - \gamma(w)\right\|^2$. So that comparing (19) and (20) and applying the proposition we see that if S is convex $A(\underline{\theta}) \le N^{\frac{1}{2}}(\underline{\theta})((2 + B^2)(1 + mB^2))^{\frac{1}{2}}$. So applying the Schwarz inequality the lhs (21) $\le N^{\frac{1}{2}} m^{(k-1)/2}((2 + B^2)(1 + mB^2))^{\frac{1}{2}}$, and the modification of Blackwell's strategy provides another solution of the k-extended game theoretic problem.

3.5. <u>A comment on the effect of play against a random perturbation of $G^k_{\alpha-1}$ in the k-extended setting</u>.

Recall the k-extended procedure suggested in §2.2 had the form

$$(10)\qquad \hat{s} = s^*(\tilde{T}_{\alpha-k} + H_{\alpha-k}Z,\ (X_{\alpha-k+1},\ldots,X_{\alpha-1}))\ .$$

The proof of Theorem 1 depends heavily on the fact that $\tilde{T}_{\alpha-k} + H_{\alpha-k}Z$ is independent of $(X_{\alpha-k+1},\ldots,X_\alpha)$. However, because of the degeneracy of the P_θ in the game theoretic situation, it is possible to replace $\tilde{T}_{\alpha-k} + H_{\alpha-k}Z$ by $\tilde{T}_{\alpha-1} + H_{\alpha-1}Z$, invoke <u>unextended</u> results for a sequence compound problem with Γ^k construct as the component problem, and improve on the bound of Theorem 1.

That is, redefine $\hat{s}$ by

$$\hat{s}_\alpha = s^*(\tilde{T}_{\alpha-1} + H_{\alpha-1}Z,\ (X_{\alpha-k+1},\ldots,X_{\alpha-1})).$$

Then almost everywhere $\underline{P}$, $\hat{s}_\alpha = s^*(G^N_{\alpha-1} + H_{\alpha-1}Z,\ (\theta_{\alpha-k+1},\ldots,\theta_{\alpha-1}))$ so that

$$(22)\qquad \sum_{\alpha=1}^{N} E\theta_\alpha s_\alpha = \sum_{\alpha=1}^{N}\mu(\theta_{\alpha-k+1},\ldots,\theta_\alpha)\tilde{\sigma}(G^N_{\alpha-1} + H_{\alpha-1}z)\ .$$

The unextended version of Theorem 1 applied to a compound problem with Γ^k component implies that (22) is bounded above by

$$\Psi^k(G_N^k) + \frac{1}{2} BH_N m^k + B(1 + 2 \sum_{\alpha=1}^{N} \frac{1}{H_\alpha}) \ .$$

In fact, with the choice $H_\alpha = (6\alpha)^{\frac{1}{2}} m^{-k/2}$ application of Hannan's result quoted in §3.3 gives the bounds of (18) for (22).

4. THE UNEXTENDED STATISTICAL PROBLEM, $O(N^{\frac{1}{2}})$ CONVERGENCE TO $\Psi^1(G_N^1)$ OF THE RISK OF THE NATURAL PROCEDURE IN THE TWO STATE CASE

Artificial randomization plays a major role in the solutions of the k-extended sequence compound problem offered in §2. In that section it was shown that under mild assumptions on $\mathcal{P}$, a procedure which at stage α uses an element of S^* corresponding to a risk point Γ^k Bayes versus an estimate of $G^k_{\alpha-k}$ plus <u>randomization</u> is an asymptotic solution to the k-extended problem. It is not possible to retain the generality of §1, delete the randomization and prove a result parallel to Theorem 1. Even in the unextended case there are trivial game theoretic examples in which the non-randomized version of Theorem 1 fails.

Gilliland and Hannan (1969b) and Helmers (1972) give smoothness conditions on σ that allow deletion of the randomization in some unextended game theoretic cases. Van Ryzin (1966b) shows that under some non-degeneracy conditions on the P_θ, in the unextended finite state finite act decision theoretic setting, neither the smoothness of σ nor the randomization is needed to obtain a result like Theorem 1. Ballard (1974) generalizes some of Van Ryzin's finite state finite act treatment to the k-extended level.

In this section we consider the unextended finite state restricted risk component statistical problem. Under non-degeneracy conditions similar to Van Ryzin's we apply a result of Bhattacharya (1970) - Mirahmedov (1974) and show that neither randomization nor the smoothness of σ are necessary in a two state problem.

4.1. Specializations for the unextended statistical problem, assumptions on the P_θ, the natural procedure $\underline{s}$.

We continue under the $k = 1$ version of the assumptions contained in §1.1 through §1.4, but will alter the assumptions on the P_θ contained in §1.5. Throughout §4 we will suppose that t is an R^m valued, $\mathcal{F}$ measurable map with the properties that

(23) $$\int t\,dP_\theta = \theta \quad \text{each} \quad \theta \in \Theta$$

and

(24) $$\exists\ \gamma < \infty \text{ such that } \int |\theta^0(t - \theta)|^3 dP_\theta < \gamma \text{ for each } \theta, \theta^0 \in \Theta.$$

In addition to (23) and (24), we will impose one or the other of the following two sets of conditions on t and $\mathcal{P}$.

(25) For each $\theta \in \Theta$ the $m \times m$ matrix $V_\theta = \int (t-\theta)(t-\theta)'dP_\theta$ is nonsingular.

(26) $\sum_\theta \theta t = 1$ and for each $\theta \in \Theta$ the $m \times m$ matrix $V_\theta = \int (t-\theta)(t-\theta)'dP_\theta$ is of rank $m-1$.

Abbreviate $t(X_\alpha)$ to t_α, $\sum_{\beta=1}^{\alpha} t_\beta$ to T_α and $\frac{1}{\alpha}\sum_{\beta=1}^{\alpha} V_{\theta_\beta}$ to $\bar{V}_\alpha$. $\alpha\bar{V}_\alpha$ is the covariance matrix of T_α. To obtain some of the elementary properties of $\bar{V}_\alpha$ we digress slightly to consider matrices which like $\bar{V}_\alpha$ are averages of nonnegative definite matrices.

Suppose $\Sigma_1, \ldots, \Sigma_n$ are nonnegative definite $m \times m$ matrices with common range $\mathcal{V}$, and $\underline{\pi} = (\pi_1, \ldots, \pi_n)$ is a probability vector in

R^n (that is, each $\pi_i \geq 0$ and $\sum_{i=1}^{n} \pi_i = 1$). Let Σ_π denote the average $\sum_{i=1}^{n} (\pi_i \Sigma_i)$. Σ_π is nonnegative definite and has range $\mathcal{X}$. With $0 \leq \eta_{1i} \leq \eta_{2i} \leq \ldots \leq \eta_{mi}$ the eigenvalues of Σ_i, let $\underline{\eta} = \wedge\{\eta_{ji} \mid j = 1,\ldots,m,\ i = 1,\ldots,n$ and $\eta_{ij} > 0\}$ and $\overline{\eta} = \vee\{\eta_{mi} \mid i = 1,\ldots,n\}$. The minimum positive eigenvalue of Σ_π is then greater than or equal to $\underline{\eta}$ and the largest eigenvalue of Σ_π is less than or equal to $\overline{\eta}$. In the case that each Σ_i is non-singular, the eigenvalues of Σ_π^{-1} are between $\overline{\eta}^{-1}$ and $\underline{\eta}^{-1}$.

Throughout §4 we will let $\underline{\lambda}$ denote $\wedge\{\lambda \mid \lambda > 0$ is an eigenvalue of V_θ for some $\theta \in \Theta\}$ and $\overline{\lambda}$ denote $\vee\{\lambda \mid \lambda$ is an eigenvalue of V_θ for some $\theta \in \Theta\}$. The last two comments above then apply with $\overline{V}_\alpha$ replacing Σ_π and $\underline{\lambda}$ and $\overline{\lambda}$ replacing $\underline{\eta}$ and $\overline{\eta}$.

The sequence compound procedure that we investigate in this section is $\underline{s} = (s_1,\ldots,s_N)$ where for each α

$$s_\alpha = \sigma(T_{\alpha-1}) \ . \tag{27}$$

The compound risk of $\underline{s}$ is

$$\begin{aligned} E\sum_1^N \theta_\alpha s_\alpha &= \sum_1^N E\, t_\alpha s_\alpha \\ &= E\sum_1^N t_\alpha s_{\alpha+1} + E\sum_1^N t_\alpha (s_\alpha - s_{\alpha+1}) . \end{aligned} \tag{28}$$

Denoting second term on the right above as A, Lemma 1 implies

$$\begin{aligned} E\sum_1^N \theta_\alpha s_\alpha &\leq E\, T_N \sigma(T_N) + A \\ &\leq \Psi^1(G_N^1) + A \ . \end{aligned}$$

The goal of §4 is to give useful bounds for A. After brief consideration of the problem for general finite m, we will specialize to the case $m = 2$ and show that under (25) or (26), where S is the lower boundary of a convex subset of $[0,B]^2$, there exists a constant $\mathcal{K}$ depending only on B, $\{\mathcal{V}_\theta\}_{\theta\in\Theta}$ and γ such that $A \le \mathcal{K} N^{\frac{1}{2}}$.

4.2. Bounds on A for general finite m.

Recall that $A = E \sum_{\alpha=1}^{N} t_\alpha(\sigma(T_{\alpha-1}) - \sigma(T_\alpha))$ and consider the problem of bounding a typical summand

$$(29) \qquad E\, t_{\alpha+1}(\sigma(T_\alpha) - \sigma(T_{\alpha+1})).$$

Iterating expectations,

$$(30) \qquad (29) = EE[t_{\alpha+1}(\sigma(T_\alpha) - \sigma(T_{\alpha+1}))|t_{\alpha+1}],$$

$$\le E|E[t_{\alpha+1}(\sigma(T_\alpha) - \sigma(T_{\alpha+1}))|t_{\alpha+1}]| .$$

Let W_α be independent of $t_{\alpha+1}$ with a normal distribution with the same mean and covariance structure as T_α, that is normal $(G_\alpha, \alpha\bar{\mathcal{V}}_\alpha)$. Abbreviate $t_{\alpha+1}$ to t and $E[\ |t_{\alpha+1}]$ to E^t through (38). Then,

$$(31) \quad \mathrm{rhs}(30) \le E|E^t t\sigma]_{W_\alpha+t}^{W_\alpha}| + E|E^t t(\sigma]_{T_\alpha+t}^{T_\alpha} - \sigma]_{W_\alpha+t}^{W_\alpha})| .$$

Denote the first term on the right above as $\mathcal{C}_\alpha$ and the second as $\mathcal{S}_\alpha$ and bound them separately.

First consider $\mathcal{C}_\alpha$. Let ν_1 be normal $(G_\alpha^1, \alpha\bar{\mathcal{V}}_\alpha)$ measure and ν_2 be normal $(G_\alpha^1 + t, \alpha\bar{\mathcal{V}}_\alpha)$ measure.

(32) $$|E^t t\sigma]_{W_\alpha+t}^{W_\alpha}| = |\int t\sigma\, d(\nu_1 - \nu_2)| \ .$$

Since $\|t\|_1 B$ is a bound for $|t\sigma|$, if we let $|\nu_1 - \nu_2|(R^m)$ stand for the total variation of the signed measure $\nu_1 - \nu_2$

$$\text{rhs}(32) \le B\|t\|_1 |\nu_1 - \nu_2|(R^m).$$

Under (25) the $q = 1$ specialization of Lemma A1 in the sppendix and the fact that the eigenvalues of $\alpha\bar{V}_\alpha$ are greater than or equal to $\alpha\lambda$ show that

$$|\nu_1 - \nu_2|(R^m) \le (\frac{2}{\pi})^{\frac{1}{2}}\|t\|(\alpha\lambda)^{-\frac{1}{2}}.$$

Thus under (25)

(33) $$\mathcal{C}_\alpha \le \alpha^{-\frac{1}{2}}B(\frac{2}{\lambda\pi})^{\frac{1}{2}}E\|t\|_1\|t\| \ .$$

Also,

(34) $$|E^t t\sigma]_{W_\alpha+t}^{W_\alpha}| = |E^t t\sigma]_{\frac{1}{\alpha+1}(W_\alpha+t)}^{\frac{1}{\alpha}W_\alpha}| = |\int t\sigma\, d(\nu_1 - \nu_2)|$$

where ν_1 is normal $(\alpha^{-1} G_\alpha^1, \alpha^{-1}\bar{V}_\alpha)$ measure and ν_2 is normal $((\alpha+1)^{-1}(G_\alpha^1 + t), \alpha(\alpha+1)^{-2}\bar{V}_\alpha)$ measure. Under (26) the range of $\bar{V}_\alpha$ is the orthogonal complement of the subspace of R^m generated by the vector $\underline{1} = \Sigma\theta$. Since $\underline{1}'((\alpha+1)^{-1}(G_\alpha^1 + t) - \alpha^{-1}G_\alpha^1) = 0$, Lemma A1 is applicable with q chosen to be $\alpha(\alpha+1)^{-1}$. So

(35) $$|\nu_1-\nu_2|(R^m) \le 2((m-1)\log \frac{1}{2}(\frac{2\alpha^2+2\alpha+1}{\alpha^2 + \alpha}) + \frac{(\alpha+1)^2}{2(2\alpha^2+2\alpha+1)}\eta^{-1}\|\frac{t}{\alpha+1} - \frac{1}{\alpha(\alpha+1)}G_\alpha^1\|^2)^{\frac{1}{2}}$$

where η is the minimum positive eigenvalue of $\alpha^{-1}\bar{V}_\alpha$. Then weakening (35) by replacement of η by $\alpha^{-1}\lambda$, $\log 2^{-1}(2\alpha^2 + 2\alpha + 1)(\alpha^2+\alpha)^{-1}$ by $2^{-1}(\alpha^2+\alpha)^{-1}$, and by noting that for $a,b > 0$, $(a+b)^{\frac{1}{2}} \le (a + 2a^{\frac{1}{2}}b^{\frac{1}{2}} + b)^{\frac{1}{2}} = (a^{\frac{1}{2}} + b^{\frac{1}{2}})$,

$$(36)\quad |\nu_1 - \nu_2|(R^m) \le \left(\frac{2(m-1)}{\alpha^2+\alpha}\right)^{\frac{1}{2}} + 2\lambda^{-\frac{1}{2}}\left(\frac{\alpha}{4\alpha^2 + 4\alpha + 2}\right)^{\frac{1}{2}}\left\|t - \frac{1}{\alpha}G_\alpha^1\right\| .$$

So under (26)

$$(37)\qquad \mathcal{C}_\alpha \le BE(\|t\|_1 \cdot \mathrm{rhs}(36)) .$$

In either case, (24) implies that there exists a real constant χ depending only on m, B, λ and γ such that $\mathcal{C}_\alpha \le \chi\alpha^{-\frac{1}{2}}$. Since T_α is the sum of the independent $t_1,\ldots,t_\alpha$ we might anticipate a central limit effect and hope to show that $\mathcal{S}_\alpha$ is also appropriately small. Recalling the form of $\mathcal{S}_\alpha$ from (31), bounding the absolute value of the coordinates of t by $\|t\|_\infty$ and applying the triangle inequality

$$(38)\quad \mathcal{S}_\alpha \le E\|t\|_\infty\left(\sum_\theta |E^t\theta\sigma]_{W_\alpha}^{T_\alpha}| + \sum_\theta |E^t\theta\sigma]_{W_\alpha+t}^{T_\alpha+t}|\right) .$$

Thus we address the problem of finding a useful bound for $|E\theta\sigma]_{W_\alpha+w}^{T_\alpha+w}|$ with $w \in R^m$.

Under (26) $t'\underline{1} = 1$, so any coordinate of t may be obtained from the remaining $m-1$ coordinates. We will take advantage of this fact to reduce by 1 the dimension of the vectors T_α and W_α in this situation. For $w \in R^m$ let $\check{w}$ denote the point $(w_1,\ldots,w_{n-1})' \in R^{m-1}$. For $r > 0$ define σ^r from R^{m-1} to S by $\sigma^r(y) = \sigma(w)$ for w the point of R^m such that $w'\underline{1} = r$ and $\check{w} = y$. Then we may write under (26)

$$(39) \qquad E\theta\sigma\Big]_{W_\alpha}^{T_\alpha} = E\theta\sigma^{\alpha}\Big]_{\check{W}_\alpha}^{\check{T}_\alpha}$$

and for $w \in R^m$ with $w'\underline{1} = 1$

$$(40) \qquad E\theta\sigma\Big]_{W_\alpha+w}^{T_\alpha+w} = E\theta\sigma^{\alpha+1}\Big]_{\check{W}_\alpha+\check{w}}^{\check{T}_\alpha+\check{w}} \quad .$$

The advantage of such a reduction of dimension will become apparent in the proof of Lemma 4.

For a function g mapping R^m to R^1 and $u \in R^m$ define the function g_u by $g_u(w) = g(u + w)$. For a set $\mathcal{A} \subset R^m$ define $\omega(g,\mathcal{A}) = \vee\{|g(w) - g(y)|\,|\,w,y \in \mathcal{A}\}$. With $y \in R^m$ and $r \geq 0$, $S(y,r)$ will denote $\{w \in R^m|\,\|w - y\| < r\}$.

Lemma 4. Let Y have a normal $(0,\alpha\bar{\mathcal{V}}_\alpha)$ distribution. There exist constants K_1 and K_2 depending only on m, B, $\{\mathcal{V}_\theta\}$ and γ such that under (25)

$$\Big|E\theta\sigma\Big]_{W_\alpha+w}^{T_\alpha+w}\Big| \leq K_1\alpha^{-\frac{1}{2}} + 2 \bigvee_{u\in R^m} E\omega(\theta\sigma,\, S(Y + u,\, K_2))$$

for any $w \in R^m$, and under (26)

$$\Big|E\theta\sigma\Big]_{W_\alpha}^{T_\alpha}\Big| \leq K_1\alpha^{-\frac{1}{2}} + 2 \bigvee_{u\in R^{m-1}} E\omega(\theta\sigma^{\alpha},\, S(\check{Y} + u,\, K_2))$$

and

$$\Big|E\theta\sigma\Big]_{W_\alpha+w}^{T_\alpha+w}\Big| \leq K_1\alpha^{-\frac{1}{2}} + 2 \bigvee_{u\in R^{m-1}} E\omega(\theta\sigma^{\alpha+1},S(\check{Y} + u,\, K_2))$$

for any $w \in R^m$ with $w'\underline{1} = 1$.

Proof. All of the asserted bounds will follow from applications of the weakened form of the Bhattacharya-Mirahmedov Theorem stated in

A.3. Consider first the situation under (25). Direct application of the result shows that with E_α a $m \times m$ nonsingular matrix such that $E'_\alpha E_\alpha = \bar{V}_\alpha^{-1}$, $\|E_\alpha\|$ the operator norm of E_α, g the function from R^m to R^1 defined by $g(y) = \theta\sigma(\alpha^{\frac{1}{2}}E_\alpha^{-1}y + w + G_\alpha^1)$, and

$$\epsilon_\alpha = C(m)\alpha^{-3/2}\|E_\alpha\|^3 m^{\frac{1}{2}} \sum_{\beta=1}^{\alpha} E(\sum_\theta |\theta(t_\beta - \theta_\beta)|^3,$$

$$|E\theta\sigma]_{W_\alpha+w}^{T_\alpha+w}| \le \omega(g,R^m)\epsilon_\alpha \quad + 2 \bigvee_u \int\omega(g_u,S(x,\epsilon_\alpha))d\Phi_m(x).$$

Since $\|E_\alpha\| \le \lambda^{-\frac{1}{2}}$, $\omega(g,R^m) \le B$, and (24) implies that $\sum_1^\alpha E(\sum_\theta|\theta(t_\beta - \theta_\beta)|^3) \le \alpha m\gamma$, we may set $\epsilon = C(m)\lambda^{-\frac{1}{2}}m^{3/2}\gamma$ and weaken the bound to

$$|E\theta\sigma]_{W_\alpha+w}^{T_\alpha+w}| \le \alpha^{-\frac{1}{2}}B\epsilon + 2\bigvee_u \int\omega(g_u,S(x,\alpha^{-\frac{1}{2}}\epsilon))d\Phi_m(x).$$

Now

$$\bigvee_u \int\omega(g_u,S(x,\alpha^{-\frac{1}{2}}\epsilon))d\Phi_m(x) = \bigvee_u\int\omega(g'_u,S(x,\alpha^{-\frac{1}{2}}\epsilon)d\Phi_m(x)$$

where $g'_u(y) = \theta\sigma(\alpha^{\frac{1}{2}}E_\alpha^{-1}y + u)$. But

$$\omega(g'_u,S(x,\alpha^{-\frac{1}{2}}\epsilon)) \le \omega(\theta\sigma,S(\alpha^{\frac{1}{2}}E_\alpha^{-1}x + u, \alpha^{-\frac{1}{2}}\epsilon\|\alpha^{\frac{1}{2}}E_\alpha^{-1}\|)).$$

$\|E_\alpha^{-1}\| = \eta^{\frac{1}{2}}$ where η is the largest eigenvalue of $(E_\alpha^{-1})'E_\alpha^{-1}$. It is always the case that for A a $m \times r$ matrix and B a $r \times m$ matrix, AB and BA have the same nonzero eigenvalues. Hence η is also the largest eigenvalue of $E_\alpha^{-1}(E_\alpha^{-1})' = \bar{V}_\alpha$. So under (25)

$$(41) \quad |E\theta\sigma]_{W_\alpha+w}^{T_\alpha+w}| \le \alpha^{-\frac{1}{2}}B\epsilon + 2\bigvee_u E\omega(\theta\sigma,S(Y + u, \epsilon\bar{\lambda}^{-\frac{1}{2}})).$$

The Bhattacharya- Mirhamedov result is not directly applicable to the left sides of (39) or (40) under (26), as the V_θ are singular. For Σ a $m \times m$ matrix, let Σ^0 be the $(m-1) \times (m-1)$ matrix obtained from Σ by deleting the mth row and column. V_θ^0 is then the covariance matrix of the random vector $\check{t}(X)$ under the distribution P_θ. It is the case that under (26) each V_θ^0 is nonsingular. To see this, suppose that Y has mean $\underline{0}$ and covariance matrix V_θ. $Y'\underline{1} = 0$ a.e. so that the coordinate random variables of $\check{Y}$ span the same $(m-1)$ dimensional subspace of $L_2(P)$ as do the coordinates of Y. Hence the coordinates of $\check{Y}$ are linearly independent in $L_2(P)$, that is V_θ is nonsingular.

Thus although the rate of weak convergence result is not directly applicable to lhs (39) or lhs (40), it is directly applicable to the right sides of the equations. And with D_α a $(m-1) \times (m-1)$ nonsingular matrix such that $D_\alpha{}'D_\alpha = (\bar{V}_\alpha^0)^{-1}$, $\lambda^0 = \wedge\{\lambda | \lambda$ is an eigenvalue of V_θ^0 for some $\theta \in \Theta\}$, $\bar{\lambda}^0 = \vee\{\lambda | \lambda$ is an eigenvalue of V_θ^0 for some $\theta \in \Theta\}$, $\epsilon^0 = C(m-1)\lambda^{0-\frac{1}{2}}(m-1)^{3/2}\gamma$, g^0 the function from R^{m-1} to R^1 defined by $g^0(y) = \theta\sigma^\alpha(\alpha^{\frac{1}{2}}D_\alpha^{-1}y + \check{G}_\alpha^1)$ and h^0 the function from R^{m-1} to R^1 defined by $h^0(y) = \theta\sigma^{\alpha+1}(\alpha^{\frac{1}{2}}D_\alpha^{-1}y + \check{G}_\alpha^1 + \check{w})$, under (26)

$$\left|E\theta\sigma\right]_{W_\alpha}^{T_\alpha}\Big| \le \alpha^{-\frac{1}{2}}B\epsilon^0 + 2 \bigvee_{u\in R^{m-1}} \int \omega(g_u^0, S(x,\alpha^{-\frac{1}{2}}\epsilon^0))d\Phi_{m-1}(x),$$

and for w with $w'\underline{1} = 1$

$$\left|E\theta\sigma\right]_{W_\alpha+w}^{T_\alpha+w}\Big| \le \alpha^{-\frac{1}{2}}B\epsilon^0 + 2 \bigvee_{u\in R^{m-1}} \int \omega(h_u^0, S(x,\alpha^{-\frac{1}{2}}\epsilon^0))d\Phi_{m-1}(x) .$$

Then by the same argument as applied under (25),

$$|E\theta\sigma]_{W_\alpha}^{T_\alpha}| \le \alpha^{-\frac{1}{2}} B\epsilon^0 + 2 \bigvee_{u\in R^{m-1}} E\omega(\theta\sigma^\alpha, S(\check{Y} + u, \epsilon^0(\bar{\lambda}^0)^{\frac{1}{2}})),$$

and

$$|E\theta\sigma]_{W_\alpha+w}^{T_\alpha+w}| \le \alpha^{-\frac{1}{2}} B\epsilon^0 + 2 \bigvee_{u\in R^{m-1}} E\omega(\theta\sigma^{\alpha+1}, S(\check{Y} + u, \epsilon^0(\bar{\lambda}^0)^{\frac{1}{2}})).$$

With $K_1 = B(\vee\{\epsilon,\epsilon^0\})$ and $K_2 = \vee\{\epsilon\bar{\lambda}^{\frac{1}{2}}, \epsilon^0(\bar{\lambda}^0)^{\frac{1}{2}}\}$, the lemma is proved. ∎

The bounds under (26) involve the functions σ^α and $\sigma^{\alpha+1}$ which involve the arbitrary choice to eliminate the final coordinate of the arguments of σ. It should then be noted that the apparent asymmetry of the bounds could be eliminated by replacing the terms involving the oscillation of $\theta\sigma^\alpha$ and $\theta\sigma^{\alpha+1}$ on spheres in R^{m-1} by terms involving the oscillation of $\theta\sigma$ on sets which are the intersections of the hyperplanes $\{w \in R^m | w'\underline{1} = \alpha\}$ or $\{w \in R^m | w'\underline{1} = \alpha + 1\}$ and appropriate spheres in R^m. For purposes of what follows however, the stated form is most manageable.

Applying Lemma 4 to (38), with Y normal $(\underline{0}, \alpha\bar{V}_\alpha)$, under (25)

$$(42)\quad \delta_\alpha \le E\|t_{\alpha+1}\|_\infty(2mK_1\alpha^{-\frac{1}{2}} + 4\sum_\theta \bigvee_{u\in R^m} E\omega(\theta\sigma, S(Y + u, K_2)))$$

and under (26)

$$(43)\quad \delta_\alpha \le E\|t_{\alpha+1}\|_\infty(2mK_1\alpha^{-\frac{1}{2}} + 2\sum_\theta \bigvee_{u\in R^{m-1}} E\omega(\theta\sigma^\alpha, S(\check{Y} + u, K_2))$$

$$+ 2\sum_\theta \bigvee_{u\in R^{m-1}} E\omega(\theta\sigma^{\alpha+1}, S(\check{Y} + u, K_2))).$$

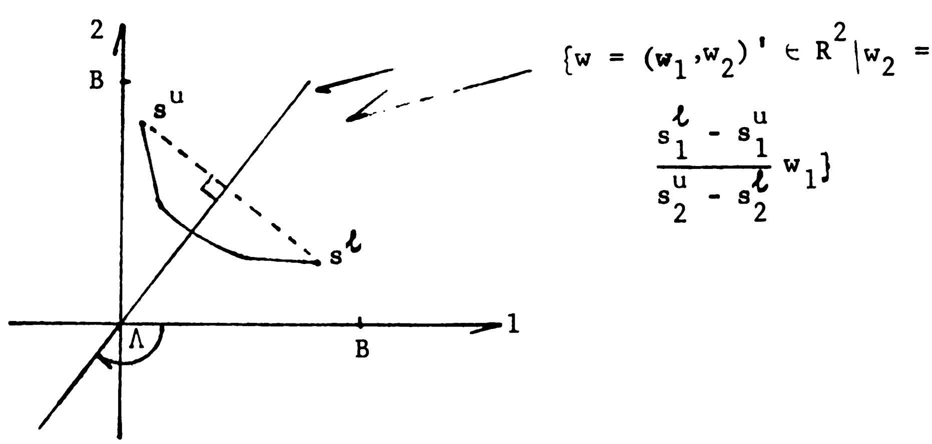
So comparing (29), (31), (33), (37), (42) and (43), provided the terms $\bigvee_u E\omega(\theta\sigma, S(Y + u, K_2))$, $\bigvee_u E\omega(\theta\sigma^{\alpha}, S(\check{Y} + u, K_2))$ and $\bigvee_u E\omega(\theta\sigma^{\alpha+1}, S(\check{Y} + u, K_2))$ are of order $O(\alpha^{-\frac{1}{2}})$, A is of order $O(N^{\frac{1}{2}})$.

4.3. <u>The two state problem</u>.

For the remainder of §4 we specialize to the case $m = 2$. Further, we suppose that S is the lower boundary of a convex subset of $[0,B]^2$. Such a choice of S arises naturally in a situation in which the risk points corresponding to all measurable decision procedures in a component decision problem with nonnegative loss function are available, and it is determined to at each stage use an admissible risk whose maximum component is bounded by B.

The assumptions on the form of S have several useful consequences. First, there are points $s^u = (s_1^u, s_2^u)'$ and $s^{\ell} = (s_1^{\ell}, s_2^{\ell})' \in S$ such that for any $s \in S$, $s_1^u \le s_1$, $s_2^u \ge s_2$, $s_1^{\ell} \ge s_1$ and $s_2^{\ell} \le s_2$.

Figure 1



It is the case that

(44) $\sigma(w) = s^{\ell}$ for w such that $w_1 \le 0$ and $w_2 > \dfrac{s_1^{\ell} - s_1^u}{s_2^u - s_2^{\ell}} w_1$,

and

(45) $\sigma(w) = s^u$ for w such that $w_2 \le 0$ and $w_2 < \dfrac{s_1^{\ell} - s_1^u}{s_2^u - s_2^{\ell}} w_1$.

Representing points $w \in R^2$ in polar form $w = \rho(\cos\phi, \sin\phi)'$, the positive homogeneity of σ guarantees that $\sigma(w)$ is independent of $\rho > 0$. With $\Lambda \in (-\pi_{\nu}, -\pi/2)$ the angle from the positive 1 axis to the ray $\{w \in R^2 \mid w_2 = (s_1^{\ell} - s_1^u)(s_2^u - s_2^{\ell})^{-1} w_1\} \cap (-\infty, 0)^2$, the coordinates of $\sigma(w)$ are monotone in $\phi \in (\Lambda, \Lambda + 2\pi)$. Because of these monotonicities, a set $\mathcal{A}^r$ of the form $\mathcal{A}^r = \{w \in R^2 \mid \theta\sigma(w) > r\}$ must be void, all of R^2 or a region bounded by two distinct rays from the origin, possibly though not necessarily including the origin and or one or both of the rays. Also, the monotonicities guarantee that for any $r > 0$, $\theta\sigma^r(y)$ is a monotone function of its real argument y. With $\sigma^r = (\sigma_1^r, \sigma_2^r)'$, σ_1^r is nonincreasing while σ_2^r is nondecreasing.

4.4. $O(N^{\frac{1}{2}})$ <u>rates in the two state problem</u>.

<u>Theorem 2</u>. Under hypothesis (25), in a situation where $m = 2$ and S is the lower boundary of a convex subset of $[0,B]^2$, there exists a real constant $\mathcal{K}_1$ depending only on B, $\{\mathcal{V}_\theta\}$ and γ such that $A \le \mathcal{K}_1 N^{\frac{1}{2}}$.

<u>Proof</u>. In view of the discussion in §4.2 it suffices to show that there exists a real constant K_3 depending only on B, $\{\mathcal{V}_\theta\}$ and γ such that for Y with a normal $(\underline{0}, \alpha\overline{\mathcal{V}}_\alpha)$ distribution, $\bigvee_{u \in R^m} E\omega(\theta\sigma, S(Y + u, K_2)) \le K_3\alpha^{-\frac{1}{2}}$, where K_2 is the constant from Lemma 4.

Define functions g and h from R^2 to R^1 by

$g(w) = \bigvee\{\theta\sigma(x) \mid x \in R^2$ with $\|x - w\| < K_2\}$ and

$$h(w) = \wedge\{\theta\sigma(x) \mid x \in R^2 \text{ with } \|x - w\| < K_2\}.$$

Notice that $\omega(\theta\sigma, S(w,K_2)) = g(w) - h(w)$. Both g and h are Borel measurable. For example, with $r > 0$ and $\mathcal{A}^r =$ $\{w \in R^2 \mid \theta\sigma(w) > r\}$,

$$\{w \in R^2 \mid g(w) > r\} = \{w \in R^2 \mid S(w,K_2) \cap \mathcal{A}^r \neq \phi\},$$

$$= \{w \in R^2 \mid \|w - x\| < K_2 \text{ for some } x \in \mathcal{A}^r\}$$

which is an open set in R^2. So with ν normal $(u, \alpha\bar{V}_\alpha)$ measure

(46) $$E\omega(\theta\sigma, S(Y + u, K_2)) = \int g - h \, d\nu .$$

By the Fubini representation of the integrals of the nonnegative functions g and h,

$$\text{rhs}(46) = \int_0^B \nu(\{w \mid g(w) > r\}) - \nu(\{w \mid h(w) > r\}) dr .$$

Letting $\mathcal{A}^r$ be as above, and for any $\mathcal{B} \subset R^2$ denoting $\{w \in R^2 \mid \|w - x\| < \varepsilon \text{ for some } x \text{ in } \mathcal{B}\}$ as $\mathcal{B}_\epsilon$ and the complement of $\mathcal{B}$ as $\mathcal{B}^c$, notice that

$$\{w \mid g(w) > r\} = \mathcal{A}^r_{K_2}$$

and

$$\{w \mid h(w) > r\} \supset (((\mathcal{A}^r)^c)_{K_2})^c .$$

So

(47) $$E\omega(\theta\sigma, S(Y + u, K_2)) \leq \int_0^B \nu(\mathcal{A}^r_{K_2} - (((\mathcal{A}^r)^c)_{K_2})^c) dr.$$

But because $\mathcal{A}^r$ has one of the forms indicated in §4.3, $\mathcal{A}^r_{K_2}$ - $(((\mathcal{A}^r)^c)_{K_2})^c$ is either void or is a subset of the union of two closed infinite strips in R^2 of width $2K_2$. Lemma A2 then implies that the integrand in (47) is bounded by $2(\frac{2}{\pi\lambda})^{\frac{1}{2}}(2K_2)$ where λ is the smallest eigenvalue of $\alpha\bar{V}_\alpha$. Hence the integrand is bounded by $4K_2(\frac{2}{\pi})^{\frac{1}{2}}(\lambda)^{-\frac{1}{2}}\alpha^{-\frac{1}{2}}$ and

$$E\omega(\theta\sigma, S(Y + u, K_2)) \leq 4BK_2(\frac{2}{\pi})^{\frac{1}{2}}(\lambda)^{-\frac{1}{2}}\alpha^{-\frac{1}{2}} .$$

The bound is uniform in u so that with $K_3 = 4BK_2(\frac{2}{\pi})^{\frac{1}{2}}(\lambda)^{-\frac{1}{2}}$ the proof is complete. ∎

In the $m = 2$ case, hypothesis (26) implies that the matrices V_1 and V_2 have the forms

$$V_1 = \begin{pmatrix} v_1^2 & -v_1^2 \\ -v_1^2 & v_1^2 \end{pmatrix} \quad \text{and} \quad V_2 = \begin{pmatrix} v_2^2 & -v_2^2 \\ -v_2^2 & v_2^2 \end{pmatrix}$$

for v_1 and v_2 nonzero real numbers. Note then that with $v = (v_1^2, v_2^2)'$, if Y has a normal $(\underline{0}, \alpha\bar{V}_\alpha)$ distribution, $\check{Y}$ is univariate normal $(0, v'G_\alpha^1)$.

<u>Theorem</u> <u>3</u>. Under hypothesis (26), in a situation where $m = 2$ and S is the lower boundary of a convex subset of $[0,B]^2$, there exists a real constant $\mathcal{K}_2$ depending only on B, $\{V_\theta\}$ and γ such that $A \leq \mathcal{K}_2 N^{\frac{1}{2}}$.

<u>Proof</u>. In view of the discussion in §4.2 it suffices to show that there exists a real constant K_4 depending only on B, $\{V_\theta\}$ and γ such that for Y with a normal $(\underline{0}, \alpha\bar{V}_\alpha)$ distribution

$$\bigvee_{u \in R} E\omega(\theta\sigma^\alpha, S(\check{Y} + u, K_2)) < K_4\alpha^{-\frac{1}{2}}$$

and

$$\bigvee_{u \in R} E\omega(\theta\sigma^{\alpha+1}, S(\check{Y} + u, K_2)) < K_4\alpha^{-\frac{1}{2}}$$

where K_2 is the constant from Lemma 4.

By the monotonicities of σ^{α} and $\sigma^{\alpha+1}$, for $y \in R$

$$\omega(\theta\sigma^{\alpha}, S(y, K_2)) \leq (-1)^{\theta-1}(\theta\sigma^{\alpha}(y - K_2) - \theta\sigma^{\alpha}(y + K_2)),$$

and

$$\omega(\theta\sigma^{\alpha+1}, S(y, K_2)) \leq (-1)^{\theta-1}(\theta\sigma^{\alpha+1}(y - K_2) - \theta\sigma^{\alpha+1}(y + K_2)).$$

So that with ν_1 univariate normal $(u - K_2, v'G^1_{\alpha})$ measure and ν_2 normal $(u + K_2, v'G^1_{\alpha})$ measure,

$$E\omega(\theta\sigma^{\alpha}, S(\check{Y} + u, K_2))) \leq (-1)^{\theta-1}\int \theta\sigma^{\alpha} d(\nu_1 - \nu_2)$$

and

$$E\omega(\theta\sigma^{\alpha+1}, S(\check{Y} + u, K_2)) \leq (-1)^{\theta-1}\int \theta\sigma^{\alpha+1} d(\nu_1 - \nu_2).$$

But the right sides above are bounded in absolute value by $B|\nu_1 - \nu_2|(R^1)$. And Lemma A1 shows that $|\nu_1 - \nu_2|(R^1) \leq 2^{\frac{1}{2}}\pi^{-\frac{1}{2}}\eta^{-\frac{1}{2}}(2K_2)$ where η is the variance of $\check{Y}$. With $v = \wedge\{v_1^2, v_2^2\}$ we may then bound

$$E\omega(\theta\sigma^{\alpha}, S(\check{Y} + u, K_2)) \leq 2^{3/2}\pi^{-\frac{1}{2}}v^{-\frac{1}{2}}K_2\alpha^{-\frac{1}{2}}$$

and

$$E\omega(\theta\sigma^{\alpha+1}, S(\check{Y} + u, K_2)) \leq 2^{3/2}\pi^{-\frac{1}{2}}v^{-\frac{1}{2}}K_2\alpha^{-\frac{1}{2}}.$$

Since the bounds above are uniform in u the proof is complete. ∎

APPENDIX

APPENDIX

A.1. A bound for the risk of Blackwell's unextended strategy.

The following proof referred to in §3.4 derives from a 1957 note of Hannan.

Proposition. If S is convex, then

$$\sum_{\alpha=1}^{N} \theta_\alpha \check{s}_\alpha - \Psi^1(G_N^1) \geq N^{\frac{1}{2}}((2 + B^2)(1 + mB^2))^{\frac{1}{2}}$$

uniform in $\underline{\theta}_N$.

Proof. Note that the concavity of Ψ^1 and the definition of $w_{\alpha-1}$ imply that if $\rho_{\alpha-1} > 0$

(1) $$(\gamma(w) - \gamma(w_{\alpha-1}))(\bar{\phi}_{\alpha-1} - \gamma(w_{\alpha-1})) \geq 0 \quad \text{for any } w$$

and

(2) $$\bar{r}_{\alpha-1} - \Psi^1(w_{\alpha-1}) > 0 \ .$$

We first show that

(3) $$\rho_\alpha^2 \leq (1 - \frac{2}{\alpha})\rho_{\alpha-1}^2 + \frac{2 + B^2}{\alpha^2} \ .$$

If $\rho_{\alpha-1} = 0$, then $\rho_\alpha^2 \leq \|\bar{\phi}_\alpha - \bar{\phi}_{\alpha-1}\|^2$. But $\bar{\phi}_\alpha - \bar{\phi}_{\alpha-1} = \frac{1}{\alpha}(\phi_\alpha - \bar{\phi}_{\alpha-1})$ and $\|\phi_\alpha - \bar{\phi}_{\alpha-1}\|^2 = \|\theta_\alpha - \bar{\theta}_{\alpha-1}\|^2 + (\theta_\alpha s_\alpha - \bar{r}_{\alpha-1})^2$
$\leq 2 + B^2$,

so that (3) holds in this case. In the case that $\rho_{\alpha-1} > 0$ we first observe $\rho_\alpha^2 \leq \|\bar{\phi}_\alpha - \gamma(w_{\alpha-1})\|^2$. Abbreviating $\gamma(w_\alpha)$ to γ_α,

(4) $$\rho_\alpha^2 \leq \|\bar{\phi}_{\alpha-1} - \gamma_{\alpha-1}\|^2 + 2(\bar{\phi}_{\alpha-1} - \gamma_{\alpha-1})(\bar{\phi}_\alpha - \bar{\phi}_{\alpha-1}) + \|\bar{\phi}_\alpha - \bar{\phi}_{\alpha-1}\|^2 .$$

Using the identity

$$\bar{\phi}_\alpha - \bar{\phi}_{\alpha-1} = \frac{1}{\alpha}((\phi_\alpha - \gamma_{\alpha-1}) + (\gamma_{\alpha-1} - \bar{\phi}_{\alpha-1}))$$

on the right hand side of (4)

(5) $$\rho_\alpha^2 \leq (1 - \frac{2}{\alpha})\|\bar{\phi}_{\alpha-1} - \gamma_{\alpha-1}\|^2 + \frac{2}{\alpha}(\bar{\phi}_{\alpha-1} - \gamma_{\alpha-1})(\phi_\alpha - \gamma_{\alpha-1}) + \|\bar{\phi}_\alpha - \bar{\phi}_{\alpha-1}\|^2 .$$

We can show that the cross product term above is non-positive and hence weaken (5) by its omission. To accomplish this, consider a game where I has pure strategies $\theta \in \Theta$, II has pure strategies $s \in S$, and the risk R is taken to be $R(w,s) = (w,ws)(\bar{\phi}_{\alpha-1} - \gamma_{\alpha-1})$ for probability vectors w. $\check{s}_\alpha$ was chosen to minimize

$$\bigvee_{\theta\in\Theta} (\theta, \theta s)(\bar{\phi}_{\alpha-1} - \gamma_{\alpha-1}) = \bigvee_{w} (w,ws)(\bar{\phi}_{\alpha-1} - \gamma_{\alpha-1}) .$$

Thus

(6) $$\bigvee_{\theta\in\Theta} (\theta,\theta\check{s}_\alpha)(\bar{\phi}_{\alpha-1} - \gamma_{\alpha-1}) = \bigvee_{w} (w,w\check{s}_\alpha)(\bar{\phi}_{\alpha-1} - \gamma_{\alpha-1})$$

$$= \bigwedge_{s\in S} \bigvee_{w} (w,ws)(\bar{\phi}_{\alpha-1} - \gamma_{\alpha-1}) .$$

Applying the Minimax Theorem to the game described above we have

$$\text{rhs}(6) = \bigvee_{w} \bigwedge_{s\in S} (w(\bar{\theta}_{\alpha-1} - w_{\alpha-1}) + ws(\bar{r}_{\alpha-1} - \Psi^1(w_{\alpha-1}))).$$

Applying (2) and then (1),

$$\text{rhs}(6) = \sup_{w} \gamma(w)(\bar{\phi}_{\alpha-1} - \gamma_{\alpha-1})$$

$$= \gamma(w_{\alpha-1})(\bar{\phi}_{\alpha-1} - \gamma_{\alpha-1}).$$

So that the product term in (5) is non-positive and we may write

$$(7) \qquad \rho_{\gamma}^{2} \le (1 - \tfrac{2}{\alpha})\|\bar{\phi}_{\alpha-1} - \gamma_{\alpha-1}\|^2 + \|\bar{\phi}_{\alpha} - \bar{\phi}_{\alpha-1}\|^2 .$$

Since $\|\bar{\phi}_{\alpha} - \bar{\phi}_{\alpha-1}\|^2 \le \alpha^{-2}(2 + B^2)$, the definition of $\rho_{\alpha-1}$ and (7) yield (3) in the $\rho_{\alpha-1} > 0$ case as well as in the $\rho_{\alpha-1} = 0$ case.

Iteration of (3) shows that

$$(8) \quad \rho_N^2 \le (2+B^2)\left(\frac{2}{N(N-1)}(\tfrac{1}{2})^2 + \frac{(3)(2)}{N(N-1)}(\tfrac{1}{3})^2 + \dots + \frac{N(N-1)}{N(N-1)}(\tfrac{1}{N})^2\right) .$$

$$\le (2 + B^2)\left(\frac{1}{N(N-1)}\right)\left(\frac{N-1}{N} + \frac{N-2}{N-1} + \dots + \frac{1}{2}\right)$$

$$< (2 + B^2)\frac{1}{N} .$$

Consider $\bar{r}_N - \Psi^1(\bar{\theta}_N) = \bar{r}_N - \Psi^1(w_N) + \Psi^1(w_N) - \Psi^1(\bar{\theta}_N)$.

$$\Psi^1(w_N) - \Psi^1(\bar{\theta}_N) = w_N\sigma(w_N) - \bar{\theta}_N\sigma(\bar{\theta}_N)$$

$$\le w_N\sigma(\bar{\theta}_N) - \bar{\theta}_N\sigma(\bar{\theta}_N)$$

$$\le \|\bar{\theta}_N - w_N\| m^{\frac{1}{2}}B .$$

So

$$(9) \qquad \bar{r}_N - \Psi^1(\bar{\theta}_N) \le \|\bar{\theta}_N - w_N\| m^{\frac{1}{2}}B + (\bar{r}_N - \Psi^1(w_N)).$$

Recalling $\|\bar{\theta}_N - w_N\|^2 + (\bar{r}_N - \Psi^1(w_N))^2 = \rho_N^2 < N^{-1}(2 + B^2)$ the problem of bounding $\bar{r}_N - \Psi^1(\bar{\theta}_N)$ has been reduced to the problem of bounding the function of two real variables $f(u,v) = u + \sqrt{m}\, Bv$ on the ball $u^2 + v^2 \le \rho_N^2$. For such u and v, $f(u,v) \le \rho_N(1 + mB^2)^{\frac{1}{2}}$. Hence

$$(10) \qquad \bar{r}_N - \Psi^1(\bar{\theta}_N) \le N^{-\frac{1}{2}}(2 + B^2)^{\frac{1}{2}}(1 + mB^2)^{\frac{1}{2}}$$

and multiplication of (10) by N completes the proof. ∎

A.2. Two lemmas concerning properties of multivariate normal distributions.

The first lemma in this section is used in §4.2 and provides bounds on the total variation of the difference of certain multivariate normal measures.

Lemma A1. Let Σ be a nonnegative definite $m \times m$ matrix with range $\mathcal{X}$. With w and v elements of R^m such that $w-v \in \mathcal{X}$, $q > 0$, and $r = \operatorname{rank} \Sigma$, let ν_1 be normal (v,Σ) measure on R^m, ν_2 be normal $(w, q^2\Sigma)$ measure on R^m, and N be any $r \times m$ matrix such that $N\Sigma N' = I_r$, the $r \times r$ identity matrix. Then

$$|\nu_1 - \nu_2|(R^m) \leq 2(r \log(\frac{q}{2} + \frac{1}{2q}) + \frac{1}{2(1+q^2)}\|N(w-v)\|^2)^{\frac{1}{2}} .$$

In the case $q = 1$

$$|\nu_1 - \nu_2|(R^m) = 2(\Phi(\frac{\|N(w-v)\|}{2}) - \Phi(\frac{-\|N(w-v)\|}{2}))$$

$$\leq 2^{\frac{1}{2}}\pi^{-\frac{1}{2}}\|N(w-v)\| .$$

Further, if η is the minimum positive eigenvalue of Σ, $\|N(w-v)\| \leq \eta^{-\frac{1}{2}}\|w-v\|$.

Proof. Let $Y = (Y_1,\ldots,Y_m)'$ have a normal $(\underline{0},\Sigma)$ distribution.

$$|\nu_1 - \nu_2|(R^m) = 2 \bigvee_{\mathcal{A}\in\mathcal{B}^m} (P((Y + v) \in \mathcal{A}) - P((qY + w) \in \mathcal{A})),$$

$$= 2 \bigvee_{\mathcal{A}\in\mathcal{B}^m} (P(Y \in \mathcal{A}-v) - P(qY + (w-v) \in \mathcal{A}-v)),$$

$$= |\nu_1' - \nu_2'|(R^m),$$

where ν_1' is normal $(\underline{0},\Sigma)$ measure and ν_2' is normal $(w-v, q^2\Sigma)$ measure. So for the rest of the proof we suppose $v = \underline{0}$ and $w \in \mathcal{N}$.

Since Σ is of rank r, the subspace of $L_2(P)$ spanned by the coordinate random variables of Y is of dimension r. The coordinates of $Z = NY$ form an orthonormal basis for this subspace. Because each Y_i is a linear combination of $\{Z_i\}_{i=1}^r$ the coordinates of Z, there exists a $m \times r$ matrix M such that $Y = MZ$. Then $Z = NMZ$ and the uniqueness of representation of elements of a subspace of $L_2(P)$ as linear combinations of elements of a basis for that subspace guarantees that $NM = I_r$. Thus, since $MM' = \Sigma$, the range of M is $\mathcal{N}$. While MN is not I_m it does act like the identity on $\mathcal{N}$. That is, if $h \in \mathcal{N}$, $h = Mx$ for some $x \in R^r$ and $MNh = MNMx = Mx = h$.

Both ν_1 and ν_2 concentrate on $\mathcal{N}$, for if $x \in \mathcal{N}^\perp$, both $x'Y$ and $x'(qY + w)$ are 0 a.e.. Thus

$$|\nu_1 - \nu_2|(R^m) = 2 \bigvee_{\mathcal{A}\in\mathcal{B}^m} (\nu_1 - \nu_2)(\mathcal{A}),$$

$$= 2 \bigvee_{\mathcal{A}\in\mathcal{B}^m\cap\mathcal{N}} (\nu_1 - \nu_2)(\mathcal{A}),$$

where $\mathcal{B}^m \cap \mathcal{N}$ is the sigma algebra of Borel subsets of $\mathcal{N}$. Now for

$\mathcal{A} \in \mathcal{B}^m \cap \mathcal{X}$

$$(\nu_1 - \nu_2)(\mathcal{A}) = P(Y \in \mathcal{A}) - P((qY + w) \in \mathcal{A}).$$

Also, $P(Y \in \mathcal{A}) = P(NY \in N\mathcal{A})$ and $P((qY + w) \in \mathcal{A}) = P((qNY + Nw) \in N\mathcal{A})$, the equalities following because $MN\mathcal{A} = \mathcal{A}$, $w \in \mathcal{X}$, and $Y \in \mathcal{X}$ with probability one. So

$$(\nu_1 - \nu_2)(\mathcal{A}) = P(NY \in N\mathcal{A}) - P((qNY + Nw) \in N\mathcal{A}),$$

and

(11) $$|\nu_1 - \nu_2|(R^m) = 2 \bigvee_{\mathcal{A} \in \mathcal{B}^m \cap \mathcal{X}} (P(NY \in N\mathcal{A}) - P((qNY + Nw) \in N\mathcal{A}).$$

But as $\mathcal{A}$ ranges over $\mathcal{B}^m \cap \mathcal{X}$, $N\mathcal{A}$ ranges over $\mathcal{B}^r$. So with μ_1 normal $(\underline{0}, I_r)$ measure on R^r and μ_2 normal $(Nw, q^2 I_r)$ measure on R^r

$$|\nu_1 - \nu_2|(R^m) = \text{rhs}(11) = |\mu_1 - \mu_2|(R^r),$$

and we bound $|\mu_1 - \mu_2|(R^r)$.

Letting f_1 be the density of μ_1 with respect to Lebesgue measure on R^r and f_2 be the density of μ_2,

(12) $$|\mu_1 - \mu_2|(R^r) = \int |f_1 - f_2| dx .$$

It is well known (see for example section 3 of Hannan (1960)) that with $\rho = \int f_1^{\frac{1}{2}} f_2^{\frac{1}{2}} dx$,

(13) $$(\text{rhs}(12))^2 \le 4\int (f_1^{\frac{1}{2}} - f_2^{\frac{1}{2}})^2 dx = 8(1-\rho) \le -8 \log \rho .$$

Abbreviating Nw to d,

$$\rho = (\det q^2 I_r)^{-\frac{1}{4}} \int (2\pi)^{-\frac{1}{2}} \exp -\tfrac{1}{2}(\tfrac{1}{2}(1+\frac{1}{q^2})x'x - \frac{1}{q^2}x'd + \frac{1}{2q^2}d'd)dx,$$

$$= q^{-r/2}(\tfrac{1}{2}(1+\frac{1}{q^2}))^{-r/2} \int \exp(\frac{x'd}{2q^2} - \frac{d'd}{4q^2}) f_3(x)dx,$$

where f_3 is the normal $(\underline{0}, (\tfrac{1}{2}(1+\frac{1}{q^2}))^{-1} I_r)$ density. So

$$(14) \qquad \rho = (\frac{q}{2} + \frac{1}{2q})^{-r/2} \exp(-\frac{d'd}{4q^2} + \tfrac{1}{2}(\tfrac{1}{2}(1+\frac{1}{q^2}))^{-1}(\frac{d}{2q^2})'(\frac{d}{2q^2})),$$

$$= (\frac{q}{2} + \frac{1}{2q})^{-r/2} \exp(-\frac{d'd}{4(1+q^2)}) \ .$$

Combining (12) through (14) the first bound holds.

Now consider the special case where $q = 1$. Here

$$|\mu_1 - \mu_2|(R^r) = \int |1 - \frac{f_2}{f_1}| d\Phi_r$$

$$= \int |1 - \exp(-\frac{d'd}{2} + x'd)| d\Phi_r(x),$$

$$= \int |1 - \exp(-\frac{\|d\|^2}{2} + \|d\|z)| d\Phi(z),$$

$$= \int_{-\infty}^{\frac{\|d\|}{2}} (1 - \exp(-\frac{\|d\|^2}{2} + \|d\|z) d\Phi(z)$$

$$+ \int_{\frac{\|d\|}{2}}^{\infty} (\exp(-\frac{\|d\|^2}{2} + \|d\|z) - 1) d\Phi(z),$$

$$= 2(\Phi(\frac{\|d\|}{2}) - \Phi(-\frac{\|d\|}{2})),$$

and the special bound for the $q = 1$ case holds.

Finally, notice that $\frac{\|Nw\|}{\|w\|} = \frac{\|d\|}{\|Md\|}$. For any $a \in R^m$, $\|Ma\|^2 = a'M'Ma \geq \|a\|^2\lambda$ where λ is the minimum eigenvalue of $M'M$. It is always the case that for A a $m \times r$ matrix and B a $r \times m$ matrix AB, and BA have the same nonzero eigenvalues, so $M'M$ has the same nonzero eigenvalues as MM'. Also $r \leq \text{rank } MN = \text{rank } N'M'MN \leq \text{rank } M'M \leq r$. Thus $M'M$ is nonsingular, $\lambda = \eta$, $\frac{\|d\|^2}{\|Md\|^2} \leq \frac{1}{\eta}$ and the proof is complete. ∎

The second lemma is used in §4.4 to bound the probability of an infinite strip in R^2 under a nonsingular bivariate normal measure.

Lemma A2. Let L be a hyperplane in R^m. For $\varepsilon > 0$ let $L_\varepsilon = \{w \in R^2 | \|w - y\| < \varepsilon$ for some $y \in L\}$. Let $v \in R^m$, Σ be a $m \times m$ nonnegative definite matrix and ν be normal (v,Σ) measure on R^m. Then

$$\nu(L) \leq \left(\frac{2}{\pi\eta}\right)^{\frac{1}{2}}\varepsilon$$

where η is the smallest eigenvalue of Σ.

Proof. Let W have a normal (v,Σ) distribution. Denote by $d(w)$ the distance from a $w \in R^m$ to L. Then $\nu(L_\varepsilon) = P(d(W) < \varepsilon)$. Let y_0 be a point of L and u be a unit vector orthogonal to $L - y_0$. uu' is then the matrix of orthogonal projection onto the line generated by u. So $d(W) = \|uu'(W - y_0)\|$. But $\|uu'(W - y_0)\| = |u'(W - y_0)|$. Since $u'(W - y_0)$ is univariate normal $(u'(v - y_0), u'\Sigma u)$, $P(|u'(W - y_0)| < \varepsilon)$ is the standard normal probability of some interval of length $2\varepsilon(u'\Sigma u)^{-\frac{1}{2}}$. The stated bound is then a weakening. ∎

A.3. A rate of weak convergence result of Bhattacharya-Mirahmedov.

Mirahmedov (1974) proves the following improvement of a rate of weak convergence result due to Bhattacharya (1970).

Theorem A1. (Bhattacharya-Mirahmedov) Let $X_1, X_2, \ldots, X_k, \ldots, X_n$ be independent random vectors in R^d with $E(X_k) = \underline{0}$ and $D_k = \mathrm{Cov}(X_k, X_k)$. Assume that $B_n = n^{-1}\Sigma_1^n D_k$ is non-singular, and let $B_n^{-1} = E_n' E_n$, $S_n = n^{-\frac{1}{2}} E_n \Sigma_1^n X_k$, and $L_{3n} = n^{-3/2}\Sigma_1^n E(\|E_n X_k\|^3)$. For any measurable function $g(x)$ on R^d, set $\omega(g,A) = \vee\{|g(x) - g(y)| : x,y \in A\}$, and $g_u(x) = g(x+u)$. With $S(x,r) = \{y \in R^d \,|\, \|x-y\| < r\}$,

$$\left|\int g(y)(P_n(dy) - \Phi_d(dy))\right| \leq C(d)\omega(g,R^d)L_{3n}$$

$$+ 2 \sup_u \int \omega(g_u, S(x, C(d)L_{3n}))\Phi_d(dx)$$

where P_n is the distribution of S_n, Φ_d is the d-variate standard normal distribution and $C(d)$ is a constant depending only on d.

Notice that with $\|E_n\|$ the operator norm of the matrix E_n,

$$L_{3n} \leq n^{-3/2}\|E_n\|^3 \Sigma_1^n E(\|X_k\|)^3 ,$$

$$= n^{-3/2}\|E_n\|^3 \Sigma_1^n E(\sum_{i=1}^d X_{ki}^2)^{3/2} ,$$

$$\leq n^{-3/2} \|E_n\|^3 d^{\frac{1}{2}} \Sigma_1^n E(\sum_{i=1}^d |X_{ki}|^3) ,$$

so the theorem may be weakened by the replacement of L_{3n} by the last line above.

REFERENCES

REFERENCES

Ballard, Robert John (1974). Extended rules for the sequence compound decision problem with $m \times n$ component. Doctoral thesis at Michigan State University.

Ballard, Robert John, Gilliland, Dennis C. and Hannan, James (1974). $O(N^{-\frac{1}{2}})$ convergence to k-extended Bayes risk in the sequence compound decision problem with $m \times n$ component. RM-333, Statistics and Probability, MSU.

Bhattacharya, R.N. (1970). Rates of weak convergence for the multidimensional central limit theorem. Theory of Probability and its Applications **15** 68-86.

Blackwell, D. (1956). Controlled random walks. Proc. Internat. Congress Math. Amsterdam **3** 336-338.

Brown, L.D. and Purves, R. (1973). Measurable selections of extrema. Ann. Statist. **1** 902-912.

Cover, Thomas M. and Shenar, Aaron (1974). A compound sequential Bayes predictor for sequences with an empirical Markov structure. TR-11, Department of Statistics, Stanford University.

Gilliland, Dennis C. (1969). Approximation to Bayes risk in sequences of non-finite games. Ann. Math. Statist. **40** 467-474.

Gilliland, Dennis C. and Hannan, James F. (1969a). On the extended compound decision problem. Ann. Math. Statist. **40** 1536-1541.

Gilliland, Dennis C. and Hannan, James F. (1969b). On continuity of the Bayes response and play against the past in a sequence of decision problems. RM-216, Statistics and Probability, MSU.

Gilliland, Dennis C. and Hannan, James F. (1974). The finite state compound decision problem, equivariance and restricted risk components. RM-317, Statistics and Probability, MSU.

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

Hannan, James (1957). Approximation to Bayes risk in repeated play. Contributions to the Theory of Games **3** 97-139. Princeton University Press.

Hannan, James F. (1960). Consistency of maximum likelihood estimation of discrete distributions. Contributions to Probability and Statistics, 249-257. Stanford University Press.

Helmers, Merrilee (1972). On the continuity of the Bayes response. RM-305, Statistics and Probability, MSU.

Johns, M.V., Jr. (1967). Two-action compound decision problems. Proc. Fifth Berkeley Symp. Math. Statist. Prob. **1** 463-478. University of California Press.

Mirahmedov, S.A. (1974). The rate of weak convergence in the multidimensional limit theorem. Izv. Akad. Nauk UzSSR Ser. Fiz.-Mat. Nauk **18** no. 2, 23-28, 92-93.

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

Robbins, Herbert (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Statist. **35** 1-20.

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

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

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

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

Van Ryzin, J.R. (1966b). The sequential compound decision problem with $m \times n$ finite loss matrix. Ann. Math. Statist. **37** 954-975.