EXTENDED RULES FOR THE SEQUENCE COMPOUND DECISION PROBLEM WITH m x n COMPONENT

Dissertation for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY
ROBERT JOHN BALLARD

1974

This is to certify that the

thesis entitled

EXTENDED RULES FOR THE SEQUENCE COMPOUND
DECISION PROBLEM WITH $m \times n$ COMPONENT

presented by

Robert John Ballard

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics and Probability

Dennis C. Gilliland
Major professor

Date August 12, 1974

O-7639

ABSTRACT

EXTENDED RULES FOR THE SEQUENCE COMPOUND DECISION PROBLEM WITH $m \times n$ COMPONENT

By

Robert John Ballard

For a sequence compound decision rule $\underset{\sim}{\varphi} = (\underset{\sim}{\varphi}_1, \ldots, \underset{\sim}{\varphi}_N)$, where $\underset{\sim}{\varphi}_\alpha$ is a function of the first α observations, $1 \le \alpha \le N$, let $\underset{\sim}{R}_N(\underset{\sim}{\theta}, \underset{\sim}{\varphi})$ denote the compound (average) risk at state $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_N)$. The usual standard for compound risk is $R(G_N)$ where R is the Bayes envelope for the component problem (the simple envelope) and G_N is the empirical distribution of component states $\theta_1, \ldots, \theta_N$. Much of the literature in compound decision theory has dealt with the construction of rules satisfying

$$\overline{\lim}_N \sup_{\underset{\sim}{\theta}} [\underset{\sim}{R}_N(\underset{\sim}{\theta}, \underset{\sim}{\varphi}) - R(G_N)] \le 0$$

for various components.

More stringent standards for compound risk are $R^k(G_N^k)$, $k = 1, 2, \ldots$, where R^k is the Bayes envelope for a construct called the Γ_k game and G_N^k is the empirical distribution of the k-tuples $\underset{\sim}{\theta}_1^k = (\theta_1, \ldots, \theta_k)$, $\underset{\sim}{\theta}_2^k = (\theta_2, \ldots, \theta_{k+1}), \ldots, \underset{\sim}{\theta}_{N-k+1}^k = (\theta_{N-k+1}, \ldots, \theta_N)$. The $k+1$ standard is asymptotically more stringent than the k standard where $R^1(G_N^1) = R(G_N)$.

We will consider the $m \times n$ component and demonstrate for each k, a k-extended sequence compound rule $\underset{\sim}{\varphi}$, $\underset{\sim}{\varphi}_\alpha$ being Γ_k

Bayes versus an estimate of G_α^k based on the first α-k observations, $k \leq \alpha \leq N$, which satisfies

$$\overline{\lim}_N \sup_{\underline{\theta}}[\underline{R}_N(\underline{\theta},\underline{\varphi}) - R^k(G_N^k)] \leq 0$$

with a rate of $N^{-1/5}$. Furthermore, we compute the envelope R^2 where the component is discrimination between N(-1,1) and N(1,1) and use Monte Carlo methods to estimate the compound risks for some k = 1 and k = 2 procedures and various $\underline{\theta}$ and N in order to determine possible small N advantages of extended procedures. In addition, we compute some Bayes compound risks versus strictly stationary $\underline{\theta}$ for various N.

EXTENDED RULES FOR THE SEQUENCE COMPOUND DECISION PROBLEM WITH $m \times n$ COMPONENT

By

Robert John Ballard

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

1974

TO MY PARENTS AND CINDY

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to Professor Dennis C. Gilliland for his guidance of the research for this dissertation. His comments, observations and suggestions were invaluable. Further, I am indebted to Professor James Hannan for suggesting the problem and many helpful observations.

The financial support provided by the Department of Statistics and Probability and the National Science Foundation made my graduate studied possible. I wish to thank them and Noralee Barnes who demonstrated both patience and skill in typing this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The compound decision problem introduced by Robbins (1951) involves N independent repetitions of a component problem. For a compound decision rule $\underset{\sim}{\varphi} = (\varphi_1, \ldots, \varphi_N)$, where φ_α is a function of all N observations, $1 \le \alpha \le N$, let $R_N(\underset{\sim}{\theta}, \underset{\sim}{\varphi})$ denote the compound (average) risk at state $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_N)$. The usual standard for compound risk is $R(G_N)$ where R is the Bayes envelope for the component problem (the simple envelope) and G_N is the empirical distribution of component states $\theta_1, \ldots, \theta_N$. Robbins (1951) demonstrated a "bootstrap" compound rule $\underset{\sim}{\varphi}$ satisfying the two-sided (limit) version of

$$(1) \qquad \overline{\lim}_N \sup_{\underset{\sim}{\theta}} [R_N(\underset{\sim}{\theta}, \underset{\sim}{\varphi}) - R(G_N)] \le 0$$

for the featured example where the component is discrimination between $N(-1,1)$ and $N(1,1)$. The term "bootstrap" refers to the fact that each φ_α is component Bayes versus an estimate of G_N based on all N observations. Much of the literature in compound decision theory has dealt with the construction of bootstrap rules satisfying (1) for various components. Recently, Gilliland, Hannan and Huang (1974) have shown that a large class of Bayes compound rules satisfy (1) for 2-state, compact risk component.

The sequence compound decision problem introduced by Hannan (1956),(1957a) restricts the class of compound rules to $\underset{\sim}{\varphi}$ where

$\underline{\varphi}_\alpha$ is a function of the first α observations, $1 \leq \alpha \leq N$. Hannan (1956), (1957a) and Samuel (1963) were the first to propose sequence compound rules to satisfy (1) for various components. The rules are bootstrap in nature, $\underline{\varphi}_\alpha$ being component Bayes versus an estimate of G_α based on the first α observations and possible artificial randomization, $1 \leq \alpha \leq N$. In the meantime, many additional components have been considered in the sequence compound problem.

The notion of using more stringent standards for compound risk is found in Johns (1967, §4). He introduces a set of standards $R^k(G_N^k)$, $k = 1,2,\ldots$, where $R^1(G_N^1) = R(G_N)$ and the $k+1$ standard is asymptotically more stringent than the k standard. (Here we have used notation similar to that of Gilliland and Hannan (1969) who give the most general treatment of these standards. R^k is the Bayes envelope for a construct called the Γ_k game and G_N^k is the empirical distribution of the k-tuples $\underline{\theta}_1^k = (\theta_1,\ldots,\theta_k)$, $\underline{\theta}_2^k = (\theta_2,\ldots,\theta_{k+1}),\ldots,\underline{\theta}_{N-k+1}^k = (\theta_{N-k+1},\ldots,\theta_N)$.) Johns (1967) for a two action component and fixed k gives a sequence bootstrap rule $\underline{\varphi}$, $\underline{\varphi}_\alpha$ being Γ_k Bayes versus an estimate of G_α^k based on the first α observations, $k \leq \alpha \leq N$, which satisfies the two-sided version of

$$(2) \qquad \overline{\lim}_N \sup_{\underline{\theta}}[\underline{R}_N(\underline{\theta},\underline{\varphi}) - R^k(G_N^k)] \leq 0 \ .$$

Swain (1965) in a thesis under Johns treats some squared error loss estimation components, obtaining (2) with rates later improved by Yu (1971). Swain calls the compound problem with standard $R^k(G_N^k)$ the extended compound decision problem.

We treat the extended sequence problem with $m \times n$ component and, thus, cover the original featured example of Robbins. We demonstrate for each k, a sequence compound rule which satisfies (2) with rate. For proving this asymptotic result we rely heavily on Van Ryzin's (1966b) analysis of the unextended version of the problem. Furthermore, we compute the envelope R^2 for Robbins' featured example and use Monte Carlo methods to estimate the compound risks for some $k = 1$ and $k = 2$ procedures and various $\underline{\theta}$ and N in order to determine the possible small N advantages of extended procedures. In addition, we compute some Bayes compound risks versus strictly stationary $\underline{\theta}$ for various N.

CHAPTER I

ASYMPTOTIC SOLUTIONS TO THE EXTENDED SEQUENCE COMPOUND PROBLEM

1. Preliminaries.

Consider a component decision problem with m states $\Theta = \{1,2,\ldots,m\}$ indexing $\mathcal{P} = \{P_1, P_2, \ldots, P_m\}$ where the P_i are distinct probability measures on $(\mathfrak{X}, \mathcal{B})$. Let μ be a finite measure dominating $\mathcal{P}$ such that

$$(3) \qquad f_i = dP_i/d\mu \le K, \quad i \in \Theta$$

for some positive finite constant K. There is no loss of generality in making this assumption since we may always take $\mu = \sum_{i=1}^{m} P_i$ and $K = 1$.

Suppose the action space is $\mathcal{A} = \{1,2,\ldots,n\}$ and the $m \times n$ loss matrix L has non-negative finite elements. A (randomized) component decision rule, φ, is a $\mathcal{B}$-measurable mapping into $\mathcal{A}^*$, the $(n-1)$-dimensional simplex of probability measures on $\mathcal{A}$. That is, $\varphi = (\varphi_1, \ldots, \varphi_n)$ where $\varphi_j \ge 0$, $1 \le j \le n$, and $\Sigma_1^n \varphi_j = 1$. The risk of φ at state i is

$$(4) \qquad R(i,\varphi) = \int \left(\sum_{j=1}^{n} L(i,j)\varphi_j \right) dP_i \, .$$

Let $\underline{P} = P_{\theta_1} \times P_{\theta_2} \times \ldots \times P_{\theta_N}$ and $\underline{\varphi} = (\underline{\varphi}_1, \underline{\varphi}_2, \ldots, \underline{\varphi}_N)$ where for each α, $\underline{\varphi}_\alpha$ is a $\mathcal{B}^\alpha$-measurable mapping into $\mathcal{A}^*$. The risk incurred in the α component decision by the sequence compound rule

$\underset{\sim}{\varphi}$ is

(5) $$R_\alpha(\underline{\theta},\underset{\sim}{\varphi}) = \int\Big(\sum_{j=1}^{n} L(\theta_\alpha, j)\underset{\sim}{\varphi}_{\alpha,j}\Big)d\underline{P}$$

and compound risk at N repetitions and state $\underline{\theta}$ is

(6) $$\underset{\sim}{R}_N(\underline{\theta},\underset{\sim}{\varphi}) = N^{-1}\sum_{\alpha=1}^{N} R_\alpha(\underline{\theta},\underset{\sim}{\varphi}) \ .$$

In order to motivate the construction of rules satisfying (2) we describe the Γ_k decision problem. In our case it is an $m^k \times n$ problem with states $\underline{i}^k = (i_1,\ldots,i_k) \in \Theta^k$ indexing product distributions $P_{\underline{i}^k} \in \mathscr{P}^k$. The loss matrix L_k is $m^k \times n$ with $L_k(\underline{i}^k, j) = L(i_k, j)$. A (randomized) decision rule φ in the Γ_k problem is a $\mathscr{B}^k$-measurable mapping into $\mathscr{A}^*$. Letting $R^k(\underline{i}^k,\varphi)$ denote its risk at state $\underline{i}^k$, the Bayes risk of φ versus a prior G on Θ^k is

(7) $$\begin{aligned} R^k(G,\varphi) &\equiv \sum_{\underline{i}^k} R^k(\underline{i}^k,\varphi)G_{\underline{i}^k} \\ &= \int\Big[\sum_{j=1}^{n}\varphi_j(\underline{x}^k)\sum_{\underline{i}^k} L(i_k,j)f_{\underline{i}^k}(\underline{x}^k)G_{\underline{i}^k}\Big]d\mu^k(\underline{x}^k) \end{aligned}$$

where $f_{\underline{i}^k}(\underline{x}^k) \equiv f_{i_1}(x_1)f_{i_2}(x_2)\ \ldots\ f_{i_k}(x_k)$ and μ^k is the k-fold product of μ. Let L_k^j denote the $j^{\underline{th}}$ column of L_k, for $1 \le j \le n$, $\underline{f}$ denote the $m^k \times 1$ matrix of densities $f_{\underline{i}^k}$ and $L_k^j\underline{f}$ denote the $m^k \times 1$ matrix with components $L_k(\underline{i}^k, j)f_{\underline{i}^k}$. Letting $(\ ,\ )_k$ denote the usual inner product in E^{m^k}-space, it follows from (7) that a Bayes rule in the Γ_k problem places all its mass on the j's which minimize $\Delta_j(\underline{x}^k) \equiv (L_k^j\underline{f}(\underline{x}^k), G)_k$. A particular version denoted by $\varphi^k\{G\}$ is

$$\varphi_j^k\{G;\ \underline{x}^k\} = \begin{cases} 1 & \text{if } j \text{ is the smallest integer such that } \Delta_j(\underline{x}^k) = \min_{1\le j\le n} \Delta_j(\underline{x}^k) \\ 0 & \text{for the other } j\ . \end{cases} \tag{8}$$

The specification $k = 1$ in the Γ_k construct gives the component decision problem. For simplicity of notation we abbreviate $k = 1$ by omission whenever it is both possible and convenient.

Theorem 2 of Gilliland and Hannan (1969) suggests that for the compound problem the sequence rule which plays Γ_k Bayes versus an estimate of G_α^k, $k \le \alpha \le N$, may satisfy (2). Let $\underline{x}_\alpha = (x_1,\ldots,x_\alpha)$ and $\underline{x}_\nu^k = (x_\nu,\ldots,x_{\nu+k-1})$ for $\alpha,\nu = 1,2,\ldots$. The sequence compound rules $\underline{\varphi}$ that we investigate have

$$\underline{\varphi}_\alpha(\underline{x}_\alpha) = \varphi^k\{\hat{G}_\alpha^k;\ \underline{x}_{\alpha-k+1}^k\},\ \alpha \ge k \tag{9}$$

where for each $\alpha \ge k$, $\hat{G}_\alpha^k = \hat{G}_\alpha^k(\underline{x}_\alpha)$ is a E^{m^k}-valued estimator of G_α^k. When $\hat{G}_\alpha^k$ is a function of $\underline{x}_{\alpha-k}$ only we refer to $\underline{\varphi}$ of (9) as a delete bootstrap rule.

Since the Γ_k construct is itself a finite state, finite action decision problem, many of Van Ryzin's (1966b) results apply to the analysis of extended rules (9). Consequently, notations, a preliminary lemma and the main theorem to follow are patterned after his work on the $k = 1$ case. However, we first treat some of the problems related to estimation of higher order empirical distributions G_α^k.

2. Estimation of the Empiric G_α^k.

For general results on the estimation of finite mixtures see Hannan (1957b), Teicher (1963), Robbins (1964, §7) and Van Ryzin (1966a, §3). Here we discuss the estimability of mixtures of the special finite class of product measures $\mathcal{P}^k$ and the structure of the dual basis estimators.

The class $\mathcal{P}^k$ is estimable if there exists a function $\underline{h} = (h_{1,\ldots,1},\ldots,h_{m,\ldots,m})$ on $\mathfrak{X}^k$ with components in $L_1(\mu^k)$ such that $E_\omega \underline{h} = \omega$ for all mixtures P_ω of $\mathcal{P}^k$ and $\underline{h}$ is said to be an unbiased estimator of $\mathcal{P}^k$. (E_ω denotes expectation with respect to the mixture P_ω.) Such functions provide kernels for unbiased estimators of G_α^k since

$$(10) \qquad E_{\underline{\theta}^k} h_{\underline{i}^k} = \delta(\underline{\theta}^k, \underline{i}^k) \quad \text{for all } \underline{\theta}^k, \underline{i}^k \in \Theta^k$$

where δ is the Kronecker delta function. From Van Ryzin (1966a, §3), $\mathcal{P}^k$ is estimable if and only if $\mathcal{P}^k$ is identifiable, if and only if $\mathcal{F}^k \equiv \{f_{\underline{i}^k} | \underline{i}^k \in \Theta^k\}$ is linearly independent in $L_1(\mu^k)$.

Remark 1. $\mathcal{F} = \{f_1,\ldots,f_m\}$ linearly independent in $L_1(\mu)$ implies $\mathcal{F}^k$ linearly independent in $L_1(\mu^k)$.

Proof: Suppose

$$(11) \qquad \sum_{\underline{i}^k} a_{\underline{i}^k} f_{\underline{i}^k} = 0 \quad \text{a.e. } \mu^k .$$

Almost every $\underline{x}^{k-1}$-section must be 0 a.e. μ, so that from $f_{\underline{i}^k} = f_{\underline{i}^{k-1}} f_{i_k}$ and the linear independence of $\mathcal{F}$ it follows that

$$(12) \qquad \sum_{\underline{i}^{k-1}} a_{\underline{i}^{k-1},i} f_{\underline{i}^{k-1}} = 0 \quad \text{a.e. } \mu^{k-1}, \ i \in \Theta .$$

By induction on k it follows that $a_{\underline{i}^k} = 0$ for all $\underline{i}^k \in \Theta^k$.

If $\underline{h}$ satisfies (10), then

$$\bar{h}_\alpha(\underline{x}_\alpha) \equiv (\alpha - k + 1)^{-1} \sum_{i=1}^{\alpha-k+1} \underline{h}(\underline{x}_i^k) \tag{13}$$

is an unbiased estimator for G_α^k; and if its components are in $L_2(\mu^k)$, then it is consistent. Henceforth, we take $\mathcal{F}$ to be linearly independent which assures the existence of $\underline{h}$ satisfying (10). From Robbins (1964, §7) and Van Ryzin (1966a, Theorem 1) if $\underline{h}$ satisfies (10) with components in $L_2(\mu^k)$, then

$$h_{\underline{i}^k} = f^*_{\underline{i}^k} + g_{\underline{i}^k} \tag{14}$$

where the $f^*_{\underline{i}^k}$ form the (unique) dual basis of S^k, the subspace of $L_2(\mu^k)$ spanned by $\mathcal{F}^k$, and $g_{\underline{i}^k} \perp S^k$ for all $\underline{i}^k \in \Theta^k$.

<u>Remark</u> $\underline{2}$. $f^*_{\underline{i}^k} = f^*_{i_1} f^*_{i_2} \ldots f^*_{i_k}$ for all $\underline{i}^k$ where $\{f^*_1, f^*_2, \ldots, f^*_m\}$ is the (unique) dual basis of S, the subspace of $L_2(\mu)$ spanned by $\mathcal{F}$.

<u>Proof</u>: The dual basis of S^k is the (unique) subset of m^k element from the span of $\mathcal{F}^k$ with elements satisfying (10). The products $f^*_{i_1} f^*_{i_2} \ldots f^*_{i_k}$ are in the span of $\mathcal{F}^k$ since $f^*_1, f^*_2, \ldots, f^*_m$ are in the span of $\mathcal{F}$ and satisfy (10).

From any unbiased estimator $h = (h_1, \ldots, h_m)$ of θ, one obtains an unbiased estimator $\underline{h} = (h_{11\ldots1}, \ldots, h_{mm\ldots m})$ of θ^k by taking

$$h_{\underline{i}^k}(\underline{x}^k) = h_{i_1}(x_1) h_{i_2}(x_2) \ldots h_{i_k}(x_k) \quad \text{for all } \underline{i}^k \in \Theta^k. \tag{15}$$

We call such an estimator a product estimator. Our theorem concerns an extended compound procedure based on an unbiased bounded product estimator of θ^k and Remark 2 shows that such an estimator is provided by $\underline{f}^* \equiv (f_1^* f_1^* \ldots f_1^*, \ldots, f_m^* f_m^* \ldots f_m^*)$. (In general, the class of unbiased product estimators of θ^k does not exhaust the class of all unbiased estimators of θ^k.)

The following remark can be applied to determine the rank of the covariance matrix of a product estimator.

<u>Remark 3</u>. Let $h = (h_1, \ldots, h_m)$ be unbiased for θ with components in $L_2(\mu)$ and let $V(i)$ denote its covariance matrix under P_i, $i \in \Theta$. Let $\underline{h}$ denote the product estimator (15) and let $V(\underline{i}^k)$ denote its covariance matrix under $P_{\underline{i}^k}$, $\underline{i}^k \in \Theta^k$. Then Rank $V(i) = m$ for all $i \in \Theta$ implies Rank $V(\underline{i}^k) = m^k$ for all $\underline{i}^k \in \Theta^k$.

<u>Proof</u>: We use the fact that the rank of the covariance matrix is the dimension of the span of the centered components in L_2. Let $\underline{i}^k \in \Theta^k$ and

(16) $$\sum_{\underline{j}^k} a_{\underline{j}^k}(h_{\underline{j}^k} - E_{\underline{i}^k} h_{\underline{j}^k}) = 0 \quad \text{a.e. } P_{\underline{i}^k}.$$

Taking $E_{\underline{i}^{k-1}}$ expectation gives

(17) $$\sum_{\underline{j}^k} a_{\underline{j}^k}(E_{\underline{i}^{k-1}} h_{\underline{j}^{k-1}})(h_{j_k} - E_{i_k} h_{j_k}) = 0 \quad \text{a.e. } P_{i^k}.$$

Since $E_{\underline{i}^{k-1}} h_{\underline{j}^{k-1}} = \delta(\underline{i}^{k-1}, \underline{j}^{k-1})$, it follows that

(18) $$\sum_{j=1}^{m} a_{\underline{i}^{k-1},j}(h_j - E_{i_k} h_j) = 0 \quad \text{a.e. } P_{i_k}.$$

Since $V(i_k)$ has rank m, (18) implies $a_{\underline{i}^{k-1},j} = 0$, $j \in \Theta$.

3. Main Result.

We will now state a lemma which is used to obtain rates for our decision procedure. Let $\underline{X} \equiv (X_1, X_2, \cdots)$ be a sequence of k-dependent random variables with means zero and finite variances. (Here k-dependent means that the variables $(X_1, \ldots, X_r)$ and $(X_s, \ldots, X_t)$ are independent for all $1 \le r < s \le t$ and $s - r \ge k$.) We will use the notation

$$S_n = \sum_{j=1}^{n} X_j, \; B_n = E(S_n^2), \; F_n(x) = \Pr[S_n \le x\sqrt{B_n}],$$
$$\Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-t^2/2} dt .$$

As a corollary to the proof of Theorem 1 of Egorov (1970) it follows that if there exist constants a_k, $b_k > 0$ independent of n such that

$$(19) \qquad |X_n| \le a_k, \quad n \ge 1$$

$$(20) \qquad B_n \ge b_k n, \quad n \ge 1$$

then there exists a constant $c_k > 0$ independent of n and the distribution of $\underline{X}$ such that

$$(21) \qquad \sup_x |F_n(x) - \Phi(x)| \le c_k (b_k n)^{-1/5}, \quad n \ge 1 .$$

From this result, Lemma 1 easily follows.

Lemma 1. Using the notation and conditions (19) and (20) from above,

$$\Pr[d \le S_n \le d + a] \le (2\pi\, b_k n)^{-\frac{1}{2}} a + 2c_k (b_k n)^{-1/5}, \; n \ge 1$$

for all real d and $a \ge 0$.

In this section we consider the extended compound rule $\underset{\sim}{\varphi}$ of (9) with $\hat{G}^k_\alpha = \bar{h}_{\alpha-k}$, $\alpha \geq k$ where

$$(22) \qquad \bar{h}_\alpha(\underline{x}_\alpha) = \begin{cases} (\alpha - k + 1)^{-1} \sum_{i=1}^{\alpha-k+1} \underline{h}(\underline{x}_i^k) & \alpha \geq k \\ \underline{0} & \alpha < k \end{cases}$$

and $\underline{h}$ is defined by (15). We take $h = (h_1, \ldots, h_m)$ to be an unbiased estimator of Θ with

$$(23) \qquad \max_{1 \leq j \leq m} |h_j| \leq H < \infty \text{ a.e. } \mu .$$

For convenience in establishing the asymptotic result (2) for the extended delete bootstrap procedure

$$(24) \qquad \underset{\sim}{\varphi}_\alpha(\underline{x}_\alpha) = \varphi^k\{\bar{h}_{\alpha-k}(\underline{x}_{\alpha-k}); \underline{x}^k_{\alpha-k+1}\}, \ \alpha \geq k;$$

we consider the average risk over the decisions concerning $\theta_k, \theta_{k+1}, \ldots, \theta_N$, i.e.,

$$(25) \qquad \underset{\sim}{R}^k_N(\underline{\theta}, \underset{\sim}{\varphi}) = (N - k + 1)^{-1} \sum_{\alpha=k}^{N} R_\alpha(\underline{\theta}, \underset{\sim}{\varphi}) .$$

With $\underset{\sim}{\varphi}_\alpha$ of (24), the independence of $\bar{h}_{\alpha-k}$ and $\underline{x}^k_{\alpha-k+1}$ implies

$$(26) \qquad R_\alpha(\underline{\theta}, \underset{\sim}{\varphi}) = \underline{E}\ R^k(\underline{\theta}^k_{\alpha-k+1}, \varphi^k\{\bar{h}_{\alpha-k}\}), \quad k \leq \alpha \leq N$$

where here and throughout this section $\underline{E}$ denotes expectation with respect to $\underline{P} = P_{\theta_1} \times P_{\theta_2} \times \ldots \times P_{\theta_N}$ measure on $(x_1, x_2, \ldots, x_N)$. Using the notation of Van Ryzin (1966b) we let $\epsilon_{\underline{i}^k}$ denote the E^{m^k}-vector of all 0's except 1 in position $\underline{i}^k$ and let $\rho_{\underline{i}^k}(\varphi) = R^k(\underline{i}^k, \varphi)$, $\underline{i}^k \in \Theta^k$. Since $\underline{x}^k_{\alpha-k+1}$ is independent of $\bar{h}_{\alpha-k}$ and $\underline{h}$ is unbiased for Θ^k we have

$$(27)\qquad R_\alpha(\underline{\theta},\underline{\varphi}) = \underline{E}(\varepsilon_{\underline{\theta}^k_{\alpha-k+1}}, \underline{\rho}(\varphi^k\{\bar{h}_{\alpha-k}\}))_k$$

$$= \underline{E}(\underline{h}(\underline{x}^k_{\alpha-k+1}), \underline{\rho}(\varphi^k\{\bar{h}_{\alpha-k}\}))_k,\ \alpha \geq k\ .$$

Using (27) in (25) it follows that

$$(28)\qquad \underline{R}^k_N(\underline{\theta},\underline{\varphi}) = (N - k + 1)^{-1}\sum_{\alpha=k}^{N} \underline{E}(\underline{h}(\underline{x}^k_{\alpha-k+1}), \underline{\rho}(\varphi^k\{\bar{h}_{\alpha-k}\}) - \underline{\rho}(\varphi^k\{\bar{h}_\alpha\}))_k$$

$$+ (N - k + 1)^{-1}\sum_{\alpha=k}^{N} \underline{E}(\underline{h}(\underline{x}^k_{\alpha-k+1}), \underline{\rho}(\varphi^k\{\bar{h}_\alpha\}))_k\ .$$

The left inequality of Lemma 3.2 and equality (3.4) of Van Ryzin (1966b) applied to the Γ_k decision problem, together with (28) gives

$$(29)\qquad \underline{R}^k_N(\underline{\theta},\underline{\varphi}) \leq A_N(\underline{\theta}) + \underline{E}\ R^k(\bar{h}_N)$$

where

$$(30)\quad A_N(\underline{\theta}) = (N - k+1)^{-1}\sum_{\alpha=k}^{N}\sum_{j<j'} \underline{E}\int|(\underline{h}(\underline{x}^k_{\alpha-k+1}), L^j_k\underline{f}(\underline{y}^k) - L^{j'}_k\underline{f}(\underline{y}^k))_k|\cdot$$

$$\{\varphi^k_j(\bar{h}_{\alpha-k};\ \underline{y}^k)\varphi^k_{j'}(\bar{h}_\alpha;\ \underline{y}^k) + \varphi^k_j(\bar{h}_\alpha;\ \underline{y}^k)\varphi^k_{j'}(\bar{h}_{\alpha-k};\ \underline{y}^k)\}d\mu^k(\underline{y}^k)$$

and $\underline{y}^k = (y_1,\ldots,y_k)$. By the unbiasedness of $\bar{h}_N$ for G^k_N

$$\underline{E}\{R^k(\bar{h}_N) - R^k(G^k_N)\} \leq \underline{E}(\bar{h}_N - G^k_N, \underline{\rho}(\varphi^k\{G^k_N\}))_k = 0$$

so that by (29)

$$(31)\qquad \underline{R}^k_N(\underline{\theta},\underline{\varphi}) - R^k(G^k_N) \leq A_N(\underline{\theta})\ .$$

We now proceed to show that if the kernel $h = (h_1,\ldots,h_m)$ has full rank covariance matrix under P_i, $i \in \Theta$, then $A_N(\underline{\theta}) = O(N^{-1/5})$ uniformly in $\underline{\theta}$. Thus we establish (2) with rate for the procedure (24).

Fixing $\underline{y}^k$, j and j' and letting $\underline{u}$ denote the $m^k \times 1$ matrix $L_k^j \underline{f}(\underline{y}^k) - L_k^{j'} \underline{f}(\underline{y}^k)$, we see from definition (8) that the factor in curly brackets in (30) is bounded by

$$(32) \qquad [-\sum_{i=\alpha-2k+2}^{\alpha-k+1} (\underline{h}(\underline{x}_i^k), \underline{u})_k < S \le 0] + [0 < S \le -\sum_{i=\alpha-2k+2}^{\alpha-k+1} (\underline{h}(\underline{x}_i^k), \underline{u})_k]$$

where the square brackets denote indicator functions and

$$(33) \qquad S = \sum_{i=1}^{\alpha-2k+1} (\underline{h}(\underline{x}_i^k), \underline{u})_k .$$

A calculation shows that

$$(34) \qquad (\underline{h}(\underline{x}_i^k), \underline{u})_k = Z_{i,1} Z_{i+1,2} \cdots Z_{i+k-1,k}$$

where

$$(35) \qquad Z_{\beta,\gamma} = \begin{cases} (h(x_\beta), f(y_\gamma)) & , \quad \gamma = 1,\dots,k-1 \\ (h(x_\beta), u(y_\gamma)) & , \quad \gamma = k , \end{cases}$$

$f(y_\gamma) = (f_1(y_\gamma),\dots,f_m(y_\gamma))$ and $u(y_\gamma) = L^j f(y_\gamma) - L^{j'} f(y_\gamma)$. With $\|\ \|$ denoting the Euclidean norm in E^m-space, (34), (35) and the Schwarz inequality for E^m applied to the $Z_{\beta,\gamma}$, it follows that

$$(36) \qquad |(\underline{h}(\underline{x}_i^k), \underline{u})_k| \le m^{k/2} H^k \|u(y_k)\| \prod_{j=1}^{k-1} \|f(y_j)\| .$$

In order to apply Lemma 1 to the sum S of k-dependent variables we need to investigate the variance of S.

Lemma 2. Suppose h has full rank covariance matrix V(i) under P_i, $i \in \Theta$ and let $\lambda^2 = \min_{1\le i\le m} \lambda_i^2$ where λ_i^2 is the minimum eigenvalue of V(i), $1 \le i \le m$, (necessarily positive since V(i) is positive definite). Then for all $\underline{\theta}$,

$$\mathrm{Var}_{\underline{\theta}}(S) \geq \left[\frac{\alpha - 2k+1}{k}\right]\lambda^{2k}\|u(y_k)\|^2 \prod_{j=1}^{k-1} \|f(y_j)\|^2 \tag{37}$$

where $[x]$ denotes the integer part of x.

Proof: Letting $r = \left[\frac{\alpha - 2k+1}{k}\right]$, we write

$$S = \sum_{i=1}^{\alpha-2k+1} Z_{i,1}Z_{i+1,2} \cdots Z_{i+k-1,k} = \sum_{j=1}^{r} S(j) + \left(S - \sum_{j=1}^{r} S(j)\right) \tag{38}$$

where

$$S(j) = \sum_{i=jk-k+1}^{jk} Z_{i,1}Z_{i+1,2} \cdots Z_{i+k-1,k}, \quad j = 1,\ldots,r. \tag{39}$$

By letting $\tilde{}$ denote conditioning on all x_i except $x_k, x_{2k}, \ldots, x_{rk}$ we have

$$\tilde{\mathrm{Var}}(S) = \sum_{j=1}^{r} \tilde{\mathrm{Var}}(S(j)) .$$

Defining

$$z_j = (Z_{jk-k+1,1}Z_{jk-k+2,2}\cdots Z_{jk-1,k-1}, Z_{jk-k+2,1}\cdots Z_{jk-1,k-2}Z_{jk+1,k},$$

$$\ldots, Z_{jk+1,2}Z_{jk+2,3} \cdots Z_{jk+k-1,k})_{1\times k}$$

$$A = \begin{pmatrix} u(y_k) \\ f(y_{k-1}) \\ \vdots \\ f(y_1) \end{pmatrix}_{k\times m}$$

and $\delta_j = z_j A$, it follows that

$$\tilde{\mathrm{Var}}(S(j)) = \mathrm{Var}(\delta_j, h(x_{jk})) \geq \lambda^2\|\delta_j\|^2 .$$

Hence,

(40) $$\mathrm{Var}_{\underline{\theta}}(S) \geq \lambda^2 \sum_{j=1}^{r} \underline{E}\|\delta_j\|^2 .$$

Letting subscript $jk+k-1$ on E and Var denote conditioning on all x_i except x_{jk+k-1} and $\delta_{j,i}$ be the $i^{\underline{th}}$ component of δ_j, we have for $j = 1,2,\ldots,r$

(41)
$$\begin{aligned}
\underline{E}\|\delta_j\|^2 &= \underline{E}\, E_{jk+k-1}\|\delta_j\|^2 \\
&= \underline{E} \sum_{i=1}^{m} E_{jk+k-1}(\delta_{j,i}^2) \\
&\geq \underline{E} \sum_{i=1}^{m} \mathrm{Var}_{jk+k-1}(\delta_{j,i}) \\
&= \underline{E} \sum_{i=1}^{m} f_i^2(y_1) Z_{jk+1,2}^2 \cdots Z_{jk+k-2,k-1}^2\, \mathrm{Var}(Z_{jk+k-1,k}) .
\end{aligned}$$

The right hand side of (41) is bounded below by

(42)
$$\begin{aligned}
&\underline{E}(\lambda^2\|u(y_k)\|^2\|f(y_1)\|^2 Z_{jk+1,2}^2 \cdots Z_{jk+k-2,k-1}^2) \\
&= \lambda^2\|u(y_k)\|^2\|f(y_1)\|^2 \underline{E}(Z_{jk+1,2}^2)\cdots\underline{E}(Z_{jk+k-2,k-1}^2) \\
&\geq \lambda^2\|u(y_k)\|^2\|f(y_1)\|^2 \mathrm{Var}(Z_{jk+1,2})\cdots\mathrm{Var}(Z_{jk+k-2,k-1}) \\
&\geq \lambda^{2(k-1)}\|u(y_k)\|^2 \prod_{i=1}^{k-1}\|f(y_i)\|^2
\end{aligned}$$

where use is made of

$$\mathrm{Var}(Z_{\beta,k}) = \mathrm{Var}(u(y_k),\, h(x_\beta)) \geq \lambda^2\|u(y_k)\|^2$$

and

$$\mathrm{Var}(Z_{\beta,\gamma}) = \mathrm{Var}(f(y_\gamma),\, h(x_\beta)) \geq \lambda^2\|f(y_\gamma)\|^2 \quad \text{for} \quad \gamma = 2,\ldots,k-1 .$$

Thus, (40) - (42) imply (37).

By (36) the bound (32) is seen to be zero if $\|u(y_k)\| = 0$ or $\|f(y_j)\| = 0$ for any $j = 1,2,\ldots,k-1$. Otherwise, Lemma 2 implies

$$\text{(43)} \qquad \mathrm{Var}_{\underline{\theta}}(S) \geq b_k(\alpha - 2k+1), \quad \alpha \geq 3k$$

where

$$\text{(44)} \qquad b_k = \frac{\lambda^{2k}\|u(y_k)\|^2 \prod_{i=1}^{k-1} \|f(y_i)\|^2}{k(k+1)} > 0 .$$

Further, using (3) and (36) and defining $D = \max_{i,j,s,t} |L(i,j) - L(s,t)|$, we have

$$|(\underline{h}(\underline{x}_i^k), \underline{u})_k| \leq a_k$$

where

$$a_k = m^k H^k D\ K^k .$$

Hence, conditions (19) and (20) of Lemma 1 are satisfied for $S_n = S$.

Using (32), Lemma 1, (36) and (3), it follows that for $\alpha \geq 3k$ the α, j, j' summand of (30) is bounded by

$$\text{(45)} \qquad B_1(\alpha - 2k+1)^{-\frac{1}{2}} + B_2(\alpha - 2k+1)^{-1/5}$$

where

$$B_1 = 2(2\pi)^{-\frac{1}{2}} D[k^3(k+1)]^{\frac{1}{2}}\, m^{3k/2} H^{2k} \lambda^{-k} [\mu(\mathfrak{X})]^k\, K^k$$

$$B_2 = 4c_k\, m^{4k/5} H^k \lambda^{-2k/5} D^{3/5} [k(k+1)]^{1/5} [\mu(\mathfrak{X})]^k\, K^{3k/5} .$$

For $k \leq \alpha \leq 3k$, we bound the summands of (30) by $B_3 = a_k[\mu(\mathfrak{X})]^k$ which together with (45) gives

$$(46)\quad A_N(\underset{\sim}{\theta}) \le (N-k+1)^{-1}\binom{n}{2}\{(2k+1)B_3 + B_1 \sum_{\alpha=3k+1}^{N} (\alpha-2k+1)^{-\frac{1}{2}}$$

$$+ B_2 \sum_{\alpha=3k+1}^{N} (\alpha-2k+1)^{-1/5}\} .$$

Noting

$$\sum_{\alpha=3k+1}^{N} (\alpha - 2k+1)^{p} \le \int_1^N x^p d_x \le (p+1)^{-1} N^{p+1}$$

for $-1 < p < 0$, we can state

$$(47)\qquad A_N(\underset{\sim}{\theta}) \le C_1 N^{-1} + C_2 N^{-\frac{1}{2}} + C_3 N^{-1/5}$$

where $C_1 = k(2k+1)\binom{n}{2}B_3$, $C_2 = 2\binom{n}{2}kB_1$ and $C_3 = 5/4\binom{n}{2}kB_2$.

The above analysis establishes the following theorem.

<u>Theorem</u>. If the bounded unbiased estimator $h = (h_1,\ldots,h_m)$ is such that $V(i)$, the covariance matrix of h under P_i, is of full rank for all $i \in \Theta$ and $\underset{\sim}{\varphi}$ is the sequence procedure (24) then there exists a positive constant C_k independent of $\underset{\sim}{\theta}$ such that

$$\underset{\sim}{R}_N^k(\underset{\sim}{\theta},\underset{\sim}{\varphi}) - R^k(G_N^k) \le C_k\, N^{-1/5} .$$

CHAPTER 2

COMPUTATIONS

1. Introduction.

The target for k-extended procedures is $R^k(G_N^k)$. We showed in Chapter 1 for the compound problem with $m \times n$ component that this standard is achieved asymptotically by certain extended sequence compound rules. Gilliland and Hannan (1969, Corollary 1) showed that $R^{k+1}(G_N^{k+1})$ is asymptotically more stringent than $R^k(G_N^k)$ so that these extended rules will have asymptotically lower risk than the unextended sequence and set procedures constructed by Van Ryzin (1966a,b) for the same component.

In this chapter we use Robbins' original component where $m = n = 2$, $L(1,1) = L(2,2) = 0$, $L(1,2) = L(2,1) = 1$, $P_1 = N(-1,1)$ and $P_2 = N(1,1)$. We calculate R^2 and compare it with R^1. Furthermore, we compute the compound risks for four unextended and four $k = 2$ extended sequence compound rules for various N and $\underline{\theta}$ to determine possible finite N advantages of extended procedures. Finally, for a selected unextended rule and a selected extended rule we compute the Bayes compound risks with respect to various distributions on $\underline{\theta}$ for $N = 50, 200$ to study the risk behavior of the rules.

2. <u>Computation of</u> $R^2(G)$.

Consider the Γ_2 decision problem based on the Robbins' component. Let f_1 and f_2 be the usual normal densities of $N(-1,1)$ and $N(1,1)$ with respect to Lebesgue measure μ. (In Chapter I μ was chosen to be a finite measure only for convenience in establishing the asymptotic result.) Then the Bayes rule (8) versus $G = (p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2})$ is

$$\varphi_1^2\{G; (x_1,x_2)\} = 1 \quad \text{if and only if} \quad \Delta_1(x_1,x_2) \le \Delta_2(x_1,x_2)$$

where

$$\Delta_1(x_1,x_2) = f_2(x_2) \sum_{i=1}^{2} p_{i,2} f_i(x_1)$$

and

$$\Delta_2(x_1,x_2) = f_1(x_2) \sum_{i=1}^{2} p_{i,1} f_i(x_1) .$$

By defining

$$p(x) = \frac{\sum_{i=1}^{2} p_{i,2} f_i(x)}{\sum_{i=1}^{2} \sum_{j=1}^{2} p_{i,j} f_i(x)}$$

it follows that $\Delta_1(x_1,x_2) \le \Delta_2(x_1,x_2)$ if and only if $p(x_1)f_2(x_2) \le (1 - p(x_1))f_1(x_2)$ which holds if and only if $x_2 \le c(x_1)$ where $c(x_1)$ is the cutoff value of the component Bayes rule versus $1 - p(x)$, $p(x)$, namely $c(x) = \frac{1}{2}\log((1 - p(x))/p(x))$.

Hence

$$R^2(G) = \sum_{i=1}^{2} \sum_{j=1}^{2} p_{i,j} \int R(j,\ d(x)) f_i(x)\,dx \tag{48}$$

where

$$R(1,\ d(x)) = 1 - \Phi(c(x) + 1)$$

and

$$R(2,\ d(x)) = \Phi(c(x) - 1)\ .$$

We will compute $R^2(G)$ for various G using the trapezoidal rule for the integrations indicated in (48). Since the marginals of G_N^2 are equal in the limit and $R^2(G_N^2)$ is the standard for compound risk in the $k = 2$ extended compound decision problem, we will compute $R^2(G)$ for G which satisfy

(49) $$p_{1,2} + p_{2,2} = p_{2,1} + p_{2,2}\ .$$

Letting p denote the common marginal probability (49) and defining $\delta = p_{2,2}/p$ for $p > 0$, we obtain a convenient parameterization for G satisfying (49),

$$G = G(p,\delta) \equiv (1 - p(2-\delta),\ p(1-\delta),\ p(1-\delta),\ p\delta),$$

and we will, henceforth, abbreviate $R^2(G(p,\delta))$ to $R^2(p,\delta)$. Since each component of $G(p,\delta)$ is linear in δ, $\alpha G(p,\delta_1) + (1-\alpha)G(p,\delta_2) = G(p,\ \alpha\delta_1 + (1-\alpha)\delta_2)$ for $0 \le \alpha \le 1$, the concavity of R^2 implies $R^2(p,\delta)$ is concave in δ for fixed p. Furthermore, by Remark 1 of Gilliland and Hannan (1969), $R^2(p,\delta)$ as a function of δ is maximum at $\delta = p$; the maximum value being $R(p)$, the $k = 1$ envelope evaluated at the prior $1-p$, p on states 1 and 2 respectively.

The $k = 1$ envelope $R(p)$ is easily computed by hand and is given in Table 1 to 5 place accuracy.

Table 1 - Values of $R(p)$ for $p = .1(.1).5$

p	R(p)
.1	.07006
.2	.11207
.3	.13875
.4	.15378
.5	.15866

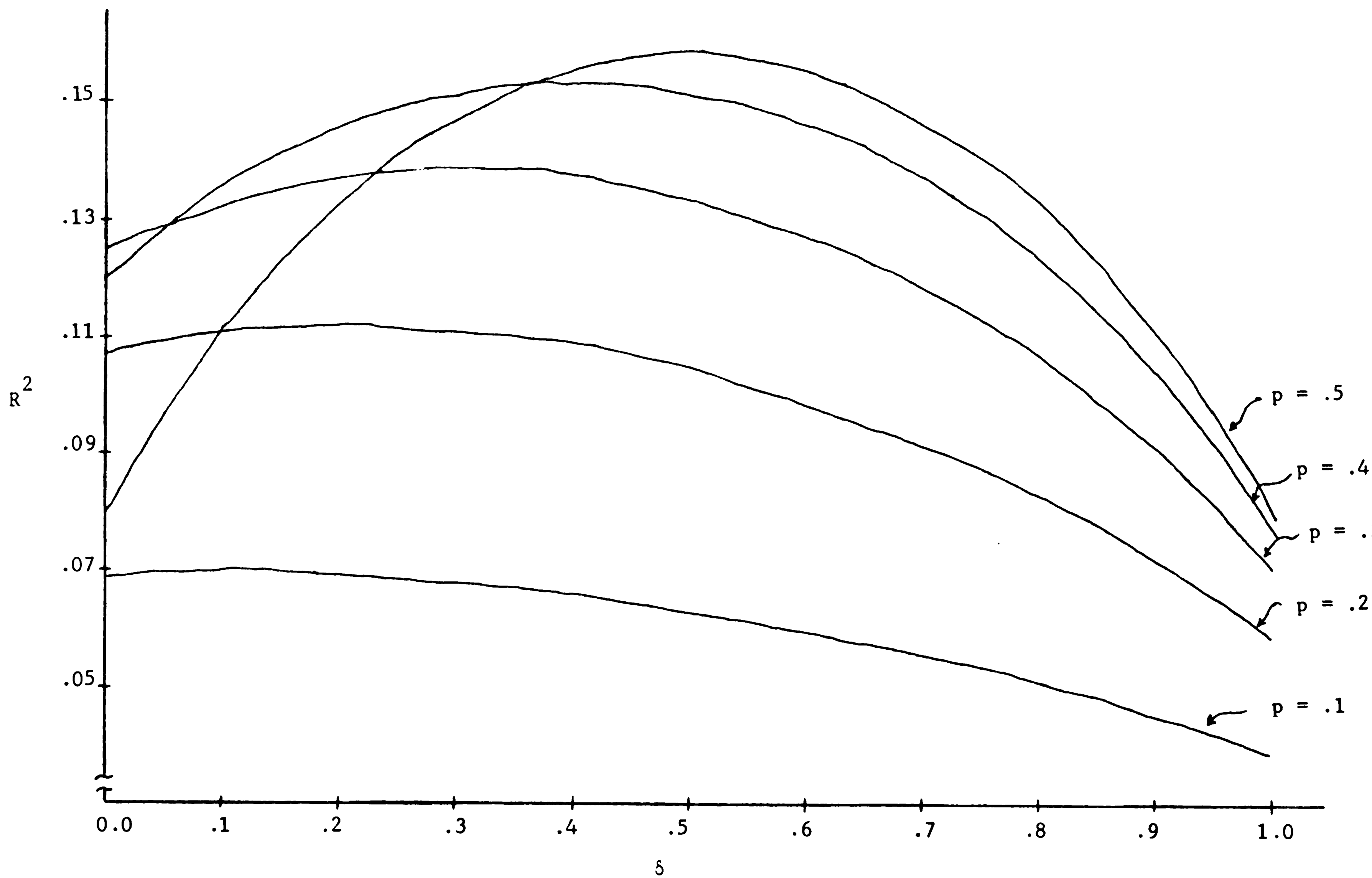
Values of $R^2(p,\delta)$ for $p = .1(.1).5$ and $\delta = 0(.05).1$ were computed on a CDC 6500 computer using (48) and the trapezoidal rule of numerical integration. These are given in Table 2 with maximum column values underlined. The grid used in these computations was sufficiently fine to guarantee that the error terms are bounded by .005. However, for most values, we feel these errors are less than .0001.

In order to more clearly understand the behavior of $R^2(p,\delta)$ as a function of δ for fixed p, we have plotted the values from Table 2 in Figure 1. The relatively flat nature of the curves corresponding to small p indicates that the extended rules will not be much better than the less complicated unextended rules at parameter sequences $\underline{\theta}$ with a small proportion of states 2. (By symmetry also at $\underline{\theta}$ with a large proportion of states 2.)

Table 2 - Values of $R^2(p,\delta)$ for $p = .1(.1).5$ and $\delta = 0(.05)1$

δ \ p	.1	.2	.3	.4	.5
0	.06939	.10757	.12474	.11999	.07866
.05	.06993	.10975	.12957	.12898	.09705
.10	<u>.07008</u>	.11111	.13312	.13608	.11167
.15	.06995	.11186	.13569	.14177	.12365
.20	.06958	<u>.11210</u>	.13744	.14623	.13348
.25	.06901	.11188	.13845	.14961	.14146
.30	.06824	.11123	<u>.13878</u>	.15196	.14780
.35	.06729	.11019	.13846	.15336	.15262
.40	.06617	.10876	.13751	<u>.15382</u>	.15601
.45	.06489	.10695	.13594	.15336	.15802
.50	.06344	.10477	.13375	.15199	<u>.15869</u>
.55	.06183	.10220	.13093	.14970	.15802
.60	.06005	.09925	.12747	.14647	.15601
.65	.05811	.09589	.12334	.14226	.15263
.70	.05599	.09212	.11852	.13702	.14781
.75	.05369	.08791	.11294	.13069	.14147
.80	.05121	.08322	.10655	.12316	.13348
.85	.04852	.07801	.09926	.11429	.12365
.90	.04562	.07220	.09093	.10387	.11168
.95	.04246	.06570	.08133	.09154	.09705
1.00	.03901	.05831	.07007	.07658	.07867

Figure 1 - Plot of $R^2(p,\delta)$ as a Function of δ for Fixed p



3. Simulation Study of the Risk Performance of Some Compound Rules.

In this section we compute and compare the compound risks of various unextended (k = 1) and extended (k = 2) compound procedures through the use of computer simulation. We consider two different unbiased kernels for estimating the empirical distributions in both the delete rule (24) and the corresponding non-delete version where $\underline{h}_{\alpha-k}$ is replaced by $\underline{h}_{\alpha}$. Thus, a total of 8 different rules are considered.

A bounded unbiased estimator is provided by the dual basis of $\{f_1, f_2\}$ in $L_2(\mu)$. In the example under consideration, it is $b(x) = (b_1(x), b_2(x))$ where

$$b_1(x) = (A^2 - B^2)^{-1}(Af_1(x) - Bf_2(x))$$

$$b_2(x) = (A^2 - B^2)^{-1}(Af_2(x) - Bf_1(x))$$

$A = \int f_1^2(x)\,dx = \int f_2^2(x)\,dx = (2\sqrt{\pi})^{-1}$ and $B = \int f_1(x) f_2(x)\,dx = A/e$. Estimator b has full rank covariance matrix under both $P_1 = N(-1,1)$ and $P_2 = N(1,1)$. The second unbiased estimator we consider is $r(x) = (r_1(x), r_2(x))$ where

$$r_1(x) = (1 - x)/2$$

$$r_2(x) = (1 + x)/2 \; ,$$

which is the kernel function used by Robbins (1951). Estimator r is unbounded and its covariance matrix has rank 1 under both P_1 and P_2.

The four unextended rules we consider are given by (9) with $k = 1$ and $\hat{G}_{\alpha}^{1}$ one of the following:

$$(\alpha - 1)^{-1}\Sigma_1^{\alpha-1} b(x_i) \quad , \quad \alpha \ge 2$$

$$(\alpha - 1)^{-1}\Sigma_1^{\alpha-1} r(x_i) \quad , \quad \alpha \ge 2$$

$$\alpha^{-1} \Sigma_1^{\alpha} b(x_i) \quad , \quad \alpha \ge 1$$

$$\alpha^{-1} \Sigma_1^{\alpha} r(x_i) \quad , \quad \alpha \ge 1 .$$

We refer to these compound rules as udb, udr, unb and unr respectively where, for example, unr denotes the unextended, non-delete rule with kernel r.

The four extended rules we consider are given by (9) with $k = 2$ and $\hat{G}_\alpha^2$ one of the following:

$$(\alpha - 3)^{-1}\Sigma_1^{\alpha-3} \underline{b}(\underline{x}_i^2) \quad , \quad \alpha \ge 4$$

$$(\alpha - 3)^{-1}\Sigma_1^{\alpha-3} \underline{r}(\underline{x}_i^2) \quad , \quad \alpha \ge 4$$

$$(\alpha - 1)^{-1}\Sigma_1^{\alpha-1} \underline{b}(\underline{x}_i^2) \quad , \quad \alpha \ge 2$$

$$(\alpha - 1)^{-1}\Sigma_1^{\alpha-1} \underline{r}(\underline{x}_i^2) \quad , \quad \alpha \ge 2$$

where $\underline{b}$ and $\underline{r}$ are the product estimators based on b and r. We refer to these rules as edb, edr, enb and enr respectively.

For our computations the compound losses for the rules edb and edr are calculated as average loss over the last $N - 3$ components; for udb, udr, enb and enr as average loss over the last $N - 1$ components and for unb and unr as average loss over all N components. In practice one might use the component minimax rule as the initial segment of the sequence compound rule. In defining the rule for the Theorem of Chapter 1 we took for convenience the estimator of G_α^k to be $\underline{0}$ (cf. (22)) in the initial segment forcing all initial decisions to be action 1 (cf. (8)).

The behavior of our eight procedures will be examined for $N = 20$, 50, 100 and 200 components and for two extreme types of parameter sequences.

Type I - Means of 1 occurring uniformly along the sequence such that $\delta = 0$. The proportion of these means will take values $p = .1(.1).5$.

Type II - All means of 1 occur in a group after means of -1. The proportion of means of 1 will take values $p = 0(.1).5$. In this type of sequence $\delta = 1 - (pN)^{-1}$.

Rayment (1971) has used these types of parameter sequences in an investigation of the compound risk behavior of the unextended delete sequence rule with $\hat{G}^1_\alpha = \bar{r}_\alpha$ truncated to the range of G^1_α.

One hundred simulations were made for each given $\underline{\theta}$ and N. All eight rules operated on the normal variables generated in a simulation. The estimated compound risk of a rule was obtained by averaging the one hundred compound losses. These averages along with error ranges of twice the standard deviation are given in the following tables. Envelope values from Tables 1 and 2 are given to indicate the unextended and extended asymptotic risk standards for each p and parameter sequence type.

Table 3 - Estimated Risks for $N = 20$ and Parameter Sequences of Type I

Rule \ p	.1	.2	.3	.4	.5
udb	.1116 $\pm$ .012	.1695 $\pm$ .014	.2132 $\pm$ .015	.2247 $\pm$ .018	.2279 $\pm$ .016
udr	.0974 $\pm$ .010	.1621 $\pm$ .013	.1911 $\pm$ .015	.2279 $\pm$ .017	.2216 $\pm$ .018
unb	.1160 $\pm$ .013	.1540 $\pm$ .014	.1745 $\pm$ .016	.1750 $\pm$ .017	.1700 $\pm$ .017
unr	.0960 $\pm$ .011	.1510 $\pm$ .014	.1585 $\pm$ .015	.1980 $\pm$ .017	.1860 $\pm$ .018
R(p)	.0701	.1121	.1387	.1538	.1587
edb	.1441 $\pm$ .012	.2265 $\pm$ .017	.2100 $\pm$ .020	.2112 $\pm$ .022	.1376 $\pm$ .022
edr	.1218 $\pm$ .011	.2006 $\pm$ .018	.1712 $\pm$ .019	.1980 $\pm$ .021	.1059 $\pm$ .018
enb	.1242 $\pm$ .013	.1658 $\pm$ .015	.1821 $\pm$ .018	.1700 $\pm$ .018	.1116 $\pm$ .019
enr	.1016 $\pm$ .012	.1558 $\pm$ .015	.1563 $\pm$ .016	.1726 $\pm$ .019	.1042 $\pm$ .018
$R^2(p,0)$	.0694	.1076	.1247	.1200	.0787

Table 4 - Estimated Risks for N = 20 and Parameter Sequences of Type II

Rule \ p	0	.1	.2	.3	.4	.5
udb	.0274 ± .016	.1163 ± .015	.1589 ± .015	.2158 ± .017	.2005 ± .021	.2063 ± .022
udr	.0126 ± .006	.1095 ± .012	.1495 ± .013	.1947 ± .016	.2216 ± .021	.2242 ± .021
unb	.0465 ± .016	.1220 ± .016	.1535 ± .014	.1990 ± .016	.1770 ± .017	.1875 ± .020
unr	.0330 ± .009	.1195 ± .014	.1370 ± .014	.1740 ± .016	.2020 ± .020	.1965 ± .017
R(p)	0	.0701	.1121	.1387	.1538	.1587
edb	.0565 ± .017	.1371 ± .016	.1953 ± .020	.2206 ± .020	.2429 ± .024	.2647 ± .025
edr	.0459 ± .011	.1312 ± .013	.1776 ± .015	.2124 ± .018	.2106 ± .023	.2159 ± .022
enb	.0626 ± .016	.1100 ± .016	.1363 + .019	.1600 ± .017	.1537 ± .019	.1684 ± .020
enr	.0563 ± .011	.1195 ± .015	.1179 ± .016	.1489 ± .016	.1642 ± .022	.1353 ± .017
$R^2(p,1)$	0	.0390	.0583	.0701	.0766	.0787

Table 5 - Estimated Risks for $N = 50$ and Parameter Sequences of Type I

Rule \ p	.1	.2	.3	.4	.5
udb	.0890 ± .006	.1390 ± .008	.1590 ± .010	.1867 ± .011	.1920 ± .011
udr	.0888 ± .005	.1447 ± .008	.1673 ± .009	.1798 ± .010	.2022 ± .011
unb	.0912 ± .007	.1322 ± .009	.1462 ± .010	.1690 ± .010	.1704 ± .010
unr	.0874 ± .007	.1414 ± .008	.1532 ± .009	.1644 ± .009	.1846 ± .010
R(p)	.0701	.1121	.1387	.1538	.1587
edb	.1062 ± .007	.1619 ± .010	.1498 ± .012	.1598 ± .012	.1153 ± .012
edr	.1028 ± .008	.1543 ± .008	.1430 ± .010	.1543 ± .010	.1045 ± .011
enb	.0970 ± .008	.1394 ± .009	.1424 ± .011	.1453 ± .012	.1030 ± .011
enr	.0976 ± .008	.1439 ± .009	.1433 ± .009	.1439 ± .011	.1065 ± .012
$R^2(p,0)$	.0694	.1076	.1247	.1200	.0787

Table 6 - Estimated Risks for $N = 50$ and Parameter Sequences of Type II

Rule \ p	0	.1	.2	.3	.4	.5
udb	.0139 ± .006	.0873 ± .005	.1318 ± .009	.1616 ± .011	.1849 ± .012	.1878 ± .011
udr	.0102 ± .005	.0835 ± .005	.1351 ± .009	.1598 ± .010	.1851 ± .013	.1933 ± .014
unb	.0302 ± .007	.0920 ± .007	.1302 ± .008	.1580 ± .010	.1768 ± .011	.1786 ± .010
unr	.0200 ± .006	.0846 ± .006	.1330 ± .009	.1530 ± .009	.1748 ± .012	.1884 ± .014
R(p)	0	.0701	.1121	.1387	.1538	.1587
edb	.0302 ± .006	.0930 ± .008	.1268 ± .011	.1383 ± .011	.1432 ± .012	.1523 ± .012
edr	.0234 ± .006	.0832 ± .007	.1062 ± .010	.1309 ± .011	.1323 ± .011	.1530 ± .014
enb	.0427 ± .007	.0861 ± .010	.1102 ± .010	.1184 ± .010	.1180 ± .011	.1218 ± .012
enr	.0278 ± .006	.0712 ± .007	.0931 ± .010	.1082 ± .010	.1114 ± .010	.1327 ± .013
$R^2(p,1)$	0	.0390	.0583	.0701	.0766	.0787

Table 7 - Estimated Risks for $N = 100$ and Parameter Sequences of Type I

Rule \ p	.1	.2	.3	.4	.5
udb	.0865 ± .005	.1286 ± .007	.1609 ± .006	.1721 ± .008	.1782 ± .007
udr	.0846 ± .004	.1320 ± .007	.1580 ± .006	.1714 ± .007	.1731 ± .007
unb	.0864 ± .005	.1253 ± .007	.1518 ± .007	.1622 ± .007	.1671 ± .007
unr	.0861 ± .005	.1277 ± .006	.1496 ± .006	.1646 ± .007	.1658 ± .007
R(p)	.0701	.1121	.1387	.1538	.1587
edb	.1018 ± .006	.1440 ± .008	.1454 ± .007	.1497 ± .008	.0971 ± .007
edr	.0996 ± .005	.1384 ± .008	.1432 ± .006	.1464 ± .008	.0965 ± .007
enb	.0962 ± .006	.1368 ± .008	.1392 ± .008	.1423 ± .002	.0931 ± .007
enr	.0955 ± .006	.1297 ± .007	.1409 ± .006	.1430 ± .008	.0960 ± .007
$R^2(p,0)$	.0694	.1076	.1247	.1200	.0787

Table 8 - Estimated Risks for $N = 100$ and Parameter Sequences of Type II

Rule \ p	0	.1	.2	.3	.4	.5
udb	.0071 $\pm$ .003	.0811 $\pm$ .004	.1296 $\pm$ .007	.1514 $\pm$ .007	.1719 $\pm$ .008	.1721 $\pm$ .010
udr	.0059 $\pm$ .003	.0811 $\pm$ .004	.1243 $\pm$ .006	.1498 $\pm$ .008	.1718 $\pm$ .008	.1660 $\pm$ .009
unb	.0135 $\pm$ .004	.0842 $\pm$ .005	.1316 $\pm$ .007	.1478 $\pm$ .006	.1694 $\pm$ .008	.1684 $\pm$ .010
unr	.0102 $\pm$ .003	.0820 $\pm$ .004	.1237 $\pm$ .006	.1482 $\pm$ .006	.1701 $\pm$ .008	.1638 $\pm$ .009
$R(p)$	0	.0701	.1121	.1387	.1538	.1587
edb	.0161 $\pm$ .004	.0797 $\pm$ .006	.1014 $\pm$ .007	.1041 $\pm$ .008	.1174 $\pm$ .009	.1208 $\pm$ .008
edr	.0117 $\pm$ .003	.0684 $\pm$ .006	.0909 $\pm$ .007	.0972 $\pm$ .006	.1179 $\pm$ .008	.1139 $\pm$ .009
enb	.0231 $\pm$ .005	.0730 $\pm$ .007	.0938 $\pm$ .007	.0930 $\pm$ .008	.1051 $\pm$ .008	.1100 $\pm$ .008
enr	.0170 $\pm$ .004	.0640 $\pm$ .006	.0843 $\pm$ .007	.0900 $\pm$ .006	.1081 $\pm$ .008	.1041 $\pm$ .008
$R^2(p,1)$	0	.0390	.0583	.0701	.0766	.0787

Table 9 - Estimated Risks for $N = 200$ and Parameter Sequences of Type I

Rule \ p	.1	.2	.3	.4	.5
udb	.0812 ± .003	.1199 ± .003	.1538 ± .004	.1636 ± .005	.1727 ± .005
udr	.0775 ± .002	.1259 ± .004	.1485 ± .005	.1624 ± .005	.1675 ± .006
unb	.0818 ± .003	.1182 ± .004	.1502 ± .004	.1595 ± .005	.1668 ± .005
unr	.0763 ± .003	.1237 ± .004	.1446 ± .005	.1585 ± .005	.1639 ± .006
R(p)	.0701	.1121	.1387	.1538	.1587
edb	.0903 ± .004	.1260 ± .004	.1409 ± .005	.1340 ± .005	.0968 ± .005
edr	.0839 ± .003	.1258 ± .005	.1346 ± .005	.1332 ± .005	.0898 ± .005
enb	.0882 ± .004	.1215 ± .004	.1388 ± .005	.1295 ± .005	.0948 ± .005
enr	.0816 ± .003	.1232 ± .005	.1342 ± .005	.1313 ± .006	.0900 ± .006
$R^2(p,0)$	.0694	.1076	.1247	.1200	.0787

Table 10 - Estimated Risks for N = 200 and Parameter Sequences of Type II

Rule \ p	0	.1	.2	.3	.4	.5
udb	.0060 ± .002	.0775 ± .003	.1231 ± .004	.1479 ± .005	.1656 ± .007	.1657 ± .007
udr	.0055 ± .002	.0768 ± .003	.1201 ± .005	.1448 ± .006	.1712 ± .006	.1664 ± .006
unb	.0097 ± .002	.0785 ± .003	.1219 ± .004	.1469 ± .005	.1648 ± .006	.1652 ± .007
unr	.0084 ± .002	.0773 ± .003	.1196 ± .005	.1440 ± .006	.1699 ± .006	.1655 ± .006
R(p)	0	.0701	.1121	.1387	.1538	.1587
edb	.0131 ± .002	.0620 ± .004	.0841 ± .005	.0957 ± .006	.1035 ± .005	.1008 ± .006
edr	.0088 ± .002	.0579 ± .004	.0817 ± .005	.0892 ± .006	.1001 ± .006	.0984 ± .007
enb	.0171 ± .003	.0579 ± .004	.0791 ± .005	.0892 ± .006	.0976 ± .005	.0951 ± .006
enr	.0112 ± .002	.0556 ± .004	.0781 ± .005	.0846 ± .006	.0968 ± .006	.0933 ± .006
$R^2(p,1)$	0	.0390	.0583	.0701	.0766	.0787

The following observations can be made from the above tables.

1) When $R(p)$ and $R^2(p,\delta)$ are nearly equal and $N > 20$, the unextended and extended procedures appear to have similar behavior. However, at points where $R^2(p,\delta)$ is considerably less than $R(p)$ and $N > 20$, the extended procedures are significantly better. The results for $N = 20$ are somewhat inconclusive except when the parameter sequence is of Type I and $p = .5$ the extended procedures are a great improvement.

2) The performance of the nondelete rules appears on the average to be better then the delete rules at 20 and 50 components. But for 100 and 200 components this advantage seems to disappear. At $p = 0$ for $N = 20, 50, 100$ and 200, the delete unextended rules have uniformly lower estimated risks.

3) Generally, the behavior of the rules based on the dual basis kernel is the same as the behavior of the rules based on Robbins' original kernel.

Observations 1 and 2 are cretainly consistent with the theory and intuition. When the extended envelope is significantly below the unextended envelope, one would expect the extended procedures to be better. Further, it is consistent that the advantage of nondeletion would become negligible as the number of components increases. The last two observations seem to indicate that Theorems 4.2 and 4.3 of Van Ryzin (1966b) may be generalized to the extended setting.

4. Performance of Compound Rules when the Parameter Sequence is a Strictly Stationary Process.

Gilliland and Hannan (1969, Theorem 3) show that if $\underline{\theta} = (\theta_1, \theta_2, \ldots)$ is a strictly stationary stochastic process then any asymptotic solution of the k-extended sequence compound decision problem $\underline{\varphi} = (\underline{\varphi}_1, \underline{\varphi}_2, \ldots)$ satisfies

$$\overline{\lim}_N \int R_N(\underline{\theta}, \underline{\varphi})\, dG(\underline{\theta}) \leq R^k(G_*^k)$$

where G denotes the measure on infinite sequences $\underline{\theta}$ and G_*^k denotes the marginal on $\underline{\theta}_i^k = (\theta_i, \theta_{i+1}, \ldots, \theta_{i+k-1})$, $i = 1, 2, \ldots$.

This theorem serves as the motivation for our next set of calculations. We modified the computer program used in the above computations so that the sequence of parameters is generated by a Markov process. The distribution of the initial parameter is

$$\Pr[\theta_1 = 2] = p$$

$$\Pr[\theta_1 = 1] = 1 - p$$

and the transition probabilities are

$$\Pr[\theta_{i+1} = 2 \mid \theta_i = 2] = \delta$$

$$\Pr[\theta_{i+1} = 1 \mid \theta_i = 2] = 1 - \delta$$

$$\Pr[\theta_{i+1} = 2 \mid \theta_i = 1] = p(1-\delta)/(1 - p)$$

$$\Pr[\theta_{i+1} = 1 \mid \theta_i = 1] = 1 - p(1-\delta)/(1 - p).$$

It is not difficult to show that this process is strictly stationary.

In our calculations we compared the Bayes performance of udb and edb for 50 and 200 components. One hundred simulations were made for each case and both rules operated on the same hundred samples. The estimated risks of the two rules were obtained by averaging the one hundred compound losses. These averages, along with error ranges of twice the standard deviation are given in the following tables.

From these tables we observe that for 50 components the un-extended procedure, udb, performs better at every (p,δ)-value while for 200 components the extended procedure, edb, is significantly better at many (p,δ)-values. Further, the estimated risks for edb indicate that its convergence to or below $R^2(p,\delta)$ is relatively slow.

Table 11 - Estimated Bayes Risk of udb for N = 50

δ \ p	.1	.2	.3	.4	.5
0	.0841 ± .009	.1384 ± .010	.1755 ± .011	.1963 ± .011	.1971 ± .012
.1	.0886 ± .007	.1400 ± .009	.1698 ± .011	.1865 ± .012	.2051 ± .010
.2	.0890 ± .008	.1425 ± .010	.1818 ± .011	.1894 ± .012	.1971 ± .012
.3	.0908 ± .009	.1425 ± .010	.1690 ± .010	.1879 ± .011	.1975 ± .010
.4	.0863 ± .011	.1329 ± .010	.1630 ± .011	.2010 ± .013	.1918 ± .010
.5	.0967 ± .010	.1235 ± .011	.1629 ± .010	.1861 ± .011	.1892 ± .010
.6	.0833 ± .011	.1331 ± .011	.1608 ± .011	.1869 ± .011	.1806 ± .011
.7	.0794 ± .013	.1329 ± .013	.1582 ± .011	.1771 ± .010	.1898 ± .011
.8	.0812 ± .014	.1096 ± .013	.1467 ± .012	.1602 ± .012	.1729 ± .011
.9	.0718 ± .015	.1090 ± .017	.1269 ± .015	.1606 ± .014	.1710 ± .012
1.0	.0129 ± .006	.0102 ± .004	.0096 ± .005	.0125 ± .004	.0157 ± .006

Table 12 - Estimated Bayes Risk of edb for N = 50

δ \ p	.1	.2	.3	.4	.5
0	.1066 ± .010	.1578 ± .011	.1766 ± .013	.1717 ± .011	.1136 ± .014
.1	.1038 ± .009	.1591 ± .012	.1951 ± .012	.1947 ± .012	.1566 ± .012
.2	.1081 ± .010	.1677 ± .010	.2049 ± .013	.2108 ± .013	.1968 ± .012
.3	.1170 ± .012	.1725 ± .012	.2074 ± .012	.2091 ± .012	.2060 ± .012
.4	.1130 ± .013	.1668 ± .011	.1985 ± .012	.2343 ± .011	.2136 ± .012
.5	.1164 ± .011	.1564 ± .012	.2036 ± .011	.2179 ± .013	.2264 ± .013
.6	.0974 ± .011	.1549 ± .011	.1940 ± .013	.2268 ± .012	.2294 ± .011
.7	.0996 ± .014	.1462 ± .013	.1802 ± .013	.2153 ± .011	.2219 ± .012
.8	.0872 ± .012	.1249 ± .014	.1553 ± .011	.1745 ± .013	.1962 ± .011
.9	.0760 ± .013	.1183 ± .017	.1355 ± .015	.1551 ± .014	.1734 ± .012
1.0	.0285 ± .008	.0268 ± .006	.0210 ± .006	.0347 ± .008	.0262 ± .007

Table 13 - Estimated Bayes Risk of udb for N = 200

ȯ p	.1	.2	.3	.4	.5
0	.0791 ± .004	.1202 ± .005	.1520 ± .005	.1626 ± .005	.1700 ± .005
.1	.0774 ± .004	.1225 ± .004	.1481 ± .005	.1623 ± .005	.1675 ± .005
.2	.0781 ± .004	.1219 ± .004	.1478 ± .005	.1650 ± .005	.1649 ± .005
.3	.0790 ± .004	.1200 ± .005	.1500 ± .005	.1618 ± .004	.1688 ± .005
.4	.0784 ± .004	.1195 ± .005	.1493 ± .005	.1668 ± .004	.1712 ± .005
.5	.0727 ± .005	.1223 ± .005	.1501 ± .005	.1642 ± .005	.1700 ± .005
.6	.0758 ± .005	.1259 ± .005	.1419 ± .005	.1617 ± .005	.1720 ± .006
.7	.0707 ± .005	.1245 ± .006	.1467 ± .006	.1661 ± .005	.1686 ± .005
.8	.0771 ± .007	.1209 ± .006	.1415 ± .006	.1614 ± .005	.1686 ± .006
.9	.0720 ± .009	.1079 ± .009	.1371 ± .008	.1598 ± .005	.1644 ± .006
1.0	.0056 ± .002	.0044 ± .001	.0057 ± .002	.0058 ± .002	.0058 ± .002

Table 14 - Estimated Bayes Risk of edb for N = 200

δ \ p	.1	.2	.3	.4	.5
0	.0857 ± .004	.1226 ± .005	.1481 ± .005	.1363 ± .005	.0932 ± .005
.1	.0880 ± .005	.1295 ± .004	.1522 ± .005	.1550 ± .005	.1299 ± .005
.2	.0880 ± .005	.1335 ± .005	.1560 ± .005	.1690 ± .005	.1540 ± .005
.3	.0882 ± .004	.1305 ± .005	.1624 ± .006	.1702 ± .005	.1681 ± .005
.4	.0872 ± .005	.1293 ± .005	.1626 ± .006	.1793 ± .005	.1771 ± .006
.5	.0787 ± .005	.1284 ± .006	.1597 ± .005	.1747 ± .006	.1814 ± .006
.6	.0744 ± .005	.1257 ± .006	.1441 ± .004	.1641 ± .005	.1820 ± .006
.7	.0730 ± .005	.1173 ± .006	.1441 ± .006	.1609 ± .005	.1716 ± .005
.8	.0685 ± .006	.1035 ± .006	.1276 ± .006	.1477 ± .005	.1575 ± .006
.9	.0597 ± .007	.0899 ± .007	.1055 ± .006	.1264 ± .005	.1400 ± .006
1.0	.0117 ± .003	.0107 ± .002	.0120 ± .002	.0118 ± .002	.0129 ± .003

CONCLUSIONS

In this thesis we have shown that the product estimators of G_N^k are natural and logical extensions of the estimators of G_N in the unextended problem. Further, using these estimators, we presented a sequence extended delete compound decision procedure which satisfies (2) with a rate of $N^{-1/5}$.

Our computer simulations for the featured example in Robbins (1951) showed that the $k = 2$ extended envelope is significantly below the simple envelope for many parameter sequences. For these sequences, the four extended rules investigated have estimated risks that are significantly better then those of the corresponding unextended procedures for as few as 20 repetitions of the component problem. The calculations indicated that the risks of extended rules converge relatively rapidly to $R^2(G_N^2)$. From a comparison of extensive tables of component by component loss not published here, one can conclude that for most parameter sequences the average number of errors made at component α has reached the envelope value for $\alpha \geq 50$. However, it takes some time to average out the large number of errors made in the first few components. Our last set of calculations showed that the extended rules perform quite effectively when the parameter sequence is being generated by a Markov process.

REFERENCES

REFERENCES

Egorov, V.A. (1970). Some limit theorems for m-dependent random variables. Litovsk Mat. Sb. **10** 51-59.

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

Gilliland, Dennis C., Hannan, James and Huang, J.S. (1974). Asymptotic solutions to the two state component compound decision problem, Bayes versus diffuse priors on proportions. RM-320, Dept. of Statist. and Prob., M.S.U.

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

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

Hannan, James F. (1957b). Unpublished lecture notes. M.S.U.

Johns, M.V., Jr. (1967). Two-action compound decision problems. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, **1** 463-478. University of California Press.

Rayment, P.R. (1971). The performance of established and modified compound decision rules. Biometrika **58** 183-194.

Robbins, Herbert (1951). Asymptotically subminimax solutions of compound statistical decision problems. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability 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.

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.

Teicher, Henry (1963). Identifiability of finite mixtures. Ann. Math. Statist. **34** 1265-1269.

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.

Yu, Benito (1971). Rates of convergence in empirical Bayes two-action and estimation problems and in extended sequence-compound estimation problems. RM-279, Dept. of Statist. and Prob., M.S.U.