This is to certify that the

dissertation entitled

Finite State k-Extended Set Compound
Decision Problem

presented by

Chitra Gunawardena

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

James Hannan
Major professor

Date 8/21/1989

FINITE STATE k–EXTENDED SET COMPOUND DECISION PROBLEM

By

Chitra Gunawardena

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

1989

ABSTRACT

FINITE STATE k–EXTENDED SET COMPOUND DECISION PROBLEM

By

Chitra Gunawardena

In compound decision theory the usual standard for evaluating compound decision rules is $R(G_N)$, where R is the Bayes envelope in the component problem and G_N is the empirical distribution of the component states $\underline{\theta} = (\theta_1,\ldots\ldots,\theta_N)$. As introduced by (Johns and) Swain (1965), a more stringent standard for evaluating compound rules is $R^k(G_N^k)$, where R^k is the Bayes envelope of a construct called Γ^k game by Gilliland and Hannan (1969) and G_N^k is the empirical distribution of the overlapping k–tuples $(\theta_1,\ldots,\theta_k)$, $(\theta_2,\ldots,\theta_{k+1})$, ... $(\theta_{N-k+2},\ldots,\theta_{N-1},\theta_N,\theta_1)$, $(\theta_{N-k+3},\ldots,\theta_N,\theta_1,\theta_2)$, ..., $(\theta_{N-2},\ldots,\theta_{k-1})$. The $k+1$ standard is more stringent than the k standard and $R^1(G_N^1) = R(G_N)$.

Ballard's thesis (1974) considered the sequence version of the finite state finite act compound decision problem with $R^k(G_N^k)$ as its risk standard. He exhibited procedures which play Γ^k Bayes against a delete estimate of G_α^k in the α^{th} component problem, $\forall\ 1 \leq \alpha \leq N$, and showed that, on the average risk scale, the excess compound risk over $R^k(G_N^k)$ for his procedures has rate $O(N^{-1/5})$. Ballard, Gilliland and Hannan (1974) improved the rate to $O(N^{-\frac{1}{2}})$.

We here consider the set version of the finite state compound decision problem with $R^k(G_N^k)$ as its risk standard and treat both delete and nondelete procedures which play Γ^k Bayes against corresponding estimates of G_N^k in each of the component problems. In both cases we show that, on the average risk scale, the excess compound risk over $R^k(G_N^k)$ for our procedures has rate $O(N^{-\frac{1}{2}})$, when the action space is finite. Similar, but weaker results are obtained in Section 2.4 when the action space is infinite.

In addition, we characterize extrema of the expected value of a function of a generalized Binomial random variable, under constant variance; an analogue to a work of Hoeffding (1956), under constant mean. We show that extrema are attained at points whose coordinates take on at most four different values, only two of which are distinct from 0 and 1.

To the memory of my Father

ACKNOWLEDGMENTS

I am truly grateful to my advisor Professor James Hannan for his continuous encouragement and guidance in the preparation of this dissertation. His patience and willingness to discuss any problem at any time are greatly appreciated.

I would also like to thank Professors Dennis C. Gilliland and R. V. Ramamoorthi for careful reading of the dissertation. Conversations with Professor Ramamoorthi concerning Chapter 3 were especially helpful.

I would like to thank Professors R. V. Ericson and H. Salehi for serving on my guidance committee.

I wish to thank my husband, K.L.D., and my daughter, Kalpanee, for their continued patience and support throughout.

Finally, I wish to thank the Department of Statistics and Probability for providing the financial support which made my graduate studies in Statistics possible.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION TO THE k–EXTENDED COMPOUND DECISION PROBLEM

This chapter presents the general k–extended compound decision problem. We begin with the introduction of some notations that will be used throughout Chapters 1 and 2. In Sections 1.1 and 1.2 we describe the compound decision problem with its usual standard (1.6) for evaluating compound procedures. In Section 1.4 we describe a more stringent standard (1.11) for evaluating compound procedures and with it we introduce the k–extended compound decision problem. In order to describe the concepts in Section 1.4 we devote Section 1.3 to present the Γ^k decision problem introduced by Gilliland and Hannan (1969).

Notations:

k and N will denote integers with $k \leq N$.

The square brackets will be used to denote the indicator function.

If f_i are functions defined on some sets A_i for $i = 1,2,\ldots,j$ then $\bigotimes_{i=1}^{j} f_i$ will denote the function; $x \in A_1 \times \ldots \times A_j \rightsquigarrow \prod_{i=1}^{j} f_i(x_i)$.

For a sequence $\underline{u}^{\infty} = (u_1,u_2,\ldots\ldots)$, $\underline{u}_i$ will denote $(u_1,u_2,\ldots,u_i)$; the subscript N will be abbreviated by omission. With indices arithmetic mod N, $\forall\ 1 \leq i,j \leq N$ $\underset{\sim}{u}_i$ will denote the k–tuple $(u_{i-k+1},\ldots\ldots,u_i)$ and ${}_i\underline{u}_j$ the (j–i)–tuple (mod N) $(u_{i+1},\ldots\ldots,u_j)$.

1.1 The Component Problem

The component problem has the structure of a usual statistical decision problem, which is composed of a parameter set Θ, indexing a family of probability measures $\{P_\theta : \theta \in \Theta\}$ over a σ–field $\mathcal{B}$ of a sample space $\mathcal{X}$, an action space $\mathcal{A}$, a loss function $L : \Theta \times \mathcal{A} \to [0,\infty)$, decision rules $\varphi : \mathcal{X} \to \mathcal{A}$ such that $L(\theta,\varphi)$ is measurable for each θ, with risk

$$R(\theta,\varphi) = E_\theta L(\theta,\varphi) \tag{1.1}$$

where E_θ denotes the expectation with respect to P_θ.

1.2 The Set Compound Problem

When N independent problems each having the same structure of the component problem described in Section 1.1 are considered simultaneously, the N–fold global problem is called a compound decision problem. The loss in the compound problem is taken to be the sum of the losses in the N decision problems.

Thus for each N, in the compound decision problem we have the parameter set Θ^N indexing the family of probability measures $\{ P_{\underline{\theta}} = \times_{i=1}^{N} P_{\theta_i} : \underline{\theta} \in \Theta^N \}$ over $(\mathcal{X}^N, \mathcal{B}^N)$, the action space $\mathcal{A}^N$, compound decision rules $\underline{\varphi} = (\varphi_1,\ldots,\varphi_N)$ where for each $1 \leq \alpha \leq N$ $\varphi_\alpha : \mathcal{X}^N \to \mathcal{A}$ is such that $L(\theta,\varphi_\alpha)$ is measurable for each θ with loss

$$L_N(\underline{\theta},\underline{\varphi}) = \sum_{\alpha=1}^{N} L(\theta_\alpha,\varphi_\alpha) , \tag{1.2}$$

α^{th} component risk

$$R_\alpha(\underline{\theta},\underline{\varphi}) = \int L(\theta_\alpha,\varphi_\alpha) \, dP_{\underline{\theta}} \tag{1.3}$$

and compound risk

$$\underline{R}(\underline{\theta},\underline{\varphi}) = \sum_{\alpha=1}^{N} R_\alpha(\underline{\theta},\underline{\varphi}) . \tag{1.4}$$

As standard in compound decision theory, we say that a compound decision rule $\underline{\varphi}$ is simple symmetric if $\varphi_\alpha(\underline{x}) = \varphi(x_\alpha)$ $\forall$ $\underline{x} \in \mathcal{X}^N$ and $\forall$ $1 \le \alpha \le N$, for some component decision rule φ. For a simple symmetric rule $\underline{\varphi}$

$$\underline{R}(\underline{\theta},\underline{\varphi}) = \sum_{\alpha=1}^{N} R(\theta_\alpha,\varphi) . \tag{1.5}$$

This is the same as the component problem Bayes risk of φ against the non-normalized empirical distribution G_N of $\theta_1,\ldots\ldots,\theta_N$. Thus, with R(G) denoting the Bayes risk versus G in the component problem

$$\underline{D}(\underline{\theta},\underline{\varphi}) = \underline{R}(\underline{\theta},\underline{\varphi}) - R(G_N) \tag{1.6}$$

can be considered as a standard for evaluating compound decision procedures. $\underline{D}(\underline{\theta},\underline{\varphi})$ is called the modified regret of the compound decision procedure $\underline{\varphi}$ at $\underline{\theta}$. We say that a rule $\underline{\varphi}$ is <u>asymptotically optimal</u> (a.o.) if

$$\sup_{\underline{\theta}} \underline{D}(\underline{\theta},\underline{\varphi}) = o(N). \tag{1.7}$$

1.3 Γ^k Decision Problem

The decision problem concerning the last component of a k–tuple with the decision based on independent observations on all k parameters is called a Γ^k decision problem by Gilliland and Hannan (1969). Specifically, for an integer $k \geq 1$, the Γ^k decision problem has states $\underset{\sim}{\theta}_k = (\theta_1,\theta_2,..,\theta_k) \in \Theta^k$ indexing possible probability measures $\{ P_{\underset{\sim}{\theta}_k} = \times_{i=1}^{k} P_{\theta_i} : \underset{\sim}{\theta}_k \in \Theta^k \}$ on $(\mathscr{X}^k,\mathscr{B}^k)$, action space $\mathscr{A}$, loss function $L^k(\underset{\sim}{\theta}_k,a) = L(\theta_k,a)$, decision rules $\varphi: \mathscr{X}^k \to \mathscr{A}$ such that $L^k(\underset{\sim}{\theta}_k,\varphi)$ is measurable for each θ_k with risk when state $\underset{\sim}{\theta}_k$ holds defined by

$$R^k(\underset{\sim}{\theta}_k,\varphi) = \int L(\theta_k,\varphi)\, dP_{\underset{\sim}{\theta}_k} . \tag{1.8}$$

For a prior G^k on Θ^k the Bayes risk of φ against G^k is

$$R^k(G^k,\varphi) = \int R^k(\underset{\sim}{\theta}_k,\varphi)\, dG^k(\underset{\sim}{\theta}_k) \tag{1.9}$$

and the Γ^k Bayes envelope evaluated at G^k is

$$R^k(G^k) = \inf_{\varphi} R^k(G^k,\varphi) . \tag{1.10}$$

The decision problem Γ^1 is the component game in the compound decision problem.

One of the important facts about the Γ^k decision problem is given by the Remark (1) of Gilliland and Hannan (1969); which states that if G_* is the marginal of G^k on any ordered subset of the coordinates of $(i_1,i_2,\ldots,i_j)$ with $i_j = k$ then $R^k(G^k) \leq R^j(G_*)$. If, in addition, G^k is the product of G_* and the marginal H_* on the other coordinates then $R^k(G^k) = R^j(G_*)$. If G^k is not the product of G_* and H_* then the difference between $R^k(G^k)$ and $R^j(G_*)$ could be substantial as was demonstrated by Ballard and Gilliland (1978).

1.4 k–Extended Set Compound Decision Problem

The k–extended version of the compound decision problem was first introduced by Johns (1967) and has a more stringent standard for the compound risk than $R(G_N)$. Gilliland and Hannan (1969) have given the most general treatment of these standards.

In order to preserve lower case letters for dummy variables under consideration, w.l.o.g. we will assume that the domain of the random observations $\underline{X}$ in the compound problem is its range space $\mathscr{X}^N$. Thus the X_i will viewed as the coordinate functions of $\underline{X}$.

To introduce the k–extended risk standards we consider a compound decision procedure $\underline{\varphi}$ of the form

$$\underline{\varphi}(\underline{X}) = (\varphi(\underset{\sim}{X}_1),\varphi(\underset{\sim}{X}_2),\ldots\ldots,\varphi(\underset{\sim}{X}_N))$$

for a fixed Γ^k decision rule φ. For such φ it follows from (1.3) and (1.4)

$$\underline{R}(\underline{\theta},\underline{\varphi}) = \sum_{\alpha=1}^{N} R^k(\underset{\sim}{\theta}_\alpha,\varphi)$$

which is the same as Γ^k Bayes risk of φ against the non–normalized empiric G_N^k of the N overlapping k–tuples $\underset{\sim}{\theta}_i$,$1 \le i \le N$.

The compound decision problem with $R^k(G_N^k)$ as its risk standard is called the k–extended set compound decision problem. Let

$$\text{(1.11)} \qquad \underline{D}^k(\underline{\theta},\underline{\varphi}) = R(\underline{\theta},\underline{\varphi}) - R^k(G_N^k) \qquad 1 \le k \le N.$$

$\underline{D}^k(\underline{\theta},\underline{\varphi})$ is called the modified regret of the compound procedure $\underline{\varphi}$ at $\underline{\theta}$ in the k–extended compound problem. Since G_N is the marginal of G_N^k on the last coordinate, it is immediate from the previously mentioned remark of Gilliland and Hannan (1969) that $R^k(G_N^k) \le R(G_N)$. Thus $R^k(G_N^k)$ is more stringent than $R(G_N)$. Hence producing compound rules satisfying

(1.12) $$\sup_{\underline{\theta}} \underline{D}^k(\underline{\theta},\underline{\varphi}) = o(N)$$

is more ambitious than producing rules satisfying (1.7). Set compound rules $\underline{\varphi}$ where φ_α plays Γ^k Bayes against an estimate of G_N^k in the α^{th} component problem may provide asymptotic solutions to the k–extended problem.

1.5 Literature Review

The compound decision problem was introduced by Robbins (1951). In his featured example involving N independent discriminations between N(1,1) and N(–1,1), he exhibited a bootstrap compound procedure satisfying (1.7). The bootstrap refers to the fact that each φ_α is component Bayes against an estimate of G based on all observations. Since Robbins original paper there has developed a large literature and much of it has dealt with the construction of bootstrap rules satisfying (1.7) with rates for various component problems. The most general results available in the literature for finite Θ are those of Gilliland and Hannan (1986,1974) in which they reduce the problem of a.o. of the unextended compound problem to that of the consistency of the estimates. Vardeman (1980) successfully used these results to obtain admissible a.o. rules for the k–extended problem. Vardeman (1980) used a clever separation technique and the concavity of Bayes risk that allowed direct application of existing unextended results to k–extended problems.

One of the most important developments in compound decision theory can be traced back to Hannan (1956, 1957) for the introduction of the sequence compound problem. The sequence compound problem restricts the compound rules to $\underline{\varphi}$, where each φ_α is a function of the first α observations, $1 \leq \alpha \leq N$. Hannan's procedures in the sequence problem involves artificial randomization and Van Ryzin (1966a, 1966b) showed that in many finite state finite act statistical problems the extra randomization is not necessary. Ballard (1974) in his thesis generalized Van Ryzin's (1966b) procedures in a finite state finite act statistical setting to achieve k–extended risk objectives in the sequence version of the compound problem. Ballard (1974) showed, on the scale of average risk, that the excess compound risk over $R^k(G_N^k)$ for his procedures has rate $O(N^{-1/5})$. By taking advantage of the special product structure of the estimator of G_N^k that was used in Ballard's (1974) procedures, Ballard, Gilliland and Hannan (1975) improved the rate of convergence to the rate $O(N^{-1/2})$ obtained for the unextended case by Van Ryzin (1966b).

In Chapter 2 we consider the set version of this compound problem and obtain an analogue of the Ballard, Gilliland and Hannan (1975) result thereby generalizing Van Ryzin (1966a) procedures to produce solutions to the set version of the k–extended problem.

CHAPTER 2

SET COMPOUND DECISION PROBLEM WITH $m \times n$ COMPONENT

In this Chapter we consider the set version of the k–extended problem in the finite state finite act statistical setting (as in Ballard (1974)) and exhibit two compound procedures that satisfy (1.12) with rate $O(N^{\frac{1}{2}})$. The chapter is organized as follows. In Section 2.1 we describe our Γ^k problem and establish some useful results related to Γ^k Bayes rules (Remark 2.2) and Γ^k risk (Lemma 2.1). In Section 2.2 we give a brief review of the estimator of G_N^k that we use in our procedures. In Section 2.3 we define our procedures (2.14) and (2.15). In Section 2.4 we prove that they satisfy (1.12) with rate $O(N^{\frac{1}{2}})$. In Section 2.5 we consider the extension to compact action space and prove that the procedures satisfy (1.12) by adapting some results in Oaten (1972).

2.1 Preliminaries

We consider a set compound problem as described in Section 1.1 with $\Theta = \{1,2,\ldots,m\}$ indexing $\mathscr{P} = \{P_1,P_2,\ldots,P_m\}$ and $\mathscr{A} = \{1,2,\ldots,n\}$. Under this set up, the Γ^k decision problem has $m^k \times n$ loss matrix L^k satisfying $L^k(\underset{\sim}{\theta}_k,j) = L(\theta_k,j)$ and a randomized Γ^k decision rule $\varphi = (\varphi_1,\varphi_2,\ldots,\varphi_n)$ is a mapping into $\mathscr{A}^*$, the set of probability measures on $\mathscr{A}$ such that $L^k(\underset{\sim}{\theta}_k,\varphi)$ is measurable for each θ_k, with risk

$$(2.1) \qquad R^k(\underset{\sim}{\theta}_k,\varphi) = \int \sum_{j=1}^{n} L(\theta_k,j)\ \varphi_j\ dP_{\underset{\sim}{\theta}_k}.$$

Let $\mu = \sum_{i=1}^{m} P_i$, $0 \le f_\theta \le 1$ be a density of P_θ with respect to μ $\forall\ \theta \in \Theta$ and

(2.2) $$f_{\underset{\sim}{\theta}_k} = \bigotimes_{j=1}^{k} f_{\theta_j} \qquad \forall\ \underset{\sim}{\theta}_k \in \Theta^k .$$

Also let $\Omega = (\mathbb{R}_+)^{m^k}$ and u^j be the Ω–valued function with

(2.3) $$u^j_{\underset{\sim}{\theta}_k} = L(\theta_k, j)\, f_{\underset{\sim}{\theta}_k} \qquad \forall\ \underset{\sim}{\theta}_k \in \Theta^k.$$

Then the Γ^k risk $R^k(\underset{\sim}{\theta}_k, \varphi)$ in (2.1) can be written as

(2.4) $$R^k(\underset{\sim}{\theta}_k, \varphi) = \int \sum_{j=1}^{n} u^j_{\underset{\sim}{\theta}_k}\, \varphi_j\, d\mu^k .$$

Remark 2.1. With

(2.5) $$u^{ij} = u^i - u^j ,$$

for any two Γ^k decision rules φ and φ'

(2.6) $$R^k(\underset{\sim}{\theta}_k, \varphi) - R^k(\underset{\sim}{\theta}_k, \varphi') = \int \sum_{ij} u^{ij}_{\underset{\sim}{\theta}_k}\, \varphi_i\, \varphi'_j\, d\mu^k$$

by writing the φ and φ' integrals as the product integral of the difference of their integrands.

Let $\bar{L} = \sup_{\theta, \mathcal{A}} L(\theta, a)$. Then

$$|u^{ij}_{\underset{\sim}{\theta}_k}| \le \bar{L}\, f_{\underset{\sim}{\theta}_k} \qquad \forall\ \underset{\sim}{\theta}_k \in \Theta^k ,$$

hence

$$\text{(2.7)} \qquad \text{R.H.S. (2.6)} \le \bar{L} \int \sum_{ij} f_{\underset{\sim}{\theta}_k} [\varphi_i \varphi_j' > 0] d\mu^k .$$

Let $S \subseteq [0,\infty)^{m^k}$ be the risk set of this Γ^k problem. Then for each Γ^k decision rule φ we can associate a point s in S, with coordinates of s given by (2.4). For $\omega \in \Omega$ and $s \in S$, let ωs denote the vector inner product of ω and s. We will also identify $(\theta_1,\theta_2,\ldots,\theta_k) \in \Theta^k$ with the basis vector in Ω with 1 in the $(\theta_1,\theta_2,\ldots,\theta_k)$ position. Thus if s is the risk vector associated with φ then

$$R^k(\underset{\sim}{\theta}_k,\varphi) = \underset{\sim}{\theta}_k s$$

and the Bayes risk of φ versus $\omega \in \Omega$ is

$$\omega s = \int \sum_{j=1}^{n} \omega u^j \varphi_j \, d\mu^k$$

and

$$R^k(\omega) = \omega\sigma(\omega) = \bigwedge_{s \in S} \omega s.$$

That is, $\sigma(\omega)$ is the risk vector associated with a $\varphi(\omega)$ satisfying

$$\text{(2.8)} \qquad \varphi_j(\omega) = 0 \quad \text{if } j \text{ is not a minimizer of } \omega u^j.$$

Remark 2.2

For every $< \omega_1,\omega_2,\underset{\sim}{x}_k > \in \Omega \times \Omega \times \mathscr{X}^k$ and φ satisfying (2.8) ,

$$\text{(2.9)} \quad \varphi_i(\omega_1)(\underset{\sim}{x}_k)\, \varphi_j(\omega_2)(\underset{\sim}{x}_k) > 0 \ \text{ only if } \ u^{ij}(\underset{\sim}{x}_k)\omega_1 \le 0 \le u^{ij}(\underset{\sim}{x}_k)\omega_2 .$$

Let $E_{\underline{\theta}_i}$ denote the expectation w.r.t. $P_{\underline{\theta}_i}$ and $\underline{E}$ denote the expectation w.r.t. $\underline{P} = P_{\underline{\theta}}$.

The following lemma gives a useful upper bound for the risk of Γ^k Bayes rules.

Lemma 2.1

If H and H' are mappings from $\mathscr{X}^N$ into Ω and $\varphi(\omega)$ is Γ^k Bayes against $\omega \in \Omega$, then for all $\underline{X} \in \mathscr{X}^N$ and $\underset{\sim}{\theta}_k \in \Theta^k$

$$\text{(2.10)} \qquad E_{\underset{\sim}{\theta}_k} L^k(\underset{\sim}{\theta}_k, \varphi(H^\alpha(\cdot))(\cdot)) - \underset{\sim}{\theta}_k \sigma(H'(\underline{X}))$$
$$\leq \overline{L} \int \sum_{ij} f_{\underset{\sim}{\theta}_\alpha} [\varphi_i(H^\alpha)\ \varphi_j(H'(\underline{X})) > 0]\ d\mu^k ,$$

with $H^\alpha(\cdot) = H(\underline{X}_{\alpha-k}, \cdot\ , {}_\alpha\underline{X}_N)$.

Proof

The assertion (2.10) is that of (2.7) in Remark (2.1) with φ and φ' replaced by $\varphi(H^\alpha(\cdot))(\cdot)$ and $\varphi(H'(\underline{X}))$ respectively. □

2.2 Bootstrap Procedures and Estimation of the Empiric G_N^k

Definition 2.1

A set compound rule $\underline{\varphi}$ is called k–order non–delete bootstrap rule associated with the Ω–valued estimator W_N based on $\underline{X}$ if for each $1 \leq \alpha \leq N$ $\varphi_\alpha(\underline{X}) = \varphi(W_N(\underline{X}))(\underset{\sim}{X}_\alpha)$ where $\varphi(\omega)$ is Γ^k Bayes against ω. The rule $\underline{\varphi}$ will be called k–order delete bootstrap rule associated with the Ω–valued estimator ${}_{\alpha}W_{\alpha-k}$ based on ${}_{\alpha}\underline{X}_{\alpha-k}$ if for each $1 \leq \alpha \leq N$ $\varphi_{\alpha}(\underline{X}) = \varphi({}_{\alpha}W_{\alpha-k}({}_{\alpha}\underline{X}_{\alpha-k}))(\underset{\sim}{X}_{\alpha})$.

In order to find k–order bootstrap rules in the k–extended compound problem we need to estimate the empirics G_N^k.

The question of estimating the empiric G_N^k has already been solved in Ballard (1974) in the following sense.

If the estimator $\mathbf{h}$ on $\mathscr{X}^k$ to Ω: $\mathbf{h}(\underset{\sim}{X}_k) = \{\underline{h}_{\underset{\sim}{\theta}_k}(\underset{\sim}{X}_k) : \underset{\sim}{\theta}_k \in \Theta^k\}$ is such that $E_{\underset{\sim}{i}_k} h_{\underset{\sim}{\theta}_k}(\underset{\sim}{X}_k) = [\underset{\sim}{\theta}_k = \underset{\sim}{i}_k]$, then the estimate $\sum_{\alpha=1}^{r} \mathbf{h}(\underset{\sim}{X}_{\alpha})$ is an unbiased estimate of G_r^k , $\forall\ 1 \leq k \leq r$.

It has been shown that such a function $\mathbf{h}$ exists if the set of densities $\{f_1, f_2, \ldots, f_m\}$ are linearly independent in $L_1(\mu)$. One such bounded $\mathbf{h}$ can be obtained by taking bounded unbiased estimators $h = (h_1, h_2, \ldots, h_m)$ of $\mathscr{P}$, and defining the mapping $\mathbf{h}$ from $\mathscr{X}^k$ to Ω componentwise by

$$h_{\underset{\sim}{\theta}_k} = \bigotimes_{i=1}^{k} h_{\theta_i} \qquad \underset{\sim}{\theta}_k \in \Theta^k. \tag{2.11}$$

Such an estimator is called a product estimator. Further the covariance matrix of $\mathbf{h}$ has full rank under $P_{\underset{\sim}{\theta}_k}$ $\forall\ \underset{\sim}{\theta}_k \in \Theta^k$ if the covariance matrix of h has full rank under P_{θ} $\forall\ \theta \in \Theta$.

The details of the results stated above and the method of obtaining such functions $\mathbf{h}$ are given in Section 3 of Ballard (1974).

Our theorems concern k–order bootstrap rules based on the bounded unbiased product estimator $\mathbf{h}$ of $\mathscr{X}^k$ defined in (2.11).

The estimators of G_N^k we will be using in our procedures (Definition 2.1) are W_N with

$$(2.12) \qquad W_N(\underline{X}) = \sum_{\alpha=1}^{N} h(\underset{\sim}{X}_\alpha) = H_N \text{ (say)}$$

for the non-delete rules, ${}_\alpha W_{\alpha-k}$ with

$$(2.13) \qquad {}_\alpha W_{\alpha-k}({}_\alpha\underline{X}_{\alpha-k}) = \sum_{i=\alpha+k}^{\alpha-k} h(\underset{\sim}{X}_\alpha) = {}_\alpha H_{\alpha-k} \text{ (say)} \qquad 1 \le \alpha \le N$$

for the delete rules.

The estimators W_N and ${}_\alpha W_{\alpha-k}$ has (k–1)–dependent summands for $N > 2k$, W_N is unbiased for G_N^k and ${}_\alpha W_{\alpha-k}$ is independent of $\underset{\sim}{X}_\alpha$ for each α and, on the average scale, is asymptotically unbiased for G_N^k .

2.3 Definition of the Procedures and a Useful Upper Bound for the Modified Regret

With H_N of (2.12) and ${}_\alpha H_{\alpha-k}$ of (2.13) the set compound procedures we investigate are

$$(2.14) \qquad \underset{\sim}{\varphi} \text{ with } \varphi_\alpha(\underline{X}) = \varphi(H_N)(\underset{\sim}{X}_\alpha) \quad \text{for} \quad 1 \le \alpha \le N$$

and

$$(2.15) \qquad \underset{\sim}{\varphi}^* \text{ with } \varphi_\alpha^*(\underline{X}) = \varphi({}_\alpha H_{\alpha-k})(\underset{\sim}{X}_\alpha) \quad \text{for} \quad 1 \le \alpha \le N \text{ .}$$

The following lemma gives useful upper bounds for the modified regret of the k–extended compound problem evaluated at $\underset{\sim}{\varphi}$ and $\underset{\sim}{\varphi}^*$.

Lemma 2.2.

With W_N defined in (2.12) let

(2.16) $$W_N^{\alpha}(\cdot) = W_N(\underset{\sim}{X}_{\alpha-k}, \cdot, {}_{\alpha}\underset{\sim}{X}_N).$$

Then, for $\underset{\sim}{\varphi}$ and $\underset{\sim}{\varphi}^*$ defined in (2.14) and (2.15)

(2.17) $$\underline{D}^k(\underset{\sim}{\theta},\underset{\sim}{\varphi}) \le A_N + B_N$$

and

(2.17)* $$\underline{D}^k(\underset{\sim}{\theta},\underset{\sim}{\varphi}^*) \le A_N^* + B_N$$

where

$$A_N = \overline{L} \sum_{\alpha=1}^{N} \int \sum_{ij} f_{\underset{\sim}{\theta}_\alpha} \underline{E}[u^{ij}\, W_N^{\alpha} \le 0 \le u^{ij}\, H_N]\, d\mu^k$$

$$A_N^* = \overline{L} \sum_{\alpha=1}^{N} \int \sum_{ij} f_{\underset{\sim}{\theta}_\alpha} \underline{E}[u^{ij}\, {}_{\alpha}H_{\alpha-k} \le 0 \le u^{ij}\, H_N]\, d\mu^k$$

and

$$B_N = \underline{E}\, G_N^k\, (\sigma(H_N) - \sigma(G_N^k)) .$$

Proof

For each $1 \le \alpha \le N$

(2.18) $$R_\alpha(\underset{\sim}{\theta},\underset{\sim}{\varphi}) = \underline{E}\, E_{\underset{\sim}{\theta}_\alpha} L^k(\underset{\sim}{\theta}_\alpha, \varphi(W_N^{\alpha}(\cdot))(\cdot))$$

and

(2.18)* $$R_\alpha(\underset{\sim}{\theta},\underset{\sim}{\varphi}^*) = \underline{E}\, E_{\underset{\sim}{\theta}_\alpha} L^k(\underset{\sim}{\theta}_\alpha, \varphi({}_{\alpha}H_{\alpha-k})(\cdot)) .$$

We will apply (2.10) of Lemma 2.1 with $H = H' = W_N$ to the inner integral of the R.H.S. of (2.18) and with $H = {}_{\alpha}W_{\alpha-k}$ and $H' = W_N$ to the inner integral of the R.H.S. of (2.18)*, noting the abbreviations in (2.12) and (2.13) and (2.16), to obtain

$$(2.19)\quad R_\alpha(\underset{\sim}{\ell},\underset{\sim}{\varphi}) \le \bar{L}\ \underline{E} \int \sum_{ij} f_{\underset{\sim\alpha}{\theta}} [\varphi_i(W_N^\alpha)\varphi_j(H_N) > 0]\, d\mu^k + \underline{E}\ \underset{\sim\alpha}{\ell}\sigma(H_N),$$

$$(2.19)^*\quad R_\alpha(\underset{\sim}{\ell},\underset{\sim}{\varphi}^*) \le \bar{L}\ \underline{E} \int \sum_{ij} f_{\underset{\sim\alpha}{\theta}} [\varphi_i({}_\alpha H_{\alpha-k})\ \varphi_j(H_N) > 0]\, d\mu^k + \underline{E}\ \underset{\sim\alpha}{\ell}\ \sigma(H_N).$$

Taking
$$< \omega_1,\omega_2, \cdot > = \begin{cases} < W_N^\alpha(\ \cdot\ ), H_N, \cdot > \\ < {}_\alpha H_{\alpha-k}\ , H_N, \cdot > \end{cases}$$

in (2.9) of Remark 2.2, we bound each summand in the integrand of the first term in R.H.S. (2.19) by

$$f_{\underset{\sim\alpha}{\theta}} [u^{ij}\ W_N^\alpha \le 0 \le u^{ij}\ H_N]$$

and that of $(2.19)^*$ by

$$f_{\underset{\sim\alpha}{\theta}} [u^{ij}\ {}_\alpha H_{\alpha-k} \le 0 \le u^{ij}\ H_N]\ .$$

With these bounds substituted in (2.19) and $(2.19)^*$, followed by summation over all α and the interchange of $\underline{E}$ and $\int \sum_{ij}$,

$$(2.20)\quad \underline{R}(\underset{\sim}{\ell},\underset{\sim}{\varphi}) \le A_N + \sum_{\alpha=1}^{N} \underline{E}\ \underset{\sim\alpha}{\ell}\ \sigma(H_N)$$

and

$$(2.20)^*\quad \underline{R}(\underset{\sim}{\ell},\underset{\sim}{\varphi}^*) \le A_N^* + \sum_{\alpha=1}^{N} \underline{E}\ \underset{\sim\alpha}{\ell}\ \sigma(H_N)\ .$$

Since the second term in the R.H.S. of (2.20) and $(2.20)^*$ is $\underset{\sim}{E}\, G_N^k\, \sigma(H_N)$, (2.17) and $(2.17)^*$ follow by (1.11). □

2.4 Asymptotic Optimality

Theorem 2.1

Let λ_θ denote the minimum eigenvalue of the covariance matrix of $h = (h_1, h_2, \ldots, h_m)$ under P_θ ; $\theta \in \Theta$ and $\lambda = \min\{\lambda_\theta : \theta \in \Theta\}$. Suppose the kernel **h** of (2.12) and (2.13) is the bounded unbiased product estimator (2.11) and λ defined above is positive. Then, for the compound procedures $\underline{\varphi}$ and $\underline{\varphi}^*$ defined in (2.12) and (2.13),

$$\bigvee_{\underline{\theta}} \underline{D}^k(\underline{\theta}, \underline{\varphi}) = O(N^{\frac{1}{2}}) \tag{2.21}$$

and

$$\bigvee_{\underline{\theta}} \underline{D}^k(\underline{\theta}, \underline{\varphi}^*) = O(N^{\frac{1}{2}})\ . \tag{2.21*}$$

Proof

In view of (2.17) and $(2.17)^*$ it is enough to show that $A_N = O(N^{\frac{1}{2}})$, $A_N^* = O(N^{\frac{1}{2}})$ and $B_N = O(N^{\frac{1}{2}})$. We establish these results in Lemma 2.3 and Lemma 2.4. □

Lemma 2.3

Assume all the conditions of Theorem 2.1. Fix $\underset{\sim}{x}_k \in \mathscr{X}^k$ and $a, b \in \mathscr{A}$.

Let $v_{\theta_1}(j) = f_{\theta_1}(x_j), \ldots\ldots, v_{\theta_{k-1}}(j) = f_{\theta_{k-1}}(x_j)$

$v_{\theta_k}(j) = f_{\theta_k}(x_j)\,(L(\theta_k, a) - L(\theta_k, b)) \quad \forall\ j = 1, 2, \ldots, k;\ \forall\ \underset{\sim}{\theta}_k \in \Theta^k$ and

$\|v\| = \prod_{j=1}^{k} (v(j), v(j))^{\frac{1}{2}} > 0\ ,\ (\sum_{j=1}^{m} h_j^2\ )^{\frac{1}{2}} \le M^{1/k}$. Then

(2.22) $$|u^{ab}(\underset{\sim}{x}_k)\ \mathbf{h}| \le M\ \|v\|$$

and, for H_N, ${}_\alpha H_{\alpha-k}$, W_N^α defined in (2.12), (2.13), (2.16) and, for $N > 2k$, $\exists$ a constant C_1 independent of N and $\underset{\sim}{\theta}$ such that

(2.23) $$\underline{E}[u^{ab}(\underset{\sim}{x}_k)\ W_N^\alpha(\underset{\sim}{x}_k) \le 0 \le u^{ab}(\underset{\sim}{x}_k)\ H_N] \le C_1 N^{-\frac{1}{2}}\ ,$$

(2.24) $$\underline{E}[u^{ab}(\underset{\sim}{x}_k)\ {}_\alpha H_{\alpha-k} \le 0 \le u^{ab}(\underset{\sim}{x}_k)\ H_N] \le C_1 N^{-\frac{1}{2}}\ .$$

Lemma 2.4

For H_N defined in (2.12)

(2.25) $$\underline{E}\ G_N^k(\sigma(H_N) - \sigma(G_N^k)) \le C_2 N^{\frac{1}{2}}\ .$$

The proofs of Lemma 2.3 and 2.4 depend on the following Proposition 1 and the Theorem 2 of Section 4 in Ballard, Gilliland and Hannan (1975).

(B.G.H.) Proposition 1

Let $k \ge 1$ and suppose $U_1, U_2, \ldots\ldots$ are (k–1)–dependent random variables. Then

$$\mathrm{Var}\ (\sum_{i=1}^{n} U_i) \le knv_n$$

where $v_n = \max\ \{\ \mathrm{var}\ U_i : 1 \le i \le n\ \}$.

(B.G.H.) Theorem 2

Let $v(1),\ldots\ldots,v(k)$ be fixed vectors in $\mathbb{R}^m$ and $h = (h_1,\ldots.,h_m)$ be an $\mathbb{R}^m$–valued function on $\mathcal{X}$, the range space of the independent random variables $X_1,X_2,\ldots.$ and $(X_1,X_2,\ldots.) \sim P_{\theta_1} \times P_{\theta_2} \times\ldots.$ for $\underline{\theta}^{\infty} \in \Theta^{\infty}$ with $\Theta = \{1,2,\ldots.,m\}$. Let $\pi_\alpha = \prod_{j=1}^{k} (v(j),h(X_{j+\alpha}))$ $\alpha = 0,1,\ldots..$ and $\|v\| = \prod_{j=1}^{k} (v(j),v(j))^{\frac{1}{2}}$. Let λ_θ denote the minimum eigenvalue of the covariance matrix of h under P_θ, $\forall$ $\theta \in \Theta$ and $\lambda = \min\{\lambda_\theta : \theta \in \Theta\}$. If $\lambda > 0$ and $(h,h)^{\frac{1}{2}} \leq M^{1/k} < \infty$ then

$$\gamma^{\frac{1}{2}} \underline{P}\{a \leq \sum_{\alpha=0}^{n-1} \pi_\alpha \leq b\} \leq A\ (\|v\|)^{-1} + B \qquad n \geq k,\ \underline{\theta}^{\infty} \in \Theta^{\infty}$$

where $A = (\pi k\lambda^k)^{-\frac{1}{2}}(b-a)$ and $B = 2\ 2^{\frac{1}{2}}\ kM[C(k\lambda^k)^{-\frac{1}{2}} + M\lambda^{-k}]$; C is the Berry Esseen constant in the independent summand case and γ is the greatest integer in nk^{-1}.

Proof of Lemma 2.3

Since $u^{ab}_{\underset{\sim}{\theta}_k}(\underset{\sim}{x}_k) = v_{\theta_1}(1)\ldots.v_{\theta_k}(k)$ and $h_{\underset{\sim}{\theta}_k} = h_{\theta_1}\otimes\ldots\ldots\otimes h_{\theta_k}$

$$(2.26)\ u^{ab}(\underset{\sim}{x}_k)h = \sum_{\underset{\sim}{\theta}_k \in \Theta^k} (v_{\theta_1}(1)\ldots.v_{\underset{\sim}{\theta}_k}(k))(h_{\theta_1}\otimes\ldots.\otimes h_{\theta_k}) = \bigotimes_{j=1}^{k} (v(j),h).$$

Applying Schwartz inequality to each of the inner products $(v(j),h)$ and using the definition of v with the fact that $(\sum_{j=1}^{m} h_j^2)^{\frac{1}{2}} \leq M^{1/k}$ we obtain (2.22).

By (2.12) and (2.26)

$$(2.27)\qquad u^{ab}(\underset{\sim}{x}_k)H_N = \sum_{\alpha=1}^{N} u^{ab}(\underset{\sim}{x}_k)\, h(\underset{\sim}{X}_\alpha)$$
$$= \sum_{\alpha=1}^{N} (v(1),h(X_{\alpha+1}))\ldots\ldots(v(k),h(X_{\alpha+k})) \ .$$

Similarly, by (2.13) and (2.26)

$$(2.28)\qquad u^{ab}(\underset{\sim}{x}_k)\,{}_\alpha H_{\alpha-k} = \sum_{i=\alpha+k}^{\alpha-k} (v(1),h(X_{i+1}))\ldots\ldots(v(k),h(X_{i+k})) \ .$$

Each of (2.27) and (2.28) has (k–1)–dependent summands of the type π_α of (B.G.H.) Theorem 2 . Also,

$$(2.29)\qquad u^{ab}(\underset{\sim}{x}_k)\ (H_N - W_N^\alpha(\underset{\sim}{x}_k)) = \sum_{i=\alpha-k+1}^{\alpha+k-1} u^{ab}(\underset{\sim}{x}_k)\ (h(\underset{\sim}{X}_i) - h(\underset{\sim}{Z}_i))$$

with

$$\underset{\sim}{Z}_i = \begin{cases} (X_{i-(k-1)},\ldots,X_{\alpha-k}\ ,x_1,..x_{i-(\alpha-k)}) & \alpha-k+1 \le i \le \alpha \\ (x_{i+1-\alpha},\ldots,x_k\,,X_{\alpha+1},\ldots,X_i) & \alpha+1 \le i \le \alpha+k-1 \end{cases}$$

and,

$$(2.30)\qquad u^{ab}(\underset{\sim}{x}_k)\ ({}_\alpha H_{\alpha-k} - H_N) = -\sum_{i=\alpha+N-k+1}^{\alpha+N+k+1} u^{ab}(\underset{\sim}{x}_k)\ h(\underset{\sim}{X}_i) \ .$$

Hence an application of the bound in (2.22) to the R.H.S. of (2.29) and (2.30) respectively yields

$$(2.31)\qquad u^{ab}(\underset{\sim}{x}_k)\ (H_N - W_N^\alpha(\underset{\sim}{x}_k)) \ge -4kM\|v\| \ ,$$

and

$$(2.32)\qquad u^{ab}(\underset{\sim}{x}_k)\ ({}_\alpha H_{\alpha-k} - H_N) \ge -4kM\|v\| \ .$$

Note that

(2.33) L.H.S. (2.23)

$$= \underline{E}[u^{ba}(\underset{\sim}{x}_k)\ (H_N - W_N^{\alpha}(\underset{\sim}{x}_k) \le u^{ba}(\underset{\sim}{x}_k)\ H_N \le 0]$$

and

(2.34) L.H.S. (2.24)

$$= \underline{E}[u^{ab}(\underset{\sim}{x}_k)\ ((_{\alpha}H_{\alpha-k}) - H_N) \le u^{ab}(\underset{\sim}{x}_k)\ {}_{\alpha}H_{\alpha-k} \le 0]\ .$$

Replacing the lower bound in each of the integrands in (2.33) and (2.34) by the bounds in (2.31) and (2.32)

(2.35) $\text{R.H.S. } (2.33) \le \underline{E}[-4kM\|v\| \le u^{ba}(\underset{\sim}{x}_k)H_N \le 0]$

and

(2.36) $\text{R.H.S. } (2.34) \le \underline{E}[-4kM\|v\| \le u^{ab}(\underset{\sim}{x}_k)\,{}_{\alpha}H_{\alpha-k} \le 0]$.

Each of the summands in R.H.S. of (2.35) and (2.36), (cf. (2.27) and (2.28)) are summands of (k–1)–dependent random variables of the type π_α in (B.G.H.) Theorem 2. For $x_1,\ldots\ldots,x_k$ such that $\|v\| > 0$ we can apply (B.G.H.) Theorem 2 with $a = -4kM\|v\|$ and $b = 0$ to the R.H.S. of (2.35) and (2.36), to obtain (2.23) and (2.24) with $C_1 = (\pi k\lambda^k)^{-\frac{1}{2}}4kM + B$. □

Proof of Lemma 2.4

Since $H_N\ \sigma(H_N) \le H_N\ \sigma(G_N^k)$

$$(2.37)\qquad G_N^k(\sigma(H_N) - \sigma(G_N^k)) \le (G_N^k - H_N)(\sigma(H_N) - \sigma(G_N^k)) .$$

Taking φ and φ' in (2.6) as $\varphi(H_N)$ and $\varphi(G_N^k)$ and bounding $u^{ij}_{\underset{\sim}{\theta}_k}$ by $\bar{L}\ f_{\underset{\sim}{\theta}_k}$ we obtain

$$|\underset{\sim}{\theta}_k \sigma(H_N) - \underset{\sim}{\theta}_k \sigma(G_N^k)| \le \bar{L} \qquad \forall\ \underset{\sim}{\theta}_k \in \Theta^k.$$

Hence

$$(2.38)\qquad \text{R.H.S. } (2.37) \le \bar{L} \sum_{\underset{\sim}{\theta}_k \in \Theta^k} |G^k_{N\underset{\sim}{\theta}_k} - H^k_{N\underset{\sim}{\theta}_k}| .$$

Integrating both sides of (2.38) with respect to $\underline{E}$ and using the fact that H_N is unbiased for G_N^k and the moment inequality, we obtain

$$(2.39)\qquad \text{L.H.S. } (2.25) \le \bar{L} \sum_{\underset{\sim}{\theta}_k \in \Theta^k} (\text{Var } H_{N\underset{\sim}{\theta}_k})^{\frac{1}{2}} .$$

Since $H_{N\underset{\sim}{\theta}_k} = \sum_{\alpha=1}^{N} h_{\theta_1}(X_{\alpha-k}).....h_{\theta_k}(X_\alpha)$ is a sum of (k–1)–dependent random variables, from (B.G.H.) Proposition 1 and the fact that $(\sum_{j=1}^{n} h_j^2)^{\frac{1}{2}} \le M^{1/k}$

$$\text{Var } H_{N\underset{\sim}{\theta}_k} \le kNM^2 \qquad \forall\ \underset{\sim}{\theta}_k \in \Theta^k.$$

Thus (2.39) yields (2.25) with $C_2 = \bar{L}\ m^k(kM^2)^{\frac{1}{2}}$. □

2.4 Infinite Action Space

In this section we replace the assumption that $\mathcal{A}$ is finite by

(2.40) $\quad \mathcal{A}$ is totally bounded in the metric $d(a,b) = \sup_{\theta} |L(\theta,a) - L(\theta,b)|$

and obtain results analogous to Theorem 2.1.

For each $\epsilon > 0$ let $D_\epsilon = \{a_1,\dots\dots,a_r\} \subset \mathcal{A}$ be such that disjoint $A_j \subseteq B_\epsilon(a_j)$, $1 \le j \le r$ covers $\mathcal{A}$. Consider the problem obtained from the original (henceforth called the $\mathcal{A}$-problem) when we restrict the action space to D_ϵ.

For any decision rule φ in the $\mathcal{A}$-problem let φ^ϵ denote the decision rule in the sub-problem given by $\varphi^\epsilon_j = \varphi(A_j)$, $i \le j \le r$.

Then, if φ is a Γ^k decision rule in the $\mathcal{A}$-problem $\forall\ \underset{\sim}{\theta}_k \in \underset{\sim}{\Theta}^k$

$$
\begin{aligned}
|L^k(\underset{\sim}{\theta}_k,\varphi) - L^k(\underset{\sim}{\theta}_k,\varphi^\epsilon)| &= |\sum_{j=1}^{r} \int_{A_j} L(\theta_k,a)\ \varphi(da) - \sum_{j=1}^{r} L(\theta_k,a_j)\ \varphi^\epsilon_j| \\
&= |\sum_{j=1}^{r} \int_{A_j} (L(\theta_k,a) - L(\theta_k,a_j))\ \varphi(da)| \\
&\le \epsilon.
\end{aligned}
$$

By integrating this inequality with respect to $E_{\underset{\sim}{\theta}_k}$,

(2.41) $$|R^k(\underset{\sim}{\theta}_k,\varphi) - R^k(\underset{\sim}{\theta}_k,\varphi^\epsilon)| \le \epsilon \qquad \forall\ \underset{\sim}{\theta}_k \in \underset{\sim}{\Theta}^k.$$

In particular, for a Γ^k Bayes rule $\varphi(\omega)$ against a $\omega \in \Omega$ in the $\mathcal{A}$-problem and using $\varphi^\epsilon(\omega)$ to denote $(\varphi(\omega))^\epsilon$, (2.41) implies

(2.42) $$R^k(\underset{\sim}{\theta}_k \cdot \varphi^\epsilon(\omega)) - \underset{\sim}{\theta}_k\, \sigma(\omega) \le \epsilon \qquad \forall\ \underset{\sim}{\theta}_k \in \underset{\sim}{\Theta}^k ,$$

where σ denotes the risk of $\varphi(\omega)$.

The following remark regarding a Γ^k Bayes rule φ in the $\mathcal{A}$-problem is an adaptation from Section 6 of Oaten (1972).

Remark 2.3

For every $< \omega_1, \omega_2, \underset{\sim}{x}_k > \in \Omega \times \Omega \times \mathcal{X}^k$ and, Γ^k Bayes rule φ in the $\mathcal{A}$-problem

(2.43) $$\varphi_i^\epsilon(\omega_1)(\underset{\sim}{x}_k)\ \varphi_j^\epsilon(\omega_2)(\underset{\sim}{x}_k) > 0 \quad \text{only if}$$
$$u^{b_i b_j}(\underset{\sim}{x}_k)\omega_1 \le 0 \le u^{b_i b_j}(\underset{\sim}{x}_k)\omega_2 \quad \text{for some } \{ b_i, b_j \} \in \{ A_i, A_j \}.$$

Proof

By the definition of φ^ϵ, for any $\omega \in \Omega$ and $\underset{\sim}{x}_k \in \mathcal{X}^k$
$$\varphi_j^\epsilon(\omega)(\underset{\sim}{x}_k) > 0 \quad \text{only if} \quad \varphi(\omega)(\underset{\sim}{x}_k)(A_j) > 0 .$$

Since $\varphi(\omega)$ is Γ^k Bayes versus ω in the $\mathcal{A}$-problem $\varphi(\omega)(\underset{\sim}{x}_k)(A_j) > 0$ only if $\exists$ an b_j (a Bayes rule against ω) $\in A_j$ such that $\omega\, u^{b_j a}(\underset{\sim}{x}_k) \le 0 \ \ \forall\ a \in \mathcal{A}$.

Using this fact with $\omega = \omega_1$ and $\omega = \omega_2$ we obtain (2.43). $\square$

Theorem 2.2

Consider the compound procedures $\underline{\varphi}$ and $\underline{\varphi}^*$ defined in (2.14) and (2.15) but assuming $\varphi(\omega)$ is Γ^k Bayes against ω in the $\mathcal{A}$-problem. Then, under (2.40) and assumptions of Theorem 2.1

$$\bigvee_{\underline{\theta}} \underline{D}^k(\underline{\theta},\underline{\varphi}) = o(N) \tag{2.44}$$

$$\bigvee_{\underline{\theta}} \underline{D}^k(\underline{\theta},\underline{\varphi}^*) = o(N). \tag{2.44*}$$

Proof

With $\omega_1 = W_N^\alpha(\cdot)$ (cf. (2.16)) and $\omega_2 = H_N$ (cf. (2.12)) in (2.43), an application of (2.23) of Lemma 2.3 yields

$$\underline{E}\ [\ \varphi_i^\epsilon(W_N^\alpha)\ \varphi_j^\epsilon(H_N) > 0\ ] \ \leq C_1\ N^{-\frac{1}{2}}, \tag{2.45}$$

and with $\omega_1 = {}_\alpha H_{\alpha-k}$ (cf. (2.13)) and $\omega_2 = H_N$ in (2.43), (2.24) gives

$$\underline{E}\ [\ \varphi_i^\epsilon({}_\alpha H_{\alpha-k})\ \varphi_j^\epsilon(H_N) > 0\ ] \ \leq C_1\ N^{-\frac{1}{2}}. \tag{2.45*}$$

For any compound decision rule $\underline{\delta} = (\delta_1,\ldots,\delta_N)$ in the $\mathcal{A}$-problem let $\underline{\delta}^\epsilon$ be the compound rule in the sub-problem given by $\underline{\delta}^\epsilon = (\delta_1^\epsilon,\ldots,\delta_N^\epsilon)$.

Then $\forall\ \underline{\theta} \in \Theta^N$ and $1 \leq \alpha \leq N$

$$|L(\theta_\alpha,\delta_\alpha) - L(\theta_\alpha,\delta_\alpha^\epsilon)| \leq \epsilon$$

and, by integrating this with respect to $P_{\underline{\theta}}$

(2.46) $$|R_\alpha(\underline{\theta},\underline{\delta}) - R_\alpha(\underline{\theta},\underline{\delta}^\epsilon)| \leq \epsilon \qquad \forall\ \underline{\theta} \in \Theta^N.$$

Express $\underline{D}^k(\underline{\theta},\underline{\delta})$ as

(2.47) $$\underline{D}^k(\underline{\theta},\underline{\delta}) = A_1 + A_2 + A_3 + A_4$$

with

$$A_1 = \sum_{\alpha=1}^{N} \{ R_\alpha(\underline{\theta},\underline{\delta}) - R_\alpha(\underline{\theta},\underline{\delta}^\epsilon) \}$$

$$A_2 = \sum_{\alpha=1}^{N} \{ R_\alpha(\underline{\theta},\underline{\delta}^\epsilon) - \underline{E}\ R^k(\underset{\sim}{\theta}_\alpha,\varphi^\epsilon(H_N)) \}$$

$$A_3 = \sum_{\alpha=1}^{N} \underline{E}\ \{R^k(\underset{\sim}{\theta}_\alpha,\varphi^\epsilon(H_N)) - \underset{\sim}{\theta}_\alpha\ \sigma(H_N)\}$$

and

$$A_4 = \sum_{\alpha=1}^{N} \underline{E}\ \{\underset{\sim}{\theta}_\alpha\ \sigma(H_N) - \underset{\sim}{\theta}_\alpha\ \sigma(G_N^k)\} .$$

By (2.46) $A_1 \leq N\epsilon$. Applying (2.42) with $\omega = H_N$, $A_3 \leq N\epsilon$.

Note that the proof of Lemma 2.4 remains valid when σ is the risk vector associated with a Γ^k Bayes rule in the $\mathscr{A}$-problem.

Hence, an application of this generalized version of Lemma 2.4 gives $A_4 \leq C_2 N^{\frac{1}{2}}$.

Since $R_\alpha(\underline{\theta},\underline{\delta}^\epsilon) = \underline{E}\ R^k(\underset{\sim}{\theta}_\alpha,\delta_\alpha^\epsilon)$, on replacing φ and φ' in (2.7) of Remark 2.1 by δ_α^ϵ and $\varphi^\epsilon(H_N)$ and applying Fubini, each summand in A_2 is bounded by

(2.48) $$\overline{L} \int \sum_{ij} f_{\underset{\sim}{\theta}_\alpha}\ \underline{E}\ [\ \delta_{\alpha\, i}^\epsilon\ ;\ \varphi_j^\epsilon(H_N) > 0\ ]\ d\mu^k .$$

Taking $\underline{\delta} = \underline{\varphi}$ in (2.48), we apply (2.45) to the inner integral and bound the iterated integral by $C_1\ r^2\ N^{-\frac{1}{2}}$.

Similarly, we obtain the same bound by applying $(2.45)^*$ with $\underline{\delta} = \underline{\varphi}^*$ in (2.48).

Hence $$A_2 \le \bar{L}\ C_1\ r^2\ N^{\frac{1}{2}} \qquad \text{when } \underline{\delta} = \begin{cases} \underline{\varphi} \\ \underline{\varphi}^* \end{cases}.$$

With these bounds substituted into the R.H.S. (2.47)

(2.49) $$\left.\begin{array}{l} \underline{D}^k(\underline{\theta},\underline{\varphi}) \\ \underline{D}^k(\underline{\theta},\underline{\varphi}^*) \end{array}\right\} \le 2N\epsilon + C_2 N^{\frac{1}{2}} + \bar{L}\ C_1\ r^2\ N^{\frac{1}{2}}\ .$$

Since the R.H.S. (2.49) is asymptotically equal to $2N\epsilon$ we have proved (2.44) and $(2.44)^*$. □

CHAPTER 3

EXTREMA OF Eg(X) FOR GENERALIZED BINOMIAL X WITH CONSTANT VARIANCE

3.1 Introduction and Statement of the Problem

Let S be the number of successes in n independent trials, and let $\mathbf{p} = (p_1,p_2,\ldots,p_n)$ with p_j denoting the probability of success of the j^{th} trial.

Hoeffding (1956) considered the problem of finding the extremum of $Eg(S)$, the expected value of a given real valued function g on the range of S when $\Sigma\, p_i$ is fixed and proved that extrema are attained when $p_1,p_2,\ldots,p_n$ take on at most three different values, only one of which is distinct from 0 and 1. In this chapter we consider the analogous problem of finding the extrema of $Eg(S)$ when $\operatorname{var} S = \Sigma\, p_i(1-p_i) = \lambda$ is fixed and prove (Corollary 3.2) that extrema are attained when $p_1,p_2,\ldots,p_n$ take on at most four different values only two of which are distinct from 0 or 1.

The proof basically depends on the functions $f_{n-k,i}$ defined in (3.5) and the representation of $f = Eg$ given in (3.6). The characterization of extrema in Theorem 3.3 asserts that if $\mathbf{a}$ is an extrema of f and has at least three unequal coordinates in (0,1) then, any point $\mathbf{b} \in D_\lambda$ (cf. (3.3)) having the same number of zero coordinates and unit coordinates as $\mathbf{a}$ and satisfying $\sum_{i=1}^{n} b_i = \sum_{i=1}^{n} a_i$, is also an extrema of f. To prove this assertion, first we show, inductively (Lemma 3.2), the functions $f_{n-m,i}(\mathbf{a}^{1,\ldots,m})$ in (3.6) with $m = \#$ of coordinates of $\mathbf{a}$ in (0,1), are zero $\forall\ 3 \le i \le m$. Theorem 3.1 covers the case $m = 3$ and $i = 3$. Then we use the fact that, $\mathbf{a} \in D_\lambda$ and $\mathbf{b} \in D_\lambda$ and $\sum_{i=1}^{m} b_i = \sum_{i=1}^{m} a_i$ with

$f_{n-m,i}(\mathbf{a}^{1,\ldots,m}) = 0 \;\; \forall \;\; 3 \le i \le m$, in (3.6), to show $f(\mathbf{a}) = f(\mathbf{b})$ (cf. proof of (3.23)). In Corollary 3.2 we exhibit such a point $\mathbf{b} \in D_\lambda$ of the form stated.

Theorem 3.1 is a corollary to Lemma 3.1 which in turn is a consequence of the Implicit Function Theorem. Theorem 3.2 depends on a simple basic result on the intersection of circles and lines, and is helpful in evaluating the maximum.

3.2 Notations and Preliminaries

The notations used will be consistent with that of Hoeffding (1956).

For a $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ with $0 \le p_i \le 1$

$$f(\mathbf{p}) = E(g) = \sum_{k=0}^{n} g(k)\, b_{n,k}(\mathbf{p}) \tag{3.1}$$

with $b_{n,k}(\mathbf{p}) = P(S = k)$ given by

$$b_{n,k}(\mathbf{p}) = \sum_{\{\mathbf{i} \,:\, \sum_{j=1}^{n} i_j = k\}} \prod_{j=1}^{n} p_j^{i_j} (1 - p_j)^{1-i_j} \tag{3.2}$$

where $\mathbf{i} = (i_1, \ldots, i_n)$ with $i_j \in \{0,1\}$.

For $0 \le \lambda \le .25n$,

$$D_\lambda = \{ \mathbf{p} \mid 0 \le p_i \le 1 \,,\, \sum_{i=1}^{n} (p_i - .5)^2 = .25n - \lambda \} . \tag{3.3}$$

For any given $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{a}^{i_1,\ldots,i_m}$ will denote the $(n-m)$-dimensional vector obtained from $\mathbf{a}$ by deleting the coordinates $i_1, \ldots, i_m$.

For any given $\mathbf{a}$ and $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{a}(b_{i_1},\ldots,b_{i_m})$ will denote the vector obtained from $\mathbf{a}$ when $a_{i_1},\ldots,a_{i_m}$ is replaced by $b_{i_1},\ldots,b_{i_m}$.

. Since f is symmetric under permutation of its coordinates and linear in each coordinate, we can write

$$(3.4) \qquad f(\mathbf{p}) = p_i \, f_{n-1,1}(\mathbf{p}^i) + f_{n-1,0}(\mathbf{p}^i) \qquad \forall \;\; 1 \le i \le n$$

where the functions $f_{n-1,1}$ and $f_{n-1,0}$ are independent of the index i and symmetric and linear in the components of $\mathbf{p}^i$. In general, we will define the functions $f_{n-k,i}$ by

$$f_{n,0}(\mathbf{p}) = f(\mathbf{p})$$

(3.5)

$$f_{n-k,i}(\mathbf{p}^{1,\ldots,k}) = p_{k+1} \, f_{n-k-1,i+1}(\mathbf{p}^{1,\ldots,k+1}) + f_{n-k-1,i}(\mathbf{p}^{1,\ldots,k+1})$$
$$0 \le i \le k \; ; \; 0 \le k \le n-1.$$

By repeated application of (3.5) to the R.H.S. of (3.4)

$$(3.6) \qquad f(\mathbf{p}) = \sum_{i=0}^{m} c_{mi}(p_1,\ldots,p_m) \, f_{n-m,i}(\mathbf{p}^{1,\ldots,m})$$
$$m = 1,\ldots,n \; ; \quad 0 \le i \le m$$

where

$$c_{m0}(p_1,\ldots,p_m) = 1$$

and, for $1 \le i \le m$

$$c_{mi}(p_1,\ldots,p_m) = \text{the } i^{th} \text{ symmetric sum of } p_1,\ldots,p_m.$$

With $(0^r,1^s)$ denoting the point in $\mathbb{R}^{r+s}$ whose first r coordinates are 0 and the remaining s coordinates are 1, let

$$(p_1,\ldots,p_m) \in \{(0^{m-h},1^h) : h = 0,1,\ldots,m\}.$$

Evaluating (3.6) at $(0^{m-h},1^h,p^{1,2,\ldots,m})$ $\forall$ $h = 0,\ldots,i$ we obtain a system of linear equations in $f_{n-m,h}(p^{1,\ldots,m})$; $0 \le h \le i$. By solving this system for $f_{n-m,i}(p^{1,\ldots,m})$

$$(3.7) \qquad f_{n-m,i}(p^{1,\ldots,m}) = \sum_{h=0}^{i} (-1)^{i-h} \begin{bmatrix} i \\ h \end{bmatrix} f(0^{m-h},1^h,p^{1,\ldots,m}) .$$

Evaluating (3.6) at $p(a_1,\ldots,a_m)$ and substracting it from (3.6)

$$(3.8) \quad f(p) - f(p(a_1,\ldots,a_m)) = \sum_{i=1}^{m} \left[c_{m,i} \left| \begin{matrix} (p_1,\ldots,p_m) \\ (a_1,\ldots,a_m) \end{matrix} \right. \right] f_{n-m,i}(p^{1,\ldots,m}) .$$

In particular when $m = 3$

$$(3.9) \qquad f(p) - f(p(a_1,a_2,a_3)) = (p_1p_2p_3 - a_1a_2a_3)\, f_{n-3,3}(p^{1,2,3})$$

if

$$p_1 + p_2 + p_3 = a_1 + a_2 + a_3$$

and

$$p_1^2 + p_2^2 + p_3^2 = a_1^2 + a_2^2 + a_3^2 .$$

3.3 Necessary Conditions for Extrema of Eg

In this section we state and prove two sets of necessary conditions (Theorem 3.1, Theorem 3.2) for the maxima of Eg. The following lemma is a consequence of the Implicit Function Theorem and will be used in the proof of Theorem 3.1.

Lemma 3.1

Let h and k be functions from $\mathbb{R}^3 \rightsquigarrow \mathbb{R}$ defined by

$$h(x,y,z) = x + y + z$$

and

$$k(x,y,z) = x^2 + y^2 + z^2 .$$

Suppose $\mathbf{a} = (a_i,a_j,a_k) \in [0,1]^3$ is a solution to $h = \alpha$ and $k = \beta$, where α and β are known constants. If $a_i \neq a_j$, then there exists an interval J containing a_k and unique continuous functions u_i and u_j defined on J such that $\forall\ x \in J$ the point

$$\mathbf{u}(x) = (u_i(x),\ u_j(x),\ x) \in [0,1]^3$$

satisfy $h(\mathbf{u}(x)) = \alpha$ and $k(\mathbf{u}(x)) = \beta$ with $\mathbf{u}(a_k) = (a_i,\ a_j,\ a_k)$.
The interval J has the form

$$(3.10) \qquad J = \begin{cases} (a_k - \delta,\ a_k] \\ (a_k - \delta,\ a_k + \delta) \\ [a_k,\ a_k + \delta) \end{cases} \quad \text{if} \quad \mathbf{a} \in \begin{cases} (0,1)^2 \times \{1\} \\ (0,1)^3 \\ (0,1)^2 \times \{0\} . \end{cases}$$

Proof

The functions h and k have the following properties.

(i) h and k are C^1 functions on

$$B_\epsilon(\mathbf{a}) = \{\mathbf{x} \in \mathbb{R}^3 \mid \|\mathbf{x}-\mathbf{a}\| < \epsilon\}, \quad \epsilon>0.$$

(ii) $h(\mathbf{a}) = \alpha \quad k(\mathbf{a}) = \beta$

(iii) the Jacobian $\dfrac{\partial(h,k)}{\partial(x,y)}$ at $\mathbf{a}$ is non zero.

Therefore by Implicit Function Theorem $\exists\ \delta$ such that $0 < \delta < \epsilon$ and unique continuous functions u_i and u_j defined on the interval $(a_k-\delta,\ a_k+\delta)$ such that $\forall\ x \in (a_k-\delta,\ a_k+\delta)$ the point $\mathbf{u}(x) = (u_i(x), u_j(x), x) \in B_\epsilon(\mathbf{a})$ with $h(\mathbf{u}(x)) = \alpha$, $k(\mathbf{u}(x)) = \beta$ and $u_i(a_k) = a_i$, $u_j(a_k) = a_j$.

Suppose $(a_i,a_j,a_k) \in (0,1)^3$. Then for small enough $\epsilon > 0$, $B_\epsilon(\mathbf{a}) \subset (0,1)^3$. But the fact that $0 < \delta < \epsilon$ implies $\forall\ x \in (a_k-\delta,\ a_k+\delta)$, $\mathbf{u}(x) \in (0,1)^3$.

Suppose $(a_i,a_j,a_k) \in (0,1)^2 \times \{1\}$, then for small enough $\epsilon > 0$, $B_\epsilon(\mathbf{a}) \subset (0,1)^2 \times (0,1+\epsilon)$. Therefore $\forall\ x \in (a_k-\delta,\ a_k+\delta)$ the first and second coordinates of $\mathbf{u}(x)$ are in $(0,1)$. Hence $\forall\ x \in (a_k-\delta, a_k]$, $\mathbf{u}(x) \in (0,1]^3$.

By a similar argument we can show that $\forall\ x \in [a_k,a_k+\delta)$ $\mathbf{u}(x) \in [0,1)^3$ if $(a_i,a_j,a_k) \in (0,1)^2 \times \{0\}$.

Thus, we have shown the existence of an interval J of the form stated in the lemma. □

Remark 3.1

If two of a_i, a_j, a_k are on the boundary of $[0,1]$ the only solutions in $[0,1]$ for $h = \alpha$ and $k = \beta$ are the permutations of a_i, a_j, a_k.

Remark 3.2

As a consequence of (3.10) of Lemma 3.1 for any $(a_i,a_j,a_k,a_l) \in (0,1)^4$ with $o < a_i \neq a_j \neq a_k < 1$, we can always find a $(b_i,b_j,b_k) \in (0,1)^3$ such that

$$b_i \neq b_j \neq b_k \neq a_l,$$

$$a_i + a_j + a_k = b_i + b_j + b_k$$

and

$$a_i^2 + a_j^2 + a_k^2 = b_i^2 + b_j^2 + b_k^2.$$

Theorem 3.1

Let $\mathbf{a}$ maximizes f on D_λ and suppose for $i \neq j \neq k$ $a_i \neq a_j \neq a_k$ with at least two of a_i,a_j,a_k are in $(0,1)$. Then

$$(3.11) \qquad f_{n-3,3}(\mathbf{a}^{i,j,k}) \begin{cases} \geq 0 & \text{if one of the coordinates is } 1 \\ = 0 & \text{if } (a_i,a_j,a_k) \in (0,1)^2,\ a_i \neq a_j \neq a_k \\ \leq 0 & \text{if one of the coordinates is } 0. \end{cases}$$

Proof

Let $a_i + a_j + a_k = \alpha$ and $a_i^2 + a_j^2 + a_k^2 = \beta$.

Since f is symmetric with respect to the permutation of a_i,a_j,a_k without loss of generality we will separate the assumptions of the theorem into the following cases.

(i) $(a_i,a_j,a_k) \in (0,1)^2 \times \{1\}$ with $a_i \neq a_j$

(ii) $0 < a_i \neq a_j \neq a_k < 1$

(iii) $(a_i,a_j,a_k) \in (0,1)^2 \times \{0\}$ with $a_i \neq a_j$.

If $\mathbf{a}$ maximizes f on D_λ, then, by (3.6), with $f_{n-3,k}(\mathbf{a}^{ijk})$ abbreviated to f_k, (a_i,a_j,a_k) maximizes

$$F(x,y,z) = xyzf_3 + (xy + yz + zx)f_2 + (x + y + z)f_1$$

on

$$D = [0,1]^3 \cap \{(x,y,z) \mid x^2+y^2+z^2 = a_i^2+a_j^2+a_k^2 \text{ , } x+y+z = a_i+a_j+a_k \} \text{ .}$$

On D, $xy + yz + zx = a_ia_j + a_ja_k + a_ka_i$ and $xyz = (z-a_i)(z-a_j)(z-a_k)$. Hence $F(x,y,z) = G(u) = (u-a_i)(u-a_j)(u-a_k)f_3 + F(a_i,a_j,a_k)$ on $J = \{u \mid u = z \text{ with } z \in D\}$.

Since $a_i \neq a_j$, by Lemma 3.1

$$J = \begin{cases} (a_k - \delta,\ a_k] \\ (a_k - \delta,\ a_k + \delta) \\ [a_k,\ a_k + \delta) \end{cases} \quad \text{if } (a_i,a_j,a_k) \text{ satisfy } \begin{cases} \text{(i)} \\ \text{(ii)} \\ \text{(iii)} \text{ .} \end{cases}$$

Therefore we must have

$$G'(a_k) \begin{cases} \geq 0 \\ = 0 \\ \leq 0 \end{cases} \quad \text{if } a_k \in \begin{cases} \text{right boundary point of } J^0 \\ J^0 \\ \text{left boundary point of } J^0 \text{ .} \end{cases}$$

Since $G'(a_k) = (a_k-a_i)(a_k-a_j)f_3$ we have shown (3.11). □

Theorem 3.2

If $\mathbf{a}$ maximizes f on D_λ then for any $i \neq j$

$$(3.12)\qquad (a_i+a_j-.5)f_{n-2,2}(\mathbf{a}^{i,j}) + f_{n-2,1}(\mathbf{a}^{i,j}) \begin{cases} \leq 0 \\ = 0 \\ \geq 0 \end{cases}$$

$$\text{if} \begin{cases} (a_i,a_j) \text{ or } (a_j,a_i) \in \{0\} \times (.5,1) \cup (.5,1) \times \{1\} \text{ or } 0 < a_i = a_j < .5 \\ a_i \neq a_j \text{ with } (a_i,a_j) \in (0,1)^2 \\ (a_i,a_j) \text{ or } (a_j,a_i) \in \{0\} \times (0,.5) \cup (0,.5) \times \{1\} \text{ or } .5 < a_i = a_j < 1. \end{cases}$$

Proof

Since f is symmetric with respect to the permutation of a_i, a_j without loss of generality we will assume that $a_i \leq a_j$.

If $\mathbf{a}$ maximizes f on D_λ, then by (3.6), with f_k denoting $f_{n-2,k}(\mathbf{a}^{ij})$, (a_i,a_j) maximizes

$$F(x,y) = xyf_2 + (x + y)f_1$$

on

$$D = [0 \leq x \leq y \leq 1] \cap [(x - .5)^2 + (y - .5)^2 = r^2]$$

with

$$r^2 = (a_i - .5)^2 + (a_j - .5)^2 < .5.$$

On D

$$2xy = (x + y)^2 - (x + y) + .5 - r^2$$

so that

$$F(x,y) = G(x + y)$$

with

$$G(z) = .5z^2 f_2 + z(f_1 - .5f_2) + .5(.5 - r^2)f_2$$

on

$$Z = \{z \mid z = x + y \text{ with } (x,y) \in D\} .$$

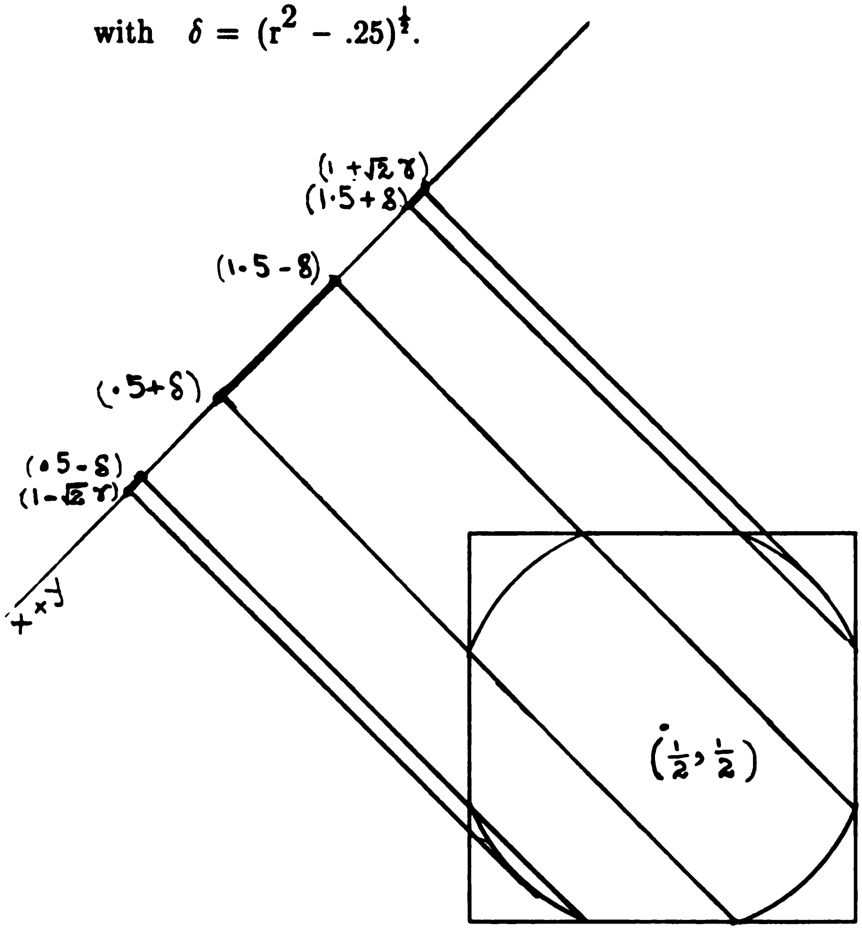
Next we will show that Z is the union of [1 or 3] closed intervals. The line $x + y = z$ is a tangent to the circle $(x - .5)^2 + (y - .5)^2 = r^2$ when $z = 1 + 2^{\frac{1}{2}}r$ and intersects the circle when $z \in (1 - 2^{\frac{1}{2}}r, 1 + 2^{\frac{1}{2}}r)$. Hence if $D \subset [0,1]^2$; that is, if $0 < r^2 \leq .25$, then

$$z \in [1 - 2^{\frac{1}{2}}r, 1 + 2^{\frac{1}{2}}r].$$

If $.25 < r^2 < .5$, then

$$z \in [1 - 2^{\frac{1}{2}}r, .5 - \delta] \cup [.5 + \delta, 1.5 - \delta] \cup [1.5 + \delta, 1 + 2^{\frac{1}{2}}r]$$

with $\delta = (r^2 - .25)^{\frac{1}{2}}$.



Since Z is the union of [1 or 3] closed intervals, with $G'(a_i + a_j)$ denoting $G'(z)\big|_{a_i+a_j}$, we must have

$$G'(a_i + a_j) \begin{cases} \leq 0 \\ = 0 \\ \geq 0 \end{cases}$$

$$\text{if} \quad a_i + a_j \in \begin{cases} \text{left boundary point of } Z^0 \\ Z^0 \\ \text{right boundary point of } Z^0 \ . \end{cases}$$

In terms of $(a_i,a_j) \in D$

$$G'(a_i + a_j) \begin{cases} \leq 0 \\ = 0 \\ \geq 0 \end{cases}$$

$$\text{if} \begin{cases} (a_i,a_j) \in \{0\} \times (.5,1) \cup (.5,1) \times \{1\} \text{ or } 0 < a_i = a_j < .5 \\ 0 < a_i < a_j < 1 \\ (a_i,a_j) \in \{0\} \times (0,.5) \cup (0,.5) \times \{1\} \text{ or } .5 < a_i = a_j < 1. \end{cases}$$

Since $G'(a_i + a_j) = (a_i + a_j - .5)f_2 + f_1$ the proof of the theorem is complete. □

Corollary 3.1

If $\mathbf{a}$ maximizes f on D_λ and if $0 < a_1 \neq a_2 \neq a_3 < 1$, then

$$(3.13) \qquad (a_1 + a_2 + a_3 - .5)\, f_{n-3,2}(\mathbf{a}^{1,2,3}) + f_{n-3,1}(\mathbf{a}^{1,2,3}) = 0 \ .$$

Proof

Since $0 < a_1 \neq a_2 < 1$, by (3.12)

$$(3.14) \qquad (a_1 + a_2 - .5)\, f_{n-2,2}(\mathbf{a}^{1,2}) + f_{n-2,1}(\mathbf{a}^{1,2}) = 0 \ .$$

By (3.5)

$$(3.15) \qquad f_{n-2,2}(\mathbf{a}^{1,2}) = a_3\, f_{n-3,3}(\mathbf{a}^{1,2,3}) + f_{n-3,2}(\mathbf{a}^{1,2,3})$$

and

$$(3.16) \qquad f_{n-2,1}(\mathbf{a}^{1,2}) = a_3\, f_{n-3,2}(\mathbf{a}^{1,2,3}) + f_{n-3,1}(\mathbf{a}^{1,2,3}) .$$

Since $0 < a_1 \neq a_2 \neq a_3 < 1$, by (3.11) $f_{n-3,3}(\mathbf{a}^{1,2,3}) = 0$. Using this fact in (3.15) and then, substituting for $f_{n-2,2}(\mathbf{a}^{1,2})$ and $f_{n-2,1}(\mathbf{a}^{1,2})$ in (3.14), yields (3.13). □

3.4 Characterization of Extrema

Theorem 3.3.

Suppose $\mathbf{a}$ maximizes f on D_λ and, has at least three unequal coordinates in (0,1).

(i) If $\mathbf{b}$ is any other point in D_λ with the same number of zero and unit coordinates as $\mathbf{a}$ and $\sum_{i=1}^{n} b_i = \sum_{i=1}^{n} a_i$, then $\mathbf{b}$ also maximizes f on D_λ.

(ii) If $\mathbf{a}$ has exactly r_1 unit coordinates and $\lambda_1 = \sum_{\{i|0<a_i<1\}} a_i$ then

$$(3.17) \quad f(\mathbf{a}) = .5(\lambda - \lambda_1)^2 \{g(r_1) - 2g(r_1 + 1) + g(r_1 + 2)\} + g(r_1) .$$

Lemma 3.3

Let $3 \leq m \leq n$ and $\mathbf{A_k}$ denote the statement that

$$f_{n-k,i}(a_{k+1},\ldots,a_n) = 0 \qquad \forall\ 3 \leq i \leq k$$

and

$$(\sum_{i=1}^{k} a_i - .5)\, f_{n-k,2}(a_{k+1},\dots,a_n) + f_{n-k,1}(a_{k+1},\dots,a_n) = 0 \qquad \forall\, k \geq 2\ ,$$

if $\mathbf{a}$ is a maximum of f such that $(a_1,a_2,\dots,a_m) \in (0,1)^m$ with at least three unequal coordinates and others 0 or 1.

Then A_k holds if $3 \leq k \leq m$.

Proof

The proof is by induction. Without loss of generality we will assume $a_1 \neq a_2 \neq a_3$.

That A_3 is true follows from (3.11) of Theorem 3.1 and (3.13) of Corollary 3.1.

Assume that A_k is true for a k with $3 \leq k < m$.

Suppose that

$$\mathbf{a}(b_1,\dots,b_k) = (b_1,\dots,b_k,a_{k+1},\dots,a_n)$$

satisfy

(3.18) $$\sum_{i=1}^{k} b_i(1 - b_i) = \sum_{i=1}^{k} a_i(1 - b_i)$$

(3.19) $$\sum_{i=1}^{k} b_i = \sum_{i=1}^{k} a_i\ ,$$

(3.20) $$0 \leq b_i \leq 1 \quad \forall \quad 1 \leq i \leq k\ .$$

Then, from (3.18) and (3.20) it follows that $\mathbf{a}(b_1,\dots,b_k) \in D_\lambda$, from (3.18) and (3.19) we obtain

(3.21) $$\sum_{\{(i,j\,|\,i\neq j\}}^{k} b_i b_j = \sum_{\{(i,j\,|\,i\neq j\}}^{k} a_i a_j\ .$$

From (3.8) with m replaced by k there and the induction hypotheses (i.e. $f_{n-k,i} = 0 \ \forall\ 3 \le i < k$) we obtain

$$(3.22)\quad f(\mathbf{a}) - f(\mathbf{a}(b_1,\dots,b_k)) = \sum_{i=1}^{2} \left[c_{k,i}\begin{vmatrix} a_1,\dots,a_k \\ b_1,\dots,b_k \end{vmatrix}\right] f_{n-k,i}(a_{k+1},\dots,a_n), \quad \forall\ 1 \le i \le k .$$

But (3.19) and (3.21) implies that each of the differences of the symmetric sums in the R.H.S. (3.22) is zero. Hence

$$(3.23)\qquad f(\mathbf{a}) = f(\mathbf{a}(b_1,\dots,b_k)) \text{ for every such } \mathbf{a}(b_1,\dots,b_k) .$$

Since $(a_1,a_2,a_3) \in (0,1)^3$ with $a_1 \ne a_2 \ne a_3$, by Remark 3.1 we can find a $(b_1,b_2,b_3) \in (0,1)^3$ such that

$$b_1 + b_2 + b_3 = a_1 + a_2 + a_3$$

$$b_1^2 + b_2^2 + b_3^2 = a_1^2 + a_2^2 + a_3^2$$

and

$$0 < b_1 \ne b_2 \ne b_3 \ne a_{k+1} < 1 .$$

For such a choice of b_1, b_2, b_3, $\mathbf{b}_3 = (b_1,b_2,b_3,a_4,\dots,a_n) \in D_\lambda$ and satisfies (3.23). Since f is symmetric with respect to the permutation of its coordinates, f will also be maximized by the permuted points $\mathbf{b}_3'$, $\mathbf{b}_3''$ of $\mathbf{b}_3$ given by

$$\mathbf{b}_3' = (a_{k+1},b_2,b_3,a_4,\dots,a_k,b_1,a_{k+2},\dots,a_n)$$
$$\mathbf{b}_3'' = (a_{k+1},b_1,b_3,a_4,\dots,a_k,b_2,a_{k+2},\dots,a_n) .$$

The first k coordinates of $\mathbf{b}_3'$ and $\mathbf{b}_3''$ are in (0,1) with at least three unequal (mainly the first three) coordinates. Hence, by the induction hypotheses , $\forall\ 3 \leq i \leq k$

$$f_{n-k,i}(b_1,a_{k+2},\dots,a_n) = 0$$

and

$$f_{n-k,i}(b_2,a_{k+2},\dots,a_n) = 0 .$$

By (3.5) we can write the above two equations as

$$b_1\ f_{n-k-1,i+1}(a_{k+2},\dots,a_n) + f_{n-k-1,i}(a_{k+2},\dots,a_n) = 0$$

and

$$b_2\ f_{n-k-1,i+1}(a_{k+2},\dots,a_n) + f_{n-k-1,i}(a_{k+2},\dots,a_n) = 0 \quad \forall\ 3 \leq i \leq k.$$

Since $b_1 \neq b_2$, this implies that

$$(3.24) \qquad \begin{cases} f_{n-k-1,i+1}(a_{k+2},\dots,a_n) = 0 \\ \\ f_{n-k-1,i}(a_{k+2},\dots,a_n) = 0 \qquad \forall\ i \leq 3 \leq k. \end{cases}$$

Also by the induction hypotheses

$$(3.25) \qquad (\sum_{i=1}^{k} a_i - .5)\ f_{n-k,2}(a_{k+1},\dots,a_n) + f_{n-k,1}(a_{k+1},\dots,a_n) = 0 .$$

Applying (3.5) for the functions $f_{n-k,2}$ and $f_{n-k,1}$ in (3.25)

$$(3.26)\qquad \begin{aligned}&\Big(\sum_{i=1}^{k} a_i - .5\Big)\, a_{k+1}\, f_{n-k-1,3}(a_{k+2},\dots,a_n)\\ &\quad + \Big(\sum_{i=1}^{k} a_i - .5\Big)\, f_{n-k-1,2}(a_{k+2},\dots,a_n)\\ &\quad + a_{k+1}\, f_{n-k-1,2}(a_{k+2},\dots,a_n) + f_{n-k-1,1}(a_{k+2},\dots,a_n) = 0\ .\end{aligned}$$

By (3.24) the first term in the L.H.S. (3.26) is zero and the rest of the equation simplifies to

$$(3.27)\ \Big(\sum_{i=1}^{k+1} a_i - .5\Big)\, f_{n-k-1,2}(a_{k+2},\dots,a_n) + f_{n-k-1,1}(a_{k+2},\dots,a_n) = 0\ .$$

(3.24) and (3.27) implies $\mathbf{A}_{k+1}$ is true, thus proving the lemma. □

Proof of Theorem 3.3

Suppose $\mathbf{a}$ maximizes f over D_λ with r_0 zero coordinates and $m = n - r_0 - r_1$ coordinates in (0,1). We will take $a_1,\dots,a_m$ to be those coordinates and assume $a_1 \neq a_2 \neq a_3$. Then with $k = m$ in Lemma 3.3, $\mathbf{A}_m$ holds for any $\mathbf{b}$ that satisfies the hypotheses (i) of Theorem 3.3 and by (3.23) $f(\mathbf{a}) = f(\mathbf{b})$, thus proving (i) of Theorem 3.3.

To prove (ii) of Theorem 3.3 we first observe that

$$(3.28)\qquad \lambda = \lambda_1 - \sum_{i=1}^{m} a_i^2$$

and we can put $\mathbf{a}^{1,\dots,m} = (0^{r_0},1^{r_1})$.

Then by $\mathbf{A}_m$ of Lemma 3.3

$$(3.29)\qquad f_{n-m,i}(0^{r_0},1^{r_1}) = 0 \qquad \forall\ 3 \le i \le m$$

and

$$(3.30)\qquad (\sum_{i=1}^{m} a_i - .5)\, f_{n-m,2}(0^{r_0},1^{r_1}) + f_{n-m,1}(0^{r_0},1^{r_1}) = 0\ .$$

Applying (3.29) with (3.6)

$$(3.31)\qquad f(\mathbf{a}) = \sum_{i \neq j} a_i a_j\, f_{n-m,2}(0^{r_0},1^{r_1}) + \sum_{i=1}^{m} a_i\, f_{n-m,1}(0^{r_0},1^{r_1}) + f_{n-m,0}(0^{r_0},1^{r_1}).$$

From (3.28), (3.30) and (3.31) it follows that

$$(3.32)\qquad f(\mathbf{a}) = .5(\lambda - \lambda_1)^2\, f_{n-m,2}(0^{r_0},1^{r_1}) + f_{n-m,0}(0^{r_0},1^{r_1})\ .$$

Applying (3.7) with $p^{1,\ldots,m} = (0^{r_0},1^{r_1})$

$$f_{n-m,i}(0^{r_0},1^{r_1}) = \sum_{h=0}^{i} (-1)^{i-h} \begin{bmatrix} i \\ h \end{bmatrix} f(0^{m-h+r_0},1^{h+r_1})$$

and from (3.1)

$$f(0^{m-h+r_0},1^{h+r_1}) = g(h + r_1).$$

Therefore,

$$(3.33)\qquad f_{n-m,i}(0^{r_0},1^{r_1}) = \sum_{h=0}^{i} (-1)^{i-h} \begin{bmatrix} i \\ h \end{bmatrix} g(h + r_1).$$

(3.33), with $i = 2$ gives

$$f_{n-m,0}(0^{r_0},1^{r_1}) = g(r_1)$$

and, with $i = 2$ gives

$$f_{n-m,2}(0^{r_0},1^{r_1}) = g(r_1) - 2g(r_1 + 1) + g(r_1 + 2) .$$

The appropriate substitutions from the above two equations into (3.32) gives (3.17), thus proving (ii) of Theorem 3.3 . □

Corollary 3.2

Let $D_0 = \{ p \in D_\lambda \mid p_1, p_2, \ldots, p_n$ take at most four different values only two of which are distinct from 0 and 1$\}$.

Then,
$$\underset{p \in D_0}{\text{extremum}}\, f(p) = \underset{p \in D_\lambda}{\text{extremum}}\, f(p) .$$

Proof

Since, Corollary 3.2 is trivially true when $\lambda = 0$ or $\lambda = .25$, without loss of generality we will assume $0 < \lambda < .25$ By (i) of Theorem 3.3 f will also be maximized by any point b with $b^{1,\ldots,m} = (0^{r_0},1^{r_1})$ and $\sum_{i=1}^{m} b_i = \sum_{i=1}^{m} a_i$. In the following we show that there exists such a b of the form $(c^{n-s_0-s_1-1},d,0^{s_0},1^{s_1})$ with $s_0 \geq r_0$ and $s_1 \geq r_1$.

Let $\lambda_1 = \sum_{i=1}^{m} a_i$ so that $\sum_{i=1}^{m} a_i^2 = \lambda_1 - \lambda$.

Since $0 < a_i < 1 \;\; \forall\, 1 \leq i \leq m$, $m > \lambda_1 > \lambda_1 - \lambda > 0$ and $m(\lambda_1 - \lambda) > \lambda_1^2 > \lambda_1 - \lambda > 0$.

For a suitably chosen k, $1 < k \leq m$ we will define

$$(b_1,\ldots,b_m) = (c^k,d,0^{m_0},1^{m_1})$$

where

$$m_0 \geq 0 \;,\; m_1 = m - (m_0 + k + 1)$$
$$c = (\lambda_1 - m_1)(k + 1)^{-1} + \delta, \quad d = c - (k + 1)\delta$$

with

$$\delta = \{(k(k+1))^{-1}(\lambda_1 - m_1 - \lambda - (\lambda_1 - m_1)^2(k+1)^{-1})\}^{\frac{1}{2}}.$$

Then by definition of c and d

(i) $kc + d = \lambda_1 - m_1$

(ii) $kc^2 + d^2 = \lambda_1 - m_1 - \lambda$

(iii) $d < c$

(iv) c and d are real numbers if

$$0 \leq (\lambda_1 - m_1)^2 (\lambda_1 - m_1 - \lambda)^{-1} \leq k + 1$$

(v) $d > 0$ if $k < (\lambda_1 - m_1)(\lambda_1 - m_1 - \lambda)^{-1}$

and (vi) $c < 1$ if $(\lambda_1 - m_1 - k - .5)^2 + \lambda - .25 > 0$.

The conditions in (iv) – (vi) can be met by choosing

$$k = \begin{cases} [\lambda_1^2(\lambda_1 - \lambda)^{-1}] \\ n - 1 \\ 1 \end{cases}, \quad m_1 = \begin{cases} 0 \\ 0 \\ r - 1 \end{cases}, \quad m_0 = \begin{cases} m - k - 1 \\ 0 \\ m - r - 1 \end{cases}$$

if

$$\begin{cases} \lambda \geq .25 \\ m - 1 + (1 - 2\lambda)^{\frac{1}{2}} < \lambda_1 < m \quad \text{with} \quad \lambda < .25 \\ \lambda \vee (r - 1 + (1 - 2\lambda)^{\frac{1}{2}}) < \lambda_1 \leq r + (1 - 2\lambda)^{\frac{1}{2}} \quad \text{with} \quad \lambda < .25 \quad \text{for} \\ \qquad \text{some} \quad 1 \leq r < m - 1. \end{cases}$$

For these choices of k, m_1, m_0 the point $b = (c^k, d, 0^{s_0}, 1^{s_1})$ with $s_0 = r_0 + m_0$ and $s_1 = r_1 + m_1$ is in D_λ thus proving the existence of a point b of the form stated in Corollary 3.2 and hence proving it.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Ballard, Robert John (1974). Extended rules for the sequence compound decision problem with m×n component. Ph.D. dissertation, Dept. of Statist. and Prob., M. S. U.

2. Ballard, Robert John and Gilliland, Dennis C. (1978). On the risk performance of extended sequence compound rules for classification between N(−1,1) and N(1,1). *J. Statist. Comput. Simul.*, **6**, 265–280.

3. Ballard, Robert John, Gilliland, Dennis C. and Hannan, James (1975). $O(N^{-\frac{1}{2}})$ convergence to k–extended Bayes risk in the sequence compound decision problem with m×n components. RM–333, Dept. of Statist. and Prob., M. S. U.

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

5. Gilliland, Dennis C. and Hannan, James (1986). The finite state compound decision decision problem, equivariance and restricted risk components. *Adaptive Statistical Procedures and Related Topics. IMS Lecture Notes – Monograph Series* **8**, 129–145. Also (1974) RM 317, Dept. of Statist. and Prob., M. S. U.

6. 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.

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

8. Hoeffding, Wassily (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Statist.*, **27**, 713–721.

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

10. Oaten, Allen (1972). Approximation to Bayes risk in compound decision problems. *Ann. Math. Statist.*, **43**, 1164–1183.

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

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

13. Van Ryzin, J. R. (1966 a). The compound decision problem with m×n finite loss matrix. *Ann. Math. Statist.*, **37**, 412–424.

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

15. Vardeman, S. (1980). Admissible solutions of k–extended finite set and sequence compound decision problems. *Journ. Multivariate Anal.*, **10**, 426–441.