SOME ADMISSIBILITY CONSIDERATIONS IN THE FINITE STATE COMPONENT COMPOUND AND EMPIRICAL BAYES DECISION PROBLEMS

Dissertation for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

JOHN ELVIN BOYER, JR.

1976

This is to certify that the

thesis entitled

SOME ADMISSIBILITY CONSIDERATIONS IN
THE FINITE STATE COMPONENT COMPOUND AND
EMPIRICAL BAYES DECISION PROBLEMS

presented by

John Elvin Boyer, Jr.

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics and Probability

Dennis C. Gilliland

Major professor

Date August 10, 1976

ABSTRACT

SOME ADMISSIBILITY CONSIDERATIONS IN
THE FINITE STATE COMPONENT COMPOUND AND
EMPIRICAL BAYES DECISION PROBLEMS

By

John Elvin Boyer, Jr.

We consider the compound and empirical Bayes decision problems with finite state component. Relationships between admissibility of a compound rule and the admissibility of the component decision rules it selects are established. Analogous results are obtained in the empirical Bayes decision problem. The main result is the demonstration of an admissible Bayes (Λ) empirical Bayes decision rule which is asymptotically optimal for a large class of Λ.

SOME ADMISSIBILITY CONSIDERATIONS IN THE FINITE STATE COMPONENT COMPOUND AND EMPIRICAL BAYES DECISION PROBLEMS

By

John Elvin Boyer, Jr.

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

1976

ACKNOWLEDGEMENTS

I wish to express my sincerest thanks to Professor Dennis C. Gilliland for his guidance and encouragement in my studies and in the preparation of this thesis. His patience and willingness to discuss any problem at any time are greatly appreciated.

I also wish to thank Professor James Hannan for his critical reading of this thesis and his encouragement as it progressed. Special thanks are due to Mrs. Noralee Burkhardt for her typing of the manuscript, and the patience with which she did it.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1. Component Decision Problem

Consider the following statistical decision problem called the component decision problem. Let $\{P_\theta: \theta \in \Theta\}$ be a family of probability measures over a σ-field $\mathcal{B}$ of subsets of $\mathcal{X}$. Θ is called the parameter or state space. X denotes an $\mathcal{X}$-valued, P_θ-distributed random variable. Let $\mathcal{A}$ denote the action space and $L \geq 0$ the loss function defined on $\Theta \times \mathcal{A}$. Let $\mathcal{C}$ be a σ-field of subsets of $\mathcal{A}$, with respect to which the θ-sections of L are measurable. A (randomized) decision rule t is a function having domain $\mathcal{X} \times \mathcal{C}$ such that each $\mathcal{X}$-section of t is a probability measure on $\mathcal{C}$ and each $\mathcal{C}$-section of t is $\mathcal{B}$-measurable. The risk of t at state θ is

(1) $$R(\theta,t) = \int\int L(\theta,a)t(x,da)P_\theta(dx).$$

Let $\mathcal{G}$ denote a class of probability measures (priors) on $\mathcal{F}$ a σ-field of subsets of Θ with respect to which the t-sections of $R(\theta,t)$ are measurable. The Bayes risk of t at $G \in \mathcal{G}$ is

(2) $$R(G,t) = \int R(\theta,t)G(d\theta).$$

Let T be a specified class of component decision rules. The infimum Bayes risk at G is

$$(3) \qquad R(G) = \inf_{t \in T} R(G,t)$$

and $R(\cdot)$ defined by (3) is called the Bayes envelope. For t^* such that $R(G,t^*) = R(G)$ we write $t^* = t_G$ and call t_G a Bayes rule with respect to G. Note that t_G need not exist and, if it does exist, it need not be unique. However, if it does exist at each G and a minimizer is specified at each G, then the decision rule valued function $t_{(\cdot)}$ defined on $\mathcal{G}$ is called a Bayes response.

This thesis considers the finite Θ case almost exclusively. Here we write $\Theta = \{0,1,\ldots,m\}$ and assume that $P_0,P_1,\ldots,P_m$ are distinct probability measures on $\mathcal{B}$. We write $P_G = \Sigma_0^m G_i P_i$ for $G = (G_0,G_1,\ldots,G_m) \in \mathcal{G}$, here the m-dimensional simplex in R^{m+1}, Euclidean (m+1)-space. We let f_i denote a (bounded) density of P_i with respect to $\mu = \Sigma_0^m P_i$, $i = 0,1,\ldots,m$, whenever densities are used. In the finite Θ case, the risk function $R(\cdot,t)$ defined by (1), is a vector $s = (s^0,s^1,\ldots,s^m)$ in $[0,\infty]^{m+1}$. The collection of all such s is termed the risk set and is denoted by S. The Bayes risk of s at G is $\Sigma_0^m G_i s^i$. If S is closed from below, then t_G exists for every $G \in \mathcal{G}$.

2. Compound Decision Problem.

In the compound decision problem the component problem occurs repeatedly and independently n times presenting a sequence (θ_i, X_i), $i = 1,2,\ldots,n$. The θ_i are unobservable. We write $\underline{\theta} = (\theta_1,\theta_2,\ldots,\theta_n)$ and $\underline{X} = (X_1,X_2,\ldots,X_n) \sim P_{\underline{\theta}} = P_{\theta_1} \times P_{\theta_2} \times \ldots \times P_{\theta_n}$. $\check{\underline{\theta}}_\alpha$ denotes $\underline{\theta}$ with the α-component deleted and, similarly, for $\check{\underline{X}}_\alpha$, $\alpha = 1,2,\ldots,n$. For the class of compound decision rules $\underline{t} = (t_1,\ldots,t_n)$ we take all functions $\underline{t}$ such that for each α in $\{0,1,\ldots,m\}$ t_α is a function on $X^n \times C$ with the property that every $\check{\underline{X}}_\alpha$-section of t_α is a component decision rule (as in section 1) and such that $s_\alpha(\check{\underline{X}}_\alpha) = \iint L(\theta,a)t_\alpha(\underline{X}_n,da)P_\theta(dX_\alpha)$ is a measurable function of $\check{\underline{X}}_\alpha$. Of course $s_\alpha(\check{\underline{X}}_\alpha) = (s_\alpha^0(\check{\underline{X}}_\alpha),\ldots,s_\alpha^m(\check{\underline{X}}_\alpha))$ is then a measurable function of $\check{\underline{X}}_\alpha$ into S. The unconditional component α risk is denoted by

$$(4) \qquad R_\alpha(\underline{\theta},\underline{t}) = \int R(\theta_\alpha, t_\alpha(\check{\underline{x}}_\alpha))P_{\check{\underline{\theta}}_\alpha}(d\check{\underline{x}}_\alpha), \ \alpha = 1,2,\ldots,n$$

and the compound (average risk across components) risk is denoted by

$$(5) \qquad R(\underline{\theta},\underline{t}) = \frac{1}{n}\sum_{\alpha=1}^{n} R_\alpha(\underline{\theta},\underline{t}) \ .$$

In the $\Theta = \{0,1,\ldots,m\}$ case, $s_\alpha^i(\check{\underline{X}}_\alpha) = R(i,t_\alpha(\check{\underline{X}}_\alpha))$, $i = 0,1,\ldots,m$, and the unconditional component α risk is a $P_{\check{\underline{\theta}}_\alpha}$-average of the S-valued function $s_\alpha(\check{\underline{X}}_\alpha)$.

The compound decision problem is invariant under the group of permutations of coordinates. $\underline{t}$ is equivariant if $t_\alpha(\underline{X}) = t(X_\alpha;\check{\underline{X}}_\alpha)$, $\alpha = 1,2,\ldots,n$, where t as written is a symmetric function of its second argument. For such $\underline{t}$, $s_\alpha(\check{\underline{X}}_\alpha) = s(\check{\underline{X}}_\alpha)$, $\alpha = 1,2,\ldots,n$, where s is a symmetric S-valued function on X^{n-1}. Equivariant $\underline{t}$ have compound risk depending on $\underline{\theta}$ only through its empirical distribution which is denoted by $\underline{n} = (n_0, n_1, \ldots, n_m)$ in the $\Theta = \{0,1,\ldots,m\}$ case.

There has been increased concern over the component risk behavior of compound decision rules. In Chapter 2 we explore the relationships between compound admissibility of $\underline{t}$, the admissibility of the $\check{\underline{X}}_\alpha$-section of t_α, $\alpha = 1,2,\ldots,n$ and the admissibility of the $\check{\underline{\theta}}_\alpha$-sections of the risk functions $R_\alpha(\underline{\theta},\underline{t})$, $\alpha = 1,2,\ldots,n$. Examples 1-3 and Theorems 1 and 2 serve to delineate the implications in finite Θ case.

3. Empirical Bayes Decision Problem.

In the empirical Bayes decision problem $\theta_1, \theta_2, \ldots$ are iid G (unknown) distributed random variables and the conditional distribution of $\underline{X} = (X_1, X_2, \ldots)$ given $\underline{\theta} = (\theta_1, \theta_2, \ldots)$ is $P_{\theta_1} \times P_{\theta_2} \times \ldots$. The marginal distribution of $\underline{X}$ is $P_G \times P_G \times \ldots$, hereafter, simply P_G^∞.

In the empirical Bayes problem in the finite Θ case we let $f_G = \Sigma_0^m G_i f_i$ which is a density of P_G with respect to μ. We also assume the linear independence of the densities $f_0, f_1, \ldots, f_m$ in $L_2(\mu)$ ensuring, among other things, the identifiability of the mixtures P_G and the existence of unbiased bounded estimators of G.

Let $\underline{t} = (t_1, t_2, \ldots)$ where for each n, the action selected for component n, is by t_n, where t_n is the n^{th} component of a compound rule, that is t_n is a function on $\mathcal{X}^n \times \mathcal{C}$ such that every $\underline{X}_{n-1}$-section is a component rule and

$$R(\theta, t_n(\underline{X}_{n-1})) = \iint L(\theta, a) t_n(\underline{X}_n, da) P_\theta(dX_n)$$

is a measurable function of $\underline{X}_{n-1}$. Robbins (1956) introduced the empirical Bayes problem showing examples of constructions of empirical Bayes rules whose component n Bayes risks converge to the component minimum Bayes risk R(G), whatever be G.

The conditional on $\check{\underline{X}}_n$ component n Bayes risk of $\underline{t}$ is

$$R(G, t_n(\check{\underline{X}}_n)) = \sum_{i=0}^{m} G_i R(i, t_n(\check{\underline{X}}_n)), \quad n = 1, 2, \ldots \tag{6}$$

and the unconditional component n Bayes risk of $\underline{t}$ is

$$(7) \qquad R_n(G,t_n) = \int R(G,t_n(\check{\underline{x}}_n))P_G^{n-1}(d\check{\underline{x}}_n), \quad n = 1,2,\ldots$$

Of course, for each n and G

$$(8) \qquad R(G) \leq R(G,t_n(\check{\underline{X}}_n))$$

from which $R(G) \leq R_n(G,t_n)$ for all $\underline{t}$, n and G.

An empirical Bayes rule $\underline{t}$ is said to be <u>strongly Bayes risk consistent</u> if

$$R(G,t_n(\check{\underline{X}}_n)) \to R(G) \quad \text{a.s.} \quad [P_G^\infty] \quad \text{for all} \quad G \in \mathcal{G}$$

and simply <u>Bayes risk consistent</u> (usually termed asymptotically optimal) if

$$R_n(G,t_n) \to R(G) \quad \text{for all} \quad G \in \mathcal{G}.$$

Chapter 3 concerns the empirical Bayes case. In section 1, Theorem 3 and Example 4 delineate relationships between the admissibility of t_n as a decision rule for decisions concerning G and the admissibility of the $\check{\underline{X}}_n$-sections in the component decision problem. Theorem 4 shows that for equivariant compound $\underline{t}$, the empirical Bayes admissibility of t_n implies the compound admissibility of $\underline{t}$. Theorem 5 shows that admissible empirical Bayes rules result from playing Bayes versus a second level prior Λ on $\mathcal{G}$.

In Section 2 of Chapter 3, Lemma 1 exposes the structure of Bayes (Λ) empirical Bayes procedures. This lemma together with a series of lemmas in Section 3 culminate in a proof of the main result (Theorem 7), namely, the strong Bayes risk consistency of the Bayes (Λ) empirical Bayes procedure, provided each point $G \in \mathcal{G}$ is a point of support of Λ.

CHAPTER 2

RELATIONSHIPS BETWEEN COMPOUND AND COMPONENT ADMISSIBILITY

1. Introduction.

In this chapter we consider some admissibility questions in the compound decision problem. In particular, we explore the relationship between admissibility of a compound decision procedure and the admissibility of the risk functions it selects for component decisions. We define three criteria which we are interested in comparing.

Definition 1. (A) A compound rule $\underline{t}$ is admissible if there does not exist a compound rule $\underline{t}^*$ such that $R(\underline{\theta},\underline{t}^*) \le R(\underline{\theta},\underline{t})$ for all $\underline{\theta}$ in Θ^n and $R(\underline{\theta},\underline{t}^*) < R(\underline{\theta},\underline{t})$ for some $\underline{\theta}$ in Θ^n.

Definition 2. (CA) A compound rule $\underline{t}$ is component admissible if, for every α in $\{1,2,\ldots,n\}$, every $\check{\underline{\theta}}_\alpha$-section of $R_\alpha(\underline{\theta},\underline{t})$ is an admissible risk function in the component.

Definition 3. (CCA) A compound rule $\underline{t}$ is conditionally component admissible, if for each $\alpha = 1,2,\ldots,n$ and $\check{\underline{\theta}}_\alpha$ in Θ^{n-1}, almost all $[P_{\check{\underline{\theta}}_\alpha}]$, $\check{\underline{x}}_\alpha$-sections of t_α are admissible decision rules in the component.

Of course, (A) is the standard admissibility definition applied to compound decision rules $\underline{t}$. Gilliland and Hannan (1974) discuss the restricted compound decision problem wherein only compound rules taking values in a specified restricted component risk set are considered. For example, if the component risk set is restricted to the admissible risk functions, then all $\underset{\sim}{\check{X}}_\alpha$-sections of t_α are admissible decision rules in the component, $\alpha = 1,2,\ldots,n$. In the risk function (S-game) notation of Gilliland and Hannan, the risk function $s_\alpha(\underset{\sim}{\check{X}}_\alpha)$ corresponding to the $\underset{\sim}{\check{X}}_\alpha$-section of t_α is an admissible risk function for all $\underset{\sim}{\check{X}}_\alpha$, $\alpha = 1,2,\ldots,n$. Our definition of (CCA) imposes this condition for almost every $[P_{\underset{\sim}{\check{\theta}}_\alpha}]$ $\check{X}_\alpha$. The unconditional component α risk is the average $R_\alpha(\underline{\theta},\underline{t}) = \int s_\alpha(\underset{\sim}{\check{x}}_\alpha)P_{\underset{\sim}{\check{\theta}}_\alpha}(d\underset{\sim}{\check{x}}_\alpha)$ and (CA) requires that $\underset{\sim}{\check{\theta}}_\alpha$-sections be admissible risk functions.

Any simple rule $\underline{t}$ where $t_1,t_2,\ldots,t_n$ are admissible component decision functions is both (CA) and (CCA). The Stein example of the inadmissibility of the usual estimator of the mean of a multivariate normal, squared error loss, $n \geq 3$, shows that, in general, (A) need not follow from (CA) and (CCA).

Copas (1974) has established necessary conditions for the admissibility of an equivariant compound decision rule when the component is a 2-state, 2-act decision problem. One such necessary condition is that the rule be cut-point in nature, that is, that there exist a symmetric function λ^* so that component decisions 0 and 1 are made according to whether $(f_1(x_\alpha)/f_0(x_\alpha)) < \lambda^*(\underline{x})$ or $(f_1(x_\alpha)/f_0(x_\alpha)) > \lambda^*(\underline{x})$ where f_1/f_0 is the likelihood ratio in the component. In the next section we establish a necessary condition and a sufficient condition for (A) for some finite state components.

Our definitions (A), (CA), (CCA) refer to fixed n. Usually the term compound decision rule refers to a specification of decision procedures for each $n = 1,2,\ldots$ so that (A), (CA), (CCA) could then be required for every n.

2. Relationships Between Compound and Component Admissibility.

We now summarize the results concerning (A), (CA) and (CCA) to be established in this section for finite state components.

Theorem 2

$$\begin{array}{ccc} (CA) & \Rightarrow & (A) \\ & \searrow & \Downarrow \text{ Theorem 1} \\ & \text{Remark 1} & (CCA) \end{array}$$

The implication $(CA) \Rightarrow (CCA)$ will be seen to be trivial. The implication $(CA) \Rightarrow (A)$ will be proved for two-state components and equivariant $\underline{t}$ only. All other pairwise implications do not hold as shown by Examples 1-3.

Example 1. $(CCA) \not\Rightarrow (A)$. Consider the 2×2 component testing problem used by Robbins (1951) to introduce compound decision theory. Here P_θ is $N(2\theta - 1,1)$, $\theta \in \{0,1\}$. The compound decision rule exhibited for this problem by Robbins (1951, (37)) is $\underline{t}$ where

$$t_\alpha(\underline{x}) = \begin{cases} 1, & x_\alpha > c(\underline{x}) \\ 0, & x_\alpha < c(\underline{x}) \end{cases}, \quad \alpha = 1,2,\ldots,n,$$

and c is the symmetric function

$$c(\underline{x}) = \begin{cases} \infty, & \bar{x} \le -1 \\ \frac{1}{2}\ln\left(\frac{1-\bar{x}}{1+\bar{x}}\right), & -1 < \bar{x} < 1 \\ -\infty, & \bar{x} \ge 1 \end{cases} .$$

Each $\check{\underline{x}}_\alpha$-section of c is a decreasing function of x_α when finite valued. It follows that each $\check{\underline{x}}_\alpha$-section of t_α is a monotone rule and, therefore, Bayes versus some component prior putting positive mass on each element of Θ. Hence, $\underline{t}$ is (CCA).

Several authors including Robbins (1951, p. 138), Hannan and Robbins (1955, §8) and Huang (1972, p. 350) have indicated that $\underline{t}$ is not admissible if $n \geq 2$. Robbins states that inadmissibility follows since $\underline{t}$ is not of the form of a Bayes compound rule, but a rigorous demonstration that $\underline{t}$ is inadmissible has not appeared in the literature. One such does exist based on the fact that there does not exist a Bayes compound rule equal to $\underline{t}$ a.e. [Lebesgue in R^n]. This demonstration is found in Appendix A for the case $n = 2$.

<u>Example 2</u>. (CCA) $\not\Rightarrow$ (CA). Consider the compound decision problem and rule $\underline{t}$ of Example 1. As indicated there, every $\check{\underline{x}}_\alpha$-section of t_α has component risk $s_\alpha(\check{\underline{x}}_\alpha)$ which is an admissible risk point of the component risk set S. Hence, s_α takes values on the lower boundary B of S. Here the likelihood ratio f_1/f_0 has a continuous distribution under both P_0 and P_1 so by Appendix B, B is strictly convex. Also, we see that s_α is not equal to a constant almost everywhere $[P_{\check{\underline{\theta}}_\alpha}]$ so that the $\check{\underline{\theta}}_\alpha$-section of $R_\alpha(\underline{\theta},\underline{t})$ being a $P_{\underline{\theta}_\alpha}$-average of s_α belongs to the interior of S and, therefore, is not admissible. Hence $\underline{t}$ is not (CA).

<u>Example 3</u>. (A) $\not\Rightarrow$ (CA). Again we consider the Robbins' component of Example 1 but here we take $\underline{t}$ to be an equivariant Bayes rule with respect to the diffuse prior (see Robbins (1951, §3)). Gilliland and Hannan (1974, p. 11) show in the setting of a general

finite state component that a version of $\underline{t}$ is delete bootstrap in form, that is, each $\check{\underline{x}}_\alpha$-section of t_α is Bayes versus some $\underline{w}(\check{\underline{x}}_\alpha)$ on $\Theta = \{0,1,\ldots,m\}$, $\alpha = 1,2,\ldots,n$. For the Robbins component such a $\underline{t}$ is (CCA) and, by the strict convexity of the lower boundary B as indicated in Example 2, such a $\underline{t}$ is not (CA).

Remark 1. Suppose $\Theta = \{0,1,\ldots,m\}$ and that the risk set S is convex. That CA $\Rightarrow$ CCA is trivial because the points selected as $\underline{\theta}_\alpha$-sections for the CA rules must be on the lower boundary of the risk set S. These points are probability-weighted averages of the $s(\check{\underline{X}}_\alpha)$, and the only way for that average to be on the lower boundary of the convex set S, is for the $s(\check{\underline{X}}_\alpha)$ to be on the lower boundary a.e. $[P_{\check{\underline{\theta}}_\alpha}]$, i.e. to be CCA.

Theorem 1. Suppose that $\Theta = \{0,1,\ldots,m\}$ and that the risk set S is a compact convex subset of $[0,\infty)^{m+1}$. Then (A) $\Rightarrow$ (CCA).

Proof. Suppose $\underline{t}$ is not (CCA). Then there exists an α, $\check{\underline{\theta}}_\alpha^o$ and a set $C \subset X^{n-1}$ such that $P_{\check{\underline{\theta}}_\alpha^o}(C) > 0$, and, for all $\check{\underline{x}}_\alpha \in C$, $s_\alpha(\check{\underline{x}}_\alpha)$ is not in B, the lower boundary of S. Therefore, for each $\check{\underline{x}}_\alpha \in C$ there exists $s_\alpha^*(\check{\underline{x}}_\alpha) \in S$ such that $s_\alpha^*(\check{\underline{x}}_\alpha) < s_\alpha(\check{\underline{x}}_\alpha)$ where for vectors a and b, $a < b$ means that $a^i \leq b^i$ for all i with $a^i < b^i$ for some i. Extend the domain of definition of s_α^* to all of X^{n-1} by defining $s_\alpha^*(\check{\underline{x}}_\alpha) = s_\alpha(\check{\underline{x}}_\alpha)$ for $\check{\underline{x}}_\alpha \notin C$. Now let $C_i = \{\check{\underline{x}}_\alpha : s_\alpha^{*i}(\check{\underline{x}}_\alpha) < s_\alpha^i(\check{\underline{x}}_\alpha)\}$, $i = 0,1,\ldots,m$ and note that $C = \cup C_i$. Hence, at least one of the C_i has positive probability which we suppose to be C_0 without loss of generality. Note that

(i) $s_\alpha^{*i}(\check{\underline{x}}_\alpha) \leq s_\alpha^i(\check{\underline{x}}_\alpha)$ for all $i \in \Theta$, $\check{\underline{x}}_\alpha \in X^{n-1}$

(ii) $s_\alpha^{*0}(\check{\underline{x}}_\alpha) < s_\alpha^0(\check{\underline{x}}_\alpha)$ for $\check{\underline{x}}_\alpha \in C_0$

(iii) $P_{\check{\underline{\theta}}^o}(C_0) > 0$.

Now consider $\underline{t}^* = (t_1,\ldots,t_{\alpha-1},t_\alpha^*,t_{\alpha+1},\ldots,t_n)$ where t_α^* is a decision rule whose $\check{\underline{x}}_\alpha$-section has risk point $s_\alpha^*(\check{\underline{x}}_\alpha)$ for all $\check{\underline{x}}_\alpha$. Then $R(\underline{\theta},\underline{t}^*) \leq R(\underline{\theta},\underline{t})$ for all $\underline{\theta} \in \Theta^n$ and $R(\underline{\theta}^o,\underline{t}^*) < R(\underline{\theta}^o,\underline{t})$ where $\underline{\theta}^o = (\theta_1^o,\ldots,\theta_{\alpha-1}^o,0,\theta_{\alpha+1}^o,\ldots,\theta_n^o)$.

That measurable s_α^* can be chosen ensuring the existence of t_α^* is demonstrated by the following argument. For a point s in S that is not in B, let $c(s)$ be the real number such that $s - c(s)\cdot\underline{1}$ is in B where $\underline{1}$ is the vector $\langle 1,1,\ldots,1\rangle$. Such a $c(s)$ exists because of the compactness and convexity of S. Then let $s_\alpha^*(\check{\underline{X}}_\alpha) = s_\alpha(\check{\underline{X}}_\alpha)$ for $\check{\underline{X}}_\alpha$ not in C and $s_\alpha^*(\check{\underline{X}}_\alpha) = (s - c(s)\cdot 1)\ (\check{\underline{X}}_\alpha)$ for $\check{\underline{X}}_\alpha$ in C. The $s_\alpha^*(\check{\underline{X}}_\alpha)$ is then a measurable function of $\check{\underline{X}}_\alpha$ since $s_\alpha(\check{\underline{X}}_\alpha)$ is measurable. □

Whereas (CCA) is a necessary condition for (A), our next result shows that (CA) is a sufficient condition for (A) for equivariant $\underline{t}$ in the two-state component case. A proof or counter example for the general finite state component has not been found. The (CA) condition is a very strong condition and, as can be seen in Examples 2 and 3, when the lower boundary of S is strictly convex, only simple rules are (CA).

Theorem 2. Suppose that $\Theta = \{0,1\}$. If $\underline{t}$ is an equivariant rule which is (CA), then $\underline{t}$ is (A).

<u>Proof</u>. The compound risk of equivariant $\underline{t}$ depends on $\underline{\theta}$ only through the number n_1 of $\theta_\alpha = 1$, $\alpha = 1,2,\ldots,n$. Displaying this through the notation $R(\underline{\theta},\underline{t}) = R((n_0,n_1),\underline{t})$ where $n_0 = n - n_1$, we have

$$\begin{aligned}
nR((n_0,n_1),\underline{t}) &= \sum_{\alpha=1}^{n} R_\alpha((n_0,n_1),\underline{t}) \\
&= \sum_{\alpha=1}^{n} [\theta_\alpha = 0]R_\alpha((n_0,n_1),\underline{t}) + \sum_{\alpha=1}^{n} [\theta_\alpha = 1]R_\alpha((n_0,n_1),\underline{t}) \\
&= n_0 \int s^0(\underline{\check{x}}_n)(P_0^{n_0-1} \times P_1^{n_1})(d\underline{\check{x}}_n) \\
&\qquad + n_1 \int s^1(\underline{\check{x}}_n)(P_0^{n_0} \times P_1^{n_1-1})(d\underline{\check{x}}_n) \\
&= n_0 R^0((n_0 - 1,n_1),\underline{t}) + n_1 R^1((n_0,n_1-1),\underline{t})
\end{aligned}$$

where R^0 and R^1 are defined by position in the last line. Then

$$(*)\quad \begin{cases} R((0,n),\underline{t}) = R^1((0,n-1),\underline{t}) \\ R((1,n-1),\underline{t}) = \frac{1}{n} R^0((0,n-1),\underline{t}) + \frac{n-1}{n} R^1((1,n-2),\underline{t}) \\ \quad\vdots \\ R((n,0),\underline{t}) = R^0((n-1,0),\underline{t}). \end{cases}$$

Suppose that $\underline{t}$ is not (A), but $\underline{t}$ is (CA). Since $\underline{t}$ is not (A), then there is an equivariant $\underline{t}^*$ such that $R((j,n-j),\underline{t}^*) \leq R((j,n-j),\underline{t})$ for all $j = 0,1,\ldots,n$, with $R((j,n-j),\underline{t}^*) < R((j,n-j),\underline{t})$ for some such j, say j_0. On the other hand, since $\underline{t}$ is (CA), if, for any j, $R^0((j,n-1-j),\underline{t}^*) <$ $R^0((j,n-1-j),\underline{t})$, then $R^1((j,n-1-j),\underline{t}^*) > R^1((j,n-1-j),\underline{t})$ and similarly, if for any j, $R^1((j,n-1-j),\underline{t}^*) < R^1((j,n-1-j),\underline{t})$ then $R^0((j,n-1-j),\underline{t}^*) > R^0((j,n-1-j),\underline{t})$.

Since $R((j_0,n-j_0),\underline{t}^*) < R((j_0,n-j_0),\underline{t})$ then either $R^0((j_0-1,n-j_0),\underline{t}^*) < R^0((j_0-1,n-j_0),\underline{t})$ or $R^1((j_0,n-1-j_0),\underline{t}^*) < R^1((j_0,n-1-j_0),\underline{t})$. Suppose that $R^0((j_0-1,n-j_0),\underline{t}^*) < R^0((j_0-1,n-j_0),\underline{t})$. (The proof with the other assumption is exactly analogous.) Then, since $\underline{t}$ is (CA), $R^1((j_0-1,n-j_0),\underline{t}^*) > R^1((j_0-1,n-j_0),\underline{t})$. Because $R((j_0-1,n-j_0+1),\underline{t}^*) \leq R((j_0-1,n-j_0+1),\underline{t})$ we also get $R^0((j_0-2,n-j_0+1),\underline{t}^*) < R^0((j_0-2,n-j_0+1),\underline{t})$. Then, again because $\underline{t}$ is (CA), $R^1((j_0-2,n-j_0+1),\underline{t}^*) > R^1((j_0-2,n-j_0+1),\underline{t})$. Proceeding inductively until the first argument is zero, we get $R^1((0,n-1),\underline{t}^*) > R^1((0,n-1),\underline{t})$, or that $R((0,n),\underline{t}^*) > R((0,n),\underline{t})$, which contradicts the assumption that $R((j,n-j),\underline{t}^*) \leq R((j,n-j),\underline{t})$ for all j. Thus if $\underline{t}$ is not (A), $\underline{t}$ is not (CA). □

Theorem 2 concerns $m = 1$, the two-state problem. The following numerical example shows that the method of proof used for $m = 1$ will not work for $m = 2$ and $n = 2$; thus a new proof must be devised if Theorem 2 is to be generalized to arbitrary finite m and n. Using notation exactly analogous to that in the proof of Theorem 2 suppose

$$(R^0((1,0,0),\underline{t}),\ R^1((1,0,0),\underline{t}),\ R^2((1,0,0),\underline{t})) = (6,2,2)$$

$$(R^0((0,1,0),\underline{t}),\ R^1((0,1,0),\underline{t}),\ R^2((0,1,0),\underline{t})) = (2,6,2)$$

$$(R^0((0,0,1),\underline{t}),\ R^1((0,0,1),\underline{t}),\ R^2((0,0,1),\underline{t})) = (2,2,6)\ .$$

Further suppose

$$(R^0((1,0,0),\underline{t}^*),\ R^1((1,0,0),\underline{t}^*),\ R^2((1,0,0),\underline{t}^*)) = (5,4,0)$$

$$(R^0((0,1,0),\underline{t}^*),\ R^1((0,1,0),\underline{t}^*),\ R^2((0,1,0),\underline{t}^*)) = (0,6,4)$$

$$(R^0((0,0,1),\underline{t}^*),\ R^1((0,0,1),\underline{t}^*),\ R^2((0,0,1),\underline{t}^*)) = (4,0,6)$$

Using the analog to (*), $R((2,0,0),\underline{t}^*) = 5 < 6 = R((2,0,0),\underline{t})$ and for all other possible $\underline{\theta}$, $R(\underline{\theta},\underline{t}^*) = R(\underline{\theta},\underline{t})$. Thus, even though $\underline{t}$ may well be (CA) (it is not dominated by t^*) it is not (A).

CHAPTER 3

ADMISSIBLE (BAYES) SOLUTIONS TO THE EMPIRICAL BAYES PROBLEM

1. Admissibility and Component Admissibility in the Empirical Bayes Problem.

The empirical Bayes decision problem was introduced in Section 3 of Chapter 1. An empirical Bayes decision rule $\underline{t}$ involves specification of a decision procedure t_n for use in component n, $n \geq 1$.

Definition 4. (A) t_n is admissible if there does not exist a t_n^* such that $R_n(G,t_n^*) \leq R_n(G,t_n)$ for all $G \in \mathcal{G}$ with $R_n(G,t_n^*) < R_n(G,t_n)$ for some $G \in \mathcal{G}$.

Definition 5. (CCA) t_n is conditionally component admissible if for each $G \in \mathcal{G}$, almost all $[P_G^{n-1}]$ $\underline{\check{x}}_n$-sections of t_n are admissible decision rules in the component.

We refer to an empirical Bayes rule $\underline{t}$ as (A) or (CCA) if t_n is (A) or (CCA) for every $n \geq 1$.

Meeden (1972, p. 97) defines admissibility (A) of $\underline{t}$ as above. In his Section 3 he demonstrates the inadmissibility of some classical empirical Bayes procedures for certain linear loss testing and squared error loss estimation components having discrete exponential distributions. The demonstrations exploit the non-(CCA) property of these classical empirical Bayes procedures.

Recently, Van Houwelingen (1973, 1976) has demonstrated (CCA) rules for the linear loss testing component which have improved rates of Bayes risk consistency. Gilliland and Hannan (1976) have extended some of these results to the general multiple decision problem component.

Out next theorem shows that (CCA) is implied by (A) in the finite state, compact convex risk set component empirical Bayes problem.

Theorem 3. Suppose that $\Theta = \{0,1,\ldots,m\}$ and that the risk set S is a compact convex subset of $[0,\infty)^{m+1}$. If $\underline{t}$ is an (A) empirical Bayes procedure, then it is (CCA).

Proof. Suppose $\underline{t}$ is not (CCA). Then there exists an n, $H \in \mathcal{G}$ and a set $C \subset \mathcal{X}^{n-1}$ such that $P_H^{n-1}(C) > 0$, and, for all $\check{\underline{X}}_n \in C$, $s_n(\check{\underline{X}}_n)$ does not belong to B, the lower boundary of S. Since the uniform prior $U = (\frac{1}{m+1},\ldots,\frac{1}{m+1})$ dominates every G so that P_U^{n-1} dominates every P_G^{n-1} we have $P_U^{n-1}(C) > 0$. Let t_n^* be t_α^* as constructed in the proof of Theorem 1 (Section 2 of Chapter 2). Since $s_n^* \le s_n$ and $s_n^* < s_n$ on C it follows that $E_G^{n-1} s_n^*(\check{\underline{X}}_n) \le E_G^{n-1} s_n(\check{\underline{X}}_n)$ for all $G \in \mathcal{G}$ and $E_U^{n-1} s_n^*(\check{\underline{X}}_n) < E_U^{n-1} s_n(\check{\underline{X}}_n)$. Since $R_n(G,t_n) = \Sigma_0^m G^i E_G^{n-1} s_n(\check{\underline{X}}_n)$ for $G \in \mathcal{G}$, and similarly for t_n^*, and since U puts positive mass on every coordinate, it follows that $R(G,t_n^*) \le R_n(G,t_n)$ for all $G \in \mathcal{G}$ and $R_n(U,t_n^*) < R_n(U,t_n)$. □

Our next result shows that for finite Θ and equivariant $\underline{t}$, the empirical Bayes admissibility of t_n (Definition 4) implies the compound admissibility of $\underline{t}$ (Definition 1).

Theorem 4. Suppose that $\Theta = \{0,1,\ldots,m+1\}$. Let $\underline{t}$ be an equivariant compound decision rule. If t_n is empirical Bayes admissible, then $\underline{t}$ is compound admissible.

Proof. The compound risk of $\underline{t}$ at $\underline{\theta}$ is a function of $\underline{\theta}$ through $\underline{n} = (n_0, n_1, \ldots, n_m)$ where for $i = 0,1,\ldots,n$, n_i is the number of $\theta_\alpha = i$, $\alpha = 1,2,\ldots,n$. Let the compound risk of $\underline{t}$ at $\underline{\theta}$ be denoted by $R(\underline{n},\underline{t})$. If $\underline{t}$ is inadmissible, there exists a $\underline{t}^*$ such that $R(\underline{\theta},\underline{t}^*) \leq R(\underline{\theta},\underline{t})$ for all $\underline{\theta} \in \Theta^n$ with $R(\underline{\theta}^o,\underline{t}^*) < R(\underline{\theta}^o,\underline{t})$ for some $\underline{\theta}^o \in \Theta^n$. By the finiteness of the group of $n!$ permutations and Theorem 4.3.2 of Ferguson (1967), $\underline{t}^*$ can be taken to be equivariant. Therefore, it follows that $R(\underline{n},\underline{t}^*) \leq R(\underline{n},\underline{t})$ for all $\underline{n}$ with $R(\underline{n}^o,\underline{t}^*) < R(\underline{n}^o,\underline{t})$. Let G be any distribution on Θ. The G^n-average of $R_\alpha(\underline{\theta},\underline{t})$ is constant with respect to α for equivariant $\underline{t}$ so that $R_n(G,t_n^*) \leq R_n(G,t_n)$. Furthermore, if H^n puts positive mass on $\underline{n}^o$, then $R_n(H\ ,t_n^*) < R_n(H\ ,t_n)$. Therefore, t_n is inadmissible according to Definition 4. □

Example 4. Empirical Bayes (CCA) $\not\Rightarrow$ (A). Consider the Robbins component and bootstrap rule $\underline{t}$ of Example 1. As demonstrated there, each $\check{\underline{x}}_n$-section of t_n is an admissible decision rule in the component. Hence, t_n is (CCA) according to Definition 5. However, $\underline{t}$ is equivariant and, as shown in Appendix A, $\underline{t}$ is inadmissible in the compound problem if $n \geq 2$. Therefore, by Theorem 4, t_n is not (A).

As seen by example, conditional component admissibility is not sufficient for the admissibility of an empirical Bayes procedure. The following observation leads to a characterization of empirical Bayes admissibility with which it is easy to show that certain Bayes empirical Bayes procedures are admissible. Let n be given and consider the statistical decision problem with states $G \in \mathcal{G}$, observation $\check{\underline{X}} = (X_1,\ldots,X_{n-1}) \sim P_G^{n-1}$, actions $d \in \mathcal{D}$, where $\mathcal{D}$ is class of all component decision rules, decision rules t which are $\check{\underline{X}}$-measurable mappings into $\mathcal{D}$, and loss function $R(G,d)$. The risk of t is $E_G^{n-1}R(G,t(\check{\underline{X}})) = R_n(G,t_n)$ where t_n is the empirical Bayes rule $t_n(X_1,\ldots,X_n) = (t(\check{\underline{X}}))(X_n)$. Thus, t_n is empirical Bayes admissible (Definition 4) if and only if t is admissible in the usual sense in the decision problem $(\mathcal{G},\mathcal{D},R,R_n)$ defined above.

Theorem 5. Let $\Theta = \{0,1,\ldots,m\}$, $S \subset [0,\infty)^{m+1}$, and suppose that $n \geq 2$ is given. Let Λ be a prior distribution on $\mathcal{G}$, the m-dimensional simplex of distributions on Θ, and suppose that the support of Λ is all of $\mathcal{G}$. If t is a decision rule in the decision problem $(\mathcal{G},\mathcal{D},R,R_n)$ which is Bayes with respect to Λ and $\int R_n(G,t)\Lambda(dG)$ is finite, then $t_n = t$ is an admissible empirical Bayes decision rule (Definition 4).

Proof. The proof used by Ferguson (1967, Theorem 2.3.3) covers the present situation. Here $\mathcal{G}$ is the m-dimension simplex in R^{m+1} whereas the parameter set is the real line in Theorem 2.3.3. Here the decision problem $(\mathcal{G},\mathcal{D},R,R_n)$ has risk functions $E_G^{n-1}R(G,t(\check{\underline{X}})) = E_G^{n-1}(\Sigma_0^m G_i R(i,t(\check{\underline{X}})))$ which are continuous functions of G. $\square$

2. Bayes Procedures in the Finite State Component Empirical Bayes Problem.

One can regard the empirical Bayes decision problem as a classical decision problem with parameter $G \in \mathcal{G}$. Therefore, it is not surprising that Bayesian decision rules have been suggested for use in the empirical Bayes problem, for example, Lindley (1971, §12.1), Tucker (1963), Meeden (1972) and Shapiro (1974). Tucker (1963) and Meeden (1972) for certain infinite state component decision problems have demonstrated the (empirical) Bayes risk consistency of Bayes procedures. Since asymptotic optimality in empirical Bayes problems is Bayes risk consistency (at every G), the Tucker and Meeden Bayes rules "solve" certain empirical Bayes problems.

Robbins (1951, §3) conjectured that Bayes procedures might be solutions to the compound decision problem. Since for equivariant $\underline{t}$, compound risk convergence to the simple envelope for every $\underline{\theta}$ in the compound problem implies that t_n is (empirical) Bayes risk consistent (for every G), as observed by Gilliland and Hannan (1974), Robbins' conjecture has implications to the Bayesian approach to empirical Bayes. In fact the two state results of Gilliland, Hannan and Huang (1974) demonstrate the (empirical) Bayes risk consistency for a large class of priors and two state components. Shapiro (1974) considered a testing component and investigated the asymptotic properties of the Bayes procedures and the average loss across component decisions.

Rolph (1968) and Ferguson (1974) have investigated the problems of placing prior distributions on classes of distributions

G. This problem is trivial in the case $\Theta = \{0,1,\ldots,m\}$ where G is a subset of Euclidean $(m+1)$-space. This problem is difficult for non-finite Θ, particularly, when tractable Bayes procedures are sought in order to make the consistency question tractable. Ferguson (1974) claims there are at least two desirable characteristics of such prior distributions: (1) the support of the prior with respect to some suitable topology on the space of probability measures should be large, and (2) the posterior distribution given a sample from the true probability measure should be manageable analytically. The resulting Bayes rules then are desirable since they are generally admissible and have nice large sample properties.

Throughout this chapter $\Theta = \{0,1,\ldots,m\}$, S is a compact subset of $[0,\infty)^{m+1}$, G is the m-dimensional simplex of probability distributions $G = (G_0, G_1, \ldots, G_m)$ on Θ and Λ is a probability distribution on G. The conditional on $\underset{\sim}{\check{X}}_n$ risk of the empirical Bayes rule $t = t_n$ at G is

$$(10) \qquad R(G,t) = \sum_{i=0}^{m} G_i R(i,t)$$

and the (empirical) Bayes risk at G is

$$(11) \qquad R_n(G,t) = E_G^{n-1} R(G,t).$$

The "second level Bayes risk" of t with respect to Λ is

$$(12) \qquad B(\Lambda,t) = \int R_n(G,t)\Lambda(dG).$$

Lemma 1. $B(\Lambda,t)$ is minimized by $t = t_{G^*}$ where $G^* = E[G|\underset{\sim}{\check{X}}_n]$ and $t_{(\cdot)}$ denotes a Bayes response in the component problem.

<u>Proof</u>. Let P denote marginal distribution of $\check{\underline{X}}_n$, hereafter denoted by $\check{\underline{X}}$, and let Q denote the conditional distribution of G given $\check{\underline{X}}$. By (12) and (11),

$$B(\Lambda,t) = \iint R(G,t(\check{\underline{x}}))Q(dG)P(d\check{\underline{x}}). \tag{13}$$

Substituting (10) and interchanging integral and finite sum we obtain

$$\begin{aligned} B(\Lambda,t) &= \int \sum_{i=0}^{m} \int G_i Q(dG)\; R(i,t\;(\check{\underline{x}}))P(d\check{\underline{x}}) \\ &= \int R(G^*,t(\check{\underline{x}}))P(d\check{\underline{x}}) \end{aligned} \tag{14}$$

which is minimized by the choice $t(\check{\underline{x}}) = t_{G^*}$. □

It is interesting to note that the result of Lemma 5 can also be obtained as a corollary to results of Gilliland and Hannan (1974). They show on pp. 10-11 that a Bayes compound rule versus a symmetric prior β on Θ^n in the compound decision problem has $t_\alpha(\check{\underline{x}}_\alpha) = t_{\underline{w}(\check{\underline{x}}_\alpha)}$, $\alpha = 1,2,\ldots,n$ where $\underline{w} = (\underline{w}_0,\underline{w}_1,\ldots,\underline{w}_n)$ is given by their (40), namely

$$\underline{w}_i = \sum_{\underline{n}} \beta_{\underline{n}} n_i \Big(\mathop{X}_{j=0}^{m} f_j^{n_{ji}} \Big)^*, \; i = 0,1,\ldots,m\;. \tag{15}$$

Here the sum ranges over all $\binom{n+m}{m}$ empirical distributions $\underline{n}$ of $\underline{\theta}$, $f_j = dP_j/d\mu$, $*$ denotes symmetrization, and $n_{ji} = n_i$ if $j \neq i$, $n_{ji} = n_i - 1$ if $j = i$, $i,j = 0,1,\ldots,m$. Consider $\beta = \beta(\Lambda)$ defined by

$$\beta_{\underline{n}}(\Lambda) = \int \binom{n}{\underline{n}} \Big(\prod_{j=0}^{m} G_j^{n_j} \Big)\Lambda(dG); \tag{16}$$

Gilliland and Hannan (1974, Remark 2) show that for such β, $t_{\underline{w}(\check{\underline{x}}_n)}$

is a Bayes (Λ) empirical Bayes procedure. Using (16) in (15) we find that

$$\text{(17)} \qquad \underline{w}_i = n \int G_i \binom{n-1}{\check{\underline{n}}_i} \left(\prod_{j=0}^{m} G_j^{n_{ji}} \right) \left(\sum_{j=0}^{m} f_j^{n_{ji}} \right)^{*} \Lambda(dG)$$

where $\check{\underline{n}}_i = (n_{0i}, \ldots, n_{mi})$. Since the marginal density of $\check{\underline{X}}_n$ is $\prod_{\alpha=1}^{n-1} \left(\sum_{i=0}^{m} G_i f_i(x_\alpha) \right)$, the integrand of (17) is proportional to G_i times the conditional density of G given $\check{\underline{X}}_n$. Thus, $\underline{w}$ is proportional to $G^* = E[G|\check{\underline{X}}_n]$.

The empirical Bayes rule which is second level Bayes (Λ) has been seen to be the rule which is first level (component) Bayes versus the induced estimator G^*. We note that G^* depends on Λ through the conditional expectation Q. By Theorem 5, $t_n = t_{G^*}$ is an admissible empirical Bayes procedure. In the next section we give conditions on Λ sufficient for the Bayes risk consistency of t_{G^*} at every $G \in \mathcal{G}$.

3. Bayes Risk Consistency of the Bayes Procedure.

The following lemma is a corollary to Lemma 1 of Oaten (1972).

Lemma 2 (Oaten (1972, p. 1167)). Let $R(i,t) \leq M < \infty$ for all $i \in \{0,1,\ldots,m\}$ and component decision rules t, that is, suppose $S \subset [0,M]^{m+1}$. Then for all $F,G \in \mathcal{G}$,

$$(18) \qquad 0 \leq R(G,t_F) - R(G) \leq M \sum_{i=0}^{m} |F_i - G_i|.$$

It follows from Lemmas 1 and 2 that if $S \subset [0,M]^{m+1}$ then the Bayes (Λ) empirical Bayes procedure $t_n = t_{G^*}$ satisfies

$$(19) \qquad 0 \leq R(G,t_n) - R(G) \leq (m+1)^{\frac{1}{2}}M \, \|G^* - G\| \text{ a.s. } [P_G^{n-1}]$$

where $\| \; \|$ denotes the usual Euclidean norm on R^{m+1}. Hence, the a.s. $[P_G^\infty]$ consistency of the conditional expectation $G^* = E[G|\underset{\sim}{\check{X}}_n]$ at G implies the strong empirical Bayes risk consistency of t_n at G. In turn, by the boundedness of $\|G^* - G\|$, this implies the mean or usual empirical Bayes consistency of t_n at G.

We will use Theorem 6.1 of Schwartz (1965) to establish the a.s. $[P_G^\infty]$ consistency of $E[G|\underset{\sim}{\check{X}}_n]$ at G for all $G \in \mathcal{G}$. For this purpose we establish some lemmas which serve to verify the hypotheses of the Schwartz theorem for our application. In what follows superscript c indicates complementation.

Lemma 3. Let G be any point of $\mathcal{G}$ and V any $\mathcal{G}$ neighborhood of G. There is a uniformly consistent test of $P = P_G$ versus $P \in \{P_F: F \in V^c\}$.

Proof. Fix $G \in \mathcal{G}$. It suffices to prove the result for neighborhoods V of the form

$V = \{(\underline{G}_0,\overline{G}_0) \times (\underline{G}_1,\overline{G}_1) \times \ldots \times (\underline{G}_m,\overline{G}_m)\} \cap G$ where $\underline{G}_i < G_i < \overline{G}_i$ for $i = 0,1,\ldots,m$. If we identify $[0,\underline{G}_i]$ with a subset of R^{m+1}, namely, $\{(x_0,x_1,\ldots,x_m)\colon x_i \in [0,\underline{G}_i]\}$ and similarly for $[\overline{G}_i,1]$, then $V^c = \bigcup_{i=0}^{m} \{([0,\underline{G}_i] \cup [\overline{G}_i,1]) \cap G\}$. Let $[0,\underline{G}_i] \cap G = \underline{U}_i$ and $[\overline{G}_i,1] \cap G = \overline{U}_i$ for $i = 0,1,\ldots,m$. Note then that if G is on the boundary of G, then one or more of the corresponding $\underline{U}_i$ or $\overline{U}_i$ is empty. Let $\mathcal{U} = \{\underline{U}_i\colon i = 0,1,\ldots,m\} \cup \{\overline{U}_i\colon i = 0,1,\ldots,m\}$. Then, since $\mathcal{U}$ is a finite collection of sets and, $\cup\mathcal{U} = V^c$, by Kraft (1955, Theorem 7) it suffices to show that for each nonempty $U \in \mathcal{U}$, there is a uniformly consistent test of $P = P_G$ against the alternative $P \in \{P_F\colon F \in U\}$.

Take $U = \overline{U}_i$ nonempty. (The case $U = \underline{U}_i$ is exactly analogous.) Let h_i be a function such that

$$(20) \qquad E_j h_i = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases}, \quad i,j = 0,1,\ldots,m .$$

Suppose that h_i is bounded by the constant M_i . (For example, one can take $h_0,h_1,\ldots,h_m$ to be a basis dual to the densities $f_0,f_1,\ldots,f_m$ in $L_2(u)$ as observed, e.g. by Robbins (1964).) Now for each n and $\underline{x} \in X^\infty$ define

$$(21) \qquad \overline{h}_{in}(\underline{x}) = \frac{1}{n} \sum_{\alpha=1}^{n} h_i(x_\alpha)$$

and the test function

$$(22)\qquad \varphi_n = \begin{cases} 1 & \text{if } \bar{h}_{in} > c \\ 0 & \text{if } \bar{h}_{in} \le c \end{cases}$$

where $c = \frac{1}{2}(G_i + \bar{G}_i)$. We will show that φ_n is a uniformly consistent test of $P = P_G$ versus $P \in \{P_F: F \in \bar{U}_i\}$. Clearly $E_G^n \varphi_n \to 0$ since $E_G h_i = G_i < c$. Now let $F \in \bar{U}_i$. Then $F_i \ge \bar{G}_i$ so that for each $n = 1,2,\ldots$

$$(23)\qquad E_F^n(1 - \varphi_n) \le P_F^n[-(\bar{h}_{in}(\underline{x}) - F_i) \ge \bar{G}_i - c].$$

By the Hoeffding bound (1963, Theorem 2), RHS (23) is bounded by $\exp\{-B_i n\}$ where B_i is a constant depending only on M_i and $\bar{G}_i - c$. Hence φ_n is a uniformly consistent test of $P = P_G$ versus $P \in \{P_F: F \in \bar{U}_i\}$. □

For $F,G \in \mathcal{G}$ we define

$$(24)\qquad KL(G,F) = E_G\left(\ln \frac{f_G}{f_F}\right),$$

the Kullback-Liebler information number between the mixture densities f_G and f_F

<u>Lemma 4</u>. Suppose that the support of Λ is all of $\mathcal{G}$. Let G be any point of $\mathcal{G}$ and V any $\mathcal{G}$ neighborhood of G. Given $\varepsilon > 0$ there exists a subset $W \subset V$ such that $\Lambda(W) > 0$ and $KL(G,F) < \varepsilon$ for all $F \in W$.

<u>Proof</u>. Fix $G \in \mathcal{G}$ and let V be a $\mathcal{G}$ neighborhood of G. Let $\varepsilon > 0$ be given. For each $\delta > 0$ we define

$$U_\delta = \{F: G_i e^{-\delta} \le F_i \le G_i e^{-\delta} + 1 - e^{-\delta},\ i = 1,2,\ldots,m\},$$

$$V_\delta = \{F: G_i e^{-\delta} \le F_i,\ i = 1,2,\ldots,m \text{ and } \sum_{i=1}^{m} F_i \le e^{-\delta} \sum_{i=1}^{m} G_i + 1 - e^{-\delta}\},$$

$$W_\delta = \{F: G_i e^{-\delta} \le F_i \le G_i e^{-\delta} + \frac{1}{m}(1 - e^{-\delta}),\ i = 1,2,\ldots,m\}$$

and note that $W_\delta \subset V_\delta \subset U_\delta$. We see that $U_\delta \subset V$ for sufficiently small δ. Let δ_0 be such that $U_{\delta_0} \subset V$ and $\delta_0 < \varepsilon$. Since on V_δ, $F_0 \ge e^{-\delta} G_0$, we see that $f_F \ge e^{-\delta} f_G$ on V_δ so that $KL(G,F) < \varepsilon$ for $F \in V_{\delta_0}$. Since each W_δ contains a nonempty open subset of $\mathcal{G}$ and every point of $\mathcal{G}$ is a point of support of Λ, $\Lambda(V_{\delta_0}) > 0$. □

Lemmas 3 and 4 verify hypotheses (ii) and (iii), respectively, of the following theorem which has been converted to our notation. Recall that Q denotes the conditional distribution of G given $\check{\underline{X}}_n = (X_1,\ldots,X_{n-1})$ and $G^* = E[G|\check{\underline{X}}_n]$.

Theorem 6. (Schwartz (1965, Theorem 6.1)). Suppose that (i) the densities $f_G(x)$ may be chosen to be jointly measurable in G and x, (ii) V is a neighborhood of G and there is a uniformly consistent test of $P = P_G$ versus $P \in \{P_F: F \in V^c\}$, and (iii) for every $\varepsilon > 0$, V contains a subset W such that $\Lambda(W) > 0$ and $KL(G,F) < \varepsilon$ on W. Then $Q(V^c) \to 0$ a.s. $[P_G^\infty]$.

Our final lemma will be used to complete the last step of the proof of the Bayes risk consistency of the Bayes (Λ) empirical Bayes rule $t_n = t_{G^*}$.

Lemma 5. Suppose that Λ is a distribution on $\mathcal{G}$. Let $G \in \mathcal{G}$ be such that for every neighborhood V of G, $Q(V^c) \to 0$ a.s. $[P_G^\infty]$. Then $\|G^* - G\| \to 0$ a.s. $[P_G^\infty]$.

Proof. Let $G \in \mathcal{G}$. Note that

(25) $$\|G^* - G\| = \|\int(F - G)Q(dF)\| \leq \int\|F - G\|Q(dF)$$

where the inequality is the Minkowski integral inequality (e.g., see Fabian and Hannan (1973, §1.5)). For each $\varepsilon > 0$ let $V_\varepsilon = \{F: \|F - G\| < \varepsilon\}$. Partitioning the integration into integration over V_ε and over V_ε^c and using the bound $\|F - G\| \leq 1$, (25) yields

(26) $$\|G^* - G\| \leq \varepsilon + Q(V_\varepsilon^c) \quad \text{a.s.} \quad [P_G^\infty].$$

Since $\varepsilon > 0$ is arbitrary the proof is complete. □

Lemmas 1-5 and the Schwartz theorem are used to prove the main result of this section, namely

Theorem 7. Suppose $\Theta = \{0,1,\ldots,m\}$ and S is a compact set. Suppose Λ is a probability distribution with support $\mathcal{G}$. Then a Bayes (Λ) empirical Bayes procedure t_n is Bayes risk consistent at every $G \in \mathcal{G}$, i.e.

(27) $$R_n(G,t_n) \to R(G) \quad \text{for all} \quad G \in \mathcal{G}.$$

Proof. By Lemma 1, $t_n = t_{G^*}$ is a Bayes (Λ) empirical Bayes procedure. By Lemma 2, this t_n satisfies (19) for each n. By Lemmas 3-5 and the Schwartz theorem (Theorem 6), $\|G^* - G\| \to 0$ a.s. $[P_G^\infty]$ for all $G \in \mathcal{G}$ from which (27) follows by the bounded convergence theorem. □

APPENDICES

APPENDIX A

In this Appendix we show that the Robbins (1951) bootstrap rule $\underline{t}$ defined in Example 1 is an inadmissible compound rule if $n = 2$. The demonstration for general $n \geq 2$ is similar but notationally complicated.

Since Θ^2 is finite, if the equivariant rule $\underline{t}$ is admissible it is Bayes versus some invariant prior β on the four states $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$. Invariance of β requires $\beta_{01} = \beta_{10}$. In Huang (1972) only β with $\beta_{00} = \beta_{11}$ are considered. It can be shown for general invariant β, a Bayes compound rule with respect to β must be equal a.e. [Lebesgue on R^2] to the equivariant compound rule $\underline{\tilde{t}}$ with

$$\tilde{t}_2(x_1,x_2) = \begin{cases} 1 & \text{if } \tfrac{1}{2}(1 - \beta_{00} - \beta_{11})(A_2 - A_1) + \beta_{11}A_1A_2 - \beta_{00} \geq 0 \\ 0 & \text{if } \tfrac{1}{2}(1 - \beta_{00} - \beta_{11})(A_2 - A_1) + \beta_{11}A_1A_2 - \beta_{00} < 0 \end{cases}$$

where $A_i \equiv \exp\{2x_i\}$, $i = 1,2$. Note that if $\beta_{11} > \beta_{00}$ then $(x_1,x_2) = (0,0)$ is in the interior of the region deciding $(\theta_1 = 1, \theta_2 = 1)$. The Robbins bootstrap rule $\underline{t}$ of Example 1 has $(0,0)$ on the boundary of the partition it induces in R^2 so $\underline{t}$ is not equal a.e. [Lebesgue on R^2] to a rule $\underline{\tilde{t}}$ Bayes with respect to β, $\beta_{11} > \beta_{00}$. Similarly, for β with $\beta_{11} < \beta_{00}$. The partitions induced by $\underline{t}$

and by $\underset{\sim}{\tilde{t}}$ for $\beta_{11} = \beta_{00} = \frac{1}{3}$ are shown in Figure 1 of Huang (1972). The separating curves for the $\underset{\sim}{\tilde{t}}$ partition have vertical and horizontal asymptotes for every β. The $\underset{\sim}{t}$ partition has separating curves with asymptotes $x_1 + x_2 = 2$, $x_1 + x_2 = -2$ so that $\underset{\sim}{t}$ is not equal a.e. [Lebesgue on R^2] to any $\underset{\sim}{\tilde{t}}$. Thus, $\underset{\sim}{t}$ is inadmissible.

APPENDIX B

Theorem. In the 2×2 component testing problem, if $Z = \frac{f_1(X)}{f_0(X)}$ has continuous cumulative distribution function on $(0,\infty)$ under both P_0 and P_1, then the lower boundary B of the risk set S is strictly convex.

Proof. Let $s_1 = (s_1^0, s_1^1)$ and $s_2 = (s_2^0, s_2^1)$ be distinct points on B, corresponding to rules t_1 and t_2 respectively and assume without loss of generality that $s_1^0 > s_2^0$. Thus, $s_2^1 > s_1^1$.

From the Neyman-Pearson lemma, we know that t_1 is of the form

$$t_1 = \begin{cases} 1 & \text{if } Z > k_1 \\ \gamma_1 & \text{if } Z = k_1 \\ 0 & \text{if } Z < k_1 \end{cases}$$

where $k_1 \geq 0$, and similarly, there is a $k_2 > k_1$ associated with t_2.

We assume for the moment that $k_1 \neq 0$, $k_2 \neq \infty$. Then, by the continuity of the distribution of Z, we know that the probability that $Z = k_1$ or $Z = k_2$ is zero under either P_0 or P_1. We arbitrarily assign $\gamma_i = 1$, $i = 1,2$ and write

$$s_i = (P_0[Z \geq k_i], P_1[Z < k_i]), \; i = 1,2.$$

Let β be given, $0 < \beta < 1$, and define

$$s_\beta = \beta s_1 + (1 - \beta) s_2$$

and note that s_β is the risk point associated with the rule $t_\beta = t_1 + (1 - \beta)t_2$, namely,

$$t_\beta = \begin{cases} 1 & \text{if } Z \geq k_2 \\ \beta & \text{if } k_1 \leq Z < k_2 \\ 0 & \text{if } Z < k_1 \end{cases} .$$

Since t_β does not possess Neyman-Pearson structure, s_β is not on the lower boundary B. (s_β is dominated by any Neyman-Pearson test of size s_β^0.)

The cases $k_1 = 0$ and $k_2 = \infty$ can be handled in analogous ways the only difference being the choice of γ_1, γ_2. (The points of B on the two axes are $(P_0[Z \geq \infty], P_1[Z < \infty])$ and $(P_0[Z > 0], P_1[Z \leq 0])$.)

REFERENCES

REFERENCES

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

Copas, J.B. (1974). On symmetric compound decision rules for dichotomies. *Ann. Statist.* **2** 199-204.

Fabian, Václav and Hannan, James (1973). *Introduction to Probability and Mathematical Statistics*. Lecture notes, Statistics and Probability, MSU.

Ferguson, Thomas (1967). *Mathematical Statistics, A Decision Theoretic Approach*. Academic Press, New York and London.

Ferguson, Thomas S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2** 615-629.

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

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, Statistics and Probability, MSU.

Gilliland, Dennis C. and Hannan, James (1976). Improved rates in the empirical Bayes monotone multiple decision problem with MLR family. RM-352, Statistics and Probability, MSU.

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

Hoeffding, Wassily (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13-30.

Huang, J.S. (1972). A note on Robbins' compound decision procedure. Ann. Math. Statist. 43 348-350.

Kraft, Charles (1955). Some conditions for consistency and uniform consistency of statistical procedures. Univ. of Calif. Publications in Statist.

Lindley, D.V. (1971). Bayesian statistics, a review. Regional Conference Series in Applied Mathematics No. 2, SIAM, Philadelphia.

Meeden, Glen (1972). Some admissible empirical Bayes procedures. Ann. Math. Statist. 43 96-101.

Oaten, Allan (1972). Approximation to Bayes risk in compound decision problems. Ann. Math. Statist. 43 1164-1184.

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

Robbins, H. (1956). An empirical Bayes approach to statistics. Proc. Third Berkeley Symp. Math. Statist. Prob. 1, University of California Press, 157-163.

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

Rolph, John E. (1968). Bayesian estimation of mixing distributions. Ann. Math. Statist. 39 1289-1302.

Schwartz, L. (1965). On Bayes' procedures. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete. 4 10-26.

Shapiro, C.P. (1974). Bayesian classification: asymptotic results. Ann. Statist. 2 763-774.

Tucker, Howard G. (1963). An estimate of the compounding distribution of a compound Poisson distribution. Theor. Prob. Appl. 8 195-200.

Van Houwelingen, J.C. (1973). On empirical Bayes rules for the continuous one-parameter exponential family. Doctoral thesis. Rijksuniversiteit te Utrecht, Netherlands (to appear in Ann. Statist. (1976)).