This is to certify that the

dissertation entitled

Empirical Bayes
with
Sequential Components

presented by

Rohana Jith Karunamuni

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

Dennis C. Gilliland
Major professor

Date Nov. 6, 1985

MSU is an Affirmative Action/Equal Opportunity Institution O-12771

EMPIRICAL BAYES WITH SEQUENTIAL COMPONENTS

by

Rohana Jith Karunamuni

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

1985

ABSTRACT

EMPIRICAL BAYES WITH SEQUENTIAL COMPONENTS

by

Rohana Jith Karunamuni

We consider the empirical Bayes decision problem where the component problem is an m-truncated sequential decision problem. The Bayes risk envelope of this component is smaller than that of the fixed m-sample size component when there is a cost for each observation. Empirical Bayes methods are presented that result in empirical Bayes risk approaching envelope risk as the number of components increases.

Consider an empirical Bayes decision problem with fixed sample size components. Let $(\theta,\underline{X})$, $(\theta_1,\underline{X}_{1m})\ldots(\theta_n,\underline{X}_{nm})\ldots$ be i.i.d. with the common distribution being $\theta \sim G$, with $\underline{X} \sim P_\theta x P_\theta x \ldots x P_\theta = P_\theta^{(m)}$ conditional on θ. An empirical Bayes decision procedure is a sequence $\{t_n\}$ where $t_n = t_n(\underline{X}_{1m},\ldots,\underline{X}_{nm})$ is a function taking values in the set of component decision rules and $EL[\theta,t_n(\underline{X}_{1m},\underline{X}_{2m},\ldots,\underline{X}_{nm})(\underline{X})]$ is its n^{th} stage risk for making a decision about θ. The convergence of this risk to the component envelope risk whatever be G is termed asymptotic optimality (a.o.).

With the sequential components, t_n selects both a stopping rule function and a terminal decision function for use in the component with parameter θ. This results in the observations available at stage n being $\underline{X}_{1N_1},\underline{X}_{2N_2},\ldots,\underline{X}_{nN_n}$, where N_1, $N_2,\ldots,N_n$

are the numbers of component observations taken in the first n components. The random sample size feature adds a degree of complexity to the development of a.o. of empirical Bayes procedures. Our results include the demonstration of a.o. procedures for some finite state components and an empirical Bayes approach to one-step look ahead procedures for some infinite state multiple decision components.

ACKNOWLEDGEMENTS

I express my sincere thanks and deep gratitude to Professor Dennis C. Gilliland for his continuous encouragement and guidance in the preparation of this thesis. I would like to place my appreciation for the time that he spent in the discussions.

I am very grateful to Professor R.V. Erickson, Professor H. Salehi and Professor J.C. Gardiner for their continuing help and for serving on my guidance committee.

I would also like to thank Professor James Hannan for many fruitful discussions during my studies at MSU.

I wish to acknowledge the Department of Statistics and Probability for providing the financial support during my stay at MSU.

Finally, I wish to thank Cathy Sparks for her patience, efficiency and great care in typing this thesis.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1. Introduction

This thesis is concerned with empirical Bayes decision theory with a sequential statistical decision problem as the component. Robbins (1949, 1956) introduces empirical Bayes decision theory, and Robbins (1964) summarizes results for certain components and develops some general methods in regard to estimation of the prior (mixing) distribution. Our development is reasonably self-contained but does presuppose that the reader is somewhat familiar with both sequential analysis and empirical Bayes decision theory.

The components to which empirical Bayes methods have been applied are, with few exceptions, given fixed sample size identical statistical decision problems. Exceptions are the varying (non-stochastic) sample size components considered by O'Bryan (1972, 1976), O'Bryan and Susarla (1975, 1976, 1977) and Susarla and O'Bryan (1975). Another exception is found in the works of Laippala (1979, 1980, 1983, 1985). In his case the varying sample sizes are random.

In this thesis we define a construct within which we can treat both the random sample size and the usual fixed sample size cases. We believe that the somewhat informal style with which this is accomplished will expedite understanding of a fairly complicated mathematical construct which mixes sequential analysis and empirical Bayes decision theory.

In Section 1.2 we introduce the sequential component problem which is the kernel of the empirical Bayes decision problem. In Section 1.3 we give a brief introduction to Robbins' empirical Bayes decision problem and develop new results for some finite parameter space component problems. Finite action linear loss testing and multiple decision component problems are treated in Chapter 2.

1.2. The Component Problem

The component problem we consider is the m-truncated sequential decision problem with constant cost per observation. This problem is described in detail in various books concerning decision theory, e.g., Berger (1980, Chapter 7). However, in order to make this presentation reasonably self-contained, we will use this section to develop notations to describe the component problem that is the kernel of the empirical Bayes decision problem that is the subject of our investigation.

The parameter space is the measurable space (Θ,A) and the class of all (prior) probability distributions on A is denoted by $\mathcal{G}$. The parameter θ indexes transitions; specifically, P_θ is a distribution on $(\mathcal{X},\mathcal{B})$ where $\mathcal{X}$ is the real line and $\mathcal{B}$ is the Borel σ-field. Conditional on θ, the observable random variables $X_1,\ldots,X_m$ are iid P_θ; for $k = 1,2,\ldots,m$, $\underline{x}^k = (x_1,\ldots,x_k) \in \mathcal{X}^k$, $\underline{X}^k \sim P_\theta^k = P_\theta \times \ldots \times P_\theta$ (k times) and $\mathcal{B}^k$ denotes the Borel σ-field in $\mathcal{X}^k$.

Suppose that the component decision problem has (terminal) action space A and loss function $L \geq 0$ defined on $\Theta \times A$. Let $c \geq 0$ denote the constant cost per observation.

For $k = 0,1,\ldots,m$, let $\mathcal{D}^k$ denote a set of mappings δ from $\mathcal{X}^m$ into A that are constant with respect to the last $m-k$ coordinates and are such that $L(\theta,\delta)$ is $A \times \mathcal{B}^m$ measurable. $\mathcal{D}^0$ consists of constant functions. We will regard the domain of $\delta \in \mathcal{D}^k$ as $\mathcal{X}^k$ when it is convenient to do so, $k = 1,2,\ldots,m$.

We are to sample the X_k sequentially. The observations are taken one at a time, with a decision being made after each observation, to either stop sampling and choose an action in A or to take another observation. Therefore, a sequential decision procedure consists of two components, a stopping rule and a terminal decision rule.

The <u>stopping rule</u> $\underline{\tau}$ is a sequence $\underline{\tau} = (\tau_0, \tau_1, \ldots, \tau_m)$ where τ_0 is a constant function representing the probability of making a decision without sampling and, for $k = 1,2,\ldots,m$ $\tau_k \colon X^m \to [0,1]$ is an $\underline{x}^k$ - measurable function representing the conditional probability of stopping at stage k given that sampling did not stop at stages $0,1,\ldots,k-1$ and given the observation $\underline{x}^k$. With the m-truncated sequential problem, $\tau_m \equiv 1$. For a <u>nonrandomized stopping rule</u> $\underline{\tau}$, τ_k takes values in $\{0,1\}$ for $k = 0,1,\ldots,m$. Associated with $\underline{\tau}$ is the stopping time variable N.

We will be concerned only with nonrandomized $\underline{\tau}$ and will take $N = \min\{k \colon \tau_k = 1\}$. Such $\underline{\tau}$ partition X^m into $\{[N=0], [N=1], \ldots, [N=m]\}$ where $[N=k]$ is a function of $\underline{x}^k$ only. If $\tau_0 = 1$, then $[N=0] = X^m$ and $N \equiv 0$. If $\tau_0 = 0$, then $[N=0] = \phi$.

The <u>terminal decision rule</u> $\underline{\delta}$ is a sequence $\underline{\delta} = (\delta_0, \delta_1, \ldots, \delta_m) \in \mathcal{D} = \mathcal{D}^0 \times \mathcal{D}^1 \times \ldots \times \mathcal{D}^m$. The interpretation is that the decision function δ_k is used on the set $[N=k]$, $k = 0,1,\ldots,m$.

A <u>sequential decision procedure</u> is a pair $(\underline{\tau}, \underline{\delta})$, where $\underline{\tau}$ is a stopping rule and $\underline{\delta}$ is a terminal decision rule. We will restrict consideration to the class of procedures $T \times \mathcal{D}$, where

T denotes the class of all nonrandomized stopping rules $\underline{\tau}$ such that $\tau_0 \equiv 0$. (The reason for working with stopping rules such that $N \geq 1$ will be made clear in the next section.) Thus, we may drop first coordinate functions and take $\underline{\tau} = (\tau_1,\ldots,\tau_m)$, $\underline{\delta} = (\delta_1,\ldots,\delta_m)$.

For a sequential procedure $(\underline{\tau},\underline{\delta})$ the risk (including cost for observations) given θ is

$$\begin{aligned} R(\theta,(\underline{\tau},\underline{\delta})) &= \int \sum_{k=1}^{m} [N=k]\{L(\theta,\delta_k) + ck\} \ \underline{P}_\theta^m(d\underline{x}^m) \\ &= \sum_{k=1}^{m} \int_{[N=k]} L(\theta,\delta_k)\underline{P}_\theta^k(d\underline{x}^k) + cE_\theta N; \end{aligned} \tag{1.1}$$

the Bayes risk of $(\underline{\tau},\underline{\delta})$ with respect to $G\varepsilon\mathcal{G}$ is

$$R(G,(\underline{\tau},\underline{\delta})) = \int R(\theta,(\underline{\tau},\underline{\delta}))G(d\theta); \tag{1.2}$$

and the (infimum) Bayes risk at G is

$$R_B(G) = \inf\{R(G,(\underline{\tau},\underline{\delta})): \ (\underline{\tau},\underline{\delta})\varepsilon T\times\mathcal{D}\}. \tag{1.3}$$

Under certain conditions, the Bayes envelope value $R_B(G)$ is achieved by a choice $(\underline{\tau}^B(G),\underline{\delta}(G))$ where the components of the Bayes stopping rule $\underline{\tau}^B(G)$ are defined by backward induction and the components of the Bayes decision rule $\underline{\delta}(G)$ are Bayes in the respective fixed sample size problems. (cf. Berger (1980, Chapter 7).)

In the typical empirical Bayes problem, the decision-maker is faced with a sequence of independent, identical component problems with a common (unknown) G. The idea is to pool sample information from the components in the sequence to estimate G, or the Bayes procedure with respect to G, so that as the number of components increases, the empirical Bayes decision procedure has risk approaching the envelope value, whatever be G, i.e., is asymptotically optimal.

Our formulation of the component problem enables one to produce the fixed sample size and other component problems by restricting $\underline{\tau}$. Gilliland and Hannan (1974, 1976, 1985) have introduced the idea of producing different envelopes by restriction of the class of decision procedures. For example, if we define

$$(1.4) \qquad T_F = \{\underline{\tau} \varepsilon T: \ \underline{\tau} \text{ constant}\}$$

and

$$(1.5) \qquad R_F(G) = \inf \{R(G,(\underline{\tau},\underline{\delta})): \ (\underline{\tau},\underline{\delta}) \varepsilon T_F \times \mathcal{D}\},$$

we have the envelope risk for fixed sample size procedures. We refer to (1.5) as the <u>optimal fixed sample size risk</u>. If we define T_k to be a subset of T_F consisting of rules $\underline{\tau}$ where $\tau_j = 1$ if $j = k$ and $= 0$ if $j < k$, then the corresponding envelope $R_k(G)$ is that of the fixed sample size k component. We refer to $R_k(G)$ as the <u>optimal fixed sample size k risk</u>. Note that for each G,

$$(1.6) \qquad R_F(G) = \inf \{R_k(G): k = 1,2,\ldots,m\}.$$

Since

$$(1.7) \qquad R_B(G) \leq R_F(G) \leq R_k(G), \ k = 1,2,\ldots,m,$$

achieving the Bayes envelope risk for the m-truncated sequential component is preferred to achieving envelope risk in either the optimal fixed sample size or given fixed sample size component. However, the construction of empirical Bayes procedures that achieve the most stringent envelope is the most difficult construction.

We will also be dealing with a sequential procedure which in the literature is called a one-step look ahead procedure. Here the decision to continue or to stop the sampling is made after observation x_k by comparing the conditional on $\underline{x}^k$ (posterior) Bayes risk of stopping and making a decision with that of making a decision after taking one more observation and playing Bayes. (See Berger (1980, §7.4.6) for details.) The one-step look ahead stopping rule is much simpler than the Bayes stopping rule. (The latter can be defined by backward induction; the decision to stop is based on a comparison of the conditional on $\underline{x}^k$ Bayes risk of stopping and making a decision with that of not stopping and playing an optimal sequential strategy from that point.)

Formally, the one-step look ahead stopping rule is defined as follows. For $k = 1,2,\ldots,m$, let G_k denote a posterior distribution of θ given $\underline{x}^k$ in the component with $\theta \sim G$ and, conditional on θ, $X_1,\ldots,X_k$ iid P_θ. (We assume such G_k exist for each $\underline{x}^k \in X^k$, $k = 1,2,\ldots,m$.). We suppose that

(1.8) $$r(G) = \inf\{\int L(\theta,a)G(d\theta) : a \in A\}$$

is attained at each G and that a Bayes decision function $\delta_k(G)$ attains the infimum posterior Bayes risk, i.e.,

$$(1.9) \qquad r(G_k) = \int L(\theta,\delta_k(G))\, G_k(d\theta)$$

for all $\underline{x}^k$ and $k = 1,2,\ldots,m$.

The <u>one-step look ahead stopping rule</u> is $\underline{\tau}^L(G) = (\tau_1^L(G),\ldots,\tau_m^L(G))$ where $\tau_m^L(G) = 1$ and, for $k = 1,2,\ldots,m-1$,

$$(1.10) \qquad \tau_k^L(G) = \begin{cases} 1 & \text{if} \quad E^* r(G_{k+1}) + c - r(G_k) \geq 0 \\ 0 & \text{if} \quad E^* r(G_{k+1}) + c - r(G_k) < 0 \end{cases}$$

with E^* denoting conditional expectation on X_{k+1} given $\underline{X}^k = \underline{x}^k$. The <u>one-step look ahead sequential decision procedure</u> is $(\underline{\tau}^L(G), \underline{\delta}(G))$; i.e., it uses the Bayes terminal decision rule with the stopping rule (1.10). If $m = 2$, $\underline{\tau}^L(G)$ is a Bayes stopping rule.

The Bayes risk of $(\underline{\tau}^L(G), \underline{\delta}(G))$ is denoted by $R_L(G)$. This is not an envelope risk, rather, the Bayes risk associated with a particularly tractable sequential decision rule. We will consider the empirical Bayes approach to achieving risk $R_L(G)$ in Chapter 2 for particular classes of sequential components.

We now give three examples to illustrate some of the concepts introduced to this point. The examples are sequential m-truncated components with testing simple versus simple, testing with linear loss and estimation with squared error loss.

<u>Example 1.1</u> (Testing Simple vs. Simple). Let $\Theta = \{0,1\}$, $A = \{0,1\}$ and $L(0,0) = L(1,1) = 0$, $L(0,1) = L(1,0) = L > 0$, a constant. We

identify a prior G on Θ by the mass π it puts on the state 1 so that G can be identified with the unit interval. Let P_0 be $N(-1,1)$ and P_1 be $N(1,1)$. (This example is the sequential version of that used by Robbins (1951) to introduce the idea of compound decision theory.) The posterior probability of $\theta = 1$ given $\underline{x}^k$ is $\pi_k = \pi\exp(\sum_1^k x_j)/\{\pi\exp(\sum_1^k x_j) + (1-\pi)\exp(-\sum_1^k x_j)\}$ and a Bayes decision rule is

$$(1.11)\qquad \delta_k = \begin{cases} 1 & \text{if } \pi_k \geq 1/2 \\ 0 & \text{if } \pi_k < 1/2 \;. \end{cases}$$

The event $\pi_k \geq 1/2$ is equivalent to $\Sigma_1^k x_j \geq c(\pi)$ where $c(\pi) = (1/2)\ln((1-\pi)/\pi)$.

Consider the case of truncation at $m = 2$. In this case, the one-step look ahead procedure is Bayes with respect to π so that $R_L(\pi) = R_B(\pi)$, $0 \leq \pi \leq 1$. Let the loss for misclassification be $L = 1$ and the cost per observation be $c = .05$.

The one-step look ahead stopping rule $\underline{\tau}(\pi)$ (cf. (1.10)) is defined by $\tau_2(\pi) = 1$ and

$$(1.12)\qquad \tau_1(\pi) = \begin{cases} 1 & \text{if } r_1(\pi_1) + .05 - r_0(\pi_1) \geq 0 \\ 0 & \text{if } r_1(\pi_1) + .05 - r_0(\pi_1) < 0 \;. \end{cases}$$

Here

$$r_0(\pi) = \pi[\pi \leq 1/2] + (1-\pi)[\pi > 1/2],$$

and

$$r_1(\pi) = \pi\Phi(c(\pi) - 1) + (1-\pi)\{1 - \Phi(c(\pi) + 1)\}$$

where Φ denotes the standard normal cdf. Calculations show that (1.12) is equivalent to

$$\tau_1(\pi) = \begin{cases} 1 & \text{if } |\pi_1 - .5| \leq .361567 \\ 0 & \text{if } |\pi_1 - .5| > .361567 \end{cases} \tag{1.13}$$

and

$$\tau_1(\pi) = \begin{cases} 1 & \text{if } |x_1 - c(\pi)| \leq c(.138433) \\ 0 & \text{if } |x_1 - c(\pi)| > c(.138433). \end{cases} \tag{1.14}$$

The envelope risk $R_B(\tau)$ resulting from the Bayes procedure $\underline{d}(\tau) = (\underline{\tau}(\pi),\underline{\delta}(\pi))$ was calculated for selected values of π and is plotted below along with the fixed sample size envelopes $R_0(\pi) = r_0(\pi)$ and $R_k(\pi)$, $k = 1, 2$, where

$$R_k(\pi) = \pi\Phi((c(\pi) - k)/\sqrt{k}) + (1 - \pi)\{1 - \Phi((c(\pi) + k)/\sqrt{k})\} + .05k. \tag{1.15}$$

Of course, the optimal fixed sample size risk envelope is $R_F(\pi) = \min\{R_1(\pi), R_2(\pi)\}$. Note that $R_B(\pi)$ is considerably less than $R_F(\pi)$ for priors π near .5 . Also note that for π's near 0 and 1, $R_0(\pi) < R_F(\pi)$.

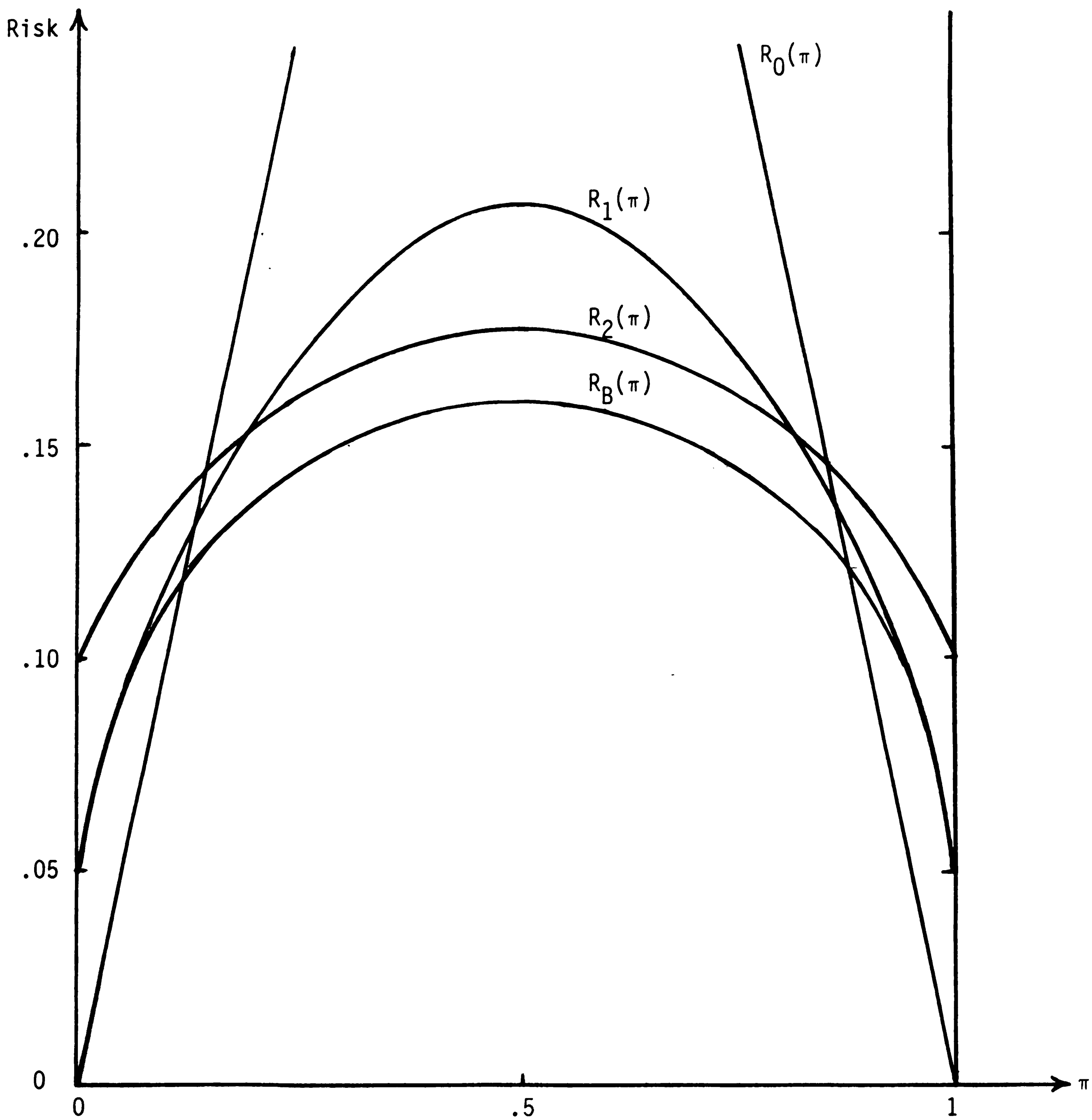

Figure 1.1 Envelope Risk Functions

Theorem 1.1 of the next section shows that the Bayes envelope risk $R_B(\pi)$ for the truncated sequential component is achieved in the limit by empirical Bayes decision procedures $\underline{d}^n = (\underline{\tau}^n, \underline{\delta}^n)$ where $\underline{\tau}^n$ and $\underline{\delta}^n$ are Bayes with respect to consistent estimates of π. The theorem as stated and proved subsumes the usual fixed sample size component case and the optimal floating fixed sample size case through restriction of the class of component stopping rules.

In the next two examples, $P_\theta = N(\theta,1)$, $\theta \in \Theta = (-\infty,\infty)$ and the prior on Θ is $G = N(0,\sigma^2)$. In this case the posterior distribution G_k of θ given $\underline{X}^k = \underline{x}^k$ is $N(\sigma^2 S_k/(1 + k\sigma^2), \sigma^2/(1 + k\sigma^2))$ where $S_k = \Sigma_1^k x_j$, $k = 1,2,\ldots$. In the following examples, μ_k denotes the conditional mean.

Example 1.2 (Testing with Linear Loss). Let $\theta_0 \in \Theta$. Let $A = \{a_0, a_1\}$, where a_0 and a_1 correspond to actions "decide $\theta \leq \theta_0$" and "decide $\theta > \theta_0$", respectively. Let $L(\theta,a_0) = (\theta - \theta_0)^+$ and let $L(\theta,a_1) = (\theta - \theta_0)^-$. Let $\underline{\delta}(G) = (\delta_1,\ldots,\delta_m)$ be the Bayes terminal decision rule where δ_k is a Bayes decision rule with respect to G for the fixed sample size k decision problem. If $\delta_k = \Pr(a_0|\underline{X}^k = \underline{x}^k)$, then we take

$$(1.16) \qquad \delta_k = \begin{cases} 1 & \text{if} \quad \mu_k \leq \theta_0 \\ 0 & \text{if} \quad \mu_k > \theta_0 . \end{cases}$$

We now derive the one-step look ahead stopping rule. Posterior minimum Bayes risk at stage $k(k \geq 1)$ is given by

$$r(G_k) = \inf_{\delta} \int_\Theta L(\theta,\delta)\, G_k(d\theta) = \int_\Theta L(\theta,\delta_k)\, G_k(d\theta)$$

$$= \begin{cases} \int_{\theta_0}^{\infty} (\theta - \theta_0) G_k(d\theta) & \text{if } \mu_k \le \theta_0 \\ \int_{-\infty}^{\theta_0} (\theta_0 - \theta) G_k(d\theta) & \text{if } \mu_k > \theta_0 \,. \end{cases}$$

Define $\psi(t) = \int_t^\infty (x - t)\phi(x)dx = \phi(t) - t(1 - \Phi(t))$, where ϕ denotes the density function and Φ denotes the cdf of the standard normal distribution. Then

$$r(G_k) = \begin{cases} a_k^{-1/2}\, \psi\, (a_k^{1/2}(\theta_0 - \mu_k)) & \text{if} \quad \mu_k \le \theta_0 \\ a_k^{-1/2}\, \psi\, (a_k^{1/2}(\mu_k - \theta_0)) & \text{if} \quad \mu_k > \theta_0 \end{cases}$$

where $a_k = (1 + k\sigma^2)/\sigma^2$. In other words,

(1.17) $$r(G_k) = a_k^{-1/2}\, \psi\, (a_k^{1/2}|\theta_0 - \mu_k|).$$

The conditional Bayes risk $E^* r(G_{k+1})$ from taking X_{k+1} given $\underline{X}^k = \underline{x}^k$ is (see DeGroot (1970) p. 286)

$$E^* r(G_{k+1}) = r(G_k) - b_k^{-1/2}\, \psi\, (b_k^{1/2}|\theta_0 - \mu_k|)$$

where

$$b_k = a_k\, (\sigma^2 a_k + 1) \,.$$

Therefore, in defining the stopping rule (1.10),

$$E^* r(G_{k+1}) + c - r(G_k) = c - b_k^{1/2}\, \psi\, (b_k^{1/2}|\theta_0 - \mu_k|) \,.$$

Thus, the stopping rule $\underline{\tau}^L$ for the m-truncated problem stops sampling for the first $k(k = 1,2,\ldots,m-1)$ for which $c \ge b_k^{1/2}\, \psi\, (b_k^{1/2}|\theta_0 - \mu_k|)$ or $k = m$, whichever is smaller.

<u>Example 1.3</u> (Estimation with Squared Error Loss). Let $A = \Theta$ and let $L(\theta,a) = (\theta - a)^2$. Then the Bayes terminal decision rule $\underline{\delta}(G) = (\delta_1,\ldots,\delta_m)$ for the m-truncated problem is given by the estimators $\delta_k = \mu_k$, $k = 1,2,\ldots,m$.

Posterior minimum Bayes risk is the posterior variance $r(G_k) = \int (\theta - \mu_k)^2 G_k(d\theta) = \sigma^2/(1 + k\sigma^2)$. Therefore, $r(G_k)$ depends only on the number of observations that have been taken and not on the observed values $X_1,\ldots,X_k$. Thus $E^* r(G_{k+1}) = \sigma^2/(1 + (k+1)\sigma^2)$. Therefore $E^* r(G_{k+1}) + c - r(G_k) = \sigma^2/(1 + (k+1)\sigma^2) + c - \sigma^2/(1 + k\sigma^2)$. Then the stopping rule $\underline{\tau}^L$ for the m-truncated problem stops sampling for the first k $(k = 1,\ldots,m-1)$ for which $c \geq \sigma^4/(1 + k\sigma^2)(1 + (k+1)\sigma^2)$ or for $k = m$, whichever is smaller.

1.3 The Empirical Bayes Problem

In the last section, we introduced the basic idea behind Robbins' empirical Bayes formulation after defining the sequential m-truncated component problem. This component has envelope risk which is no greater than that of any fixed sample size k version, $k = 1,2,\ldots,m$. By suitable restriction of the class of stopping rules, the component specializes to the customary fixed sample size component.

In its full generality, the sequential component presents some unique problems when it is the kernel of the empirical Bayes decision problem. Let $\underline{X}^{N_1},\ldots,\underline{X}^{N_n},\ldots$ be the observation vectors from the sequence of independent repetitions of the component. For each n, the stopping rule $\underline{\tau}^n$ used in the nth component can depend on $(\underline{X}^{N_1},\ldots,\underline{X}^{N_{n-1}})$. Thus, the sequence of observed random vectors is not the usual iid sequence.

This fact raises interesting problems in regard to the efficient use of such random vectors in estimating stopping rules and decision rules for the empirical Bayes application. This thesis will not address these problems. Rather, it will address the issues of the convergence of empirical Bayes risk to envelope risks such as $R_B(G)$ and $R_F(G)$ and to one-step look ahead risk $R_L(G)$ for selected empirical Bayes decision procedures based on (assumed) consistent estimators of G or of $\underline{\tau}^B(G),\underline{\tau}^F(G)$, $\underline{\tau}^L(G)$ and $\underline{\delta}(G)$.

We will now indicate the reason why we restrict T to $\underline{\tau}$ such that $\tau_0 = 0$. Consider the empirical Bayes problem with the sequential m-truncated component and envelope R_B. For specificity, take $\Theta = \{0,1,\ldots,a\}$ and G to be the a-dimensional simplex of probability vectors on Θ with norm and Borel σ-field that induced from X^{a+1}. For certain component loss structure, transitions P_θ and cost c, there will exist an open set $G_0 \subseteq G$ such that, for $G\varepsilon G_0$, $\tau_0(G) = 1$, i.e., G such that a Bayes stopping rule with respect to G in unrestricted T calls for taking no observations. (See Example 1.1.) Hence, if $\hat{G}_n$ is an a.s. consistent estimator of $G\varepsilon G$ and if $G\varepsilon G_0$, then the empirical Bayes strategy based on the estimate $\hat{G}_n$ will a.s. cause sampling to terminate after a finite number of repetitions of the component and, therefore, will not enable consistent estimation of the mixing distribution. Hence, it will not be possible to achieve the envelope risk $R_B(G)$. For this reason, we have restricted consideration to the class of stopping rules T which take at least one observation, i.e., such that $\tau_0 = 0$. (In this case note, that the first coordinates of the observation vectors $\underline{X}^{N_1},\ldots,\underline{X}^{N_n},\ldots$ are iid so that the decision-maker does have standard data available with which to construct consistent estimates of G.)

Suppose that the stopping rules are restricted to a class $T_\star \subseteq T$. The envelope risk associated with $T_\star$ is

$$\begin{aligned} R_\star(G) &= \inf\,\{R(G,(\underline{\tau},\underline{\delta})):\ (\underline{\tau},\underline{\delta})\varepsilon T_\star \times \mathcal{D}\} \\ &= \inf\,\{R(G,(\underline{\tau},\underline{\delta}(G)):\ \underline{\tau}\varepsilon T_\star\} \end{aligned} \tag{1.18}$$

which we assume is attained at $\underline{\tau}^*(G)$, $G\varepsilon\mathcal{G}$. The $n\underline{th}$ stage empirical Bayes risk based on the estimate $\hat{G} = G_n(\underline{X}^{N_1},\ldots,\underline{X}^{N_{n-1}})$ and conditional on the data satisfies

$$
(1.19) \qquad \begin{aligned} 0 \le R(G,\underline{d}(\hat{G})) - R_*(G) \le\ & R(G,\underline{d}(\hat{G})) - R(G,\underline{d}(G)) \\ & + R(\hat{G},\underline{d}(G)) - R(\hat{G},\underline{d}(\hat{G})) \quad \text{for } G,\ \hat{G}\varepsilon\mathcal{G}, \end{aligned}
$$

where we have used abbreviations $\underline{d}(H) = (\underline{\tau}^*(H),\underline{\delta}(H))$, $H\varepsilon\mathcal{G}$. Bounds like (1.19) are basic in empirical Bayes analysis and go back at least to Hannan (1957). Hence,

$$
(1.20) \qquad 0 \le R(G,\underline{d}(\hat{G})) - R_*(G) \le 2 \sup \{|R(\hat{G},\underline{d}) - R(G,\underline{d})| : \underline{d}\varepsilon T_* x \mathcal{D}\}.
$$

Thus, if the convergence of the estimator is in the metric defined in RHS (1.20), the empirical Bayes procedure based on $\hat{G}$ will be asymptotically optimal (a.o.).

Theorem 1.1. Suppose that $\Theta = \{0,1,\ldots,a\}$ and that the loss function L is bounded. Suppose that T_* is a specified subset of T, R_* denotes the associated envelope risk function and that $R_*(G)$ is attained by $(\underline{\tau}^*(G),\underline{\delta}(G))$, $G\varepsilon\mathcal{G}$. Suppose further that $\hat{G}_n = G_n(\underline{X}^{N_1},\ldots,\underline{X}^{N_{n-1}})$, $n = 2,3,\ldots$ is an a.s. consistent estimator of $G\varepsilon\mathcal{G}$. (Here we identify $\mathcal{G}$ with the a-dimensional simplex of probability vectors on Θ and we denote the sup norm on $\mathcal{X}^{a+1}$ by $||\ ||$). Then the empirical Bayes procedure $\underline{d}^n = (\underline{\tau}^*(\hat{G}_n),\underline{\delta}(\hat{G}_n))$ is a.o. on $\mathcal{G}$.

<u>Proof.</u> Since L is bounded, the risk set associated with the component is bounded; specifically, $0 \leq R(\theta,(\underline{\tau},\underline{\delta})) \leq L + c\,m$ for $\theta = 0,1,\ldots,a$. Hence, RHS(1.20) $\leq 2(L + c\,m)||\hat{G}_n - G||$ so that (1.20) with $\hat{G} = \hat{G}_n$ implies $R(G,\underline{d}^n) \to R_\star(G)$ a.s., $G\varepsilon\mathcal{G}$. Taking expectation with respect to $\underline{X}^{N_1},\ldots,\underline{X}^{N_{n-1}}$ establishes the L_1 convergence, which is sense in which asymptotic optimality is usually defined. □

Of course, choices $T_\star = T_i$, $T_\star = T_F$, $T_\star = T$ result in envelopes $R_\star = R_i$, $R_\star = R_F$, $R_\star = R_B$ defined in the last section and displayed for a particular component in Example 1.1. For a component where $T_\star$ includes truly sequential stopping rules, the implementation of the empirical Bayes stopping rule $\underline{\tau}(\hat{G}_n)$ may require considerable calculation involving backward induction.

We will not attempt to generalize Theorem 1.1 to cover a general infinite Θ component. In the next chapter we investigate empirical Bayes one-step look ahead sequential procedures for certain components where the structure is sufficiently tractable to analysis, namely certain finite action components.

CHAPTER 2

TWO ACTION AND MULTIPLE DECISION PROBLEMS

2.1 Two Action Problem

In this chapter we will establish asymptotic results for empirical Bayes sequential decision procedures in the cases of linear loss testing and the multiple decision problem components.

In this section we describe the linear loss testing problem where two composite hypotheses are tested against each other and derive the one-step look procedure $\underline{d}^L = (\underline{\tau}^L(G), \underline{\delta}(G))$ relative to G for the linear loss two-action problem. An empirical Bayes sequential decision procedure will be constructed in the next section and asymptotic results will be given in Section 2.3. In Section 2.4 we formulate the one-step look ahead procedure for the general multiple decision problem and exhibit an empirical Bayes sequential decision procedure with asymptotic results. Throughout this chapter, we will assume that the parameter space Θ is a subset of the real line and $f_\theta \geq 0$ is a density function of the distribution P_θ with respect to a given σ-finite measure μ on $(X, \mathcal{B})$. To conserve notation we will also let $f_\theta(\underline{x}^k)$ denote the product $f_\theta(x_1)\ldots f_\theta(x_k)$ for $\underline{x}^k \varepsilon X^k$, $k \geq 1$.

We wish to test the hypothesis

$$H_0:\ \theta \leq \theta_0 \qquad \text{against} \qquad H_1:\ \theta > \theta_0$$

where $\theta_0 \varepsilon \Theta$. Consequently, the action space A consists of two actions only, namely $\{a_0,a_1\}$ where a_0 and a_1 denote the actions of deciding H_0 and H_1, respectively. We assume a linear loss function, specifically,

$$(2.1) \qquad L(\theta,a_0) = (\theta - \theta_0)^+, \quad L(\theta,a_1) = (\theta - \theta_0)^-, \ \theta\varepsilon\Theta \ .$$

We assume that the first moment of θ is finite with respect to G, where G is the prior distribution on Θ. This is sufficient to ensure that the Bayes risk of the m-truncated one-step look ahead procedure defined in Section 1.2 is finite.

In the literature many authors have studied the two-action problem from the standard empirical Bayes point of view. Samuel (1963) discussed the two-action problem and exhibited a.o. empirical Bayes tests under various loss structures and, in part, dealt specifically with certain types of discrete exponential families. Yu (1970), Johns and Van Ryzin (1971, 1972) considered the linear loss two-action problem with exponential families and developed rates of convergence in the regret $ER(G,\delta_n) - R_B(G)$ where δ_n is an empirical Bayes test for the fixed sample size linear loss two-action problem. Van Houwelingen (1976) proposed monotonizing empirical Bayes tests defined in Johns and Van Ryzin (1971, 1972). O'Bryan (1972), O'Bryan and Susarla (1975) treated the testing problem where the sequence of component problems consists of independent but not identical decision problems, all having the same unknown prior distribution. These sequences of decision problems are identical except for the sample size. Laippala (1979, 1980, 1983, 1985) discussed the

two-action problem with varying random sample sizes in Binomial and exponential conditional distributions.

Now let us determine the m-truncated one-step look ahead procedure $\underline{d}^L = (\underline{\tau}^L(G), \underline{\delta}(G))$ with respect to G for our testing problem (see Section 1.2). The terminal decision rule $\underline{\delta}(G)$ is a finite sequence $(\delta_1, \ldots, \delta_m)$ where $\delta_k \varepsilon \mathcal{D}^k (k=1,\ldots,m)$ is a Bayes decision function relative to G for the fixed sample size k testing problem (Θ, A, L) with the loss function (2.1).

For $k \geq 1$, δ_k can be determined as follows.

Let $\delta_k(\underline{x}^k) = \Pr\{\text{choosing } a_0 | X_1 = x_1, \ldots, X_k = x_k\}$ be a randomized decision for the two-action problem with the loss function (2.1) given that the observations are $X_1 = x_1, \ldots, X_k = x_k$. Then the Bayes risk of decision function δ_k relative to the prior distribution G is given by

$$r(G, \delta_k) = \int_{\mathcal{X}^k} \int_\Theta \{L(\theta, a_0)\delta_k(\underline{x}^k) + L(\theta, a_1)(1-\delta_k(\underline{x}^k))\} P_\theta^k(d\underline{x}^k) G(d\theta).$$

Since $f_\theta(\underline{x}^k)$ is a conditional μ^k-density of $\underline{X}^k = (X_1, \ldots, X_k)$, and $L(\theta, a_0) - L(\theta, a_1) = \theta - \theta_0$ (see (2.1)), one can write

$$r(G, \delta_k) = \int_{\mathcal{X}^k} \alpha_k(\underline{x}^k)\, \delta_k(\underline{x}^k)\, \mu^k\, (d\underline{x}^k) + C_G \tag{2.2}$$

where $C_G = \int_\Theta L(\theta, a_1) G(d\theta)$ and

$$\alpha_k(\underline{x}^k) = \int_\Theta (\theta - \theta_0)\, f_\theta(\underline{x}^k)\, G(d\theta)\ . \tag{2.3}$$

From (2.2) and (2.3) it is clear that a Bayes rule (a minimizer of $r(G, \delta_k)$ for given G) is provided by the nonrandomized rule

$$(2.4)\qquad \delta_k(\underline{x}^k) = \begin{cases} 1 & \text{if } \alpha_k(\underline{x}^k) \le 0 \\ 0 & \text{if } \alpha_k(\underline{x}^k) > 0 . \end{cases}$$

Note that we have suppressed the display of dependence of the Bayes rule $\delta_k(\underline{x}^k)$ on G. Henceforth, the terminal decision rule $\underline{\delta}(G) = (\delta_1,\ldots,\delta_m)$ for the m-truncated one-step look ahead procedure $\underline{d}^L = (\underline{\tau}^L(G),\underline{\delta}(G))$ defined in Section 1.2 is given by (2.4) and (2.3), $k = 1,2,\ldots,m$. Then the minimum posterior Bayes risk $r(G_k)$ with respect to G_k, for $k \ge 1$, is given by

$$r(G_k) = \int_\Theta L(\theta,\delta_k(\underline{x}^k))\; G_k(d\theta)$$

where G_k is the posterior distribution of θ given $\underline{X}^k = \underline{x}^k$. Thus for the linear loss testing component,

$$(2.5)\qquad \begin{aligned} r(G_k) &= \int_\Theta \{L(\theta,a_0)\delta_k(\underline{x}^k) + L(\theta,a_1)(1-\delta_k(\underline{x}^k))\}\; G_k(d\theta) \\ &= \int_\Theta \{(L(\theta,a_0) - L(\theta,a_1))\delta_k(\underline{x}^k) + L(\theta,a_1)\}\; G_k(d\theta). \end{aligned}$$

Since $L(\theta,a_0) - L(\theta,a_1) = \theta - \theta_0$, (2.5) can be written as

$$(2.6)\qquad \begin{aligned} r(G_k) &= C_{G_k} + \delta_k(\underline{x}^k) \int_\Theta (\theta-\theta_0)G_k(d\theta) \\ &= \frac{1}{f_k(\underline{x}^k)} \int_\Theta L(\theta,a_1)f_\theta(\underline{x}^k)G(d\theta) + \frac{\delta_k(\underline{x}^k)}{f_k(\underline{x}^k)}\; \alpha_k(\underline{x}^k) \end{aligned}$$

provided $f_k(\underline{x}^k) > 0$ and $r(G_k) = 0$ if $f_k(\underline{x}^k) = 0$ where for $k \ge 1$, $f_k(\underline{x}^k) = \int_\Theta f_\theta(\underline{x}^k)G(d\theta)$, and α_k is given by equation (2.3).

The expected risk $E^*r(G_{k+1})$ from taking X_{k+1} observation and playing Bayes conditional on $\underline{X}^k = \underline{x}^k$ is given by

$$E^{*}r(G_{k+1}) = \int_{X} r(G_{k+1})f^{*}(x_{k+1})\mu(dx_{k+1}) \tag{2.7}$$

where $f^{*}(x_{k+1})$ is a conditional density of X_{k+1} given $\underline{X}^{k} = \underline{x}^{k}$. Let $f_{k+1}(\underline{x}^{k+1})$ and $f_{k}(\underline{x}^{k})$ be unconditional marginal densities of $\underline{X}^{k+1}$ and $\underline{X}^{k}$ respectively. Then a conditional density of X_{k+1} given $\underline{X}^{k} = \underline{x}^{k}$ is given by $f^{*}(x_{k+1}) = f_{k+1}(\underline{x}^{k+1})/f_{k}(\underline{x}^{k})$ if $f_{k}(\underline{x}^{k}) > 0$ and $f^{*}(x_{k+1}) = 0$ if $f_{k}(\underline{x}^{k}) = 0$. Using this fact and changing k to $k + 1$ in equation (2.6) and then substituting back in (2.7) we get

$$\begin{aligned} E^{*}r(G_{k+1}) = {} & \frac{1}{f_{k}(\underline{x}^{k})} \int_{\Theta} L(\theta,a_{1})f_{\theta}(\underline{x}^{k})G(d\theta) + \\ & \frac{1}{f_{k}(\underline{x}^{k})} \int_{X} \delta_{k+1}(\underline{x}^{k+1})\alpha_{k+1}(\underline{x}^{k+1})\mu(dx_{k+1}) \end{aligned} \tag{2.8}$$

provided $f_{k}(\underline{x}^{k}) > 0$ and $E^{*}r(G_{k+1}) = 0$ if $f_{k}(\underline{x}^{k}) = 0$ where δ_{k+1} and α_{k+1} are given by (2.4) and (2.3), respectively, with k replaced by $k+1$. Letting

$$\rho_{k}(\underline{x}^{k}) = \int_{X} [\alpha_{k+1} \leq 0]\, \alpha_{k+1}\, \mu(dx_{k+1}) + cf_{k}(\underline{x}^{k}) - [\alpha_{k} \leq 0]\, \alpha_{k}(\underline{x}^{k}), \tag{2.9}$$

we observe that $\rho_{k}(\underline{x}^{k}) = 0$ when $f_{k}(\underline{x}^{k}) = 0$ and

$$E^{*}r(G_{k+1}) + c - r(G_{k}) = \begin{cases} \rho_{k}(\underline{x}^{k})/f_{k}(\underline{x}^{k}) & \text{if } f_{k}(\underline{x}^{k}) > 0 \\ c & \text{if } f_{k}(\underline{x}^{k}) = 0 . \end{cases}$$

Hence, the stopping rule $\underline{\tau}^{L} = (\tau_{1}^{L},\ldots,\tau_{m}^{L})$ of m-truncated one-step look ahead procedure (1.10) for our testing problem is defined by $\tau_m^L \equiv 1$, and for $k = 1,\ldots,m-1$ by

$$(2.10) \qquad \tau_k^L(\underline{x}^k) = \begin{cases} 1 & \text{if } \rho_k(\underline{x}^k) \geq 0 \\ 0 & \text{if } \rho_k(\underline{x}^k) < 0 \end{cases}$$

where we have suppressed the display of dependence on G.

Let N^L be the stopping time of m-truncated one-step look ahead procedure $\underline{d}^L = (\underline{\tau}^L(G),\underline{\delta}(G))$; then

$$N^L(\underline{X}^m) = \min\,\{k|\tau_k^L(\underline{X}^k) = 1\}$$

and since $\tau_m^L \equiv 1$, sampling will be stopped after X_m has been observed if it had not been stopped earlier. The risk of this procedure at G is (see (1.1) and (1.2)),

$$(2.11) \qquad R_L(G) = \sum_{k=1}^{m} \int_\Theta \int_{X^m} [N^L=k](L(\theta,\delta_k(\underline{x}^k)) + ck)\; f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

where $\delta_k = [\alpha_k \leq 0]$ by (2.4), α_k is given by the equation (2.3), and L is defined by (2.1). By the definition of L we get $L(\theta,\delta_k) = L(\theta,a_1) + (\theta - \theta_0)\delta_k$, since $L(\theta,a_0) = L(\theta,a_1) = \theta - \theta_0$. Using this fact and the definition of C_G following (2.2), we have

$$(2.12) \qquad R_L(G) = C_G + \sum_{k=1}^{m} \int_\Theta \int_{X^m} [N^L = k]([\alpha_k \leq 0](\theta - \theta_0) + ck)\; f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)\;.$$

The empirical Bayes approach applied to this problem can be based upon estimation of the functions $[N^L = k]$ and α_k, $k = 1,2,\ldots,m$. For this purpose, it is useful to decompose and represent the indicator functions as follows. Note that for $k = 1,2,\ldots,m-1$ $[N^L = k] = A_k + B_k$

and $[N^L = m] = A_m$ where

$$A_k = [\rho_1 < 0]..........[\rho_{k-1} < 0][\rho_k > 0] \quad \text{for} \quad k = 1,.....m-1$$

(2.13)
$$B_k = [\rho_1 < 0]..........[\rho_{k-1} < 0][\rho_k = 0] \quad \text{for} \quad k = 1,.....m-1$$

$$A_m = [\rho_1 < 0]..........[\rho_{m-1} < 0].$$

Thus, the Bayes risk of m-truncated one-step look ahead procedure $\underline{d}^L = (\underline{\tau}^L(G),\underline{\delta}(G))$ relative to G for our testing problem can be written in the following form (see (2.12)):

(2.14)
$$\begin{aligned} R_L(G) = C_G &+ \sum_{k=1}^{m} \int_\Theta \int_{X^m} A_k([\alpha_k \leq 0](\theta - \theta_0) + ck)\ f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) \\ &+ \sum_{k=1}^{m-1} \int_\Theta \int_{X^m} B_k([\alpha_k \leq 0](\theta - \theta_0) + ck) f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) \ . \end{aligned}$$

The inequalities defining the indicators correspond to open sets for the A_k and boundary sets for the B_k. In the empirical Bayes application, the functions ρ_k are estimated so that a separate treatment of boundaries is important in so far as convergences are concerned.

2.2 An Empirical Bayes One-step Sequential Decision Procedure for the Two Action Component Case

Suppose that the prior distribution G is unknown but fixed; then the classical Bayes quantities (2.3)-(2.9) of the Section 2.1 are not available to the statistician. However, suppose that we are experiencing independent repetitions of the same component problem. Then applying the empirical Bayes approach introduced by Robbins (1956), we may derive empirical Bayes estimates of the classical Bayes quantities (2.3)-(2.9), and, hence, an empirical Bayes one-step sequential decision procedure $\underline{d}^n = (\underline{\tau}^n, \underline{\delta}^n)$, where $\underline{\tau}^n$ is an empirical Bayes stopping rule and $\underline{\delta}^n$ is an empirical Bayes terminal decision rule.

In order to construct an empirical Bayes sequential decision procedure $\underline{d}^n = (\underline{\tau}^n, \underline{\delta}^n)$, we will make the following assumptions on conditional density $f_\theta(x)$ and the parameter space Θ.

(A1) For each x, $f_\theta(x)$ is a continuous function of θ.

(A2) Θ is compact.

At the nth problem of the repetitions, we will have observed the random vectors $\underline{X}_1^{N_1}, \ldots, \underline{X}_{n-1}^{N_{n-1}}$ from the past (n-1) repetitions of the component problem 2.1, where $N_1, \ldots, N_{n-1}$ are the respective stopping times of the past repetitions. Let $\{G_n\}$ be a sequence of distribution functions on Θ, where $G_n(\theta) = G_n(\theta, \underline{X}_1^{N_1}, \ldots, \underline{X}_{n-1}^{N_{n-1}})$ depends only on the random vectors $\underline{X}_1^{N_1}, \ldots, \underline{X}_{n-1}^{N_{n-1}}$, which converges

weakly to the prior distribution function G with probability one as $n \to \infty$, that is,

$$\Pr \{\lim_{n\to\infty} G_n(\theta) = G(\theta), \text{ any continuity point of } G\} = 1.$$

<u>Remark 2.1.</u> Robbins (1964) showed the existence of such a sequence of distribution functions on $\Theta = (a,b)$, $-\infty \le a < b \le \infty$, under the following assumptions.

(a) For each x, $F_\theta(x)$ is a continuous function of θ, where $F_\theta(x)$ is the conditional distribution function of X.

(b) If G_1 and G_2 are two distributions on Θ such that $F_{G_1} = F_{G_2}$ then $G_1 = G_2$, where $F_G(x) = \int_\Theta F_\theta(x)G(d\theta)$.

(c) The limits $\lim_{\theta\to a} F_\theta(x)$ and $\lim_{\theta\to b} F_\theta(x)$ exist for each x and neither $\lim_{\theta\to a} F_\theta(x)$ nor $\lim_{\theta\to b} F_\theta(x)$ is a distribution function.

He showed that when Θ is a compact subset of R, condition (c) can be relaxed.

Now we define our empirical Bayes sequential decision (EBSD) procedure $\underline{d}^n = (\underline{\tau}^n, \underline{\delta}^n)$ as follows.

Let $\underline{\delta}^n$ be a finite sequence of functions $(\delta_1^n,\ldots,\delta_m^n)$, where δ_k^n is such that $\delta_k^n(\underline{x}^k) = \Pr\{\text{choosing } a_0 | \underline{X}^k = \underline{x}^k\}$ and, motivated by (2.4) and (2.3),

$$(2.15) \qquad \delta_k^n(\underline{x}^k) = \begin{cases} 1 & \text{if } \alpha_k^n(\underline{x}^k) \le 0 \\ 0 & \text{if } \alpha_k^n(\underline{x}^k) > 0 \end{cases}$$

and

$$(2.16) \qquad \alpha_k^n(\underline{x}^k) = \int_\Theta (\theta - \theta_0) f_\theta(\underline{x}^k) G_n(d\theta) .$$

Let $\underline{\tau}^n$ be a stopping rule consisting of a finite sequence of functions $(\tau_1^n,\ldots,\tau_m^n)$ where, motivated by (2.10) and (2.9), $\tau_m^n \equiv 1$ and, for $k = 1,\ldots,m-1$,

$$(2.17) \qquad \tau_k^n(\underline{x}^k) = \begin{cases} 1 & \text{if} \quad \rho_k^n(\underline{x}^k) \geq 0 \\ 0 & \text{if} \quad \rho_k^n(\underline{x}^k) < 0 \end{cases}$$

where

$$(2.18) \qquad \rho_k^n(\underline{x}^k) = \int_X \delta_{k+1}^n(\underline{x}^{k+1})\, \alpha_{k+1}^n(\underline{x}^{k+1})\mu(dx_{k+1}) + cf_k^n(\underline{x}^k) - \delta_k^n(\underline{x}^k)\alpha_k^n(\underline{x}^k)$$

with $f_k^n(\underline{x}^k) = \int_\Theta f_\theta(\underline{x}^k)G_n(d\theta)$. Since $\tau_m^n \equiv 1$, sampling will be stopped just after X_m has been observed if it had not been stopped earlier.

For investigating the risk of the EBSD it is useful to define

$$(2.19) \qquad \begin{aligned} C_k^n &= [\rho_1^n < 0]\ldots\ldots\ldots[\rho_{k-1}^n < 0][\rho_k^n \geq 0] \quad \text{for} \quad k = 1,\ldots,m-1 \\ C_m^n &= [\rho_1^n < 0]\ldots\ldots\ldots[\rho_{m-1}^n < 0] \ . \end{aligned}$$

Then $[N^n = k] = C_k^n$ for $k = 1,\ldots,m-1$ and $[N^n = m] = C_m^n$, where N^n denotes the stopping time of the EBSD procedure $\underline{d}^n = (\underline{\tau}^n,\underline{\delta}^n)$. Note that $\sum_{k=0}^{m} [N^n = k] = 1$ implies $\sum_{k=1}^{m} C_k^n = 1$.

Let $R(G,\underline{d}^n)$ denote the conditional Bayes risk of $\underline{d}^n = (\underline{\tau}^n,\underline{\delta}^n)$ with respect to G. Then since the C_k^n partition X^m ,

$$(2.20) \qquad R(G,\underline{d}^n) = C_G + \sum_{k=1}^{m} \int_\Theta \int_{X^m} C_k^n([\alpha_k^n \leq 0](\theta - \theta_0) + $$

$$+ ck)\ f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)\ .$$

In the next section we will treat the difference between the empirical Bayes risk $R(G,\underline{d}^n)$ and the one-step look ahead risk $R_L(G)$ using the decompositions (2.20) and (2.14).

2.3 Asymptotic Results for the Two Action Component Case

In this section we compare the asymptotic behavior of unconditional Bayes risk of the Section 2.2 EBSD procedure $\underline{d}^n$ with the Bayes risk of m-truncated one-step look ahead procedure $\underline{d}^L$ discussed in Section 2.1. First we prove the following useful lemma. Convergence of sequences of functions on X^m is understood to be pointwise convergence.

Lemma 2.1: Under the assumptions (A1) and (A2) for $k \geq 1$,

$$\rho_k^n \to \rho_k \quad \text{w.p.1} \quad \text{as} \quad n \to \infty \ .$$

Proof: From (2.9) and (2.18) with dependence on $\underline{x}^k$ suppressed,

$$\rho_k = \int_X [\alpha_{k+1} \leq 0]\alpha_{k+1}\mu(dx_{k+1}) + cf_k - [\alpha_k \leq 0]\alpha_k$$

and

$$\rho_k^n = \int_X [\alpha_{k+1}^n \leq 0]\alpha_{k+1}^n\mu(dx_{k+1}) + cf_k^n - [\alpha_k^n \leq 0]\alpha_k^n \ .$$

By assumptions (A1) and (A2), $f_\theta(\underline{x}^k)$ is a bounded and continuous function of θ for each $\underline{x}^k$. Then recalling the definitions of α_k^n, α_k, f_k^n, f_k (see (2.16), (2.3), below (2.18) and below (2.6)) and the assumptions on the sequence $\{G_n\}$, we get for $k \geq 1$,

$$\text{(i)} \quad \alpha_k^n \to \alpha_k \quad \text{w.p.1} \quad \text{as} \quad n \to \infty$$

and

$$\text{(ii)} \quad f_k^n \to f_k \quad \text{w.p.1} \quad \text{as} \quad n \to \infty \ .$$

From (i) it follows that

(2.21) $\quad [\alpha_k^n \le 0]\alpha_k^n \to [\alpha_k \le 0]\alpha_k$ w.p.1 as $n \to \infty$, $k \ge 1$.

To show that $\int_X [\alpha_{k+1}^n \le 0]\alpha_{k+1}^n \mu(dx_{k+1}) \to \int_X [\alpha_{k+1} \le 0]\alpha_{k+1}\mu(dx_{k+1})$

w.p.1 as $n \to \infty$, we use the Generalized Dominated Convergence Theorem (GDCT).

Note that $|\alpha_{k+1}^n(\underline{x}^{k+1})[\alpha_{k+1}^n \le 0]| \le h_{k+1}^n(\underline{x}^{k+1})$ where $h_{k+1}^n(\underline{x}^{k+1}) = \int|\theta - \theta_0| f_\theta(\underline{x}^{k+1})G_n(d\theta)$. Then $\lim_{n\to\infty} \int h_{k+1}^n(\underline{x}^{k+1})\mu(dx_{k+1}) =$ $\lim_{n\to\infty} \int \int|\theta - \theta_0| f_\theta(\underline{x}^{k+1})G_n(d\theta)\mu(dx_{k+1}) = \int|\theta - \theta_0| f_\theta(\underline{x}^k)G(d\theta)$ w.p.1 by the assumptions (A1) and (A2). But $\int|\theta - \theta_0| f_\theta(\underline{x}^k)G(d\theta)$ is equal to $\int \int|\theta - \theta_0| f_\theta(\underline{x}^{k+1})G(d\theta)\mu(dx_{k+1})$ and note that the following equality is satisfied:

$$\int \int |\theta - \theta_0| f_\theta(\underline{x}^{k+1})G(d\theta)\mu(dx_{k+1}) = \int \lim_{n\to\infty} \int |\theta - \theta_0| f_\theta(\underline{x}^{k+1})G_n(d\theta)\mu(dx_{k+1})$$

$$= \int \lim_{n\to\infty} h_{k+1}^n(\underline{x}^{k+1})\mu(dx_{k+1}) \text{ w.p.1} .$$

Now use the GDCT and (2.21) to get the required results. $\square$

Now we state and prove a theorem which concerns the asymptotic behavior of the conditional Bayes risk of empirical Bayes sequential decision procedure $\underline{d}^n = (\underline{\tau}^n, \underline{\delta}^n)$ defined in the previous section.

<u>Theorem 2.1</u>: Under the assumptions (A1) and (A2)

(2.22) $\quad \limsup_{n\to\infty} R(G,\underline{d}^n) \le R_L(G)$ w.p.1

where $R(G,\underline{d}^n)$, and $R_L(G)$ are given by the equations (2.20) and (2.14), respectively.

Proof: The difference of $R(G,\underline{d}^n)$ and $R_L(G)$ can be written as

$$R(G,d^n) - R_L(G) = \sum_{j=1}^{m}\sum_{i=1}^{m} \int\int C_i^n A_j\{([\alpha_i^n \le 0] - [\alpha_j \le 0])(\theta - \theta_0) + c(i - j)\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

$$(2.23) \qquad + \sum_{j=1}^{m-1}\sum_{i=1}^{m} \int\int C_i^n B_j\{([\alpha_i^n \le 0] - [\alpha_j \le 0])(\theta - \theta_0) + c(i - j)\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) .$$

The first double sum in (2.23) can be written as $J_n^1 + J_n^2 + J_n^3$, where

$$J_n^1 = \sum_{j=2}^{m}\sum_{i=1}^{j-1} \int\int C_i^n A_j\{([\alpha_i^n \le 0] - [\alpha_j \le 0])(\theta - \theta_0) + c(i - j)\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

$$J_n^2 = \sum_{j=1}^{m-1}\sum_{i=j+1}^{m} \int\int C_i^n A_j\{([\alpha_i^n \le 0] - [\alpha_j \le 0])(\theta - \theta_0) + c(i - j)\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

and

$$J_n^3 = \sum_{i=1}^{m} \int\int C_i^n A_i\{([\alpha_i^n \le 0] - [\alpha_i \le 0])(\theta - \theta_0)\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) .$$

We will show that $J_n^i \to 0$ w.p.1 as $n \to \infty$, $i = 1,2,3$, and, hence, the first double sum in (2.23) goes to zero w.p.1.

By the definitions of C_i^n, A_j and B_j note that for $i < j$, $C_i^n A_j \le [\rho_i^n \ge 0][\rho_i < 0]$, and for $i > j$, $C_i^n A_j \le [\rho_j^n < 0][\rho_j > 0]$. Then observe that

$$|J_n^1| \le \sum_{j=2}^{m} \sum_{i=1}^{j-1} \int \int [\rho_i^n \ge 0][\rho_i < 0]\{2|\theta - \theta_0| + c|i - j|\}f_\theta(\underline{x}^m) G(d\theta)\mu^m(d\underline{x}^m)$$

(2.24)

$$\le \sum_{j=2}^{m} \sum_{i=1}^{j-1} \int \int [\rho_i < 0][|\rho_i^n - \rho_i| \ge |\rho_i|]\{2|\theta - \theta_0| + c|i - j|\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) .$$

But $\int|\theta - \theta_0|G(d\theta) < \infty$, and $[\rho_i < 0][|\rho_i^n - \rho_i| \ge |\rho_i|] \to 0$ w.p.1 as $n \to \infty$ by Lemma 2.1, and hence the DCT gives $|J_n^1| \to 0$ w.p.1 as $n \to \infty$. Similarly, one can show that $|J_n^2| \to 0$ w.p.1 as $n \to \infty$.

Since C_i^n, A_i, α_i^n and α_i depend only on the first i observations $(X_1,\ldots,X_i)$, J_n^3 can be written as

$$J_n^3 = \sum_{i=1}^{m} \int_{X^i} C_i^n A_i \{[\alpha_i^n \le 0] - [\alpha_i \le 0]\}\alpha_i \mu^i(d\underline{x}^i) .$$

But for $i \ge 1$, $|([\alpha_i^n \le 0] - [\alpha_i \le 0])\alpha_i| \le [|\alpha_i^n - \alpha_i| \ge |\alpha_i|]|\alpha_i|$

so that

$$|J_n^3| \le \sum_{i=1}^{m} \int_{X^i} [|\alpha_i^n - \alpha_i| \ge |\alpha_i|] |\alpha_i|\mu^i(d\underline{x}^i) .$$

But $\int |\theta - \theta_0|G(d\theta) < \infty$ implies $\int_{X^i}|\alpha_i|\mu^i(d\underline{x}^i) < \infty$, and for $i \ge 1$, $\alpha_i^n \to \alpha_i$ w.p.1 as $n \to \infty$ (see proof of Lemma 2.1) implies for $i \ge 1$,

(2.25) $\quad |\alpha_i| [|\alpha_i^n - \alpha_i| \ge |\alpha_i|] \to 0$ w.p.1 as $n \to \infty$.

Now apply the DCT to get $|J_n^3| \to 0$ w.p.1 as $n \to \infty$.

To derive the asymptotic behaviour of the second double sum in (2.23), observe that it can be written as $K_n^1 + K_n^2 + K_n^3$ where

$$K_n^1 = \sum_{j=2}^{m-1} \sum_{i=1}^{j-1} \int \int C_i^n B_j \{([\alpha_i^n \le 0] - [\alpha_j \le 0])(\theta - \theta_0) + c(i - j)\} f_\theta(\underline{x}^m) G(d\theta) \mu^m(d\underline{x}^m) ,$$

$$K_n^2 = \sum_{i=1}^{m-1} \int \int C_i^n B_i \{([\alpha_i^n \le 0] - [\alpha_i \le 0])(\theta - \theta_0)\} f_\theta(\underline{x}^m) G(d\theta) \mu^m(d\underline{x}^m)$$

and

$$K_n^3 = \sum_{j=1}^{m-1} \sum_{i=j+1}^{m} \int \int C_i^n B_j \{([\alpha_i^n \le 0] - [\alpha_j \le 0])(\theta - \theta_0) + c(i - j)\} f_\theta(\underline{x}^m) G(d\theta) \mu^m(d\underline{x}^m) . \tag{2.26}$$

We will show that $K_n^i \to 0$ w.p.1 as $n \to \infty$, $i = 1,2,$ and $\limsup_{n\to\infty} K_n^3 \le 0$ w.p.1 .

For $i < j$, $C_i^n B_j \le [\rho_i^n \ge 0][\rho_i < 0]$ so that

$$|K_n^1| \le \sum_{j=2}^{m-1} \sum_{i=1}^{j-1} \int \int [\rho_i^n \ge 0][\rho_i < 0] \{2|\theta - \theta_0| + c|i - j|\} f_\theta(\underline{x}^m) G(d\theta) \mu^m(d\underline{x}^m)$$

$$\le \sum_{j=2}^{m-1} \sum_{i=1}^{j-1} \int \int [\rho_i < 0][|\rho_i^n - \rho_i| \ge |\rho_i|] \{2|\theta - \theta_0| + c|i - j|\} f_\theta(\underline{x}^m) G(d\theta) \mu^m(d\underline{x}^m) .$$

This is the bound (2.24) for $|J_n^1|$ which was shown to converge to zero w.p.1 as $n \to \infty$. Hence, $|K_n^1| \to 0$ w.p.1 as $n \to \infty$.

Now note that K_n^2 can be written as

$$K_n^2 = \sum_{i=1}^{m-1} \int C_i^n B_i \{[\alpha_i^n \le 0] - [\alpha_i \le 0]\}\alpha_i \mu^i(d\underline{x}^i) .$$

The above equality follows from the fact that the functions C_i^n, B_i, α_i^n and α_i depend only on the first i observations $(X_1,\ldots,X_i)$. Thus,

$$|K_n^2| \le \sum_{i=1}^{m-1} \int [|\alpha_i^n - \alpha_i| \ge |\alpha_i|]|\alpha_i|\mu^i(d\underline{x}^i) .$$

This is the bound for $|J_n^3|$ which was shown to converge to zero w.p.1.

We summarize the results concerning the difference (2.21) obtained so far with

$$R(G,\underline{d}^n) - R_L(G) = \sum_{j=1}^{3} J_n^j + \sum_{j=1}^{3} K_n^j \tag{2.27}$$

where J_n^1, J_n^2, J_n^3, K_n^1, K_n^2 have been shown to converge to zero w.p.1.

From (2.26) we can write

$$K_n^3 = \sum_{j=1}^{m-1} (L_1(n,j) + L_2(n,j) + L_3(n,j)) \tag{2.28}$$

where

$$L_1(n,j) = \sum_{i=j+1}^{m} \int\int C_i^n B_j [\alpha_i^n \le 0](\theta - \theta_0)f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) \tag{2.29}$$

$$L_2(n,j) = -\sum_{i=j+1}^{m} \int\int C_i^n B_j [\alpha_j \le 0](\theta - \theta_0)f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) \tag{2.30}$$

and

$$L_3(n,j) = \sum_{i=j+1}^{m} \int\int C_i^n B_j c(i - j)f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) . \tag{2.31}$$

We define $\hat{C}_k^n = \sum_{i=k}^{m} C_i^n$ for $k = j+1,\ldots,m$, $j = 1,\ldots, m-1$ and observe that $\hat{C}_k^n = [\rho_1^n < 0]\ldots\ldots\ldots[\rho_{k-1}^n < 0]$ for $k = j+1,\ldots, m$, and, hence, $\hat{C}_k^n$ depends only on the first $k-1$ observations $(X_1,\ldots,X_{k-1})$. Now using the definition of $\hat{C}_k^n$, $k \geq 1$, we can write $L_1(n,j)$, $L_2(n,j)$ and $L_3(n,j)$ as follows:

$$
\begin{aligned}
L_3(n,j) &= c \sum_{k=j+1}^{m} \int\int \hat{C}_k^n B_j f_\theta(\underline{x}^m) G(d\theta)\mu^m(d\underline{x}^m) \\
&= c \sum_{k=j+1}^{m} \int \hat{C}_k^n B_j f_{k-1}(\underline{x}^{k-1})\mu^{k-1}(d\underline{x}^{k-1}) \qquad (2.32)\\
&= \sum_{k=j+1}^{m} \Delta_k(n,j)
\end{aligned}
$$

where

$$\Delta_k(n,j) = c \int \hat{C}_k^n B_j f_{k-1}(\underline{x}^{k-1})\mu^{k-1}(d\underline{x}^{k-1}) .$$

Also

$$
\begin{aligned}
L_2(n,j) &= -\int\int (C_{j+1}^n + \ldots + C_m^n) B_j[\alpha_j \leq 0](\theta - \theta_0) f_\theta(\underline{x}^m) \\
&\qquad\qquad G(d\theta)\mu^m(d\underline{x}^m) \\
&= -\int\int \hat{C}_{j+1}^n B_j[\alpha_j \leq 0](\theta - \theta_0) f_\theta(\underline{x}^m) G(d\theta)\mu^m(d\underline{x}^m) \qquad (2.33)\\
&= -\int \hat{C}_{j+1}^n B_j [\alpha_j \leq 0]\alpha_j \mu^j(d\underline{x}^j)
\end{aligned}
$$

and

$$
\begin{aligned}
L_1(n,j) &= \sum_{i=j+1}^{m} \int C_i^n B_j [\alpha_i^n \leq 0]\alpha_i \mu^i(d\underline{x}^i) \qquad (2.34)\\
&= \sum_{i=j+1}^{m} \int \hat{C}_i^n B_j [\alpha_i^n \leq 0]\alpha_i \mu^i(d\underline{x}^i) - \sum_{i=j+1}^{m-1} \int \hat{C}_{i+1}^n B_j \\
&\qquad\qquad [\alpha_i^n \leq 0]\alpha_i \mu^i(d\underline{x}^i) .
\end{aligned}
$$

The last equality follows from the fact that $C_i^n = \hat{C}_i^n - \hat{C}_{i+1}^n$ for $i=j+1,\ldots,m-1$, $j=1,\ldots,m-1$ and $C_m^n = \hat{C}_m^n$. We can write

$$(2.35) \qquad L_1(n,j) = \sum_{i=j+1}^{m} T_i(n,j) - \sum_{i=j+1}^{m-1} S_i(n,j)$$

where $T_i(n,j) = \int \hat{C}_i^n B_j [\alpha_i^n \le 0]\alpha_i\mu^i(d\underline{x}^i)$ and

$S_i(n,j) = \int \hat{C}_{i+1}^n B_j [\alpha_i^n \le 0]\alpha_i\mu^i(d\underline{x}^i)$. Thus the sum of $L_1(n,j)$, $L_2(n,j)$ and $L_3(n,j)$ can be written as

$$L_1(n,j) + L_2(n,j) + L_3(n,j) = \sum_{i=j+1}^{m} T_i(n,j) - \sum_{i=j+1}^{m-1} S_i(n,j)$$

$$- \int \hat{C}_{j+1}^n B_j [\alpha_j \le 0]\alpha_j\mu^j(d\underline{x}^j)$$

$$+ \sum_{k=j+1}^{m} \Delta_k(n,j)$$

$$(2.36) \qquad = T_{j+1}(n,j) - \int \hat{C}_{j+1}^n B_j [\alpha_j \le 0]\alpha_j\mu^j(d\underline{x}^j)$$

$$+ \Delta_{j+1}(n,j)$$

$$+ \sum_{i=j+2}^{m} [T_i(n,j) - S_{i-1}(n,j)$$

$$+ \Delta_i(n,j)]$$

$$= M_1(n,j) + M_2(n,j)$$

where

$$M_1(n,j) = T_{j+1}(j,n) - \int \hat{C}^n_{j+1}\ B_j\ [\alpha_j \le 0]\alpha_j \mu^j(d\underline{x}^j) + \Delta_{j+1}(n,j) \tag{2.37}$$

and

$$M_2(n,j) = \sum_{i=j+2}^{m} [T_i(n,j) - S_{i-1}(n,j) + \Delta_i(n,j)] \ . \tag{2.38}$$

Now observe that $M_1(n,j)$ is equal to

$$\int \hat{C}^n_{j+1}\ B_j\ \{\int[\alpha^n_{j+1} \le 0]\alpha_{j+1}\mu(dx_{j+1}) - [\alpha_j \le 0]\alpha_j + cf_j\}\mu^j(d\underline{x}^j) \ . \tag{2.39}$$

But $0 \le B_j \le [\rho_j = 0]$ and by the definition of ρ_j (see (2.9)) $cf_j - [\alpha_j \le 0]\alpha_j = -\int [\alpha_{j+1} \le 0]\alpha_{j+1}\mu(dx_{j+1})$ on $[\rho_j = 0]$. Then from (2.39),

$$\begin{aligned} |M_1(n,j)| &\le \int [\rho_j = 0](|\int [\alpha^n_{j+1} \le 0]\alpha_{j+1}\mu(dx_{j+1}) \\ &\qquad - \int [\alpha_{j+1} \le 0]\alpha_{j+1}\mu(dx_{j+1})|)\mu^j(d\underline{x}^j) \\ &\le \int [|\alpha^n_{j+1} - \alpha_{j+1}| \ge |\alpha_{j+1}|]|\alpha_{j+1}|\mu^{j+1}(d\underline{x}^{j+1}) . \end{aligned} \tag{2.40}$$

The R.H.S. of (2.40) goes to zero w.p.1 as $n \to \infty$, since $\alpha^n_{j+1} \to \alpha_{j+1}$ w.p.1 (see the proof of Lemma 2.1), $\int |\alpha_{j+1}|\mu^{j+1}(d\underline{x}^{j+1}) < \infty$ and by an application of the DCT. Thus, $M_1(n,j) \to 0$ w.p.1 as $n \to \infty$.

Now it remains to consider $M_2(n,j)$ in (2.38). Observe that for $i=j+2,\dots,m$, $T_i(n,j) - S_{i-1}(n,j) + \Delta_i(n,j)$ is equal to

$$\begin{aligned} &\int \hat{C}^n_i\ B_j\ [\alpha^n_i \le 0]\alpha_i\mu^i(d\underline{x}^i) - \int \hat{C}^n_i\ B_j\ [\alpha^n_{i-1} \le 0]\alpha_{i-1}\mu^{i-1}(d\underline{x}^{i-1}) \\ &\qquad + \int \hat{C}^n_i\ B_j\ cf_{i-1}(\underline{x}^{i-1})\mu^{i-1}(d\underline{x}^{i-1}) \ . \end{aligned}$$

The above sum can be written as

$$(2.41)\qquad \int \hat{C}_i^n B_j \left(\int [\alpha_i^n \le 0]\alpha_i \mu(dx_i) - [\alpha_{i-1}^n \le 0]\alpha_{i-1} + cf_{i-1}\right)\mu^{i-1}(d\underline{x}^{i-1}) .$$

Adding and subtracting the term $[\alpha_{i-1} \le 0]\alpha_{i-1}$ into the integrand of the above integral (2.41), we get

$$\int \hat{C}_i^n B_j \left(\int [\alpha_i^n \le 0]\alpha_i \mu(dx_i) - [\alpha_{i-1} \le 0]\alpha_{i-1} + cf_{i-1}\right)\mu^{i-1}(d\underline{x}^{i-1})$$

$$(2.42)\qquad + \int \hat{C}_i^n B_j \left([\alpha_{i-1} \le 0]\alpha_{i-1} - [\alpha_{i-1}^n \le 0]\alpha_{i-1}\right) \mu^{i-1}(d\underline{x}^{i-1}) .$$

Now use (2.25), $\int |\alpha_{i-1}|\mu^{i-1}(d\underline{x}^{i-1}) < \infty$, and the DCT to show that $\int \hat{C}_i^n B_j ([\alpha_{i-1} \le 0]\alpha_{i-1} - [\alpha_{i-1}^n \le 0]\alpha_{i-1})\mu^{i-1}(d\underline{x}^{i-1})$ goes to zero w.p.1 as $n \to \infty$. First integral in (2.42) can be rewritten as

$$\int \hat{C}_i^n B_j [\rho_{i-1} \le 0]\left(\int [\alpha_i^n \le 0]\alpha_i \mu(dx_i) - [\alpha_{i-1} \le 0]\alpha_{i-1} + cf_{i-1}\right)\mu^{i-1}(d\underline{x}^{i-1})$$

$$(2.43)\qquad + \int \hat{C}_i^n B_j [\rho_{i-1} > 0]\left(\int [\alpha_i^n \le 0]\alpha_i \mu(dx_i) - [\alpha_{i-1} \le 0]\alpha_{i-1} + cf_{i-1}\right)\mu^{i-1}(d\underline{x}^{i-1}) .$$

But $0 \le \hat{C}_i^n [\rho_{i-1} > 0] \le [\rho_{i-1}^n < 0][\rho_{i-1} > 0]$ and therefore the absolute value of the second integral in (2.43) is less than or equal to

$$\int [\rho^n_{i-1} < 0][\rho_{i-1} > 0]\{\int |\alpha_i|\mu(dx_i) + |\alpha_{i-1}| + cf_{i-1}\}\mu^{i-1}(d\underline{x}^{i-1}) , \tag{2.44}$$

and (2.44) is less than or equal to

$$\int [|\rho^n_{i-1} - \rho_{i-1}| \geq |\rho_{i-1}|][\rho_{i-1} > 0]\{\int|\alpha_i|\mu(dx_i) + |\alpha_{i-1}| + cf_{i-1}\}\mu^{i-1}(d\underline{x}^{i-1}) . \tag{2.45}$$

We use Lemma 2.1, $\int|\alpha_i| \mu^i(d\underline{x}^i) < \infty$,$\int|\alpha_{i-1}|\mu^{i-1}(d\underline{x}^{i-1}) < \infty$ and the DCT to conclude that (2.45) goes to zero w.p.1 as $n \to \infty$, and, hence, the second integral in (2.43) goes to zero w.p.1 as $n \to \infty$. By the definition of ρ_{i-1} (see (2.9)), observe that $cf_{i-1} - [\alpha_{i-1} \leq 0]\alpha_{i-1} \leq -\int [\alpha_i \leq 0]\alpha_i\mu(dx_i)$ on $[\rho_{i-1} \leq 0]$. Then the first integral in (2.43) is less than or equal to the following expression

$$\int \hat{C}^n_i B_j [\rho_{i-1} \leq 0](\int [\alpha^n_i \leq 0]\alpha_i\mu(dx_i) -\int [\alpha_i \leq 0]\alpha_i\mu(dx_i)) \mu^{i-1}(d\underline{x}^{i-1}) , \tag{2.46}$$

and the absolute value of (2.46)

$$\leq \int |\alpha_i||[\alpha^n_i \leq 0] - [\alpha_i \leq 0]|\mu^i(d\underline{x}^i)$$

$$\leq \int |\alpha_i|[|\alpha^n_i - \alpha_i| \geq |\alpha_i|]\mu^i(d\underline{x}^i) ,$$

which goes to zero w.p.1 by (2.25), $\int |\alpha_i|\mu^i(dx^i) < \infty$ and the DCT.

Now combining (2.39) - (2.46) we get for $i=j+2,\ldots,m$

$$\limsup_{n\to\infty} (T_i(n,j) - S_{i-1}(n,j) + \Delta_i(n,j)) \leq 0 \quad \text{w.p.1} .$$

Hence, by the definition of $M_2(n,j)$ (see (2.38)) for $j=1,\ldots,m-1$, $\limsup_{n\to\infty} M_2(n,j) \leq 0$ w.p.1 , and by (2.36) $\limsup_{n\to\infty} (L_n(n,j) + L_2(n,j) + L_3(n,j)) \leq 0$ w.p.1 , and by (2.28) $\limsup_{n\to\infty} K_n^3 \leq 0$ w.p.1 . This completes the proof of the Theorem 2.1 . □

The next corollary compares the asymptotic behaviour of the unconditional Bayes risk of EBSD procedure $\underline{d}^n$ with $R_L(G)$.

<u>Corollary 2.1</u>: Under the assumptions (A1) and (A2)

$$\limsup_{n\to\infty} E\, R(G,\underline{d}^n) \leq R_L(G) ,$$

where E denotes expectation with respect to random vectors $\underline{X}_1^{N_1},\ldots,\underline{X}_{n-1}^{N_{n-1}}$.

<u>Proof</u>. The proof follows from Theorem 2.1 and Fatou's lemma. □

<u>Corollary 2.2</u>: If $m=2$, then $\lim_{n\to\infty} R(G,d^n) = R_L(G)$ w.p.1. and $\lim_{n\to\infty} E\, R\,(G,\underline{d}^n) = R_L(G)$.

<u>Proof</u>: If $m=2$, then $R_L(G) = R_B(G)$ and, therefore, $R(G,\underline{d}^n) \geq R_L(G)$ for all n and G. Hence, $\lim_{n\to\infty} R(G,\underline{d}^n) = R_L(G)$ w.p.1 follows from (2.22) and then $\lim_{n\to\infty} E\, R(G,\underline{d}^n) = R_L(G)$ follows from the DCT. □

Corollary 2.3: If G is such that $\int\int B_j f_\theta(\underline{x}^i)\mu^j(d\underline{x}^j)G(d\theta) = 0$, $j = 1,\ldots,m-2$, then

$$\lim_{n\to\infty} R(G,\underline{d}^n) = R_L(G) \quad \text{w.p.1} .$$

Proof: From (A2) and (2.3) there exists a constant B such that $|\alpha_i| \leq Bf_i$, $i = 1,\ldots,m$ and $\int|\alpha_i|\mu(dx_i) \leq Bf_{i-1}$, $i = 2,\ldots,m$. Hence, by (2.38) and (2.41) for $j = 1,\ldots,m-2$,

$$|M_2(n,j)| \leq \sum_{i=j+2}^{m} \int B_j\,(2B + c)f_{i-1}(\underline{x}^{i-1})\mu^{i-1}(d\underline{x}^{i-1}) .$$

But B_j is $\underline{x}^j$ measurable so RHS of the above inequality is zero by the hypothesis of the corollary. Combining this result with $\lim M_1(n,j) = 0$ w.p.1 as $n \to \infty$, we obtain $\lim K_n^3 = 0$ w.p.1 as $n \to \infty$ so that from (2.27) the proof is complete. □

Corollary 2.4: Let N^n be the stopping time associated with the EBSD procedure $\underline{d}^n$ and let N^L denote the stopping time associated with the one-step look ahead procedure $\underline{d}^L$. Then w.p.1, N^n is stochastically larger than N^L as $n \to \infty$; specifically, for $i = 1,2,\ldots,m$,

$$\liminf_{n\to\infty} \int\int[N^n \geq i]f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) \geq \int\int [N^L \geq i]f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) \text{ w.p.1}$$

with convergence to the RHS if $\int\int B_j f_\theta(\underline{x}^j)\mu^j(d\underline{x}^j)G(d\theta) = 0$, $j = 1,2,\ldots,m-1$.

Proof: By the definitions, $[N^n \geq i] = [\rho_i^n < 0]\,\ldots\ldots\ldots\,[\rho_{i-1}^n < 0]$ and $[N^L \geq i] = [\rho_1 < 0]\,\ldots\ldots\ldots\,[\rho_{i-1} < 0]$. Thus, we have

$$[\rho_1 < 0]\ldots\ldots[\rho_{i-1} < 0] \leq \liminf_{n\to\infty} [N^n \geq i] \leq \limsup_{n\to\infty} [N^n \geq i]$$

(2.47)

$$\leq [\rho_1 \leq 0]\ldots\ldots[\rho_{i-1} \leq 0] ,$$

and the inequality follows from Fatou's Lemma. Observe that

$$0 \leq \mathrm{RHS}(2.47) - \mathrm{LHS}(2.47) \leq \sum_{j=1}^{i-1} B_j .$$

Hence, by the hypothesis concerning the B_j, we have $\lim\, [N^n \geq i] = [N^L \geq i]$ w.p.1 as $n \to \infty$, $i = 1,2,\ldots,m$. The proof is completed by taking expectation and applying the DCT. □

2.4 Multiple Decision Problem

In this section we treat a component decision problem that subsumes that treated in the last section. Assume that the action space $A = \{a_0, a_1, \ldots, a_\ell\}$ consists of a finite number of distinct actions and let $L(\theta,a) \geq 0$ on $\Theta \times A$ be the loss function associated with the problem.

The m-truncated one-step look ahead sequential decision procedure $\underline{d}^L = (\underline{\tau}^L(G), \underline{\delta}(G))$ with respect to G, $G\varepsilon\mathcal{G}$, for this multiple decision problem (Θ,A,L) can be defined as follows. Let $\underline{\delta}(G)$ be the decision rule consisting of a finite sequence of functions $(\delta_1,\ldots,\delta_m)$ where δ_k is a Bayes decision function with respect to G for the fixed sample size k decision problem based on the sample $(X_1,\ldots,X_k)$. δ_k can be derived as follows (see Van Ryzin and Susarla (1977), Gilliland and Hannan (1977) or Ferguson (1967) Chapter 6).

At the stage k of the sequential decision procedure, suppose we use a decision rule $t_k(\underline{x}^k) = (t_k(0|x^k),\ldots,t_k(\ell|\underline{x}^k))$ where $t_k(j|\underline{x}^k) = \Pr\{\text{choosing action } a_j | \underline{X}^k = \underline{x}^k\}$ and $\sum_{j=0}^{\ell} t_k(j|\underline{x}^k) = 1$. Then the Bayes risk of $t_k(\cdot)$ w.r. to G is

$$r(G,t_k) = \sum_{j=0}^{\ell} \int L(\theta, a_j)(\int t_k(j|\underline{x}^k)P_\theta^k(d\underline{x}^k))G(d\theta)$$

$$= \sum_{j=0}^{\ell} \int t_k(j|\underline{x}^k)\{\int(L(\theta,a_j) - L(\theta,a_0))f_\theta(\underline{x}^k)G(d\theta)\}\mu^k(d\underline{x}^k)$$

$$+ \int\int L(\theta,a_0)G(d\theta)f_\theta(\underline{x}^k)\mu^k(d\underline{x}^k) .$$

$r(G,t_k)$ is minimized by $t_k(j|\underline{x}^k) = \delta_k(j|\underline{x}^k)$, $j = 0,1,\ldots,m$ where

$$(2.48)\qquad \delta_k(j|\underline{x}^k) = \begin{cases} 1 & \text{if} \quad \underline{x}^k \in S_j \\ 0 & \text{if} \quad \underline{x}^k \notin S_j \end{cases}$$

with

$$S_j = \{\underline{x}^k | j = \min\{t: \Delta_k(a_t,\underline{x}^k) = \min_i \Delta_k(a_i,\underline{x}^k)\}\}$$

(2.49)

$$\Delta_k(a_i,\underline{x}^k) = \int(L(\theta,a_i) - L(\theta,a_0))f_\theta(\underline{x}^k)G(d\theta) .$$

When arguments are displayed we will delete the subscripts on δ_k and Δ_k. Therefore, the Bayes terminal decision rule $\underline{\delta}(G) = (\delta_1,\ldots,\delta_m)$ is defined componentwise by (2.48). Then the minimum posterior Bayes risk w.r. to G_k is given by

$$\begin{aligned} r(G_k) &= \int L(\theta,\delta_k(\underline{x}^k))G_k(d\theta) \\ &= \sum_{j=0}^{\ell} \int \delta(j|\underline{x}^k)L(\theta,a_j)G_k(d\theta) \\ &= \sum_{j=0}^{\ell} \delta(j|\underline{x}^k) \int (L(\theta,a_j) - L(\theta,a_0))G_k(d\theta) \\ &\qquad\qquad + \int L(\theta,a_0)G_k(d\theta) \\ &= \sum_{j=0}^{\ell} \delta(j|\underline{x}^k) \int (L(\theta,a_j) - L(\theta,a_0)) \frac{f_\theta(\underline{x}^k)G(d\theta)}{f_k(\underline{x}^k)} \\ &\qquad\qquad + \frac{1}{f_k(\underline{x}^k)} \int L(\theta,a_0)f_\theta(\underline{x}^k)G(d\theta) \end{aligned}$$

provided $f_k(\underline{x}^k) > 0$ and $r(G_k) = 0$ if $f_k(\underline{x}^k) = 0$. In the above derivations we use the fact that $\sum_{j=0}^{\ell} \delta(j|\underline{x}^k) = 1$. Now use (2.49) to get $r(G_k)$ in the following form:

$$(2.50) \qquad r(G_k) = \frac{1}{f_k(\underline{x}^k)} \sum_{j=0}^{\ell} \delta(j|\underline{x}^k)\Delta(a_j,\underline{x}^k) + \frac{1}{f_k(\underline{x}^k)} \int L(\theta,a_0)f_\theta(\underline{x}^k)G(d\theta) \ .$$

The stopping rule $\underline{\tau}^L(G)$ of $\underline{d}^L$ consists of a finite sequence of functions $(\tau_1^L,\ldots,\tau_m^L)$ and stops sampling for the first k (k=1,2,...,m) for which $\tau_k^L(\underline{x}^k) = 1$ where $\tau_m^L \equiv 1$ and for $k=1,\ldots,m-1$,

$$\tau_k^L(\underline{x}^k) = \begin{cases} 1 & \text{if} \quad E^* \, r(G_{k+1}) + c - r(G_k) \geq 0 \\ 0 & \text{if} \quad E^* \, r(G_{k+1}) + c - r(G_k) < 0 \end{cases}$$

and E^* denotes the conditional expectation over X_{k+1} given $\underline{X}^k = \underline{x}^k$. It is easy to show that

$$E^* \, r(G_{k+1}) = \frac{1}{f_k(\underline{x}^k)} \int \sum_{j=0}^{\ell} \delta(j|\underline{x}^{k+1})\Delta(a_j,\underline{x}^{k+1})\mu(dx_{k+1})$$

$$+ \frac{1}{f_k(\underline{x}^k)} \int L(\theta,a_0)f_\theta(\underline{x}^k)G(d\theta)$$

provided $f_k(\underline{x}^k) > 0$ and $E^* \, r(G_{k+1}) = 0$ if $f_k(\underline{x}^k) = 0$.

Then $E^* \, r(G_{k+1}) + c - r\,(G_k) \geq 0$ if and only if

$$\sum_{j=0}^{\ell} \int \delta(j|\underline{x}^{k+1})\Delta(a_j,\underline{x}^{k+1})\mu(dx_{k+1}) + cf_k(\underline{x}^k) \geq \sum_{j=0}^{\ell} \delta(j|\underline{x}^k)\Delta(a_j,\underline{x}^k)$$

when $f_k(\underline{x}^k) > 0$.

We define

$$(2.51) \qquad \beta_k(\underline{x}^k) = \sum_{j=0}^{\ell} \int \delta(j|\underline{x}^{k+1})\Delta(a_j,\underline{x}^{k+1})\mu(dx_{k+1}) + cf_k(\underline{x}^k) - \sum_{j=0}^{\ell} \delta(j|\underline{x}^k)\Delta(a_j,\underline{x}^k)$$

if $f_k(\underline{x}^k) > 0$ and $\beta_k(\underline{x}^k) = 0$ if $f_k(\underline{x}^k) = 0$, and the functions

$$A_k = [\beta_1 < 0].....[\beta_{k-1} < 0][\beta_k > 0] \quad \text{for} \quad k=1,\ldots,m-1$$

$$B_k = [\beta_1 < 0].....[\beta_{k-1} < 0][\beta_k = 0] \quad \text{for} \quad k=1,\ldots,m-1$$

and

$$A_m = [\beta_1 < 0].....[\beta_{m-1} < 0] \; .$$

Then $[N^L = k] = A_k + B_k$, $k = 1,\ldots,m-1$, $[N^L = m] = A_m$ and $\sum_{k=1}^{m-1} (A_k + B_k) + A_m = 1$. Therefore, $(\tau_1^L,\ldots,\tau_m^L)$ is defined by

$$(2.52) \qquad \tau_k^L(\underline{x}^k) = \begin{cases} 1 & \text{if} \quad \beta_k(\underline{x}^k) \geq 0 \\ 0 & \text{if} \quad \beta_k(\underline{x}^k) < 0 \; , \end{cases}$$

$k=1,\ldots,m-1$ and $\tau_m \equiv 1$. The Bayes risk of $\underline{d}^L = (\underline{\tau}^L(G),\underline{\delta}(G))$ with respect to G is

$$R_L(G) = \sum_{k=1}^{m} \iint [N^L = k]\{L(\theta,\delta(\underline{x}^k)) + ck\} f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

$$= \sum_{k=1}^{m} \iint [N^L = k]\{\sum_{j=0}^{\ell} \delta(j|\underline{x}^k)L(\theta,a_j) + ck\} f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) \; .$$

Now add and subtract $L(\theta,a_0)$ into the integrand of the above integral and use the facts $\sum_{j=0}^{\ell} \delta(j|\underline{x}^k) = 1$ and $\sum_{j=1}^{m} [N^L = j] = 1$ to obtain

$$R_L(G) = \sum_{k=1}^{m} \iint [N^L = k]\{\sum_{j=0}^{\ell} \delta(j|\underline{x}^k)(L(\theta,a_j) - L(\theta,a_0))$$

$$+ ck\} f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

$$+ \sum_{k=1}^{m} \iint [N^L = k]L(\theta,a_0) f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m),$$

that is,

$$R_L(G) = \sum_{k=1}^{m} \int\int A_k \sum_{j=0}^{\ell} \delta(j|\underline{x}^k)(L(\theta,a_j) - L(\theta,a_0))$$

$$+ ck\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

(2.53)
$$+ \sum_{k=1}^{m-1} \int\int B_k\{ \sum_{j=0}^{\ell} \delta(j|\underline{x}^k)(L(\theta,a_j) - L(\theta,a_0))$$

$$+ ck\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

$$+ \int\int L(\theta,a_0)f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

Let $\underline{d}^n = (\underline{\tau}^n,\underline{\delta}^n)$ denote an empirical Bayes one-step sequential decision procedure for the above multiple decision problem. We shall construct $\underline{d}^n$ based on the past data $\underline{X}^{N_1},\ldots,\underline{X}_{n-1}^{N_{n-1}}$ and the present data $\underline{X}$ at the stage n. Assume that there exists a sequence of distribution functions $\{G_n\}$ on Θ, where $G_n(\theta) = G_n(\theta,\underline{X}^{N_1},\ldots,\underline{X}_{n-1}^{N_{n-1}})$ depends only on the past data $\underline{X}^{N_1},\ldots,\underline{X}_{n-1}^{N_{n-1}}$, which converges weakly to G with probability one as $n \to \infty$. Let $\underline{\delta}^n$ be the decision rule consisting of a finite sequence of functions $(\delta_1^n,\ldots,\delta_m^n)$ where $\delta_k^n(\underline{x}^k) =(\delta^n(0|\underline{x}^k),\ldots,\delta^n(\ell|\underline{x}^k))$ subject to $\sum_{j=0}^{\ell} \delta^n(j|\underline{x}^k) = 1$ and, for $j = 0,1,\ldots,\ell$, if $\delta^n(j|\underline{x}^k) = \Pr\{\text{choosing } a_j|\underline{X}^k = \underline{x}^k\}$, then

(2.54)
$$\delta^n(j|\underline{x}^k) = \begin{cases} 1 & \text{if } \underline{x}^k \in \hat{S}_j \\ 0 & \text{if } \underline{x}^k \notin \hat{S}_j \end{cases}$$

(2.55)
$$\hat{S}_j = \{\underline{x}^k|j = \min\{t:\Delta^n(a_t,\underline{x}^k) = \min_i \Delta^n(a_i,\underline{x}^k)\} \text{ , and}$$

$$\Delta^n(a_i,\underline{x}^k) = \int (L(\theta,a_i) - L(\theta,a_0)f_\theta(\underline{x}^k)G_n(d\theta) .$$

The stopping rule $\underline{\tau}^n$ is defined by a finite sequence of functions $(\tau_1^n,\ldots,\tau_m^n)$ which stops sampling for the procedure $\underline{d}^n$ for the first $k(k = 1,\ldots,m)$ for which $\tau_k^n(\underline{x}^k) = 1$ where $\tau_m^n \equiv 1$ and for $k=1,\ldots,m-1$,

$$\tau_k^n(\underline{x}^k) = \begin{cases} 1 & \text{if} \quad \beta_k^n(\underline{x}^k) \geq 0 \\ 0 & \text{if} \quad \beta_k^n(\underline{x}^k) < 0 \,, \end{cases}$$

where

$$\beta_k^n(\underline{x}^k) = \sum_{j=0}^{\ell} \int_X \delta^n(j|\underline{x}^{k+1})\Delta^n(a_j,\underline{x}^{k+1})\mu(dx_{k+1}) + cf_k^n(\underline{x}^k) - \sum_{j=0}^{\ell} \delta^n(j|\underline{x}^k)\Delta^n(a_j,\underline{x}^k) \tag{2.56}$$

with $f_k^n(\underline{x}^k) = \int f_\theta(\underline{x}^k)G_n(d\theta)$.

Define

$$C_k^n = [\beta_1^n < 0].\ldots.[\beta_{k-1}^n < 0][\beta_k^n \geq 0] \quad \text{for} \quad k=1,\ldots,m-1$$

and

$$C_m^n = [\beta_1^n < 0].\ldots.[\beta_{m-1} < 0] \,.$$

Then $[N^n = k] = C_k^n$ for $k=1,\ldots,m$ where N^n is the stopping time of $\underline{d}^n = (\underline{\tau}^n,\underline{\delta}^n)$ for the multiple decision problem. The conditional Bayes risk of $\underline{d}^n = (\underline{\tau}^n,\underline{\delta}^n)$ with respect to G is then equal to

$$R(G,\underline{d}^n) = \sum_{k=1}^{m} \int\int C_k^n \{ \sum_{j=0}^{\ell} \delta^n(j|\underline{x}^k)(L(\theta,a_j) - L(\theta,a_0)) + ck\} f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) + \int\int L(\theta,a_0)f_\theta(\underline{x}^m)G(d\theta)\mu^m(dx^m) \,. \tag{2.57}$$

Now we prove a lemma which is used to prove the next theorem on the asymptotic behaviour of conditional Bayes risk of empirical Bayes sequential decision procedure $\underline{d}^n = (\underline{\tau}^n, \underline{\delta}^n)$ for the multiple decision problem.

Lemma 2.2: Assume $L(\theta,a)$ is a continuous function of θ for each $a \in A$; then under the assumptions (A1) and (A2) for $k \geq 1$ $\beta^n_k \to \beta_k$ w.p.1 as $n \to \infty$ where β_k and β^n_k are given by (2.51) and (2.56), respectively.

Proof: Observe that for $j \geq 1$, $k \geq 1, \Delta^n(a_j,\underline{x}^k) \to \Delta(a_j,\underline{x}^k)$ w.p.1 as $n \to \infty$ and $f^n_k(\underline{x}^k) \to f_k(\underline{x}^k)$ w.p.1 as $n \to \infty$ by the assumptions in the lemma. Then by the definitions of $\delta^n(j|\underline{x}^k)$ and $\delta(j|\underline{x}^k)$ we get

$$(2.58) \qquad \sum_{j=0}^{\ell} \delta^n(j|\underline{x}^k)\Delta^n(a_j,\underline{x}^k) \to \sum_{j=0}^{\ell} \delta(j|\underline{x}^k)\Delta(a_j,\underline{x}^k) \quad \text{w.p.1 as} \quad n \to \infty .$$

Now to complete the proof of $\beta^n_k(\underline{x}^k) \to \beta_k(\underline{x}^k)$ w.p.1 as $n \to \infty$, it remains to prove $\sum_{j=0}^{\ell} \int_X \delta^n(j|\underline{x}^{k+1})\Delta^n(a_j,\underline{x}^{k+1})\mu(dx_{k+1}) \to \sum_{j=0}^{\ell} \int_X \delta(j|\underline{x}^{k+1})$ $\Delta(a_j,\underline{x}^{k+1})\mu(dx_{k+1})$ w.p.1 as $n \to \infty$. To prove this statement we use the GDCT. From (2.58), $\sum_{j=0}^{\ell} \delta^n(j|\underline{x}^{k+1})\Delta^n(a_j,\underline{x}^{k+1}) \to \sum_{j=0}^{\ell} \delta(j|\underline{x}^{k+1})$ $\Delta(a_j,\underline{x}^{k+1})$ w.p.1 as $n \to \infty$, and observe that by the definition of $\Delta^n(a_j,\underline{x}^{k+1})$ and the boundedness of $L(\theta,a), |\Delta^n(a_j,\underline{x}^{k+1})| \leq 2J\, f^n_{k+1}(\underline{x}^{k+1})$ where J is a finite constant and $f^n_{k+1}(\underline{x}^{k+1}) = \int f_\theta(\underline{x}^{k+1})G_n(d\theta)$. But $\int f^n_{k+1}(\underline{x}^{k+1})\mu(dx_{k+1}) = f^n_k(\underline{x}^k)$ by Fubini's theorem. Then

$\lim_{n\to\infty} \int f^n_{k+1}(\underline{x}^{k+1})\mu(dx_{k+1}) = f_k(\underline{x}^k)$ w.p.1 by (A1) and (A2). Therefore,

$\lim_{n\to\infty} \int f_{k+1}^n(\underline{x}^{k+1})\mu(dx_{k+1}) = \int \lim_{n\to\infty} f_{k+1}^n(\underline{x}^{k+1})\mu(dx_{k+1})$ w.p.1 follows from the equality $f_k(\underline{x}^k) = \int f_{k+1}(\underline{x}^{k+1})\mu(dx_{k+1}) = \int \lim_{n\to\infty} f_{k+1}^n(\underline{x}^{k+1})\mu(dx_{k+1})$ w.p.1. It follows from the GDCT that

$$\lim_{n\to\infty} \int_X \Big(\sum_{j=0}^{\ell} \delta^n(j|\underline{x}^{k+1})\Delta^n(a_j,\underline{x}^{k+1})\Big)\mu(dx_{k+1}) = \int_X \Big(\sum_{j=0}^{\ell} \delta(j|\underline{x}^{k+1}) \Delta(a_j,\underline{x}^{k+1})\Big)\mu(dx_{k+1}) \quad \text{w.p.1} .$$

This completes the proof of Lemma 2.2. □

The next theorem gives us the asymptotic behaviour of the conditional Bayes risk of our empirical Bayes sequential decision procedure $\underline{d}^n = (\underline{\tau}^n, \underline{\delta}^n)$ for the multiple decision problem.

Theorem 2.2: If $L(\theta,a)$ is continuous function of θ for each $a \in A$, then under the assumptions (A1) and (A2),

$$\limsup_{n\to\infty} R(G,\underline{d}^n) \le R_L(G) \quad \text{w.p.1} ,$$

where $R_L(G)$ and $R(G,\underline{d}^n)$ are given by (2.53) and (2.57), respectively.

Proof: Write

$$R(G,\underline{d}^n) - R_L(G) = \sum_{i=1}^{m}\sum_{j=1}^{m}\int\int C_i^n A_j\{\sum_{k=0}^{\ell}\delta^n(k|\underline{x}^i)g_k(\theta) -$$

$$-\sum_{k=0}^{\ell}\delta(k|\underline{x}^j)g_k(\theta) + c(i-j)\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

(2.59)

$$+\sum_{i=1}^{m}\sum_{j=1}^{m-1}\int\int C_i^n B_j\{\sum_{k=0}^{\ell}\delta^n(k|\underline{x}^i)g_k(\theta)$$

$$-\sum_{k=0}^{\ell}\delta(k|\underline{x}^j)g_k(\theta) + c(i-j)\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

where $g_k(\theta) = L(\theta,a_k) - L(\theta,a_0)$, $k = 1,2,\ldots,\ell$.

The first double sum in (2.59) can be written as $J_n^1 + J_n^2 + J_n^3$, where

$$J_n^1 = \sum_{j=2}^{m}\sum_{i=1}^{j-1}\int\int C_i^n A_j\{\sum_{k=0}^{\ell}\delta^n(k|\underline{x}^i)g_k(\theta) -\sum_{k=0}^{\ell}\delta(k|\underline{x}^j)g_k(\theta) +$$

$$+ c(i-j)\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(dx^m),$$

$$J_n^2 = \sum_{j=1}^{m-1}\sum_{i=j+1}^{m}\int\int C_i^n A_j\{\sum_{k=0}^{\ell}\delta^n(k|\underline{x}^i)g_k(\theta) -\sum_{k=0}^{\ell}\delta(k|\underline{x}^j)g_k(\theta) +$$

$$+ c(i-j)\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m),$$

and

$$J_n^3 = \sum_{i=1}^{m}\int\int C_i^n A_j\{\sum_{k=0}^{\ell}\delta^n(k|\underline{x}^i)g_k(\theta) -\sum_{k=0}^{\ell}\delta(k|\underline{x}^i)g_k(\theta)\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m).$$

Then observe that

$$|J_n^1| \le \sum_{j=2}^{m}\sum_{i=1}^{j-1}\int\int C_i^n A_j\{\sum_{k=0}^{\ell}|g_k(\theta)| + c|i-j|\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m),$$

and

$$|J_n^2| \le \sum_{j=1}^{m-1} \sum_{i=j+1}^{m} \int \int C_i^n A_j \{ \sum_{k=0}^{\ell} |g_k(\theta)| + c|i-j| \} f_\theta(\underline{x}^m) G(d\theta) \mu^m(d\underline{x}^m) .$$

Recall that for $i \ge 1$, $C_i^n = [\beta_1^n < 0] \ldots\ldots [\beta_{i-1}^n < 0][\beta_i^n \ge 0]$ and for $j \ge 1$, $A_j = [\beta_1 < 0] \ldots\ldots [\beta_{j-1} < 0][\beta_j > 0]$, $B_j = [\beta_1 < 0] \ldots\ldots [\beta_{j-1} < 0]$ $[\beta_j = 0]$. Then by Lemma 2.2 for $i < j$, $C_i^n A_j^n \le [\beta_i^n \ge 0][\beta_i < 0] \to 0$ w.p.1 as $n \to \infty$ and for $i > j$, $C_i^n A_j^n \le [\beta_j^n < 0][\beta_j > 0] \to 0$ w.p.1 as $n \to \infty$. Now use the DCT and the assumptions in the theorem to conclude that $J_n^1 \to 0$ and $J_n^2 \to 0$ w.p.1 as $n \to \infty$. Now to finish the proof that the first double sum in (2.59) goes to zero w.p.1 as $n \to \infty$, it is enough to show that J_n^3 goes to zero w.p.1 as $n \to \infty$, where

$$J_n^3 = \sum_{i=1}^{m} \int \int C_i^n A_i \{ \sum_{k=0}^{\ell} (\delta^n(k|\underline{x}^i) - \delta(k|\underline{x}^i)) g_k(\theta) \} f_\theta(\underline{x}^m) G(d\theta) \mu^m(d\underline{x}^m) .$$

From the definition (2.49) of $\Delta(a_j, \underline{x}^k)$ we can write J_n^3 in the following form:

$$J_n^3 = \sum_{i=1}^{m} \int_{X^i} C_i^n A_i \{ \sum_{k=0}^{\ell} (\delta^n(k|\underline{x}^i) - \delta(k|\underline{x}^i)) \Delta(a_k, \underline{x}^i) \} \mu^i(d\underline{x}^i) .$$

From (2.58) and $\Delta^n(a_k, \underline{x}^i) \to \Delta(a_k, \underline{x}^i)$ w.p.1 as $n \to \infty$, $k \ge 1$, $i \ge 1$, we get for $i \ge 1$,

$$\text{(2.60)} \qquad \sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i) \Delta(a_k, \underline{x}^i) \to \sum_{k=0}^{\ell} \delta(k|\underline{x}^i) \Delta(a_k, \underline{x}^i) \quad \text{w.p.1 as } n \to \infty .$$

Now use $\int_{X^i} |\Delta(a_k, \underline{x}^i)| \mu^i(d\underline{x}^i) < \infty$, $k = 0,1,\ldots,\ell$, $i \ge 1$, and the DCT to conclude that $J_n^3 \to 0$ w.p.1 as $n \to \infty$.

The second double sum in (2.59) can be written as $K_n^1 + K_n^2 + K_n^3$ where

$$K_n^1 = \sum_{j=2}^{m-1} \sum_{i=1}^{j-1} \int\int C_i^n B_j\{\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i)g_k(\theta) - \sum_{k=0}^{\ell} \delta(k|\underline{x}^j)g_k(\theta) + c(i-j)\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) ,$$

$$K_n^2 = \sum_{i=1}^{m-1} \int\int C_i^n B_i\{\sum_{k=0}^{\ell} (\delta^n(k|\underline{x}^i) - \delta(k|\underline{x}^i))g_k(\theta)\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) ,$$

and

$$K_n^3 = \sum_{j=1}^{m-1} \sum_{i=j+1}^{m} \int\int C_i^n B_j\{\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i)g_k(\theta) - \sum_{k=0}^{\ell} \delta(k|\underline{x}^j)g_k(\theta) + c(i-j)\}f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) .$$

K_n^1 and K_n^2 have similar forms as J_n^1 and J_n^3 respectively, and one can show that $K_n^1 \to 0$ and $K_n^2 \to 0$ w.p.1 as $n \to \infty$ using similar arguments as for J_n^1 and J_n^3.

We now summarize the results concerning the difference (2.59) obtained so far.

$$R(G,\underline{d}^n) - R_L(G) = \sum_{j=1}^{3} J_n^j + \sum_{j=1}^{3} K_n^j \tag{2.61}$$

where J_n^1, J_n^2, J_n^3, K_n^1, K_n^2 have been shown to converge to zero w.p.1 as $n \to \infty$.

Now we will show that $\limsup_{n\to\infty} K_n^3 \le 0$ w.p.1. Write

$$K_n^3 = \sum_{j=1}^{m-1} (L_1(n,j) + L_2(n,j) + L_3(n,j)) \tag{2.62}$$

where

$$L_1(n,j) = \sum_{i=j+1}^{m} \int\int C_i^n B_j(\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i)g_k(\theta))f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m) \tag{2.63}$$

$$(2.64)\qquad L_2(n,j) = -\sum_{i=j+1}^{m} \int\int C_i^n\, B_j\Big(\sum_{k=0}^{\ell} \delta(k|\underline{x}^j) g_k(\theta)\Big) f_\theta(\underline{x}^m) G(d\theta)\mu^m(d\underline{x}^m)$$

and

$$(2.65)\qquad L_3(n,j) = \sum_{i=j+1}^{m} \int\int C_i^n\, B_j\; c(i-j) f_\theta(\underline{x}^m) G(d\theta)\mu^m(d\underline{x}^m)\ .$$

Define $\hat{C}_k^n = \sum_{i=k}^{m} C_i^n$ for $k = j+1,\dots,m$ and $j = 1,\dots,m-1$; then $C_i^n = \hat{C}_i^n - \hat{C}_{i+1}^n$ for $i = 1,\dots,m-1$ and $C_m^n = \hat{C}_m^n$. Now $L_1(n,j)$, $L_2(n,j)$ and $L_3(n,j)$ can be simplified into the following form:

$$L_1(n,j) = \sum_{i=j+1}^{m} \int_{X^i} C_i^n\, B_j\Big(\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i)\Delta(a_k,\underline{x}^i)\Big)\mu^i(d\underline{x}^i)$$

$$(2.66)\qquad \begin{aligned} &= \sum_{i=j+1}^{m} \int_{X^i} \hat{C}_i^n\, B_j\Big(\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i)\Delta(a_k,\underline{x}^i)\Big)\mu^i(d\underline{x}^i) \\ &\qquad - \sum_{i=j+1}^{m-1} \int_{X^i} \hat{C}_{i+1}^n\, B_j\Big(\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i)\Delta(a_k,\underline{x}^i)\Big)\mu^i(d\underline{x}^i) \\ &= \sum_{i=j+1}^{m} T_i(n,j) - \sum_{i=j+1}^{m-1} S_i(n,j) \end{aligned}$$

where

$$T_i(n,j) = \int_{X^i} C_i^n\, B_j\Big(\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i)\Delta(a_k,\underline{x}^i)\Big)\mu^i(d\underline{x}^i)$$

and

$$S_i(n,j) = \int_{X^i} \hat{C}_{i+1}^n\, B_j\Big(\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i)\Delta(a_k,\underline{x}^i)\Big)\mu^i(d\underline{x}^i)\ ;$$

$$L_2(n,j) = -\sum_{i=j+1}^{m} \int\int C_i^n B_j(\sum_{k=0}^{\ell} \delta(k|\underline{x}^k)g_k(\theta))f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

$$= -\int\int B_j(C_{j+1}^n + C_{j+2}^n \cdots + C_m^n)(\sum_{k=0}^{\ell} \delta(k|\underline{x}^i)g_k(\theta))$$

$$(2.67) \qquad f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

$$= -\int\int B_j \hat{C}_{j+1}^n(\sum_{k=0}^{\ell} \delta(k|\underline{x}^i)g_k(\theta))f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

$$= -\int_{X^i} B_j \hat{C}_{j+1}^n(\sum_{k=0}^{\ell} \delta(k|\underline{x}^j)\Delta(a_k,\underline{x}^j))\mu^j(d\underline{x}^j) \ ;$$

and

$$L_3(n,j) = c\sum_{i=j+1}^{m} \int\int \hat{C}_i^n B_j f_\theta(\underline{x}^m)G(d\theta)\mu^m(d\underline{x}^m)$$

$$(2.68) \qquad = \sum_{i=j+1}^{m} c\int_{X^{i-1}} \hat{C}_i^n B_j f_{i-1}(\underline{x}^{i-1})\mu^{i-1}(d\underline{x}^{i-1})$$

$$= \sum_{i=j+1}^{m} U_i(n,j)$$

where

$$U_i(n,j) = c\int_{X^{i-1}} \hat{C}_i^n B_j f_{i-1}(\underline{x}^{i-1})\mu^{i-1}(d\underline{x}^{i-1}) \ .$$

The sum $L_1(n,j) + L_2(n,j) + L_3(n,j)$ now takes the following form:

$$\sum_{i=j+1}^{m} T_i(n,j) - \sum_{i=j+1}^{m-1} S_i(n,j) - \int_{X^j} \hat{C}_{j+1}^n B_j(\sum_{k=0}^{\ell} \delta(k|\underline{x}^j)$$

$$\Delta(a_k,\underline{x}^j))\mu^j(d\underline{x}^j) + \sum_{i=j+1}^{m} U_i(n,j) \ ,$$

that is,

$$T_{j+1}(n,j) - \int_{X^j} \hat{C}^n_{j+1}\; B_j\Big(\sum_{k=0}^{\ell} \delta(k|\underline{x}^j)\Delta(a_k,\underline{x}^j)\Big)\mu^j(d\underline{x}^j)$$

(2.69)

$$+\; U_{j+1}(n_1,j) + \sum_{i=j+2}^{m} (T_i(n,j) - S_{i-1}(n,j) + U_i(n,j)).$$

The sum of the first three terms in the expression (2.69) is equal to

$$\int_{X^j} \hat{C}^n_{j+1}\; B_j\Big\{\int_X \sum_{k=0}^{\ell} \delta^n(k|\underline{x}^{j+1})\Delta(a_k,\underline{x}^{j+1})\mu(dx_{j+1}) \tag{2.70}$$

$$-\sum_{k=0}^{\ell} \delta(k|\underline{x}^j)\Delta(a_k,\underline{x}^j) + cf_j(\underline{x}^j)\Big\}\mu^j(d\underline{x}^j)\ .$$

By the definition of β_j (see (2.51)),

$$-\sum_{k=0}^{\ell} \delta(k|\underline{x}^j)\Delta(a_k,\underline{x}^j) + cf_j(\underline{x}^j) = -\int_X \sum_{k=0}^{\ell} \delta(k|\underline{x}^{j+1})\Delta(a_k,\underline{x}^{j+1})$$

$$\mu(dx_{j+1}) \quad \text{on} \quad [\beta_j = 0]\ .$$

But $0 \le B_j \le [\beta_j = 0]$ so the absolute value of the expression (2.70) is less than or equal to

$$\int_{X^j} [\beta_j = 0]\Big\{\Big|\int_X \Big(\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^{j+1})\Delta(a_k,\underline{x}^{j+1}) \tag{2.71}$$

$$-\sum_{k=0}^{\ell} \delta(k|\underline{x}^{j+1})\Delta(a_k,\underline{x}^{j+1})\Big)\mu(dx_{j+1})\Big|\Big\}\mu^j(d\underline{x}^j)\ .$$

Expression (2.71) goes to zero w.p.1 as $n \to \infty$ by (2.60) , $\sum_{k=0}^{\ell} \int_{X^{j+1}} |\Delta(a_k,\underline{x}^{j+1})|\mu^{j+1}(d\underline{x}^{j+1}) < \infty$ and the DCT.

Finally let us consider the sum $\sum_{i=j+2} \{T_i(n,j) - S_{i-1}(n,j) + U_i(n,j)\}$.

Observe that for $i=j+2,\ldots,m$, $T_i(n,j) - S_{i-1}(n,j) + U_i(n,j)$ is equal to

$$\int_{X^i} \hat{C}_i^n B_j(\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i)\Delta(a_k,\underline{x}^i))\mu^i(dx^i) -$$

$$\int_{X^{i-1}} \hat{C}_i^n B_j(\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^{i-1})\Delta(a_k,\underline{x}^{i-1}))\mu^{i-1}(d\underline{x}^{i-1})$$

$$+ c \int_{X^{i-1}} \hat{C}_i^n B_j f_{i-1}(\underline{x}^{i-1})\mu^{i-1}(d\underline{x}^{i-1}) .$$

Rearranging the terms in the above integrals we get

$$\begin{aligned} &\int_{X^{i-1}} \hat{C}_i^n B_j \{\int (\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i)\Delta(a_k,\underline{x}^i))\mu(dx_i) \\ &- \sum_{k=0}^{\ell} \delta^n(k|\underline{x}^{i-1})\Delta(a_k,\underline{x}^{i-1}) + cf_{i-1}(\underline{x}^{i-1})\}\mu^{i-1}(d\underline{x}^{i-1}) . \end{aligned} \tag{2.72}$$

Now add and subtract the term $\sum_{k=0}^{\ell} \delta(k|\underline{x}^{i-1})\Delta(a_k,\underline{x}^{i-1})$ into the integrand of (2.72) to get

$$\begin{aligned} &\int_{X^{i-1}} \hat{C}_i^n B_j\{\int_X(\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i)\Delta(a_k,\underline{x}^i))\mu(dx_i) \\ &- \sum_{k=0}^{\ell} \delta(k|\underline{x}^{i-1})\Delta(a_k,\underline{x}^{i-1}) + cf_{i-1}(\underline{x}^{i-1})\mu^{i-1}(d\underline{x}^{i-1}) \\ &+ \int_{X^{i-1}} C_i^n B_j\{\sum_{k=0}^{\ell} \delta(k|\underline{x}^{i-1})\Delta(a_k,\underline{x}^{i-1}) - \sum_{k=0}^{\ell} \delta^n(k|\underline{x}^{i-1}) \\ &\Delta(a_k,\underline{x}^{i-1}\}\mu^{n-1}(d\underline{x}^{i-1}) . \end{aligned} \tag{2.73}$$

The second integral term in (2.73) goes to zero w.p.1 as $n \to \infty$ from (2.60), $\sum_{k=0}^{\ell} \int_{X^{i-1}}|\Delta(a_k,\underline{x}^{i-1})|\mu^{i-1}(d\underline{x}^{i-1}) < \infty$, and the DCT.

The first integral in (2.73) can be rewritten as

$$\int_{X^{i-1}} \hat{C}_i^n B_j[\beta_{i-1} \le 0]\{\int_X (\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i)\Delta(a_k,\underline{x}^i))\mu(dx_i)$$

$$- \sum_{k=0}^{\ell} \delta(k|\underline{x}^{i-1})\Delta(a_k,\underline{x}^{i-1}) + cf_{i-1}(\underline{x}^{i-1})\}\mu^{i-1}(d\underline{x}^{i-1})$$

(2.74)
$$+ \int_{X^{i-1}} \hat{C}_i^n B_j[\beta_{i-1} > 0]\{\int (\sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i)$$

$$\Delta(a_k,\underline{x}^i)\mu(dx_i)$$

$$- \sum_{k=0}^{\ell} \delta(k|\underline{x}^{i-1})\Delta(a_k,\underline{x}^{i-1}) + cf_{i-1}(\underline{x}^{i-1}))$$

$$\mu^{i-1}(d\underline{x}^{i-1}) .$$

But $0 \le \hat{C}_i^n [\beta_{i-1} > 0] \le [\beta_{i-1}^n < 0][\beta_{i-1} > 0] \to 0$ w.p.1 as $n \to \infty$ by Lemma 2.2 . Therefore, the second term in (2.74) goes to zero w.p.1 as $n \to \infty$ by the DCT and $\int \Delta(a_k,\underline{x}^i)\mu(d\underline{x}^i) < \infty$, $i \ge 1$, $k \ge 1$. The first term in (2.74) is

(2.75)
$$\le \int_{X^{i-1}} \hat{C}_i^n B_j [\beta_{i-1} \le 0]\{\int \sum_{k=0}^{\ell} \delta^n(k|\underline{x}^i)\Delta(a_k,\underline{x}^i))\mu(dx_i)$$

$$- \int_X (\sum_{k=0}^{\ell} \delta(k|\underline{x}^i)\Delta(a_k,\underline{x}^i))\mu(dx_i)\mu^{i-1}(d\underline{x}^{i-1}) .$$

The inequality (2.75) follows from the fact that (see 2.51) on $[\beta_{i-1} \le 0]$ we have

$$-\sum_{k=0}^{\ell} \delta(k|\underline{x}^{i-1})\Delta(a_k,\underline{x}^{i-1}) + cf_{i-1}(\underline{x}^{i-1}) \le -\int_X (\sum_{k=0}^{\ell} \delta(k|\underline{x}^i)\Delta(a_k,\underline{x}^i))\mu(dx_i) .$$

Now it is easy to see that the

R.H.S. of the inequality (2.75) goes to zero w.p.1 as $n \to \infty$ from (2.60), $\sum_{k=0}^{\ell} \int_{X^i} |\Delta(a_k, \underline{x}^i)| \mu^i(d\underline{x}^i) < \infty$, and the DCT. Now combine (2.74) and (2.75) to get for $i=j+2,\ldots,m$

$$\limsup_{n\to\infty} \{T_i(n,j) - S_{i-1}(n,j) + U_i(n,j)\} \leq 0 \quad \text{w.p.1 as } n \to \infty$$

and then with (2.69) and (2.70), we have for $j=1,\ldots,m-1$,

$$\limsup_{n\to\infty} (L_1(n,j) + L_2(n,j) + L_3(n,j)) \leq 0 \quad \text{w.p.1}.$$

Therefore, $\limsup_{n\to\infty} K_n^3 \leq 0$ w.p.1 follows from (2.62). This completes the proof of Theorem 2.2. □

Remark 2.2: Corollaries analogous to Corollaries 2.1 - 2.4 of the previous section can be stated and proved in the more general multiple decision problem context as well.

2.5 Final Remarks

Using arguments very similar to those of Sections 2.3 and 2.4, we can show that the result $\limsup R(G,\underline{d}^n) \leq R_L(G)$ w.p.1 as $n \to \infty$ also holds for natural empirical Bayes sequential decision procedures for the squared error loss estimation component under (A1) and (A2).

A curious feature of the EBSD's in approximating one-step look ahead risk is the inequality in the asymptotic result. (The asymptotic optimality that is typically proved is the convergence to $R_B(G)$.) There are examples of two-action linear loss components with Poisson distributions and priors G for which $\limsup R(G,\underline{d}^n) < R_L(G)$ on a set of positive probability.

BIBLIOGRAPHY

BIBLIOGRAPHY

Berger, James O. (1980). Statistical Decision Theory. Foundations, Concepts and Methods. Springer-Verlag, New York.

DeGroot, Morris H. (1970). Optimal Statistical Decisions. McGraw-Hill Book Company, New York.

Ferguson, Thomas S. (1967). Mathematical Statistics. A Decision Theoretical Approach. Academic Press, New York.

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

Gilliland, Dennis, Hannan, James and Huang, J.S. (1976). Asymptotic solutions to the two state component compound decision problem, Bayes versus diffuse priors on proportions. Ann. Statist. 4, 1101-1112.

Gilliland, Dennis and Hannan, James (1977). Improved rates for empirical Bayes monotone decision problems. Ann. Statist. 5, 516-521.

Gilliland, Dennis and Hannan, James (1985). The finite state compound decision problem, equivariance and restricted risk components. To appear in Proceedings of the Robbins' Symposium 1985.

Hannan, James (1957). Approximations to Bayes risk in repeated play. Contributions to the theory of games, Vol. 3, pp. 97-139, Annals of Mathematics Studies, no. 39. Princeton University Press. Princeton, N.J.

Johns, M.V. and Van Ryzin, J. (1971). Convergence rates for empirical Bayes two-action problems I. Discrete case. Ann. Math. Statist., 42, 1521-1539.

Johns, M.V. and Van Ryzin, J. (1972). Convergence rates for empirical Bayes two-action problems II. Continuous case. Ann. Math. Statist. 43, 934-947.

Laippala, P. (1979). The empirical Bayes approach with floating sample size in Binomial experimentation. Scand. J. Statist. 6, 113-118. Correction note SJS 7, 105.

Laippala, P. (1980). Ph.D. Dissertation. University of Turku, Finland.

Laippala, P. (1983). Simulation results concerning the empirical Bayes two-action rules with floating optimal sample size. Scand. J. Statist. 10, 87-96.

Laippala, P. (1985). The empirical Bayes rules with floating optimal sample size for exponential conditional distributions. Ann. Inst. Statist. Math. 37, 315-327.

O'Bryan, T. (1972). Empirical Bayes results in the case of non-identical components. Ph.D. Thesis, Michigan State University.

O'Bryan, T. (1976). Some empirical Bayes results in the case of component problems with varying sample sizes for discrete exponential families. Ann. Statist. 4, 1290-1293.

O'Bryan, T. and Susarla, V. (1975). An empirical Bayes estimation problem with non-identical components involving normal distributions. Communications in Statistics 4, 1033-1042.

O'Bryan, T. and Susarla, V. (1976). Rates in the empirical Bayes estimation problem with non-identical components. Case of normal distributions. Ann. Inst. Statist. Math., 28, 389-397.

O'Bryan, T. and Susarla, V. (1977). Empirical Bayes estimation with non-identical components. Continuous case. Austral. J. Statist., 19, 115-125.

Robbins, H. (1949). Unpublished lecture notes. UNC, Department of Statistics.

Robbins, H. (1951). Asymptotically subminimal 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. I, 157-163.

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

Samuel, Ester (1963). An empirical Bayes approach to the testing of certain parametric hypotheses. Ann. Math. Statist., 34, 1370-1385.

Susarla, V. and O'Bryan, T. (1975). An empirical Bayes two-action problem with non-identical components for a translated exponential distribution. Communications in Statistics 4, 767-775.

Van Houwelingen, J.C. (1976). Monotone empirical Bayes tests for the continuous one parameter exponential family. Ann. Statist. 4, 981-989.

Van Ryzin, J. and Susarla, V. (1977). On the empirical Bayes approach to multiple decision problems. Ann. Statist. 5, 172-181.

Yu, Benito Ong (1970). Rates of convergence in empirical Bayes two-action and estimation problems and in extended sequence-compound estimation problems. Ph.D. Thesis, Michigan State University.