ON THE RISK PERFORMANCE OF BAYES EMPIRICAL BAYES PROCEDURES IN THE FINITE STATE COMPONENT CASE

By

How Jan Tsao

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

1980

ABSTRACT

ON THE RISK PERFORMANCE OF BAYES EMPIRICAL
BAYES PROCEDURES IN THE FINITE STATE
COMPONENT CASE

By

How Jan Tsao

Since Robbins' introduction of the empirical Bayes approach to a sequence of decision problems, a large literature has evolved treating a variety of component problems. Most of the papers advance empirical Bayes procedures which are asymptotically optimal, and some establish rates of convergence.

In empirical Bayes decision making, the Bayes empirical Bayes approach is discussed by Gilliland and Boyer (1979). In the finite state component case, the Bayes empirical Bayes procedures are shown to have optimal properties in a fairly general setting and believed to have small sample advantage over the classical rules. The flexibility of making desirable adjustments for these decision procedures by choice of prior enables one to set a proper strategy when dealing with actual problems.

In this thesis, a complete class theorem is proved to show that, at each sample stage, the class of Bayes empirical Bayes rules is complete, and, under some regularity conditions, that it is minimal complete. In the two state component case the posterior

mean which generates the Bayes empirical Bayes rules is shown to be asymptotically normal under certain assumptions.

The use of Bayes empirical Bayes procedures creates some interesting theoretical and computational problems as the Bayes procedures are fairly complicated in structure. The thesis also develops methods of computing Bayes empirical Bayes rules and determining their small sample risk behavior. In some cases risk functions are evaluated by numerical methods, and, in other cases, Monte Carlo simulation is used to estimate risk.

To my parents and Grace

ACKNOWLEDGMENTS

I would like to take this opportunity to thank my advisor, Professor Dennis C. Gilliland for his invaluable guidance, assistance, encouragement and caretaking during the entire course of this study. The financial support provided by the Department of Statistics and Probability and National Science Foundation made my graduate studies possible. I wish to thank them and Clara who accurately typed the whole manuscript with great patience and skill.

My deep gratitude is extended to my parents and my wife, Grace, for their understanding, encouragement and support.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

FINITE STATE BAYES EMPIRICAL BAYES PROCEDURES

Section 1.1. The component and empirical Bayes decision problems.

Consider the following component statistical decision problem with which we shall be concerned. This comprises

(i) A sample space $(X,\underline{X})$ and a parameter space $(\Omega,\underline{\Omega})$ where $\underline{X},\underline{\Omega}$ are σ-algebras on X,Ω respectively. $\{P_\theta: \theta \in \Omega\}$ is a family of probability measures on $(X,\underline{X})$ dominated by some σ-finite measure μ. f_θ is a density for P_θ with respect to μ, $\theta \in \Omega$. X denotes an X-valued random variable distributed P_θ, conditional on θ.

(ii) An action space $(A,\underline{A})$ where $\underline{A}$ is a σ-algebra on A containing the singleton sets.

(iii) A loss function $L: \Omega \times A \to [0,\infty)$ representing the loss of taking action a in A with $\theta \in \Omega$. $L(\theta,\cdot)$ is measurable for each $\theta \in \Omega$.

(iv) The (behavioral) decision rules $t(\cdot,\cdot)$, each a function of the pair (x,B) where $x \in X$ and $B \in \underline{A}$, having the measurability properties below:

  (a) for each x, $t(x,\cdot)$ is a probability measure on $\underline{A}$.

  (b) for each B, $t(\cdot,B)$ is $\underline{X}$-measurable.

A nonrandomized decision rule is one where for each x, $t(x,\cdot)$ is degenerate. The set of (behavioral) decision rules is denoted by A.

For any t, the expected loss when θ is the true parameter is

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

Let G denote the class of all probability measures (priors) on $\underline{\Omega}$ with respect to which the t-sections of $R(\theta,t)$ are measurable. The Bayes risk of t versus G is

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

t_G is called a Bayes rule with respect to G if its Bayes risk attains the infimum Bayes risk

$$R(G) = \inf_{t\in A} R(G,t).$$

We will assume that $R(G)$ is attained for each $G \in G$.

Throughout our discussions we will consider $\Omega = \{0,1,\ldots,m\}$, $\underline{\Omega} = 2^\Omega$, and assume that $P_G = \sum_{\theta=0}^{m} g_\theta P_\theta$ is identified by $G = (g_0,\ldots,g_m)$. G is the m-dimensional simplex in E_{m+1}, the m+1-dimensional Euclidean space. We will call $R(\cdot,t)$ the risk function of t and $R(\cdot)$ the Bayes envelope defined on G.

Consider the <u>empirical Bayes</u> decision problem. In it the component decision problem just described occurs repeatedly and independently. Thus, let (θ_1,X_1), $(\theta_2,X_2),\ldots,(\theta_n,X_n)$, $(\theta_{n+1},X_{n+1}),\ldots$ be iid with θ_i having distribution G and, conditional on

θ_i, X_i having distribution P_{θ_i}. The marginal distribution of X_i is the mixture P_G. Based on the initial observations $\underline{x}_n = (x_1,\ldots,x_n)$, a component decision rule $T_n(\underline{x}_n)$ is selected and evaluated at x_{n+1} to reach a decision about θ_{n+1}, $n \geq 1$. Thus, an empirical Bayes decision rule for reaching a decision about θ_{n+1} is

$$T_n(\underline{X}_n)(X_{n+1},\cdot),\ n \geq 1.$$

The goal is to use the information about G from the initial observations to construct a rule T_n whose risk behavior is close to that of the Bayes rule $t_G(X_{n+1},\cdot)$. In general, more information about G will be available with increasing number of observations. We will consider an empirical Bayes procedure as a sequence $T = (T_1,T_2,\ldots)$ of empirical Bayes decision rules where for each n, T_n is a function on $X^{n+1} \times \underline{A}$ such that every $\underline{x}_n = (x_1,\ldots,x_n)$ -section is an element of A, the class of component decision rules, and such that for each $\theta \in \Omega$, $R(\theta,T_n(\underline{x}_n))$ is a measurable function in $\underline{x}_n$.

For each n, we let $\mathcal{T}_n$ denote the collection of all possible T_n defined as above. The use of T_n against prior G incurs the unconditional component Bayes risk

$$R_n(G,T_n) = \int R(G,T_n(\underline{x}_n))P_G^n(d\underline{x}_n),\quad n \geq 1$$

where here and throughout a symbol for a measure with a superscript indicates a product measure. Since $T_n(\underline{x}_n) \in A$ for each $\underline{x}_n \in X^n$, we see that for all n, $R_n(G,T_n) \geq R(G)$, the minimum component

Bayes risk. Observe that

$$R_n(G,T_n) = ER(G,T_n(\underline{X}_n))$$

$$= \sum_{\theta=0}^{m} g_\theta \int R(\theta,T_n(\underline{x}_n))P_G^n(d\underline{x}_n)$$

$$= \sum_{\theta=0}^{m} g_\theta \int R(\theta,T_n(\underline{x}_n))\{\prod_{i=1}^{n} [\sum_{j=0}^{m} f_j(x_i)g_j]\}\, \mu^n(d\underline{x}_n)$$

$$= \sum_{\theta=0}^{m} \sum_{\ell_0+\ldots+\ell_m = n} g_\theta(g_0^{\ell_0}\ldots g_m^{\ell_m})H_n(\theta,\ell_0,\ldots,\ell_m) \qquad (1.1)$$

where

$$H_n(\theta,\ell_0,\ldots,\ell_m) = \sum_{\substack{B_0,\ldots,B_m \\ |B_i| = \ell_i \\ i = 0,\ldots,m}} \int R(\theta,T_n(\underline{x}_n))\{\prod_{j=0}^{m} \prod_{i\in B_j} f_j(x_i)\}\, \mu^n(d\underline{x}_n).$$

The summation above is over partitions $\{B_0,B_1,\ldots,B_m\}$ of $\{1,2,\ldots,n\}$, and the second summation in (1.1) is over all partitions $\ell_0,\ell_1,\ldots,\ell_m$ of the integer n, i.e., integers $\ell_i \geq 0$ with $\sum_{i=0}^{m} \ell_i = n$. From (1.1) we see that the risk function $R_n(\cdot,T_n)$ is determined by the collection of coefficients

$$\{H_n(\theta,\ell_0,\ldots,\ell_m) \mid \theta = 0,\ldots,m;\ \sum_{i=0}^{m} \ell_i = n,\ \ell_i \geq 0,\ i = 0,..,m\} \qquad (1.2)$$

which in turn, can be identified by an element of of the space E_N, where by Feller (1975, (II.5.2)), $N = (m+1)\binom{m+n}{n}$. This remark will prove useful in Section 1.4.

Definition 1.1. If $\lim_n R_n(G,T_n) = R(G)$ we say that T is asymptotically optimal relative to G(a.o.[G]). If T is a.o.[G] for all $G \in \mathcal{G}$, we say that T is asymptotically optimal (a.o.).

Definition 1.2. For $T_n, T_n^* \in \mathcal{T}_n, T_n$ is as good as T_n^* if $R_n(G,T_n) \leq R_n(G,T_n^*)$ for all $G \in \mathcal{G}$. T_n is better than T_n^* if $R_n(G,T_n) \leqq R_n(G,T_n^*)$ for all $G \in \mathcal{G}$ and $R_n(G,T_n) < R_n(G,T_n^*)$ for at least one $G \in \mathcal{G}$. T_n is equivalent to T_n^* if $R_n(G,T_n) = R_n(G,T_n^*)$ for all $G \in \mathcal{G}$.

Definition 1.3. T_n is said to be admissible if there does not exist an empirical Bayes decision rule in $\mathcal{T}_n$ that is better than T_n. T is called an admissible empirical Bayes procedure if T_n is admissible, $n \geqq 1$.

Listed below are some desirable properties of an empirical Bayes decision procedure $T = (T_1, T_2, \ldots)$.

(i) T is a.o.

(ii) $R_n(G,T_n)$ converges to $R(G)$ rapidly for all $G \in \mathcal{G}$.

(iii) T is admissible.

(iv) T_n has good risk behavior for small to moderate values of n, that is, T_n is suitable for use even when large numbers of observations are not available.

(v) An algorithm for computing the decision rules is available and can be executed economically.

(vi) T can be adjusted systematically to improve its performance on many specified subsets of G.

We will judge the performance of an empirical Bayes procedure on the basis of properties (i) - (vi) mentioned above.

Section 1.2. Bayes empirical Bayes

Let $\underline{G}$ be the Borel σ-algebra of subsets of G. The Bayes approach to the empirical Bayes decision problem considers possible priors on $(G,\underline{G})$. First we give the following definitions.

Definition 1.4. An empirical Bayes rule $T_n \in \mathcal{T}_n$ is Bayes with respect to a prior Λ on $\underline{G}$, if it is a infimizer (across $\mathcal{T}_n$) of

$$R_n(\Lambda,T_n) = \int R_n(G,T_n)\Lambda(dG)$$

Definition 1.5. T is said to be a Bayes empirical Bayes procedure if T_n is Bayes, $n \geq 1$. T is said to be a Bayes procedure with respect to a prior Λ if T_n is Bayes with respect to Λ, $n \geq 1$.

To construct a Bayes empirical Bayes rule at stage n, it is convenient to introduce the component risk set, $S = \{\underline{s} = (s_0,\ldots,s_m) |$ for some $t \in A$, $s_i = R(i,t)$, $i = 0,\ldots,m\}$. S is a convex subset of E_{m+1} which we will assume is compact throughout this thesis.

We will use the following theorem (LeCam (1956, Theorem 3.3.2)).

Theorem 1.1. Let $(X,\underline{X})$ be a measurable space and let Θ be a compact metric space. Let $f(x,\theta)$ be a function from $X \times \Theta$ to the real line. Assume that f is measurable in x for each θ

and continuous in θ for each x. Then it is possible to find a function $\hat{\theta}(x)$ which is measurable in x and such that

$$f(x,\hat{\theta}(x)) = \inf_{t\in\Theta} f(x,t). \qquad \square$$

For a given Λ, $T_n \in \mathcal{T}_n$

$$\begin{aligned} R_n(\Lambda,T_n) &= ER(G,T_n(\underline{X}_n)) \\ &= E_{(\Lambda)}E_{\Lambda}(R(G,T_n(\underline{X}_n))|\underline{X}_n) \\ &= E_{(\Lambda)} \sum_{\theta=0}^{m} R(\theta,T_n(\underline{X}_n))E_{\Lambda}(g_\theta|\underline{X}_n) \end{aligned}$$

Here $E_{\Lambda}(G|\underline{X}_n)$ is G-valued conditional expectation corresponding to the conditional distribution of G given $\underline{X}_n$ and $E_{(\Lambda)}$ corresponds to the mixture $P_{(\Lambda)}(\cdot) = \int P_G^n(\cdot)\Lambda(dG)$. Since $(R(0,T_n(\underline{x}_n)),\ldots,R(m,T_n(\underline{x}_n))) \in S$ for all $\underline{x}_n$ and T_n, to minimize $R_n(\Lambda,T_n)$ we seek a function $\delta\colon X^n \to S$ such that δ is measurable and

$$\sum_{\theta=0}^{m} \delta^{\theta}(\underline{X}_n)E_{\Lambda}(g_\theta|\underline{X}_n) = \inf_{\underline{s}\in S} \sum_{\theta=0}^{m} s_\theta E_{\Lambda}(g_\theta|\underline{X}_n)$$

where $\delta(\underline{X}_n) = (\delta^0(\underline{X}_n),\ldots,\delta^m(\underline{X}_n))$. By Theorem 1.1, such a δ exists.

Suppose δ is a measurable version, and for each $\underline{x}_n$, $T_{n,\Lambda}(\underline{x}_n) \in A$ is such that

$$\delta^{\theta}(\underline{x}_n) = R(\theta,T_{n,\Lambda}(\underline{x}_n)), \qquad \theta = 0,\ldots,m. \tag{1.3}$$

Then

$$T_{n,\Lambda} \in \mathcal{T}_n \quad \text{and} \quad R_n(\Lambda, T_{n,\Lambda}) = \inf_{T_n \in \mathcal{T}_n} R_n(\Lambda, T_n).$$

Also note that the Bayes empirical Bayes rule $T_{n,\Lambda}$ is pointwise component Bayes with respect to $E_\Lambda(G|\underline{X}_n) = (E_\Lambda(g_0|\underline{X}_n),\ldots,E_\Lambda(g_m|\underline{X}_n))$.

In what follows we sometimes will use the notation $G_\Lambda(\underline{X}_n)$ instead of $E_\Lambda(G|\underline{X}_n)$.

For a given prior Λ, let $T_{n,\Lambda} \in \mathcal{T}_n$ denote a Bayes empirical Bayes rule with respect to Λ which has the above form. We first discuss conditions that assure that $T_{n,\Lambda}$ is a.o.. Oaten (1972, (1.6)) shows that

$$0 \leqq R(G,t_F) - R(G) \leqq M \|G-F\| \tag{1.4}$$

for all $F,G \in \mathcal{G}$. Here M is a bound on the component risk, $\|\cdot\|$ is the ℓ_1 (total variation) norm on $\mathcal{E}_{m+1}$. Under the assumption that Λ has support all of $\mathcal{G}$, Gilliland and Boyer (1979) prove that

$$\lim_n \|G_\Lambda(\underline{X}_n) - G\| = 0 \quad \text{a.s. } P_G^\infty \quad \text{for all} \quad G \in \mathcal{G}, \tag{1.5}$$

where P_G^∞ is the probability distribution of $X_1,X_2,\ldots$. Thus, in the bounded risk case, (1.4) with $F = G_\Lambda(\underline{X}_n)$ and (1.5) establish that $T_{n,\Lambda}$ is a.o., that is,

$$\lim_n R_n(G,T_{n,\Lambda}) = R(G) \quad \text{for all} \quad G \in \mathcal{G}.$$

The above shows how the question of asymptotic optimality in the finite Ω Bayes empirical Bayes problem can often be reduced to a question of the consistency of the estimator $G_\Lambda(\underline{X}_n)$ for G.

To obtain the form of $G_\Lambda(\underline{X}_n)$, note that the conditional density of $\underline{X}_n$ is

$$f^n(\underline{x}_n|G) = \prod_{i=1}^{n} \sum_{j=0}^{m} f_j(x_i)g_j.$$

Hence the conditional probability of G given $\underline{X}_n$ has density

$$f(G|\underline{x}_n) = f^n(\underline{x}_n|G) / \int f^n(\underline{x}_n|G)\Lambda(dG)$$

with respect to Λ and

$$G_\Lambda(\underline{X}_n) = \int Gf(G|\underline{X}_n)\Lambda(dG). \tag{1.6}$$

We denote the components of $G_\Lambda(\underline{X}_n)$ by $g^\theta_\Lambda(\underline{X}_n)$, $\theta = 0,1,\ldots,m$.

We now develop an algorithm for computing $G_\Lambda(\underline{X}_n)$. Here we will represent G by the m-dimensional simplex in E_m

$$S_m = \{\underline{s}_m = (s_1,\ldots,s_m) | s_i \geq 0,\ i = 1,\ldots,m,\ \sum_{j=1}^{m} s_j \leq 1\}$$

and for $\underline{s}_m \in S_m$ let $s_0 = 1 - \sum_{j=1}^{m} s_j$. By (1.6)

$$g^\theta_\Lambda(\underline{X}_n) = \frac{\int s_\theta \prod_{i=1}^{n} \{\sum_{j=0}^{m} s_j f_j(X_i)\}\ \Lambda\ d(\underline{s}_m)}{\int \prod_{i=1}^{n} \{\sum_{j=0}^{m} s_j f_j(X_i)\}\ \Lambda\ (d\underline{s}_m)} \tag{1.7}$$

$$= \frac{\sum_{\ell_0+\ell_1+\ldots+\ell_m = n} S_{\ell_0,\ldots,\ell_m}(\underline{X}_n)\mu^{\theta}_{\ell_0,\ldots,\ell_m}}{\sum_{\ell_0+\ell_1+\ldots+\ell_m = n} S_{\ell_0,\ldots,\ell_m}(\underline{X}_n)\,\mu_{\ell_0,\ldots,\ell_m}}, \quad \theta=1,\ldots,m \tag{1.8}$$

where for each nonnegative integer partition $\ell_0,\ell_1,\ldots\ell_m$ of n,

$$S_{\ell_0,\ldots,\ell_m}(\underline{X}_n) = \sum_{\substack{B_0,\ldots,B_m \\ |B_0| = \ell_0,\ldots,|B_m| = \ell_m}} \Big\{ \prod_{j=1}^{m} \prod_{i\in B_j} [f_j(X_i)-f_0(X_i)] \Big\} \prod_{i\in B_0} f_0(X_i)\Big\}. \tag{1.9}$$

Here $B_1 U \ldots U B_m = \{1,\ldots,n\}$, $B_i \cap B_j = \phi$, $i,j = 0,\ldots,m$,

$$\mu_{\ell_0,\ldots,\ell_m} = \int_{S_m} s_1^{\ell_1}\ldots s_m^{\ell_m} \Lambda\, (d\underline{s}_m)$$

and

$$\mu^{\theta}_{\ell_0,\ldots,\ell_m} = \int_{S_m} (s_1^{\ell_1}\ldots s_m^{\ell_m})\, s_\theta \Lambda\, (d\underline{s}_m), \quad \theta = 1,\ldots,m.$$

The following theorem leads to a convenient way of computing (1.9).

Theorem 1.2. For each $n \geq 1$ and set of real numbers $\{a_{ij} | i=1,\ldots,n;\ j=0,\ldots,m\}$ define the function Q_n on S_m by

$$Q_n(s_1,\ldots,s_m) = \prod_{i=1}^{n} (a_{i0} + a_{i1}s_1 + \ldots + a_{im}s_m).$$

For each nonnegative integer partition $\ell_0,\ell_1,\ldots,\ell_m$ of n, let $C^n_{\ell_0,\ldots,\ell_m}$ denotes the coefficient for the term $s_1^{\ell_1} s_2^{\ell_2}\ldots s_m^{\ell_m}$ in the polynomial expansion for Q_n. Then

$$C^n_{\ell_0,\ldots,\ell_m} = \sum_{j=0}^{m} a_{nj}\, C^{n-1}_{\ell_0,\ldots\ell_j-1,\ldots\ell_m},\ n \geq 2, \tag{1.10}$$

with the convention $C^{n-1}_{k_0,\ldots,k_m} = 0$ if some $k_j = -1$.

Proof. The proof follows from the uniqueness of the coefficients $C^n_{\ell_0,\ldots,\ell_m}$ in the polynomial Q_n. □

To find all coefficients of $Q_n(s_1,\ldots,s_m)$, $n \geq 2$, we go through equation (1.10)

$$\sum_{k=2}^{n} |\{(\ell_0,\ldots,\ell_m) | \sum_{i=0}^{m} \ell_i = k;\ \ell_i \geq 0\}|$$

$$= \binom{n+m+1}{m+1} - (m+2)$$

$$\sim \frac{n^{m+1}}{(m+1)!}$$

times (see Feller (1957, (II.5.2) and (II.12.8)), where the sign $\sim$ is used to indicate that the ratio of the two sides tends to unity as $n \to \infty$. The limiting form is obtained by applying the Stirling's formula (Feller (1957)) and the l'Hôpital's rule. To apply Theorem 1.2 in computing (1.9), for each $\underline{x}_n \in X^n$, we let $a_{ij} = f_j(x_i) - f_0(x_i)$ and $a_{i0} = f_0(x_i)$, $i = 1,..,n$, $j = 1,\ldots,m$.

Then for each nonnegative integer partition $\ell_0,\ldots,\ell_m$ of n, we have $S_{\ell_0,\ldots,\ell_m}(\underline{x}_n) = C^n_{\ell_0,\ldots,\ell_m}$. Hence by (1.8),

$$g_{\Lambda}^{\theta}(\underline{X}_n) = \frac{\sum_{\ell_0+\ldots+\ell_m = n} C^{n}_{\ell_0,\ldots,\ell_m} \mu^{\theta}_{\ell_0,\ldots,\ell_m}}{\sum_{\ell_0+\ldots+\ell_m = n} C^{n}_{\ell_0,\ldots,\ell_m} \mu_{\ell_0,\ldots,\ell_m}}, \quad \theta = 1,\ldots,m. \qquad (1.11)$$

Note that $g_{\Lambda}^{\theta}(\underline{X}_n)$ depends on Λ only through a finite number of general moments of Λ. The computation of (1.11) is, in most cases, more efficient and accurate than a direct numerical integration in (1.7) or a direct evaluation of (1.8). Even in the case $m = 1$, a direct evaluation of $g_{\Lambda}^{\theta}(\underline{x}_n)$ through (1.9) is not feasible; in most cases, however, the application of (1.11) results in an efficient and accurate evaluation. Chapter 2 provides a detailed example.

Of course, computation with (1.11) is simplified when those general moments of Λ can be evaluated easily. Here we consider one such example:

EXAMPLE. (Bayes empirical Bayes with Dirichlet priors.) Let $\mathcal{D}(\alpha_1,\ldots,\alpha_m,\alpha_0)$ denote the m-variate Dirichlet distribution on the simplex S_m which has probability density function

$$f(\underline{s}_m) = \frac{\Gamma(\alpha_0+\ldots+\alpha_m)}{\Gamma(\alpha_0)\ldots\Gamma(\alpha_m)} s_1^{\alpha_1-1}\ldots s_m^{\alpha_m-1}(1-s_1-\ldots-s_m)^{\alpha_0-1}, \quad \underline{s}_m \in S_m,$$

where the α_i are all real and positive. If we let $\Lambda = \mathcal{D}(\alpha_1,\ldots,\alpha_m,\alpha_0)$, then it can be verified (Wilks (1962), (7.7.6)) that the general moment $\mu_{\ell_0,\ldots,\ell_m}$ of the m-variate Dirichlet prior Λ has the following value

$$\mu_{\ell_0,\ldots,\ell_m} = \frac{\Gamma(\alpha_1+\ell_1)\ldots\Gamma(\alpha_m+\ell_m)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_m)} \frac{\Gamma(\alpha_0+\ldots+\alpha_m)}{\Gamma(\alpha_0+\ldots+\alpha_m+\ell_1+\ldots+\ell_m)}.$$

Section 1.3. A complete class theorem

Gilliland and Boyer (1979) have suggested that, for each n, the study of empirical Bayes rules in T_n can be viewed as a study of the class of nonrandomized decision rules in a decision problem (G,D,R_n), so that the class B_n of Bayes empirical Bayes rules is the class of Bayes rules in (G,D,R_n). In this section we will prove that, in a large number of empirical Bayes problems, B_n is a complete class for (G,D,R_n). The results apply to each stage n, $n \geq 1$.

Definition 1.6. A class C of decision rules $C \subset D$, is said to be complete, if, given any rule t in D not in C, there exists a rule t^* in C that is better than t. A class C of decision rules is said to be essentially complete, if, given any rule t not in C, there exists a rule t^* in C that is as good as t.

Consider the decision problem (G,D,R_n) with sample space $(X^n,\underline{X}^n)$, parameter space $(G,\underline{G})$, $\{P_G^n: G \in G\}$ a family of probability measures on $(X^n,\underline{X}^n)$ dominated by $\mu^n,\underline{X}_n$ distributed P_G^n conditional on G, action space $(S,\underline{S})$ where $\underline{S}$ is the Borel σ-algebra on S, loss R: $G \times S \to [0,\infty)$ with $R(G,\underline{s}) = \sum_{\theta=0}^{m} g_\theta s_\theta$. The class of nonrandomized rules D is represented by the class of measurable transformations form $(X^n,\underline{X}^n)$ to $(S,\underline{S})$. Using rule $d = (d^0,d^1,\ldots d^m) \in D$, the expected loss when G is the true parameter is

$$R_n(G,d) = \int R(G,d(\underline{x}_n))P_G^n(d\underline{x}_n)$$

$$= \sum_{\theta=0}^{m} g_\theta \int d^\theta(\underline{x}_n)P_G^n(d\underline{x}_n) \qquad (1.12)$$

Note here (1.12) and the fact $d(\underline{x}_n) \in S$ implies that each $T_n \in \mathcal{T}_n$ determines a $d \in D$ such that T_n, d have the same risk function; conversely, for each $d \in D$ there exists a $T_n \in \mathcal{T}_n$ with the same risk function.

Let Λ be a prior on $\underline{G}$; the Bayes risk of $d \in D$ is

$$R_n(\Lambda,d) = \int R_n(G,d)\ \Lambda\ (dG).$$

A Bayes rule with respect to Λ is a rule $d_\Lambda \in D$ such that

$$R_n(\Lambda,d_\Lambda) = \inf_{d\in D} R_n(\Lambda,d).$$

Our discussion is restricted to nonrandomized rules because for $t \in \mathcal{D}$, $\mathcal{D}$ denoting the class of behavioral rules, we have

$$R_n(G,t) = \int\int R(G,\underline{s})t(\underline{x}_n,d\underline{s})P_G^n(d\underline{x}_n)$$

$$= \sum_{\theta=0}^{m} g_\theta \int\int \underline{s}t(\underline{x}_n,d\underline{s})P_G^n(d\underline{x}_n).$$

For each $\underline{x}_n$, define

$$d(\underline{x}_n) = \int \underline{s}t(\underline{x}_n,d\underline{s}).$$

Then, according to Lemma 2.7.3. of Ferguson (1967), we have $d(\underline{x}_n) \in S$. This implies $d \in D$ and d,t are equivalent (Definition 1.2). Therefore, the discussion of $t \in \mathcal{D}$ is redundant

as far as the risk behavior is concerned and we may restrict our consideration to the class of non-randomized rules D.

Assume that $\underline{X}$ is generated by a countable number of sets and recall S is a compact, convex subset of E_{m+1}. From previous discussions we know that the question as to whether B_n is complete is reduced to the question whether B, the class of nonrandomized Bayes rules in the game (G,D,R_n), is complete. To study this, we need the following lemma.

Lemma 1.1. A compact convex subset S of E_{m+1} is an intersection of countably many closed half spaces which contain it.

Proof:

Let Q_{m+1},Q_1 denote the rational points in E_{m+1},E_1, respectively, and define the countable collection of closed half-spaces,

$$S = \{H = \{\underline{x}|\underline{b}'\underline{x} \leq c\}|\underline{b} \in Q_{m+1},\ c \in Q_1 \text{ and } S \subset H\}.$$

We will show $S = \cap S$. Obviously, $S \subset \cap S$ so it remains to show $\cap S \subset S$.

Let $\underline{a} \notin S$. The separating hyperplane theorem (e.g., Rockafellar (1972), Corollary 11.4.1.), implies there exist $\underline{b}_0$, c_0 such that for all $\underline{s} \in S$, $\underline{b}_0'\underline{s} \leq c_0$ and $c_0 < \underline{b}_0'\underline{a}$. Let $\Delta = (\underline{b}_0'\underline{a} - c_0)/3$. The fact that S is bounded and Q_{m+1} is dense implies the existence of a $\underline{b} \in Q_{m+1}$ such that for all $\underline{s} \in S$, $\underline{b}'\underline{s} \leq c_0 + \Delta$ and $\underline{b}_0'\underline{a} - \Delta \leq \underline{b}'\underline{a}$. The denseness of Q_1 establishes the existence of a $c \in Q_1$ such that $c_0 + \Delta < c < \underline{b}_0'\underline{a} - \Delta$.

Then $S \subset \{\underline{x} | \underline{b}'\underline{x} \leq c\} = H$ and $c < \underline{b}'\underline{a}$ so $a \notin H$. Since $H \in S$, $\underline{a} \notin \cap S$. □

The remark preceding Definition 1.1 of Section 1.1 implies that in the bounded S case, all empirical Bayes risk functions $R_n(G,d)$ are polynomials in $G = (g_0, g_1, \ldots g_m)$ on $\mathcal{G}$. Identify a $d \in D$ by the risk function which in turn is identified by a vector $\underset{\sim}{d} = (\underline{d}^1, \ldots, \underline{d}^N)$ of coefficients in the polynomial. Let $\underset{\sim}{D} = \{\underset{\sim}{d} | d \in D\}$ and let $\|\cdot\|$ denote the usual Euclidean norm in E_N.

Theorem 1.3. $\underset{\sim}{D}$ is a compact subset of E_N.

Proof: The proof will be based on the fact that D is (componentwise) weakly compact.

Since $\underset{\sim}{D}$ is a subset of a metric space it is sufficient to show that $\underset{\sim}{D}$ is sequentially compact (Munkres (1975) p. 181), that is, every sequence in $\underset{\sim}{D}$ has a convergent subsequence. Let $\{\underset{\sim}{d}_i\} \subset \underset{\sim}{D}$. Let $\{d_i\}$ be a corresponding sequence in D such that $\underset{\sim}{d}_i$ is generated by d_i. Since $\{d_i\}$ is bounded, by the weak compactness theorem (Lehmann (1959), p. 354), there exists a real valued measurable function d_0^0 and a subsequence $\{d_{n_i}^0\}$ of $\{d_i^0\}$ such that

$$\lim_i \int d_{n_i}^0(\underline{x}_n) h(\underline{x}_n) \mu^n(d\underline{x}_n) = \int d_0^0(\underline{x}_n) h(\underline{x}_n) \mu^n(d\underline{x}_n)$$

for all integrable h.

Again, apply the weak compactness theorem to the sequence $\{d_{n_i}^1\}$. There exists a real valued masurable function d_0^1 and a further subsequence $\{d_{m_i}^1\}$ of $\{d_{n_i}^1\}$ such that

$$\lim_i \int d^1_{m_i}(\underset{\sim}{x}_n)h(\underset{\sim}{x}_n)\mu^n(d\underset{\sim}{x}_n) = \int d^1_0(\underset{\sim}{x}_n)h(\underset{\sim}{x}_n)\mu^n(d\underset{\sim}{x}_n)$$

for all integrable h.

Repeat the above process we obtain a measurable transformation from X^n to E_{m+1}, $d_0 = (d^0_0,\ldots,d^m_0)$ and a subsequence $\{d_{k_i}\}$ of $\{d_i\}$ such that

$$\lim_i \int d^\theta_{k_i}(\underset{\sim}{x}_n)h(\underset{\sim}{x}_n)\mu^n(d\underset{\sim}{x}_n) = \int d^\theta_0(\underset{\sim}{x}_n)h(\underset{\sim}{x}_n)\mu^n(d\underset{\sim}{x}_n) \tag{1.13}$$

for all integrable h and for $\theta = 0,\ldots,m$. By the expression following (1.1), each coefficient of d_{k_i} in the polynomial $R_n(G,d_{k_i})$ is of the form (1.13) so from (1.13)

$$\lim_i \|\underset{\sim}{d}_{k_i} - \underset{\sim}{d}_0\| = 0$$

It remains to be shown that $\underset{\sim}{d}_0 \in \underset{\sim}{\mathcal{D}}$ or equivalently, $P^n_G[d_0 \in S] = 1$ for all $G \in \mathcal{G}$. Let $S = \{H = \{\underset{\sim}{x} | \underset{\sim}{b}'\underset{\sim}{x} \leq c\} | \underset{\sim}{b} \in \mathcal{Q}_{m+1}, c \in \mathcal{Q}_1$ and $S \subset H\}$. We claim that $P^n_G[d_0 \in H] = 1$ for all $G \in \mathcal{G}$ and $H \in S$. To see this, suppose $\underset{\sim}{b} \in E_{m+1}$, $c \in E_1$, and $\underset{\sim}{b}'\underset{\sim}{s} \leq c$ for all $\underset{\sim}{s} \in S$. Then

$$\begin{aligned}
&cP^n_G[\underset{\sim}{b}'d_0 > c] \\
&\leq \int_{[\underset{\sim}{b}'d_0 > c]} \underset{\sim}{b}'d_0(\underset{\sim}{x}_n)f^n_G(\underset{\sim}{x}_n)\mu^n(d\underset{\sim}{x}_n) \quad (\text{with } < \text{ if } P^n_G[\underset{\sim}{b}'d_0 > c] > 0) \\
&= \sum_{\theta=0}^m \int_{[\underset{\sim}{b}'d_0 > c]} b_\theta d^\theta_0(\underset{\sim}{x}_n)f^n_G(\underset{\sim}{x}_n)\mu^n(d\underset{\sim}{x}_n) \\
&= \sum_{\theta=0}^m \lim_i \int_{[\underset{\sim}{b}'d_0 > c]} b_\theta d^\theta_{k_i}(\underset{\sim}{x}_n)f^n_G(\underset{\sim}{x}_n)\mu^n(d\underset{\sim}{x}_n)
\end{aligned}$$

$$= \lim_i \int_{[\underline{b}'d_0 > c]} \underline{b}'d_{k_i}(\underline{x}_n) f_G^n(\underline{x}_n) \mu^n(d\underline{x}_n)$$

$$\leqq c P_G^n [\underline{b}'d_0 > c] ,$$

where the last inequality follows from the fact $\underline{b}'d_{k_i}(\underline{x}_n) \leqq c$ for all i. Therefore, we have $P_G^n [\underline{b}'d_0 > c] = 0$, i.e., $P_G^n [\underline{b}'d_0 \leqq c] = 1$. This proves the claim.

From the fact that S is countable and the above claim, we obtain the result

$$1 = P_G^n \{ \bigcap_{H \in S} [d_0 \in H] \} = P_G^n [d_0 \in \cap S] \quad \text{for all} \quad G \in \mathcal{G}.$$

But Lemma 1.1 shows that $S = \cap S$. Therefore $P_G^n [d_0 \in S] = 1$ for all $G \in \mathcal{G}$. This completes our proof. □

Corollory 1.1 There exists a topology on D such that (a) D is compact and (b) $R_n(G,d)$ is continuous in $d \in D$ for all $G \in \mathcal{G}$.

Proof: Define φ to be a function on D such that $\varphi(d) = \underset{\sim}{d}$ for all $d \in D$. Then the collection of sets

$$F = \{\varphi^{-1}(A) | A \text{ open in } \underset{\sim}{D}\}$$

is a topology on D such that φ: $D \to \underset{\sim}{D}$ is continuous. Since φ is onto, if $\mathcal{Q}$ is a covering of $\underset{\sim}{D} = \varphi(D)$ then $\{\varphi^{-1}(A) | A \in \mathcal{Q}\}$ is a covering of D. Hence the compactness of $\underset{\sim}{D}$ from Theorem 1.3. implies that (D,F) is compact.

Form the polynomial form of $R_n(G,d)$, $R_n(G,d)$ is a linear combination of $\underset{\sim}{d}^i = \pi_i \circ \varphi(d)$, $i = 1,\ldots,N$ where

Π_i: $E_N \to E_1$, $i = 1,\ldots,N$, is the projection map. Therefore the continuity of $\Pi_i \circ \varphi$, $i = 1,\ldots,N$, implies that $R_n(G,d)$ is continuous in $d \in D$ for all $G \in \mathcal{G}$. □

Definition 1.7. A rule $d \in D$ is extended Bayes if for every $\varepsilon > 0$ there is a prior distribution Λ such that

$$R_n(\Lambda,d) \leqq \inf_{d\in D} R_n(\Lambda,d) + \varepsilon$$

The following theorem follows immediately from Corollary 1.1 and Theorem 2.10.3 of Ferguson (1967).

Theorem 1.4. The class of extended Bayes rules in D is essentially complete.

Theorem 1.5. Any extended Bayes rule in D is a Bayes rule.

Proof: For $d \in D$, $R_n(\cdot,d)$ is continuous in $\mathcal{G}$. Let $d \in D$ be an extended Bayes procedure. Then for each positive integer N, there exists a prior distribution Λ_N such that

$$\int R_n(G,d_{\Lambda_N})\Lambda_N(dG) \leqq \int R_n(G,d)\Lambda_N(dG)$$

$$\leqq \int R_n(G,d_{\Lambda_N})\Lambda_N(dG) + 1/N. \qquad (1.14)$$

Since $\mathcal{G}$, a closed subset of $[0,1]^m$, is compact, the class $\{\Lambda_N\}_{N=1}^{\infty}$ is tight. By the Prohorov theorem (Billingsley (1968)) $\{\Lambda_N\}_{N=1}^{\infty}$ is relatively compact which means that there exists a prior Λ and a subsequence $\{\Lambda_N'\}_{N=1}^{\infty} \subset \{\Lambda_N\}_{N=1}^{\infty}$ such that Λ_N' converges weakly to Λ as $N \to \infty$.

Con

The

1.

Th

Pr

Ba

fo

R_n

in

d_{λ}

De

co

co

en

Consequently,

$$\begin{aligned}
\int R_n(G,d_\Lambda)\Lambda(dG) &\leqq \int R_n(G,d)\Lambda(dG) \\
&= \lim_N \int R_n(G,d)\Lambda'_N(dG) \\
&\leqq \overline{\lim_N} \int R_n(G,d_{\Lambda'_N})\Lambda'_N(dG) \qquad \text{by} \quad (1.14) \\
&\leqq \overline{\lim_N} \int R_n(G,d_\Lambda)\Lambda'_N(dG) \\
&= \int R_n(G,d_\Lambda)\Lambda(dG).
\end{aligned}$$

The above shows that d is Bayes with respect to Λ. □

Our complete class theorem follows directly from Theorem 1.4. and Theorem 1.5.

Theorem 1.6. The class of Bayes empirical Bayes rules is complete.

Proof: From Theorems 1.4., 1.5. we know that the class of extended Bayes rules is equal to B and is essentially complete. Therefore, for $d \notin B$, there exists a Bayes rule d_Λ such that $R_n(G,d_\Lambda) \leqq R_n(G,d)$ for all $G \in \mathcal{G}$. If "=" holds for all G in $\mathcal{G}$ then d is Bayes with respect to Λ, a contradiction, so d_Λ is better than d. This implies that B is complete. □

Definition 1.7. A class C of decision rules is said to be minimal complete if C is complete and if no proper subclass of C is complete.

It is also of interest to know when the class of Bayes empirical Bayes rules will constitute a minimal complete class.

The minimal complete class, when it exists, is exactly the class of admissible rules. Since $\mathcal{B}_n$ has been proved to be a complete class, any admissible rule will be in $\mathcal{B}_n$. It is then sufficient to find conditions under which the Bayes empirical Bayes rules are admissible. The following remark is needed in the proof of Theorem 1.7.

Remark 1.1. If the members of $\{P_\theta | \theta \in \Omega\}$ are mutually absolutely continuous then so are the $\{P_G | G \in \mathcal{G}\}$ which implies that the product measures $\{P_G^n | G \in \mathcal{G}\}$ are mutually absolutely continuous and equivalent to any mixture $P_{(\Lambda)}$.

Theorem 1.7. Suppose that $\{P_\theta, \theta \in \Omega\}$, are mutually absolutely continuous and that the Bayes component decision rules are unique up to risk equivalence. Then the class of Bayes empirical Bayes rules is minimal complete.

Proof: Since the class of Bayes empirical Bayes rules is complete, if we show that the Bayes empirical Bayes rules are admissible, then $\mathcal{B}_n$ is minimal complete.

For a given Λ, let $T_{n,\Lambda}$ be the Bayes empirical Bayes rule with respect to Λ as defined in (1.3). Then for $T_n \in \mathcal{T}_n$,

$$R(G_\Lambda(\underline{X}_n), T_n(\underline{X}_n)) \geq R(G_\Lambda(\underline{X}_n), T_{n,\Lambda}(\underline{X}_n)) \tag{1.15}$$

Suppose $T_n \in \mathcal{T}_n$ is Bayes with respect to Λ. Then

$$R_n(\Lambda, T_n) = R_n(\Lambda, T_{n,\Lambda}) \tag{1.16}$$

and (1.15) and (1.16) implies

$$R(G_\Lambda(\underline{x}_n), T_n(\underline{x}_n)) = R(G_\Lambda(\underline{x}_n), T_{n,\Lambda}(\underline{x}_n)) \quad \text{a.s.} \quad P_{(\Lambda)}. \tag{1.17}$$

By our hypothesis, the Bayes component rules are unique up to risk equivalence, which means if $t_1, t_2 \in A$ are Bayes with respect to G, then $R(\theta,t_1) = R(\theta,t_2)$, $\theta = 0,\ldots,m$. This and (1.17) implies that

$$R(\theta,T_n(\underline{x}_n)) = R(\theta,T_{n,\Lambda}(\underline{x}_n)), \quad \theta = 0,\ldots,m. \quad \text{a.s.} \quad P_{(\Lambda)}$$

By Remark 1.1, the above equalities holds a.s. P_G^n for all $G \in \mathcal{G}$, so that

$$R_n(G,T_n) = R_n(G,T_{n,\Lambda}) \quad \text{for all} \quad G \in \mathcal{G}.$$

i.e., T_n is equivalent to $T_{n,\Lambda}$. Thus, the Bayes rule with respect to Λ is unique up to risk equivalence. It is well known that if a Bayes rule is unique up to risk equivalence then it is admissible. □

Empirical Bayes classification between $N(-1,1)$ and $N(1,1)$ is a decision problem satisfying the hypothesis of Theorem 1.7. This example is the subject of computation and study in Section 2.3.

Boyer and Gilliland (1980, Theorem 4) point out how the continuity of risk functions $R_n(G,T_n)$ in G ensures that $T_{n,\Lambda}$ is admissible if Λ has support all of $\mathcal{G}$.

Section 1.4. The classification problem

In this section we will derive the form of the Bayes empirical Bayes rules for classification problems. A classification problem will provide an example for the application of the algorithm developed in (1.11) for computing Bayes empirical Bayes rules. In a classification problem, an observation is to be classified as comming from one of $m + 1$ distributions. Specifically, we let $A = \{0,1,\ldots,m\} = \Omega$ and the loss be α if an incorrect classification is made and β if a correct classification is made, $\alpha > \beta \geqq 0$.

Recall, $G = (g_0,\ldots,g_m)$ represents a probability measure on Ω. Conditional on $X = x$, the distribution of θ has density

$$f(\theta|x) = f_\theta(x)g_\theta / \sum_{j=0}^{m} g_j f_j(x) \qquad \theta = 0,\ldots,m.$$

For each $a \in \{0,\ldots,m\}$ and $x \in X$,

$$\begin{aligned} E(L(\theta,a)|x) &= \sum_{\theta=0}^{m} L(\theta,a)f(\theta|x) \\ &= \alpha - (\alpha - \beta)f(a|x) \\ &\geqq \alpha - (\alpha - \beta) \max_{i\in\Omega} f(i|x) \end{aligned}$$

Define $d_G(X) = \max \{\theta | f(\theta|X) = \max_{i\in\Omega} f(i|X),\ \theta \in \Omega\}$

$$= \max \{\theta | f_\theta(X)g_\theta = \max_{i\in\Omega} f_i(X)g_i,\ \theta \in \Omega\} \qquad (1.18)$$

Then d_G is a non-randomized component decision rule which is Bayes with respect to G.

From the discussions in last section we know that $T_{n,\Lambda}(\underline{x}_n)$ choses a Bayes component rule with respect to $G_\Lambda(\underline{x}_n)$. Therefore, to implement the Bayes empirical Bayes rule with respect to Λ, first evaluate $G_\Lambda(\underline{x}_n)$ and then replace $f_\theta(x)g_\theta$ in (1.18) by $f_\theta(x_{n+1})g^\theta_\Lambda(\underline{x}_n)$.

It is known (Ferguson (1967)) that when both Ω and A are finite, the risk set S is compact. Hence in classification problems, Theorem 1.6. implies that the class of Bayes empirical Bayes rules $\mathcal{B}_n$ is complete at each stage $n \geq 1$.

CHAPTER II

TWO STATE BAYES EMPIRICAL BAYES PROCEDURES

The studies based on a two state component decision problem have a long history. Robbins (1951) studied the compound decision problem and discussed both bootstrap and Bayes rules. After that, Hannan and Robbins (1955), Hannan and Van Ryzin (1965), Huang (1970), Van Houwelingen (1974), Shapiro (1974), Gilliland, Hannan and Huang (1976), Snijder (1977) have studied two state component decision problems through either the compound or empirical Bayes approach. Some of these discuss the rate of convergence for a.o., some discuss finite state risk behavior. Only Snijder found a complete class among a class of decision procedures under consideration.

In this chapter, we will study the two state component Bayes empirical Bayes procedures. Section 2.1. formulates the classification problem between two subpopulations and applies the computing algorithm developed in Section 1.2. to evaluate the Bayes empirical Bayes rules. In Section 2.2. we give sufficient conditions under which the limiting distributions of the posterior means are asymptotically normally distributed. In order to illustrate the properties of Bayes empirical Bayes procedures we examine empirical Bayes classification between $N(-1,1)$ and $N(1,1)$ in

Section 2.3, where we know that the class of Bayes empirical Bayes procedures is minimal complete at each stage n. With the help of the algorithm developed from (1.11), the computations of the risk functions are simplified, so that in Section 2.4. we are able to compare the risk performance of Bayes empirical Bayes procedures with other empirical Bayes procedures for selected priors Λ and n. Van Houwelingen (1974) has discussed the empirical Bayes approach to the classical problem of testing a simple hypothesis against a simple alternative. He has proposed a non-Bayes empirical Bayes procedure as an improvement over the original Robbins rule. The fact that the Robbins and Van Houwelingen rules are inadmissible is also established in Section 2.4. Section 2.5 explores their risk behavior and compares it with that of selected Bayes empirical Bayes procedures.

Section 2.1. Testing simple hypothesis against simple alternative

In order to demonstrate the feasibility of Bayes empirical Bayes approach we study a two state classification problem, i.e., $m = 1$, with the following component model:

$\Omega = \{0,1\} = A$

P_0, P_1 have densities f_0, f_1 respectively

Loss function

$$L(\theta,a) = \begin{cases} 1 & \text{if } \theta \neq a \\ 0 & \text{if } \theta = a. \end{cases}$$

Identify G by the mass p it puts on 1, so e.g., we write $R(p,t)$ in place of $R(G,t)$. Thus P_p has density $f_p = pf_1 + (1-p)f_0$, $P \in [0,1] = S_1$. From (1.18), a nonrandomized version of a component Bayes rule is

$$t_p(x) = 1 \quad \text{if} \quad pf_1(x) \geq (1-p)f_0(x) \qquad (2.1)$$

$$0 \quad \text{if} \quad pf_1(x) < (1-p)f_0(x).$$

Now let $p_\Lambda(\underline{X}_n)$ denote $g^1_\Lambda(\underline{X}_n)$. By (1.7)

$$p_\Lambda(\underline{X}_n) = \frac{\int_0^1 p \prod_{i=1}^n [pf_1(X_i) + (1-p)f_0(X_i)]\Lambda(dp)}{\int_0^1 \prod_{i=1}^n [pf_1(X_i) + (1-p)f_0(X_i)]\Lambda(dp)}$$

$$= \frac{S_n(\underline{X}_n)m_{n+1} + \ldots + S_0(\underline{X}_n)m_1}{S_n(\underline{X}_n)m_n + \ldots + S_1(\underline{X}_n)m_1 + S_0(\underline{X}_n)} \qquad (2.2)$$

where $m_i = \int_0^1 p^i \Lambda(dp)$; and, for $B \subset \{1,\ldots,n\}$, $|B|$ = cardinal number of B and

$$S_k(\underline{X}_n) = \sum_{\substack{B \\ |B|=k}} \prod_{i\in B} [f_1(X_i) - f_0(X_i)] \prod_{j\notin B} f_0(X_j) \quad k = 0,1,\ldots,n.$$

Direct computation of $p_\Lambda(\underline{X}_n)$ by (2.2) involves the search of $\binom{n}{k}$ subsets of $\{1,\ldots,n\}$ for each value of $S_k(\underline{X}_n)$ (2^n in total) and by Stirling's formula $\binom{2n}{n} \sim (\pi n)^{-\frac{1}{2}} 2^{2n}$ increases quite rapidly.

To apply Theorem 1.2. in this special case ($m=1$), observe that the function Q_n on $S_1 = [0,1]$ is

$$Q_n(s) = \prod_{i=1}^{n} (a_{i0} + a_{i1}s), \quad n \geq 1.$$

Let C_k^n denotes the coefficient for the term s^k in the polynomial for Q_n. Then (1.10) becomes

$$C_k^n = a_{n0}C_k^{n-1} + a_{n1}C_{k-1}^{n-1}, \quad n \geq 2 \tag{2.3}$$

with $C_{-1}^{n-1} = C_n^{n-1} = 0$. Hence, to compute all the coefficients of $Q_n(s)$, we only need to go through the recursive definition (2.3) $(n^2 + 3n - 4)/2 \sim n^2/2$ times.

Let $a_{i1} = f_1(x_i) - f_0(x_i)$; $a_{i0} = f_0(x_i)$, $i = 1,\ldots,m$. Then $S_k(\underline{x}_n) = C_k^n$; $k = 0,\ldots,m$, and, therefore,

$$p_\Lambda(\underline{x}_n) = \frac{C_n^n m_{n+1} + \ldots + C_1^n m_2 + C_0^n m_1}{C_n^n m_n + \ldots + C_1^n m_1 + C_0^n} \tag{2.4}$$

has the form of (1.11). (Note that $p_\Lambda(\underline{x}_n)$ depends on Λ only through the first $n + 1$ moments of Λ.) The Bayes empirical Bayes rule is given by (2.1) with $p = p_\Lambda(\underline{X}_n)$ as in (2.4); this was discussed in Section 1.4.

We now turn to risk behavior. For $T_{n,\Lambda}$ the conditional expected loss given $\underline{X}_n$ is

$$R(p,T_{n,\Lambda}(\underline{X}_n)) = R(p,t_{p_\Lambda(\underline{X}_n)})$$

$$= p\int_{[p_\Lambda(\underline{X}_n)f_1 < (1-p_\Lambda(\underline{X}_n))f_0]} f_1(t)dt + (1-p)\int_{[p_\Lambda(\underline{X}_n)f_1 \geq (1-p_\Lambda(\underline{X}_n))f_0]} f_0(t)dt \tag{2.5}$$

Also, the risk function $R_n(\cdot,T_{n,\Lambda})$ has the form

$$R_n(p,T_{n,\Lambda}) = \int R(p,T_{n,\Lambda}(\underline{x}_n)) \prod_{i=1}^{n} f_p(x_i)d\underline{x}_n \qquad (2.6)$$

which is a polynomial in p with degree at most equal to $n + 1$.

From the observations in Chapter 1, we see that a Bayes empirical Bayes procedure $T_\Lambda = (T_{1,\Lambda},T_{2,\Lambda},\ldots)$ has the properties

(i) T_Λ is a.o. if Λ has support $(0,1)$

(ii) $T_{n,\Lambda}$ is admissible, if Λ has support $(0,1)$

(iii) $T_{n,\Lambda}$ is admissible, if $\{P_\theta|\theta \in \Omega\}$ are mutually absolutely continuous and if the Bayes component rules t_p are unique up to risk equivalence. Admissibility results in a good risk performance for small values of n.

(iv) $T_{n,\Lambda} \in \mathcal{B}_n$, $\mathcal{B}_n$ is a complete class, $n \geq 1$.

(v) An algorithm for computing the decision procedure is based on (2.4) which can be executed economically.

(vi) The performance of $T_{n,\Lambda}$ can be adjusted by choice of Λ. Low risk over a region of $[0,1]$ is obtained by choice of Λ concentrating on that region.

In later sections we will study the risk performance of $T_{n,\Lambda}$ along with other empirical Bayes decision procedures. All the properties (i) - (vi) above will be demonstrated in a class of examples. Also, for notational convenience n will not be displayed in denoting empirical Bayes rules. It will be clear from the context whether a sequence of decision rules (procedure) or a decision rule is being discussed.

Section 2.2. Asymptotic property of $p_\Lambda(\underline{X}_n)$

This section is a slight digression in which asymptotic properties of the posterior mean $p_\Lambda(\underline{X}_n)$ and maximum likelihood estimator $\hat{p}(\underline{X}_n)$ are stated. The proofs are deferred until Appendix B. A theorem of LeCam (1956) is used to prove the asymptotic normality of $\sqrt{n}\,(\hat{p}(\underline{X}_n) - p)$ and one of Johnson (1970) to prove the asymptotic equivalence $\sqrt{n}\,(p_\Lambda(\underline{X}_n) - \hat{p}(\underline{X}_n)) \to 0$. The approach is similar to that of Shapiro (1972) in establishing the asymptotic normality of the cut point in the Bayes empirical Bayes rule.

The product of mixtures density $f_p^n(\underline{x}_n)$ is continuous in $p \in [0,1]$, a compact subset of E_1. Theorem 1.1 ensures the existence of a maximum likelihood estimator $\hat{p}$. Whereas the evaluation of $\hat{p}(\underline{X}_n)$ is a difficult computation, the Bayes estimator is easily computed by the recursive formula developed in Section 2.1.

Gilliland, Hannan and Huang (1976) show that the maximum likelihood estimator $\hat{p}$ is consistent for the empirical proportion of states "$\theta_i = 1$" in the independent non-identically distributed compound model and the consistency result is inherited by the empirical Bayes model. Likewise, their results on the consistency of p_Λ transfer to the empirical Bayes model. The theorems to follow place stronger assumptions on the model and prior but yield the asymptotic normality in addition to the consistency for p in the interval $(0,1)$.

Theorem 2.1 If $\int |\log f_i(x)| P_j(dx) < \infty$ for $i,j \in \{0,1\}$ and the true parameter $p_0 \in (0,1)$, then

$$\hat{p}(\underline{X}_n) \to p_0 \quad \text{a.s.} \quad P^{\infty}_{p_0}$$

and

$$\sqrt{n}(\hat{p}(\underline{X}_n) - p_0) \to N(0, I(p_0)^{-1}) \quad \text{in distribution}$$

where

$$I(p) = -E_p \frac{\partial^2}{\partial p^2} \log f_p(x)$$

Proof: (In Appendix B) □

Theorem 2.2 Suppose Λ is a prior on $[0,1]$ which has density λ with respect to the Lebesgue measure where $\lambda(p_0) > 0$ and $\lambda(\cdot)$ has three continuous derivatives in a neighborhood of the true parameter $p_0 \in (0,1)$. If P_0, P_1 are mutually absolutely continuous and if $\int |\log f_i(x)| P_j(dx) < \infty$ for $i,j \in \{0,1\}$ then

$$\sqrt{n}(p_{\Lambda}(\underline{X}_n) - \hat{p}(\underline{X}_n)) \to 0 \quad \text{a.s.} \quad P^{\infty}_{p_0}$$

Proof: (In Appendix B) □

As a consequence of Theorem 2.1 and Theorem 2.2, under the hypothesis of Theorem 2.2, $p_{\Lambda}(\underline{X}_n) \to p_0$ a.s. $P^{\infty}_{p_0}$ and $\sqrt{n}(p_{\Lambda}(\underline{X}_n) - p_0) \to N(0, I(p_0)^{-1})$ in distribution.

Section 2.3. Optimal properties and risk performance of Bayes empirical Bayes procedures for classification between $N(-1,1)$ and $N(1,1)$.

To illustrate the risk performance of Bayes empirical Bayes procedures we will study the following example.

EXAMPLE: Testing $N(-1,1)$ against $N(1,1)$.

In this example we have $X = E_1$, $f_0(x) = (2\pi)^{-\frac{1}{2}}\exp\{-(x+1)^2/2\}$ and $f_1(x) = (2\pi)^{-\frac{1}{2}}\exp\{-(x-1)^2/2\}$. By (2.1) a nonrandomized version of a component Bayes rule is:

$$t_p(x) = 1 \quad \text{if} \quad x \geq C_p$$
$$0 \quad \text{if} \quad x < C_p \tag{2.7}$$

where $C_p = \frac{1}{2}\ell n\left(\frac{1-p}{p}\right)$. The Bayes empirical Bayes rule $T_\Lambda(\underline{X}_n)$ simply replaces p in (2.7) with $p_\Lambda(\underline{X}_n)$. By (2.4) and (2.7) an algorithm for computing the Bayes empirical Bayes rule is already available. If Λ is chosen as the probability measure corresponding to a mass 1 at p, then $T_\Lambda \equiv t_p$ is the Bayes empirical Bayes procedure with respect to Λ. In particular, with $\Lambda(\frac{1}{2}) = 1$, $T_\Lambda = t_{\frac{1}{2}}$ is the minimax procedure with constant risk $R_n(p,T_\Lambda) = P_0(X \geq 0) = 0.1587$ for all $p \in [0,1]$ and $n \geq 1$.

Observe that the risk set of t_p, $p \in [0,1]$, is $\{(s_0,s_1) | s_0 = P_0[X \geq a],\ s_1 = P_1[X < a]$ for some $a \in [-\infty,\infty]\}$; this together with the form of (2.7) implies that the component Bayes rules are unique up to risk equivalence. By Theorem 1.7

we see that at each stage n, the class of Bayes empirical Bayes rules is minimal complete in this example.

In our applications we will deal with those Λ that belong to a given parametric family $\mathcal{B} = \{B(\gamma) | \gamma > 0\}$ where $B(\gamma)$ denotes a symmetric beta distribution on $(0,1)$ with density

$$g_{B(\gamma)}(p) = \frac{\Gamma(2\gamma)}{[\Gamma(\gamma)]^2} p^{\gamma-1}(1-p)^{\gamma-1} \quad \text{for} \quad 0 < p < 1.$$

From previous discussions we note that $\{T_\Lambda: \Lambda \in \mathcal{B}\}$ are asymptotically optimal procedures and are admissible at each stage n. Also note that assumptions in Theorem 2.2 are satisfied, so that $p_\Lambda(\underline{X}_n)$ is asymptotically normally distributed. The variance of the limiting distribution of $\sqrt{n}\,(p_\Lambda(\underline{X}_n) - p_0)$ is $I(p_0)^{-1}$. (Behboodian (1972) discussed the conditional moments of p for Beta priors.)

Remark 2.1. If Λ has a density $g_\Lambda(p)$ which is symmetric about 1/2, then $R_n(p,T_\Lambda) = R_n(1-p, T_\Lambda)$ for $p \in [0,1]$. To see this, observe

(i) $f_p(-x) = f_{1-p}(x)$

(ii) $p_\Lambda(-\underline{x}_n) = 1-p_\Lambda(\underline{x}_n)$ (by elementary calculus)

(iii) $C_\Lambda(-\underline{x}_n) = -C_\Lambda(\underline{x}_n)$ (a direct result of (ii))

where $C_\Lambda(\underline{X}_n) = \frac{1}{2} \ell m\,[(1-p_\Lambda(\underline{X}_n))/p_\Lambda(\underline{X}_n)]$. Since (iii) implies

$$R(p,T_\Lambda(-\underline{x}_n)) = R(1-p, T_\Lambda(\underline{x}_n)),$$

the remark is verified by appealing to (2.6), (i) and (iii).

Since $R_n(p,T_\Lambda)$ is a polynomial in p (see (2.6)), the Remark 2.1. implies that for $\Lambda \in B$, $R_n(p,T_\Lambda)$ is a function of $(p-\frac{1}{2})^2$; hence, it has an even degree less than or equal to $n+1$. With $n = 1$ or 2 one can readily see that $R_n(p,T_\Lambda)$ will be a horizontal line or a parabola with extremum at 1/2.

We will compute the values $R_n(p,T_\Lambda)$ for $p \in [0,1]$ when $n = 1$ or 2. Using (2.5), (2.6) and results (i), (ii), (iii) of the Remark, elementary calculus shows

$$R_1(p,T_\Lambda) = 2(a-b)p^2 + 2(b-a)p + a$$

with

$$a = \int_{-\infty}^{\infty} \int_{-\infty}^{C_\Lambda(x_1)} f_1(x)dx\, f_1(x_1)dx_1$$

$$b = \int_{-\infty}^{\infty} \int_{-\infty}^{C_\Lambda(x_1)} f_1(x)dx\, f_0(x_1)dx_1.$$

Also, a tedious calculation shows that

$$R_2(p,T_\Lambda) = [3c - (2d + e)]p^2 + [(2d + e) - 3c]p + c$$

with

$$c = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{C_\Lambda(x_1,x_2)} f_1(x)dx\, f_1(x_1)f_1(x_2)dx_1dx_2$$

$$d = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{C_\Lambda(x_1,x_2)} f_1(x)dx\, f_1(x_1)f_0(x_2)dx_1dx_2$$

$$e = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{C_\Lambda(x_1,x_2)} f_1(x)dx\, f_0(x_1)f_0(x_2)dx_1dx_2.$$

For the case $\Lambda = B(1)$, the uniform distribution on $(0,1)$, numerical computations supported by softwares from IMSL (1979) (Table A.6) subroutines were used to compute: $a = 0.12071$, $b = 0.21212$, $c = 0.09576$, $d = 0.16486$, $e = 0.25720$. Using MSU CDC 6500 computer the accuracy of computing a,b,c,d,e was controlled at 3 to 4 significant decimal digits. Therefore:

$$R_1(p,T_{B(1)}) = -0.1828p^2 + 0.1828p + 0.1207 \qquad (2.8)$$

$$R_2(p,T_{B(1)}) = -0.2997p^2 + 0.2997p + 0.0958 \qquad (2.9)$$

are parabolas concave downward with extremum at $p = \frac{1}{2}$.

The direct numerical computations for $n \geq 2$ and $\Lambda \in B$ are in general not feasible; to overcome this difficulty, Monte Carlo integration method was used to evaluate $R_n(p,T_\Lambda)$. For $\Lambda \in B$, we generate independently L sample sequences of independent random variables $X_1,\ldots,X_n$ from a population having $f_p(x)$ as density. For each of the L sequences generated, we then compute $R(p,T_\Lambda(\underline{x}_n))$ based on (2.4) and (2.5). An estimate of $R_n(p,T_\Lambda)$ is obtained by averaging the L computed values of $R(p,T_\Lambda(\underline{x}_n))$. An estimate of two standard deviations of the average is also obtained based on these L samples. L is made large enough to make the two standard deviations width acceptable in each experiment. Within each constructed table in this paper, the numbers following the $\pm$ signs are estimates of two standard deviations of the Monte Carlo estimates. (See Table A.5 for computing program.)

To examine the accuracy of our Monte Carlo estimates, Table 1 compares the values of $R_n(p,T_{B(1)})$ with $p = 0.0(0.05)0.5$

and $n = 1,2$ obtained by (2.8), (2.9) and by Monte Carlo integrations. Table 2 explores the risk behavior of $R_n(p,T_{B(1)})$ for $n = 1,2,5,10,25,50$ and for $p = 0.0(0.05)0.5$. (also see Table A.3 for $R_n(p,T_{B(2)})$). It can be seen that $R_n(p,T_{B(1)})$ converges to $R(p)$ quite rapidly and has steady small sample size risk behavior. Values of $R_n(p,T_{B(1)})$ for $p > 0.5$ need not be computed because of the symmetry about 0.5.

Table 1. $R_n(p,T_{B(1)})$

	n = 1		n = 2		
p	Monte* Carlo	Numerical Computing	Monte* Carlo	Numerical Computing	R(p)
0.0	0.122±0.006	0.121	0.093±0.005	0.096	0
0.05	0.128±0.006	0.129	0.107±0.006	0.110	0.0405
0.10	0.137±0.005	0.137	0.125±0.006	0.123	0.0701
0.15	0.146±0.005	0.144	0.135±0.006	0.134	0.0934
0.20	0.151±0.005	0.150	0.147±0.005	0.144	0.1121
0.25	0.155±0.004	0.155	0.152±0.005	0.152	0.1270
0.30	0.160±0.003	0.159	0.159±0.004	0.159	0.1387
0.35	0.162±0.003	0.162	0.164±0.003	0.164	0.1476
0.40	0.165±0.002	0.165	0.168±0.002	0.168	0.1538
0.45	0.166±0.001	0.166	0.170±0.002	0.170	0.1574
0.50	0.166±0.001	0.166	0.170±0.002	0.171	0.1587

*200 replications for each estimate

Table 2. Risk behavior of $R_n(p,T_{B(1)})$

p	n=1*	n=2*	n=5**	n=10**	n=25**	n=50***	R(p)
0.00	0.121	0.096	0.059±0.005	0.033±0.003	0.015±0.001	0.006±0.001	0
0.05	0.129	0.110	0.077±0.005	0.062±0.003	0.048±0.001	0.044±0.001	0.0405
0.10	0.137	0.123	0.101±0.005	0.088±0.004	0.076±0.001	0.073±0.001	0.0701
0.15	0.144	0.134	0.120±0.005	0.105±0.002	0.100±0.001	0.098±0.002	0.0934
0.20	0.150	0.144	0.139±0.005	0.127±0.004	0.118±0.001	0.116±0.001	0.1121
0.25	0.155	0.152	0.146±0.004	0.141±0.003	0.134±0.001	0.130±0.001	0.1270
0.30	0.159	0.159	0.159±0.004	0.152±0.003	0.145±0.001	0.145±0.002	0.1387
0.35	0.162	0.164	0.161±0.003	0.160±0.003	0.156±0.002	0.153±0.002	0.1476
0.40	0.165	0.168	0.171±0.003	0.168±0.003	0.163±0.002	0.157±0.001	0.1538
0.45	0.166	0.170	0.173±0.003	0.171±0.003	0.166±0.002	0.160±0.001	0.1574
0.50	0.166	0.171	0.177±0.003	0.173±0.002	0.166±0.001	0.163±0.002	0.1587

* estimates obtained by numerical computation

** 200 replications for each estimate (Monte Carlo)

*** 50 replications for each estimate (Monte Carlo)

At this point it is important to note that in the Bayes empirical Bayes approach the presence of Λ does not restrict the construction of Bayes empirical Bayes procedures but adds the flexibility which enables one to access a family of decision procedures with predictable risk behavior. In particular, consider procedures T_Λ for $\Lambda \in \mathcal{B}$. While the mass of $B(1)$ is evenly distributed over $[0,1]$, $B(\gamma)$ puts more weight to those p values close to 0.5 as γ increases, and conversely, puts more weight to those p values close to 0 and 1 as γ decreases. From the fact that T_Λ is admissible and T_Λ is Bayes with respect to Λ, we expect that for $a < b$, $R_n(p,T_{B(a)}) > R_n(p,T_{B(b)})$ for p close to 0.5 and $R_n(p,T_{B(a)}) < R_n(p,T_{B(b)})$ for p close to 0 or 1. Table 3 shows the flexibility with choices among $B(\gamma)$; $\gamma = 0.25,1,2,3,10$ and gives values of $R_1(p,T_{B(\gamma)})$ for $p = 0.0(0.05)0.5$. The fact that $B(\gamma)$ has mean $\frac{1}{2}$ and variance $1/4(2\gamma + 1)$ implies that as $\gamma \to \infty$, $B(\gamma)$ converges weakly to the distribution degenerated in $p = \frac{1}{2}$, and hence $T_{B(\gamma)}$ converges to the minimax rule with constant risk .1587. This is also reflected in Table 3.

Section 2.4. Other empirical Bayes procedures

Robbins (1951) in his original example of the related compound decision problem uses the estimator

$$p_1(\underline{X}_n) = \max\{0,\min\{1,\ 0.5 + (\sum_{i=1}^{n} 0.5\ X_i)/n\}\} \qquad (2.10)$$

Table 3. The flexibility of $R_1(p,T_\Lambda)$ with a prior Λ in $\{B(\gamma)|\gamma > 0\}$

		parameter of $\Lambda = B(\gamma)$				
p	R(p)	γ=0.25*	γ=1**	γ=2**	γ=3**	γ=10**
0.00	0	0.093±0.005	0.122±0.006	0.135±0.003	0.140±0.002	0.152±0.001
0.05	0.0405	0.111±0.006	0.128±0.006	0.140±0.003	0.146±0.002	0.153±0.001
0.10	0.0701	0.128±0.005	0.137±0.005	0.145±0.003	0.149±0.002	0.155±0.001
0.15	0.0934	0.147±0.005	0.146±0.005	0.148±0.003	0.152±0.002	0.156±0.001
0.20	0.1121	0.159±0.005	0.151±0.005	0.151±0.002	0.153±0.002	0.157±0.001
0.25	0.1270	0.169±0.004	0.155±0.004	0.155±0.003	0.156±0.002	0.158±0.001
0.30	0.1387	0.181±0.004	0.160±0.003	0.157±0.002	0.157±0.001	0.158±0.001
0.35	0.1476	0.187±0.003	0.162±0.003	0.159±0.002	0.159±0.001	0.158±0.001
0.40	0.1538	0.190±0.002	0.165±0.002	0.160±0.001	0.159±0.001	0.159±0.001
0.45	0.1574	0.194±0.002	0.166±0.001	0.161±0.001	0.160±0.000	0.159±0.001
0.50	0.1587	0.196±0.002	0.166±0.001	0.161±0.000	0.160±0.000	0.159±0.001

* 800 replications for each estimate

** 200 replications for each estimate

and the corresponding decision procesure T_1 constructed by replacing p in (2.7) with $p_1(\underline{X}_n)$. Van Houwelingen (1974) modified Robbins' procedure by estimating p with an improved estimator

$$p_2(\underline{X}_n) = \max\{0, \min\{1, 0.5 + (\sum_{i=1}^{n} 0.908429 \tan h(X_i))/n\}\} \quad (2.11)$$

and constructing a decision procedure T_2 by replacing p in (2.7) with $p_2(\underline{X}_n)$. Both p_1, p_2 are consistent estimators of p, and consequently from (1.4) and (1.5), the corresponding decision procedures T_1 and T_2 are asymptotically optimal. Also as observed in Van Houwelingen (1974), the rate of convergence for both T_1 and T_2 is proportional to $(n)^{-1}$.

However, T_1 and T_2 are not Bayes empirical Bayes rules. To see this, note that from (2.2) it follows that the conditional mean $p_\Lambda(\underline{X}_n)$ is 0 if Λ is degenerate at $p = 0$, is 1 if Λ is degenerate at $p = 1$, and satisfies $0 < p_\Lambda(\underline{X}_n) < 1$ otherwise. Thus, apart from the trivial procedures $T(\underline{X}_n) \equiv t_0$ and $T(\underline{X}_n) \equiv t_1$, there are no Bayes empirical Bayes procedures taking values t_0 or t_1, with positive P_p^n probability. However, since $p_1(\underline{X}_n)$ and $p_2(\underline{X}_n)$ take on both 0 and 1 with positive P_p^n probability, T_1 and T_2 take on values t_0 and t_1 with positive P_p^n probability demonstrating that T_1 and T_2 are not Bayes empirical Bayes procedures. Moreover, the fact that $\mathcal{B}_n$ is complete implies that T_1, T_2 are not admissible empirical Bayes rules.

In fact, if $p(\underline{X}_n)$ is an unbiased estimator for $p = 0$ other than the estimator $p(\underline{X}_n) = 0$ a.s., then the corresponding decision rule $\hat{T}$ choses t_0 with a positive probability less than one, which implies that $\hat{T}$ is not Bayes and hence not admissible.

Before leaving this section we assess the accuracy of equation (28) of Van Houwelingen (1974) which gives an approximation to the risk functions of T_1 and T_2 for large n and $p \in (0,1)$. The approximation formula has a faster convergence rate for points p close to 0.5 than those points close to 0 or 1 and is not defined at $p = 0,1$. Table 4 compares the values of $R_{50}(p,T_2)$ estimated by (28) of Van Houwelingen (1974) with the Monte Carlo estimates. Note there is a significant difference for the two estimated values at $p = 0.05$ and agreement otherwise.

Since $p_\Lambda(\underline{X}_n)$ is known to be asymptotically equivalent to the M.L.E., using the asymptotic second moment of $p_\Lambda(\underline{X}_n)$ about p in (28) of Van Houwelingen (1974) provides an alternative estimate for the large n risk $R_n(p,T_{B(\gamma)})$. A numerical computation showed the fairly close agreement of the results with those reported in Table 2 for $n = 50$.

Section 2.5. Monte Carlo Comparisons of T_Λ, T_1 and T_2

T_1 and T_2 are neither Bayes nor admissible rules. At stage n, we were able to choose a Λ in B such that $R_n(p,T_\Lambda)$ behaves as a good competitor against $R_n(p,T_\alpha)$; $\alpha = 1,2$. In some cases that follow, the risk values of T_1 will not be listed

Table 4. $R_{50}(p,T_2)$

p	Van Houwelingen approximation	Monte Carlo*
0.0		0.004±0.001
0.05	0.050	0.046±0.001
0.10	0.077	0.078±0.001
0.15	0.099	0.101±0.002
0.20	0.117	0.119±0.002
0.25	0.132	0.133±0.001
0.30	0.144	0.144±0.001
0.35	0.152	0.153±0.001
0.40	0.158	0.158±0.001
0.45	0.162	0.162±0.001
0.50	0.163	0.163±0.001

* 200 replications for each estimate

if our results showed that the risk behaviors of T_1 and T_2 were very similar. (See Table A.1 and Table A.2 for complete data, Table A.7 for computing program.)

From Table 5 we see that the estimates of $R_1(p,T_{B(.25)})$ dominate those of $R_1(p,T_\alpha)$, $R_2(p,T_{B(.10)})$ dominates $R_2(p,T_\alpha)$, $R_5(p,T_{B(.15)})$ dominates $R_5(p,T_\alpha)$. From Table 6 the estimates of $R_{10}(p,T_{B(.35)})$ dominate those of $R_{10}(p,T_\alpha)$. The estimates of $R_{25}(p,T_{B(.37)})$ come within one standard deviation of the estimates for $R_{25}(p,T_2)$ when $p = 0.25(0.05)0.5$ and $p = 0.0$ but significantly less than $R_{25}(p,T_2)$ at $p = 0.05(0.05)0.2$. For $n = 50$, $R_{50}(p,T_{B(1)})$ dominates $R_{50}(p,T_\alpha)$ except at $p = 0.0$. The small difference may be adjusted by carefully choosing some $B(\gamma)$ with γ slightly less than 1. This will improve the risk function at $p = 0.0$ with a little sacrifice at $p = 0.5$.

Table 2 shows the rule $T_{B(1)}$ has good small sample performance. However, this is not true for T_1 and T_2 at $n = 1,2$. Table A.1, Table A.2 and Table 5 entries indicate that the Bayes empirical Bayes rule with respect to the uniform prior $T_{B(1)}$ has lower risk than Robbins and Van Houwelingen empirical Bayes rules T_1 and T_2 except near $p = 0$ (and by symmetry, near $p = 1$). Copas (1969), p. 413) reports a similar finding in regard to $T_{B(1)}$ and T_1.

It is interesting to note that the estimates of $R_1(p,T_1)$ dominate $R_1(p,T_2)$, but, estimates of $R_{50}(p,T_1)$ are dominated by $R_{50}(p,T_2)$. This means that small sample properties may not be guaranteed by a fast convergence rate and vice versa.

Table 5. Comparisons of risk behaviors for decision procedures

T_1, T_2, T_Λ; $\Lambda \varepsilon \{B(\gamma) \mid \gamma > 0\}$, when n = 1, 2, 5

		n = 1			n = 2		n = 3	
p	R(p)	$T_{B(.25)}$*	Van Houwe-@ lingen	Robbins@	$T_{B(.10)}$#	Van Houwe-% lingen	$T_{B(.15)}$*	Van Houwe-* lingen
0.00	0	0.093±0.005	0.111±0.008	0.093±0.006	0.046±0.005	0.050±0.006	0.016±0.002	0.021±0.003
0.05	0.0405	0.111±0.006	0.166±0.008	0.143±0.007	0.084±0.006	0.099±0.006	0.056±0.003	0.064±0.003
0.10	0.0701	0.128±0.005	0.215±0.008	0.185±0.007	0.116±0.006	0.142±0.006	0.091±0.004	0.101±0.003
0.15	0.0934	0.147±0.005	0.267±0.008	0.232±0.007	0.152±0.007	0.184±0.007	0.125±0.005	0.134±0.004
0.20	0.1121	0.159±0.005	0.302±0.007	0.265±0.007	0.174±0.006	0.219±0.007	0.152±0.004	0.162±0.005
0.25	0.1270	0.169±0.004	0.343±0.007	0.300±0.006	0.197±0.006	0.246±0.007	0.169±0.004	0.181±0.005
0.30	0.1387	0.181±0.004	0.372±0.006	0.324±0.006	0.210±0.005	0.271±0.007	0.188±0.004	0.202±0.006
0.35	0.1476	0.187±0.003	0.393±0.005	0.344±0.005	0.230±0.005	0.282±0.007	0.204±0.004	0.215±0.007
0.40	0.1538	0.190±0.002	0.409±0.005	0.358±0.005	0.233±0.005	0.302±0.007	0.219±0.005	0.222±0.007
0.45	0.1574	0.194±0.002	0.418±0.004	0.365±0.005	0.245±0.004	0.307±0.007	0.222±0.005	0.234±0.008
0.50	0.1587	0.196±0.002	0.422±0.004	0.369±0.005	0.246±0.004	0.310±0.007	0.222±0.005	0.232±0.007

* 800 replications for each estimate
\# 1000 replications for each estimate
% 2000 replications for each estimate
@ 4000 replications for each estimate

Table 6. Comparisons of risk behaviors for decision procedures

T_1, T_2, T_Λ; $\Lambda \varepsilon \{B(\gamma) \mid \gamma > 0\}$, when n = 10, 25, 50

		n = 10		n = 25		n = 50		
p	R(p)	$T_{B(.35)}$*	Van Houwe-* lingen	$T_{B(.37)}$*	Van Houwe-* lingen	$T_{B(1)}$@	Van Houwe-# lingen	Robbins#
0.00	0	0.014±0.002	0.015±0.003	0.006±0.001	0.006±0.001	0.006±0.001	0.004±0.001	0.005±0.001
0.05	0.0405	0.050±0.002	0.055±0.002	0.045±0.001	0.049±0.001	0.044±0.001	0.046±0.001	0.047±0.001
0.10	0.0701	0.081±0.003	0.090±0.002	0.076±0.001	0.082±0.001	0.073±0.001	0.078±0.001	0.078±0.001
0.15	0.0934	0.108±0.002	0.120±0.003	0.102±0.001	0.106±0.002	0.098±0.002	0.101±0.002	0.102±0.001
0.20	0.1121	0.136±0.003	0.143±0.004	0.122±0.-01	0.125±0.002	0.116±0.001	0.119±0.002	0.120±0.002
0.25	0.1270	0.149±0.004	0.162±0.005	0.138±0.001	0.138±0.002	0.130±0.001	0.133±0.001	0.134±0.002
0.30	0.1387	0.163±0.003	0.169±0.004	0.151±0.002	0.152±0.002	0.143±0.002	0.144±0.001	0.146±0.003
0.35	0.1476	0.171±0.003	0.183±0.006	0.159±0.002	0.158±0.002	0.153±0.002	0.153±0.001	0.153±0.002
0.40	0.1538	0.177±0.003	0.187±0.009	0.165±0.002	0.164±0.002	0.157±0.001	0.158±0.001	0.158±0.001
0.45	0.1574	0.185±0.004	0.188±0.006	0.168±0.001	0.169±0.002	0.160±0.001	0.162±0.001	0.162±0.001
0.50	0.1587	0.188±0.004	0.191±0.005	0.169±0.002	0.169±0.002	0.163±0.002	0.163±0.001	0.164±0.001

@ 50 replications for each estimate

200 replications for each estimate

* 400 replications for each estimate

APPENDICES

APPENDIX A

Table A.1. Risk behavior of $R_n(p,T_1)$.

p	n=1 (4,000)	n=2 (2,000)	n=3 (1,600)	n=5 (800)	n=10 (400)	n=25 (200)	n=50 (200)
0.00	0.0927±0.006	0.0496±0.005	0.0313±0.003	0.0223±0.003	0.0156±0.003	0.0072±0.002	0.0048±0.001
0.05	0.1431±0.007	0.0993±0.006	0.0797±0.004	0.0654±0.003	0.0556±0.002	0.0499±0.002	0.0471±0.001
0.10	0.1848±0.007	0.1370±0.006	0.1181±0.004	0.1020±0.003	0.0899±0.002	0.0834±0.002	0.0781±0.001
0.15	0.2323±0.007	0.1839±0.007	0.1592±0.006	0.1362±0.005	0.1215±0.003	0.1089±0.003	0.1025±0.002
0.20	0.2648±0.007	0.2147±0.007	0.1870±0.005	0.1645±0.005	0.1448±0.004	0.1263±0.003	0.1202±0.002
0.25	0.2998±0.006	0.2435±0.007	0.2182±0.006	0.1824±0.005	0.1625±0.004	0.1419±0.004	0.1336±0.002
0.30	0.3242±0.006	0.2668±0.007	0.2411±0.007	0.2079±0.006	0.1714±0.005	0.1507±0.003	0.1458±0.003
0.35	0.3442±0.005	0.2845±0.007	0.2570±0.007	0.2189±0.007	0.1853±0.006	0.1600±0.003	0.1534±0.002
0.40	0.3583±0.005	0.2967±0.007	0.2672±0.007	0.2238±0.007	0.1923±0.007	0.1664±0.003	0.1582±0.001
0.45	0.3654±0.005	0.3053±0.007	0.2699±0.007	0.2380±0.008	0.1920±0.006	0.1682±0.002	0.1621±0.001
0.50	0.3694±0.005	0.3087±0.007	0.2803±0.007	0.2367±0.008	0.1942±0.006	0.1701±0.003	0.1638±0.001

(.) number of replications for each estimate (Monte Carlo)

Table A.2. Risk behavior of $R_n(p,T_2)$.

p	n=1 (4,000)	n=2 (2,000)	n=3 (1,600)	n=5 (800)	n=10 (400)	n=25 (200)	n=50 (200)
0.00	0.1105±0.008	0.0504±0.006	0.0301±0.004	0.0207±0.003	0.0148±0.003	0.0063±0.002	0.0038±0.001
0.05	0.1664±0.008	0.0991±0.006	0.0791±0.005	0.0640±0.003	0.0551±0.002	0.0491±0.002	0.0463±0.001
0.10	0.2155±0.008	0.1416±0.006	0.1168±0.004	0.1010±0.003	0.0897±0.002	0.0829±0.002	0.0775±0.001
0.15	0.2675±0.008	0.1842±0.007	0.1568±0.005	0.1340±0.004	0.1203±0.003	0.1091±0.003	0.1014±0.002
0.20	0.3023±0.007	0.2195±0.007	0.1876±0.005	0.1624±0.005	0.1433±0.004	0.1260±0.003	0.1192±0.002
0.25	0.3428±0.007	0.2460±0.007	0.2142±0.006	0.1813±0.005	0.1624±0.005	0.1416±0.004	0.1326±0.001
0.30	0.3717±0.006	0.2708±0.007	0.2361±0.006	0.2020±0.006	0.1690±0.004	0.1496±0.003	0.1436±0.001
0.35	0.3925±0.005	0.2823±0.007	0.2524±0.006	0.2153±0.007	0.1826±0.006	0.1574±0.002	0.1526±0.001
0.40	0.4089±0.005	0.3024±0.007	0.2613±0.007	0.2220±0.007	0.1865±0.009	0.1651±0.003	0.1577±0.001
0.45	0.4177±0.004	0.3065±0.007	0.2632±0.007	0.2341±0.008	0.1877±0.006	0.1675±0.002	0.1616±0.001
0.50	0.4225±0.004	0.3104±0.007	0.2764±0.007	0.2320±0.007	0.1906±0.005	0.1686±0.002	0.1633±0.001

(.) number of replications for each estimate (Monte Carlo)

Table A.3. Risk behavior of $R_n(p,T_{B(2)})$.

p	n=1*	n=2*	n=3*	n=5*	n=10*	n=25*	n=50**
0.00	0.1347±0.003	0.1186±0.004	0.1030±0.004	0.0819±0.003	0.0546±0.003	0.0264±0.002	0.0132±0.002
0.05	0.1402±0.003	0.1251±0.004	0.1141±0.004	0.0979±0.004	0.0786±0.003	0.0565±0.002	0.0472±0.002
0.10	0.1449±0.003	0.1327±0.004	0.1188±0.003	0.1157±0.005	0.0959±0.003	0.0813±0.002	0.0768±0.002
0.15	0.1481±0.003	0.1395±0.003	0.1350±0.004	0.1232±0.003	0.1114±0.002	0.1015±0.001	0.0973±0.001
0.20	0.1508±0.002	0.1444±0.003	0.1448±0.003	0.1369±0.004	0.1262±0.003	0.1187±0.001	0.1158±0.002
0.25	0.1546±0.003	0.1519±0.003	0.1499±0.003	0.1440±0.003	0.1400±0.003	0.1346±0.001	0.1298±0.001
0.30	0.1568±0.002	0.1559±0.002	0.1549±0.003	0.1524±0.002	0.1479±0.002	0.1451±0.001	0.1436±0.002
0.35	0.1588±0.002	0.1590±0.002	0.1586±0.002	0.1587±0.002	0.1587±0.002	0.1532±0.001	0.1505±0.001
0.40	0.1602±0.001	0.1616±0.001	0.1617±0.001	0.1641±0.002	0.1624±0.002	0.1597±0.001	0.1574±0.002
0.45	0.1613±0.001	0.1623±0.001	0.1633±0.001	0.1655±0.001	0.1655±0.002	0.1636±0.001	0.1610±0.002
0.50	0.1613±0.000	0.1626±0.001	0.1647±0.001	0.1652±0.001	0.1666±0.002	0.1639±0.001	0.1612±0.001

* 200 replications for each estimate (Monte Carlo)
** 50 replications for each estimate (Monte Carlo)

Table A.4. Evaluation of the Bayes envelope R(p)

```
LIST,F.
100=       PROGRAM ENVLOP(OUTPUT)
105=       REAL P,CP,U,V,A,B,RP
110=       P=0.
120=       DO 2 I=1,50
130=       P=P+0.01
140=       CP=0.5*ALOG((1.-P)/P)
150=       U=CP+1.
160=       V=CP-1.
170=       CALL MDNOR(V,A)
180=       CALL MDNOR(U,B)
190=       RP=P*A+(1.-P)*(1.-B)
200=       PRINT 7,P,RP
210=7      FORMAT(3X,F5.2,3X,F10.6)
220=2      CONTINUE
230=       END
```

```
EXEC BEGUN.09.24.20.
   .01      .009511
   .02      .018136
   .03      .026090
   .04      .033503
   .05      .040459
   .06      .047018
   .07      .053226
   .08      .059118
   .09      .064722
   .10      .070061
   .11      .075155
   .12      .080019
   .13      .084670
   .14      .089117
   .15      .093373
   .16      .097446
   .17      .101345
   .18      .105077
   .19      .108649
   .20      .112067
   .21      .115336
   .22      .118461
   .23      .121447
   .24      .124298
   .25      .127017
   .26      .129608
   .27      .132074
   .28      .134417
   .29      .136642
   .30      .138749
   .31      .140741
   .32      .142620
   .33      .144388
   .34      .146047
   .35      .147598
   .36      .149042
   .37      .150382
   .38      .151618
   .39      .152751
   .40      .153783
   .41      .154714
   .42      .155545
   .43      .156276
   .44      .156909
   .45      .157443
   .46      .157880
   .47      .158219
   .48      .158462
   .49      .158607
   .50      .158655
  END  ENVLOP
```

Table A.5. Monte Carlo simulation of $R_n(p,T_\Lambda)$, $\Lambda \in B$

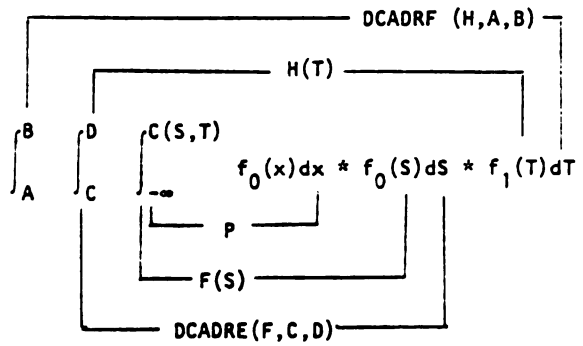
```
READY 22.11.33
OK.
OK-ATTACH,A,BAYES6.
 ATTACH,A,BAYES6.
OK-FTN,I=A,OPT=2.
 COMPILING BERISK
 COMPILING COEFICT
 COMPILING BETA
       .531 CP SECONDS COMPILATION TIME
OK-PROMPT.
OK-LISTTY,I=A,NS.
      PROGRAM BERISK(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
      DOUBLE A(100),B(100),C(100),M(100),X(100)
      DOUBLE DA,DB,DSEED,GE1,SUM1,SUM2
      REAL ACLIMIT,MEAN,P,P1,P2,PAR,R(1000),RISK1,RISK2,SD,Y1,Y2
      REAL RNUM,SSD
      INTEGER COUNT,NEXP,NUM
      COMMON A,B,C/MOMENT/M
      WRITE(6,100)
100   FORMAT(*1*,*THIS PROGRAM IS WRITTEN BY HOW JAN TSAO*)
C
      READ(5,400)S,T
400   FORMAT(F5.2,F5.2)
1000  IF(EOF(5).NE.0)STOP
C
      READ(5,500)P,NEXP,N,DSEED
C
500   FORMAT(F5.2,I4,I4,D25.18)
C
      WRITE(6,550)P,NEXP,N,DSEED
550   FORMAT(*0*,F5.2,3X,I4,3X,I4,D25.18)
C
      K=N+1
      CALL BETA(K,T,S)
C
C FOR EACH EXPERIMENT WE ASSIGN A UNIFORM(0,1) RANDOM VARIABLE
C
      CALL GGUBS(DSEED,NEXP,R)
C
      COUNT=0
C COUNT IS THE NUMBER OF ILLIGAL DATAS FOR GE1
C
      RISK1=0.0
      RISK2=0.0
C
      DO 1 L=1,NEXP
C
      DSEED=2147483647.D0*R(L)+1.
C
      DO 10 I=1,N
      PAR=GGBIR(DSEED,1,P)
C
C WE HAVE GENERATED A BERNOULLI(P) RANDOM VARIABLE
C
      X(I)=GGNOF(DSEED)+(2.*PAR-1.)
C
C     WE HAVE GENERATED A NORMAL(2*PAR-1.,1) RANDOM VARIABLE
C
10    CONTINUE
C
```

```
C NOW WE COMPUTE GE1,ACLIMIT
C
      DO 20 I=1,N
      DA=-0.5*(X(I)-1.)*(X(I)-1.)
      DB=-0.5*(X(I)+1.)*(X(I)+1.)
      B(I)=DEXP(DB)
      A(I)=DEXP(DA)-B(I)
20    CONTINUE
C
      CALL COEFICT(N)
C
      SUM1=C(1)*M(1)
      SUM2=C(1)
C
      DO 30 I=2,K
      SUM1=SUM1+C(I)*M(I)
      SUM2=SUM2+C(I)*M(I-1)
30    CONTINUE
C
      GE1=SUM1/SUM2
C
C NOW WE SCREEN OUT ALL ILLIGAL DATA
C
      IF((GE1.LE.0.D0).OR.(GE1.GE.1.D0))GO TO 4321
C
      ACLIMIT=0.5D0*DLOG((1.D0-GE1)/GE1)
C
C
C  NOW WE COMPUTE CONDITIONAL BAYES RISK GIVEN X(1),...,X(N)
C
      Y1=ACLIMIT-1.
      Y2=ACLIMIT+1.
      CALL MDNOR(Y1,P1)
      CALL MDNOR(Y2,P2)
      RISK=P*P1+(1.-P)*(1.-P2)
      RISK1=RISK+RISK1
      RISK2=RISK*RISK+RISK2
C
      GO TO 1
C
4321  WRITE(6,650)(I,A(I),B(I),C(I),I=1,N)
650   FORMAT(* *,4X,I4,3D25.18)
C
      COUNT=COUNT+1
C
1     CONTINUE
C
      NUM=NEXP-COUNT
      MEAN=RISK1/NUM
      SD=SQRT((RISK2-NUM*MEAN*MEAN)/(NUM-1.))
      RNUM=NUM
      SSD=2.*SD/SQRT(RNUM)
      WRITE(6,700)P,N,NEXP,MEAN,SD,SSD
700   FORMAT(*0*,*P= *,F5.2,*  N=*,I4,*  NEXP=*,I4,
     + *  RISK=*,F10.5,*  SD=*,F10.5,*  SSD= *,F5.3)
      GO TO 1000
      END
C
```

```
      SUBROUTINE COEFICT(N)
      DOUBLE A(100),B(100),C(100),D(100)
      COMMON A,B,C
      C(1)=B(1)
      C(2)=A(1)
C
      IF(N.EQ.1)GO TO 5
      DO 10 I=2,N
      D(1)=B(I)*C(1)
      DO 20 J=2,I
      D(J)=A(I)*C(J-1)+B(I)*C(J)
      C(J-1)=D(J-1)
20    CONTINUE
      D(I+1)=A(I)*C(I)
      C(I)=D(I)
      C(I+1)=D(I+1)
10    CONTINUE
5     RETURN
      END
C
C
      SUBROUTINE BETA(K,T,S)
      REAL T,S
      DOUBLE M(100),PROD1,PROD2
      COMMON /MOMENT/M
C
C THIS SUBROUTINE GENERATES 1 THRU K TH MOMENTS OF
C BETA(T,S)
C
      PROD1=1.D0
      PROD2=1.D0
      DO 10 I=1,K
      PROD1=PROD1*(T+(I-1.))
      PROD2=PROD2*(T+S+(I-1.))
      M(I)=PROD1/PROD2
10    CONTINUE
      RETURN
      END
*EOR00
*EOI
OK-HAL.
 HAL 5.38
L?LGO.
 EXEC BEGUN.22.16.42.


THIS PROGRAM IS WRITTEN BY HOW JAN TSAO
*0.5  0.5
*0.20   10   5 13524.D0

  .20     10      5  .13524000000000000D+05

P=   .20  N=   5  NEXP=  10  RISK=    .13207  SD=    .03410  SSD=  .022
*0.20  400   5 14326.D0

  .20    400      5  .14326000000000000D+05

P=   .20  N=   5  NEXP= 400  RISK=    .13464  SD=    .03972  SSD=  .004
*  THIS ROUTINE CONTINUES UNTIL USER ABORT.
\\\\                                        !
```

Table A.6. A numerical computation program.

This program evaluates:



where

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+1)^2}{2}}$$

```
READY 12.22.37
ATTACH,A,ROMB8.
 ATTACH,A,ROMB8.

READY 12.22.49
LISTTY,I=A,NS.
      PROGRAM ROMB3(OUTPUT)
      INTEGER IER
      REAL DCADRF,H,F0,F1,A,B,AERR,RERR,ERROR,INTEG
      EXTERNAL H
      A=-3.11
      B=5.11
      RERR=0.
      AERR=1.E-5
      INTEG=DCADRF(H,A,B,AERR,RERR,ERROR,IER)
      PRINT 7,INTEG,ERROR,IER
 7    FORMAT(1X,F17.15,3X,F10.8,3X,I3)
      END
C
C
      REAL FUNCTION H(T)
      INTEGER IER
      REAL DCADRE,F,F0,F1,C,D,AERR,RERR,ERROR,INTEG,Z
      EXTERNAL F
      COMMON /JOINT/Z
      Z=T
      C=-5.11
      D=3.11
      RERR=0.
      AERR=1.E-5
      H=DCADRE(F,C,D,AERR,RERR,ERROR,IER)*F1(T)
      RETURN
      END
C
C
      REAL FUNCTION F(S)
      DOUBLE C
      REAL S,P,Y,Z,T,F0
      COMMON /JOINT/Z
      T=Z
      Y=C(S,T)+1.
      CALL MDNOR(Y,P)
      F=P*F0(S)
      RETURN
      END
```

IMSL subroutines used:

1. MDNOR
2. DCADRE
3. DCADRF: a binary copy of DCADRE.

```
C
      DOUBLE FUNCTION C(S,T)
      DOUBLE X1,X2,DA,DB,G
      REAL S,T
      X1=S
      X2=T
      DA=1.+DEXP(2.*X2)+DEXP(2.*X1)+3.*DEXP(2.*(X1+X2))
      DB=4.+2.*DEXP(2.*X2)+2.*DEXP(2.*X1)+4.*DEXP(2.*(X1+X2))
      G=DA/DB
      C=0.5*DLOG((1.-G)/G)
      RETURN
      END
C
C
      REAL FUNCTION F0(X)
      DOUBLE Y,PI
      REAL X
      PI=3.14159265358979323846264338D0
      Y=X
      F0=DEXP(-0.5*(Y+1.)*(Y+1.))/DSQRT(2.*PI)
      RETURN
      END
C
C
      REAL FUNCTION F1(X)
      DOUBLE Y,PI
      REAL X
      PI=3.14159265358979323846264338D0
      Y=X
      F1=DEXP(-0.5*(Y-1.)*(Y-1.))/DSQRT(2.*PI)
      RETURN
      END
*EOR00
*EOI

READY 12.24.10
RETURN,DCADRF.

READY 12.25.49
REWIND,A.

READY 12.26.04
ATTACH,DCADRF,CRSDCADRF.
 ATTACH,DCADRF,CRSDCADRF.

READY 12.26.25
FTN,I=A.
 COMPILING ROMB3
 COMPILING H
 COMPILING F
 COMPILING C
 COMPILING F0
 COMPILING F1
      .220 CP SECONDS COMPILATION TIME

READY 12.26.48
HAL.
 HAL 5.37
L?LOAD,DCADRF.
L?LOAD,LGO.
L?EXECUTE.
 EXEC BEGUN.12.27.39.
 .835138858743978     .00000214      0
      END  ROMB3
      1.012 CP SECONDS EXECUTION TIME
```

Table A.7. Monte Carlo simulation of $R_n(p,T_\alpha)$, α = 1,2

```
READY 12.48.39
LIST,F,NS.
      PROGRAM RISK(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
      DOUBLE DSEED
      REAL R(6000),X(100)
C
C MONTE CARLO SIMULATION OF MIXED NORMAL RANDOM VARIABLES
C FOR TESTING N(1,1) VS N(-1,1)
C NEXP REPLICATIONS OF SAMPLES WITH SIZE N IS GENERATED
C TO ESTIMATE RISK BEHAVIORS OF (1) ROBBIN'S DECISION PROCEDURE
C  AND (2) VAN HOWELINGEN'S DECISION PROCEDURE126=C
      WRITE(6,50)
50    FORMAT(*0*,*DATA-*)
C
      READ(5,100)P,NEXP,N,DSEED
100   FORMAT(F5.2,I4,I4,D25.18)
      WRITE(6,200)P,NEXP,N,DSEED
200   FORMAT(*0*,F5.2,3X,I4,3X,I4,3X,D25.18)
C
      CALL GGUBS(DSEED,NEXP,R)
      RB1=0.0
      RB2=0.0
      RVH1=0.0
      RVH2=0.0
C
C
      DO 1 L=1,NEXP
      DSEED=2147483647.*R(L)+1.
      SUM1=0.0
      SUM2=0.0
C
      DO 10 I=1,N
      PAR=GGBIR(DSEED,1,P)
      X(I)=GGNQF(DSEED)+(2.*PAR-1.)
      SUM1=X(I)+SUM1
      SUM2=TANH(X(I))+SUM2
10    CONTINUE
C
      PRB=0.5+SUM1/(2.*N)
      PVH=0.5+0.90842942*SUM2/N
C
      IF(PRB.GE.1.)GO TO 350
      IF(PRB.LE.0.)GO TO 352
      CRB=0.5*ALOG((1.-PRB)/PRB)
      YRB1=CRB-1.
      YRB2=CRB+1.
      CALL MDNOR(YRB1,PRB1)
      CALL MDNOR(YRB2,PRB2)
      RB=P*PRB1+(1.-P)*(1.-PRB2)
      GO TO 356
350   RB=1.-P
      GO TO 356
352   RB=P
```

```
C
356     IF(PVH.GE.1.)GO TO 380
        IF(PVH.LE.0.)GO TO 382
        CVH=0.5*ALOG((1.-PVH)/PVH)
        YVH1=CVH-1.
        YVH2=CVH+1.
        CALL MDNOR(YVH1,PVH1)
        CALL MDNOR(YVH2,PVH2)
        PV=P*PVH1+(1.-P)*(1.-PVH2)
        GO TO 460
380     PV=1.-P
381     GO TO 460
382     PV=P
C
460     RB1=RB1+RB
        RB2=RB2+RB*RB
        RVH1=RVH1+PV
        RVH2=RVH2+PV*PV
1       CONTINUE
C
C
        RNEXP=NEXP
        SMRB=RB1/RNEXP
        SDRB=SQRT((RB2-RNEXP*SMRB*SMRB)/(RNEXP-1.))
        SSDRB=2*SDRB/SQRT(RNEXP)
        SMVH=RVH1/RNEXP
        SDVH=SQRT((RVH2-RNEXP*SMVH*SMVH)/(RNEXP-1.))
        SSDVH=2*SDVH/SQRT(RNEXP)
        WRITE(6,400)SMRB,SDRB,SSDRB,SMVH,SDVH,SSDVH
400     FORMAT(*0*,*ROBBIN *,F10.5,F10.5,F5.3,* V HOU *,2F10.5,F5.3)
        END

READY 12.50.09
FTN.
 COMPILING RISK
       .159 CP SECONDS COMPILATION TIME

READY 12.50.37
HAL.
 HAL 5.37
L?LGO.
 EXEC BEGUN.12.51.38.

DATA-0.3   200   5 26138.D0

  .30    200      5     .2613800000000000000D+05

ROBBIN     .19827    .08830 .012 V HOU     .19715    .09019 .013
      END  RISK
       .193 CP SECONDS EXECUTION TIME
```

APPENDIX B

APPENDIX B

In this appendix we prove Theorem 2.1 and 2.2. The notation and the following assumptions are from Johnson (1970). The model assumes $X_1, X_2, \ldots,$ i.i.d P_θ where P_θ has density $f(x,\theta)$ with respect to a given σ-finite measure μ.

B.1 The parameter space Θ is a compact subset of E_1. Let Θ^0 denote its interior and $\underline{\Theta}$ denote the Borel σ-algebra on Θ.

B.2 θ is identified by P_θ.

B.3 $f(x,\theta)$ is jointly measurable in (x,θ).

B.4 For each x, $f(x,\theta)$ admits continuous first and second partial derivatives with respect to θ.

B.5 The measures P_θ are mutually absolutely continuous.

B.6 If $\lim_{i\to\infty} |\theta_i| = \infty$, then $\lim_{i\to\infty} f(x,\theta_i) = 0$ for all x except for perhaps a null set depending on the sequence.

B.7 For all $\theta \in \Theta$, $E_\theta |\log f(X,\theta)| < \infty$ and

$$0 < I(\theta) = -E_\theta \left[\frac{\partial^2}{\partial\theta^2} \log f(X,\theta)\right]$$

B.8 For each $\theta_0 \in \Theta^0$, there exist functions $G_1(X)$ and $G_2(X)$ satisfying

$$\left|\frac{\partial}{\partial\theta} \log f(X,\theta)\right| \leq G_1(X), \quad \left|\frac{\partial^2}{\partial\theta^2} \log f(x,\theta)\right| \leq G_2(X)$$

for θ in a neighborhood of θ_0 and also $E_{\theta_0}[G_1(X)] < \infty$ and $E_{\theta_0}[G_2(X)] < \infty$. The functions G_1 and G_2 may depend on θ_0.

B.9 Let $f(x,\theta,\rho) = \sup_{|\theta-\theta'|\leq\rho} f(x,\theta')$, $\rho > 0$

and $Q(x,\gamma) = \sup_{|\theta|>\gamma} f(x,\theta)$, $\gamma > 0$.

For every $\theta \in \Theta$ and $\rho, \gamma > 0$, $f(x,\theta,\rho)$ and $Q(x,\gamma)$ are measurable functions of x. Moreover, for sufficiently small ρ and sufficiently large γ,

$$E_{\theta_0}[\log f(x,\theta,\rho)]^+ < \infty$$

$$E_{\theta_0}[\log Q(x,\gamma)]^+ < \infty \qquad \text{for each } \theta_0 \in \Theta^0$$

B.10 For each x, $\log f(x,\theta)$ had 5 continuous partial derivatives with respect to $\theta \in \Theta$.

B.11 There exists functions $G_k(x)$ with $E_{\theta_0}[G_k(X)] < \infty$ and

$$\left|\frac{\partial^k}{\partial\theta^k} \log f(x,\theta)\right| \leq G_k(x) \quad \text{for } \theta \text{ in a neighborhood of } \theta_0 \in \Theta,\ k = 3,4,5.$$

B.12 Λ is a probability measure on $(\Theta, \underline{\Theta})$, Λ has density λ with respect to the Lebesgue measure. For $\theta_0 \in \Theta^0$, $\lambda(\theta_0) > 0$ and $\lambda(\cdot)$ has 3 continuous derivatives in a neighborhood of θ_0.

B.13 $\int_\Theta |\theta|\lambda(\theta)d\theta < \infty$.

Conditions B.1 B.9 are basically those assumed by Wald (1949) to establish the strong consistency of the M.L.E. and those of LeCam (1956) to show that the M.L.E. is asymptotically normal. A weakened one-dimensional verison of LeCam's (1956) Theorem 3.4.1 is

Theorem B.II.1 Let B.1 ~ B.4, B.7, B.8 be satisfied. Then the maximum likelihood estimator $\hat{\theta}_n$ is strongly consistent and asymptotically normally distributed. The variance of the

limiting distribution of

$$\sqrt{n}\ (\hat{\theta}_n(\underline{X}_n) - \theta_0) \quad \text{is} \quad 1/[I(\theta_0)].$$

The following Theorem is a specialization of Theorem 3.1 of Johnson (1970) to $\gamma = 1$ and $k = 2$.

Let

$$b(\theta) = [-\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta^2} \log f(X_i,\theta)]^{\frac{1}{2}}$$

and

$$a_{3n}(\theta) = n^{-1} \sum_{i=1}^{n} \frac{\partial^3}{\partial\theta^3} \log f(X_i,\theta)/6, \quad \theta \in \Theta.$$

Theorem B.II.2 Under the assumptions B.1 ~ B.13, there exists a constant C such that for sufficiently large n depending on $\underset{\sim}{x} = (x_1,x_2,\ldots)$ belonging to a set of probability one,

$$|E_\Lambda(\theta|\underline{x}_n) - \hat{\theta}_n - b^{-1}(6a_{3n}(\hat{\theta}_n) + \lambda'(\hat{\theta}_n)/\lambda(\hat{\theta}_n))n^{-1}| \leq b^{-1}Cn^{-\frac{3}{2}} \qquad (1)$$

where we have abbreviated $b(\hat{\theta}_n(\underline{x}_n))$, $\hat{\theta}_n(\underline{x}_n)$ by b and $\hat{\theta}_n$.

Proof: See Theorem 3.1 and (3.4) of Johnson (1970). □

For the proofs of Theorem 2.1 and Theorem 2.2 we apply the above results with $f(x,\theta) = \theta f_1(x) + (1-\theta) f_0(x)$, $\Theta = [0,1]$, $\theta = p$.

Proof of Theorem 2.1. It is sufficient to show that the hypothesis of Theorem 2.1 implies that of Theorem B.II.1. Clearly B.1 ~ B.4 are satisfied. (Recall that identifiability B.2 is a tacit assumption in our empirical Bayes problem and is implied by P_0 and P_1 being different measures in the two state case.)

Since logarithm is strictly increasing on $(0,\infty]$, for $\theta \in [0,1]$ and for almost all x

$$|\log f(x,\theta)| \leqq \max\{|\log f_1(x)|, |\log f_0(x)|\}. \qquad (2)$$

Also,

$$I(\theta) = -\int \frac{\partial^2}{\partial\theta^2} \log f(x,\theta) P_\theta(dx) = \int \frac{(f_1(x) - f_0(x))^2}{f(x,\theta)^2} P_\theta(dx) > 0,$$

for $\theta \in \Theta$, since $\mu[f_1 \neq f_0] > 0$. Hence, B.7 is satisfied.

For $\theta_0 \in (0,1)$ pick $\varepsilon > 0$ such that $\varepsilon < \min\{\theta_0, 1-\theta_0\}$. Then for each $\theta \in (\theta_0-\varepsilon, \theta_0 + \varepsilon)$ and for almost all x

$$\left|\frac{\partial^k}{\partial\theta^k} \log f(x,\theta)\right| [(k-1)!]^{-1}$$

$$= 1/[\theta + f_0(x)/(f_1(x)-f_0(x))]^k \leqq \theta^{-k} < (\theta_0-\varepsilon)^{-k} \quad \text{if} \quad f_1(x) > f_0(x)$$

$$= \left|\frac{f_1(x)}{f_0(x)} - 1\right|^k / [\theta(f_1(x)/f_0(x)) + (1-\theta)]^k \leqq (1-\theta)^{-k} < [1-(\theta_0 + \varepsilon)]^{-k}$$

$$\text{if} \quad f_1(x) < f_0(x).$$

$$= 0 \quad \text{if} \quad f_0(x) = f_1(x) > 0, \qquad k = 1,2,3,\ldots \qquad (3)$$

Hence B.8 is satisfied.

Proof of Theorem 2.2 We first show that the hypothesis of Theorem 2.2 implies that of Theorem B.II.2.It is clear that B.5, B.6, B.12, B.13 are satisfied; also, from the proof of Theorem 2.1, we see that B.1 ~ B.4, B.7, B.8 are satisfied.

Observe that

$$f(x,\theta,\rho) = f(x,(\theta-\rho)\vee 0)I[f_0 \geqq f_1] + f(x,(\theta+\rho)\wedge 1)I[f_0 < f_1] \leqq \max\{f_0(x),f_1(x)\}$$

and

$$Q(x,\gamma) = f(x,0)I[f_0 \geqq f_1] + f(x,1)I[f_0 < f_1] \leqq \max\{f_0(x),f_1(x)\} \quad (4)$$

both are log-integrable by (2).

We see that (3) implies B.10, (4) implies B.9. Also, (3) implies that B.11 is satisfied as well.

Next we show that the result of Theorem B.II.2 leads to

$$\sqrt{n}\,(E_\Lambda(\theta|\underline{X}_n) - \hat{\theta}_n) \to 0 \qquad \text{a.s.} \quad P^\infty_{\theta_0} \qquad (5)$$

By Theorem B.II.1, $\hat{\theta}_n \to \theta_0$ a.s. $P^\infty_{\theta_0}$, θ_0 is the true parameter, $\theta_0 \in \Theta^0$.

Let

$$B(\theta|\theta_0) = \int -\frac{\partial^2}{\partial\theta^2} \log f(x,\theta) P_{\theta_0}(dx)$$

and

$$C(\theta|\theta_0) = \int \frac{\partial^3}{\partial\theta^3} \log f(x,\theta) P_{\theta_0}(dx).$$

Together B.8, B.11 and the uniform strong law (Rubin (1956)) implies that

$$b^2(\hat{\theta}_n) \to I(\theta_0) > 0; \quad a_{3n}(\hat{\theta}_n) \to \frac{1}{6} C(\theta_0|\theta_0) \quad \text{a.s.} \quad P^\infty_{\theta_0}$$

Also, B.12 implies that

$$\lambda(\hat{\theta}_n) \to \lambda(\theta_0) > 0 \quad \text{and} \quad \lambda'(\hat{\theta}_n) \to \lambda'(\theta_0) \quad \text{a.s.} \quad P^{\infty}_{\theta_0}.$$

Therefore, for some $M > 0$ and for almost all $\underline{x}$,

$$|b^{-1}(6a_{3n}(\hat{\theta}_n) + \lambda'(\hat{\theta}_n)/\lambda(\hat{\theta}_n))| + Cb^{-1} \leq M$$

for large n so that (1) implies (5). □

BIBLOGRAPHY

BIBLIOGRAPHY

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

Behboodian, Javad (1972). Bayesian estimation for the proportions in a mixture of distributions. Sankhya, Series B. 34, 15-22.

Billingsley, Patrick (1968). Convergence of probability measures. John Wiley and Sons Inc.

Boyer, John E., Jr. and Gilliland, Dennis, C. (1980). Admissibility consideration in the finite state compound and empirical decision problems. Statistica Neerlandica, 34.

Copas, J.B. (1969). Compound decisions and empirical Bayes. JRSS Series B, 31, 397-425.

Feller, William (1957). An introduction to probability theory and its applications, Vol. I., 2nd edition, John Wiley & Sons, Inc., New York.

Ferguson, Thomas, S. (1967). Mathematical statistics a decision theoretic approach. Academic Press.

Gilliland, Dennis, C., Hannan, James and Huang, J.S. (1974). Asymptotic solutions to the two state component compound decision problem, Bayes versus diffuse priors on proportions. RM-320, Statistics and Probability, MSU. (1976). Ann. Statist. 4, 1101-1112.

Gilliland, Dennis, C. and Boyer, John E., Jr. (1979). Bayes empirical Bayes. Submitted for publication.

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

Hannan, James, F. and Van Ryzin, J.R. (1965). Rate of convergence in the compound decision problem for two completely specified distributions. Ann. Math. Statist. 36, 1743-1752.

Huang, J.S. (1970). A note on Robbins' compound decision problem. RM-266, Statistics and Probability, MSU. (1972). Ann. Math. Statist. 43, 348-350

IMSL Library, 1 and 2, 7th ed. Int. Math. Stat. Libraries, Houston, TX, January 1979.

Johnson, R.A. (1970). Asymptotic expensions associated with posterior distributions. Ann. Math. Statist. 41, 851-864.

LeCam, Lucien (1956). Lecture Notes. Department of Statistics, University of California at Berkeley.

Lehmann, E.L. (1959). Testing Statistical Hypothesis. Wiley, New York.

Munkres, James, R. (1975). Topology. Prentice-Hall, Inc.

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

Robbins, Herbert (1951). Asymptotically subminimax solutions of compound statistical decision problems. Proc. Second Berkeley Symp. Math. Statist. Prob., 157-163. University of California Press.

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

Rockafellar, R. Tyrrell (1972). Convex Analysis. Princeton University Press.

Rubin, H. (1956). Uniform convergence of random functions with applications to statistics. Ann. Math. Statist. 27, 200-203.

Shapiro, Connie P. (1972). Bayesian classification. Ph.D. dissertation, Department of Statistics, University of Michigan.

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

Snijders, Tom (1977). Complete class theorems for the simplest empirical Bayes decision problems. Ann. Statist. 5, 164-171.

Van Houwelingen, J.C. (1974). An empirical Bayes rule for testing simple hypothesis versus simple alternative. Statistica Neerlandica 28, 209-221.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. Ann. Math. Statist. 20, 595-601.