113
239
THS

21401557

MICHIGAN STATE UNIVERSITY LIBRARIES

3 1293 00548 0292

LIBRARY
Michigan State
University

This is to certify that the

dissertation entitled

Parametric Empirical Bayes Problems
with
Cost For Component Observations

presented by

Inna Jung

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

Major professor

Date November 11, 1988

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

MSU LIBRARIES

RETURNING MATERIALS:
Place in book drop to remove this checkout from your record. FINES will be charged if book is returned after the date stamped below.

PARAMETRIC EMPIRICAL BAYES PROBLEMS
WITH
COST FOR COMPONENT OBSERVATIONS

By

Inna Jung

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

1988

5583391

ABSTRACT

PARAMETRIC EMPIRICAL BAYES PROBLEMS WITH COST FOR COMPONENT OBSERVATIONS

By

Inna Jung

We consider the empirical Bayes decision problem where the component problem includes a constant cost per observation and the option to choose in advance the total number of observations. The usual empirical Bayes decision problem involves identical components with a given fixed sample size for all repetitions of the component. The empirical Bayes decision approach with our component permits data accumulated over past component problems to be used in selecting both the sample size and the decision rule to be used in the current component problem. The generality introduced by allowing sample sizes that are determined stochastically makes the result more useful in applications where, typically, the choice of sample size is an option based on past data.

The empirical Bayes version involves "independent" repetitions (a sequence) of the component decision problem. With the varying sample size possible, these are not identical components. However, we impose the usual assumption that the parameter sequence $\underline{\theta} = (\theta_1, \theta_2, \ldots)$ consists of independent G–distributed parameters where G is unknown. We assume that $G \in \mathscr{G}$, a known family of distributions. The sample size N_i and the decision rule d_i for component i of

the sequence are determined in an evolutionary way. The sample size N_1 and the decision rule $d_1 \in D_{N_1}$ used in the first component are fixed and chosen in advance. The sample size N_2 and the decision rule d_2 are functions of $\underline{X}^1 = (X_{11},\ldots,X_{1N_1})$, the observations in the first component. The sample size N_3 and decision rule d_3 are functions of $(\underline{X}^1, \underline{X}^2)$. In general, N_i is an integer–valued function of $(\underline{X}^1, \underline{X}^2,\ldots,\underline{X}^{i-1})$ and, given N_i, d_i is a D_{N_i}–valued function of $(\underline{X}^1, \underline{X}^2,\ldots,\underline{X}^{i-1})$. (The action chosen in the i–th component is $d_i(\underline{X}^i)$ which hides the display of dependence on $(\underline{X}^1, \underline{X}^2,\ldots,\underline{X}^{i-1})$.) For a variety of models, we will construct empirical Bayes rules that are asymptotically optimal.

We consider both parametric models involving squared error loss estimation and linear loss testing and show how more general cost functions are covered by the work. We will simulate one model to assess the small–to–moderate i risk plus cost behavior of one of the suggested asymptotically optimal empirical Bayes procedures.

To my wife Chairan

and

Sons Sehyun, Sunggon

ACKNOWLEDGEMENTS

I wish to express my deepest appreciation to Professor Dennis C. Gilliland for his guidance throughout the preparation of this dissertation and his concern for my work. I would like to thank Professor R. V. Erickson, Professor R. V. Ramamoorthi and Professor H. Salehi for serving on my committee.

I am also thankful to Ms. JoAnn Peterson and Ms. Cathy Sparks who have helped me in many ways. My special thanks go to Ms. Loretta Ferguson for her patience and great care in typing my thesis.

TABLE OF CONTENTS

CHAPTER

1. INTRODUCTION. 1
   1.1. A Statistical Decision Problem With Cost for Observation. 1
   1.2 An Empirical Bayes Decision Problem with Random Sample Size Components. 3
   1.3 Literature Review. 8

2. ESTIMATION OF BINOMIAL PARAMETER. 11
   2.1. The Component Problem. 11
   2.2. An Empirical Bayes Decision Procedure. 16
   2.3. Some Empirical Bayes Risk Calculations. 19

3. TESTING THE BINOMIAL PARAMETER. 22
   3.1 The Component Problem. 22
   3.2 An Empirical Bayes Decision Procedure. 26

4. ESTIMATION OF THE NORMAL MEAN. 28
   4.1. The Component Problem. 28
   4.2 An Empirical Bayes Decision Procedure. 31

5. TESTING THE NORMAL MEAN. 35
   5.1. The Component Problem. 35
   5.2 An Empirical Bayes Decision Procedure. 38

REFERENCES. 41

LIST OF TABLES

TABLE

1.1. Empirical Bayes Procedure with Stochastically Determined Sample Sizes. 5

2.1. $n^*(\alpha, \beta)$ and $r(\alpha, \beta)$. 20

2.2. Estimated Empirical Bayes Risks ($m=2$, $c=.001$, $\alpha=\beta$). 21

LIST OF FIGURES

Figure

2.1. A Risk Envelope. 19

CHAPTER 1

INTRODUCTION

§ 1.1. A Statistical Decision Problem With Cost for Observations

Consider a statistical decision problem with parameter space Θ, action space $\mathscr{A}$, nonnegative loss function $L(\cdot,\cdot)$ on $\Theta \times \mathscr{A}$, unknown prior distribution G on Θ and a cost $c > 0$ per observation. Let $X_1, X_2, \ldots$ be observations which are independently and identically distributed with a distribution P_θ given θ, taking values in a set $\mathscr{X}$, the observation space. Let D_n be the set of all decision functions $d: \mathscr{X}^n \to \mathscr{A}$ where $\mathscr{X}^n$ is the observation space for the vector $\underline{X} = (X_1, \ldots, X_n)$. When θ is the parameter and a decision rule $d \in D_n$ is used, the decision loss plus cost for observing $\underline{X} = (X_1, \ldots, X_n)$ is

$$L(\theta, d(\underline{X})) + cn$$

where we assume that L is integrable for all θ, n and $d \in D_n$.

Let R_n denote the risk and Bayes risk of the decision rule $d \in D_n$, i.e.,

$$R_n(\theta, d) = \int_{\mathscr{X}^n} L(\theta, d(\underline{x}))dP^n_\theta(\underline{x}) \tag{1.1}$$

$$R_n(G, d) = \int R_n(\theta, d)\, dG(\theta) \tag{1.2}$$

and let r_n denote the risk and Bayes risk of the decision rule $d \in D_n$ including cost for observations.

Then

$$r_n(\theta, d) = R_n(\theta, d) + cn \tag{1.3}$$

$$r_n(G, d) = R_n(G, d) + cn. \tag{1.4}$$

We define minimum Bayes risk and minimum Bayes risk plus cost in the usual way. We assume for each prior G and each $n = 1,2,\ldots$ that a Bayes rule $d^n_G \in D_n$ exists. Thus,

$$\inf_{d \in D_n} R_n(G, d) = R_n(G, d^n_G). \tag{1.5}$$

Let

$$R_n(G) = R_n(G, d_G^n)$$

and

(1.6) $$r_n(G) = R_n(G) + cn.$$

Since $R_n(G)$ is nonincreasing in n, a minimizer of $r_n(G)$ exists among the integers 1, 2,.... We will denote a specified minimizer as $n^* = n^*(G)$ and refer to it as an optimal fixed sample size. Therefore, $r(G) = r_{n^*}(G)$ is the minimum Bayes risk in the component across all the possible sample sizes and the corresponding class of decision rules, i.e.,

(1.7) $$r(G)=r_{n^*}(G) = \min\{\min\{r_n(G, d) \,|\, d \in D_n\} \,|\, n = 1,2,....\}.$$

Moreover, note that $R_{n^*}(G) + cn^* \le R_1(G) + c < \infty$ so that $n^* \le (R_1(G)+c)/c < \infty$. For some components, $R_1(G)$ is a bounded function of $G \in \mathcal{G}$.

<u>Example 1.1</u>. (Estimation). Let $X_1, X_2,....,X_n$ be i.i.d. $N(\theta, A)$ given θ and let θ have prior distribution $G = N(\mu, V)$. Assume A is known. Let $\Theta = \mathcal{A} = (-\infty, \infty)$, $L(\theta, a) = (\theta - a)^2$ for $(\theta, a) \in \Theta \times \mathcal{A}$, and let $c > 0$ denote the constant cost per observation. Then a Bayes decision function for estimating θ based on observation $\underline{X}_n = (X_1,....,X_n)$ is

(1.8) $$d_G(\underline{X}_n) = \left(\frac{A}{A+nV}\right)\mu + \left(1 - \frac{A}{A+nV}\right)\bar{X}_n$$

and

(1.9) $$r_n(G) = \left(\frac{AV}{A+nV}\right) + cn.$$

The function $AV/(A+nV) + cn$ is a convex function of $n \in (-A/V, \infty)$ with a minimum at $\eta = (A/c)^{1/2} - A/V$. Therefore, we can define an optimal fixed sample size n^* as the smallest positive integer minimizer of (1.9), which is related to η by

(1.10) $$n^* = n^*(A, V) = \begin{cases} 1 & \text{if } \eta < 1 \\ \eta & \text{if } \eta \in \{1, 2, 3, ...\} \\ [\eta] \text{ or } [\eta] + 1, & \text{otherwise} \end{cases}$$

where [ ] denotes the greatest integer function.

§ 1.2. An Empirical Bayes Decision Problem with Random Sample Size Components

When a statistical decision problem occurs repeatedly and independently with the same unknown prior G, one can apply an empirical Bayes approach where G is estimated using data collected from previous repetitions and a Bayes rule with respect to the estimated G is used in the current component problem. The empirical Bayes decision approach with our component permits data accumulated over past component problems to be used in selecting both the sample size and the decision rule to be used in the current component problem. The generality introduced by allowing sample sizes that are determined stochastically makes the result more useful in applications where, typically, the choice of a sample size is an option and based on past data. We impose the usual assumption that the parameter sequence $(\theta_1, \theta_2, \ldots)$ consists of independent G–distributed parameters, where G is an unknown element of the known class of distributions $\mathcal{G}$.

The sample size N_i and the decision rule d_i for the components are determined in an evolutionary way. The sample size N_1 and the decision rule $d_1 \in D_{N_1}$ used in the first component are given nonrandom choices.

The sample size N_2 and the decision rule d_2 are functions of $\underline{X}^1 = (X_{11}, \ldots, X_{1N_1})$, the observations in the first component. The sample size N_3 and the decision rule d_3 are functions of $(\underline{X}^1, \underline{X}^2)$. In general N_i is an integer–valued function of $(\underline{X}^1, \underline{X}^2, \ldots \underline{X}^{i-1})$ and, given N_i, d_i is a D_{N_i}–valued function of $(\underline{X}^1, \underline{X}^2, \ldots, \underline{X}^{i-1})$.

Let $\underline{N} = (N_1, N_2, \ldots)$ and $\underline{d} = (d_1, d_2, \ldots)$. We will be concerned with the risk behavior of empirical Bayes procedures $(\underline{N}, \underline{d})$. (Here and henceforth, the term risk will refer to the expected loss plus cost for observations.) The risk for the decision about θ_i is

(1.11) $$Er_{N_i}(G, d_i) = ER_{N_i}(G, d_i) + cEN_i,$$

where E denotes the expectation over the earlier observations $\underline{X}^1, \underline{X}^2, \ldots \underline{X}^{i-1}$.

Definition 1.1. If the empirical Bayes procedure $(\underline{N}, \underline{d})$ possesses the property:

(1.12) $$\lim_i Er_{N_i}(G, d_i) = r(G) \text{ for all } G \in \mathcal{G},$$

we say it is asymptotically optimal. This means that in the limit, the empirical Bayes procedure has the best possible risk behavior, i.e., achieves minimum Bayes risk.

For a variety of models, we will construct empirical Bayes rules that are asymptotically optimal. All of our results concern parametric families of priors, $\mathcal{G}=\{G_\omega | \omega \in \Omega\}$ where Ω is a specified subset of a finite–dimensional Euclidean space R^p. Families of conjugate priors will be used as the parametric families of priors. We will identify G by ω and replace G accordingly in formulas for risk, etc.

Also, we will use the empirical Bayes approach wherein the prior ω is estimated, say by $\hat{\omega}$, and $\hat{n}=n^*(\hat{\omega})$ and $d_{\hat{\omega}} \in D_{\hat{n}}$ are used in defining the empirical Bayes procedure. Note that we have dropped the superscript on $d_{\hat{\omega}}$. The following table shows how the empirical Bayes procedure evolves using estimates $\hat{\omega}_0$ arbitrary, $\hat{\omega}_1=\hat{\omega}_1(\underline{X}^1)$, $\hat{\omega}_2=\hat{\omega}_2(\underline{X}^1,\underline{X}^2)$, $\hat{\omega}_3=\hat{\omega}_3(\underline{X}^1,\underline{X}^2,\underline{X}^3)$,.... The $\theta_1, \theta_2, \theta_3, \ldots$ are i.i.d. G_ω.

Table 1.1. Empirical Bayes Procedure with Stochastically Determined Sample Sizes

Stage	Parameter	Sample Size	Decision Rule	Observation	Estimated Prior	Risk
1	θ_1	$N_1=n^*(\hat{\omega}_0)$	$d_1=d_{\hat{\omega}_0}$	$\underline{X}^1$	$\hat{\omega}_1(\underline{X}^1)$	$E\{L(\theta_1,d_1(\underline{X}^1))+cN_1\} = r_{N_1}(\omega,d_1)$
2	θ_2	$N_2=n^*(\hat{\omega}_1)$	$d_2=d_{\hat{\omega}_1}$	$\underline{X}^2$	$\hat{\omega}_2(\underline{X}^1, \underline{X}^2)$	$E\{L(\theta_1,d_2(\underline{X}^2))+cN_2\} = Er_{N_2}(\omega, d_2)$
3	θ_3	$N_3=n^*(\hat{\omega}_2)$	$d_3=d_{\hat{\omega}_2}$	$\underline{X}^3$	$\hat{\omega}_3(\underline{X}^1,\underline{X}^2,\underline{X}^3)$	$E\{L(\theta_3,d_3(\underline{X}^3))+cN_3\} = Er_{N_3}(\omega, d_3)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

The convergence of the sequence of risks in the last column to the smallest possible risk $r(\omega) = r_{n^*}(\omega)$ is the asymptotic optimality property. The following remark shows how asymptotic optimality implies the convergence of the sample sizes N_i to the set of optimal fixed sample sizes.

Remark 1.1. Let $s(\omega)$ denote the set of integer minimizers of $r_n(\omega)$

(a) If $(\underline{N}, \underline{d})$ is asymptotically optimal at ω, then

$$P(N_i \in s(\omega)) \to 1 \text{ as } i \to \infty. \tag{1.13}$$

(b) If $r_{N_i}(\omega, d_i) \to r(\omega)$ a.s., then

$$P(N_i \in s(\omega) \text{ eventually}) = 1. \tag{1.14}$$

Proof. For given ω, there exists an $\epsilon > 0$ such that for all $n' \notin s(\omega)$, $r_{n'}(\omega, d) - r(\omega) \geq \epsilon$ for all $d \in D_{n'}$. On the event, $N_i \notin s(\omega)$, $r_{N_i}(\omega, d_i) - r(\omega) \geq \epsilon$ so that

$$E[r_{N_i}(\omega, d_i) - r(\omega)] \geq \epsilon \;\; P(N_i \notin s(\omega)),$$

which yields (1.13) by letting $i \to \infty$. Since $(N_i \notin s(\omega)$ i.o.$)$ implies $r_{N_i}(\omega, d_i) - r(\omega) \geq \epsilon$, i.o., (1.14) is proved. □

The following lemma will be used in subsequent chapters in establishing the asymptotic optimality property.

Lemma 1.1. For priors ω and ν, let $n = n^*(\omega)$, $m = n^*(\nu)$ be optimal fixed sample sizes and let $d^k_\omega, d^k_\nu \in D_k$ denote Bayes decision rules with respect to ω, ν for $k = 1,2,...$ Then

$$\begin{aligned} 0 \leq r_m(\omega, d^m_\nu) - r(\omega) \leq \\ \sup_k |R_k(\omega, d^k_\nu) - R_k(\nu, d^k_\nu)| + \sup_k |R_k(\omega, d^k_\omega) - R_k(\nu, d^k_\omega)|. \end{aligned} \tag{1.15}$$

Proof. The left inequality follows from the fact that $r(\omega)$ is the minimum Bayes risk over choices $d \in D_k$ and sample sizes k. Adding and subtracting $r_m(\nu, d^m_\nu)$ and noting that $r_m(\nu, d^m_\nu) \leq r_n(\nu, d^n_\omega)$ yields

$$r_m(\omega, d^m_\nu) - r_n(\omega, d^n_\omega) \leq r_m(\omega, d^m_\nu) - r_m(\nu, d^m_\nu) + r_n(\nu, d^n_\omega) - r_n(\omega, d^n_\omega) \tag{1.16}$$

which together with (1.4) implies the right inequality of (1.15). □

In Chapters 2 and 3 we develop a.o. empirical Bayes procedures for squared loss estimation and linear loss testing and a binomial component. Here the family of priors is the beta family. In Chapter 2 we give the results of computer simulations that provide estimates of risk behavior for small to moderate i. In Chapters 4 and 5 we treat the two loss functions in a normal component with normal priors.

The quadratic loss function $L(\theta, a) = b(\theta - a)^2$, where $b > 0$, is covered by our results by factoring b out and replacing c by c/b. Similarly, the linear loss function for testing with slopes $-b$ and b for its arms is covered by our work.

Our methods cover more general cost functions as well. If the cost function is $c(n)$ and $\liminf c(n) > R_1(G)$, then for any given G, $\inf \{r_n(G) \mid n = 1, 2, \ldots.\}$ is attained, and we can define $n^*(G)$ as the smallest minimizer. Moreover, the proof of Lemma 1.1. applies to give the same conclusion, that is, a bound for excess risk in terms of the supremum of differences in decision risks over varying sample size problems.

§ 1.3. Literature Review

In the usual empirical Bayes decision problem we are given a stochastic process $(\theta_1, X_1), (\theta_2, X_2), \ldots$ of independent and identically distributed random vectors with the interpretation that, at the ith component problem, observation X_i has distribution P_θ given the parameter $\theta_i = \theta$ and $\theta_1, \theta_2, \ldots$ are i.i.d. with a fixed but unknown prior distribution G in a family of distributions $\mathcal{G}$. The datum X_i may be a vector of summary statistics for the observations taken at the ith component, e.g., the sample mean or other sufficient statistic based on a sample of specified size taken at that stage. The family of priors $\mathcal{G}$ can be an unspecified subfamily of all priors on Θ or a certain parametric family, like conjugate priors. Morris (1983) uses the terminology nonparametric empirical Bayes (NPEB) for the former case and parametric empirical Bayes (PEB) for the latter case. Morris (1983) indicates that PEB is needed to deal with those cases in which number of component problems is too small for Bayes' theory to approximate well. Robbins (1951, 1955, 1966) introduced the empirical Bayes problem. Most of his work and that which followed Robbins is NPEB. It has mainly concerned constructing procedures in a variety of situations that are asymptotically optimal, i.e, such that

$$\lim_i R_k(G, d_i) = R_k(G) \; \forall \, G \in \mathcal{G}.$$

Here k indicates the common sample size taken at each component and on which both the Bayes and empirical Bayes procedure d_G and d_i are based.

Two different approaches have been used in constructing empirical Bayes procedures. The first one is to estimate G from data accumulated from previous component problems and then to construct a Bayes procedure with respect to the estimated G. The second approach is to estimate the Bayes procedure d_G with respect to G directly using data from previous component problems without estimating G from the previous component problems. The first approach gives smoother procedures since the decision rules will be conditionally component Bayes. O'Bryan (1972, 1976) introduced the nonparametric empirical Bayes decision problem with non – i.i.d. components by allowing unequal nonrandom sample sizes in the component problems. He followed the second approach in the situation that P_θ is in the discrete exponential family. O'Bryan (1976) defined asymptotic optimality for his case, which is necessarily more general than that of Robbins (1951), and showed the asymptotic optimality of his procedure. O'Bryan and Susarla (1975) studied the empirical Bayes decision problem with nonidentical components in which P_θ is normal with mean θ and known variance which is changing from component to component.

Laippala (1985), whose work is motivated by O'Bryan (1976), introduced an empirical Bayes problem with nonidentical components with cost for observations and random "floating" sample sizes for the components. Laippala (1985) defines the "optimal" sample size as

$$i_G^* = [\inf \{n \,|\, r_{n+1}(G) \geq r_n(G)\}] \wedge \hat{i}$$

where $\hat{i}$ is a given fixed integer. This is not optimal among the set of all fixed sample sizes since for all $G \in \mathcal{G}$,

$$r_{i_G^*}(G) \geq r_{n^*}(G),$$

and for some $G \in \mathcal{G}$ it is possible to have

$$r_{i_G^*}(G) > r_{n^*}(G).$$

Laippala (1979) defines a floating optimal sample size i_{n+1}^* for use at (n+1)th component problem which is a function of the observations from previous n components as well as current observations. It is pointed out in Gilliland and Karunamuni (1988) that this rule is not necessarily optimal when $\hat{i} \geq 3$ and that the first line of the proof of Theorem 1 in Laippala (1985) claiming that $i_n^* \xrightarrow{P} i_G^*$ neglects the boundary set on which the convergence may fail. Laippala's results as claimed are nonparametric in the sense of Morris (1983).

The component problems that we will consider involve squared error loss estimation and linear loss testing. Many authors have considered the empirical Bayes problem with independent and identical repetitions of these components following Robbins (1956, 1964).

Morris (1983) and Susarla (1982) give general discussions. Singh (1979) provides results on squared error loss estimation problems. Van Ryzin and Susarla (1977) and Gilliland and Hannan (1977) develop the theory for monotone multiple decision problems extending the results for linear loss testing of Johns and Van Ryzin (1971, 1972).

All empirical Bayes work cited above involves identical components with the exception of the nonrandom sample size work of O'Bryan and Laippala. The variant of O'Bryan and Susarla (1975) has a linear loss component with a translation and scale parameter exponential with the scale parameter known and changing from component to component.

Karunamuni (1985, 1988) and Gilliland and Karunamuni (1988) consider the possibility of varying stochastic sample sizes. Gilliland and Karunamuni (1988) develop the theory for finite state problems. Karunamuni (1985, 1988) studies an empirical Bayes problem with a sequential component with linear loss and multiple decision loss structures. He does not treat the optimal fixed sample size problem. Rather, assuming a consistent estimator for G, he shows that the risk of an empirical Bayes one–step sequential decision procedure converges to the Bayes risk attained by the one–step look ahead sequential decision procedure. This is not the asymptotic optimality defined by Robbins (1956).

CHAPTER 2

ESTIMATION OF THE BINOMIAL PARAMETER

§2.1. The Component Problem

Suppose that the rate θ at which defectives are produced by a given production process varies from day–to–day. On each day a random sample of at least two parts is taken at a cost of \$ .50 per part and an estimate $\hat{\theta}$ made with loss \$1000 $(\hat{\theta}-\theta)^2$. If the sequence $\theta_1, \theta_2, \ldots$ is modeled as a stochastic sequence with independent and identically G–distributed variables with G unknown, then the empirical Bayes method is appropriate. For the case G is restricted to the Beta (α, β) family and the sampling is two–at–a–time, we show how to construct a decision procedure with risk plus cost for observations converging to the lowest possibly risk, whatever be α and β. In Section 3 we find that in this case the envelope risk plus cost is no greater than \$18.00 per day, the minimax risk plus cost. Against the least favorable $\alpha=\beta=2$, the empirical Bayes risk is estimated to be below \$20.00 after 15 days. The empirical Bayes sample size converges to the optimal $8 \times 2 = 16$ parts here. Other α, β values are tested in the computational work of Section 3. In this section and the next we develop the empirical Bayes procedure and prove its asymptotic optimality.

Let $X_1, X_2, \ldots$ be i.i.d $B(m, \theta)$, where m is a given positive integer and the parameter θ have prior distribution G in the beta family $\mathcal{G} = \{\text{Beta}\ (\alpha, \beta) \mid \alpha>0, \beta>0\}$. Estimation of θ is considered for squared–error loss. Here $\Theta = \mathcal{A} = [0, 1]$. Let $c > 0$ be a constant cost per observation. Let $d \in D_n$ be a decision rule based on the observation $\underline{X}^n = (X_1, \ldots, X_n)$. The decision loss plus cost for observation is given by $[\theta - d(\underline{X}^n)]^2 + cn$.

The marginal distribution of X_i is Beta–Binomial. We let ξ and η denote the first two moments of $G = \text{Beta}\ (\alpha, \beta)$, that is,

$$\xi = E_G \theta = \frac{\alpha}{\alpha+\beta}$$

$$\eta = E_G\theta^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}, \tag{2.1}$$

and note that $0 < \xi^2 < \eta < \xi < 1$ since $\alpha > 0$, $\beta > 0$. Also

$$\begin{aligned} E(X_i) &= m\xi \\ E(X_i^2) &= m(\xi - \eta) + m^2\eta, \end{aligned} \tag{2.2}$$

and from (2.1) it follows that

$$\begin{aligned} \alpha &= \frac{\xi(\xi-\eta)}{\eta - \xi^2} \\ \beta &= \frac{(1-\xi)(\xi-\eta)}{\eta - \xi^2}. \end{aligned} \tag{2.3}$$

In the empirical Bayes application, (2.2) and (2.3) will be useful in the construction of consistent estimators for α and β. We will use the method of moments to obtain estimates of ξ and η and will use (2.3) to obtain estimates for the parameters α and β.

A Bayes rule exists and is given by the posterior mean of θ, given $\underline{X}^n$. The posterior distribution of θ, given $\underline{X}^n$, is Beta $(\alpha+nX_n, \beta + mn - nX_n)$, where X_n denotes the average of $X_1,\ldots,X_n$.

Hence, a Bayes rule $d_G \in D_n$ is

$$\begin{aligned} d_G(\underline{X}^n) &= \frac{\alpha+nX_n}{\alpha+\beta+mn} \\ &= \frac{\alpha}{\alpha+\beta+mn} + \frac{n}{\alpha+\beta+mn} X_n \end{aligned} \tag{2.4}$$

if $G =$ Beta (α, β).

<u>Remark 2.1.</u> For $G =$ Beta (α, β) and $G' =$ Beta (α', β'),

$$R_n(G, d_{G'}) = \frac{1}{(\alpha'+\beta'+mn)^2}\{[(\alpha' + \beta')^2 - mn]\eta \tag{2.5}$$

$$-[2\alpha'(\alpha'+\beta')-mn]\,\xi+(\alpha')^2\},$$

(2.6) $$|R_n(G, d_{G'})-R_n(G', d_{G'})| \leq 2\,|\xi-\xi'| + |\eta-\eta'|,$$

and

(2.7) $$R_n(G) = \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)(\alpha+\beta+mn)}.$$

Proof. In (2.4) for G', we see that

$$R_n(G, d_{G'}) = E_G E_\theta\,[\theta - \frac{\alpha'}{\alpha'+\beta'+mn} - \frac{n}{\alpha'+\beta'+mn} X_n]^2$$

$$= E_G E_\theta\,[\theta - \frac{\alpha'}{\alpha'+\beta'+mn} - \frac{n}{\alpha'+\beta'+mn}(X_n - m\theta) - \frac{mn}{\alpha'+\beta'+mn}\,\theta]^2$$

$$= E_G E_\theta\,[(\frac{\alpha'+\beta'}{\alpha'+\beta'+mn}\,\theta - \frac{\alpha'}{\alpha'+\beta'+mn}) - (\frac{n}{\alpha'+\beta'+mn}(X_n - m\theta))]^2$$

$$= \left[\frac{1}{\alpha'+\beta'+mn}\right]^2 \{E_G[(\alpha'+\beta')\theta-\alpha']^2 + n^2 E_G E_\theta (X_n - m\theta)^2\}.$$

Using (2.1)

$$E_G[(\alpha'+\beta')\theta-\alpha']^2 = (\alpha'+\beta')^2\eta - 2\alpha'(\alpha'+\beta')\xi + (\alpha')^2$$

and

$$n^2 E_G E_\theta (X_n - m\theta)^2 = n^2 E_G(\mathrm{Var}_\theta X_n)$$

$$= n^2 E_G[\tfrac{1}{n}\, m\theta(1-\theta)] = mn\,(\xi-\eta).$$

Hence

$$R_n(G, d_{G'}) = \left[\frac{1}{\alpha'+\beta'+mn}\right]^2 \{(\alpha'+\beta')^2\,\eta - 2\alpha'(\alpha'+\beta')\xi$$

$$+ (\alpha')^2 + mn\,(\xi-\eta)\}$$

$$= \left[\frac{1}{\alpha'+\beta'+mn}\right]^2 \{[(\alpha'+\beta')^2 - mn]\,\eta - [2\alpha'(\alpha'+\beta')-mn]\xi + (\alpha')^2\},$$

which proves (2.5). Letting $G' = G$ in (2.5) and using (2.1) leads to (2.7). Finally, from (2.5) we obtain (2.6) since

$$|R_n(G, d_{G'}) - R_n(G', d_{G'})| =$$
$$\left|\left[\frac{1}{\alpha'+\beta'+mn}\right]^2 \{[(\alpha' + \beta')^2 - mn](\eta-\eta') - [2\alpha'(\alpha'+\beta') - mn](\xi - \xi')\}\right|$$

and

$$\left|\frac{(\alpha'+\beta')^2 - mn}{(\alpha'+\beta'+mn)^2}\right| \leq \left|\frac{(\alpha'+\beta')^2 + mn}{(\alpha'+\beta'+mn)^2}\right| \leq 1$$

$$\left|\frac{2\alpha'(\alpha'+\beta') - mn}{(\alpha'+\beta'+mn)^2}\right| \leq \frac{2\alpha'(\alpha'+\beta') + mn}{(\alpha'+\beta'+mn)^2} \leq 2. \qquad \square$$

From (2.7) the minimum Bayes risk including cost for observations is

$$r_n(G) = \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}(\alpha + \beta + mn)^{-1} + cn. \tag{2.8}$$

We seek the optimal sample size n^*. $r_n(G)$ is a continuous and convex function of real $n > -(\alpha+\beta)/n$. Consider the equation

$$0 = \frac{d}{dn} r_n(G) = -\frac{m\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}(\alpha + \beta + mn)^{-2} + c.$$

Its larger solution is

$$\nu = \{(\frac{m}{c}\frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)})^{1/2} - (\alpha + \beta)\}/m \tag{2.9}$$

and an optimal fixed sample size $n^* = n^*(\alpha, \beta)$ is given by

$$n^* = \begin{cases} 1 & \text{if } \nu < 1 \\ \nu & \text{if } \nu \in \{1, 2, 3, \ldots\} \\ [\nu] \text{ or } [\nu] + 1 \text{ depending on which integer minimizes } r_n(G), & \\ \qquad \text{otherwise.} & \end{cases} \tag{2.10}$$

Here [ ] denotes the greatest integer function and we take $n^* = [\nu]$ if both $[\nu]$ and $[\nu] + 1$ minimize $r_n(G)$. By the comment preceding Example 1.1 and the fact $R_1(G) \leq .25$ for all G, it follows that $n^* \leq (.25 + c)/c$ for all G.

If α and β were known constants, we can use $d_G \in D_{n^*}$ to achieve minimum Bayes risk, i.e.,

$$r(G) = \min \{r_n(G) \mid n = 1, 2, \ldots\}.$$

In the next section we show how (α, β) is estimated in the empirical Bayes problem with this component and establish the asymptotic optimality for the resulting procedure.

§ 2.2. An Empirical Bayes Decision Procedure

Consider the binomial component problem of the last section. Let $\hat{\alpha}_0, \hat{\beta}_0$ be initial nonrandom estimates of α, β and let $N_1 = n^*(\hat{\alpha}_0, \hat{\beta}_0)$ be the sample size chosen for the first component. (See (2.10) for the definition of the optimal fixed sample size function n^*.) Recall that $\underline{X}^1 = (X_{11}, X_{12},\ldots,X_{1N_1})$ denotes the vector of observations from the first component.

We will define a sequence of estimates $\hat{\alpha}_i, \hat{\beta}_i$ based on $(\underline{X}^1, \underline{X}^2,\ldots,\underline{X}^i)$. Then for component i+1, the empirical Bayes sample size is $N_{i+1} = n^*(\hat{\alpha}_i, \hat{\beta}_i)$ and the empirical Bayes estimator of θ_{i+1} is

$$d_{i+1}(\underline{X}^{i+1}) = \frac{\hat{\alpha}_i + N_{i+1} Y_{i+1}}{\hat{\alpha}_i + \hat{\beta}_i + m N_{i+1}}, \; i = 0, 1,\ldots \tag{2.11}$$

(see (2.4)), where

$$Y_i = \frac{1}{N_i} \sum_{j=1}^{N_i} X_{ij}, \quad i = 1, 2,\ldots \tag{2.12}$$

We will give estimates based on the method of moments and will find it useful to consider

$$Z_i = \frac{1}{N_i} \sum_{j=1}^{N_i} X_{ij}^2, \quad i = 1, 2,\ldots \tag{2.13}$$

and denote average of $Y_j, Z_j, j = 1, 2,\ldots,i$ as $\bar{Y}_i, \bar{Z}_i, i = 1,2,\ldots$

Let $\mathcal{F}_0$ be the trivial σ–field and let $\mathcal{F}_j = \sigma(\underline{X}^1, \underline{X}^2,\ldots,\underline{X}^j), j = 1, 2,\ldots$ The sample size N_j is $\mathcal{F}_{j-1}$ measurable, j = 1,2,..., and we see that

$$\begin{aligned} E(Y_j \mid \mathcal{F}_{j-1}) &= m\,\xi, & j = 1, 2,\ldots \\ E(Z_j \mid \mathcal{F}_{j-1}) &= m\,(\xi - \eta) + m^2\eta, & j = 1, 2,\ldots \end{aligned} \tag{2.14}$$

follow from (2.2.).

Since $Y_j \le m$ and $Z_j \le m^2$, $j = 1, 2,\ldots$, the strong law for centerings at conditional expectation (see Hall and Heyde (1980, Theorem 2.19)) implies

$$(2.15)\qquad \begin{aligned} &Y_i - \frac{1}{i}\sum_{j=1}^{i} E(Y_j \mid \mathcal{F}_{j-1}) \to 0 \ \text{ a.s.} \\ &Z_i - \frac{1}{i}\sum_{j=1}^{i} E(Z_j \mid \mathcal{F}_{j-1}) \to 0 \ \text{ a.s.} \end{aligned}$$

From (2.14) and (2.15) we have

$$(2.16)\qquad \begin{aligned} &Y_i \to m\xi \ \text{ a.s.} \\ &Z_i \to m(\xi-\eta) + m^2\eta \ \text{ a.s.} \end{aligned}$$

Lemma 2.1. Let $m \geq 2$. The estimators defined for $i = 1,2,\ldots$ by

$$(2.17)\qquad \begin{aligned} &\hat{\xi}_i \equiv \frac{Y_i}{m} \\ &\hat{\eta}_i \equiv \frac{Z_i - Y_i}{m(m-1)} \end{aligned}$$

and

$$(2.18)\qquad \begin{aligned} &\hat{\alpha}_i \equiv \left[\frac{\hat{\xi}_i(\hat{\xi}_i - \hat{\eta}_i)}{\hat{\eta}_i - \hat{\xi}_i^2}\right]^+ \\ &\hat{\beta}_i \equiv \left[\frac{(1-\hat{\xi}_i)\ (\hat{\xi}_i - \hat{\eta}_i)}{\hat{\eta}_i - \hat{\xi}_i^2}\right]^+ \end{aligned}$$

are a.s. consistent. (In (2.18) take ratios 0/0 to be 0.)

Proof. The a.s. convergence of the estimates (2.17) follows from (2.16). The a.s. convergence of the estimates (2.18) then follows from (2.3). □

Refer to Table 1.1. Let $\omega = (\alpha, \beta)$, $\hat{\omega}_0$ be arbitrary and $\hat{\omega}_i = (\hat{\alpha}_i, \hat{\beta}_i)$ be defined by (2.18). Let the sample size sequence $\underline{N}$ be defined by $N_{i+1} = n^*(\hat{\alpha}_i, \hat{\beta}_i)$, $i = 0, 1,\ldots$ where n^* is defined by (2.10). Let the empirical Bayes decision rules $\underline{d}$ be defined by (2.11).

Theorem 2.1. Let $m \geq 2$. The empirical Bayes procedure $(\underline{N}, \underline{d})$ defined above is asymptotically optimal at each $G = (\alpha, \beta)$.

Proof. By Lemma 1.1 and (2.6),

$$0 \leq r_{N_{i+1}}(G, d_{i+1}) - r(G) \leq 4|\hat{\xi}_i - \xi| + 2|\hat{\eta}_i - \eta|. \tag{2.19}$$

Since $|\hat{\xi}_i - \xi| \leq 1$ and $|\hat{\eta}_i - \eta| \leq 2$ for all i, the DCT, Lemma 2.1 and (2.19) imply that $Er_{N_{i+1}}(G, d_{i+1}) \to r(G)$. □

Remark 2.2. In the component problem under consideration in this chapter and the next, the marginal distribution of a single observation is Beta–Binomial with parameters m, α, β. If $m = 1$, this is Binomial $(1, \alpha/\alpha + \beta)$ and the pair (α, β) is not identified. Our method of estimation in the empirical Bayes version requires that $m \geq 2$. This assumption can be removed if we require that the $N_i \geq 2$ and use estimators based on pooled data. Requiring $N_i \geq 2$ i.o. would suffice but details of these variations will not be presented. In Chapters 4 and 5 we optimize sample size over $n \geq 2$ for the purpose of simplifying the problem of estimating the prior.

§ 2.3. Some Empirical Bayes Risk Calculations

In this section we treat the empirical Bayes problem of the last section. All risks are multiplied by 1000, which corresponds to a component with loss function $1000(a-\theta)^2$ and cost 1000c per observation.

We have calculated the envelope risk $r(\alpha, \beta)$ and the optimal sample size(s) for various m, c, α, and β and present some of the results in Table 2.1. We have included the mean and standard deviation of the Beta (α, β) prior in each case.

Figure 2.1 below is a graph of the envelope risk function $r(\alpha, \alpha)$ plotted against α on a log scale. For this we have chosen $m = 2$ and $c = .001$.

Figure 2.1 A Risk Envelope

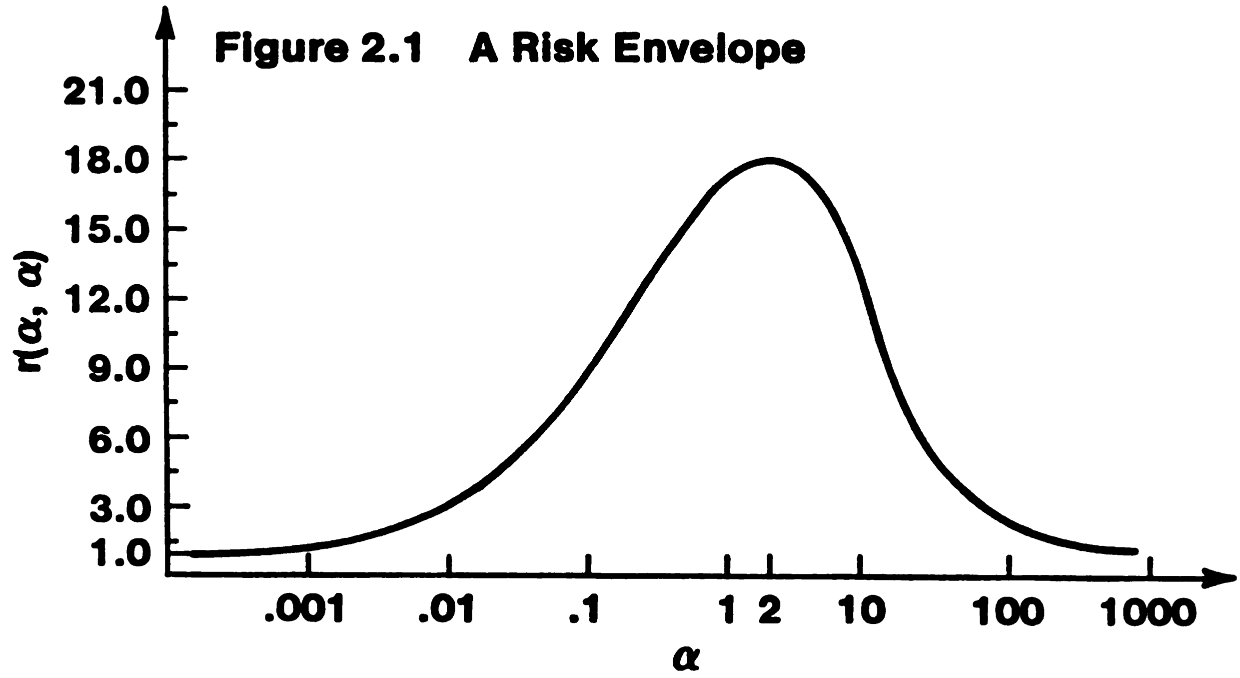

Table 2.1. n*(α, β) and r(α, β)

Prior				m = 2				m = 3			
				c=.001		c=.002		c=.001		c=.002	
α	β	μ	σ	n*	r	n*	r	n*	r	n*	r
0.1	0.1	0.50	0.456	4	9.081	3	12.720	4	7.415	3	10.529
0.1	0.3	0.25	0.366	5	10.151	3	14.371	4	8.320	3	11.699
0.1	0.9	0.10	0.212	4	9.000	3	12.429	4	7.462	2	10.429
0.1	1.9	0.05	0.126	3	6.958	2	9.278	3	5.879	2	7.958
0.2	0.2	0.50	0.423	6	11.760	4	16.503	5	9.638	3	10.529
0.2	0.6	0.25	0.323	6	12.510	4	17.470	3	10.274	3	11.699
0.2	1.2	0.14	0.226	5	11.266	4	15.599	4	9.330	2	10.429
0.2	1.8	0.10	0.173	4	10.000	3	13.500	4	8.286	3	11.455
0.3	0.3	0.50	0.395	7	13.421	5	18.844	5	11.010	4	15.440
0.3	0.6	0.33	0.342	7	14.065	5	19.657	6	11.569	4	16.160
0.3	1.2	0.20	0.253	6	13.111	4	18.105	5	10.818	4	15.111
0.3	1.8	0.14	0.199	5	11.855	4	16.213	5	9.851	3	13.473
0.5	0.5	0.50	0.354	7	15.333	5	21.364	6	12.579	4	17.615
0.5	1.0	0.33	0.298	7	15.602	5	21.594	6	12.838	4	17.877
0.5	1.5	0.25	0.250	7	14.812	5	20.417	6	12.250	4	16.929
1.0	1.0	0.50	0.289	8	17.259	5	23.889	7	14.246	5	19.804
1.0	1.5	0.40	0.262	8	17.266	5	23.714	7	14.295	5	19.796
1.0	2.0	0.33	0.236	8	16.772	5	22.821	6	13.937	4	19.111
1.5	1.5	0.50	0.250	8	17.868	5	24.423	7	14.813	5	20.417
1.5	2.0	0.43	0.233	8	17.768	5	24.109	7	14.775	4	20.289
2.0	2.0	0.50	0.224	8	18.000	5	24.286	7	15.000	4	20.500
3.0	3.0	0.50	0.189	7	17.714	4	23.306	6	14.929	4	19.905
4.0	4.0	0.50	0.167	7	17.101	3	21.873	6	14.547	3	19.072
5.0	5.0	0.50	0.151	6	16.331	3	20.205	5	14.091	3	17.962
10.0	10.0	0.50	0.109	1	11.823	1	12.823	2	11.158	1	12.352

For m=2, c=.001 and selected α, β values, we have made Monte Carlo estimates of the empirical Bayes risk of our procedure with initial starting estimates $\hat{\alpha}_0 = \hat{\beta}_0 = 1$. This is done for stages i = 10, 15, 20, 25, 50 and 100 and the results are presented in Table 2.2 along with the standard errors of the estimates.

Table 2.2 Estimated Empirical Bayes Risks (m=2, c=.001)

α	β	10	15	20	25	50	100	Envelope Risk
		Estimated Empirical Bayes Risks (Standard Errors)						
0.1	0.1	10.22 (0.18)	9.83 (0.07)	10.13 (0.14)	10.00 (0.14)	9.28 (0.05)	9.13 (0.01)	9.081
0.5	0.5	17.31 (0.67)	15.97 (0.10)	15.68 (0.05)	15.56 (0.03)	15.40 (0.01)	15.37 (0.00)	15.333
1.0	1.0	21.27 (0.73)	19.05 (0.43)	18.26 (0.25)	18.05 (0.28)	17.41 (0.02)	17.32 (0.00)	17.259
2.0	2.0	21.26 (0.43)	19.67 (0.25)	19.89 (0.30)	19.44 (0.25)	19.09 (0.20)	18.27 (0.04)	18.000
3.0	3.0	20.43 (0.28)	19.73 (0.24)	19.36 (0.21)	19.75 (0.25)	18.73 (0.14)	18.47 (0.17)	17.714
4.0	4.0	19.98 (0.29)	19.34 (0.19)	19.05 (0.16)	18.95 (0.16)	18.66 (0.15)	18.10 (0.12)	17.101
0.1	0.9	12.25 (0.27)	12.58 (0.34)	13.12 (0.42)	13.05 (0.44)	10.69 (0.31)	9.41 (0.31)	9.000
0.2	1.8	12.79 (0.19)	13.34 (0.24)	13.24 (0.29)	13.28 (0.29)	12.38 (0.28)	10.86 (0.17)	10.000

CHAPTER 3

TESTING THE BINOMIAL PARAMETER

§ 3.1. The Component Problem

In connection with the estimation problem for the binomial parameter θ presented in Chapter 2, we consider a testing problem concerning the value of θ in $B(m, \theta)$, where $m \geq 2$ is a given integer. As in Chapter 2, we assume the conjugate prior $G=\text{Beta}(\alpha, \beta)$ for the binomial parameter θ and a constant cost $c > 0$ per observation. The hypothesis to be tested is

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

for a given $\theta_0 \in \Theta = [0, 1]$. Thus the action space $\mathscr{A}$ consists of two actions a_0 and a_1, where a_0 = "accept H_0" and a_1 = "reject H_0". We assume the the linear loss function $L(\cdot, \cdot) \geq 0$ on $\Theta \times \mathscr{A}$

$$L(\theta, a_0) = (\theta - \theta_0)^+,$$
$$L(\theta, a_1) = (\theta_0 - \theta)^+.$$

Conveniently, $L(\theta, a_0) - L(\theta, a_1) = \theta - \theta_0$. Let $X_1,\ldots,X_n$ be i.i.d. P_θ, the distribution $B(m, \theta)$, with support $\mathscr{X} = \{0, 1,\ldots,m\}$. Then P_θ^n, the joint distribution function of $\underline{X} = (X_1,\ldots,X_n)$, has support $\mathscr{X}^n$.

Let Δ_n denote the set of all nonrandomized decision functions

(3.2) $$\delta\colon \mathscr{X}^n \to \{0, 1\}.$$

When $\underline{x} \in \mathscr{X}^n$ is observed, we take action $a_{\delta(\underline{x})}$ and thereby incur the loss

(3.3) $$\begin{aligned} L(\theta, a_{\delta(\underline{x})}) &= L(\theta, a_0) - \delta(\underline{x})\,[L(\theta, a_0) - L(\theta, a_1)] \\ &= L(\theta, a_0) - \delta(\underline{x})(\theta - \theta_0). \end{aligned}$$

The Bayes risk of $\delta \in \Delta_n$ at G is

(3.4) $$R_n(G, \delta) = EL(\theta, a_{\delta(\underline{x})}),$$

where E denotes the expectation with respect to the joint distribution of $(\theta, \underline{X})$.

Using (3.3), we can write

$$(3.5)\qquad R_n(G, \delta) = \int_{\theta_0}^{1}(\theta - \theta_0)dG(\theta) - \sum_{\underline{x}\in\mathscr{X}^n} \delta(\underline{x})[\int_0^1(\theta - \theta_0)p_\theta(\underline{x})\ dG(\theta)].$$

where p_θ is the conditional mass function for $\underline{X}$. We see that in (3.5),

$$(3.6)\qquad \int_0^1(\theta - \theta_0)p_\theta(\underline{x})dG(\theta) \propto E_G(\theta|\underline{x}) - \theta_0$$

and $E_G(\theta|\underline{X})$ is the Bayes estimate of θ based on $\underline{X}$ defined by (2.4):

$$d_G(\underline{X}) = E_G(\theta|\underline{X}) = \frac{\alpha + (X_1 + \ldots X_n)}{\alpha + \beta + nm}.$$

Thus (3.5) can be written as

$$(3.7)\qquad R_n(G, \delta) = \int_{\theta_0}^{1}(\theta - \theta_0)dG(\theta) - \sum_{\underline{x}\in\mathscr{X}^n} \delta(\underline{x})[d_G(\underline{X}) - \theta_0]\ p(\underline{x}),$$

where p denotes the marginal mass function for $\underline{X}$. Since $\delta(\underline{X})$ takes values 0 or 1, it is clear from (3.7) that $R_n(G, \delta)$ is minimized by taking

$$(3.8)\qquad \delta_G(\underline{X}) = \begin{cases} 1 & \text{if } d_G(\underline{X}) \geq \theta_0 \\ 0 & \text{otherwise} \end{cases}$$

which is a Bayes decision function with respect to G. From (3.8), we observe that a Bayes test $\delta_G \in \Delta_n$ is determined by comparing a Bayes estimate $d_G \in D_n$ with θ_0 for each $n = 1,2,\ldots.$ This observation is useful in that an empirical Bayes test δ_n can be obtained from the empirical Bayes estimate d_n defined in Chapter 2.

<u>Remark 3.1</u>. Let g, g′ be densities of G = Beta (α, β) and G′ = Beta (α', β'). Then we have

$$(3.9)\qquad |R_n(G, \delta_{G'}) - R_n(G', \delta_{G'})| \leq 2\int_0^1 |g'(\theta) - g(\theta)|d(\theta)$$

and

(3.10) $$|R_n(G', \delta_G) - R_n(G)| \leq 2\int_0^1 |g'(\theta) - g(\theta)| d\theta$$

for all n = 1, 2...

<u>Proof</u>. For $\delta = \delta_{G'}$ in (3.7)

(3.11) $$R_n(G, \delta_{G'}) = \int_{\theta_0}^1 (\theta - \theta_0) g(\theta) d\theta - \sum_{\underline{x} \in \mathscr{X}^n} \delta_{G'}(\underline{x})[d_G(\underline{x}) - \theta_0] p(\underline{x}).$$

Letting G = G′ in (3.11) leads to

(3.12) $$R_n(G', \delta_{G'}) = \int_{\theta_0}^1 (\theta - \theta_0) g'(\theta) d\theta - \sum_{\underline{x} \in \mathscr{X}^n} \delta_{G'}(\underline{x})[d_{G'}(\underline{x}) - \theta_0] p'(\underline{x}).$$

By subtraction,

$$\begin{aligned} |R_n(G, \delta_{G'}) - R_n(G', \delta_{G'})| &\leq \int_{\theta_0}^1 |\theta - \theta_0|\ |g(\theta) - g'(\theta)| d\theta \\ &\quad + \sum_{\underline{x} \in \mathscr{X}^n} \delta_{G'}(\underline{x}) [\int_0^1 |\theta - \theta_0|\ |g(\theta) - g'(\theta)| d\theta] \\ &\leq 2\int_0^1 |\theta - \theta_0|\ |g'(\theta) - g(\theta)| d\theta \\ &\leq 2\int_0^1 |g'(\theta) - g(\theta)| d\theta. \end{aligned}$$

The second statement (3.10) follows immediately from (3.12) by changing the roles of G and G′. □

From (3.12), Bayes decision risk of $\delta_G \in \Delta_n$ at G = Beta (α, β) can be written as

(3.13) $$\begin{aligned} R_n(G) &= R_n(G, \delta_G) \\ &= \int_{\theta_0}^1 (\theta - \theta_0)\, g(\theta) d\theta - \sum_{\underline{x} \in \mathscr{X}^n} [d_G(\underline{x}) - \theta_0]^+ p(\underline{x}), \quad n = 1, 2, \ldots. \end{aligned}$$

We seek a minimizer of $r_n(G) = R_n(G) + cn$ among the integers n = 1,2,... By the comment preceding Example 1.1 and the fact that $R_1(G) \leq 1$ for all G, it follows that a minimizer n** satisfies $n^{**} \leq (1+c)/c$ for all G. We have chosen to denote the optimal sample size function for the test as $n^{**} = n^{**}(\alpha, \beta)$ to distinguish it from the optimal sample size n^* for estimation. We do not have an explicit formula for n^{**} although it is easily computed for any given α, β. Thus, using the sample size n^{**} and using $\delta_G \in \Delta_{n^{**}}$, we achieve minimum Bayes risk $r(G)$.

§ 3.2. An Empirical Bayes Decision Procedure

In this section we consider the empirical Bayes decision problem with the linear loss testing component problem described in the last section. The prior G is assumed to be in the parametric family $\mathcal{G}$ of beta priors on $\Theta = [0, 1]$. Let $G =$ Beta (α, β), where $\alpha\, \beta > 0$ are unknown constants. In the sequence of component problems resulting from the repetition of the component, we are given a sequence of parameters $\theta_1, \theta_2, \ldots$ which are assumed to be i.i.d. $G =$ Beta (α, β).

Suppose that we have experienced i component problems by observing $\underline{X}^1 = (X_{11}, \ldots, X_{1N_1}), \ldots, \underline{X}^i = (X_{i1}, \ldots, X_{iN_i})$. At the (i+1)th component problem we will test

$$H_0\colon \theta_{i+1} \leq \theta_0 \text{ against } H_1\colon \theta_{i+1} > \theta_0$$

with the linear loss function given by (3.1). Since $\theta_{i+1} \sim$ Beta (α, β) and $\alpha > 0$, $\beta > 0$ are unknown, the optimal sample size $n^{**}(\alpha, \beta)$ and Bayes decision rule $\delta_G \in \Delta_{n^{**}(\alpha,\beta)}$ are not directly available, so that the minimum Bayes risk $r(G)$ cannot be achieved. However, if an estimate $\hat{G}_i$ of G is available at this stage, we estimate the optimal sample size $n^{**}(G)$ and the Bayes rule $\delta_G \in \Delta_{n^{**}(G)}$ at G by $N_{i+1} = n^{**}(\hat{G}_i)$ and $\delta_{i+1} = \delta_{\hat{G}_i} \in \Delta_{n^{**}(\hat{G}_i)}$ and, thus, define an empirical Bayes procedure $(\underline{N}, \underline{\delta})$ as in Table 1.1. For the estimates of α, β assume $m \geq 2$ and let $\hat{\alpha}_i, \hat{\beta}_i$ be given by (2.18). Let $\hat{\alpha}_0, \hat{\beta}_0$ be arbitrary initial estimates. Then

$$N_{i+1} = n^{**}(\hat{\alpha}_i, \hat{\beta}_i)\,, \quad i = 1, 2, \ldots \tag{3.14}$$

and

$$\delta(\underline{X}^{i+1}) = \begin{cases} 1\,, & \text{if } d_{i+1}(\underline{X}^{i+1}) \geq \theta_0 \\ 0\,, & \text{otherwise} \end{cases} \tag{3.15}$$

where d_{i+1} is defined by (2.11).

Lemma 3.1. Let $m \geq 2$ and let $\hat{\alpha}_i, \hat{\beta}_i$ be the a.s. consistent estimators, e.g., as in (2.18). Let $\hat{g}_i$ denote the Beta density with parameter $\hat{\alpha}_i, \hat{\beta}_i$ and let g be the beta density with the governing parameter values α, β. Then

$$\hat{g}_i(\theta) \to g(\theta), \quad 0 < \theta < 1, \text{ a.s.} \tag{3.16}$$

Proof. At each θ, $g(\theta)$ is a continuous function of (α, β). □

Theorem 3.1. Let $m \geq 2$. The empirical Bayes testing procedure $(\underline{N}, \underline{\delta})$ defined by (3.14) and (3.15) is asymptotically optimal at each $G = \text{Beta}(\alpha, \beta)$.

Proof. From Lemma 1.1 and (3.10), it follows that

$$0 \leq Er_{N_{i+1}}(G, \delta_{i+1}) - r(G) \leq 4\, E\int_0^1 |\hat{g}_i(\theta) - g(\theta)|\, d\theta. \tag{3.17}$$

Note that the sequence $\hat{g}_i - g \to 0$ a.s. on the probability space of the empirical Bayes problem cross Lebesgue measure on $(0, 1)$. The sequence $\hat{g}_i + g$ dominates $|\hat{g}_i - g|$ and converges to $2g(\theta)$ so by the generalized dominated convergence theorem, RHS (3.17) converges to zero. □

CHAPTER 4

ESTIMATION OF THE NORMAL MEAN

§ 4.1 The Component Problem

The component problem considered in this chapter is the one introduced in Example 1.1. Here $G = N(\mu, V)$ and, letting

$$\rho = \frac{A}{A+nV}, \tag{4.1}$$

the posterioi distribution of θ given $\underline{X} = (X_1, X_2, \ldots, X_n)$ is

$$N(\rho\mu + (1-\rho)\,\overline{X}, \frac{A}{A+nV}). \tag{4.2}$$

With this notation, the Bayes estimator (1.8) can be written

$$d_G(\underline{X}) = \rho\mu + (1-\rho)\overline{X}. \tag{4.3}$$

The following remark parallels Remark 2.1.

Remark 4.1. For $G = N(\mu, V)$ and $G' = N(\mu', V')$,

$$R_n(G, d_{G'}) = (1-\rho')^2 \frac{A}{n} + \rho'^2 [(\mu - \mu')^2 + V], \tag{4.4}$$

$$|R_n(G, d_{G'}) - R_n(G', d_{G'})| \le (\mu' - \mu)^2 + |V' - V| \tag{4.5}$$

and

$$R_n(G) = \frac{AV}{A+nV}. \tag{4.6}$$

Proof. By (4.3), $d_{G'}(\underline{X}) = \rho'\mu' + (1-\rho')\overline{X}$. Since expected squared deviation is variance plus bias squared,

$$\begin{aligned} R_n(G, d_{G'}) &= E_G E_\theta[\rho'\mu' + (1-\rho')\overline{X} - \theta]^2 \\ &= E_G \{(1-\rho')^2 \frac{A}{n} + \rho'^2 (\mu' - \theta)^2\} \\ &= (1-\rho')^2 \frac{A}{n} + \rho'^2 [V + (\mu' - \mu)^2]. \end{aligned}$$

Then (4.6) follows by replacing G' by G above and using (4.1). Since

$$R_n(G', d_{G'}) = (1-\rho')^2 \frac{A}{n} + \rho'^2 V'$$

it follows that

$$R_n(G, d_{G'}) - R_n(G', d_{G'}) = \rho'^2[(\mu' - \mu)^2 + (V - V')]$$

which yields (4.5). $\square$

We seek the optimal sample size n^* which minimizes

$$r_n(G) = R_n(G) + cn = \frac{AV}{A+nV} + cn$$

among the integers $n = 1,2,\ldots$. We consider $r_n(G)$ as a function of real n and the equation

$$0 = \frac{d}{dn} r_n(G) = -\frac{AV^2}{(A + nV)^2} + c.$$

Its larger solution is

$$\eta = (A/c)^{1/2} - A/V. \tag{4.7}$$

We see that $r_n(G)$ is convex in $n \in (-A/V, \infty)$ and that the optimal sample size $n^* = n^*(A, V)$ is given by (1.10).

In our empirical Bayes application the variance A of the conditional distribution $N(\theta, A)$ is assumed to be unknown but is assumed to be in a given bounded interval $(0, a]$. Thus we are taking A to be a nuisance parameter.

It is convenient, though not necessary, to require that at least two observations be taken in each component of the empirical Bayes problem so that the estimation of A is simple. Therefore, we will optimize sample size over choices $n = 2, 3,\ldots$ in defining the envelope risk. It follows that

$$n^* = n^*(A, V) = \begin{cases} 2 & \text{if } \eta < 2 \\ \eta & \text{if } \eta \in \{2,3,\ldots\} \\ [\eta] \text{ or } [\eta]+1, & \text{otherwise} \end{cases} \tag{4.8}$$

where η is given in (4.7).

Since $R_2(G) = E_G E_\theta (\bar{X} - \theta)^2 \leq A/2$, it follows as in the comment preceding Example 1.1 that $n^* \leq (A/2 + 2c)/c$. Letting M be the integer $[a/2c + 2] + 1$, it follows that

$$2 \leq n^*(A, V) \leq M < \infty \tag{4.9}$$

for all A and priors $G = N(\mu, V)$.

Notice that in the component problem

$$EX̄_n = \mu, \tag{4.10}$$

$$E\frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)^2 = V+A, \tag{4.11}$$

and, provided $n \geq 2$,

$$E\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X}_n)^2 = A. \tag{4.12}$$

§ 4.2. An Empirical Bayes Decision Procedure

In this section we construct a decision procedure for the empirical Bayes problem with the component of the last section. The unknown prior G is assumed to be from the family of normal distributions $\mathscr{G}$, the family of conjugate priors. Let $G = N(\mu, V)$, where $\mu \in (-\infty, \infty)$ and $V \in (0, \infty)$.

Let $\hat{A}_0$, $\hat{\mu}_0$ and $\hat{V}_0$ be initial nonrandom estimates of the component nuisance parameter A and the parameters μ, V of the prior. Let $N_1 = n^*(\hat{A}_0, \hat{V}_0)$. Then $\underline{X}^1 = (X_{11}, \ldots, X_{1N_1})$ is observed in the first component. The empirical Bayes procedure that we will study is defined through sequences of estimators $\hat{A}_i$, $\hat{\mu}_i$ and $\hat{V}_i$ that are $(\underline{X}^1, \ldots, \underline{X}^i)$ measurable with

(4.13) $$N_{i+1} = n^*(\hat{A}_i, \hat{V}_i), \quad i = 0, 1, \ldots$$

and

(4.14) $$d_{i+1}(\underline{X}^{i+1}) = \hat{\rho}_{i+1}\hat{\mu}_i + (1-\hat{\rho}_{i+1})\, Y_{i+1}, \quad i = 0, 1, \ldots$$

where

(4.15) $$\hat{\rho}_{i+1} = \frac{\hat{A}_i}{\hat{A}_i + N_{i+1}\hat{V}_i}, \quad i = 0, 1, \ldots$$

and

(4.16) $$Y_i = \frac{1}{N_i} \sum_{j=1}^{N_i} X_{ij}, \quad i = 1, 2, \ldots$$

We now define the estimators $\hat{\mu}_i$, $\hat{A}_i$ and $\hat{V}_i$, $i = 1, 2, \ldots$. Motivated by (4.10) we define

(4.17) $$\hat{\mu}_i = \overline{Y}_i = \frac{1}{i} \sum_{j=1}^{i} Y_j, \quad i = 1, 2, \ldots$$

the average of the sample means for the first i components. Motivated by (4.12) we define

(4.18) $$\hat{A}_i = \overline{S}_i \wedge a \quad i = 1, 2, \ldots$$

where

$$\bar{S}_i = \frac{1}{i} \sum_{j=1}^{i} S_j \tag{4.19}$$

is the average of the sample variances

$$S_j = \frac{1}{N_j - 1} \sum_{k=1}^{N_j} (X_{jk} - Y_j)^2 \tag{4.20}$$

for the first i components. Finally, motivated by (4.11) we define

$$\hat{V}_i = [\hat{T}_i - \hat{A}_i]^+, \quad i = 1, 2, \ldots \tag{4.21}$$

where

$$\hat{T}_i = \frac{1}{i} \sum_{j=1}^{i} T_{ji} \tag{4.22}$$

is the average of the average squared deviations from $\hat{\mu}_i = \bar{Y}_i$,

$$T_{ji} = \frac{1}{N_j} \sum_{k=1}^{N_j} (X_{jk} - \bar{Y}_i)^2. \tag{4.23}$$

In (4.23) the centerings change with i, which creates a more complicated stochastic structure than exists in (4.20). For purpose of triangulation, we introduce

$$\bar{T}_i = \frac{1}{i} \sum_{j=1}^{i} T_j, \tag{4.24}$$

where

$$T_j = \frac{1}{N_j} \sum_{k=1}^{N_j} (X_{jk} - \mu)^2. \tag{4.25}$$

Let $\mathcal{F}_0$ be the trivial σ–field and let $\mathcal{F}_j = \sigma(\underline{X}^1, \underline{X}^2, \ldots, \underline{X}^j)$, $j = 1, 2, \ldots$. The sample size N_j is $\mathcal{F}_j$–measurable and we see that

$$\begin{aligned} E(Y_j \mid \mathcal{F}_{j-1}) &= \mu, \quad j = 1,2,\ldots \\ E(S_j \mid \mathcal{F}_{j-1}) &= A, \quad j = 1,2,\ldots \\ E(T_j \mid \mathcal{F}_{j-1}) &= V + A, \quad j = 1,2,\ldots \end{aligned} \tag{4.26}$$

<u>Lemma 4.1</u>. The sequences $\hat{\mu}_i = \bar{Y}_i$, $\bar{S}_i$ and $\bar{T}_i$ are a.s. consistent for μ, A and V+A, respectively.

Proof. We will use (4.26) and the theorem on stability about conditional expectation used earlier, i.e., Hall and Heyde (1980, Theorem 2.19). The sequences Y_i, S_i and T_i are not bounded. However, we will find random variables Y, S and T that are square integrable and stochastically larger than their absolute values. This implies the hypothesis of Theorem 2.19 that is sufficient for the a.s. convergence.

Recall that $2 \le N_i \le M$, $i = 1,2,....$ Consider the component problem with sample size M and observations $X_1, X_2, ...X_M$. Let $Y = \Sigma\,|X_j|$, $S = \Sigma\,X_j^2$ and $T = \Sigma\,(X_j - \mu)^2$. From the definitions (4.16), (4.20) and (4.25) we see that Y, S and T are stochastically larger than $|Y_i|$, $|S_i|$ and $|T_i|$, $i = 1,2,....$ Also $Y^2 \le MS$ and, conditional on θ, the distributions of S and T are noncentral chi–square distributions with second moments that are integrable $N(\mu, V)$. Thus, Y, S, and T are square integrable. □

Lemma 4.2. The estimator $\hat{T}_i$ and $\hat{V}_i$ are a.s. consistent for V+A and V.

Proof. We have from (4.23) and (4.25) that

$$T_j - T_{ji} = (\bar{Y}_i - \mu)(2Y_j - \mu - \bar{Y}_i). \tag{4.27}$$

Since $\bar{Y}_i = \Sigma\,Y_j/i$, we have from (4.22) and (4.24), that

$$\mathbf{T}_i - \hat{T}_i = (\bar{Y}_i - \mu)^2. \tag{4.28}$$

It follows from Lemma 4.1 that $\hat{T}_i$ is a.s. consistent for V+A. Using (4.21) and Lemma 4.1 it follows that $\hat{V}_i$ is consistent for V. □

Theorem 4.1. Let $A \le a$. Then the empirical Bayes procedure $(\underline{N}, \underline{d})$ defined by (4.13) – (4.23) is asymptotically optimal at each $G = N(\mu, V)$.

Proof. From Lemma 1.1 and (4.5),

$$0 \le r_{N_{i+1}}(G, d_{i+1}) - r(G) \le 2(\hat{\mu}_i - \mu)^2 + 2|\hat{V}_i - V|. \tag{4.29}$$

Let Y, T be the random variables defined in the proof of Lemma 4.1. Then for $p > 0$, $E|Y_j|^{2+p} \le E(Y+1)^{2+p} < \infty$ and $E|T_j|^{1+p} \le E(T+1)^{1+p} < \infty$ for $j = 1,2,\ldots.$ Hence, the $\{Y_j^2\}$ and the $\{T_j\}$ are uniformly integrable. Thus, $\{\hat{\mu}_i^2\}$ and $\{T_i\}$ are uniformly integrable and the a.s. convergence (Lemma 4.1) implies that

$$E(\hat{\mu}_i - \mu)^2 \to 0 \tag{4.30}$$

and

$$E|T_i - (V+A)| \to 0. \tag{4.31}$$

It follows from the triangle inequality and (4.28) that

$$\begin{aligned} |\hat{V}_i - V| &\le |\hat{T}_i - T_i| + |T_i - (V+A)| + |(V+A) - (V+\hat{A}_i)| \\ &= (\hat{\mu}_i - \mu)^2 + |T_i - (V+A)| + |\hat{A}_i - A|. \end{aligned} \tag{4.32}$$

The dominated convergence theorem and Lemma 4.1 imply

$$E|\hat{A}_i - A| \to 0 \tag{4.33}$$

which together with (4.29) – (4.32) establish the result. □

CHAPTER 5

TESTING THE NORMAL MEAN

§ 5.1. The Component Problem.

In this section we consider linear loss testing of the normal mean θ in $N(\theta, A)$. Specifically, we consider the problem of testing

$$(5.1) \qquad H_0\colon\ \theta \le \theta_0 \ \text{ against } \ H_1\colon\ \theta > \theta_0$$

with

$$L(\theta, a_0) = (\theta - \theta_0)^+$$
$$L(\theta, a_1) = (\theta_0 - \theta)^+.$$

Using the analysis developed in Section 3.1 for the component of this section, we find that for any test δ,

$$(5.2) \qquad R_n(\theta, \delta) = \int_{\theta_0}^{\infty} (\theta - \theta_0) dG(\theta) - \int_{\mathscr{X}^n} \delta(\underline{x})\, [d_G(\underline{x}) - \theta_0] f(\underline{x})\, d\underline{x},$$

where $\mathscr{X} = (-\infty, \infty)$, f is the marginal density of $\underline{X} = (X_1, \ldots, X_n)$ and d_G is given by (4.3). A Bayes test versus the prior G is

$$(5.3) \qquad \delta_G(\underline{X}) = \begin{cases} 1 & \text{if } d_G(\underline{X}) \ge \theta_0 \\ 0 & \text{if } d_G(\underline{X}) < \theta_0. \end{cases}$$

Throughout this chapter we will take $\mathscr{G}$ to be the family of normal distributions $N(\mu, V)$ with

$$(5.4) \qquad E_G\, |\theta| = \int |\theta|\, g(\theta) d\theta \le K < \infty,$$

where $K > 0$ is a known constant.

Remark 5.1. Let g, g' be densities for $G = N(\mu, V)$, $G = N(\mu', V')$ in $\mathscr{G}$. Then

$$(5.5) \qquad |R_n(G, \delta_{G'}) - R_n(G', \delta_{G'})| \le 2\int |\theta - \theta_0|\, |g'(\theta) - g(\theta)| d\theta$$

and

(5.6) $$R_n(G) \le 2(K + |\theta_0|)$$

for $n = 1, 2, \ldots$.

Proof. (5.5) follows as in (3.9). Let $G' = G$ in $R_n(G', d_{G'})$. Using

$$|d_G(\underline{x}) - \theta_0| \ f(\underline{x}) \le \int_{-\infty}^{\infty} |\theta - \theta_0| f_\theta^n(\underline{x}) \ g(\theta) d\theta,$$

we have

$$\begin{aligned} R_n(G) = R_n(G, \delta_G) &\le \int_{-\infty}^{\infty} |\theta - \theta_0| \ g(\theta) \ d\theta \\ &+ \int_{\mathscr{X}^n} [\int_{-\infty}^{\infty} |\theta - \theta_0| \ |g(\theta) \ d\theta \, f_\theta^n(\underline{x}) \ d\underline{x} \\ &= 2 \int_{-\infty}^{\infty} |\theta - \theta_0| \ g(\theta) \ d\theta \\ &\le 2[\int_{-\infty}^{\infty} |\theta| \ g(\theta) \ d\theta + |\theta_0|] \le 2(K + |\theta_0|) \end{aligned}$$

for $n = 1, 2, \ldots$, which proves (5.6). □

Remark 5.2. Let ϕ be the c.d.f. of $N(0, 1)$. Then

(5.7) $$\begin{aligned} \int |\theta| \ g(\theta) d\theta = (\tfrac{2V}{\pi})^{1/2} \exp(-2\mu^2/V) \\ + \mu [1 - 2\phi(-\mu/\sqrt{V})] \end{aligned}$$

and

(5.8) $$\begin{aligned} |\int |\theta| \ g'(\theta) d\theta - \int |\theta| \ g(\theta) d\theta| &\le 3|\mu' - \mu| \\ &+ |\sqrt{V'} - \sqrt{V}| + 2|\mu'| \ |\phi(-\mu/\sqrt{V'}) - \phi(-\mu/\sqrt{V})| \\ &+ \sqrt{V} | \exp(-2\mu'^2/V') - \exp(-2\mu^2/V)|. \end{aligned}$$

Proof. A direct calculation gives (5.7). Using (5.7) for G' and subtraction, LHS of (5.8) is less than or equal to

$$\begin{aligned} &|(\tfrac{2V'}{\pi})^{1/2} \exp(-2\mu'^2/V') + \mu' (1 - 2\phi(-\mu'/\sqrt{V'})) \\ &- (\tfrac{2V}{\pi})^{1/2} \exp(-2\mu^2/V) - \mu(1 - 2\phi(-\mu/\sqrt{V}))| \\ &\le |(\tfrac{2V'}{\pi})^{1/2} - (\tfrac{2V}{\pi})^{1/2}| \exp(-2\mu'^2/V') \\ &+ (\tfrac{2V}{\pi})^{1/2} |\exp(-2\mu'^2/V') - \exp(-2\mu^2/V)| \\ &+ |\mu' - \mu| + 2|\mu' - \mu| \ \phi(-\mu'/\sqrt{V'}) \\ &+ 2|\mu| \ |\phi(-\mu'/\sqrt{V'}) - \phi(-\mu/\sqrt{V})| \end{aligned}$$

$$\leq 3\ |\mu' - \mu| + |\sqrt{V'} - \sqrt{V}| + 2|\mu|\ |\Phi(-\mu'/\sqrt{V'}) - \Phi(-\mu/\sqrt{V})|$$
$$+ \sqrt{V}|\ \exp(-2\mu'^2/V') - \exp(-2\mu^2/V)|,$$

the RHS of (5.8). □

We seek the smallest minimizer n^{**} of $r_n(G) = R_n(G) + cn$, $n = 2,3,\ldots$ (As in Chapter 4 we are optimizing over $n \geq 2$.) It follows as in the comment preceding Lemma 1.1, that $n^{**} \leq (R_2(G) + 2c)/c$, so using (5.6) and letting M denote the integer $[\{2(K + |\theta_0|) + 2c\}/c] + 1$,

$$2 \leq n^{**} \leq M < \infty \tag{5.9}$$

for all $G \in \mathcal{G}$.

§ 5.2. An Empirical Bayes Decision Procedure

Suppose that in the component problem of Section 5.1, the prior $G=N(\mu, V)$ and the variance A of the conditional distribution $N(\theta, A)$ are unknown. We assume that unknown prior G is in the subfamily $\mathcal{G}$ of normal distributions satisfying

$$E_G|\theta| \leq K < \infty$$

and that the variance A, a nuisance parameter, is in a bounded interval $(0, a]$. The constants K and a are known.

In general, the optimal sample size n^{**} defined in (5.9) is a function of A and $G = N(\mu, V)$, i.e.,

$$n^{**} = n^{**}(A, \mu, V)$$

for $(A, \mu, V,) \in (0, a] \times (-\infty, \infty) \times (0, \infty)$.

If the component problem occurs repeatedly and independently with the same unknown $G = N(\mu, V)$ and A, the empirical Bayes approach is applicable.

Suppose that we have experienced i components by observing

$$\underline{X}^1 = (X_{11},\ldots,X_{1N_1}),\ldots, \underline{X}^i = (X_{i1},\ldots,X_{iN_i})$$

from $N(\theta_1, A),\ldots,N(\theta_i, A)$, where $\theta_1, \theta_2,\ldots$ are independent G–distributed parameters. At the $(i+1)$th component, we will test

$$H_0\colon \theta_{i+1} \leq \theta_0 \text{ against } H_1\colon \theta > \theta_0.$$

This will be carried out by determining the sample size N_{i+1} and the decision rule $\delta_{i+1} \in \Delta_{N_{i+1}}$ for $i = 0, 1,\ldots$.

Let $\hat{A}_0, \hat{\mu}_0,$ and $\hat{V}_0$ be nonrandom initial estimates of A, μ and V and let $N_1 = n^{**}(\hat{A}_0, \hat{\mu}_0, \hat{V}_0)$. Then $\underline{X}^1 = (X_{11},\ldots,X_{1N_1})$ is observed in the first component. The empirical Bayes procedure that we will study is defined through $\hat{A}_i, \hat{\mu}_i, \hat{V}_i$ that are $(\underline{X}^1,\ldots,\underline{X}^i)$ measurable with

(5.9) $$N_{i+1} = n^{**}(\hat{A}_i, \hat{\mu}_i, \hat{V}_i)$$

and

$$\text{(5.10)} \qquad \delta_{i+1}(\underline{X}^{i+1}) = \begin{cases} 1 & , \text{ if } d_{i+1}(\underline{X}^{i+1}) \geq \theta_0 \\ 0 & , \text{ otherwise} \end{cases}$$

where $d_{i+1}(\underline{X}^{i+1})$ is defined by (4.14) for $i = 0, 1, \ldots.$

If we use $\hat{A}_i$, $\hat{\mu}_i$ and $\hat{V}_i$ defined by (4.8), (4.16) and (4.21) in constructing $(\underline{N}, \underline{\delta}) = ((N_1, N_2, \ldots), (\delta_1, \delta_2, \ldots))$ given by (5.9), (5.10), then it is easy to see that they satisfy all the consistency properties proved in Lemma 4.1, 4.2 and (4.30), (4.31), (4.33) in Theorem 4.1.

The following lemma is useful in proving the asymptotic optimality of $(\underline{N}, \underline{\delta})$ that has been constructed above through $\hat{A}_i$, $\hat{\mu}_i$ and $\hat{V}_i$ $i = 0, 1, \ldots.$

<u>Lemma 5.1</u>. Let $\hat{g}_i$, g be densities of $\hat{G}_i = N(\hat{\mu}_i, \hat{V}_i)$, $G = N(\mu, V)$ for $i = 0, 1, \ldots$

Then

$$\text{(5.11)} \qquad \lim_i E \int_{-\infty}^{\infty} |\hat{g}_i(\theta) - g(\theta)| d\theta = 0$$

and

$$\text{(5.12)} \qquad \lim_i E \int_{-\infty}^{\infty} |\theta - \theta_0| \, |\hat{g}_i(\theta) - g(\theta)| d\theta = 0$$

<u>Proof</u>. Note that $\hat{g}_i - g \to 0$ a.e. on the measure space of the empirical Bayes problem cross Lebesgue measure on $(-\infty, \infty)$. Using the same argument as in the proof of Theorem 3.1., we obtain (5.11).

For (5.12), it suffices to show that

$$\text{(5.13)} \qquad \lim_i E \int_{-\infty}^{\infty} |\theta| \, |\hat{g}_i(\theta) - g(\theta)| d\theta = 0.$$

Since $|\theta| \, |\hat{g}_i(\theta) - g(\theta)| \to 0$ a.e. on the measure space of the empirical Bayes problem cross Lebesgue measure on $(-\infty, \infty)$, $|\theta| \, (\hat{g}_i(\theta) + g(\theta))$ dominates the integrand $|\theta| \, |\hat{g}_i(\theta) - g(\theta)|$ and $|\theta| \, (\hat{g}_i(\theta) + g(\theta)) \to 2|\theta| \, g(\theta)$ a.e. on that product space, (5.13) will follow by generalized dominated convergence theorem by showing that

$$(5.14) \qquad E \int_{-\infty}^{\infty} |\theta| \, (\hat{g}_i(\theta) + g(\theta))d\theta \to 2\, E\int_{-\infty}^{\infty} |\theta| \, g(\theta)d\theta.$$

Using (5.8) applied to $g' = \hat{g}_i$,

$$(5.15) \qquad |E\int_{-\infty}^{\infty} |\theta| \, \hat{g}_i(\theta)d\theta - \int_{-\infty}^{\infty} |\theta| \, g(\theta)d\theta|.$$

$$\leq 3E\, |\hat{\mu}_i - \mu| + E\, |\sqrt{\hat{V}_i} - \sqrt{V}|$$

$$+ 2|\hat{\mu}_i| \, E|\Phi(-\hat{\mu}_i/\sqrt{\hat{V}_i}) - \Phi(-\mu/\sqrt{V})|$$

$$+ \sqrt{\hat{V}_i}\, E|\exp(-2\hat{\mu}_i^2/\hat{V}_i) - \exp(-2\mu^2/V)|,$$

which converges to 0 by the a.s. consistency and the mean consistency of $\hat{\mu}_i$ and $\hat{V}_i$. The proof is completed since

$$|E\int_{-\infty}^{\infty} |\theta| \, \hat{g}_i(\theta)d\theta - E\int_{-\infty}^{\infty} |\theta| \, g(\theta)d\theta|$$

$$\leq E\, |\int_{-\infty}^{\infty} |\theta| \, \hat{g}_i(\theta)d\theta - \int_{-\infty}^{\infty} |\theta| \, g(\theta)d\theta|. \qquad \square$$

<u>Theorem 5.1</u>. Let $A \leq a$. Then the empirical Bayes decision procedure $(\underline{N}, \underline{\delta})$ defined by (5.9), (5.10) through the estimates $\hat{A}_i$ $\hat{\mu}_i$ and $\hat{V}_i$ given by (4.18), (4.17) and (4.21) is asymptotically optimal for all G with $E_G|\theta| \leq K$.

<u>Proof</u>. From Lemma 1.1, (5.5), and Lemma 5.1,

$$0 \leq E\, r_{N_{i+1}}(G, d_{i+1}) - r(G)$$

$$\leq 4\, E\int_{-\infty}^{\infty} |\theta - \theta_0| \, |\hat{g}_i(\theta) - g(\theta)| d\theta \to 0 \qquad \square$$

REFERENCES

REFERENCES

Gilliland, Dennis and Hannan, James (1977). Improved rates in the empirical Bayes monotone multiple decision problem with MLR family. Ann. Statist. **5** 516–521.

Gilliland, Dennis and Karunamuni, Rohana (1988). On empirical Bayes with sequential component. Ann. Inst. Statist. Math. **40** 187–193.

Hall, P. and Heyde, C. C. (1980). Martingale limit theory and its application. Academic Press, New York.

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.

Karunamuni, Rohana (1985). Empirical Bayes with sequential components. Ph.D. Thesis. Dept. of Statistics & Probability, Michigan State University.

Karunamuni, Rohana (1988). On empirical Bayes testing with sequential components. Ann. Statist. **16** 1270–1282.

Laippala, P. (1979). The empirical Bayes approach with floating optimal sample size in binomial experimentation. Scand. J. Statist. **6** 113–118; correction note **7** 105.

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

Morris, Carl (1983). Parametric empirical Bayes inference: Theory and Applications. J. Amer. Statist. Assoc. **78** 47–65.

O'Bryan, T. (1972). Empirical Bayes results in the case of non–identical components. Ph.D. Thesis, RM–306, Statistics and Probability, Michigan State University.

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

O'Bryan, T. and Susarla, V. (1975). An empirical Bayes two–action problem with non–identical components for a translated exponential distribution. Comm. Statist. **4(8)** 767–775.

Robbins, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. Proc. 2nd Berkeley Symp. Math. Statist. Prob. 131–148, University of California Press.

Robbins, H. (1956). The empirical Bayes approach to statistics. Proc. 3rd Berkeley Symp. Math. Statist. Prob. **1** 157–163, University of California Press.

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

Singh, R. S. (1979). Empirical Bayes estimation in Lebesgue–exponential families with rates near the best possible rate. Ann. Statist. **7** 890–902.

Susarla, V. (1982). Empirical Bayes Theory, Encyclopedia of Statistical Science, (eds. Kotz and Johnson) **2** 490 – 503, Wiley, New York.

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

MICHIGAN STATE UNIV. LIBRARIES
31293005480292