EMPIRICAL BAYES RESULTS IN THE CASE OF NON-IDENTICAL COMPONENTS

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
THOMAS EUGENE O'BRYAN
1972

This is to certify that the

thesis entitled

EMPIRICAL BAYES RESULTS IN THE CASE OF
NON-IDENTICAL COMPONENTS

presented by

Thomas Eugene O'Bryan

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics and Probability

Major professor

Date August 11, 1972

O-7639

ABSTRACT

EMPIRICAL BAYES RESULTS IN THE CASE OF NON-IDENTICAL COMPONENTS

By

Thomas Eugene O'Bryan

A Bayes rule with respect to a distribution G will minimize the risk of a decision concerning a parameter θ which is distributed according to G. The infimum Bayes risk is denoted by $R(G)$. Herbert Robbins ((1956), An empirical Bayes approach to statistics. Proc. Third Berkeley Symp. Math. Statist. Prob., 157-163, University of California Press) demonstrated that, even if G is unknown, in certain cases one can construct statistical procedures based on data gathered from n independent repetitions of the decision problem for which the risk converges to $R(G)$ as $n \to \infty$ for all G. Such an empirical Bayes procedure is asymptotically optimal. Rudimentary forms of this problem had appeared prior to Robbins' unifying treatment and a huge empirical Bayes literature has evolved since.

Only sequences of identical component problems have been treated in the literature. However, it is clear that when the only difference from problem to problem is sample size, empirical Bayes methods should still be useful. In this case there is not a single Bayes envelope $R(G)$, but rather a sequence of envelopes $R^{m_n}(G)$

where m_n denotes the sample size in the nth problem. Let $\underline{\theta} = (\theta_1, \theta_2, \ldots)$ be a sequence of iid G variables and let the conditional distribution of the observations $\underline{X}_n = (X_{n,1}, \ldots, X_{n,m_n})$ given $\underline{\theta}$ be $(P_{\theta_n})^{m_n}$, $n = 1,2,\ldots$. For a decision concerning θ_n, we will investigate procedures t_n which will utilize all the data $\underline{X}_1, \ldots, \underline{X}_n$ and which, under certain conditions, are asymptotically optimal in the sense that $\lim_n [R^{m_n}(t_n, G) - R^{m_n}(G)] = 0$ for all G. In particular this paper treats squared error loss estimation and linear loss testing in certain discrete exponential families where the construction of asymptotically optimal procedures is tractable.

EMPIRICAL BAYES RESULTS IN THE CASE OF NON-IDENTICAL COMPONENTS

By

Thomas Eugene O'Bryan

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

1972

TO MY PARENTS

AND

MARY

ACKNOWLEDGMENTS

I wish to express my appreciation to Professor Dennis C. Gilliland for his guidance throughout the preparation of this manuscript and his concern for my work. I am indebted to both he and Professor James Hannan for the introduction to empirical Bayes decision theory as well as many other aspects of mathematical statistics and for their many helpful comments and suggestions which led the way to stronger theorems and simplified proofs.

The financial support provided by the Department of Statistics and Probability and the National Science Foundation made my graduate studies possible. I thank them and also wish to thank Mrs. Noralee Barnes for her patience and skill in typing this dissertation.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

§1.1 The Statistical Decision Problem

Consider the following statistical decision problem. Let $\{P_\theta : \theta \in \Theta\}$ be a family of probability measures over a σ-field $\mathcal{B}$ of subsets of $\mathfrak{X}$. Θ will be called the parameter space and θ will denote a generic element of Θ. $\mathfrak{X}$ will be called the observation space. Let $\mathcal{A}$ be an action space with generic element a. Let $L \geq 0$ be a loss function defined on $\Theta \times \mathcal{A}$. Let G be a probability measure over a σ-field $\mathcal{F}$ of subsets of Θ. G will be called the "a priori" distribution. With $\mathcal{C}$ a σ-field of subsets of $\mathcal{A}$, a randomized decision rule (decision function) t has domain $\mathfrak{X} \times \mathcal{C}$ and is such that $t(x,\cdot)$ is a probability measure on $\mathcal{C}$ for each fixed $x \in \mathfrak{X}$ and $t(\cdot,C)$ is $\mathcal{B}$ measurable for each fixed $C \in \mathcal{C}$. When θ is the parameter, the decision rule t results in expected loss

$$(1.1) \qquad R(t,\theta) = \int_{\mathfrak{X}} \int_{\mathcal{A}} L(\theta,a)\,t(x,da)\,P_\theta(dx) \ .$$

We require that $\mathcal{C}$ contain all the singleton subsets of $\mathcal{A}$ so that the class of randomized decision rules contains the class of nonrandomized decision rules. With a nonrandomized decision rule t, $t(x,\cdot)$ puts all its probability on a singleton set, say $\{t(x)\}$, for each x. For a nonrandomized decision rule t

$$(1.1') \qquad R(t,\theta) = \int_{\mathfrak{X}} L(\theta,t(x))P_\theta(dx) \ .$$

$R(t,\theta)$ is called the risk of t with respect to θ. When G is the "a priori" distribution on Θ, the overall expected loss is given by

$$(1.2) \qquad R(t,G) = \int_{\Theta} R(t,\theta)G(d\theta) \ .$$

$R(t,G)$ is called the Bayes risk of t with respect to G. Let

$$(1.3) \qquad R(G) = \inf_t R(t,G) \ .$$

$R(G)$ is called the Bayes envelope evaluated at G. If there exists a decision rule t^* such that $R(t^*,G) = R(G)$, then we write $t^* = t_G$ and call t_G a Bayes decision rule with respect to G. Note that t_G need not exist and, if it does exist, it need not be unique. However if it does exist, then there exists a nonrandomized decision rule which is also a Bayes decision rule with respect to G.

We give two examples of the preceeding which will illustrate the two types of decision problems, i.e., estimation and testing, which we will be concerned with in the following chapters. In both examples, let $\Theta = (0,\infty)$. Let $\{P_\theta : \theta \in \Theta\}$ be the Poisson family of distributions with means θ, i.e., P_θ admits a density f_θ with respect to counting measure μ on $\mathfrak{X} = \{0,1,\ldots\}$ given by $f_\theta(x) = e^{-\theta}\theta^x(x!)^{-1}$. Let G be the Gamma distribution on $(0,\infty)$ which has density with respect to Lebesgue measure λ given by $g(\theta) = [\Gamma(\beta)]^{-1}\alpha^\beta\theta^{\beta-1}e^{-\alpha\theta}$ for $\alpha,\beta > 0$.

<u>Example 1.1</u>. (Poisson) Estimation. Let $\mathcal{A} = \Theta$. Let $L(\theta,a) = (\theta-a)^2$. Then for any nonrandomized decision rule, (1.2) becomes

$$R(t,G) = \int_{\Theta} \sum_{x=0}^{\infty} (t(x)-\theta)^2 f_\theta(x) g(\theta) d\theta \ .$$

Here the nonrandomized Bayes decision rule with respect to G is given by a version of the conditional expectation of θ given $X = x$,

$$t_G(x) = \frac{\beta+x}{\alpha+1} \ .$$

<u>Example 1.2</u>. (Poisson) Testing. Let $0 < c < \infty$. Let $\mathcal{A} = \{a_1, a_2\}$ correspond to the actions "decide $\theta \le c$" and "decide $\theta > c$" respectively. Let $L(\theta,a_1) = b(\theta-c)^+$ and let $L(\theta,a_2) = b(\theta-c)^-$. Here we use the symbol δ rather than t for decision rules. Let $\delta(x)$ be the probability of choosing action a_1 given x. Here we have (1.2) as

$$R(\delta,G) = \int_{\Theta} \sum_{x=0}^{\infty} [\delta(x)L(\theta,a_1) + (1-\delta(x))L(\theta,a_2)] f_\theta(x) g(\theta) d\theta$$

and a nonrandomized Bayes decision rule with respect to G is given by

$$\delta_G(x) = [\frac{\beta+x}{\alpha+1} \le c] \ .$$

<u>Example 1.3</u>. Consider the problem of Example 1.1 with $X = (X_1,\ldots,X_m)$, a sequence of iid random variables each distributed f_θ. The statistic $Y = \sum_{i=1}^{m} X_i$ is sufficient for this family. Here $Y \sim f_{m\theta}$. The nonrandomized Bayes decision rule with respect to G is given by a version of the conditional expectation of θ given $Y = y$, $t_G(y) = \frac{y+\beta}{\alpha+m}$. Also we calculate $R(G) = \frac{\beta}{\alpha(\alpha+m)}$.

The important thing to notice here is that both the Bayes decision rule with respect to G and the Bayes envelope evaluated at G depend upon the sample size m.

Throughout this paper, we will let square brackets denote indicator functions, so that for an event A, $[A]$ denotes the indicator function of A.

§1.2 The Empirical Bayes Decision Problem

In the case where G is known and a Bayes decision rule with respect to G exists, one merely employs t_G and thereby incurs the minimum possible Bayes risk $R(G)$. But suppose G is unknown. Robbins (1956, 1963, 1964) showed that if a given statistical decision problem occurred repeatedly and independently with the same unknown G throughout, then, under certain conditions, one could exhibit a sequence of rules $\{t_n\}$ which had Bayes risk with respect to G converging to the Bayes envelope evaluated at G. As the problem repeats itself, it presents a sequence of pairs of random variables $\{(\theta_i, X_i)\}$ with each pair being independent of all other pairs. The θ_i are unobservable and iid with distribution G. The conditional distribution of X_i given that $\theta_i = \theta$ is given by P_θ. Robbins suggested that one use a decision rule t_n in the $n+1$st repetition of the problem with t_n depending on $X_1,\ldots,X_n$. Robbins' rationale was that one could use the knowledge about G gained through the variables $X_1,\ldots,X_n$ in such a way that for large n the Bayes risk with respect to G of t_n would be close to the Bayes risk with respect to G of t_G. With t_n used as the decision rule in the $n+1$st problem, the risk conditional on $X_1,\ldots,X_n$ is

$$(1.4)\qquad R(t_n,G) = \int_\Theta \int_{\mathfrak{X}} \int_{\mathcal{A}} L(\theta,a)\, t_n(x,da) P_\theta(dx) G(d\theta)$$

which satisfies $R(t_n,G) \geq R(G)$ in view of (1.3). Hence with the overall expected loss for the decision concerning θ_{n+1} denoted by

$$(1.5)\qquad R_n(t_n,G) \equiv E\, R(t_n,G)$$

we see that

(1.6) $$R_n(t_n,G) \geq R(G) \ .$$

Definition 1.1: If $\lim_{n\to\infty} R_n(t_n,G) = R(G)$, then $\{t_n\}$ is said to be asymptotically optimal relative to G and we will write $\{t_n\}$ a.o. relative to G.

§1.3 History

The search for rules $\{t_n\}$ which are a.o. relative to G for every distribution G or at least for every G within a certain class has taken basically two tracks. The first track is to use the values of $X_1,\ldots,X_n$ to form an estimate of G, call it $\hat{G}_n$, and then let t_n be a Bayes decision rule with respect to $\hat{G}_n$, i.e., let $t_n = t_{\hat{G}_n}$. The second track is to estimate the form of the Bayes decision rule directly without estimating G first.

In 1955 at the Third Berkeley Symposium on Mathematical Statistics, Robbins introduced empirical Bayes procedures and discussed both tracks mentioned above. In 1963 and 1964, two more papers by Robbins appeared which discussed the empirical Bayes problem further. Rudimentary forms of the problem had appeared prior to Robbins' unifying treatment and a huge empirical Bayes literature has evolved since. We will concern ourselves here with that segment of the literature which involves situations similar to what we will discuss in this paper.

Along the second track, Johns (1957) discussed estimation in the case where the class of probability distributions $\{P_\theta\colon \theta \in \Theta\}$ was not restricted to a particular parametric family. Macky (1966) and more recently Hannan and Macky (1971) have dealt with certain exponential families in the case of estimation and have demonstrated a.o. rules with weak restrictions on the prior distribution and the parameter space. Samuel (1963) discussed the testing problem under various loss structures and in part dealt specifically with the type of discrete exponential families which we will discuss in

Chapter II. Johns and Van Ryzin (1971) treated the testing problem with linear loss and developed rates of convergence on $R_n(\delta_n,G) - R(G)$.

Concerning the estimation of G which is part of the first track, Tucker (1963) examined the case where $\{P_\theta : \theta \in \Theta\}$ was the family of Poisson distributions. Rolph (1968) used Bayesian estimation of G in the case where the parameter space Θ was limited to $[0,1]$. Recently Meeden (1972) has looked at Bayesian estimation of G in the case where Θ may be $[0,\infty)$.

§1.4 The Non-Identical Case

The history of the empirical Bayes decision problem is such that the only problem that seems to have been considered thus far is the problem where the stages are identical repetitions of a given component problem. One could ask whether it is meaningful and useful to apply empirical Bayes procedures to sequences of independent but not identical decision problems all having the same unknown G. We will attempt to answer this question in part in the remainder of this paper. Specifically, we will address the case where the statistical decision problems in the sequence are identical except for sample size. When we observe a random vector of observations $\underline{X} = (X_1,\ldots,X_m)$, where m may vary from stage to stage, it will become necessary to consider the dependence of the Bayes decision rules and the Bayes envelopes evaluated at G upon the values of m (cf. Example 1.3). This was not necessary before when one considered problems where the sample sizes were identical at each stage.

In the situation that we are considering, where the problems occur independently with the same unknown G throughout, there is a sequence of independent random vectors $\{(\theta_i,\underline{X}_i)\}$, $i = 1,2,\ldots$ where $\underline{X}_i = (X_{i1},\ldots,X_{i,m_i})$ is the sample of size m_i from the ith problem. The random variables θ_i are unobservable and are iid with distribution G. Conditional on $\theta_i = \theta$, $X_{i1},\ldots,X_{i,m_i}$ are iid P_θ. We consider a decision rule t_n for use in the n+1st problem which depends on $\underline{X}_1,\ldots,\underline{X}_n$. Letting $m_{n+1} = m$, the risk conditional on $\underline{X}_1,\ldots,\underline{X}_n$ is given by

$$(1.7) \qquad R^m(t_n,G) = \int_{\Theta}\int_{\mathcal{X}^m}\int_{\mathcal{A}} L(\theta,a)\,t_n(\underline{x},da)P_\theta^m(d\underline{x})G(d\theta)$$

which satisfies $R^m(t_n,G) \geq R^m(G)$ where $R^m(G)$ is the Bayes envelope for a sample size m component problem. Hence, with the overall expected loss for the decision concerning θ_{n+1} denoted by

$$(1.8) \qquad R_n(t_n,G) \equiv E\, R^m(t_n,G) \ ,$$

we see that

$$(1.9) \qquad R_n(t_n,G) \geq R^m(G) \ .$$

This motivates the following definition which parallels Robbins' definition.

Definition 1.2. A sequence of decision rules $\{t_n\}$ is said to be asymptotically optimal (a.o.) relative to G if $\lim_{n\to\infty} D_n(G) = 0$ where $D_n(G) = R_n(t_n,G) - R^{m}{}_{n+1}(G)$.

The remainder of this paper treats squared error loss estimation and linear loss testing involving certain discrete exponential families and exhibits sequences of rules $\{t_n\}$ which are asymptotically optimal in the situation described in this section. We will approach the problem along the second track discussed before.

CHAPTER II

DECISION PROBLEMS INVOLVING SOME DISCRETE EXPONENTIAL FAMILIES (PRELIMINARIES)

§2.1 Introduction

We impose the following special structure on the component problem to be treated in the empirical Bayes framework. Suppose that conditional on θ, $X_1,\ldots,X_m$ are iid P_θ where P_θ has a density with respect to counting measure μ on $\mathfrak{X} = \{0,1,\ldots\}$ given by

$$(2.1) \qquad f_\theta(x) = \theta^x z(\theta) g(x) \quad \text{where} \quad g(x) > 0,\ x \in \mathfrak{X}$$

and $\theta \in \Theta \subset \Pi$ where

$$\Pi \equiv \{\theta \geq 0 : \sum_{x=0}^{\infty} \theta^x g(x) < \infty\}$$

and z is the function defined by

$$z(\theta) \equiv (\sum_{x=0}^{\infty} \theta^x g(x))^{-1},\ \theta \geq 0 \quad .$$

The function z is continuous on $[0,\infty)$, is decreasing to 0, and is positive on the interval Π. We note that $Y = \sum_{i=1}^{m} X_i$ is sufficient for this family. The random variable Y, whose distribution we will denote by $P_{\theta,m}$, takes values in $\mathfrak{X}$ and has density with respect to μ given by

$$(2.2) \qquad f_{\theta,m}(y) = \theta^y z^m(\theta) g_m(y), \ y \in \mathcal{X}$$

where

$$g_m(y) = \sum_{A_m(y)} \prod_{i=1}^{m} g(x_i) > 0, \ y \in \mathcal{X}$$

(2.3)

$$A_m(y) = \{(x_1, \ldots, x_m) : \sum_{i=1}^{m} x_i = y\} .$$

With $\theta \sim G$, the marginal density of Y with respect to μ is given by

$$(2.4) \qquad h_m(y) = q_m(y) g_m(y)$$

where

$$(2.5) \qquad q_m(y) = \int_{\Theta} \theta^y z^m(\theta) G(d\theta)$$

is the marginal density for Y with respect to the measure on $\mathcal{X}$ defined by the mass density g_m. We note that for all m, $q_m(y) > 0$ for all $y \in \mathcal{X}$ and hence $h_m(y) > 0$ for all $y \in \mathcal{X}$, unless G is degenerate at 0 in which case $q_m(0) = z^m(0) = (g^{-1}(0))^m > 0$, $h_m(0) = 1$ and $q_m(y) = h_m(y) = 0$ for $y \geq 1$.

Of course $f_{\theta,1} = f_\theta$, $g_1 = g$ and for convenience we let $h_1 = h$ and $q_1 = q$.

Two common families of the type discussed above are the Poisson family with

$$f_\theta(x) = \theta^x e^{-\theta} (x!)^{-1}$$

(2.6) and

$$f_{\theta,m}(y) = \theta^y e^{-m\theta} m^y (y!)^{-1}$$

and the Negative Binomial family ($r > 0$ known) with

$$f_\theta(x) = \theta^x(1-\theta)^r\binom{r+x-1}{x}$$

(2.7) and

$$f_{\theta,m}(y) = \theta^y(1-\theta)^{mr}\binom{mr+y-1}{y} .$$

In this paper we will consider two loss structures. In Chapter III we will consider

Estimation with $\Theta \subset \mathcal{A} \subset [0,\infty)$ and $L(\theta,a) = (\theta-a)^2$ and in Chapter IV we will consider

Testing with $\mathcal{A} = \{a_1,a_2\}$ and for $b > 0$ and $c \in \Theta$,

$$L(\theta,a_1) = b(\theta-c)^+,\ L(\theta,a_2) = b(\theta-c)^- .$$

For Estimation (hereafter understood with squared error loss), the Bayes risk of any nonrandomized rule t based on Y in the sample size m problem is given by

$$R(t,G) = \int_\Theta \sum_{y=0}^{\infty} (t(y) - \theta)^2 f_{\theta,m}(y)G(d\theta) \tag{2.8}$$

and the nonrandomized rule which is Bayes with respect to G is given by a version of the conditional expectation of θ given $Y = y$,

$$t_G(y) = \frac{q_m(y+1)}{q_m(y)} \tag{2.9}$$

where throughout this paper ratios 0/0 are to be interpreted as 0. For Testing (hereafter understood with linear loss function as given above), the Bayes risk of any randomized rule δ with respect to G, where $\delta(y)$ is the probability of taking action a_1 given

$Y = y$, is given by

$$(2.10) \qquad R(\delta,G) = \int_{\Theta} \sum_{y=0}^{\infty} [\delta(y)L(\theta,a_1) + (1-\delta(y))L(\theta,a_2)]f_{\theta,m}(y)G(d\theta)$$

and a nonrandomized rule which is Bayes with respect to G is given by

$$(2.11) \qquad \delta_G(y) = [a_G(y) \le 0]$$

where

$$(2.12) \qquad a_G(y) = q_m(y+1) - c\, q_m(y) .$$

We will add superscripts m to $R,t,t_G,\delta,\delta_G,a$ and a_G whenever it is necessary to emphasize the dependence on m.

The empirical Bayes problem that we consider involves repetitions of the component problem (either <u>Estimation</u> or <u>Testing</u>) with sample size m varying from problem to problem. We now turn our attention to the method that we will use in constructing rules which are a.o. relative to any G. Notice that the Bayes decision rules (2.9) and (2.11) depend upon y through q_m. With G unknown, q_m is unknown but can be estimated. One method is to express q_m as a function of q, a common marginal density with respect to g of all X_{ij}, $j = 1,\ldots,m_i$, $i = 1,2,\ldots$ and to use the X_{ij} data to estimate q. With the discrete exponential family (2.1) this method can sometimes be implemented. We note that

$$(2.13) \qquad q_m(y) = \int_{\Theta} z^{m-1}(\theta)\theta^y z(\theta)G(d\theta),\ y \in \mathfrak{X} .$$

If $z(\theta)$ is a polynomial then $z^{m-1}(\theta)$ is also a polynomial and we can write for each m

(2.14) $$z^{m-1}(\theta) = \Sigma \gamma_k \theta^k$$

where the γ_k are constants. Substituting (2.14) into (2.13) and interchanging the order of integration and (finite) summation yields

(2.15) $$q_m(y) = \Sigma \gamma_k q(y+k)$$

so that q_m is expressed in terms of the estimable function q.

Since $z(\theta)$ is continuous, given an interval $[0,\beta]$, $\beta < \infty$, the Weierstrass' approximation theorem allows that for each m and every $\epsilon > 0$ there exists a polynomial $\Sigma \gamma_k \theta^k$ which approximates $z^{m-1}(\theta)$ to within ϵ uniformly on $[0,\beta]$, i.e.,

(2.16) $$|z^{m-1}(\theta) - \Sigma \gamma_k \theta^k| < \epsilon \quad \text{for all} \quad \theta \in [0,\beta] .$$

Defining

(2.17) $$q_{m,\epsilon}(y) = \Sigma \gamma_k q(y+k), \; y \in \mathfrak{X},$$

we see that

(2.18) $$|q_m(y) - q_{m,\epsilon}(y)| \le$$
$$\int_{\Theta} |z^{m-1}(\theta) - \Sigma \gamma_k \theta^k| \theta^y z(\theta) g(d\theta) \le \epsilon\, q(y), \; y \in \mathfrak{X},$$

i.e., $q_{m,\epsilon}(y)$ approximates $q_m(y)$ to within $\epsilon\, q(y)$ for each $y \in \mathfrak{X}$ and each m.

For the Negative Binomial f_θ of (2.7) with r an integer, $z(\theta) = (1-\theta)^r$, a polynomial of degree r. For the Poisson f_θ of (2.6), $z^{m-1}(\theta) = e^{-\theta(m-1)}$ has a power series expansion about

0 and the series is uniformly convergent on any bounded set of θ values for each m and hence, for each $\epsilon > 0$, the polynomial approximating z^{m-1} to within ϵ uniformly on the bounded set of θ values can be found simply by truncating the power series expansion.

We will find it sufficient but not necessary for our purposes to require that the estimator of $q(y)$, $y \in \mathfrak{X}$, be

$$\bar{q}(y) \equiv \frac{1}{n} \sum_{i=1}^{n} {}_i q(y), \quad y \in \mathfrak{X}, \tag{2.19}$$

where for each i, ${}_i q(y)$ is an unbiased estimator of $q(y)$ based on $\underline{X}_i$, bounded by $1/g(y)$, and note that ${}_i q(y) = 0$ for all $y \geq 1$ if $\underline{X}_i = \underline{0}$. As an average of unbiased and bounded estimators, $\bar{q}(y)$ is unbiased and pointwise consistent for $q(y)$, $y \in \mathfrak{X}$.

We now develop an example of such an estimator. Let $X_1,\ldots,X_m$ be a sample where, conditional on θ, $X_1,\ldots,X_m$ are iid f_θ and where $\theta \sim G$. An unbiased estimator of $q(k)$ is provided by $[X_1 = k]/g(k)$ since

$$E[X_1 = k] = h(k) = q(k)g(k) \ . \tag{2.20}$$

An improved estimator based on the sufficient statistic $Y = X_1 + \ldots + X_m$ is given by the conditional expectation

$$E_y[X_1 = k] = \frac{P[X_1 = k, X_2 + \ldots + X_m = y - k]}{P[Y = y]} \ .$$

The probabilities can be computed first conditional on θ and then unconditionally to obtain the result

$$E_y[X_1 = k] = \frac{g(k)g_{m-1}(y-k)}{g_m(y)} \quad . \tag{2.21}$$

This relation (2.21) holds for all $m \geq 1$, $k,y \in \mathfrak{X}$ with the definitions $g_0(u) = [u = 0]$ and $g_m(u) = 0$ if $m \geq 1$ and $u < 0$. In view of (2.20) and (2.21) we see that the estimator

$$\frac{g_{m-1}(Y - k)}{g_m(Y)} \ , \ k \in \mathfrak{X} \tag{2.22}$$

is unbiased for $q(k)$, $k \in \mathfrak{X}$. Furthermore, $0 \leq g(k)g_{m-1}(Y - k) \leq g_m(Y)$ for each m to ensure that the estimator (2.22) is bounded by $g^{-1}(k)$, $k \in \mathfrak{X}$. When $X_1 = \ldots = X_m = 0$ then $Y = 0$ and we see that the estimator (2.22) is equal to 0 for $k \geq 1$. Therefore, with Y_i denoting the sufficient statistic in the ith component problem,

$${}_iq(y) = \frac{g_{m_i-1}(Y_i - y)}{g_{m_i}(Y_i)} \ , \ y \in \mathfrak{X}, \ i = 1,2,\ldots \tag{2.23}$$

provides an example of an estimator to be used in (2.19).

§2.2 Assumptions

The discussion in Section 2.1 helps motivate the imposition of the following assumptions. In order to exhibit rules $\{t_n\}$ which are a.o. relative to all G we will impose these assumptions from time to time in what follows.

(A1$^-$) $\Theta \subset [0,\beta]$ where $\beta < \infty$.

(A1) $\Theta \subset [0,\beta]$ where $\beta \in \Pi$.

(A2) $z(\theta)$ is a polynomial in $\theta \in \Theta$.

(A3) The sample sizes m_n form a bounded sequence, $m_n \le M < \infty$, $n=1,2,\ldots$.

In Estimation, the action space $\mathcal{A}$ is always taken to be $[0,\beta]$.

Remark 2.1. The assumption (A2) implies the assumption (A1$^-$).

Proof: Recall from the definition of $z(\theta)$ in Section 2.1 that (i) $z(\theta) \ge 0$ for all $\theta \ge 0$ and (ii) $z(\theta) \downarrow 0$ as $\theta \uparrow \infty$. If (A2) obtains, then

$$z(\theta) = \Sigma_0^K \gamma_k \theta^k, \ \theta \in \Theta,$$

for constants $\gamma_0,\ldots,\gamma_K$ where $\gamma_K \ne 0$. If $\gamma_K > 0$, $\Sigma_0^K \gamma_k \theta^k \to \infty$ as $\theta \to \infty$ so that Θ must be bounded in view of (ii). If $\gamma_K < 0$, $\Sigma_0^K \gamma_k \theta^k \to -\infty$ as $\theta \to \infty$ so that Θ must be bounded in view of (i).

Of course, assumption (A1) also implies Θ is bounded. In the presence of (A2), (A1) only adds the requirement that $\sup\{\theta | \theta \in \Theta\} \in \Pi$. Assumption (A1) need not be satisfied when (A2) is satisfied; for example, with $\Theta = [0,1) = \Pi$, the negative binomial family with r a known integer satisfies (A2); however, Θ does not satisfy (A1).

§2.3 Lemmas

We will need the following lemmas in Chapters III and IV.

Lemma 2.1. For the discrete exponential family (2.1)

$$(2.24)\quad P_{\theta,m}[Y > y] \le P_{\beta,M}[Y > y] \quad \text{for all } y \in \mathfrak{X},$$

$$0 \le \theta \le \beta \in \Pi,\ 1 \le m \le M .$$

Proof. Let $y \in \mathfrak{X}$, $0 \le \theta \le \beta \in \Pi$ and $1 \le m \le M$ be fixed. Since $\{f_{\theta,m}(y) \mid \theta \in \Pi\}$ has an increasing monotone likelihood ratio in y it follows from Lehmann (1959, Lemma 3.2) that

$$P_{\theta,m}[Y > y] \le P_{\beta,m}[Y > y] .$$

Since $\Sigma_1^M X_i = \Sigma_1^m X_i + \Sigma_{m+1}^M X_i$ where the $X_i \ge 0$, $P[\Sigma_1^M X_i \ge y] \ge P[\Sigma_1^m X_i \ge y]$ so that

$$P_{\beta,m}[Y > y] \le P_{\beta,M}[Y > y] .$$

In our constructions in Chapters III and IV it is necessary in the absence of (A3) to use the existence of decision rules L^m such that $R(L^m,G) \to 0$ as $m \to \infty$ for every G. The two lemmas to follow will establish the existence of such rules. We notice that

$$(2.25)\qquad \frac{g(0)}{g(1)}\,\frac{f_\theta(1)}{f_\theta(0)} = \theta$$

has a "natural" bounded estimator

$$(2.26)\quad T^m(X_1,\ldots,X_m) = \frac{g(0)}{g(1)}\,\frac{\sum_{i=1}^{m}[X_i = 1]}{\sum_{i=1}^{m}[X_i = 0]} \wedge \beta ,$$

under $(A1^-)$.

<u>Lemma 2.2</u>. (Estimation) Under $(A1^-)$, $R(T^m,G) \to 0$ as $m \to \infty$ for any G.

<u>Proof</u>: Let G be an arbitrary but fixed a priori distribution on Θ. Since $\frac{1}{m} \sum_{i=1}^{m} [X_i = x] \xrightarrow{P_\theta^m} f_\theta(x)$, $x \in \mathcal{X}$ and $f_\theta(0) \neq 0$ and $\theta \in [0,\beta]$, we see that

$$T^m(X_1,\dots,X_m) \xrightarrow{P_\theta^m} \frac{g(0)}{g(1)} \frac{f_\theta(1)}{f_\theta(0)} = \theta \ .$$

Since $(T^m - \theta)^2 \le \beta^2$ for all m, the dominated convergence yields

$$R(T^m,\theta) = E_\theta(T^m - \theta)^2 \to 0 \quad \text{for each } \theta$$

where E_θ denotes expectation with respect to the distribution P_θ^m. Since $E_\theta(T^m - \theta)^2 \le \beta^2$ for all θ and m, another application of the dominated convergence theorem yields

$$R(T^m,G) = \int_\Theta R(T^m,\theta)G(d\theta) \to 0 \ . \tag{2.27}$$

For the testing problem we define

$$\Delta^m(X_1,\dots,X_m) = [T^m \le c] \ . \tag{2.28}$$

<u>Lemma 2.3</u>. (Testing) Under $(A1^-)$, $R(\Delta^m,G) \to 0$ as $m \to \infty$ for all G.

<u>Proof</u>: Let G be an arbitrary but fixed a priori distribution on Θ and let E_θ be as in Lemma 2.3. We can write

$$R(\Delta^m,\theta) = b\{E_\theta([T^m \le c](\theta - c)^+)$$

$$+ E_\theta([T^m > c](\theta - c)^-)\} \le b\, E_\theta|T^m - \theta| \ .$$

Since $T^m - \theta \to 0$ in L_2 (cf. (2.27)), $T^m - \theta \to 0$ in L_1 which completes the proof.

CHAPTER III

ESTIMATION

§3.1 Estimation Under ($A1^-$) and (A3)

In this chapter we will exhibit sequences of decision rules which are a.o. relative to every G in the case of squared error loss estimation in the special discrete exponential families described in Section 2.1. As has been noted by many authors (e.g. Macky (1966, pp. 6-7)) and quite apart from distributional assumptions if $T(X)$ and θ are L_2 random variables then $T(X) - E[\theta|X]$ and $E[\theta|X] - \theta$ are orthogonal in L_2 so that $E(T(X) - \theta)^2 - E(E[\theta|X] - \theta)^2 = E(T(X) - E[\theta|X])^2$. This implies that for estimating θ in our sample size m component decision problem with an estimator t,

$$(3.1) \qquad R^m(t,G) - R^m(G) = E(t - t_G)^2 .$$

In dealing with risks concerning the estimation of θ_{n+1} in the empirical Bayes problem we will find it convenient to drop subscripts within the n+1st problem. We will let P and E denote probability and expectation taken over all random variables on which they operate and P_y and E_y denote their conditional on $Y = y$ counterparts.

The following lemma is motivated by the approach of Robbins (1964) and proves useful in establishing the asymptotic optimality of specified sequences of decision rules when assumptions (A1$^-$) and (A3) obtain.

Lemma 3.1. Suppose (A1$^-$) and (A3) obtain. Then with t_n^m defined for each m to be a y-measurable decision rule in the sample size m problem depending on $\underline{X}_1,\ldots,\underline{X}_n$ and with t_G^m Bayes with respect to G in the sample size m problem (t_G^m is given in (2.9)),

$$(3.2) \qquad t_n^m(y) - t_G^m(y) \xrightarrow{P_y} 0 \quad \text{for all } y \in \mathcal{X},\ 1 \le m \le M$$

implies that the sequence $t_n = t_n^{m_{n+1}}$ is a.o. relative to all G.

Proof: Let G be arbitrary but fixed. By applying (3.1) conditional on $\underline{X}_1,\ldots,\underline{X}_n$ and then completing the expectation, we see that $0 \le R_n(t_n^m,G) - R^m(G) = E(t_n^m(Y) - t_G(Y))^2$. Condition (3.2) implies that

$$(3.3) \qquad t_n^m(Y) - t_G^m(Y) \xrightarrow{P} 0 \quad \text{for each } m,\ 1 \le m \le M \ .$$

Under (A1$^-$), the sequence in (3.3) is bounded and for a bounded sequence, convergence in probability implies convergence in L_2 so we see that

$$(3.4) \qquad R_n(t_n^m,G) - R^m(G) \to 0 \quad \text{for each } m,\ 1 \le m \le M \ .$$

Since $1 \le m_{n+1} \le M < \infty$ for all n under (A3), (3.4) implies that $D_n(G) \to 0$ (where $D_n(G)$ is defined in Definition 1.2) as was to be proved.

This lemma shows that in order to find a.o. rules under (A1$^-$) and (A3) it suffices to be able to approximate t_G^m as $n \to \infty$ for each m. Now under (A2), (2.15) obtains, i.e., for each m

(3.5) $$q_m(y) = \Sigma\, \gamma_k q(y+k)\, ,\ y \in \mathfrak{X}\ ,$$

and from (2.9) the Bayes rule is provided by

(3.6) $$t_G^m(y) = \frac{q_m(y+1)}{q_m(y)}\ ,\ y \in \mathfrak{X}\ .$$

With $\bar{q}$ defined in (2.19),

(3.7) $$\bar{q}_m(y) \equiv \Sigma\, \gamma_k\, \bar{q}(y+k) \overset{P_y}{\to} q_m(y)$$

since $\bar{q}(y+k)$ is consistent for $q(y+k)$ for each k. So in view of Lemma 3.1 the following empirical Bayes procedure is a candidate for an a.o. rule. Let $t_n = t_n^{m_{n+1}}$ where

(3.8) $$t_n^m(y) = \frac{\bar{q}_m^{\,+}(y+1)}{\bar{q}_m^{\,+}(y)} \wedge \beta,\ y \in \mathfrak{X},\ n,m \geq 1\ .$$

<u>Remark 3.1</u>. If G is degenerate at 0, then each $\underline{X}_i$, $i \geq 1$, is degenerate at $\underline{0}$. Then $\bar{q}(y) = 0$ if $y \geq 1$ as required in (2.19) and we see that $\bar{q}_m(y) = 0$ if $y \geq 1$, $m \geq 1$. Hence, t_n^m defined by (3.8) satisfies $t_n^m(y) = 0$ for all $y \in \mathfrak{X}$, $m \geq 1$. Since the Bayes rule (3.6) satisfies $t_G^m(y) = 0$ for all $y \in \mathfrak{X}$, $m \geq 1$, we see that $D_n(G) = 0$ for all n if G is degenerate at 0, i.e., t_n is trivially asymptotically optimal relative to G degenerate at 0. All of the empirical Bayes rules defined in this chapter can be easily shown to be a.o. relative to G degenerate at 0 and we will exclude treatment of this case in the proofs of asymptotic optimality relative to all G. When G is not degenerate at 0, $q_m(y) > 0$ for all y and m.

<u>Theorem 3.1</u>. Under (A2) and (A3) and with $t_n = t_n^{m_{n+1}}$ where t_n^m is given by (3.8), $\{t_n\}$ is a.o. relative to every G.

Proof: Let G and m be fixed, $1 \le m \le M$. By (3.7) and the fact $q_m(y) > 0$, $y \in \mathcal{X}$, we see that

$$(3.9) \qquad \frac{\bar{q}_m^{+}(y+1)}{\bar{q}_m^{+}(y)} \xrightarrow{P}_y \frac{q_m(y+1)}{q_m(y)} \quad , \; y \in \mathcal{X} \; .$$

Since (A2) implies ($A1^-$) (cf. Remark 2.1) we have $0 \le t_G^m(y) \le \beta$, $y \in \mathcal{X}$, so that (3.9) implies

$$(3.10) \qquad t_n^m(y) \overset{P}{\to}_y t_G^m(y) \quad , \; y \in \mathcal{X} \; ,$$

and, hence, (3.2) is satisfied. An application of Lemma 3.1 completes the proof.

In the absence of (A2) the choice of a sequence $\{t_n\}$ which is a.o. relative to every G is more complicated. However, in Section 2.1 we saw that under ($A1^-$), for each m and each $\epsilon > 0$ there exists a polynomial $\Sigma\, \gamma_k \theta^k$ which approximates $z^{m-1}(\theta)$ to within ϵ uniformly in $\theta \in \Theta$. For a determination of such a polynomial, we define

$$(3.11) \qquad \bar{q}_{m,\epsilon}(y) = \Sigma\, \gamma_k\, \bar{q}(y+k) \quad , \; y \in \mathcal{X}$$

where $\bar{q}$ is given in (2.19). Reversing the order of summation in (3.11) we have

$$(3.12) \qquad \bar{q}_{m,\epsilon}(y) = \frac{1}{n} \sum_{i=1}^{n} \Sigma\, \gamma_k\, {}_i q(y+k) , \; y \in \mathcal{X} \; ,$$

where for each i,

$$(3.13) \qquad |\Sigma\, \gamma_k\, {}_i q(y+k)| \le \rho(\epsilon, y) \equiv \bigvee_{m=1}^{M} \Sigma \frac{|\gamma_k|}{g(y+k)}$$

in view of our requirement ${}_i q(y) \le g^{-1}(y)$ in (2.19). Hence,

$$(3.14) \qquad \mathrm{Var}_y(\bar{q}_{m,\varepsilon}(y)) \le \frac{\rho^2(\varepsilon,y)}{n}, \; y \in \mathfrak{X}, \; 1 \le m \le M .$$

Since $E_y\bar{q}_{m,\varepsilon}(y) = q_{m,\varepsilon}(y)$, $y \in \mathfrak{X}$, where $q_{m,\varepsilon}$ is defined by (2.17), we have by (2.18) and (3.14) that

$$(3.15) \qquad E_y(\bar{q}_{m,\varepsilon}(y) - q_m(y))^2 \le \frac{\rho^2(\varepsilon,y)}{n} + \varepsilon^2 q^2(y), \; y \in \mathfrak{X} .$$

The bound $\rho(\varepsilon,y)$ defined in (3.13) is independent of G. With $\varepsilon \to 0$ there exist $n = n(y,\varepsilon)$ such that $n^{-1}\rho^2(\varepsilon,y) \to 0$. By inverting the function for each fixed y we obtain a choice $\varepsilon = \varepsilon(y,n) \to 0$ with $n^{-1}\rho^2(\varepsilon,y) \to 0$. For such choices

$$(3.16) \qquad \bar{q}_{m,\varepsilon}(y) \xrightarrow{P_y} q_m(y), \; y \in \mathfrak{X}, \; 1 \le m \le M .$$

<u>Theorem 3.2</u>. Under ($A1^-$) and (A3) and with $t_n = t_n^{m}{}^{n+1}$ where t_n^m is given by

$$(3.17) \qquad t_n^m(y) = \frac{\bar{q}^{\,+}_{m,\varepsilon}(y+1)}{\bar{q}^{\,+}_{m,\varepsilon}(y)} \wedge \beta, \; y \in \mathfrak{X}, \; 1 \le m \le M,$$

with a choice $\varepsilon = \varepsilon(y,n)$ such that (3.16) obtains, $\{t_n\}$ is a.o. relative to every G.

<u>Proof</u>: Let G and m be fixed, $1 \le m \le M$. By (3.16) and the fact that $q_m(y) > 0$, $y \in \mathfrak{X}$, we see that

$$(3.18) \qquad \frac{\bar{q}^{\,+}_{m,\varepsilon}(y+1)}{\bar{q}^{\,+}_{m,\varepsilon}(y)} \xrightarrow{P_y} \frac{q_m(y+1)}{q_m(y)}, \; y \in \mathfrak{X} .$$

Under ($A1^-$), $0 \le t_G^m(y) \le \beta$ so (3.18) implies

$$(3.19) \qquad t_n^m(y) \xrightarrow{P_y} t_G^m(y), \; y \in \mathfrak{X}$$

and, hence, (3.2) is satisfied. An application of Lemma 3.1 completes the proof.

Theorem 3.2 subsumes Theorem 3.1 for in case $z(\theta)$, $\theta \in \Theta$, is a polynomial a choice corresponding to $\varepsilon = 0$ exists. Theorem 3.1 was presented because of its simple proof and because of its significance in the motivation of Theorem 3.2.

§3.2 Estimation Under (A1) and (A2)

Let these two conditions hold throughout this section. The candidate for an a.o. empirical Bayes rule will be based on t_n^m defined by (3.8). In the absence of (A3) we have found it necessary to examine in greater depth the conditional mean square error of estimation

(3.20) $$E_y(t_n^m(y) - t_G^m(y))^2 .$$

Define ${}_i\hat{q}_m(y) = \Sigma\, \gamma_k\, {}_iq(y+k)$ for each i,m,y so that $\bar{q}_m = \frac{1}{n}\sum_{i=1}^{n} {}_i\hat{q}_m$. Let f' denote the function defined by $f'(y) = f(y+1)$ and fix $m,n,$ and y. Temporarily suppressing the display of the dependence on $m,n,$ and y (e.g. $q_m(y) = q$) we can write (3.20) as

(3.21) $$\int_0^{\beta-t_G} P[\,t > t_G + c]dc^2 + \int_0^{t_G} P[\,t < t_G - c]dc^2 .$$

Since $t = [(\bar{q}')^+/(\bar{q})^+] \wedge \beta$, we see that

(3.22) $$[\,t > t_G + c] \le [\bar{q} \le 0] + [\bar{q}' - (t_G + c)\bar{q} > 0] .$$

We can write

(3.23) $$P[\bar{q}' - (t_G + c)\bar{q} > 0] = P[\bar{\omega} > 0]$$

where

(3.24) $$\bar{\omega} = \frac{1}{n}\sum_{i=1}^{n} {}_i\omega$$

and

(3.25) $${}_i\omega = {}_i\hat{q}' - (t_G + c)\,{}_i\hat{q},\ i = 1,2,\ldots .$$

In preparation for bounding the tail probability (3.23) note that, since $t_G = q'/q$ and $E\ {}_i\hat{q} = q$, we have

$$E(\bar{\omega}) = -cq \ . \tag{3.26}$$

From (3.25) and the bound for ${}_iq$ required in (2.19), we have for $c \in [0, t_G - \beta]$ and for each i that

$$|{}_i\omega| \le \Sigma|\gamma_k|\left(\frac{1}{g(y+1+k)} + \frac{\beta}{g(y+k)}\right) \equiv \rho \ . \tag{3.27}$$

By (2.3) of Theorem 1 of Hoeffding (1963) we have

$$P[\bar{\omega} \ge 0] \le \exp\{-2n[\frac{cq}{2\rho}]^2\} \tag{3.28}$$

and hence

$$\int_0^{\beta - t_G} P[\bar{\omega} \ge 0]dc^2 \le \frac{2\rho^2}{nq^2} \ . \tag{3.29}$$

A similar treatment of $P[\bar{q} \le 0]$ yields

$$P[\bar{q} \le 0] \le \exp\{-2n[\frac{\beta q}{2\rho}]^2\} \tag{3.30}$$

where use is made of the fact that ${}_i\hat{q} \le \rho/\beta$. Combining (3.22), (3.29), and (3.30) we obtain

$$\int_0^{\beta - t_G} P[t > t_G + c]dc^2 \le \frac{2\rho^2}{nq^2} + \beta^2 \exp\{\frac{-n\beta^2 q^2}{2\rho^2}\} \equiv \tfrac{1}{2}B \ . \tag{3.31}$$

The same bound holds for the second term of (3.21) so that by (3.20) and (3.21)

$$E_y(t_n^m(y) - t_G^m(y))^2 \le B_n(m,y) \tag{3.32}$$

where, with $\rho(m,y)$ defined in (3.27), $B_n(m,y)$ is defined in (3.31) with the dependence on m,n and y now displayed.

Lemma 3.2. Let N be any function from $\mathcal{I}$ to $\mathcal{I}$ such that $N(M) \to \infty$ as $M \to \infty$. There exists a sequence $\{M_n\}$ independent of G such that

$$(3.33) \qquad B_n(N) \equiv \bigvee_{m \le M_n} \bigvee_{y \le N(M_n)} B_n(m,y) \to 0 \quad \text{as} \quad n \to \infty .$$

Proof: Let N be fixed such that $N(M) \to \infty$ as $M \to \infty$. For each M, let $n = n(M)$ be any increasing sequence of integers independent of G such that

$$(3.34) \qquad \bigvee_{m \le M} \bigvee_{y \le N(M)} B_n(m,y) \to 0 \quad \text{as} \quad M \to \infty .$$

Inverting $n(M)$ to obtain $M(n)$ will allow a choice of a sequence $M_n = M(n)$ independent of G such that (3.33) obtains. To see that such a sequence $n(M)$ independent of G exists, note that $q_m(y) \ge z^m(\beta)\int_\Theta \theta^y G(d\theta) \ge z^m(\beta)\mu^y$ where $\mu = \int_\Theta \theta \, G(d\theta)$. Then

$$(3.35) \qquad \bigwedge_{m \le M} \bigwedge_{y \le N(M)} q_m(y) \ge [z(\beta) \wedge 1]^M [\mu \wedge 1]^{N(M)} .$$

With

$$(3.36) \qquad \rho_M \equiv \bigvee_{m \le M} \bigvee_{y \le N(M)} \rho(m,y) ,$$

we see by the definition of B in (3.31) and by (3.35) and (3.36) that any choice $n = n(M)$ such that

$$(3.37) \qquad n^{\frac{1}{2}}[z(\beta) \wedge 1]^M [\mu \wedge 1]^{N(M)} \rho_M^{-1} \to \infty \quad \text{as} \quad M \to \infty$$

will ensure that (3.34) obtains. Any choice

$$(3.38) \qquad n = n(M) = [z(\beta) \wedge 1]^{-2M} \rho_M^2 \exp\{aN^b(M)\}$$

where a and b are constants, $a > 0$, $b > 1$, is independent of G and guarantees (3.37) regardless of the value of μ. Hence the proof of the lemma is complete.

We are ready to define a candidate for an a.o. rule. Let L^m be decision rules such that $R(L^m,G) \to 0$ as $m \to \infty$ for every G. Such a choice is possible as was seen in Section 2.3. Let $\{M_n\}$ be any sequence of positive integers $\to \infty$ as $n \to \infty$. For each n define t_n by

$$(3.39) \qquad t_n = L^{m_{n+1}}[m_{n+1} > M_n] + t_n^{m_{n+1}}[m_{n+1} \le M_n]$$

where for each m, t_n^m is defined in (3.8).

Theorem 3.3. Under (A1) and (A2), with N any function satisfying the hypothesis of Lemma 3.2 such that $P_{\beta,k}[Y > N(k)] \to 0$ as $k \to \infty$, and with t_n defined by (3.39) with $\{M_n\}$ chosen independent of G so that (3.33) obtains, the rule $\{t_n\}$ is a.o. relative to every G.

Proof: Let G be fixed. For $m_{n+1} > M_n$,

$$(3.40) \qquad 0 \le D_n(G) \le R(L^{m_{n+1}},G) \to 0 \quad \text{as} \quad n \to \infty$$

since $M_n \to \infty$. For $m = m_{n+1} \le M_n$, by applying (3.1) conditional on $\underline{X}_1,\ldots,\underline{X}_n$ and then completing the expectation,

$$(3.41) \qquad 0 \le D_n(G) = E(t_n^m(Y) - t_G^m(Y))^2 .$$

Since $(t_n^m(y) - t_G^m(y))^2 \le \beta^2$ for each m,y, and n, the right hand side of (3.41) is bounded (with $N_n = N(M_n)$) by

$$(3.42) \qquad \sum_{y=0}^{N_n} E_y(t_n^m(y) - t_G^m(y))^2 h_m(y) + \beta^2 P[Y > N_n]$$

which in turn is bounded by

$$(3.43) \qquad B_n(N) + \beta^2 P_{\beta,M_n}[Y > N_n]$$

from (3.32), the definition of $B_n(N)$ in (3.33), and Lemma 2.1. Since (3.33) obtains for $\{M_n\}$ and N, the first term in (3.43) $\to 0$, while the second term in (3.43) $\to 0$ by the definition of the function N and the proof is complete.

This section is subsumed by the succeeding section in the same manner as Theorem 3.1 was subsumed by Theorem 3.2. However we have included this section for its significance in motivating the development of a.o. rules in the succeeding section.

An earlier construction used weaker bounds on the conditional mean square error of estimation and required the imposition of an assumption

$$(A1^{+}) \qquad \Theta \subset [\alpha,\beta] = \mathcal{A},\ \alpha > 0,\ \beta \in \Pi$$

in order to determine the choice $\{M_n\}$ independent of G. Professor James Hannan observed that an application of Hoeffding's Theorem 1 yielded a bound from which a construction could be accomplished under (A1).

§3.3 Estimation Under (A1)

Let (A1) hold throughout this section. The candidate for an a.o. empirical Bayes rule will be based on t_n^m defined in (3.17). For $m \geq 1$ and $\epsilon \geq 0$, define

$$t_{G,\epsilon}^m = \frac{q'_{m,\epsilon}}{q_{m,\epsilon}} \tag{3.44}$$

where $q_{m,\epsilon}$ is defined by (2.17) and $f'(y) = f(y+1)$ for any function f. For each m and ϵ,

$$|t_{G,\epsilon}^m - t_G^m| = q_{m,\epsilon}^{-1} \{|q'_{m,\epsilon} - q'_m| + \beta|q_{m,\epsilon} - q_m|\} . \tag{3.45}$$

Fixing m and taking $2\epsilon < z^{m-1}(\beta)$ fixed,

$$q_m \geq z^{m-1}(\beta)q \tag{3.46}$$

and from (3.46), (2.18), and the choice of ϵ

$$q_{m,\epsilon} \geq q_m - \epsilon q \geq \tfrac{1}{2} z^{m-1}(\beta)q . \tag{3.47}$$

Hence from (2.18), (3.47), and the fact that $q' \leq \beta q$, the right hand side of (3.45) is bounded by

$$4\beta \epsilon[z^{m-1}(\beta)]^{-1} . \tag{3.48}$$

For the choice of ϵ, we see

$$0 < t_{G,\epsilon}^m \leq 3\beta \tag{3.49}$$

where use is also made of (3.47), (3.48), and the fact that $0 \leq t_G^m \leq \beta$ under (A1). Now define

$$T_{n,\epsilon}^m = [(\bar{q}'_{m,\epsilon})^+/(\bar{q}_{m,\epsilon})^+] \wedge 3\beta \tag{3.50}$$

with $\bar{q}_{m,\varepsilon}$ defined in (3.11) and note that

$$(3.51) \qquad (\tau^m_{n,\varepsilon} - t^m_{G,\varepsilon})^2 \geq (t^m_n - t^m_{G,\varepsilon})^2$$

where t^m_n is defined by (3.17). Following the same procedure leading to (3.32), we have that for each m,y,n, and $\varepsilon < z^{m-1}(\beta)/2$,

$$(3.52) \qquad E_y(\tau^m_{n,\varepsilon}(y) - t^m_{G,\varepsilon}(y))^2 \leq B^*_n(m,y,\varepsilon) \equiv$$

$$\frac{4\rho^2(m,y,\varepsilon)}{nq^2_{m,\varepsilon}(y)} + 18\beta^2 \exp\{\frac{-9\beta^2 n\, q^2_{m,\varepsilon}(y)}{2\,\rho^2(m,y,\varepsilon)}\}$$

where

$$(3.53) \qquad \rho(m,y,\varepsilon) \equiv \Sigma|\gamma_k|\left(\frac{1}{g(y+1+k)} + \frac{3\beta}{g(y+k)}\right) .$$

From (3.51) and (3.52), we have

$$(3.54) \qquad E_y(t^m_n(y) - t^m_{G,\varepsilon}(y))^2 \leq B^*_n(m,y,\varepsilon)$$

for each m,n,y, and $\varepsilon < z^{m-1}(\beta)/2$.

<u>Lemma 3.3</u>. Let N be any function from $\mathcal{I}$ to $\mathcal{I}$ such that $N(M) \to \infty$ as $M \to \infty$. There exist sequences $\{M_n\}$ and $\{\varepsilon_n\}$ independent of G such that

$$(3.55) \qquad B^*_n(N) \equiv \bigvee_{m\leq M_n} \; \bigvee_{y\leq N(M_n)} B^*_n(m,y,\varepsilon_n) \to 0 \quad \text{as} \quad n \to \infty$$

and

$$(3.56) \qquad \varepsilon_n\left(\bigwedge_{m\leq M_n} z^{m-1}(\beta)\right)^{-1} \to 0 \quad \text{as} \quad n \to \infty .$$

<u>Proof</u>: Let N be fixed such that $N(M) \to \infty$ as $M \to \infty$. For each M, let $\varepsilon(M)$ be any null sequence such that $\varepsilon(M)\left(\bigwedge_{m\leq M} z^{m-1}(\beta)\right)^{-1} \to 0$ as $M \to \infty$. Let $n = n(M)$ be any increasing sequence of

integers independent of G such that

$$\bigvee_{m\le M}\ \bigvee_{y\le N(M)} B_n^*(m,y,\varepsilon(M)) \to 0 \quad \text{as} \quad M \to \infty . \tag{3.57}$$

Inverting $n(M)$ to obtain $M(n)$ will allow a choice of sequences $M_n = M(n)$ and $\varepsilon_n = \varepsilon(M_n)$ independent of G such that (3.55) and (3.56) obtain. Again such a choice of $n(M)$ independent of G is possible. Without loss of generality, $\varepsilon(M) \le \frac{1}{2}(\bigwedge_{m\le M} z^{m-1}(\beta))$ so that from (3.47), $q_{m,\varepsilon(M)} \ge \frac{1}{2} z^{m-1}(\beta)q \ge \frac{1}{2} z^m(\beta)\mu^y$ where $\mu = \int_\Theta \theta\, G(d\theta)$. Then

$$\bigwedge_{m\le M}\ \bigwedge_{y\le N(M)} q_{m,\varepsilon(M)}(y) \ge \tfrac{1}{2}[z(\beta) \wedge 1]^M[\mu \wedge 1]^{N(M)} . \tag{3.58}$$

With

$$\rho_M^* \equiv \bigvee_{m\le M}\ \bigvee_{y\le N(M)} \rho(m,y,\varepsilon(M)), \tag{3.59}$$

we see as in the proof of Lemma 3.2 that any choice

$$n = n(M) = [z(\beta) \wedge 1]^{-2M}(\rho_M^*)^2 \exp\{aN^b(M)\} \tag{3.60}$$

where a and b are constants, $a > 0$, $b > 1$, is independent of G and guarantees (3.57). Hence the proof is complete.

Define a candidate for an a.o. rule by letting L^m be decision rules such that $R(L^m,G) \to 0$ as $m \to \infty$ for every G. Let $\{M_n\}$ be any sequence of positive integers $\to \infty$ as $n \to \infty$ and let $\{\varepsilon_n\}$ be any null sequence. For each n, define t_n by

$$t_n = L^{m_{n+1}}[m_{n+1} > M_n] + t_n^{m_{n+1}}[m_{n+1} \le M_n] \tag{3.61}$$

where for each m and ε, t_n^m is defined by (3.17).

<u>Theorem 3.4</u>. Under (A1), with N any function satisfying the hypotheses of Lemma 3.3 such that $P_{\beta,k}[Y > N(k)] \to 0$ as $k \to \infty$, and with t_n defined by (3.61) with $\{M_n\}$ and $\{\epsilon_n\}$ chosen independent of G so that (3.55) and (3.56) obtain, the rule $\{t_n\}$ is a.o. relative to every G.

<u>Proof</u>: Let G be fixed. For $m_{n+1} > M_n$, (3.40) holds since $M_n \to \infty$. For $m = m_{n+1} \le M_n$, (3.41) holds and its right hand side is bounded by

$$(3.62) \qquad 2 \sum_{y=0}^{N_n} E_y(t_n^m(y) - t_{G,\epsilon_n}^m(y))^2 h_m(y) +$$

$$2 \sum_{y=0}^{N_n} (t_{G,\epsilon_n}^m(y) - t_G^m(y))^2 h_m(y) + \beta^2 P_{\beta,M_n}[Y > N_n]$$

with $N_n = N(M_n)$ by use of the c_r-inequality (cf. Loève (1963, p. 155)), the fact that $(t_n^m - t_G^m) \le \beta^2$, and Lemma 2.1. The third term of (3.62) $\to 0$ as $n \to \infty$ by the choice of N. Without loss of generality, $\epsilon_n \le \frac{1}{2} \bigwedge_{m \le M_n} z^{m-1}(\beta)$ so that from (3.54) and (3.48) the first two terms of (3.62) are bounded by

$$(3.63) \qquad 2B_n^*(N) + 8\beta\, \epsilon_n\Big(\bigwedge_{m \le M_n} z^{m-1}(\beta)\Big)^{-1}$$

which $\to 0$ as $n \to \infty$ since (3.55) and (3.56) obtain. Hence the proof is complete.

CHAPTER IV

TESTING AND FINAL REMARKS

§4.1 Testing Under ($A1^-$) and (A3)

In this chapter we will exhibit sequences of decision rules which are a.o. relative to every G in the case of linear loss testing in the special discrete exponential families described in Section 2.1. Recall from that section that a Bayes rule with respect ot G in the sample size m component problem is provided by (2.11) and (2.12). Lemma 1 of Johns and Van Ryzin (1971, p. 1524) establishes the useful inequality for each m

$$(4.1) \qquad 0 \le R_n(\delta_n^m,G) - R^m(G) \le b \sum_{y=0}^{\infty} |a_G^m(y)| g_m(y) P_y[|a_n^m(y) - a_G^m(y)| \ge |a_G^m(y)|]$$

for any empirical Bayes procedure δ_n^m defined by

$$(4.2) \qquad \delta_n^m(y) = [a_n^m(y) \le 0],\ y \in \mathcal{X}\ ,$$

where $a_n^m(y)$ is a function of $\underline{X}_1,\ldots,\underline{X}_n$.

The following lemma proves useful in establishing the asymptotic optimality of specified sequences of decision rules when assumption (A3) obtains.

<u>Lemma 4.1</u>. Suppose δ_n^m is defined for each m to be a y-measurable decision rule in the sample size m problem of the form (4.2)

where $a_n^m(y)$ is a function of $\underline{X}_1,\ldots,\underline{X}_n$ for each $y \in \mathfrak{X}$. Under (A3) and with G any prior such that $\int_\Theta \theta\, G(d\theta) < \infty$,

$$(4.3) \qquad a_n^m(y) \xrightarrow{P_y} a_G^m(y) \quad \text{for all } y \in \mathfrak{X},\ 1 \le m \le M,$$

implies that the sequence $\delta_n = \delta_n^{m_{n+1}}$ is a.o. relative to G.

Proof: Let G be fixed such that $\int_\Theta \theta\, G(d\theta) < \infty$. Fix m, $1 \le m \le M$. Since $1 \ge [a_G^m(y) \ne 0]P_y[|a_n^m(y) - a_G^m(y)| \ge |a_G^m(y)|] \to 0$ for each $y \in \mathfrak{X}$ by (4.3) and since $\sum_{y=0}^{\infty} |a_G^m(y)| g_m(y) \le \int_\Theta (\theta+c)G(d\theta) < \infty$, the dominated convergence theorem applies to show that the right hand side of (4.1) $\to 0$ and hence

$$(4.4) \qquad R_n(\delta_n^m, G) - R^m(G) \to 0 \ .$$

Since $1 \le m_{n+1} \le M$ for all n under (A3), (4.4) implies that $D_n(G) \to 0$ as was to be proved.

Theorem 4.1. Under (A1$^-$) and (A3) and with

$$(4.5) \qquad a_n^m(y) = \bar{q}_{m,\varepsilon}(y+1) - c\, \bar{q}_{m,\varepsilon}(y),\ y \in \mathfrak{X},\ m \ge 1,$$

where $\bar{q}_{m,\varepsilon}$ is defined in (3.11) with a choice of $\varepsilon = \varepsilon(y,n)$ such that (3.16) obtains, the sequence $\delta_n = \delta_n^{m_{n+1}}$ with δ_n^m defined by (4.2) is a.o. relative to all G.

Proof: Let G be fixed. Under (A1$^-$), Θ is bounded so that $\int_\Theta \theta\, G(d\theta) < \infty$. From (3.16) it follows that (4.3) is satisfied and so an application of Lemma 4.1 completes the proof.

§4.2 Testing Under (A1)

Let (A1) obtain throughout this section. With $f'(y) = f(y+1)$ for any function f, define

$$(4.6) \qquad a^m_{G,\epsilon} = g'_{m,\epsilon} - c\, q_{m,\epsilon}\,,$$

for each m and ϵ where $q_{m,\epsilon}$ is defined by (2.17). With a^m_G defined by (2.12),

$$(4.7) \qquad |a^m_{G,\epsilon} - a^m_G| \le |q'_{m,\epsilon} - q'_m|$$

$$+ c|q_{m,\epsilon} - q_m| \le \epsilon\, q' + c\, \epsilon\, q$$

from (2.18). In light of the fact that $q' \le \beta q$, the right hand side of (4.7) is bounded by

$$(4.8) \qquad \epsilon(\beta + c)q\ .$$

For each m and ϵ, define

$$(4.9) \qquad a^m_{n,\epsilon} = \bar{q}'_{m,\epsilon} - c\,\bar{q}_{m,\epsilon},\ n \ge 1,$$

where $\bar{q}_{m,\epsilon}$ is defined by (3.11). Reversing the order of summation in (4.9) yields

$$(4.10) \qquad a^m_{n,\epsilon}(y) = \frac{1}{n} \sum_{i=1}^{n} \Sigma\, \gamma_k({}_i q(y+1+k) - c\ {}_i q(y+k)),\ y \in \mathcal{X},\ n \ge 1,$$

where for each i

$$(4.11) \qquad |\Sigma\, \gamma_k({}_i q(y+1+k) - c\ {}_i q(y+k))| \le \rho(\epsilon,m,y) \equiv$$

$$\Sigma|\gamma_k|\left(\frac{1}{g(y+1+k)} + \frac{c}{g(y+k)}\right)$$

in view of our requirement that ${}_1q \le g^{-1}$ in (2.19). Noting that $E_y(a^m_{n,\varepsilon}(y)) = a^m_{G,\varepsilon}(y)$, $y \in \mathfrak{X}$, $n \ge 1$, we have from (4.10) and (4.11)

$$(4.12) \qquad E_y(a^m_{n,\varepsilon}(y) - a^m_{G,\varepsilon}(y))^2 \le \frac{\rho^2(\varepsilon,m,y)}{n}, \; y \in \mathfrak{X}, \; n \ge 1 .$$

Since

$$[|a^m_{n,\varepsilon} - a^m_G| \ge |a^m_G|] \le [|a^m_{n,\varepsilon} - a^m_{G,\varepsilon}| \ge \tfrac{1}{2}|a^m_G|] +$$
$$[|a^m_{G,\varepsilon} - a^m_G| \ge \tfrac{1}{2}|a^m_G|], \; n \ge 1,$$

the summand on the right hand side of (4.1) with $a^m_n = a^m_{n,\varepsilon}$ is bounded for each n,m,ε, and y by

$$(4.13) \quad |a^m_G(y)|g_m(y)P_y[|a^m_{n,\varepsilon}(y) - a^m_{G,\varepsilon}(y)| \ge \tfrac{1}{2}|a^m_G(y)|]$$
$$+ 2g_m(y)|a^m_{G,\varepsilon}(y) - a^m_G(y)| .$$

In view of (4.7), (4.8) and the fact that $g_m q \le [z^{m-1}(\beta)]^{-1}h_m$, the second term of (4.13) is bounded by

$$(4.14) \qquad 2(\beta+c)\varepsilon[z^{m-1}(\beta)]^{-1}h_m(y) .$$

Using a Markov bound on the first term of (4.13), this term is bounded by

$$(4.15) \qquad 2g_m(y)[E_y(a^m_{n,\varepsilon}(y) - a^m_{G,\varepsilon}(y))^2]^{\frac{1}{2}},$$

which, in light of (4.12), is itself bounded by

$$(4.16) \qquad 2g_m(y)n^{-\frac{1}{2}}\rho(\varepsilon,m,y) .$$

Combining (4.13), (4.14) and (4.16), the first N terms of the right hand side of (4.1) with $a_n^m = a_{n,\epsilon}^m$ is bounded by

$$(4.17) \qquad 2n^{-\frac{1}{2}}b \sum_{y=0}^{N} g_m(y)\rho(\epsilon,m,y) + 2b(\beta+c)\epsilon[z^{m-1}(\beta)]^{-1} .$$

The bound (4.17) motivates the following lemma.

<u>Lemma 4.1</u>. With N any function from $\mathcal{I}$ to $\mathcal{I}$, there exist sequences M_n and ϵ_n independent of G such that

$$(4.18) \qquad \rho_n(N) \equiv n^{-\frac{1}{2}} \bigvee_{m\le M_n} \sum_{y=0}^{N(M_n)} g_m(y)\rho(\epsilon_n,m,y) \to 0 \quad \text{as} \quad n \to \infty$$

and

$$(4.19) \qquad \rho_n^* \equiv \epsilon_n(\bigwedge_{m\le M_n} z^{m-1}(\beta))^{-1} \to 0 \quad \text{as} \quad n \to \infty .$$

<u>Proof</u>: For each M, let $\epsilon(M)$ be any null sequence such that $\epsilon(M)(\bigwedge_{m\le M} z^{m-1}(\beta))^{-1} \to 0$ as $M \to \infty$. Since $\rho(M,N) \equiv \bigvee_{m\le M} \sum_{y=0}^{N(M)} g_m(y)\rho(\epsilon(M),m,y)$ is independent of G, $n = n(M)$ can be chosen independent of G such that $n^{-\frac{1}{2}}\rho(M,N) \to 0$ as $M \to \infty$. Inverting $n(M)$ to obtain $M(n)$ allows the choices of $M_n = M(n)$ and $\epsilon_n = \epsilon(M_n)$ independent of G such that (4.18) and (4.19) obtain.

Now let L^m be any decision rules such that $R(L^m,G) \to 0$ as $m \to \infty$ for every G. Such a choice is possible as was seen in Section 2.3. Then for any sequences $\{M_n\}$ and $\{\epsilon_n\}$, define

$$(4.20) \qquad \delta_n = L^{m_{n+1}}[m_{n+1} > M_n] + \delta_n^{m_{n+1}}[m_{n+1} \le M_n]$$

where, for each m and n, δ_n^m is defined by (4.2) with

$$(4.21) \qquad a_n^m(y) = a_{n,\epsilon_n}^m(y), \; y \in \mathcal{I}, \; m,n \ge 1 ,$$

where $a^m_{n,\varepsilon}$ is defined by (4.9).

<u>Theorem 4.2</u>. Under (A1) and with N a function from $\mathfrak{X}$ to $\mathfrak{X}$ defined such that

$$(4.22) \qquad P_{\beta,k}[Y > N(k)] \to 0 \quad \text{as} \quad k \to \infty$$

and with δ_n defined by (4.20) with $\{M_n\}$ and $\{\varepsilon_n\}$ chosen independent of G such that (4.18) and (4.19) obtain, the rule $\{\delta_n\}$ is a.o. relative to every G.

<u>Proof</u>: Let G be fixed. For $m_{n+1} > M_n$,

$$(4.23) \qquad 0 \le D_n(G) \le R(L^{m_{n+1}}, G) \to 0 \quad \text{as} \quad n \to \infty$$

since $M_n \to \infty$. For $m = m_{n+1} \le M_n$, $0 \le D_n(G)$ is bounded by the right hand side of (4.1). In light of (4.17), the fact that $g_m(y)|a^m_G(y)| \le (\beta+c)h_m(y)$ for each m and y, and Lemma 2.1, the right hand side of (4.1) is bounded by

$$(4.24) \qquad 2b\,\rho_n(N) + 2b(\beta+c)\rho^*_n + (\beta+c)P_{\beta,M_n}[Y > N(M_n)]$$

which $\to 0$ as $n \to \infty$ from (4.18), (4.19), and (4.22) by the choice of N, $\{M_n\}$, and $\{\varepsilon_n\}$. Hence the proof is complete.

§4.3 Final Remarks

The rules presented in Chapters III and IV have several competitors. The first competitor which we shall discuss arises in the following manner. Suppose for each m that, in an empirical Bayes problem involving repetitions of a sample size m component problem, there is a sequence $\{\Psi_n^m\}$ of rules which is a.o. relative to every G. One might then consider for use in the corresponding varying sample size empirical Bayes problem, a procedure $\{\varphi_n\}$ which partitions the problem into those involving a common sample size and uses the appropriate $\{\Psi_n^m\}$ within each class, i.e.,

$$(4.25) \qquad \varphi_n = \Psi^m_{k(n,m)}[m_{n+1} = m] \quad \text{where} \quad k(n,m) \equiv \sum_{i=1}^{n} [m_i = m] \ .$$

Such a rule was suggested by Professor Hannan as a first thought as to what could be done in the varying sample size empirical Bayes problem. Under (A3), $\{\varphi_n\}$ will be a.o. relative to every G. However such a rule does not use all of the past data at each stage and so we could say that it does not use all of the available "information about G". So intuitively at least, when dealing with the particular component problem of Chapter II, our method seems better in the sense that it uses all of the past data available. In the absence of (A3) the rule (4.25) would not be a.o. since a new sample size can then appear infinitely often. Professor Hannan has also suggested more sophisticated estimators of q_m based on averaging across m-tuples within $\underline{X}_i$, $m_i \geq m$, which use more of the available data.

One can use $\{\varphi_n\}$ to obtain ϵ-asymptotic optimality relative to every G by employing the device used in Sections 3.2, 3.3, and 4.2. For if L^m is a decision rule in the sample size m problem such that $R(L^m,G) \le \epsilon$ if $m \ge M$ for every G, then

$$\varphi_n^* = \varphi_n[m_{n+1} \le M] + L^{m_{n+1}}[m_{n+1} > M]$$

will result in $\overline{\lim}_n D_n(G) \le \epsilon$ for every G. It is possible that we might obtain asymptotic optimality by replacing M by a proper choice of $M_n \to \infty$ but this idea has not been fully explored.

The second competitor that we look at is one which would arise from the first track discussed in Section 1.3. We will discuss this competitor in the context of the special component problem of Chapter II. For an estimator $\hat{G}_n$ based on $\underline{X}_1,\ldots,\underline{X}_n$ and taking values on the set of probability distributions on $[0,\beta]$, let

$$\text{(4.26)} \qquad \hat{q}_m(y) \equiv \int_{\Theta} \theta^y z^m(\theta)\hat{G}_n(d\theta) \ .$$

Since $\theta^y z^m(\theta)$ is bounded and continuous in $\theta \in \Theta$, the Helly Bray Lemma (cf. Loève (1963, p. 180)) implies that $\hat{q}_m(y) \xrightarrow{a.s.} q_m(y)$ for each $y \in \mathfrak{X}$ if $\hat{G}_n \to G$ in distribution a.s. Tucker (1963), Rolph (1968), and Meeden (1972) have demonstrated the existence of such estimators $\hat{G}_n$ in the case of identical sample sizes. We saw in the proofs of Theorems 3.1, 3.2, and 4.1 that rules based on consistent estimators of q_m can easily be shown to be a.o relative to every G under (A3). Again in the absence of (A3) the rules based on $\hat{q}_m$ might possibly be extended to a.o. rules by choice of rules L^m and a sequence $\{M_n\}$.

The assumption that the parameter space Θ is bounded is a stronger assumption than what is needed to prove asymptotic optimality in the particular case of the component problem of Chapter II where the sample sizes are identical in the empirical Bayes problem. Macky (1966) and Hannan and Macky (1971) have demonstrated a.o. procedures for estimation when $\Theta = [0,\infty)$ while Robbins (1963) and Samuel (1963) have demonstrated a.o. procedures for testing when $\Theta \subset [0,\infty)$ relative to every G for which $\int_\Theta \theta\, G(d\theta) < \infty$. So one would hope that in the future methods for establishing a.o. procedures in the varying sample size empirical Bayes problem could be found without restricting Θ to a bounded set.

BIBLIOGRAPHY

BIBLIOGRAPHY

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

Hannan, James and Macky, David W. (1971). Empirical Bayes squared error loss estimation of unbounded functionals in exponential families. Unpublished.

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

Lehmann, E.L. (1959). Testing Statistical Hypotheses. John Wiley & Sons, New York.

Loève, Michel (1963). Probability Theory. Third Edition, Van Nostrand, Princeton.

Johns, M.V., Jr. (1957). Non-parametric empirical Bayes procedures. Ann. Math. Statist. 28 649-669.

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

Macky, David W. (1966). Empirical Bayes estimation in an exponential family. RM-176, Department of Statistics and Probability, Michigan State University.

Maritz, J.S. (1970). Empirical Bayes Methods, Methuen and Co. Ltd., London.

Meeden, Glen (1972). Bayes estimation of the mixing distribution, the discrete case. Presented at 133rd Meeting of the Institute of Mathematical Statistics, Ames, Iowa, April 1972.

Neyman, J. (1962). Two breakthroughs in the theory of statistical decision making. Rev. Inst. Internat. Statist. 30 11-27.

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

Robbins, H. (1963). The empirical Bayes approach to testing statistical hypotheses. Rev. Inst. Internat. Statist. 31 195-208.

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

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

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

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