

Multiple Decision Procedures from an  
Empirical Bayes Approach

by

John J. Deely

Supported by the Office of Naval Research  
Contract NONR-1100(26) and the  
Aerospace Research Laboratories Contract AF 33(657)11737

Shanti S. Gupta, Project Director  
Purdue University

Reproduction in Whole or in Part is Permitted for  
any Purpose of the United States Government

Department of Statistics  
Division of Mathematical Statistics

Mimeograph Series No. 45

July, 1965

## INTRODUCTION

Suppose one encounters  $k$ -populations (categories, varieties, processes, candidates, etc.) and for each population one can observe a random variable whose distribution depends upon an unknown parameter  $\theta_i$ . This parameter may be the mean, variance, some quantile, or a function of these quantities. The "best" population is defined as that population with largest  $\theta_i$  (or equivalently the smallest) and hence based upon " $r$ " observations from each population, one desires an "optimal" decision procedure which will either: (1) select the "best" population, or (2) select a subset of the  $k$ -populations which contains the "best". As thus formulated this is called the multiple decision problem. See Bechhofer [4], Gupta [16], and Lehmann [22]. Such decision procedures are often called "ranking and selection" procedures.

The need to consider such a problem arose due to the shortcomings in the classical tests of homogeneity, i.e. testing the hypothesis of equality of parameters. For, if in the final analysis a ranking of the parameters is desired as is often the case, then to view the problem from the beginning as a "multiple decision problem" is a more realistic and meaningful approach than a test of the hypothesis of their equality. Bahadur [1] was one of the earliest authors to recognize this and to contribute to the theory of the  $k$ -sample problems. Many authors have since made contributions to various aspects and modifications of this basic problem. For further amplification reference could be made to Bechhofer [4], Gupta [9], Gupta and Sobel [10], Lehmann [22]. Until recently most contributions could be classified into one of two categories: (1) "selecting a subset" formulation or (2) the "indifference zone" formulation for selecting the best.

In (1) decision procedures are given which select a subset of the  $k$ -populations subject to the requirement that the probability of selecting the "best" population in the subset is at least a prescribed number  $P^*$ , regardless of the possible configurations of the parameters. The size of such selected subsets is of course a random variable and thus among decision procedures which give the required  $P^*$  condition, the one which gives the smallest "expected" size is considered in some sense desirable. Expected minimal rank and expected sum of ranks of the populations selected in the subset are also possible performance criteria for comparing various decision rules. Contributions to this problem have been made by Paulson [26]; Gupta [9], [14], [15]; Gupta and Sobel [10], [11], [12], [13]; and Seal [33], [34].

In the "indifference zone" formulation for selecting the best population, a single population is selected so as to guarantee with probability  $P^*$  that the selected population is best provided some other condition on the parameters is satisfied. This predesignated condition is thought of as a "width" of an "indifference zone" in the parameter space (see Bechhofer [4]), and thus requires the experimenter to give two quantities:  $P^*$  and this "width". Other contributions to this problem are Bechhofer and Sobel [5]; Bechhofer, Sobel, and Dunnett [6]; Sobel and Huyett [35]. Chambers and Jarratt [7] considered this problem with the assumption of an "indifference zone" replaced with assuming that the variance of each population is a known function of the unknown mean, this function being the same for each population.

Multiple decision problems have also been investigated by Dunnett [8] in which a specific "a priori" distribution is assumed on the parameter space. Guttman and Tiao [17] also considered a "selecting the best"

problem from this perspective. This approach is the so called Bayesian approach to the multiple decision problem. Whereas this is a wide divergence of opinion over the philosophical implications of such an assumption, it is not the purpose of this paper to enter such discussions. However, from a practical point of view, one can find situations in which the assumption of an "a priori" distribution seems reasonable. (See [8] and [25]). This assumption becomes significantly more realistic if further one assumes only the existence of an "a priori" distribution, the exact distribution itself remaining unknown. It is from this perspective that Robbins [28], [29], [30] has introduced the concept, empirical Bayes decision procedures. This technique has been called a "breakthrough" by Neyman [25] and has been used by several authors, Johns [18], Kagan [19], Krutchkoff [21], Miyasawa [24], Samuel [32]. It is the purpose of this paper to extend this technique to the multiple decision problem.

In the previous work done on the multiple decision problem, decision procedures were given somewhat intuitively and then desirable characteristics rigorously developed. The approach here will be to use a decision theoretic framework with a specific loss structure and then derive procedures which minimize the "overall expected" loss. (i.e. the Bayes risk). In Chapter I the problem is stated in decision theoretic terms after basic definitions and notations are given. In Chapter II the subset problem is shown to be intimately related to the "selecting the best" problem, and hence decision procedures for the subset problem arise immediately from those procedures which select the "best". A summary of the results in Chapters III and IV are thus included. In Chapters III and IV the existing prior distribution  $G$  is assumed to be a member of a specific parametric class. In Chapter III a theorem is proved relating empirical Bayes

procedures to the Bayes procedure with respect to  $G$ . Then Bayes procedures are found for the cases in which the a priori distribution is (1) normal, (2) uniform, (3) beta, and (4) gamma. In Chapter IV empirical Bayes procedures are obtained for the above cases using the theorem proved in Chapter III. Finally, in Chapter V a theorem of Robbins [30] is used to derive empirical Bayes procedures for selecting the best for various densities: (1) normal, (2) Poisson, and (3) of the form  $\theta^x g(x)h(\theta)$ ; and each  $G_i$  satisfying only the condition that it has finite absolute first moment. It is in this sense that the results are called non-parametric.

## CHAPTER I

## BASIC DEFINITIONS AND NOTATIONAL PRELIMINARIES

## A. Definition of the Problem

Suppose each of  $k$ -populations has an observable random variable with density  $f(x|\theta_i)$  where  $\theta_i$  is a parameter belonging to some prescribed set  $\mathbb{H}$ , a subset of the real numbers. Then  $\underline{\theta} = (\theta_1, \dots, \theta_k)$  belongs to a subset of Euclidean  $k$ -space,  $\mathbb{H}^k$ , which is the  $k^{\text{th}}$  product of  $\mathbb{H}$ . Let  $G_i$  be an a priori cumulative distribution function of  $\theta_i$  and thus  $G(\underline{\theta}) = \prod_{j=1}^{j=k} G_j(\theta_j)$  is an a priori cumulative distribution function on  $\mathbb{H}^k$ . (The  $\theta_i$ 's are considered independent random variables.) It is desirable to know at the time of any random observation  $\underline{x} = (x_1, \dots, x_k)$  belonging to  $\chi$ , a subset of Euclidean  $k$ -space, which population has the largest value for the parameter  $\theta_i$ ; i.e. which "i" in  $1, \dots, k$  has  $\theta_i \geq \theta_j$  for  $j = 1, \dots, k$ . To answer this question one could (1) select only one population and say it is "best", or (2) select a subset of the  $k$ -populations and say the "best" is contained in this subset. Note that the word "best" is used synonymously with "largest". This will be done throughout this paper but we remark that equivalent statements can be made if "best" were defined in terms of "smallest".

In decision theoretic terminology (1) has an action space  $A = A_1 = \{a_1, a_2, \dots, a_k\}$  in which action  $a_i$  means: "say  $\theta_i$  is largest." For (2)  $A = A_2 = \{S_1, S_2, \dots, S_p\}$  where  $p = 2^k - 1$ . (i.e. all possible subsets of  $\{1, 2, \dots, k\}$  excluding the empty set for obvious reasons). Action  $S_i$  in  $A_2$  means: "say subset  $S_i$  contains the 'best' population." Of course this requires a knowledge of the contents of  $S_i$  for  $i = 1, \dots, p$  but this will be discussed later as decision procedures are developed.

With each action there is a corresponding loss, say  $L(a_1, \underline{\theta})$ , which "measures" the loss incurred if  $\underline{\theta}$  is the "true state of nature" and action  $a_1 \in A_1$  is taken. An analogous statement (and statements to follow) can be made in regards to  $A_2$ . A decision function or procedure is a mapping "t" from  $\chi$  to A. Thus when  $\underline{x} \in \chi$  is observed, action  $t(\underline{x}) \in A$  is taken incurring a loss of  $L(t(\underline{x}), \underline{\theta})$ . If an a priori distribution is not assumed then a decision procedure is generally evaluated on the basis of its risk; i.e. its "expected loss". It is defined by

$$(1.1) \quad r(t, \underline{\theta}) = E_{\underline{x}}[L(t(\underline{x}), \underline{\theta})] = \int_{\chi} L(t(\underline{x}), \underline{\theta}) f(\underline{x}|\underline{\theta}) d\underline{x}$$

where  $f(\underline{x}|\underline{\theta}) = \prod_{j=1}^{j=k} f(x|\theta_j)$ . From this definition one usually speaks of admissible decision procedures, minimax decision procedures, or a minimal complete class of decision procedures as possible criteria for selecting a procedure. No distribution is assumed on the parameter space but often certain assumptions (such as being a bounded space) are made in order to achieve the above properties. However if one assumes an a priori distribution on the parameter space (so called Bayesian theory), then a Bayes risk can be defined by

$$(1.2) \quad R(t, G) = E[\text{Loss}] = E_G\{E_{\underline{x}}[L(t(\underline{x}), \underline{\theta})]\} = \int_{\Theta^k} r(t, \underline{\theta}) dG(\underline{\theta}) .$$

which of course takes into consideration the a priori distribution G on  $\Theta^k$ . If  $L(t(\underline{x}), \underline{\theta}) f(\underline{x}|\underline{\theta})$  is integrable on  $\chi \cdot \Theta^k$ , then by Fubini's theorem

$$(1.3) \quad R(t, G) = \int_{\chi} \varphi_G(t, \underline{x}) d\underline{x} \quad \text{in which}$$

$$(1.4) \quad \varphi_G(t, \underline{x}) = \int_k L(t(\underline{x}), \underline{\theta}), f(\underline{x}|\underline{\theta}) dG(\underline{\theta}) .$$

This "integrability" requirement will be satisfied by all loss functions and densities used in this paper. If a decision procedure  $t_G$  has the property

(1.5)  $R(t_G, G) \leq R(t, G)$  for any other decision procedure "t", then  $t_G$  is called a Bayes decision procedure with respect to  $G$ . Note that if  $t_G$  has the property

(1.6)  $\varphi_G(t_G, \underline{x}) \leq \varphi_G(t, \underline{x})$  for almost all  $\underline{x} \in \chi$  and for every decision procedure "t", then clearly  $t_G$  is Bayes with respect to  $G$ . Condition (1.6) will be used throughout this work as the definition for a Bayes procedure. For any finite action space  $A = \{d_1, d_2, \dots, d_m\}$  it is easy to see using (1.6) that  $t_G$  is defined by:

$$(1.7) \quad t_G(\underline{x}) = d_j \text{ where } j \text{ is any integer } 1, 2, \dots, m$$

$$\text{such that } \varphi_G(d_j, \underline{x}) = \min_{1 \leq i \leq m} \{\varphi_G(d_i, \underline{x})\} .$$

Of course finding the Bayes decision procedure with respect to  $G$  demands complete knowledge of  $G$ ; whereas using (1.2) ignores completely the possibility of an a priori  $G$ . An intermediate approach between these two is the case in which  $G$  is unknown but its existence is assumed. Thus  $t_G$  (the Bayes procedure) is not available to us, but the fact that an a priori  $G$  exists provides useful information which should not be ignored. For suppose observations  $(\underline{x}_1, \underline{\theta}_1), (\underline{x}_2, \underline{\theta}_2), \dots, (\underline{x}_n, \underline{\theta}_n)$  of  $\underline{X}$ , not  $\underline{\theta}$ ,



are repeatedly and independently obtained and all related to the same  $G$ . Then at the time a decision about the present observation  $\underline{x}$  of  $X$  is to be made, the prior observations  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  are available; and since the existence of  $G$  is assumed, these prior observations contain information about the form of that unknown  $G$ . If this information could be extracted in such a way so as to form a decision procedure  $t_n$  whose Bayes risk (with respect to any  $G$  in a large class  $\mathcal{G}$ ) converged to the Bayes risk of  $t_G$ , then  $t_n$  is asymptotically as good as  $t_G$ . In this sense then  $t_n$  is an optimal decision procedure and we say  $t_n$  is asymptotically optimal (a.o.) with respect to  $\mathcal{G}$ ; and thus call such a.o. procedures, empirical Bayes procedures.

Formally, let  $t_n(\cdot) = t_n(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n; \cdot)$  be a mapping from  $X^{n+1}$  to  $A$ , taking action  $t_n(\underline{x})$  in  $A$ , incurring loss  $L(t_n(\underline{x}), \theta)$ . For any given sequence  $T = \{t_n\}$  of such procedures, the Bayes risk relative to  $G$  is defined as

$$(1.8) \quad R_n(T, G) = \int_X E_n[\varphi_G(t_n, \underline{x})] d\underline{x}, \text{ where } \varphi_G \text{ is given by (1.4).}$$

Note that  $R_n(T, G)$  is an overall expected loss since  $t_n$  depends upon the "n" prior observations. Hence this explains the presence of  $E_n$  in (1.8), i.e. the expectation with respect to the "n" prior random variables which is given by

$$(1.9) \quad E_n[\varphi_G(t_n, \underline{x})] = \int_X \dots \int_X \varphi_G(t_n, \underline{x}) f_G(\underline{y}) d\underline{y} \dots f_G(\underline{y}) d\underline{y}$$

where

$$(1.10) \quad f_G(\underline{y}) = \int_k f(\underline{y}|\underline{\theta})dG(\underline{\theta}) \quad \text{and}$$

$$(1.11) \quad f(\underline{y}|\underline{\theta}) = \prod_{j=1}^k f(y_j|\theta_j)$$

(1.12) Definition. A sequence of empirical decision procedures  $T = \{t_n\}$  as defined above is said to be asymptotically optimal with respect to  $G$  if

$$\lim_{n \rightarrow \infty} R(t_n, G) = R(t_G, G)$$

for every  $G \in \tilde{G}$ , where  $\tilde{G}$  is some specified class of distributions on  $\mathbb{R}^k$ . We call  $t_n \in T$  an empirical Bayes procedure.

Therefore in decision theoretic terminology as given above, the multiple decision problem of interest throughout this paper is as follows:

Problem: For a specified loss structure, find empirical Bayes procedures for various (preferable large) classes  $G$  with  $A = A_1$  or  $A_2$ .

## B. Notation

The following notation will be used consistently for future work.

$\underline{x}$  - a k-dimensional vector with components  
 $(x_1, \dots, x_k)$ .

$\underline{\theta}$  - a k-dimensional vector with components  
 $(\theta_1, \dots, \theta_k)$ .

$\underline{x}_i$  - the  $i^{\text{th}}$  term of a sequence of k-dimensional vectors; it has components  $(x_{i1}, x_{i2}, \dots, x_{ik})$ .

$\underline{\theta}_i$  - the  $i^{\text{th}}$  term of a sequence of k-dimensional vectors; it has components  $(\theta_{i1}, \dots, \theta_{ik})$ .

$\underline{x}^*$  - an  $r \times k$  matrix, which will mean "r" observations from each of k-populations.

$$\theta_{[k]} - \max_{1 \leq j \leq k} \theta_j$$

For the  $x$ 's, a capital letter indicates the random variable while the lower case indicates the observation. Since  $\theta$  is a parameter and also a random variable but never observed, the above distinction will not be necessary. Also we emphasize that  $\underline{x}^*$  is an  $r \times k$  matrix each of whose rows is a vector  $\underline{x}^{(l)}$ ,  $l = 1, 2, \dots, r$ .

Note also that for convenience the subscripts "nj" are used in place of "n,j" as in a term such as  $f_{nj}$  or  $G_{nj}$ . Since the product of subscripts is never used in this paper, no confusion should arise.

## CHAPTER II

## SELECTING A SUBSET CONTAINING THE

## "BEST" OF k-POPULATIONS

## A. Introduction

Often it is desirable to select a subset containing the "best" of k-populations as opposed to selecting only the "best". Such authors as Paulson [26], Gupta [9], [14], [15], Gupta and Sobel [10], [11], [12], [13], and Seal [33], [34] have considered various aspects of this problem. However these authors have been more concerned with developing desirable properties satisfying some basic probability requirements. For example, the formulation in the papers by Gupta, Gupta and Sobel, and Seal, has been that of selecting a non-empty subset of random size satisfying the condition: the "best" population is included in the selected subset with probability at least  $P^*$ , a specified number. In this approach the efficiency of the decision procedure is the expected size of the selected subset and in terms of the simple zero-one loss function associated with the selection of the best, the risk does not exceed  $1 - P^*$ . In the present investigation a more basic decision theoretic framework (as described in Chapter I) is adopted with the eventual goal being the derivation of empirical Bayes decision procedures for either (1) selecting the "best" or (2) selecting a subset containing the "best". (The term empirical Bayes is intended to relay to the reader an approach toward the solution when an underlying a priori distribution  $G$  on the parameter space exists but is unknown). The main result of this chapter is to establish a connection between (1) and (2) for the Bayesian case. Then Chapters III and IV will relate Bayes and empirical Bayes procedures under certain conditions. Namely, we prove that

the Bayes procedure for selecting the "best" is precisely the Bayes procedure for selecting a subset, under a certain relationship of the corresponding loss functions. This fact will become significant in the light of Theorem 3.1 of Chapter III in which it is shown that under certain regularity conditions, the empirical Bayes procedure is just the Bayes procedure with respect to  $G_n$ ;  $G_n$  being an estimate of  $G$ , the true but unknown a priori  $G$ . (It should be pointed out that in Chapters III and IV,  $G$  is assumed to belong to a particular parametric family.)

B. Main Theorem And Corollaries

Using the notation of Chapter I, the action space  $A$  for the subset problem is given by  $A_2 = \{S_1, S_2, \dots, S_p\}$ , where  $p = 2^k - 1$ . That is,  $A_2$  is the set of all subsets of the integers  $1, 2, \dots, k$  excluding the empty set. For this problem we will assume that the loss function is of the form:

$$(2.1) \quad L(S_j; \underline{\theta}) = \sum_{q \in S_j} \alpha_{jq} (\theta_{[k]} - \theta_q), \quad (j = 1, 2, \dots, p)$$

and  $\alpha_{jq} \geq 0$  for all "j" and "q"; in which action  $S_j$  means: "say the best population is in the set  $S_j$ ." This loss function is non-negative (zero if and only if  $S_j$  has one element which is the best), continuous in  $\underline{\theta}$ , and bounded if and only if  $\theta_{[k]}$  is bounded. It also in some sense gives a balance between including more populations in order to obtain the best and having "too many" populations to allow a meaningful conclusion. Note the relationship of (2.1) to the loss function

$$(2.2) \quad L(a_i; \underline{\theta}) = \theta_{[k]} - \theta_i, \quad (i = 1, 2, \dots, k)$$

which is the loss function considered in Chapters III, IV and V for the problem of selecting the "best" of  $k$ -populations. (2.2) is the so-called linear loss structure (i.e. the "penalty" for a wrong decision is a linear function of the difference between the largest and the one selected) for the multiple decision problem of selecting the "best". It has been used by various authors. (See Dunnett [8], Bland [3], Bahadur and Robbins [2]). The theorem and corollaries below establish the relationship between the present problem and that of selecting the "best" considered in Chapters III and IV. For simplicity we assume that the first " $k$ " sets of  $A_2$  are the " $k$ " one element subsets of  $\{1, 2, \dots, k\}$ ; i.e. for  $j = 1, 2, \dots, k$  action  $S_j$  means: "say population  $j$  is the best"; and for  $j = k+1, \dots, p$  action  $S_j$  means: "say the best is in the subset  $S_j$ ." We do not require explicit knowledge of the sets  $S_j$ ,  $j > k$ , as will be seen in the theorem. Other notation and general background is the same as given in Chapter I. Recall that if  $r$  observations are taken from each of the  $k$ -populations, this observation matrix is denoted by  $\underline{x}^*$  and it is a point in  $\chi^r$ .

Theorem 2.3. In the loss function given in (2.1) let  $\alpha_{jq} = \alpha > 0$  for  $j = 1, 2, \dots, k$ . Let

$$c_q = \int_k (\theta_{[k]} - \theta_q) f(\underline{x}^* | \theta) dG(\theta), \quad c_{[1]} = \min_{1 \leq q \leq k} c_q,$$

and  $b_q = c_q - c_{[1]}$ . Using (1.4) and the above, we have

$\varphi_G(S_j, \underline{x}^*) = \sum_{q \in S_j} \alpha_{jq} c_q$ . If  $c_{[1]} \neq 0$ , then a necessary and sufficient

condition that  $\min_{1 \leq j \leq p} \varphi_G(S_j, \underline{x}^*) = \min_{1 \leq j \leq k} \varphi_G(S_j, \underline{x}^*)$  is that

$$\sum_{q \in S_j} \alpha_{jq} \geq \alpha - (1/c_{[1]}) \left( \sum_{q \in S_j} \alpha_{jq} b_q \right) \quad \text{for every } j = 1, 2, \dots, p.$$

If  $c_{[1]} = 0$ , then  $\min_{1 \leq j \leq p} \varphi_G(S_j, \underline{x}^*) = \min_{1 \leq j \leq k} \varphi_G(S_j, \underline{x}^*)$ .

Proof: Since  $\min_{1 \leq j \leq p} \varphi_G(S_j, \underline{x}^*) \leq \min_{1 \leq j \leq k} \varphi_G(S_j, \underline{x}^*)$ , where

$p = 2^k - 1$ , then  $\min_{1 \leq j \leq p} \varphi_G(S_j, \underline{x}^*) = \min_{1 \leq j \leq k} \varphi_G(S_j, \underline{x}^*)$  if and only if

$\min_{1 \leq i \leq k} \varphi_G(S_i, \underline{x}^*) \leq \varphi_G(S_j, \underline{x}^*)$  for every  $j = 1, 2, \dots, p$ . Now

$\min_{1 \leq i \leq k} \varphi_G(S_i, \underline{x}^*) = \alpha c_{[1]}$  and  $\varphi_G(S_j, \underline{x}^*) = \sum_{q \in S_j} \alpha_{jq} c_q$ , thus

$\min_{1 \leq i \leq k} \varphi_G(S_i, \underline{x}^*) \leq \varphi_G(S_j, \underline{x}^*)$  for every  $j = 1, 2, \dots, p$  if and only

if

$$(2.3.1) \quad \alpha c_{[1]} \leq \sum_{q \in S_j} \alpha_{jq} c_q, \quad (j = 1, \dots, p).$$

Note that if  $c_{[1]} = 0$ , then (2.3.1) is clearly satisfied since  $c_q \geq 0$

and  $\alpha_{jq} \geq 0$ . Thus  $\min_{1 \leq j \leq p} \varphi_G(S_j, \underline{x}^*) = \min_{1 \leq j \leq k} \varphi_G(S_j, \underline{x}^*)$ .

Now for the case  $c_{[1]} \neq 0$ , we have (2.3.1) if and only if

$$\alpha c_{[1]} \leq \sum_{q \in S_j} \alpha_{jq} (c_{[1]} + b_q);$$

if and only if  $(\alpha - \sum_{q \in S_j} \alpha_{jq}) \leq \frac{1}{c_{[1]}} \sum_{q \in S_j} \alpha_{jq} b_q$ ; if and only if

$$(2.3.2) \quad \sum_{q \in S_j} \alpha_{jq} \geq \alpha - \frac{1}{c_{[1]}} \sum_{q \in S_j} \alpha_{jq} b_q$$

for every  $j = 1, 2, \dots, p$  which completes the proof.

Remarks on Theorem 2.3: (1) The theorem could be stated in more general terms if, in (2.1),  $(\theta_{[k]} - \theta_q)$  is replaced with  $L(q, \underline{\theta})$  (i.e. the loss incurred when selecting a subset containing only the element "q",  $q = 1, 2, \dots, k$ .) and  $\alpha = 1$ . However for  $L(q, \underline{\theta})$  different from  $\alpha(\theta_{[k]} - \theta_q)$ , which is called linear loss, the computation of the Bayes procedure becomes quite complicated.

(2) In general the quantity  $c_q$  is difficult to compute and hence  $b_q$  is not available. Thus the necessary and sufficient condition given in the theorem is not too useful from a practical viewpoint. For this reason the following Corollary is significant.

Corollary 2.4: In the loss function given in (2.1), let  $\alpha_{jq} = \alpha > 0$  for  $j = 1, 2, \dots, k$ . If  $\sum_{q \in S_j} \alpha_{jq} \geq \alpha$  for every  $j = 1, 2, \dots, p$ , then

$$\min_{1 \leq j \leq p} \varphi_G(S_j, \underline{x}^*) = \min_{1 \leq j \leq k} \varphi_G(S_j, \underline{x}^*).$$

Proof. Clearly  $\sum_{q \in S_j} \alpha_{jq} c_q \geq \sum_{q \in S_j} \alpha_{jq} c_{[1]} \geq \alpha c_{[1]}$  using the notation of theorem (2.3). Then (2.3.1) is satisfied which implies the desired result.

A slightly more general loss function than (2.1) is

$$(2.5) \quad L^*(S_j, \underline{\theta}) = \sum_{q \in S_j} \alpha_{jq} (\theta_{[k]} - \theta_q) + \gamma_j h(\underline{\theta})$$

in which  $\alpha_{jq} \geq 0$  for all "j" and "q";  $\gamma_j \geq 0$  with  $\gamma_j = \gamma \geq 0$  for  $j = 1, 2, \dots, k$ ; and  $h(\underline{\theta})$  an integrable function of  $\underline{\theta}$  with respect to the measure induced by the cumulative distribution function  $G$ . Thus



$$(2.6) \quad L^*(S_j, \underline{\theta}) = L(S_j, \underline{\theta}) + \gamma_j h(\underline{\theta}) \quad \text{where}$$

$$L(S_j, \underline{\theta}) \quad \text{is given by (2.1). Then}$$

$$\varphi_G^*(S_j, \underline{x}^*) = \int_k L^*(S_j, \underline{\theta}) f(\underline{x}^* | \underline{\theta}) dG(\underline{\theta}),$$

using (1.4) and from (2.6), we have

$$(2.7) \quad \varphi_G^*(S_j, \underline{x}^*) = \varphi_G(S_j, \underline{x}^*) + \gamma_j h_G(\underline{x}^*) \quad \text{where}$$

$$(2.8) \quad h_G(\underline{x}^*) = \int_k h(\underline{\theta}) f(\underline{x}^* | \underline{\theta}) dG(\underline{\theta}).$$

We now prove an analogous corollary to (2.4) for this more general case.

Corollary 2.9: In the loss function given by (2.6) let  $\alpha_{jq} = \alpha > 0$  for  $j = 1, 2, \dots, k$ . If  $\sum_{q \in S_j} \alpha_{jq} \geq \alpha$  and  $\gamma_j \geq \gamma$  for  $j = 1, 2, \dots, p$ ,

then

$$\min_{1 \leq j \leq p} \varphi_G^*(S_j, \underline{x}^*) = \min_{1 \leq j \leq k} \varphi_G^*(S_j, \underline{x}^*).$$

Proof: Clearly  $\min_{1 \leq j \leq p} \varphi_G^*(S_j, \underline{x}^*) \leq \min_{1 \leq j \leq k} \varphi_G^*(S_j, \underline{x}^*)$ .

For the other direction, we have from (2.7) that

$$\begin{aligned} \min_{1 \leq j \leq p} \varphi_G^*(S_j, \underline{x}^*) &= \min_{1 \leq j \leq p} \{ \varphi_G(S_j, \underline{x}^*) + \gamma_j h_G(\underline{x}^*) \} \\ &\geq \min_{1 \leq j \leq p} \varphi_G(S_j, \underline{x}^*) + h_G(\underline{x}^*) \min_{1 \leq j \leq p} \gamma_j \end{aligned}$$

But  $\min_{1 \leq j \leq p} \varphi_G(S_j, \underline{x}^*) = \min_{1 \leq j \leq k} \varphi_G(S_j, \underline{x}^*)$  from Corollary 2.4, and

$\gamma_j \geq \gamma$  for every  $j = 1, 2, \dots, p$ . Hence

$$\min_{1 \leq j \leq p} \varphi_G^*(S_j, \underline{x}^*) \geq \min_{1 \leq j \leq k} \varphi_G(S_j, \underline{x}^*) + \gamma h_G(\underline{x}^*) =$$

$$\min_{1 \leq j \leq k} \{\varphi_G(S_j, \underline{x}^*) + \gamma h_G(\underline{x}^*)\} = \min_{1 \leq j \leq k} \varphi_G^*(S_j, \underline{x}^*). \text{ Thus}$$

$$\min_{1 \leq j \leq p} \varphi_G^*(S_j, \underline{x}^*) \geq \min_{1 \leq j \leq k} \varphi_G^*(S_j, \underline{x}^*) \text{ and this completes the proof.}$$

The following corollary is the key result of this chapter.

Corollary 2.10: Assume the conditions of Corollary (2.9). Then a decision procedure  $t_G$  which is Bayes with respect to  $G$  when  $A = A_1 = \{a_1, \dots, a_k\}$  and  $L(a_j, \theta)$  is given by (2.2) is also Bayes with respect to  $G$  when  $A = A_2 = \{S_1, \dots, S_p\}$ ,  $p = 2^k - 1$ , and  $L^*(S_j, \theta)$  is given by (2.5).

Proof: From (1.7)  $t$  is Bayes with respect to  $G$  for  $A_2$  if

$$(2.10.1) \quad \varphi_G^*(t, \underline{x}^*) = \min_{1 \leq j \leq p} \{\varphi_G^*(S_j, \underline{x}^*)\}$$

for each  $\underline{x}^*$ , where  $\varphi_G^*$  is given by (2.7). Let  $t_G$  be a decision procedure which is Bayes with respect to  $G$  for  $A_1$ .

Then by (1.7)

$$(2.10.2) \quad \varphi_G(t_G, \underline{x}^*) = \min_{1 \leq j \leq k} \{\varphi_G(a_j, \underline{x}^*)\}.$$

Now by (2.7) noting that  $\gamma_j = \gamma$  for  $j = 1, \dots, k$  we have

$$(2.10.3) \quad \varphi_G^*(S_j, \underline{x}^*) = \varphi_G(S_j, \underline{x}^*) + \gamma h_G(\underline{x}^*) \text{ for } j = 1, \dots, k.$$

But  $\alpha_{jq} = \alpha$  for  $j = 1, 2, \dots, k$ , hence  $L(S_j, \theta) = \alpha L(a_j, \theta)$

for  $j = 1, \dots, k$  which implies

$$\varphi_G(S_j, \underline{x}^*) = \alpha \varphi_G(a_j, \underline{x}^*)$$

and thus from (2.10.3) we have

$$(2.10.4) \quad \varphi_G^*(S_j, \underline{x}^*) = \alpha \varphi_G(a_j, \underline{x}^*) + \gamma h_G(\underline{x}^*) .$$

Then from (2.10.2) we have

$$\varphi_G^*(t_G, \underline{x}^*) = \alpha \varphi_G(t_G, \underline{x}^*) + \gamma h_G(\underline{x}^*) \leq \alpha \varphi_G(a_j, \underline{x}^*) + \gamma h_G(\underline{x}^*)$$

for every  $j = 1, \dots, k \leq \varphi_G^*(S_j, \underline{x}^*)$  for every  $j = 1, \dots, k$ . But  $t_G$  must select some one of the first  $k$  sets  $S_1, \dots, S_k$ . Thus

$$\varphi_G^*(t_G, \underline{x}^*) = \min_{1 \leq j \leq k} \varphi_G^*(S_j, \underline{x}^*)$$

and from Corollary 2.9

$$(2.10.5) \quad \varphi_G^*(t_G, \underline{x}^*) = \min_{1 \leq j \leq p} \varphi_G^*(S_j, \underline{x}^*)$$

which from (2.10.1) implies  $t_G$  is Bayes with respect to  $G$  for  $A_2$  and the proof is complete.

### C. Examples

The chart below gives several examples of loss functions which satisfy Corollary 2.9 ( $|S_j|$  indicates the number of elements in the subset  $S_j$ ). Note that (1) is just the sum of the individual losses; (2) has a "weight factor" proportional to the number of elements in the subset added to

2.11 Chart of loss functions satisfying Corollary (2.9)

$L^*(S_j, \theta)$	$\alpha$	$\alpha_{jq}$	$\sum_{q \in S_j} \alpha_{jq}$	$\gamma$	$\gamma_j^h(\theta)$
1 $\sum_{q \in S_j} (\theta_{[k]} - \theta_q)$	1	1	$ S_j $ which is $\geq 1$ for every $j=1, \dots, p$	0	0 for every $j = 1, \dots, p$
2 $\sum_{q \in S_j} (\theta_{[k]} - \theta_q) + ( S_j  - 1)$	1	1	$ S_j $ which is $\geq 1$ for every $j=1, \dots, p$	0	$ S_j  - 1 \geq 0$
3 $\frac{1}{ S_j } \sum_{q \in S_j} \theta_{[k]} - \theta_q$	1	$(1/ S_j )$ for $q \in S_j$	$\frac{ S_j }{ S_j } = 1$ for every $j = 1, \dots, p$	0	0
4 $\frac{1}{ S_j } \sum_{q \in S_j} \theta_{[k]} - \theta_q + ( S_j  - 1)$	1	$(1/ S_j )$ for $q \in S_j$	$\frac{ S_j }{ S_j } = 1$ for every $j = 1, \dots, p$	0	$ S_j  - 1 \geq 0$
5 $(k+1 -  S_j ) \sum_{q \in S_j} (\theta_{[k]} - \theta_q)$	k	$(k+1 -  S_j )$ for every $q \in S_j$	$ S_j  (k+1 -  S_j ) \geq k$ for every $j = 1, \dots, p$ (see 2.11.1).	0	0
6 $\theta_{[k]} - \sum_{q \in S_j} \theta_q + d( S_j  - 1)$ (see 2.11.2)	1	1	$ S_j  \geq 1$ for every $j = 1, \dots, p$	0	$ S_j  - 1 \geq \frac{ S_j  - 1}{(d - \theta_{[k]})^{-1}}$

(1); (3) is the average of the individual losses; (4) is (3) plus the "weight factor" of (2); (5) is a sum of individual losses with each loss "weighted" by a factor inversely proportional to the number of elements in the subset  $S_j$ ; (6) is a useful loss function in case the parameter space is bounded, but is quite similar to (2). (Items (2.11.1) and (2.11.2) below will be referred to in chart 2.11).

2.11.1 If  $|S_j| = q, q = 1, \dots, k$ , then  $|S_j|(k+1 - |S_j|) = q(k+1 - q) = qk + q - q^2$  which is  $\geq k$  if  $qk - k + q - q^2 \geq 0$ . Now  $k = q + t$  where "t" is a non negative integer. Then  $qk - k + q - q^2 \geq 0$  if  $q(q+t) - q - t + q - q^2 \geq 0$  which is true if  $t(q-1) \geq 0$ , which is true since  $t \geq 0$  and  $q \geq 1$ . Therefore  $|S_j|(k+1 - |S_j|) \geq k$ .

2.11.2 Note that  $\theta_{[k]} - \sum_{q \in S_j} \theta_q + d(|S_j| - 1) =$

$$\sum_{q \in S_j} (\theta_{[k]} - \theta_q) + (|S_j| - 1)(d - \theta_{[k]})$$

where  $d$  is such that  $|\theta| \leq d$  for every  $\theta \in \Theta$ . Thus this loss function is limited in use to the case in which the parameter space is bounded.

#### D. Summary of Bayes and Empirical Bayes

##### Procedures For Selecting A Subset

In Corollary (2.10) we have seen that for the loss functions given in (2.11) and any others satisfying Corollary (2.9) the Bayes procedure for the subset problem is precisely the Bayes procedure for selecting the "best". In Chapter III it is shown that if  $G_n$  is an estimate of  $G$  which converges

with probability one then the Bayes procedure with respect to  $G_n$  is empirical Bayes with respect to  $G$  (i.e. asymptotically optimal) provided certain other regularity conditions hold. Now if  $G$  is a member of some parametric family  $\tilde{G}$  (as it is in Chapters III and IV), then obtaining  $G_n$  hinges upon estimating this parameter. Thus  $G_n$  is also in  $\tilde{G}$  and finding the empirical Bayes procedure is immediate if the Bayes procedure with respect to any  $G$  in  $\tilde{G}$  is available. These are obtained in Chapter IV. Since these later results are applicable to the subset problem, we include a summary of those results here for convenience.

(1) Normal-normal

$f(x|\theta_i)$  is  $N(\theta_i, \sigma_i^2)$  and  $G_i(\theta)$  is a cumulative distribution function of a  $N(\lambda_i, \beta_i^2)$ .

$t_G(\underline{x}^*) = \{j\}$  where "j" is any positive integer 1, 2, ..., k such that

$$(2.12.1) \quad \frac{r\beta_j^2 \bar{x}_j + \sigma_j^2 \lambda_j}{\sigma_j^2 + r\beta_j^2} = \max_{1 \leq i \leq k} \left\{ \frac{r\beta_i^2 \bar{x}_i + \sigma_i^2 \lambda_i}{\sigma_i^2 + r\beta_i^2} \right\}.$$

(2.12.2)  $t_{G_n}(\underline{x}^*) = \{j\}$  where j is any positive integer 1, 2, ..., k

$$\frac{r\beta_j^2 \bar{x}_j + \sigma_j^2 \bar{x}_j}{\sigma_j^2 + r\beta_j^2} = \max_{1 \leq i \leq k} \left\{ \frac{r\beta_i^2 \bar{x}_i + \sigma_i^2 \bar{x}_i}{\sigma_i^2 + r\beta_i^2} \right\}$$

is the empirical Bayes procedure;  $\sigma_i^2, \beta_i^2$  are known; with

$$\bar{x}_j = \frac{1}{r} \sum_{\ell=1}^{\ell=r} x_j^{(\ell)}, \quad \text{and} \quad \bar{\bar{x}}_j = \frac{1}{rn} \sum_{i=1}^{i=mn} \sum_{\ell=1}^{\ell=r} x_{ij}^{(\ell)} .$$

(2) Normal-uniform

$f(x_j | \theta_j) = N(\theta_j, \sigma_j^2)$  and  $\theta_j$  uniformly distributed on  $(\lambda_j - d_j, \lambda_j + d_j)$  .

(2.13.1)  $t_G(\underline{x}^*) = \{j\}$ , where  $j$  is any integer  $1, 2, \dots, k$

$$\frac{\varphi(\beta_j) - \varphi(\alpha_j)}{\Phi(\alpha_j) - \Phi(\beta_j)} + \frac{r\bar{x}_j}{\sigma_j^2} = \max_{1 \leq i \leq k} \left\{ \frac{\varphi(\beta_i) - \varphi(\alpha_i)}{\Phi(\alpha_i) - \Phi(\beta_i)} + \frac{r\bar{x}_i}{\sigma_i^2} \right\}$$

is the Bayes procedure with respect to  $G = \prod_{j=1}^k G_j$  where  $\lambda_j$ 's and  $d_j$ 's are known, and

$$\alpha_j = r/\sigma_j^2 \left( \lambda_j + d_j - \bar{x}_j \right)$$

$$\beta_j = r/\sigma_j^2 \left( \lambda_j - d_j - \bar{x}_j \right) .$$

(2.13.2)  $t_{G_n}(\underline{x}^*) = \{j\}$ , where  $j$  is any integer  $1, 2, \dots, k$

$$\frac{\varphi(\beta_j) - \varphi(\alpha_j)}{\Phi(\alpha_j) - \Phi(\beta_j)} + \frac{r\bar{x}_j}{\sigma_j^2} = \max_{1 \leq i \leq k} \left\{ \frac{\varphi(\beta_i) - \varphi(\alpha_i)}{\Phi(\alpha_i) - \Phi(\beta_i)} + \frac{r\bar{x}_i}{\sigma_i^2} \right\}$$

is asymptotically optimal to  $t_G$ . It is based upon "n" prior observations and assumes  $d_j$ 's are known, with  $\alpha_j = \bar{\bar{x}}_j + d_j - \bar{x}_j$ ,  $\beta_j = \bar{\bar{x}}_j - d_j - \bar{x}_j$ .

(3) Binomial, Beta

$$\text{Let } f(x_j | \theta_j) = \binom{u_j}{x_j} \theta_j^{x_j} (1 - \theta_j)^{u_j - x_j}; 0 \leq \theta_j \leq 1,$$

$u_j$  a positive integer and  $x_j = 0, 1, \dots, u_j$ . Let  $G_j(\theta_j)$  be the

cumulative distribution function for a  $\beta$  density

$$g_j(\theta_j) = c_j \theta_j^{\lambda_j - 1} (1 - \theta_j)^{\nu_j - \lambda_j - 1}$$

where

$$c_j = \frac{\Gamma(\nu_j)}{\Gamma(\lambda_j)\Gamma(\nu_j - \lambda_j)} = \frac{1}{B(\lambda_j, \nu_j)}, \lambda_j$$

and  $\nu_j$  are non-negative.

(2.14.1)  $t_G(\underline{x}^*) = \{i\}$ , where  $i$  is any integer  $1, \dots, k$  such that

$$\frac{\bar{x}_i + \frac{1}{r} \lambda_i}{u_i + \frac{1}{r} \nu_i} = \max_{1 \leq j \leq k} \left\{ \frac{\bar{x}_j + \frac{1}{r} \lambda_j}{u_j + \frac{1}{r} \nu_j} \right\}$$

is the Bayes procedure.

(2.14.2)  $t_{G_n}(\underline{x}^*) = \{i\}$ , where  $i$  is any integer  $1, \dots, k$  such that

$$\frac{\bar{x}_i + \frac{1}{r} \frac{\nu_i}{u_i} \bar{x}_i}{u_i + \frac{1}{r} \nu_i} = \max_{1 \leq j \leq k} \frac{\bar{x}_j + \frac{1}{r} \frac{\nu_j}{u_j} \bar{x}_j}{u_j + \frac{1}{r} \nu_j}$$

is the empirical Bayes procedure based upon "n" prior observations and is asymptotically optimal to  $t_G$ .



(4) Poisson-Gamma

Let 
$$f(x_j | \theta_j) = \frac{e^{-\theta_j} \theta_j^{x_j}}{x_j!}; \quad \theta_j > 0,$$

$x_j = 0, 1, 2, \dots$  and  $G_j(\theta_j)$  the cumulative distribution function for the gamma density

$$g_j(\theta_j) = \frac{\lambda_j \cdot \theta_j^{\lambda_j - 1} e^{-\theta_j \alpha_j}}{\Gamma(\alpha_j)};$$

$\lambda_j$  and  $\alpha_j$  non negative.

(2.15.1)  $t_G(\underline{x}^*) = \{i\}$ , where  $i$  is any integer  $1, 2, \dots, k$  such that

$$\frac{r\bar{x}_i + \lambda_i}{r + \alpha_i} = \max_{1 \leq j \leq k} \left\{ \frac{r\bar{x}_j + \lambda_j}{r + \alpha_j} \right\}$$

is the Bayes procedure with respect to  $G = \pi G_j$  when  $\alpha_j, \lambda_j$  are known  $j = 1, \dots, k$ .

(2.15.2)  $t_{G_n}(\underline{x}^*) = \{i\}$ , where  $i$  is any integer  $1, 2, \dots, k$  such that

$$\frac{r\bar{x}_i + \alpha_i \bar{\bar{x}}_i}{r + \alpha_i} = \max_{1 \leq j \leq k} \left\{ \frac{r\bar{x}_j + \alpha_j \bar{\bar{x}}_j}{r + \alpha_j} \right\}$$

is the empirical Bayes procedure which is asymptotically optimal to  $t_G$ .

It is based upon "n" prior observations and assumes  $\alpha_j$  known for

$j = 1, \dots, k$ .

## CHAPTER III

## PARAMETRIC CASE - BAYES PROCEDURES

## FOR SELECTING THE BEST

## A. Introduction

(1) Background - In many situations only the "best" of  $k$ -populations is desired contrasted to the subset formulation given in Chapter II. Such authors as Bechhofer [4], Bechhofer and Sobel [5], Bechhofer, Sobel and Dunnett [6], Sobel and Huyett [35], Chambers and Jarratt [7] have made contributions to this problem. These authors usually work under the restriction of an "indifference zone." For example in the formulation by Bechhofer [4], one is interested in guaranteeing the probability of selecting the "best" population to be a specified number  $P^*$  whenever the difference between the "best" and the second best is at least equal to a specified number  $\delta^*$ . A decision procedure is given which depends upon the sample size  $r$  and the problem of explicitly determining the common sample size  $r$  to guarantee the probability  $P^*$  under the "indifference zone"  $\delta \geq \delta^*$  is solved therein. Various modifications and extensions have been made by other authors, but the approach is basically similar. A truly Bayesian approach to this problem was taken by Dunnett [8], and more recently by Guttman and Tiao [17], in which an a priori distribution on the parameter space is assumed. In [17] a tolerance region as a function of the parameter is defined and utility is then defined as some function of this tolerance region. Procedures which select the best population are derived from maximizing the expected utility.

In this investigation a basic decision theoretic framework as described in Chapter I is adopted and empirical Bayes procedures for selecting the

"best" population are sought; the difference between Bayes and empirical Bayes being that in the empirical Bayes approach, only the existence of an a priori distribution is assumed. Recall that in the context of Chapter I we have:

$A_1 = \{a_1, a_2, \dots, a_k\}$  is the action space;

$L(a_j, \theta) = \theta_{[k]} - \theta_j$  is the loss function;

$f(x|\theta_i)$  is the density of  $X_1$  with parameter  $\theta_i$ ;

$G_i(\theta)$  is an a priori distribution on  $\Theta$ , and

$G = \prod_{i=1}^{i=k} G_i$  is an a priori distribution on  $\Theta^k$ , the parameter space;

$t_G$  denotes the Bayes procedure with respect to  $G$ ;

$\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_n^*$  denote prior observations of the random matrix  $\underline{X}^*$ , with  $\underline{x}^*$  being considered the present observation of  $\underline{X}^*$  with  $G$  remaining the same for all observations (note that the "\*" indicates  $r$  observations per population for each observation vector);

$R(t, G)$  denotes the overall Bayes risk of a decision procedure  $t$  with respect to an a priori distribution  $G$ .

In this chapter the main result is to show that under certain regularity conditions the Bayes procedure with respect to an estimate  $G_n$  of  $G$  is also empirical Bayes with respect to  $G$ ; (i.e.  $R(t_{G_n}, G) \rightarrow R(t_G, G)$ ).

Of course to apply this theorem, a suitable estimate  $G_n$  is required. A completely satisfactory answer to this problem is not yet available (see Robbins [30]), but one approach is to extend the existence assumption on  $G$  to say further that  $G$  belongs to some parametric family  $\tilde{G}$ ; e.g. the class of all normal distributions. This extension is valid in many practical

situations and will be assumed in this chapter and Chapter IV as well. In Chapter V results are obtained without this additional assumption on  $G$ ; rather we assume only that

$$\int \theta_i dG_i < \infty \text{ for } i = 1, 2, \dots, k. \text{ Herein lies the motivation for}$$

calling Chapters III and IV the "parametric" case and Chapter V, the "non-parametric" case.

Now if  $G$  belongs to some parametric family having parameter  $\underline{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_k)$ , then finding an estimate  $G_n$  of  $G$  reduces to finding an estimate  $\underline{\lambda}_n^*$  of  $\underline{\lambda}$ ; and hence by the theorem below,  $t_{G_n}$  will be empirical Bayes with respect to  $G$ . But  $G_n$  is a member of the parametric family  $G$  as well, thus finding  $t_{G_n}$  (the empirical Bayes procedure) amounts to finding the Bayes procedure with respect to the family  $G$ ; i.e. finding the Bayes procedure as a function of the parameter  $\underline{\lambda}$ . Therefore besides the main theorem below, Bayes procedures with respect to specific classes  $\tilde{G}$  will be derived in this chapter. Then in Chapter IV the corresponding empirical Bayes procedures depending upon the estimates  $\underline{\lambda}_n^*$  will be obtained.

(2) Relation to the subset problem - From the above one can easily see the usefulness of the results in Chapter II in deriving empirical Bayes procedures for selecting a subset. Namely, if the empirical Bayes procedure is really a Bayes procedure with respect to some distribution  $G_n$ , then by Corollary 2.9 of Chapter II, the empirical Bayes procedure for selecting a subset is given by the empirical Bayes procedure for selecting the "best"; provided of course the conditions of Corollary 2.9 are satisfied. Hence under those conditions, the procedures derived in Chapters III and IV for selecting the "best" are also Bayes and

empirical Bayes for selecting a subset; i.e. only one population is selected even in the subset formulation.

### B. Main Theorem

We now proceed with the main theorem of this chapter and also Chapter IV.

Theorem 3.1: Let  $\underline{X}_1, \underline{X}_2, \dots$  be independent random vectors;  $\underline{X}_i$  having density  $f(\underline{x} | \underline{\theta}_i)$  and consisting of  $k$ -components which are independent random variables. Let  $G_n$  be a cumulative distribution function on  $\Theta^k$ .

Suppose  $G_n$  is a function of  $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$  such that

$P\{\lim_{n \rightarrow \infty} G_n(\underline{\theta}) = G(\underline{\theta}), \text{ for every continuity point } \underline{\theta} \text{ of } G\} = 1$ ; where  $G$

is a  $k$ -dimensional cumulative distribution function on  $\Theta^k$  and does not

depend upon the  $\underline{X}$ 's, and the probability  $P$  is taken with respect to

$\underline{X}_1, \underline{X}_2, \dots$ . Let the loss function  $L(a_i, \underline{\theta})$  and densities  $f(\underline{x} | \underline{\theta}_i)$  be

such that  $L(a_i, \underline{\theta})f(\underline{x} | \underline{\theta})$  is bounded and continuous in  $\underline{\theta}$  for every

$a_i \in A_1$  and  $\underline{x} \in \chi$ . Let  $L(\underline{\theta}) = \max\{L(a_i, \underline{\theta}) : a_i \in A\}$ . Then  $\{t_{G_n}\}$  is

asymptotically optimal with respect to  $G$  if  $\int_{\Theta^k} L(\underline{\theta}) dG(\underline{\theta}) < \infty$ .

Proof: We first prove the following lemma.

Lemma 3.2: Under the hypotheses of Theorem 3.1,

$P\{\lim_{n \rightarrow \infty} \varphi_{G_n}(t_{G_n}, \underline{x}) = \varphi_G(t_G, \underline{x})\} = 1$ ; where the function

$\varphi_G$  is defined in (1.4).

Proof: (i) Since  $L(a_i, \theta)f(\underline{x}|\theta)$  is bounded and continuous in  $\theta$  for every  $a_i \in A$  and  $\underline{x} \in \chi$ , the Helly Bray Theorem can be applied to  $\varphi_{G_n}(t, \underline{x})$  to obtain:

$$(3.2.1) \quad \lim_{n \rightarrow \infty} \varphi_{G_n}(t, \underline{x}) = \varphi_G(t, \underline{x}) \text{ for any decision procedure } t.$$

By hypothesis in the theorem above,  $P\{\lim_{n \rightarrow \infty} G_n(\theta) = G(\theta)\} = 1$ , and hence

$$P\{\lim_{n \rightarrow \infty} \varphi_{G_n}(t, \underline{x}) = \varphi_G(t, \underline{x})\} = 1. \text{ That is, there exists a set } B \subset \chi^\infty$$

with  $P\{B\} = 1$  and such that  $\varphi_{G_n}(t, \underline{x})$  converges to  $\varphi_G(t, \underline{x})$  for each element in  $B$ . (Note: an element in  $B$  is just one possible observation of the infinite sequence  $\underline{x}_1, \underline{x}_2, \dots$  for which the convergence above holds.)

(ii) Now  $t_{G_n}$  denotes the Bayes procedure with respect to  $G_n$  and hence is considered a function of  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  and  $\underline{x}$ . So for a given sequence  $\underline{x}_1, \dots, \underline{x}_n$ ,  $t_{G_n}$  is a decision procedure mapping  $\chi$  to  $A$ . Further  $t_G$  is a Bayes Procedure with respect to  $G$  and is a function of  $\underline{x}$  only. By definition of a Bayes procedure we have:

$$(3.2.2) \quad \varphi_G(t_G, \underline{x}) \leq \varphi_G(t_{G_n}, \underline{x}) \text{ and}$$

$$\varphi_{G_n}(t_{G_n}, \underline{x}) \leq \varphi_{G_n}(t_G, \underline{x})$$

for  $\underline{x} \in \chi$ . Let  $y$  be any element of  $B$  and let  $\epsilon > 0$ . Then using

(3.2.1) we obtain:

$$(3.2.3) \quad \varphi_{G_n}(t_G, \underline{x}) \leq \varphi_G(t_G, \underline{x}) + \frac{\epsilon}{2} \text{ and}$$

$$\varphi_G(t_{G_n}, \underline{x}) \leq \varphi_{G_n}(t_{G_n}, \underline{x}) + \frac{\epsilon}{2} \text{ for}$$

$n > N_\epsilon(\underline{x}, y)$  where

$$N_\epsilon(\underline{x}, y) = \max_{1 \leq i \leq k} N_{\epsilon, i}(\underline{x}, y);$$

$N_{\epsilon, i}(\underline{x}, y)$  being a number such  $n > N_{\epsilon, i}(\underline{x}, y)$  implies

$$\varphi_G(a_i, \underline{x}) - \varphi_{G_n}(a_i, \underline{x}) \leq \frac{\epsilon}{2}.$$

(Note: the decision procedure  $t_{G_n}$  must always select one of the "k" possible actions  $a_1, a_2, \dots, a_k$  for any  $n$ .)

We now return to the proof of the theorem. We want to show that:

$$(3.3.1) \quad \lim_{n \rightarrow \infty} R(t_{G_n}, G) = R(t_G, G).$$

Now

$$(3.3.2) \quad R(t_{G_n}, G) = \int_X E[\varphi_{G_n}(t_{G_n}, \underline{x})] d\underline{x}, \text{ E being taken with respect to } \underline{X}_1, \dots, \underline{X}_n, \text{ and}$$

$$(3.3.3) \quad R(t_G, G) = \int_X \varphi_G(t_G, \underline{x}) d\underline{x} \text{ by definition.}$$

Thus, if  $E[\varphi_{G_n}(t_{G_n}, \underline{x})]$  is bounded by an integrable function on  $X$  and  $E[\varphi_{G_n}(t_{G_n}, \underline{x})] \xrightarrow{P} \varphi_G(t_G, \underline{x})$ , then (3.3.1) is a result of the Dominated Convergence Theorem. (See Loeve [23], p. 152). It suffices then to show that these two conditions are satisfied.

Let  $\varphi_G(\underline{x})$  denote  $\int_k L(\theta) f(\underline{x} | \theta) dG(\theta)$ ; then we have

$$(3.4) \quad \varphi_G(t_{G_n}, \underline{x}) \leq \varphi_G(\underline{x}) \text{ for every } \underline{x} \in \chi \text{ and for every}$$

n. Also

$$(3.5) \quad \int_{\chi} \varphi_G(\underline{x}) d\underline{x} = \int_k L(\theta) \left\{ \int_{\chi} f(\underline{x}|\theta) d\underline{x} \right\} dG(\theta) \\ = \int_k L(\theta) dG(\theta) < \infty \text{ by hypothesis.}$$

(Recall the integrability assumption made earlier which allows the change of order of integration in (3.5).)

Now

$$(3.6) \quad 0 \leq E[\varphi_G(t_{G_n}, \underline{x})] \leq E[\varphi_G(\underline{x})] = \varphi_G(\underline{x}) < \infty$$

using (3.4), and noting that the expectation is taken with respect to the random vectors  $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$  and that  $\varphi_G(\underline{x})$  is a constant with respect to these random vectors. Thus (3.6) and (3.5) give the boundedness condition on  $E[\varphi_G(t_{G_n}, \underline{x})]$ . Also for a fixed  $\underline{x} \in \chi$ , we can see from (3.4) that  $\varphi_G(t_{G_n}, \underline{x})$  is bounded by an integrable function (i.e. a constant); and from Lemma 3.2  $\varphi_G(t_{G_n}, \underline{x}) \xrightarrow{\text{a.e.}} \varphi_G(t_G, \underline{x})$ . Hence we may reapply the Dominated Convergence Theorem to give:

$$(3.7) \quad \lim_{n \rightarrow \infty} E[\varphi_G(t_{G_n}, \underline{x})] = \varphi_G(t_G, \underline{x}), \text{ for each } \underline{x} \in \chi.$$

This is the second requirement stated above and completes the proof of the theorem.



## C. Remarks on Theorem 3.1

- (1) If  $r$  observations are taken from each population for each vector  $\underline{x}$ , then the same proof holds with  $f(\underline{x}|\underline{\theta})$  replaced by  $f(\underline{x}^r|\underline{\theta})$  and  $\chi$  replaced by  $\chi^r$ .
- (2) The requirement that  $L(a_i, \underline{\theta})f(\underline{x}|\underline{\theta})$  be bounded and continuous in  $\underline{\theta}$  can be weakened to that which is necessary to allow application of the Helly-Bray Theorem.
- (3) This theorem can be obtained indirectly from Robbins [30], however the above statement and proof is more directly related to the problem at hand.
- (4) As mentioned earlier, finding estimates  $G_n$  is the problem. We will ease this difficulty somewhat by assuming  $G$  to be a member of some parametric family  $\tilde{G}$  with parameter  $\underline{\lambda}$ . Then the theorem suggests the following procedure. (i) For specified  $\tilde{G}$ , find the Bayes procedure  $t_G$  as a function of the observation  $\underline{x}$  and the parameter  $\underline{\lambda}$  of  $G$ ; say  $t_G(\underline{x}) = h(\underline{x}, \underline{\lambda})$ . (ii) Verify the hypotheses; namely, that  $L(a_i, \underline{\theta})f(\underline{x}|\underline{\theta})$  is bounded and continuous in  $\underline{\theta}$  for each  $\underline{x}$  and for any  $a_i$  in  $A_1$ , and that  $\int_k L(\underline{\theta})dG(\underline{\theta}) < \infty$  where  $L(\underline{\theta}) = \max_{1 \leq i \leq k} L(a_i, \underline{\theta})$ .
- (iii) Then, using the prior observations  $\underline{x}_1, \dots, \underline{x}_n$ , find an estimate  $\underline{\lambda}_n$  of  $\underline{\lambda}$  such that  $G_n$  converges to  $G$  with probability one. If this can be done, direct application of the theorem implies  $t_{G_n}(\underline{x}) = h(\underline{x}, \underline{\lambda}_n)$  is empirical Bayes with respect to  $G$  in  $\tilde{G}$ .

The remainder of the chapter will be concerned with finding the function  $h(\underline{x}, \underline{\lambda})$  for specific situations; whereas Chapter IV will be

concerned with verifying the hypotheses and finding  $\lambda_n$  (and hence  $t_{G_n}$ ) for these situations covered in Chapter III.

#### D. Bayes Procedures for Selecting the "Best"

Each of the  $k$  populations has an associated random variable  $X_j$  with density  $f(x|\theta_j)$  where  $\theta_j$  is an unknown parameter and  $X_i, X_j$  are independent,  $i \neq j$ . But an a priori distribution  $G_j$  of each  $\theta_j$  exists and is specified by a given parameter  $\lambda_j$ . ( $\lambda_j$  itself may be a vector.) Suppose " $r$ " observations ( $r = 1, 2, \dots$ ) are taken on each  $X_j$ . Denote by  $\underline{X}^*$  the random matrix  $(X_1^*, \dots, X_k^*)$  and by  $\underline{x}^*$ , an observation of  $\underline{X}^*$  where  $\underline{x}_j^*$  is an  $r$ -dimensional column vector representing the " $r$ "

observations on  $X_j$ . Then the quantity  $f(\underline{x}|\underline{\theta}) = \prod_{j=1}^{j=k} f(x_j|\theta_j)$  in (1.4) becomes  $f(\underline{x}^*|\underline{\theta}) = \prod_{j=1}^{j=k} f(x_j^*|\theta_j)$  where  $f(x_j^*|\theta_j) = \prod_{\ell=1}^{\ell=r} f(x_j^{(\ell)}|\theta_j)$ ,  $x_j^{(\ell)}$  being the " $\ell$ "<sup>th</sup> observation on  $X_j$ . Thus from (1.4) we have

$$(3.8) \quad \varphi_G(t, \underline{x}^*) = \int_k L(t(\underline{x}^*), \underline{\theta}) f(\underline{x}^*|\underline{\theta}) dG(\underline{\theta}), \quad \text{and}$$

$$(3.9) \quad \varphi_G(a_i, \underline{x}^*) = \int_k L(a_i, \underline{\theta}) f(\underline{x}^*|\underline{\theta}) dG(\underline{\theta}), \quad \text{for}$$

$$a_i \in A_1.$$

Then from (1.6) we see that a decision procedure " $t$ " which minimizes the quantity  $\varphi_G(a_i, \underline{x}^*)$  in (3.9) over  $i = 1, 2, \dots, k$  is called a Bayes decision procedure (denoted by  $t_G$ ) with respect to the a priori distribution  $G = \prod_{j=1}^{j=k} G_j$ .

As indicated in Chapter II (2.2), the linear loss function

$$(3.10) \quad L(a_i, \underline{\theta}) = \theta_{[k]} - \theta_i \text{ for } a_i \in A_1, (i = 1, 2, \dots, k)$$

will be used both in this chapter and Chapter IV. For this loss function we see that for each  $a_i$  in  $A_1$ , (3.9) becomes

$$(3.11) \quad \varphi_G(a_i, \underline{x}^*) = \left[ \int_k \theta_{[k]} f(\underline{x}^* | \underline{\theta}) dG(\underline{\theta}) - \int_k \theta_i f(\underline{x}^* | \underline{\theta}) dG(\underline{\theta}) \right];$$

and this quantity is minimized over "i" when the second integral is maximum since the first integral is independent of "i". Thus the Bayes procedure is defined by:

$$(3.12) \quad t_G(\underline{x}^*) = a_j \text{ where } j \text{ is any integer } 1, \dots, k \text{ such that}$$

$$(3.13) \quad \int_k \theta_j f(\underline{x}^* | \underline{\theta}) dG(\underline{\theta}) = \max_{1 \leq i \leq k} \left\{ \int_k \theta_i f(\underline{x}^* | \underline{\theta}) dG(\underline{\theta}) \right\}.$$

That is, the Bayes procedure selects that population with largest

$$\int_k \theta_i f(\underline{x}^* | \underline{\theta}) dG(\underline{\theta}), \quad i = 1, 2, \dots, k.$$

Thus it remains to evaluate this integral (3.13) for various cases and obtain  $t_G$  as an explicit function of  $\underline{x}^*$  and the parameter  $\underline{\lambda}$ . This will be done for the following cases: (1) Normal-normal (2) Normal-uniform (3) Binomial-beta (4) Poisson-gamma (The first distribution referring to  $f(\underline{x} | \theta_j)$ ; the second to  $G_j$ ) A concluding remark is made regarding the general case.

3.14 Normal-Normal: Let  $f(x | \theta_i)$  be a normal density with mean  $\theta_i$  and variance  $\sigma_i^2$ , and let  $G_i$  be the cumulative distribution function for a normal density  $g_i(\theta)$  with mean  $\lambda_i$  and variance  $\beta_i^2$ . Now we

want to compute the integral given in (3.13). First observe that

$$(3.14.1) \quad \int_k \theta_i f(\underline{x}^* | \underline{\theta}) dG(\underline{\theta}) = \\ \left\{ \int_{-\infty}^{\infty} \theta_i f(\underline{x}_i^* | \theta_i) dG_i(\theta) \right\} \prod_{\substack{j=1 \\ j \neq i}}^{j=k} \left\{ \int_{-\infty}^{\infty} f(\underline{x}_j^* | \theta_j) dG_j(\theta) \right\}$$

Then looking at the first integral on the right hand side of (3.14.1),

we have

$$(3.14.2) \quad \int_{-\infty}^{\infty} \theta_i f(\underline{x}_i^* | \theta_i) dG_i(\theta) = \int_{-\infty}^{\infty} \theta f(\underline{x}_i^* | \theta) g_i(\theta) d\theta \\ = \int_{-\infty}^{\infty} \theta g_i(\theta | \underline{x}_i^*) f_{G_i}(\underline{x}_i^*) d\theta$$

where

$$(3.14.3) \quad f_{G_i}(\underline{x}_i^*) = \int_{-\infty}^{\infty} f(\underline{x}_i^* | \theta) dG_i(\theta) \quad \text{and}$$

$$(3.14.4) \quad g_i(\theta | \underline{x}_i^*) = \frac{f(\underline{x}_i^* | \theta) g_i(\theta)}{f_{G_i}(\underline{x}_i^*)}$$

For the normal case being discussed

$$f(\underline{x}_i^* | \theta_i) = \prod_{\ell=1}^{\ell=r} f(x_i^{(\ell)} | \theta_i) =$$

$$(2\pi\sigma_i^2)^{-\frac{r}{2}} \exp \left\{ -\frac{1}{2\sigma_i^2} \sum_{\ell=1}^{\ell=r} (x_i^{(\ell)} - \theta_i)^2 \right\}$$

$$(3.14.5) \quad f(\underline{x}_i^* | \theta_i) = \\ (2\pi\sigma_i^2)^{-\frac{r}{2}} \exp \left\{ -\frac{1}{2\sigma_i^2} [rs_i^2 + r(\bar{x}_i - \theta_i)^2] \right\}$$

with

$$(3.14.6) \quad s_i^2 = \frac{1}{r} \sum_{\ell=1}^{\ell=r} (x_i^{(\ell)} - \bar{x}_i)^2 \quad \text{and}$$

$$(3.14.7) \quad \bar{x}_i = \frac{1}{r} \sum_{\ell=1}^{\ell=r} x_i^{(\ell)} .$$

(Note:  $x_i^{(\ell)}$  is the " $\ell^{\text{th}}$ " observation from the " $i^{\text{th}}$ " population.)

Putting (3.14.5) into (3.14.3) we obtain

$$(3.14.8) \quad f_{G_i}(x_i^*) = \left\{ (2\pi)^{-\frac{r+1}{2}} \sigma_i^{-\frac{r}{2}} \beta_i^{-1} \right\} \cdot \left\{ \int_{-\infty}^{\infty} \exp \left[ -\frac{rs_i^2 + r(\bar{x}_i - \theta_i)^2}{2\sigma_i^2} - \frac{(\theta_i - \lambda_i)^2}{2\beta_i^2} \right] d\theta \right\} .$$

Then

$$(3.14.9) \quad f_{G_i}(x_i^*) = \left\{ (2\pi)^{-\frac{r+1}{2}} \sigma_i^{-\frac{r}{2}} \beta_i^{-1} \exp\left(-\frac{rs_i^2}{2\sigma_i^2}\right) \right\} \left\{ \int_{-\infty}^{\infty} \exp -\frac{1}{2} \left( \frac{\bar{x}_i - \theta_i}{\sigma_i/r} \right)^2 \exp -\frac{1}{2} \left( \frac{\theta_i - \lambda_i}{\beta_i} \right)^2 d\theta \right\} .$$

One can see that the last quantity in brackets on the right side of (3.14.8) is

$$(3.14.10) \quad \sqrt{2\pi} \frac{\sigma_i \beta_i}{\sigma_i^2 + r\beta_i^2} \exp \left\{ -\frac{1}{2} \frac{\bar{x}_i - \lambda_i}{\frac{\sigma_i^2}{r} + \beta_i^2} \right\} .$$

Thus putting (3.14.10) into (3.14.9) and simplifying yields

$$(3.14.11) \quad f_{G_i}(x_i^*) = \left\{ \frac{(2\pi)^{-\frac{r}{2}} \sigma_i^{-\frac{r}{2} + 1}}{\sigma_i^2 + r\beta_i^2} \right\} \left\{ \right.$$

$$\exp \left[ -\frac{rs_i^2}{2\sigma_i^2} - \frac{r(\bar{x}_i - \lambda_i)^2}{2(\sigma_i^2 + r\beta_i^2)} \right] \} .$$

Observe further that in (3.14.4) the expression

$$(3.14.12) \quad f(x_i^* | \theta_i) g_i(\theta) = (2\pi)^{-\frac{r+1}{2}} \sigma_i^{-\frac{r}{2}} \beta_i^{-1} \\ \exp \left( -\frac{rs_i^2}{2\sigma_i^2} - \frac{r(x_i - \theta_i)^2}{2\sigma_i^2} - \frac{(\theta_i - \lambda_i)^2}{2\beta_i^2} \right) .$$

Now putting (3.14.12) and (3.14.11) into (3.14.4) we obtain

$$(3.14.13) \quad g_i(\theta | x_i^*) = \left[ (2\pi)^{-\frac{1}{2}} (\sigma_i^2 + r\beta_i^2)^{\frac{1}{2}} (\sigma_i \beta_i)^{-1} \right] \left[ \right. \\ \left. \exp \left\{ -\frac{1}{2} \left( \frac{\sigma_i^2 + r\beta_i^2}{\sigma_i^2 \beta_i^2} \right) \left( \theta_i - \frac{r\beta_i^2 \bar{x}_i + \sigma_i^2 \lambda_i}{\sigma_i^2 + r\beta_i^2} \right)^2 \right\} \right]$$

which is the normal density with mean

$$(3.14.14) \quad \int_{-\infty}^{\infty} \theta g_i(\theta | x_i^*) d\theta = \frac{r\beta_i^2 \bar{x}_i + \sigma_i^2 \lambda_i}{\sigma_i^2 + r\beta_i^2}$$

Therefore (3.14.2) can be written as

$$(3.14.15) \quad \int_{-\infty}^{\infty} \theta_i f(x_i^* | \theta_i) dG_i(\theta) = \frac{r\beta_i^2 \bar{x}_i + \sigma_i^2 \lambda_i}{\sigma_i^2 + r\beta_i^2} \cdot f_{G_i}(x_i^*)$$

and thus the desired evaluation of (3.14.1) is given by:

$$(3.14.16) \quad \int_{\bullet}^k \theta_i f(x_i^* | \theta) dG(\theta) = \left\{ \frac{r\beta_i^2 \bar{x}_i + \sigma_i^2 \lambda_i}{\sigma_i^2 + r\beta_i^2} \right\} f_G(x_i^*)$$

where

$$(3.14.17) \quad f_G(x_i^*) = \prod_{j=1}^{j=k} f_{G_j}(x_j^*)$$

and  $f_{G_j}$  is given by (3.14.3). Note that in (3.14.16) the term  $f_G(x_i^*)$

is independent of "i" and hence the largest of the integrals over

$i = 1, 2, \dots, k$  is determined by the largest of the terms in parentheses in (3.14.16). The following theorem is now immediate.

**Theorem 3.14:** From each of  $k$ -populations let  $r$  independent observations be taken on a random variable which is normally distributed with mean  $\theta_i$  and variance  $\sigma_i^2$ . If  $\theta_i$  is distributed normally with mean  $\lambda_i$  and variance  $\beta_i^2$ , then the Bayes procedure for selecting the "best" population (i.e. the population with largest  $\theta_i$ ) under the linear loss function (3.10) is given by:

$t_G(\underline{x}^*) = a_j$  where  $j$  is any integer  $1, 2, \dots, k$  such that

$$\frac{r\beta_j^2 \bar{x}_j + \sigma_j^2 \lambda_j}{\sigma_j^2 + r\beta_j^2} = \max_{1 \leq i \leq k} \left\{ \frac{r\beta_i^2 \bar{x}_i + \sigma_i^2 \lambda_i}{\sigma_i^2 + r\beta_i^2} \right\}$$

where  $\bar{x}_i$  is the mean of the  $r$  observations from the  $i^{\text{th}}$  population,  $i = 1, 2, \dots, k$ .

**Proof:** The proof is immediate from the results obtained above in (3.14.16) coupled with (3.12) and (3.13).

**Corollary 3.15:** Let the conditions of Theorem 3.14 hold and in addition assume  $\sigma_i^2 = \sigma^2$  and  $\beta_i^2 = \sigma_0^2$  for all  $i = 1, 2, \dots, k$ . Then the Bayes procedure for selecting the "best" is:

$t_G(\underline{x}^*) = a_j$  where " $j$ " is any integer  $1, 2, \dots, k$  such that

$$\bar{x}_j + \frac{1}{r} \left( \frac{\sigma}{\sigma_0} \right)^2 \lambda_j = \max_{1 \leq i \leq k} \left\{ \bar{x}_i + \frac{1}{r} \left( \frac{\sigma}{\sigma_0} \right)^2 \lambda_i \right\}.$$

**Remark:** The procedure given in Corollary 3.15 is the procedure used by Dunnett [8]. A statement without proof is made therein that the above

procedure is Bayes.

3.16 Sufficiency: It may have occurred to the reader that "sufficient" statistics may be useful to simplify the calculations above. Sufficiency here will be taken in terms of the factorization criterion in which we will say:  $y_j = y_j(x_j^*)$  is sufficient for  $f(x_j^*|\theta_j)$  if

$$(3.16.1) \quad f(x^*|\theta_j) = h(y_j|\theta_j)p(x_j^*),$$

where  $h$  depends upon the observations only through  $y_j$  and  $p$  does not involve the parameter  $\theta_j$ . The following lemma gives the expected result.

Lemma 3.16.2 (Sufficiency Lemma): Suppose  $f(x_j^*|\theta_j)$  admits a sufficient statistic  $y_j = y_j(x_j^*)$  for  $j = 1, 2, \dots, k$ . Then

$$\int_k \theta_j f(\underline{x}^*|\underline{\theta}) dG(\underline{\theta}) = \max_{1 \leq i \leq k} \int_k \theta_i f(\underline{x}^*|\underline{\theta}) dG(\underline{\theta})$$

if and only if

$$\int_k \theta_j h(\underline{y}|\underline{\theta}) dG(\underline{\theta}) = \max_{1 \leq i \leq k} \int_k \theta_i h(\underline{y}|\underline{\theta}) dG(\underline{\theta})$$

where  $h(\underline{y}|\underline{\theta}) = \prod_{j=1}^{j=k} h(y_j|\theta_j)$ .

Proof: From the definition of  $f(\underline{x}^*|\underline{\theta})$  and (3.16.1) we have

$$(3.16.3) \quad f(\underline{x}^*|\underline{\theta}) = \prod_{j=1}^{j=k} f(x_j^*|\theta_j) =$$

$$\prod_{j=1}^{j=k} h(y_j|\theta_j)p(x_j^*) = \left\{ \prod_{j=1}^{j=k} p(x_j^*) \right\} \left\{ \prod_{j=1}^{j=k} h(y_j|\theta_j) \right\} =$$

$$p(\underline{x}^*) h(\underline{y}|\underline{\theta})$$



and  $p(\underline{x}^*)$  is independent of  $\underline{\theta}$  and "i". Hence using (3.16.3) we see that

$$\begin{aligned} \max_{1 \leq i \leq k} \left\{ \int_k \theta_i f(\underline{x}^* | \underline{\theta}) dG(\underline{\theta}) \right\} &= \\ \max_{1 \leq i \leq k} \left\{ p(\underline{x}^*) \int_k \theta_i h(\underline{y} | \underline{\theta}) dG(\underline{\theta}) \right\} & \\ = p(\underline{x}^*) \max_{1 \leq i \leq k} \left\{ \int_k \theta_i h(\underline{y} | \underline{\theta}) dG(\underline{\theta}) \right\} & \end{aligned}$$

and the proof is complete.

Remark: One can see the application of Lemma 3.16.2 to the case just considered by looking at (3.14.5). It is clear that  $y_j = \bar{x}_j$  is a sufficient statistic for  $f(x_j^* | \theta_j)$  when  $\sigma_j^2$ 's are known; and

$$(3.16.4) \quad h(y_j | \theta_j) = \exp \left[ - \frac{r}{2\sigma_j^2} (y_j - \theta_j)^2 \right]$$

Using the results of the lemma and  $h(\underline{y} | \underline{\theta})$  in place of  $f(\underline{x}^* | \underline{\theta})$  simplifies the calculations made in deriving Theorem 3.14. These calculations were made for illustrative purposes but in the following cases, the Sufficiency Lemma will be used when applicable in deriving the Bayes procedures.

3.17 Norman-uniform: Let  $f(x | \theta_j)$  be a normal density with mean  $\theta_j$  and variance  $\sigma_j^2$ , and let  $G_j$  be the cumulative distribution function for a uniform density  $g(\theta; \lambda_j)$  on  $(\lambda_j - d_j, \lambda_j + d_j)$ . We want to compute the integral in (3.13) which can be written as (3.14.1); but since  $y_j = \bar{x}_j$  is a sufficient statistic for  $f(x_j^* | \theta_j)$  when  $\sigma_j^2$

is known, we can apply Lemma 3.16.2 and replace  $f(x_j^*|\theta_j)$  with  $h(y_j|\theta_j)$  given in (3.16.4). Then (3.14.1) becomes

$$(3.17.1) \quad \int_k \theta_i h(\underline{y}|\underline{\theta}) dG(\theta) = \left\{ \int \theta_i h(y_i|\theta_i) dG_i(\theta) \right\} \left\{ \prod_{\substack{j=1 \\ j \neq i}}^{j=k} \left[ \int h(y_j|\theta_j) dG_j(\theta) \right] \right\}$$

and

$$(3.17.2) \quad \int \theta_i h(y_i|\theta_i) dG_i(\theta) =$$

$$\frac{1}{2d_i} \int_{\lambda_i - d_i}^{\lambda_i + d_i} \theta \exp \left\{ -\frac{r}{2\sigma_i^2} (\theta - \bar{x}_i)^2 \right\} d\theta$$

$$= \frac{\bar{x}_i}{2d_i} \int_{\lambda_i - d_i}^{\lambda_i + d_i} \exp \left\{ -\frac{r}{2\sigma_i^2} (\theta - \bar{x}_i)^2 \right\} d\theta -$$

$$\left\{ \frac{\sigma_i^2}{2d_i r} \exp -\frac{r}{2\sigma_i^2} (\theta - \bar{x}_i)^2 \right\} \Bigg|_{\theta = \lambda_i - d_i}^{\theta = \lambda_i + d_i}$$

$$= \left\{ \bar{x}_i \int h(y_i|\theta_i) dG_i(\theta) \right\} -$$

$$\left\{ \frac{\sigma_i^2}{2d_i r} \exp -\frac{r}{2\sigma_i^2} (\theta - \bar{x}_i)^2 \right\} \Bigg|_{\theta = \lambda_i - d_i}^{\theta = \lambda_i + d_i}$$

Using this expression in (3.17.1), we obtain

$$(3.17.3) \quad \int_k \theta_i h(\underline{y}|\underline{\theta}) dG(\theta) =$$

$$\left\{ \bar{x}_i - \frac{\frac{\sigma_i^2}{2d_i r} \exp \left[ -\frac{r}{2\sigma_i^2} (\theta - \bar{x}_i) \right]}{\frac{1}{2d_i} \int_{\lambda_i - d_i}^{\lambda_i + d_i} \exp \left[ -\frac{r}{2\sigma_i^2} (\theta - \bar{x}_i)^2 \right] d\theta} \right\}_{\theta = \lambda_i - d_i}^{\theta = \lambda_i + d_i} h_G(\underline{y});$$

where

$$(3.17.4) \quad h_G(\underline{y}) = \prod_{j=1}^{j=k} \int h(y_j | \theta_j) dG_j(\theta) \quad \text{and is}$$

independent of "i". Let  $\varphi(u)$  the standard normal density function and  $\Phi(u)$  the standard normal cumulative distribution function. Then

$$(3.17.5) \quad \int_k \theta_i h(\underline{y} | \theta) dG(\theta) = \left\{ \bar{x}_i + \frac{\frac{\sigma_i^2}{r} \left[ \varphi \left( \frac{r}{\sigma_i} (\lambda_i - d_i - \bar{x}_i) \right) - \varphi \left( \frac{r}{\sigma_i} (\lambda_i + d_i - \bar{x}_i) \right) \right]}{\Phi \left( \frac{r}{\sigma_i} (\lambda_i + d_i - \bar{x}_i) \right) - \Phi \left( \frac{r}{\sigma_i} (\lambda_i - d_i - \bar{x}_i) \right)} \right\} h_G(\underline{y})$$

For convenience let  $U_i = U(\bar{x}_i, \sigma_i, r, \lambda_i, d_i)$  denote the term in braces in (3.17.5). As noted in (3.17.4)  $h_G(\underline{y})$  is independent of "i", hence the largest of the  $U_i$ 's ( $i = 1, 2, \dots, k$ ) determines the Bayes procedure.

Thus we have proved:

**Theorem 3.15:** From each of  $k$ -populations let  $r$  independent observations be taken on a random variable which is normally distributed with mean  $\theta_i$  and variance  $\sigma_i^2$ . If  $\theta_i$  distributed uniformly over the interval  $(\lambda_i - d_i, \lambda_i + d_i)$ , then the Bayes procedure for selecting the "best" population (i.e. the population with the largest  $\theta_i$ ) under the linear loss function (3.10) is given by:  $t_G(\underline{x}^*) = a_j$  where  $j$  is any integer  $1, 2, \dots, k$  such that  $U_j = \max_{1 \leq i \leq k} \{U_i\}$ , where  $U_i$  is defined above.

Proof: The proof is immediate from the results (3.17.1) and (3.17.5), and recalling that (3.12) and (3.13) define a Bayes procedure.

Remark: The above theorem would be very useful in the situation in which the only prior information available is a bound on the parameter space. It would be interesting to investigate the performance of the above decision procedure when in fact the a priori distribution is something other than uniform. Such an investigation was not performed at this time.

3.18 Binomial-beta: Let  $f(x|\theta_j) = \binom{u_j}{x} \theta_j^x (1 - \theta_j)^{u_j - x}$ ,

the binomial distribution with  $u_j$  trials,  $x$  successes, and  $\theta_j$  the probability of a success on a single trial. Let  $G_j$  be the cumulative distribution function for a beta density  $g_j(\theta) = c_j \theta^{\lambda_j - 1} (1 - \theta)^{\nu_j - 1}$ .

That is,  $\theta_j$  has a beta distribution on  $(0, 1)$  with parameters  $\nu_j$  and  $\lambda_j$ . Even though a sufficient statistic exists for  $f(x_j^*|\theta_j)$  the computations are not simplified; so Lemma 3.16.2 is not used here. For this case

(3.18.1)  $f(x_i^*|\theta_i) = \prod_{\ell=1}^{\ell=r} f(x_i^{(\ell)}|\theta_i) = V_i \theta_i^{S_i} (1 - \theta_i)^{ru_1 - S_i}$

where  $V_i = \prod_{\ell=1}^{\ell=r} \binom{u_i}{x_i^{(\ell)}}$  and  $S_i = \sum_{\ell=1}^{\ell=r} x_i^{(\ell)}$ . Putting

(3.18.1) into (3.14.3) we obtain

(3.18.2)  $f_{G_i}(x_i^*) =$

$$\int_0^1 V_i c_i \theta^{S_i + \lambda_i - 1} (1 - \theta)^{ru_1 - S_i + \nu_i - \lambda_i - 1} d\theta.$$

Then using the identity:

$$(3.18.3) \quad \int_0^1 p^{a-1}(1-p)^{b-1} dp = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad \text{where}$$

$$(3.18.4) \quad \Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt, \quad \text{the gamma function,}$$

in (3.18.2) yields:

$$(3.18.5) \quad f_{G_i}(x_i^*) = V_i c_i \frac{\Gamma(S_i + \lambda_i) \Gamma(v_i + ru_i - S_i - \lambda_i)}{\Gamma(v_i + ru_i)}.$$

Now with (3.18.1) and the beta density for  $f$  in (3.14.2)

we have

$$(3.18.6) \quad \int_0^1 \theta_i f(x_i^* | \theta_i) dG_i(\theta) =$$

$$\int_0^1 \theta V_i \theta^{S_i} (1-\theta)^{ru_i - S_i} c_i \theta^{\lambda_i - 1} (1-\theta)^{v_i - \lambda_i - 1} d\theta.$$

Using (3.18.3) in (3.18.6), we have

$$(3.18.7) \quad \int \theta_i f(x_i^* | \theta_i) dG_i(\theta) =$$

$$V_i c_i \frac{\Gamma(S_i + \lambda_i + 1) \Gamma(v_i + ru_i - S_i - \lambda_i)}{\Gamma(v_i + ru_i + 1)}$$

For the gamma function  $\Gamma(a+1) = a\Gamma(a)$ , hence (3.18.7) becomes

$$(3.18.8) \quad \int \theta_i f(x_i^* | \theta_i) dG_i(\theta) = \frac{S_i + \lambda_i}{v_i + ru_i} f_{G_i}(x_i^*)$$

where  $f_{G_i}(x_i^*)$  is given by (3.18.5). Therefore putting (3.18.8) into (3.14.1) we obtain:

$$(3.18.9) \quad \int_k \theta_i f(x^* | \theta) dG(\theta) = \left\{ \frac{S_i + \lambda_i}{v_i + ru_i} \right\} f_G(x^*) \quad \text{where}$$

$$(3.18.10) \quad f_G(x^*) = \prod_{j=1}^{j=k} f_{G_j}(x_j^*) =$$

$$\prod_{j=1}^{j=k} \int f(x_j^* | \theta_j) dG_j(\theta) .$$

Since  $f_G(x^*)$  is independent of "i" the Bayes procedure is determined by the largest of the terms  $\frac{S_i + \lambda_i}{v_i + ru_i}$ . Hence the following theorem

has been proved.

**Theorem 3.18:** From each of k-populations let r independent observations be taken on a binomial random variable with probability of success  $\theta_i$  and  $u_i$  the number of trials. If  $\theta_i$  has a beta distribution on (0, 1) with parameters  $v_i$  and  $\lambda_i$ , then the Bayes procedure for selecting the "best" population (i.e. the population with the largest  $\theta_i$ ) under the linear loss function (3.10) is given by:

$t_G(x^*) = a_j$  where j is any integer 1, 2, ..., k such that

$$\frac{S_j + \lambda_j}{v_j + ru_j} = \max_{1 \leq i \leq k} \left\{ \frac{S_i + \lambda_i}{v_i + ru_i} \right\} , \quad \text{where}$$

$$S_i = \sum_{\ell=1}^r x_{i\ell} .$$

Proof: The result obtained above (3.18.9) used in the definition of a Bayes procedure in (3.12) and (3.13) completes the proof.

3.19 Poisson-Gamma: Let  $f(x|\theta_j)$  be a Poisson distribution with parameter  $\theta_j > 0$ , and  $G_j(\theta)$  be the cumulative distribution function for a gamma density  $g_j(\theta) = [\Gamma(\lambda_j)]^{-1} a_j^{\lambda_j} \theta^{\lambda_j - 1} e^{-a_j \theta}$  in which  $a_j$  and  $\lambda_j$  are non-negative parameters. Following the procedure used in the previous cases we write:

$$(3.19.1) \int_{\mathbf{k}} \theta_1 f(x^*|\theta) dG(\theta) = \left\{ \int \theta_1 f(x_1|\theta_1) dG_1(\theta) \right\} \left\{ \prod_{\substack{j=1 \\ j \neq 1}}^{j=k} \int f(x_j^*|\theta_j) dG_j(\theta) \right\} .$$

Then we compute

$$(3.19.2) f_{G_j}(x_j^*) = \int f(x_j^*|\theta_j) dG_j(\theta) = \int \left( \prod_{\ell=1}^{\ell=r} f(x_j^{(\ell)}|\theta_j) \right) dG_j(\theta) = \int_0^{\infty} c_j e^{-r\theta} e^{rx_j^* \theta} [\Gamma(\lambda_j)]^{-1} a_j^{\lambda_j} \theta^{\lambda_j - 1} e^{-a_j \theta} d\theta$$

using  $f$  and  $G_j$  as given above and letting

$$(3.19.3) c_j = \left[ \prod_{\ell=1}^{\ell=r} x_j^{(\ell)} \right]^{-1} . \quad (\text{"}\Gamma\text{" indicates the gamma function as in (3.18.4).) \quad \text{Thus, simplifying (3.19.2) yields}$$

$$(3.19.4) f_{G_j}(x_j^*) = c_j [\Gamma(\lambda_j)]^{-1} a_j^{\lambda_j} \frac{\Gamma(rx_j^* + \lambda_j)}{(r + a_j)^{rx_j^* + \lambda_j}} .$$

Next, we compute

$$\begin{aligned}
 (3.19.5) \quad & \int c_i f(x_i^* | \theta_i) dG_i(\theta) = \\
 & \int_0^{\infty} \theta c_i e^{-r\theta} \frac{r \bar{x}_i}{\theta} [\Gamma(\lambda_i)]^{-1} a_i^{\lambda_i} \theta^{\lambda_i - 1} e^{-a_i \theta} d\theta = \\
 & c_i [\Gamma(\lambda_i)]^{-1} a_i^{\lambda_i} \frac{\Gamma(r \bar{x}_i + \lambda_i + 1)}{(r + a_i)^{r \bar{x}_i + \lambda_i + 1}} .
 \end{aligned}$$

Using the property of the gamma function as in (3.18.8), and (3.19.4) in (3.19.5) gives

$$(3.19.6) \quad \int \theta_i f(x_i^* | \theta_i) dG_i(\theta) = \frac{r \bar{x}_i + \lambda_i}{r + a_i} f_{G_i}(x_i^*) .$$

Then using (3.19.6) and (3.19.2) in (3.19.1), we have

$$(3.19.7) \quad \int_k \theta_i f(\underline{x}^* | \underline{\theta}) dG(\underline{\theta}) = \left\{ \frac{r \bar{x}_i + \lambda_i}{r + a_i} \right\} f_G(\underline{x}^*)$$

where

$$(3.19.8) \quad f_G(\underline{x}^*) = \prod_{j=1}^{j=k} f_{G_j}(x_j^*) .$$

The following theorem can now be proved.

**Theorem 3.19:** From each of  $k$ -populations let  $r$  independent observations be taken on a Poisson random variable with parameter  $\theta_i$  ( $i = 1, 2, \dots, k$ ). If  $\theta_i$  has a gamma distribution with parameters  $a_i$  and  $\lambda_i$ , then the Bayes procedure for selecting the "best" (i.e. the population with the largest  $\theta_i$ ) under the linear loss function (3.10) is given by:



$t_G(\underline{x}^*) = a_j$  where  $j$  is any integer  $1, 2, \dots, k$  such that

$$\frac{r\bar{x}_j + \lambda_j}{r + a_j} = \max_{1 \leq i \leq k} \left\{ \frac{r\bar{x}_i + \lambda_i}{r + a_i} \right\}.$$

Proof: The proof follows immediately from the result (3.19.7) above and the definition of a Bayes procedure given in (3.12) and (3.13).

3.20 General case: The preceding cases indicate that a more general result is possible. For suppose that  $G_i$  possesses a density function  $g_i(\theta)$ . Then as in section 3.14 using (3.14.1), (3.14.2) and (3.14.4). (assuming of course  $f$  is such that (3.14.4) is meaningful), we can write:

$$(3.20.1) \quad \int_k \theta_i f(\underline{x}^* | \underline{\theta}) dG(\underline{\theta}) =$$

$$\left\{ \int \theta_i f(x_i^* | \theta_i) g_i(\theta) d\theta \right\} \left\{ \prod_{\substack{j=1 \\ j \neq i}}^{j=k} f_{G_j}(x_j^*) \right\} =$$

$$\int \theta g(\theta | x_i^*) d\theta \cdot f_G(\underline{x}^*) = E[\theta | x_i^*] \cdot f_G(\underline{x}^*),$$

in which we have set

$$(3.20.2) \quad E[\theta | x_i^*] = \int \theta g(\theta | x_i^*) d\theta.$$

The quantity  $E[\theta | x_i^*]$  is called the a posteriori mean for the  $i^{\text{th}}$  population,  $i = 1, 2, \dots, k$ . Then from (3.12) and (3.13) it is seen that the largest a posteriori mean defines the Bayes procedure with respect to  $G$ . This of course is intuitively reasonable since in general

estimation theory, the Bayes estimate is the a posteriori mean. (e.g. under squared error loss, if " $a$ " is a parameter of a distribution for an observable random variable  $X$ , then the Bayes estimate  $t(x)$  of " $a$ " is obtained by minimizing the overall expected loss,  $E[t(x) - a]^2$ ). But this quantity is minimized by taking  $t(x) = E[a|x]$ , which is the a posteriori mean, and can be computed of course only if the a priori distribution is known.) Thus the general theorem is:

Theorem 3.20: From each of  $k$ -populations let  $r$  independent observations be taken on a random variable which has a density function  $f(x|\theta_i)$ , ( $i = 1, 2, \dots, k$ ). If  $\theta_i$  is distributed according to the density  $g_i(\theta)$ , then the Bayes procedure for selecting the "best" (i.e. the population with largest  $\theta_i$ ) under the linear loss function (3.10) is given by:

$$t_G(\underline{x}^*) = a_j \text{ where } j \text{ is any integer } 1, 2, \dots, k \text{ such that}$$

$$E[\theta|x_j^*] = \max_{1 \leq i \leq k} \{E[\theta|x_i^*]\}, \text{ where } E[\theta|x_i^*] \text{ is}$$

the a posteriori mean of the  $i^{\text{th}}$  population and is given by (3.20.2).

Proof: Using the conclusion in (3.20.1) and the definition of the Bayes given in (3.12) and (3.13), the proof is immediate.

Remark: If  $f(x_j^*|\theta_j)$  admits a sufficient statistic  $y_j$  in the sense of (3.16.1) then by the Sufficiency Lemma (3.16.2,  $E[\theta|x_j^*]$  in the above theorem can be replaced by  $E[\theta|y_j]$  and this often will reduce the computations involved.

## CHAPTER IV

PARAMETRIC-CASE-EMPIRICAL BAYES  
PROCEDURES FOR SELECTING THE BEST

## A. Introduction

In this chapter we are continuing the investigation started in Chapter III into the problem of selecting the "best" of  $k$ -populations; and in particular we want to derive empirical Bayes procedures that will accomplish this. The notation and background here is the same as that given in Chapter III and will not be repeated here; particular reference should be made to the Introduction and Part C, (4), of that chapter. Thus for each of  $k$ -populations there exists an observable random variable  $X_j$  with distribution  $f(x|\theta_j)$  depending upon a parameter  $\theta_j$  which in turn is assumed to have an a priori distribution  $G_j$ , ( $j = 1, 2, \dots, k$ ).

$G = \prod_{j=1}^{j=k} G_j$  is then an a priori distribution on the parameter space  $\Theta^k$ ; and this product is always intended when referring to  $G$ . If  $G$  is specified completely, then the Bayes procedure for selecting the "best" under the linear loss function (3.10) can be obtained from Theorem 3.20 depending of course upon whether or not the a posteriori mean can be computed. However, if only the existence of  $G$  is assumed, then clearly the Bayes procedure is not available to us, and hence other "optimal" decision procedures must be sought. One such "optimal" decision procedure is the so called empirical Bayes procedure and was first suggested by Robbins [28] (For a general decision theoretic discussion of empirical Bayes procedures, see Robbins [30].) In this investigation we are interested in a particular application of the empirical Bayes approach to decision theory. So, what is said here will be specifically related to

the problem of selecting the "best" of  $k$ -populations.

The idea then of empirical Bayes related to this problem is first of all to assume only the existence of an a priori distribution  $G$  on the parameter space  $\Theta^k$ . Thus the Bayes procedure is not available; but suppose independent observations  $(\underline{x}_1^*, \theta_1), (\underline{x}_2^*, \theta_2), \dots, (\underline{x}_n^*, \theta_n)$  on a random variable  $X$  are available with  $\theta_i$ 's all being drawn from the same a priori distribution  $G$ . (Note: the "\*" on each  $\underline{x}_i^*$  indicates that "r" observations from each population have been taken for each  $i = 1, 2, \dots, n$ .) The "prior observations" contain information about  $G$  and thus if a decision procedure  $t_n$  based upon  $\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_n^*$  could be found such that  $R(t_n, G)$  converges to  $R(t_G, G)$ , (i.e. the Bayes risk of  $t_n$  converges to the Bayes risk of the Bayes procedure  $t_G$  which we would use if we knew  $G$  at the start), for any  $G$  in some family  $\tilde{G}$ , then the procedure  $t_n$  is asymptotically optimal to  $t_G$  and is called an empirical Bayes procedure with respect to the unknown  $G$ . Now Theorem 3.1 gives a method for finding empirical Bayes procedures for this problem. Namely, if a "suitable" estimate  $G_n$  based upon  $\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_n^*$  can be found, then the Bayes procedure with respect to  $G_n$  is empirical Bayes with respect to  $G$  provided the other conditions of the theorem are satisfied. To find  $G_n$ , we have made an additional restriction on  $G$  in Chapters III and IV; that is, we assume  $G$  to be a member of some parametric family  $\tilde{G}$  with parameter  $\underline{\lambda} = (\lambda_1, \dots, \lambda_k)$ . (Chapter V will be concerned with this same problem, only without this restriction on  $G$ .) Suppose now an estimate  $\lambda_{nj}$  depending upon the prior observations from the  $j^{\text{th}}$  population of  $\lambda_j$  can be found such that  $G_{nj}$  converges to  $G_j$  with probability one, probability being taken with respect to the random variables  $X_{1j}, X_{2j}, \dots$ . Then by the lemmas

below,  $G_{\pi,n} = \prod_{j=1}^{j=k} G_{nj}$  converges to  $G = \prod_{j=1}^{j=k} G_j$  with probability one, probability here being the product of the above probabilities. Furthermore  $G_{\pi,n}$  is also a member of  $\tilde{G}$  and if the Bayes procedure with respect to any  $G$  in  $\tilde{G}$  is available, then in particular the Bayes procedure with respect to  $G_{\pi,n}$  is available and thus an empirical Bayes procedure with respect to  $G$  is obtained via Theorem 3.1. But Chapter III has dealt with the problem of Bayes procedures for various classes  $\tilde{G}$ . Hence, using the notation and ideas established in Chapter III, it will suffice to:

(4.1) Verify the hypotheses of Theorem 3.1; namely that  $L(a_i, \theta)f(x|\theta)$  is bounded and continuous in  $\theta$  for each  $x$  and any  $a_i$  in  $A_1$ , and that  $\int_{\theta^k} L(\theta)dG(\theta) < \infty$ ; and

(4.2) Find an estimate  $\lambda_{nj}$  of  $\lambda_j$  such that  $G_{nj}$  converges to  $G_j$  with probability one for  $j = 1, 2, \dots, k$ ; the estimate  $\lambda_{nj}$  being a function of the prior observations from the  $j^{\text{th}}$  population.

Hence, after satisfying (4.1) and (4.2) above, direct application of Theorem 3.1 will yield empirical Bayes procedures for selecting the "best" in each situation discussed in Chapter III; i.e. 3.14, 3.17, 3.18, and 3.19. We proceed with two preliminary lemmas which prove that (4.2) implies  $G_{\pi,n} = \prod_{j=1}^{j=k} G_{nj}$  converges to  $G = \prod_{j=1}^{j=k} G_j$  with probability one. After the lemmas a remark is made relative to the second requirement in (4.1), and then the specific classes  $\tilde{G}$  treated in Chapter III are discussed with the corresponding empirical Bayes procedures being obtained thereby.

## B. Preliminary Lemmas

Lemma 4.3: Let  $G_j$  be a one dimensional cumulative distribution function and suppose  $G_{nj}$  ( $n = 1, 2, \dots$ ) is a sequence of cumulative distribution functions converging to  $G_j$  at the points of continuity of  $G_j$ , ( $j = 1, 2, \dots, k$ ). Then  $G_{\pi, n} = \prod_{j=1}^{j=k} G_{nj}$  converges to  $G = \prod_{j=1}^{j=k} G_j$  at the points of continuity of  $G$ .

Proof: Let  $\underline{\theta} = (\theta_1, \dots, \theta_k)$  be a point of continuity of  $G$ . If  $G(\underline{\theta}) \neq 0$  then clearly  $\theta_j$  is a point of continuity of  $G_j$  for  $j = 1, 2, \dots, k$  and thus  $\prod_{j=1}^{j=k} G_{nj} \rightarrow \prod_{j=1}^{j=k} G_n$  at  $\underline{\theta}$ . Suppose then that  $G(\underline{\theta}) = 0$ . Then  $G_j(\theta_j) = 0$  for  $j \in K$ , a subset of  $\{1, 2, \dots, k\}$ . Certainly  $K$  is non-empty, otherwise  $G(\underline{\theta}) \neq 0$ ; and for some  $j$  in  $K$ ,  $\theta_j$  must be a point of continuity of  $G_j$ . For otherwise  $\underline{\theta}$  cannot be a point of continuity of  $G$ . Thus  $G_{nj}(\theta_j) \rightarrow G_j(\theta_j) = 0$  which gives  $G_{\pi, n} \rightarrow G$  at  $\underline{\theta}$  and the proof is complete.

Lemma 4.4: Let  $X_1, X_2 \dots$  be independent  $k$ -dimensional random vector consisting of  $k$  components which are independent random variables. Let  $G_{nj}$  be a cumulative distribution function on and suppose  $G_{nj}$  is a function of  $X_{1j}, X_{2j}, \dots, X_{kj}$  such that  $G_{nj}$  converges with probability one to  $G_j$ , a cumulative distribution function on , ( $j = 1, 2, \dots, k$ ); probability being taken with respect to the random variables  $X_{1j}, X_{2j}, \dots$ . Then  $G_{\pi, n} = \prod_{j=1}^{j=k} G_{nj}$  converges with probability one to  $G = \prod_{j=1}^{j=k} G_j$ , probability here being the product of the above probabilities.

Proof: Let  $B_j = \{y = (x_{1j}, x_{2j}, \dots): G_{nj} \rightarrow G_j \text{ at } y\}$ . Then  $P\{B_j\} = 1$  by hypothesis, ( $j = 1, 2, \dots, k$ ). Then by Lemma 4.3  $G_{\pi, n} = \prod_{j=1}^{j=k} G_{nj} \rightarrow G = \prod_{j=1}^{j=k} G_j$  for  $y \in B = \prod_{j=1}^{j=k} B_j$ , and hence  $P\{B\} = \prod_{j=1}^{j=k} P\{B_j\} = 1$  which completes the proof.

Now as indicated in (4.2) if an estimate  $\lambda_{nj}$  of  $\lambda_j$ , the parameter of  $G_j$ , is available with the property that  $G_{nj}$ , the cumulative distribution function corresponding to  $\lambda_{nj}$ , converges to  $G_j$  with probability one, then by Lemma 4.4  $G_{\pi, n} = \prod_{j=1}^{j=k} G_{nj}$  converges to  $G = \prod_{j=1}^{j=k} G_j$  with probability one. Examples of such estimates  $\lambda_{nj}$  will be given in Part D for the various classes  $\tilde{G}$  discussed in Chapter III. Note that Lemma 4.4 assumes  $r = 1$  (i.e. only one observation on the random variable  $X_{ij}$ ), but it can be extended to the case in which  $r$  is any positive integer.

### C. Discussion of (4.1)

To verify the hypotheses of Theorem 3.1 two quantities must be considered. They are given in (4.1). The boundedness and continuity requirement on the function  $L(a_i, \underline{\theta}) f(\underline{x}|\underline{\theta})$  will be discussed in each example in Part D. Now we will show what assumptions are necessary in order that  $\int_{\mathbb{N}^k} L(\underline{\theta}) dG(\underline{\theta}) < \infty$ . As stated in Theorem 3.1

$$(4.5) \quad L(\underline{\theta}) = \max\{L(a_i, \underline{\theta}): a_i \in A_1\},$$

and the loss function being used is the linear loss function (3.10).

Thus

$$(4.6) \quad \int_{\mathbb{N}^k} L(\underline{\theta}) dG(\underline{\theta}) = \int_{\mathbb{N}^k} \max_{1 \leq i \leq k} \{\theta_{[k]} - \theta_i\} dG(\underline{\theta})$$

$$= \int_{\Theta^k} (\theta_{[k]} - \theta_{[1]}) dG(\theta), \text{ where}$$

$$(4.7) \quad \theta_{[1]} = \min_{1 \leq i \leq k} \theta_i.$$

Now

$$(4.8) \quad \int_{\Theta^k} (\theta_{[k]} - \theta_{[1]}) dG(\theta) \leq \int_{\Theta^k} 2 \left\{ \sum_i |\theta_i| \right\} dG(\theta) \\ \leq 2 \sum_i \left\{ \int_{\Theta} |\theta| dG_i(\theta) \right\} \\ \leq 2 \sum_{i=1}^{i=k} E[|\theta_i|].$$

Therefore, using (4.8) in (4.6), gives

$$(4.9) \quad \int_{\Theta^k} L(\theta) dG(\theta) < \infty \text{ if } E[|\theta_i|] < \infty \text{ for all } i = 1, 2, \dots, k.$$

That is, the condition:  $\int_{\Theta^k} L(\theta) dG(\theta)$  is finite, will be satisfied provided each a priori distribution  $G_i$  has finite first absolute moment.

#### D. Empirical Bayes Procedures

We now turn our attention to the various classes  $\tilde{G}$  discussed in Chapter III and derive the corresponding empirical Bayes procedures.

4.10 Normal-normal: Let  $f(x|\theta_i)$  be a normal density with unknown mean  $\theta_i$  and known variance  $\sigma_i^2$ , and let  $G_i$  be the cumulative distribution function for a normal density  $g_i(\theta)$  with unknown but finite mean  $\lambda_i$  and known variance  $\rho_i^2$ . Note that the finite  $\lambda_i$ 's for the normal case implies (4.9) which is one part of (4.1). To check



the other part of (4.1) we look at:

$$(4.10.1) \quad L(a_i, \underline{\theta}) f(\underline{x}^* | \underline{\theta}) = (\theta_{[k]} - \theta_i) \prod_{j=1}^{j=k} f(x_j^* | \theta_j) \\ = M(\theta_{[k]} - \theta_i) \exp \left\{ -\frac{1}{2} \sum_{j=1}^{j=k} \left( \frac{\bar{x}_j - \theta_j}{\sigma_j} \right)^2 \right\}$$

where

$$(4.10.2) \quad M = \prod_{j=1}^{j=k} (2\pi\sigma_j^2)^{-\frac{r}{2}} \exp \left( -\frac{rs_j^2}{2\sigma_j^2} \right) \quad \text{which is a constant}$$

with respect to  $\underline{\theta}$  (for fixed  $\underline{x}^*$ );  $\bar{x}_j$  and  $s_j^2$  given (3.14.7) and (3.14.6) respectively. From (4.10.1) one can see that the function  $L(a_i, \underline{\theta}) f(\underline{x}^* | \underline{\theta})$  is bounded and continuous in  $\underline{\theta}$  for a given  $\underline{x}^*$  and any  $a_i$  in  $A_1$ . Therefore (4.1) is satisfied.

To find an estimate  $\lambda_{nj}$  of  $\lambda_j$  as suggested in (4.2) we look at the prior observations  $x_1^*, x_2^*, \dots, x_n^*$ . From (3.12.11) with  $r$  equal one, one sees that the random variable  $X_j$  has a normal unconditional density with mean  $\lambda_j$  and variance  $\sigma_j^2 + \theta_j^2$ . Thus the  $nr$  prior independent observations on  $X_j$  provided a suitable estimate of  $\lambda_j$ ; namely define  $\lambda_{nj}$  by:

$$(4.10.3) \quad \lambda_{nj} = \bar{\bar{x}}_j = \frac{1}{nr} \left( \sum_{i=1}^{i=n} \sum_{\ell=1}^{\ell=r} x_{ij}^{(\ell)} \right),$$

which is merely the overall average of the  $nr$  observations on  $X_j$ , ( $j = 1, 2, \dots, k$ ). The Strong Law of Large Numbers guarantees that  $\lambda_{nj}$  converges to  $\lambda_j$  almost surely; and thus if we define  $G_{nj}$  by

$G_j$  with  $\lambda_j$  replaced with  $\lambda_{nj}$ , we have  $G_{nj}$  converging to  $G_j$  with probability one which satisfies (4.2). This gives the following theorem.

Theorem 4.10: Let  $f(x|\theta_j)$  be normal with mean  $\theta_j$  and known variance  $\sigma_j^2$ , and let  $\theta_j$  be distributed normally with unknown but finite mean  $\lambda_j$  and known variance  $\beta_j^2$ . Denote this latter cumulative distribution function by  $G_j$  and define  $G(\theta) = \prod_{j=1}^{j=k} G_j(\theta)$ . Let  $\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_n^*$  be independent prior observations of the random matrix  $\underline{X}^*$  and call  $\underline{x}^*$  the present observation of  $\underline{X}^*$ . Then the procedure  $t_{G_{\pi,n}}(\underline{x}^*) = a_j$  where "j" is any integer 1, 2, ..., k such that

$$\frac{r\beta_j^2 \bar{x}_j + \sigma_j^2 \bar{x}_j}{\sigma_j^2 + r\beta_j^2} = \max_{1 \leq i \leq k} \left\{ \frac{r\beta_i^2 \bar{x}_i + \sigma_i^2 \bar{x}_i}{\sigma_i^2 + r\beta_i^2} \right\}$$

is an empirical Bayes procedure with respect to  $G$  for selecting the "best" population under the linear loss function (3.10);  $\bar{x}_j$  being given by (4.10.3).

Proof: By the remarks above we see that (4.1) and (4.2) are satisfied hence Theorem 3.1 can be applied and thus  $t_{G_{\pi,n}}$ , the Bayes procedure

with respect to  $G_{\pi,n}$  is an empirical Bayes procedure. But  $G_{\pi,n} = \prod_{j=1}^{j=k} G_{nj}$  and each  $G_{nj}$  is a normal distribution with mean  $\bar{x}_j$  and variance  $\beta_j^2$ . Thus Theorem 3.14 gives the Bayes procedure under these conditions and  $t_{G_{\pi,n}}$  is obtained by replacing  $\lambda_j$  with  $\bar{x}_j$ . This completes the proof.

4.11 Normal-uniform: Suppose that in 4.10 the a priori distribution  $G_i$  is uniform on  $(\lambda_i - d_i, \lambda_i + d_i)$  where  $\lambda_i$  is unknown but  $d_i$  is a known positive number, ( $i = 1, 2, \dots, k$ ). The same argument as that used in 4.10 shows that (4.1) is satisfied. To find  $\lambda_{nj}$  (using the prior observations  $\underline{x}_1^*, \dots, \underline{x}_n^*$ ), we compute

$$\begin{aligned}
 (4.11.1) \quad E[X_j] &= \int_{\lambda_j - d_j}^{\lambda_j + d_j} (1/2d_j) \left\{ \int_{-\infty}^{\infty} xf(x|\theta_j)dx \right\} d\theta \\
 &= (1/2d_j) \int_{\lambda_j - d_j}^{\lambda_j + d_j} \theta d\theta \\
 &= \lambda_j .
 \end{aligned}$$

Hence by the Strong Law of Large Numbers the estimate  $\lambda_{nj}$  defined by (4.10.3), based upon the  $nr$  independent observations on  $X_j$ , converges almost to  $\lambda_j$ , ( $j = 1, 2, \dots, k$ ). Thus  $G_{nj}$  defined by  $G_j$  with  $\lambda_j$  replaced with  $\bar{x}_j$  (see 4.10.3) converges to  $G_j$  with probability one. We can now prove:

Theorem 4.11: Let the conditions of Theorem 4.10 with the exception that  $G_j$  is a uniform cumulative distribution function on  $(\lambda_j - d_j, \lambda_j + d_j)$ ;  $\lambda_j$  unknown but finite and  $d_j$  a known positive number. Then the procedure

$$t_{G_{\pi,n}}(\underline{x}^*) = a_j \text{ where "j" is any integer } 1, 2, \dots, k \text{ such that}$$

$$\frac{\varphi(\beta_j) - \varphi(\alpha_j)}{\bar{\Phi}(\alpha_j) - \bar{\Phi}(\beta_j)} + \frac{r\bar{x}_j}{\sigma_j^2} = \max_{1 \leq i \leq k} \left\{ \frac{\varphi(\beta_i) - \varphi(\alpha_i)}{\bar{\Phi}(\alpha_i) - \bar{\Phi}(\beta_i)} + \frac{r\bar{x}_i}{\sigma_i^2} \right\}$$

is an empirical Bayes procedure with respect to  $G$  for selecting the "best" population under the linear loss function (3.10);  $\varphi(x)$ ,  $\bar{\Phi}(x)$  being the standard normal density and cumulative distribution functions respectively;  $\alpha_j = (r/\sigma_j^2)^{\frac{1}{2}}(\bar{x}_j + d_j - \bar{x}_j)$ ,  $\beta_j = (r/\sigma_j^2)^{\frac{1}{2}}(\bar{x}_j - d_j - \bar{x}_j)$ , and  $\bar{x}_j$  given by (4.10.3), ( $j = 1, 2, \dots, k$ ).

Proof: From the remarks of 4.11 we see that (4.1) is satisfied.

Furthermore the work above shows that (4.2) is satisfied. Hence Theorem 3.1 can be applied and thus  $t_{G_{\pi,n}}$ , the Bayes procedure with respect to

$G_{\pi,n}$ ; and empirical Bayes procedure with respect to  $G$ ; and Theorem

3.15 gives the Bayes procedure with respect to any uniform cumulative distribution function. Therefore  $t_{G_{\pi,n}}$  is defined as in Theorem 3.15

with  $\lambda_j$  replaced by  $\bar{x}_j$ . This completes the proof.

4.12 Binomial-beta: The distributions here are as given in 3.16 with  $\theta_j$ 's unknown, and  $\lambda_j$ 's unknown but finite. Note that the finiteness of the  $\lambda_j$ 's gives  $\int L(\theta)dG(\theta) < \infty$  by (4.0), which is one part of (4.1). For the other part of (4.1) we look at:

$$(4.12.1) \quad L(a_i, \underline{\theta})f(\underline{x}^*|\underline{\theta}) = (\theta_{[k]} - \theta_i) \prod_{j=1}^{j=k} f(x_j^*|\theta_j)$$

$$= (\theta_{[k]} - \theta_i)(U) \left( \prod_{j=1}^{j=k} \theta_j^{r\bar{x}_j} (1 - \theta_j)^{ru_j - r\bar{x}_j} \right)$$

where

$$(4.12.2) \quad U = \prod_{j=1}^{j=k} [\prod_{\ell=1}^{\ell=r} \binom{u_j}{x_j(\ell)}], \text{ a constant with respect to } \underline{\theta}$$

for fixed  $\underline{x}^*$ . The function in (4.12.1) is continuous in  $\underline{\theta}$  and since  $\mathcal{M}^k$  is bounded, it is also bounded. Thus (4.1) is satisfied.

To find  $\lambda_{nj}$  (using the prior observations  $\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_n^*$ ), we compute

$$(4.12.3) \quad \begin{aligned} E[X_j] &= E_{\theta} \{ E_x[X_j | \theta_j] \} = E_{\theta} \{ u_j \theta_j \} = u_j E_{\theta} \{ \theta_j \} \\ &= u_j (\lambda_j / v_j) \end{aligned}$$

As before the Strong Law of Large Numbers implies that  $\bar{x}_j$  in (4.10.3) converges almost surely to  $(u_j \lambda_j / v_j)$ , and since  $u_j$  and  $v_j$  are known,  $v_j \neq 0$ ,  $\lambda_{nj}$  defined by:

$$(4.12.4) \quad \lambda_{nj} = (v_j / u_j) \bar{x}_j, \text{ converges}$$

almost surely to  $\lambda_j$ . Thus  $G_{nj}$  defined by  $G_j$  with  $\lambda_j$  replaced with  $\lambda_{nj}$  given in (4.12.4) converges to  $G_j$  with probability one.

We can now prove:

Theorem 4.12: Let  $f(x|\theta_j)$  be a binomial distribution,

$\binom{u_j}{x} \theta_j^x (1 - \theta_j)^{u_j - x}$ , and  $\theta_j$  be distributed according to a beta density,  $g_j(\theta) = c_j \theta^{\lambda_j - 1} (1 - \theta)^{v_j - \lambda_j - 1}$  where

$c_j = \Gamma(v_j) / (\Gamma(\lambda_j) \Gamma(v_j - \lambda_j))$ . Let  $G_j$  denote the cumulative distribution

function for  $g_j$  and define  $G = \prod_{j=1}^{j=k} G_j$ . Let  $\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_n^*$  be independent prior observation of the random matrix  $\underline{X}^*$  and consider  $\underline{x}^*$  to be a present observation of  $\underline{X}^*$ . Then the procedure

$$t_{G_{\pi,n}}(\underline{x}^*) = a_j \text{ where "j" is any integer } 1, 2, \dots, k$$

such that

$$\frac{r\bar{x}_j + (v_j/u_j)\bar{x}_j}{ru_j + v_j} = \max_{1 \leq i \leq k} \left\{ \frac{r\bar{x}_i + (v_i/u_i)\bar{x}_i}{ru_i + v_i} \right\}$$

is an empirical Bayes procedure with respect to  $G$  for selecting the "best" population under the linear loss function (3.10);  $\bar{x}$  given in (4.10.3).

Proof: The remarks of this section preceding the theorem show that (4.1) and (4.2) are satisfied, hence Theorem 3.1 guarantees that

$t_{G_{\pi,n}}$  (the Bayes procedure with respect to  $G_{\pi,n}$ ) is empirical Bayes with respect to  $G$ . But Theorem 3.18 gives the Bayes procedure under the given hypotheses on  $f$  and  $G$ ; hence  $t_{G_{\pi,n}}$  is obtained therein by replacing  $\lambda_j$  with  $\bar{x}_j$ . This completes the proof.

4.13 Poisson-gamma: Let  $f(x|\theta_j)$  be a Poisson distribution with parameter  $\theta_j > 0$  and assume  $\theta_j$  is distributed according to  $G_j$ , the cumulative distribution function for a gamma density

$$g_j(\theta) = \frac{1}{\Gamma(\lambda_j)} \alpha_j \lambda_j \theta^{\lambda_j-1} e^{-\alpha_j \theta}; \lambda_j \text{ unknown but finite, } \alpha_j \text{ a}$$

known non-negative number, ( $j = 1, 2, \dots, k$ ). As before the finite

$\lambda_j$ 's with (4.9) given the finiteness of  $\int_k L(\theta) dG(\theta)$ . To see the condition on  $L(a_i, \theta)f(\underline{x}^*|\theta)$ , we write:

$$\begin{aligned}
 (4.13.1) \quad L(a_i, \theta)f(\underline{x}^*|\theta) &= (\theta_{[k]} - \theta_i) \prod_{j=1}^{j=k} f(x_j^*|\theta_j) \\
 &= c \frac{(\theta_{[k]} - \theta_i) (\prod_{j=1}^{j=k} \theta_j^{\bar{x}_j})}{(\exp \sum_{j=1}^{j=k} \theta_j)^r},
 \end{aligned}$$

where

$$(4.13.2) \quad c = \left[ \prod_{j=1}^{j=k} \prod_{\ell=1}^{\ell=r} (r_j^{(\ell)} !)^{-1} \right] \text{ is a constant with}$$

respect to  $\theta$  and for fixed  $\underline{x}^*$ . Thus from (4.13.1) one can observe that the function  $L(a_i, \theta)f(\underline{x}^*|\theta)$  is bounded and continuous in  $\theta$ . Therefore (4.1) is satisfied.

Now to find  $\lambda_{nj}$ , consider the unconditional expectation:

$$(4.13.3) \quad E[X_j] = E_{\theta} \{E_x[X_j|\theta_j]\} = E_{\theta} \{\theta_j\} = (\lambda_j/\alpha_j).$$

Thus as before, by the Strong Law of Large Numbers the  $rn$  independent observations of  $X_j$  provide the proper estimate of  $\lambda_j$ , namely  $\alpha_j \bar{\bar{x}}_j$  given in (4.10.3). Then  $G_{nj}$  with  $\lambda_j$  replaced by  $\alpha_j \bar{\bar{x}}_j$  converges to  $G_j$  with probability one and hence (4.2) is satisfied. The following theorem can now be proved.

Theorem 4.13: Let  $f$  and  $G_j$  be as given above, and define  $G = \Pi G_j$ .

Let  $\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_n^*$  be independent prior observations of the random matrix  $\underline{X}^*$  and consider  $\underline{x}^*$  to be the present observation of  $\underline{X}^*$ . Then the procedure

$$t_{G, \pi, n}(\underline{x}^*) = a_j \text{ where "j" is any integer } 1, 2, \dots, k \text{ such that}$$

$$\frac{r\bar{x}_j + \alpha_j \bar{\bar{x}}_j}{r + \alpha_j} = \max_{1 \leq i \leq k} \left\{ \frac{r\bar{x}_i + \alpha_i \bar{\bar{x}}_i}{r + \alpha_i} \right\}$$

is an empirical Bayes procedure with respect to  $G$  for selecting the "best" population under the linear loss function (3.10);  $\bar{\bar{x}}_j$  being given by (4.10.3).

Proof: The remarks preceding the theorem show that (4.1) and (4.2) are satisfied, and hence Theorem 3.1 can be applied to guarantee that

$t_{G, \pi, n}$ , the Bayes procedure with respect to  $G_{\pi, n}$ , is empirical

Bayes with respect to  $G$ . But Theorem 3.19 gives the Bayes procedures with respect to the gamma family, hence  $t_{G, \pi, n}$  is obtained by replacing  $\lambda_j$  in the procedure of Theorem 3.19 with  $\alpha_j \bar{\bar{x}}_j$  and the proof is complete.

4.14 General case: As indicated in 3.20 the Bayes procedure for selecting the "best" for a known a priori distribution  $G$  is to choose the population with the largest a posteriori mean,  $E[\theta_i | \underline{x}_i^*]$ , ( $i = 1, 2, \dots, k$ ). In the situation in which  $G$  is unknown but prior observations  $\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_n^*$  are available, an empirical Bayes procedure is available via Theorem 3.1 provided an estimate  $G_n$  (a function of the prior observations) of  $G$  with probability one can be found; namely, the



Bayes procedure with respect to  $G_n$ . Therefore the procedure which selects the largest a posteriori mean  $E_n[\theta_i | x_i^*]$ , where  $E_n$  denotes expectation with respect to  $G_n$ , is Bayes with respect to  $G_n$  and empirical Bayes with respect to  $G$ ; provided of course all the conditions of Theorem 3.1 are satisfied. This gives the following theorem (Note: we draw particular attention to the fact the "\*" indicates  $r$  observations are taken from each of  $k$ -populations for a particular value of the parameter  $\theta$ ).

Theorem 4.15: Let  $f(x|\theta_j)$  be the density of an observable random variable  $X_j$  with parameter  $\theta_j \in \mathcal{M}$  and such that  $L(a_i, \theta)f(x|\theta)$  is bounded and continuous in  $\theta$ ;  $f(x|\theta) = \prod_{j=1}^{j=k} f(x_j|\theta_j)$  and  $L(a_i, \theta)$  being given by (3.10). Let  $G_j$  be an a priori distribution on  $\mathcal{M}$  with finite absolute mean. Suppose prior observations on the random vector  $\underline{X} = (X_1, \dots, X_k)$  are available; i.e.  $(\underline{x}_1^*, \theta_1)$ ,  $(\underline{x}_2^*, \theta_2), \dots, (\underline{x}_n^*, \theta_n)$ , with  $(\underline{x}^*, \theta)$  being the present observation. If a cumulative distribution function  $G_{n,j}$  on  $\mathcal{M}$  can be found which is a function of  $x_{1j}^*, x_{2j}^*, \dots, x_{nj}^*$  ( $j = 1, 2, \dots, k$ ) such that  $G_{nj}$  converges to  $G_j$  at the points of continuity of  $G_j$  with probability one, then the procedure:

$$t_{G_{\pi,n}}(\underline{x}^*) = a_j \text{ where "j" is any integer } 1, 2, \dots, k \text{ such that}$$

$$E_n[\theta_j | x_j^*] = \max_{1 \leq i \leq k} \{E_n[\theta_i | x_i^*]\}$$

is an empirical Bayes procedure with respect to  $G$  for selecting the "best" population.

Proof: With  $G_{\pi,n}$  taken as in Lemma 4.4, we see that the conditions of Theorem 3.1 and Theorem 3.20 are satisfied. Since Theorem 3.20 gives the Bayes procedure with respect to an a priori  $G$ , the Bayes procedure with respect to  $G_{\pi,n}$  is given by  $t_{G_{\pi,n}}$  as above.

But Theorem 3.1 asserts  $t_{G_{\pi,n}}$  to be asymptotically optimal to  $t_G$  and hence an empirical Bayes procedure with respect to  $G$ . The proof is thus complete.

## CHAPTER V

### NON-PARAMETRIC CASE - EMPIRICAL BAYES

#### PROCEDURES FOR SELECTING THE BEST

##### A. Introduction

In this chapter we continue our investigation into the general problem of selecting the "best" of  $k$ -populations. The work already done in Chapters III and IV on this same problem has relied heavily upon the assumption that the a priori distribution  $G$  be a member of some parametric family. However, in many situations this may be an undesirable restriction. Thus, using the notation and terminology established in earlier chapters, it is our specific purpose in this chapter to derive empirical Bayes procedures for selecting the "best" of  $k$ -populations under as general assumptions as possible on  $G$ , the a priori distribution, and  $f(x|\theta_j)$  the density of an observable random variable  $X_j$ .

In the strict sense of the term, the work to follow is not really "non-parametric" since it is always assumed that the density  $f(x|\theta_j)$  has an "indexing parameter"  $\theta_j$  which is of primary importance; that is, the "best" population is defined in terms of  $\theta_j$ . But what is intended here is a "non-parametric" approach relative to the distribution  $G$  in contrast to the "parametric" assumption on  $G$  made in Chapters III and IV; and it is in this sense that the term "non-parametric" is used.

The basis of the work done in this chapter is a result due to Robbins [30], and his proof is included for completeness.

## B. Main Theorem

The notation and definitions below use the same background as already established in previous material. Before stating the theorem we introduce the following quantities:

$$(5.1.1) \quad \tilde{G} = \left\{ G: \int_k L(a_i, \underline{\theta}) dG(\underline{\theta}) < \infty \text{ for } a_i \in A_1 \right\}.$$

$$(5.1.2) \quad \Delta_G(a_i, \underline{x}) = \int_k [L(a_i, \underline{\theta}) - L(a_1, \underline{\theta})] f(\underline{x}|\underline{\theta}) dG(\underline{\theta})$$

for arbitrary  $a_1$  in  $A_1$ , ( $a_1$  is chosen for convenience only), and  $\underline{x}$  in  $\chi$ .

$$(5.1.3) \quad \Delta_{i,n}(\underline{x}) = \Delta_{i,n}(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n; \underline{x}) \quad \text{for}$$

$n = 1, 2, \dots$ , and  $i = 2, 3, \dots, k$  is a function of the prior observations of the random variable  $\underline{X}$ , and  $\underline{x}$ , the present observation of  $\underline{X}$ .

Theorem 5.1: Suppose  $\Delta_{i,n}(\underline{x}) \xrightarrow{p} \Delta_G(a_i, \underline{x})$  for  $i = 1, 2, \dots, k$ .

Define  $\Delta_{1,n}(\underline{x}) \equiv 0$  and a decision procedure  $t_n(\underline{x})$  by:

$$t_n(\underline{x}) = a_j \quad \text{where } j \text{ is any positive integer such that}$$

$$\Delta_{j,n}(\underline{x}) = \min_{1 \leq i \leq k} \{ \Delta_{i,n}(\underline{x}) \}.$$

Then  $t_n$  is an empirical Bayes procedure with respect to any  $G$  in  $\tilde{G}$ .

Proof: We must show that

$$(5.1.4) \quad \lim_{n \rightarrow \infty} R(t_n, G) = R(t_G, G) \quad \text{for } G \in \tilde{G}$$

where  $t_G$  denotes the Bayes procedure with respect to  $G$ , and recalling from Chapter I that:

$$(5.1.5) \quad R(t_G, G) = \int_X \varphi_G(t_G, \underline{x}) dx;$$

$$(5.1.6) \quad R(t_n, G) = \int_X E[\varphi_G(t_n, \underline{x})] dx, \quad E \text{ being}$$

taken with respect to the random variable  $X_1, X_2, \dots, X_n$ ;

and

$$(5.1.7) \quad \varphi_G(t, \underline{x}) = \int_{\Theta^k} L(t, \theta) f(\underline{x}|\theta) dG(\theta) \quad \text{for any}$$

decision procedure  $t$ . Thus (5.1.4) follows from the Lebesgue Dominated Convergence Theorem provided:

$$(5.1.8) \quad \lim_{n \rightarrow \infty} E[\varphi_G(t_n, \underline{x})] = \varphi_G(t_G, \underline{x}) \quad \text{for all } \underline{x}, \text{ and}$$

$$(5.1.9) \quad E[\varphi_G(t_n, \underline{x})] \leq H(\underline{x}) \quad \text{for all } n \text{ with}$$

$$\int_X H(\underline{x}) dx < \infty. \quad \text{But if}$$

$$(5.1.10) \quad \varphi_G(t_n, \underline{x}) \xrightarrow{p} \varphi_G(t_G, \underline{x}) \quad \text{for each } \underline{x} \text{ and}$$

$$\varphi_G(t_n, \underline{x}) < \infty \quad \text{for all } n,$$

then (5.1.8) is true. Thus it suffices to show (5.1.9) and (5.1.10).

(i) Let  $\underline{x} \in \chi$ . Define

$$(5.1.11) \quad H(\underline{x}) = \int_{\Theta^k} L(\underline{\theta})f(\underline{x}|\underline{\theta})dG(\underline{\theta}) \quad \text{with}$$

$$(5.1.12) \quad L(\underline{\theta}) = \max \{L(a_i, \underline{\theta}) : a_i \in A_1\} .$$

Then for  $G \in \tilde{\mathcal{G}}$  we have

$$\int_{\Theta^k} L(\underline{\theta})dG(\underline{\theta}) \leq \sum_{i=1}^{i=k} \int_{\Theta^k} L(a_i, \underline{\theta})dG(\underline{\theta}) < \infty ;$$

and by Fubini's Theorem

$$(5.1.13) \quad \int_{\chi} H(\underline{x})d\underline{x} = \int_{\Theta^k} L(\underline{\theta}) \left\{ \int_{\chi} f(\underline{x}|\underline{\theta})d\underline{x} \right\} dG(\underline{\theta}) \\ = \int_{\Theta^k} L(\underline{\theta})dG(\underline{\theta}) < \infty$$

Also from (5.1.7) we see that  $\varphi_G(t_n, \underline{x}) \leq H(\underline{x})$  for all  $n$  and hence  $E[\varphi_G(t_n, \underline{x})] \leq H(\underline{x})$  for all  $n$ . Thus (5.1.9) is satisfied.

(ii) To show (5.1.10) is suffices to prove that  $\varphi_G(t_n, \underline{x})$  converges in probability to  $\varphi_G(t_G, \underline{x})$  since we have shown above that  $\varphi_G(t_n, \underline{x}) < \infty$  for all  $n$ .

Note that from (5.1.2) we can write

$$(5.1.14) \quad \varphi_G(a_i, \underline{x}) = \Delta_G(a_i, \underline{x}) + \varphi_G(a_1, \underline{x}), \quad i = 1, 2, \dots, k.$$

Let  $t_n$  be a decision procedure as defined in the theorem. Then

$$(5.1.15) \quad \Delta_G(t_n, \underline{x}) - \Delta_G(t_G, \underline{x}) = \varphi_G(t_n, \underline{x}) - \varphi_G(t_G, \underline{x})$$

which is non-negative by definition of  $t_G$ . Thus

$$(5.1.16) \quad 0 \leq \Delta_G(t_n, \underline{x}) - \Delta_G(t_G, \underline{x}) = \Delta_G(t_n, \underline{x}) -$$

$$\Delta_{t_n, n}(\underline{x}) + \Delta_{t_n, n}(\underline{x}) - \Delta_{t_G, n}(\underline{x}) + \Delta_{t_G, n}(\underline{x}) -$$

$$\Delta_G(t_n, \underline{x}).$$

(Note that  $\Delta_{t, n}(\underline{x})$  is  $\Delta_{i, n}(\underline{x})$  for some  $i = 1, 2, \dots, k$  if  $t$  is any decision procedure taking action in  $A_1$ .) Let  $\epsilon > 0$ . Then for  $n$  sufficiently large, we have

$$(5.1.17) \quad \Delta_G(t_n, \underline{x}) - \Delta_{t_n, n}(\underline{x}) \leq \epsilon \quad \text{and}$$

$$(5.1.18) \quad \Delta_{t_G, n}(\underline{x}) - \Delta_G(t_G, \underline{x}) \leq \epsilon$$

with probability near one since  $\Delta_{i, n}(\underline{x}) \xrightarrow{P} \Delta_G(a_i, \underline{x})$

for  $i = 1, 2, \dots, k$ . Furthermore, by definition of  $t_n$ , we have

$$(5.1.19) \quad \Delta_{t_n, n}(\underline{x}) \leq_{t_G, n}(\underline{x}) \quad \text{for all } n.$$

Hence using (5.1.7), (5.1.18), and (5.1.19) in (5.1.15), we obtain

$$(5.1.20) \quad 0 \leq \Delta_G(t_n, \underline{x}) - \Delta_G(t_G, \underline{x}) \leq 2 \epsilon$$

with probability near one for  $n$  sufficiently large; and therefore by (5.1.15) we have the desired result that  $\varphi_G(t_n, \underline{x}) \xrightarrow{P} \varphi_G(t_G, \underline{x})$ .

The proof is thus completed.

Remarks: (1) We indicate how the above theorem could be used to prove Theorem 3.1. If  $G_n$  is an estimate of  $G$  such that  $G_n \xrightarrow{\text{a.s.}} G$ , then  $\Delta_{i,n}(\underline{x})$  taken as  $\Delta_{G_n}(a_i, \underline{x})$  defined in (5.1.2) converges in probability to  $\Delta_G(a_i, \underline{x})$ . Then it follows from Theorem 5.1 that  $t_n$ , defined therein, is an empirical Bayes procedure. Further it can be shown that  $t_n$  is also Bayes with respect to  $G_n$ . Theorem 3.1 is now an immediate result of Theorem 5.1. However, it was felt that the proof as given for Theorem 3.1 was more directly related to the problem at hand at that point and was more instructive relative to the development of later work which was based upon the theorem.

(2) The above proof is for the case  $r = 1$  but can be extended for any positive integer  $r$  by replacing  $f(\underline{x}|\underline{\theta})$  with  $f(\underline{x}^*|\underline{\theta})$  and  $\underline{x}$  with  $\underline{x}^r$ .

Theorem 5.1 gives rise to an empirical Bayes procedure for selecting the "best" of  $k$ -populations provided of course the hypotheses of the theorem are satisfied; namely,

$$(5.2) \quad \int_{\Theta^k} L(a_i, \underline{\theta}) dG(\underline{\theta}) < \infty \text{ for } a_i \in A_1, \text{ and}$$

$$(5.3) \quad \Delta_{i,n}(\underline{x}) \xrightarrow{P} \Delta_G(a_i, \underline{x}) \text{ for } i = 2, 3, \dots, k$$



where these terms are defined in (5.1.2) and (5.1.3) respectively. These two conditions will be discussed below using for the loss function the linear loss given in (3.10). This discussion will then be followed by some specific examples in which the conditions below are satisfied. In particular we will consider (1) the normal case. (2) the Poisson, and (3) the class  $\theta^x g(x)h(\theta)$ .

### C. Discussion of the Conditions of Theorem 5.1

5.2 First condition: For any loss function we have

$$(5.2.1) \quad \int_k L(a_i, \theta) dG(\theta) \leq \int L(\theta) dG(\theta)$$

where  $L(\theta)$  is given by (5.1.12). But for the loss function given by (3.10) it has already been shown that the integral on the right in (5.2.1) is finite if the first absolute moment of  $G_i$  ( $i = 1, 2, \dots, k$ ) is finite. (i.e. if the a priori distribution for each population has finite absolute mean.) Thus we define:

$$(5.2.2) \quad \bar{G} = \{G: G = \prod_{j=1}^k G_j \text{ and } G_j \text{ is a cumulative distribution function on } \mathbb{Q} \text{ such that } \int_{\mathbb{Q}} \theta dG_j(\theta) < \infty \text{ for } j = 1, 2, \dots, k\}.$$

Then, if an a priori distribution  $G$  on  $\mathbb{Q}^k$  belongs to  $\bar{G}$ , we will say (5.2) is satisfied.

5.3 Second condition: In accordance with (5.3) our task is to find a suitable estimate of  $\Delta_G(a_i, \underline{x})$  for  $i = 2, 3, \dots, k$ . We will now expand (5.3) for the given loss function (3.10) and arrive at simpler quantities for which we need suitable estimates. Observe that (5.3) can be written as:

$$\begin{aligned}
 (5.3.1) \quad \Delta_G(a_i, \underline{x}^*) &= \int_{\mathbb{M}^k} [L(a_i, \underline{\theta}) - L(a_1, \underline{\theta})] f(\underline{x}^* | \underline{\theta}) dG(\underline{\theta}) \\
 &= \int_{\mathbb{M}^k} (\theta_1 - \theta_i) f(\underline{x}^* | \underline{\theta}) dG(\underline{\theta}) \quad \text{for the}
 \end{aligned}$$

given loss function (3.10) and  $i = 2, 3, \dots, k$ . (Recall that  $a_1 \in A_1$  is arbitrary and was chosen as  $a_1$  merely for convenience.) As in (3.14.1) we can write:

$$\begin{aligned}
 (5.3.2) \quad \int_{\mathbb{M}^k} \theta_i f(\underline{x}^* | \underline{\theta}) dG(\underline{\theta}) &= \left\{ \int_{\mathbb{M}} \theta_i f(x_i^* | \theta_i) dG_i(\theta_i) \right\} \left\{ \right. \\
 &\quad \left. \prod_{\substack{j=1 \\ j \neq i}}^{j=k} \int_{\mathbb{M}} f(x_j^* | \theta_j) dG_j(\theta_j) \right\}
 \end{aligned}$$

Let

$$(5.3.3) \quad \gamma_i(x_i^*) = \int_{\mathbb{M}} \theta_i f(x_i^* | \theta_i) dG_i(\theta_i) \quad \text{and}$$

$$(5.3.4) \quad f_{G_j}(x_j^*) = \int_{\mathbb{M}} f(x_j^* | \theta_j) dG_j(\theta) .$$

Then using (5.3.2), (5.3.3) and (5.3.4) in (5.3.1), we have

$$\begin{aligned}
 (5.3.5) \quad \Delta_G(a_i, \underline{x}^*) &= \left\{ \gamma_1(x_1^*) \prod_{j=2}^{j=k} f_{G_j}(x_j^*) \right\} - \\
 &\quad \left\{ \gamma_i(x_i^*) \prod_{\substack{j=1 \\ j \neq i}}^{j=k} f_{G_j}(x_j^*) \right\} ,
 \end{aligned}$$

for  $i = 2, 3, \dots, k$ . As before let  $\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_n^*$  be prior observations of the random matrix  $\underline{X}^*$ . If sequences  $f_{nj}(x_j^*)$  and  $\gamma_{nj}(x_j^*)$  which

are functions of the prior observations can be found such that

$$f_{nj}(x_j^*) \xrightarrow{p} f_{G_j}(x_j^*) \quad \text{and} \quad \gamma_{nj}(x_j^*) \xrightarrow{p} \gamma_j(x_j^*) \quad \text{for} \\ j = 1, 2, \dots, k, \quad \text{then}$$

$$(5.3.6) \quad \Delta_{i,n}(\underline{x}^*) = \left\{ \gamma_{ni}(x_i^*) \prod_{\substack{j=1 \\ j \neq i}}^{j=k} f_{nj}(x_j^*) \right\} - \\ \left\{ \gamma_{ni}(x_i^*) \prod_{j=1}^{j=k} f_{nj}(x_j^*) \right\}$$

converges in probability to  $\Delta_G(a_i, \underline{x}^*)$  for  $i = 2, 3, \dots, k$ . Thus (5.3) will be satisfied if the sequences of (5.3.6) can be found.

We will now show that a sequence  $f_{nj}(x_j^*)$  can always be found thus reducing the problem to that of finding a sequence  $\gamma_{nj}(x_j^*)$  which converges in probability to  $\gamma_j(x_j^*)$ . Note that due to independence, (5.3.4) can also be written as:

$$(5.3.7) \quad f_{G_j}(x_j^*) = f_{G_j}(x_j^{(1)}, x_j^{(2)}, \dots, x_j^{(r)}) = \prod_{\ell=1}^{\ell=r} f_{G_j}(x_j^{(\ell)}),$$

where

$$(5.3.8) \quad f_{G_j}(x_j^{(\ell)}) = \int f(x_j^{(\ell)} | \theta_j) dG_j(\theta_j), \quad \text{the marginal density} \\ \text{of } X_j^{(\ell)}.$$

Thus we want to estimate this marginal density  $f_{G_j}(x)$  at  $r$  points; namely  $x_j^{(1)}, x_j^{(2)}, \dots, x_j^{(r)}$ . Since the random variable  $X_{ij}$  for each  $(i = 1, 2, \dots, n)$  has the same marginal

density  $f_{G_j}(x)$ , we can use the prior observations as well as the present observation to find our estimate. In particular define

$$(5.3.9) \quad F_{nj}(x_j^{(\ell)}) = ((n+1)r)^{-1} \text{ (total number of prior observations from the } j^{\text{th}} \text{ population which are } \leq x_j^{(\ell)});$$

that is, the so called empirical cumulative distribution function for the  $j^{\text{th}}$  population. Then

$$(5.3.10) \quad f_{nj}(x_j^{(\ell)}) = \frac{F_{nj}(x_j^{(\ell)} + n^{-1/5}) - F_{nj}(x_j^{(\ell)} - n^{-1/5})}{2n^{-1/5}} > 0$$

converges in probability to  $f_{G_j}(x_j^{(\ell)})$  for  $\ell = 1, 2, \dots, r$

and for  $j = 1, 2, \dots, k$ . (See [31], [32], [29].) Note that

other estimates of the density function could be made. However for our purposes here, we require only that the estimate be consistent.

Hence

$$(5.3.11) \quad f_{nj}(x_j^*) = \prod_{\ell=1}^{\ell=r} f_{nj}(x_j^{(\ell)}) \xrightarrow{P} f_{G_j}(x_j^*),$$

for  $j = 1, 2, \dots, k$  and  $f_{nj}(x_j^{(\ell)})$  given by (5.3.10).

Therefore if a sequence  $\gamma_{nj}(x_j^*)$  can be found which converges in probability to  $\gamma_j(x_j^*)$  defined in (5.3.3), then by using such a sequence and  $f_{nj}(x_j^*)$  defined in (5.3.11) in the expression

(5.3.6), we have that  $\Delta_{i,n}(\underline{x}^*)$  converges in probability to  $\Delta_G(a_i, \underline{x})$  for  $i = 2, 3, \dots, k$  thus satisfying condition (5.3).

Such sequences will now be obtained for the specific cases mentioned earlier.

#### D. Empirical Bayes Procedures for Selecting the "Best"

Once the sequence mentioned in 5.3 can be found, the empirical Bayes procedure  $t_n$  is given by Theorem 5.1 in terms of the  $\min\{\Delta_{i,n}(\underline{x}) : i = 1, 2, \dots, k\}$ . Observe that from (5.3.10) and (5.3.11),  $f_{nj}(\underline{x}_j^*) > 0$ ; hence we can write (5.3.6) as:

$$(5.3.13) \quad \Delta_{i,n}(\underline{x}^*) = \left\{ \frac{Y_{ni}(\underline{x}_i^*)}{f_{ni}(\underline{x}_i^*)} - \frac{Y_{ni}(\underline{x}_i^*)}{f_{ni}(\underline{x}_i^*)} \right\} f_{ni}(\underline{x}_i^*)$$

where we have set

$$(5.3.14) \quad f_{ni}(\underline{x}_i^*) = \prod_{j=1}^{j=k} f_{nj}(\underline{x}_j^*) .$$

Hence from (5.3.13) one can see that:

$$(5.3.15) \quad \Delta_{j,n}(\underline{x}^*) = \min_{1 \leq i \leq k} \Delta_{i,n}(\underline{x}^*) \quad \text{if and only if}$$

$$(5.3.16) \quad \frac{Y_{nj}(\underline{x}_j^*)}{f_{nj}(\underline{x}_j^*)} = \max_{1 \leq i \leq k} \left\{ \frac{Y_{ni}(\underline{x}_i^*)}{f_{ni}(\underline{x}_i^*)} \right\} .$$

Therefore we can define  $t_n$  in Theorem 5.1 as follows:

$$(5.3.17) \quad t_n(\underline{x}^*) = a_j \quad \text{where } j \text{ is any integer } 1, 2, \dots, k$$

such that

$$\frac{\gamma_{nj}(x_j^*)}{f_{nj}(x_j^*)} = \max_{1 \leq i \leq k} \left\{ \frac{\gamma_{ni}(x_i^*)}{f_{ni}(x_i^*)} \right\} .$$

5.4. Now in view of 5.2, 5.3 and (5.3.17), it remains only to find conditions under which sequences  $\gamma_{ni}(x_i^*)$  converging in probability to  $\gamma_i(x_i^*)$  can be obtained for  $G$  in  $\bar{G}$ . Once such a sequence is found an empirical Bayes procedure is given by (5.3.17). Three specific situations in which such sequences are found will now be given. First,  $f(x|\theta_j)$  is taken to be a normal density; secondly,  $f(x|\theta_j)$  is taken to be a Poisson distribution with parameter  $\theta_j$ ; and thirdly,  $f(x|\theta_j)$  is assumed to belong to the class of densities expressible as  $\theta_j^x g(x)h(\theta_j)$ . We remark that a general solution to this estimation problem is not available at this time.

5.5 Normal case: Let  $f(x|\theta_j)$  be a normal distribution with mean  $\theta_j$  and variance  $\sigma_j^2$ . Let  $G_j$  be an a priori cumulative distribution function for  $\theta_j$  with finite absolute mean. Then  $G = \prod_{j=1}^{j=k} G_j$  is in  $\bar{G}$ . We desire a sequence  $\gamma_{nj}(x_j^*)$  converging in probability to  $\gamma_j(x_j^*)$  given in (5.3.3). To do this, we first observe that the Lemma 3.16.2 can be used in this case since  $\bar{x}_j$  is a sufficient statistic for

$f(x_j^* | \theta_j)$ . Hence

$$(5.5.1) \quad h(\bar{x}_j | \theta_j) = (r/2\pi\sigma_j^2)^{\frac{1}{2}} \exp[-(r/2\sigma_j^2)(\bar{x}_j - \theta_j)^2]$$

may be used in place of  $f(x_j^* | \theta_j)$  in Theorem 5.1 and in other pertinent equations. Thus we seek a sequence  $\gamma_{nj}(\bar{x}_j)$  converging in probability to  $\gamma_j(\bar{x}_j)$  where

$$(5.5.2) \quad \gamma_j(\bar{x}_j) = \int_{\mathcal{D}} \theta_j h(\bar{x}_j | \theta_j) dG_j(\theta_j) .$$

For this case it can be seen that (5.5.2) gives rise to

$$(5.5.3) \quad \gamma_j(\bar{x}_j) = \bar{x}_j h_{G_j}(\bar{x}_j) + \left(\frac{\sigma_j^2}{r}\right) h'_{G_j}(\bar{x}_j) , \quad \text{where}$$

$$(5.5.4) \quad h_{G_j}(\bar{x}_j) = \int_{\mathcal{D}} h(\bar{x}_j | \theta_j) dG_j(\theta_j) \quad \text{and}$$

$h'_{G_j}(\bar{x}_j)$  is the derivative of (5.5.4) with respect to  $\bar{x}_j$ . To find

$\gamma_{nj}(\bar{x}_j)$ , it remains to find  $g_{nj}(\bar{x}_j)$  which converges in probability

to  $h'_{G_j}(\bar{x}_j)$ . Using the sufficient statistic  $\bar{x}_j$  will change (5.3.9)

and (5.3.10) to

$$(5.5.5) \quad H_{nj}(\bar{x}_j) = (n+1)^{-1} \text{ (total number of } \bar{x}_{ij} \text{'s which} \\ \leq \bar{x}_j \text{ for } i = 1, 2, \dots, n \text{ and including the present} \\ \text{observation } \bar{x}_j \text{),}$$

$$(5.5.6) \quad h_{nj}(\bar{x}_j) = \frac{H_{nj}(\bar{x}_j + n^{-1/5}) - H_{nj}(\bar{x}_j - n^{-1/5})}{2n^{-1/5}}$$

respectively, where  $\bar{x}_{ij} = \frac{1}{r} \sum_{\ell=1}^{\ell=r} x_{ij}^{(\ell)}$ ,  $i = 1, 2, \dots, n$ .

Since  $G$  is in  $\bar{G}$ , it can be shown that the derivative of  $h_{G_j}(\bar{x}_j)$ , exists in a neighborhood of  $\bar{x}_j$  and further using  $f_{nj}$  as in (5.5.6), that the sequence

$$(5.5.7) \quad g_{nj}(\bar{x}_j) = \frac{h_{nj}(\bar{x}_j + n^{-1/5}) - h_{nj}(\bar{x}_j - n^{-1/5})}{2n^{-1/5}}$$

converges in probability to  $h'_{G_j}(\bar{x}_j)$ . (See Robbins [29], p. 204.) Therefore the sequence  $\gamma_{nj}(\bar{x}_j)$  converging to  $\gamma_j(\bar{x}_j)$  in probability is given by

$$(5.5.8) \quad \gamma_{nj}(\bar{x}_j) = \bar{x}_j h_{nj}(\bar{x}_j) + (\sigma_j^2/r) g_{nj}(\bar{x}_j).$$

This allows us to give an empirical Bayes procedure using (5.3.17) which becomes:

$$(5.5.9) \quad t_n(\bar{x}_j) = a_j \text{ where } j \text{ is any integer } 1, 2, \dots, k$$

such that

$$\bar{x}_j + \left( \frac{\sigma_j^2}{r} \right) \frac{g_{nj}(\bar{x}_j)}{h_{nj}(\bar{x}_j)} = \max_{1 \leq i \leq k} \left\{ x_i + \frac{\sigma_i^2}{r} \frac{g_{ni}(\bar{x}_i)}{h_{ni}(\bar{x}_i)} \right\},$$

the functions  $g_{nj}$  and  $h_{nj}$  being given by (5.5.7) and (5.5.6) respectively.



This establishes the following theorem.

**Theorem 5.5:** Let  $f(x|\theta_j)$  be a normal density with mean  $\theta_j$  and variance  $\sigma_j^2$ . Let  $\theta_j$  be distributed according to  $G_j$  such that  $G = \prod_{j=1}^{j=k} G_j$  belongs to  $\bar{G}$  defined in (5.2.2). Let  $\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_2^*$  be independent prior observations of the random matrix  $\underline{X}^*$  with  $\underline{x}^*$  being the present observation of  $\underline{X}^*$ . Then the procedure (5.5.9) is an empirical Bayes procedure for selecting the "best" of k-populations under the linear loss function (3.10).

**5.6 Poisson case:** Let  $f(x|\theta_j)$  be a Poisson distribution with parameter  $\theta_j > 0$ , and  $\theta_j$  distribution according to  $G_j$ . As indicated in 5.4 we seek a sequence  $\gamma_{nj}(x_j^*)$  converging in probability to  $\gamma_{nj}(x_j^*)$  of (5.3.3). For this case we can write (5.3.3) as:

$$\begin{aligned}
 (5.6.1) \quad \gamma_j(x_j^*) &= \int_0^\infty \theta_j f(x_j^* | \theta_j) dG_j(\theta_j) \\
 &= \int_0^\infty \theta_j \frac{(e^{-\theta_j})^r \theta_j^{\sum_{\ell=1}^{\ell=r} r_j^{(\ell)}}}{\prod_{\ell=1}^{\ell=r} (x_j^{(\ell)}!) } dG_j(\theta_j) \\
 &= (x_j^{(1)} + 1) \int_0^\infty \frac{e^{-\theta_j} \theta_j^{x_j^{(1)} + 1}}{(x_j^{(1)} + 1)!} \cdot \frac{(e^{-\theta_j})^{r-1} \theta_j^{\sum_{\ell=2}^{\ell=r} x_j^{(\ell)}}}{\prod_{\ell=2}^{\ell=r} (x_j^{(\ell)}!) } dG_j(\theta_j) \\
 &= (x_j^{(1)} + 1) \int_0^\infty f(x_j^{(1)} + 1 | \theta_j) \prod_{\ell=2}^{\ell=r} f(x_j^{(\ell)} | \theta_j) dG_j(\theta_j) .
 \end{aligned}$$

(Note: the selection of  $x_j^{(1)}$  was determined by convenience only; any  $x_j^{(\ell)}$ ,  $\ell = 1, 2, \dots, r$ , could have been used in the above.)  
 Now due to independence

$$(5.6.2) \quad f_{G_j}(x_j^*) = \prod_{\ell=1}^{\ell=r} f_{G_j}(x_j^{(\ell)}); \text{ but we also have}$$

$$(5.6.3) \quad f_{G_j}(x_j^*) = \int \prod_{\ell=1}^{\ell=r} f(x_j^{(\ell)} | \theta_j) dG_j(\theta_j). \text{ Thus using}$$

(5.6.2) and (5.6.3) in (5.6.1) we have

$$(5.6.4) \quad \gamma_j(x_j^*) = (x_j^{(1)} + 1) f_{G_j}(x_j^{(1)} + 1) \prod_{\ell=2}^{\ell=r} f_{G_h}(x_j^{(\ell)}).$$

However we have established that  $f_{n_j}(x_j^{(\ell)})$  defined in (5.3.10) converges in probability to  $f_{G_j}(x_j^{(\ell)})$ . Therefore, if we define

$$(5.6.5) \quad \gamma_{n_j}(x_j^*) = (x_j^{(1)} + 1) f_{n_j}(x_j^{(1)} + 1) \prod_{\ell=2}^{\ell=r} f_{n_j}(x_j^{(\ell)}),$$

then  $\gamma_{n_j}(x_j^*)$  converges in probability to  $\gamma_j(x_j^*)$  in (5.6.4). Note that we can write

$$(5.6.6) \quad \gamma_{n_j}(x_j^*) = \left\{ (x_j^{(1)} + 1) \frac{f_{n_j}(x_j^{(1)} + 1)}{f_{n_j}(x_j^{(1)})} \right\} f_{n_j}(x_j^*)$$

where  $f_{n_j}(x_j^*) = \prod_{\ell=1}^{\ell=r} f_{n_j}(x_j^{(\ell)})$ , since  $f_{n_j}(x_j^{(1)}) > 0$ .

Then statement 5.4 implies that an empirical Bayes procedure with respect to  $G \in \bar{G}$  is given by

(5.6.7)  $t_n(\underline{x}^*) = a_j$  where  $j$  is any integer  $1, 2, \dots, k$  such that

$$(x_j^{(1)} + 1) \frac{f_{nj}(x_j^{(1)} + 1)}{f_{nj}(x_j^{(1)})} = \max_{1 \leq i \leq k} \left\{ (x_i^{(1)} + 1) \frac{f_{ni}(x_i^{(1)} + 1)}{f_{ni}(x_i^{(1)})} \right\},$$

using (5.6.6) in (5.3.17) and  $f_{nj}$  defined by (5.3.10). We have thus proved:

Theorem 5.6: Let  $f(x|\theta_j)$  be a Poisson distribution with parameter

$\theta_j$ . Let  $\theta_j$  be distributed according to  $G_j$  such that  $G = \prod_{j=1}^k G_j$

belongs to  $\bar{G}$  defined in (5.2.2). Let  $\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_n^*$  be independent prior observations of the random matrix  $\underline{X}^*$ ;  $\underline{x}^*$  being the present

observation. Then the procedure (5.6.7) is an empirical Bayes procedure for selecting the "best" of  $k$ -populations under the linear loss function (3.10).

5.7 A Class of densities: Let  $f(x|\theta_j)$  be expressible as  $\theta_j^x g(x)h(\theta)$ . Then we write  $v_j(x_j^*)$  as:

$$\begin{aligned} (5.7.1) \quad v_j(x_j^*) &= \int_{\mathbb{N}} \theta_j f(x_j^*|\theta_j) dG_j(\theta_j) \\ &= \int_{\mathbb{N}} \theta_j \theta_j^{\sum_{\ell=1}^r x_j^{(\ell)}} \left( \prod_{\ell=1}^r g(x_j^{(\ell)}) \right) (h(\theta_j))^r dG_j(\theta_j) \\ &= \frac{g(x_j^{(1)})}{g(x_j^{(1)} + 1)} \int_{\mathbb{N}} f(x_j^{(1)} + 1|\theta_j) \prod_{\ell=2}^r f(x_j^{(\ell)}|\theta_j) dG_j(\theta_j) \\ &= \frac{g(x_j^{(1)})}{g(x_j^{(1)} + 1)} f_{G_j}(x_j^{(1)} + 1) \prod_{\ell=2}^r f_{G_j}(x_j^{(\ell)}) \end{aligned}$$

using arguments similar to that in 5.6. Since the function  $g$  is known, finding a sequence  $\gamma_{nj}$  depends upon finding a sequence  $f_{nj}$  converging in probability to  $f_{G_j}$  which we have already found in (5.3.10). Thus by defining

$$(5.7.2) \quad \gamma_{nj}(x_j^*) = \frac{g(x_j^{(1)})}{g(x_j^{(1)} + 1)} f_{nj}(x_j^{(1)} + 1) \prod_{\ell=2}^{\ell=r} f_{nj}(x_j^{(\ell)}) ,$$

we have our desired sequence which converges to  $\gamma_j(x_j^*)$  in probability.

Notice that (5.7.2) may be written

$$(5.7.3) \quad \gamma_{nj}(x_j^*) = \left\{ \frac{g(x_j^{(1)})}{g(x_j^{(1)} + 1)} \cdot \frac{f_{nj}(x_j^{(1)} + 1)}{f_{nj}(x_j^{(1)})} \right\} f_{nj}(x_j^*)$$

where  $f_{nj}(x_j^*) = \prod_{\ell=1}^{\ell=r} f_{nj}(x_j^{(\ell)})$ , since  $f_{nj}(x_j^{(1)}) > 0$ . In accord with statement 5.4 the following procedure is empirical Bayes with respect to  $G \in \bar{G}$ :

$$(5.7.4) \quad t_n(\underline{x}^*) = a_j \text{ where } j \text{ is any integer } 1, 2, \dots, k$$

such that

$$\frac{g(x_j^{(1)})}{g(x_j^{(1)} + 1)} \cdot \frac{f_{nj}(x_j^{(1)} + 1)}{f_{nj}(x_j^{(1)})} = \max_{\bar{1} \leq i \leq k} \left\{ \frac{g(x_i^{(1)})}{g(x_i^{(1)} + 1)} \cdot \frac{f_{ni}(x_i^{(1)} + 1)}{f_{ni}(x_i^{(1)})} \right\} ,$$

where  $f_{nj}$  is given by (5.3.10). Thus we have proved:

Theorem 5.7: Let  $f(x|\theta_j)$  be a density such that:

$$f(x|\theta_j) = \theta^x g(x)h(\theta), \quad x > a ;$$

$$= 0, \quad x \leq a \text{ for some constant}$$

"a." Let  $\theta_j$  be distributed according to  $G_j$  such that

$G = \prod_{j=1}^{j=k} G_j$  belongs to  $\bar{G}$ . Let  $\underline{x}_1^*, \underline{x}_2^*, \dots, \underline{x}_n^*$  be independent

prior observations of the random matrix  $\underline{X}^*$ ;  $\underline{x}^*$  being the present

observation. Then the procedure (5.7.4) is an empirical Bayes procedure

for selecting the "best" of  $k$ -populations under the linear loss function (3.10).

Remark on Theorem 5.7: After a suitable transformation a large class of exponential densities can be written as in Theorem 5.7. (See Samuel [32].)

#### ACKNOWLEDGEMENT

The author would like to express his sincere gratitude to his Major Professor S. S. Gupta for suggesting this area of research and for his patience, encouragement and inspiration; and the Office of Naval Research and the Aerospace Research Laboratory for their financial assistance.

REFERENCES

1. Bahadur, R.R. (1950). On the problem in the theory of  $k$  populations, Ann. Math. Statist. (21) 362-375.
2. Bahadur, R.R. and Robbins, H. (1950). The problem of the greater mean, Ann. Math. Statist. (21), 469-487.
3. Bland, R.P. (1961). A minimum average risk solution for the problem of choosing the largest mean. Mimeograph Series No. 280, Institute of Statistics, University of North Carolina, Chapel Hill, N.C.
4. Bechhofer, R.E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist. (25), 16-29.
5. Bechhofer, R.E. and Sobel, M. (1954). A single sample decision procedure for ranking variances of normal populations. Ann. Math. Statist. (25), 273-289
6. Bechhofer, R.E., Dunnett, C.W. and Sobel, M. (1954). A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. Biometrika (41), 170-176
7. Chambers, M.L. and Jarratt, P. (1964). Use of double sampling for selecting best population. Biometrika (51), 49-64.
8. Dunnett, C.W. (1960). On selecting the largest of  $k$  normal population means. J. Roy. Stat. Soc. Ser. B 22, 1-40.
9. Gupta, S.S. (1956). On a decision rule for a problem in ranking means. Mimeograph Series No. 150, Institute of Statistics, University of North Carolina, Chapel Hill, N.C.
10. Gupta, S.S. and Sobel, M. (1957). On a statistic which arises in selection and ranking problems. Ann. Math. Statist. (28), 957-967.
11. Gupta, S.S. and Sobel, M. (1960). Selecting a subset containing the best of several binomial populations. Contributions to Probability and Statistics. Ch. XX. Stanford University Press.
12. Gupta, S.S., and Sobel, M. (1962a). On selecting a subset containing the population with the smallest variance. Biometrika (49) 495-507.

13. Gupta, S.S. and Sobel, M. (1962b). On the smallest of several correlated F statistics. Biometrika (49), 509-523.
14. Gupta, S.S. (1963b). On a selection and ranking procedure for gamma populations. Ann. Inst. Statist. Math. Tokyo (14), 199-216.
15. Gupta, S.S. (1963c). Selection and ranking procedures and order statistics for the binomial distribution. To appear in Proceedings of the International Symposium on discrete distributions held at McGill University, Aug. 1963.
16. Gupta, S.S. (1965). On some multiple decision (selection and ranking) rules, Technometrics (7).
17. Guttman, I and Tiao, G.C. (1964). A Bayesian Approach to some best population problems, Ann. Math. Stat. (35), 825-835.
18. Johns Jr., M.V. (1957). Non-parametric empirical Bayes procedures, Ann. Math. Stat. (28), 649-669.
19. Kagan, A.M. (1962). An empirical Bayes approach to the estimation problem. Doklady of the Academy of Sciences of the USSR (147:5). 1020-1021. See also same Doklady (150:4) 1963.
20. Krutchkoff, R.G. (1964). A two sample non-parametric empirical Bayes approach to some problems in decision theory. Submitted to Ann. Math. Statist.
21. Lehmann, E.L. (1957). A theory of some multiple decision problems, I. Ann. Math. Statist. (28), 1-25.
22. Lehmann, E.L. (1961). Some Model I problems of selection. Ann. Math. Statist. (32), 990-1012.
23. Loeve, M., Probability Theory, University Series in Higher Mathematics, 2nd Edition 1960.
24. Miyasawa, K. (1961). An empirical Bayes estimator of the mean of a normal population, Bulletin of the International Statistical Institute (38:4), 181-188.
25. Neyman, J. (1962). Two breakthroughs in the theory of statistical decision making. Rev. Inst. Internat. Statistics (30) 11-27.
26. Paulson, E. (1949). A multiple decision procedure for certain problems in analysis of variance. Ann. Math. Statist. (20), 95-98

27. Rizvi, M.H. (1963). Ranking and selection problems of normal populations using the absolute values of their means: fixed sample size case. Department of Statistics, University of Minnesota, Technical Report #31.
28. Robbins, H. (1955). An empirical Bayes approach to statistics, Proceedings of the Third Berkeley Symposium on Statistics and Probability, I, 157-164
29. Robbins, H. (1963) The empirical Bayes approach to testing statistical hypotheses. Rev. Inst. Internat. Statist. No. 2, 195-208.
30. Robbins, H. (1964). The empirical Bayes approach to statistical decision problems, Annals of Math. Stat. (35) 1-20
31. Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. Annals Math. Stat. (27) 832-837.
32. Samuel, E. (1963). An empirical Bayes approach to the testing of certain parametric hypotheses. Ann. Math. Stat. (34) 1370-1385.
33. Seal, K. C. (1955). On a class of decision procedures for ranking means of normal populations. Ann. Math. Statist. (26) 387-398.
34. Seal, K. C. (1958). On ranking parameters of scale in type III populations. J. Amer. Stat. Assn. (53), 164-175.
35. Sobel, M. and Huyett, M. M. (1957). Selecting the best one of several binomial populations. Bell System Tech. J. (36), 537-576.



DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) <b>Purdue University</b>		2a. REPORT SECURITY CLASSIFICATION <b>Unclassified</b>	
		2b. GROUP	
3. REPORT TITLE <b>Multiple Decision Procedures from an Empirical Bayes Approach</b>			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) <b>Technical Report - July, 1965</b>			
5. AUTHOR(S) (Last name, first name, initial) <b>Deely, John J.</b>			
6. REPORT DATE <b>July, 1965</b>		7a. TOTAL NO. OF PAGES <b>87</b>	7b. NO. OF REFS <b>35</b>
8a. CONTRACT OR GRANT NO. <b>NONR-1100(26)</b>		9a. ORIGINATOR'S REPORT NUMBER(S) <b>Mimeograph Series Number 45</b>	
b. PROJECT NO.		9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
c.			
d.			
10. AVAILABILITY/LIMITATION NOTICES <b>Releasable without limitations on dissemination</b>			
11. SUPPLEMENTARY NOTES <b>Also supported by Aerospace Research Laboratories Contract AF 33(657)11737</b>		12. SPONSORING MILITARY ACTIVITY <b>Logistics and Mathematical Sciences Branch Office of Naval Research Washington, D.C. 20360</b>	
13. ABSTRACT The problem of selecting the "best" or a subset containing the "best" of k-populations has been discussed in the past mainly from a non-Bayesian approach. Recently a few Bayesian developments have been produced relative to these problems. As an alternate to these two extremes, it is the purpose of this paper to use the so-called empirical Bayes approach in selecting the "best" or a subset containing the "best" under specified loss functions. Three main areas are discussed. Section II deals with the problem of selecting a subset and it is shown that the Bayes procedure for selecting a subset is given by the Bayes procedure for selecting the "best" provided the corresponding loss fcts bear a certain relationship. This result becomes useful in obtaining empirical Bayes procs in the light of results in Sections III and IV. Sections III and IV, as the second main area of investigation, deal with the parametric case for selecting the "best". That is, the a priori distn on the parameter space is assumed to belong to a parametric family. Empirical Bayes procs are obtained from the Bayes procs provided the prior obsns give rise to a suitable estimate of the parameter of the a priori distn. Such estimates are found and hence empirical Bayes procs are obtained for the specific situations: (1) Normal-normal, (2) Normal-uniform, (3) Binomial-beta, (4) Poisson-gamma. (The first distn referring to the density of the observable r.v. and the second referring to the a priori distn.) The third area of investigation, the non-parametric case for selecting the "best", is discussed in Section V. The a priori distn is assumed to possess finite absolute mean. Under this assumption empirical Bayes procs are obtained for the cases in which the density of the observable r.v. is: (1) Normal, (2) Poisson, (3) of the form $\theta^x g(x) h(\theta)$ .			

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

Bayes procedures  
 empirical Bayes procedures  
 multiple decision procedures  
 linear loss function  
 expected loss or risk  
 best of k-populations

**INSTRUCTIONS**

1. **ORIGINATING ACTIVITY:** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (*corporate author*) issuing the report.
- 2a. **REPORT SECURITY CLASSIFICATION:** Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.
- 2b. **GROUP:** Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.
3. **REPORT TITLE:** Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.
4. **DESCRIPTIVE NOTES:** If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.
5. **AUTHOR(S):** Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.
6. **REPORT DATE:** Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.
- 7a. **TOTAL NUMBER OF PAGES:** The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.
- 7b. **NUMBER OF REFERENCES:** Enter the total number of references cited in the report.
- 8a. **CONTRACT OR GRANT NUMBER:** If appropriate, enter the applicable number of the contract or grant under which the report was written.
- 8b, 8c, & 8d. **PROJECT NUMBER:** Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.
- 9a. **ORIGINATOR'S REPORT NUMBER(S):** Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.
- 9b. **OTHER REPORT NUMBER(S):** If the report has been assigned any other report numbers (*either by the originator or by the sponsor*), also enter this number(s).
10. **AVAILABILITY/LIMITATION NOTICES:** Enter any limitations on further dissemination of the report, other than those

imposed by security classification, using standard statements such as:

- (1) "Qualified requesters may obtain copies of this report from DDC."
- (2) "Foreign announcement and dissemination of this report by DDC is not authorized."
- (3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through \_\_\_\_\_."
- (4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through \_\_\_\_\_."
- (5) "All distribution of this report is controlled. Qualified DDC users shall request through \_\_\_\_\_."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. **SUPPLEMENTARY NOTES:** Use for additional explanatory notes.
12. **SPONSORING MILITARY ACTIVITY:** Enter the name of the departmental project office or laboratory sponsoring (*paying for*) the research and development. Include address.
13. **ABSTRACT:** Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **KEY WORDS:** Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.