PROCEDURES FOR SELECTING THE BEST OF SEVERAL POPULATIONS

by

Peter Charles O'Brien

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1970

TABLE OF CONTENTS

PART I. INTRODUCTION

In this thesis we shall consider the problem of selecting the best of several populations assuming that all populations belong to a common known family. In Part II we shall consider selecting the multinomial category with the largest (or smallest) probability. In Part III we apply the results of Part II to selecting the best population when the populations under consideration belong to an incompletely specified class. Selecting the best gamma population is considered in Part IV, while selecting the best normal population is considered in Part V. The problem of selecting the Koopman Darmois population with largest parameter is discussed in Part VI with special reference to gamma, normal, Poisson, Bernoulli, and negative binomial populations. In Part VII we pursue the problem of selecting the best negative binomial population. A modified selection problem which includes a test for homogeneity (i. e. that all populations are identical) is also included. Finally, in Part VIII we consider the interval estimation of the shape parameter of the gamma distribution.

Before pursuing the discussion further it is desirable to state the selection problem precisely.

CHAPTER A. GENERAL MATHEMATICAL FORMULATION OF A SELECTION PROBLEM

In this section we consider the mathematical formulation of a selection problem. It is assumed that one may observe a sequence of random variables $(X_{1j}, \ldots, X_{kj})$ $j = 1, 2, \ldots$, having joint probability density function (p.d.f.) $f_{X_{1j}, \ldots, X_{kj}}(x_1, \ldots, x_k | \theta_1, \ldots, \theta_k)$ where $\{X_{ij}\}_{j=1}^{\infty}$ is said to arise from population i, denoted by π_i. The experimenter specifies constants θ^* and P^*, a function g, and a definition of best population. The problem is then to select the best population using a procedure which is such that the probability of correct selection (PRCS) is no less than P^* whenever $g(\theta_1, \ldots, \theta_k) \gtreqless \theta^*$ (the direction of the inequality is also specified by the experimenter). Thus to completely specify a selection problem, it suffices to specify 1) $f_{X_{1j}, \ldots, X_{kj}}(x_1, \ldots, x_k | \theta_1, \ldots, \theta_k)$ 2) g, together with direction of the inequality, and 3) the definition of best population. Generally, g will depend on the ordered θ_i, denoted by $\theta_{(i)}$ $(\theta_{(1)} \leq \ldots \leq \theta_{(k)})$. The population associated with $\theta_{(i)}$ is denoted by $\pi_{(i)}$.

The constant θ^* is referred to as the indifference constant while g is called the indifference function. The parameter space (H) is assumed to be a k dimensional Euclidean space. A point θ belonging to (H) is called a parametric configuration. That part of (H) in which no requirements are made regarding PRCS, $(H)_1$ say, is called the indifference region. As indicated previously, the indifference region is determined by the experimenter by specifying the indifference function g and the indifference constant θ^*.

Least favorable configuration

A parametric configuration θ which lies outside the indifference region is a least favorable configuration (LFC) with respect to a procedure R if θ minimizes the PRCS over all parametric configurations lying outside the indifference region when procedure R is used.

Slippage configuration

θ is said to be a slippage configuration (SC) if it lies on the boundary of the indifference region and has k-1 equal valued components.

Generalized least favorable configuration

θ is said to be a generalized least favorable configuration (GLFC) if it lies outside the indifference region and has k-1 equal valued components.

Generalized slippage configuration

θ is said to be a generalized slippage configuration (GSC) of type t if t components are equal to c_1, one component is equal to c_2, and the remainder are equal to c_3, where c_1, c_2, and c_3 are any constants such that $c_1 \neq c_3$, the k-dimensional vectors $(c_1, \ldots, c_1, c_2)$ and $(c_3, \ldots, c_3, c_2)$ both lie outside the indifference region, and $(c_3, \ldots, c_3, c_2)$ lies on the boundary of the indifference region.

The selection problems will be indexed using the format "Problem X.y". X will be the capital of the first letter of the p.d.f. (listed in specification 1) and y will be $\underline{a}$ if $\pi_{(k)}$ is to be selected or $\underline{b}$ if $\pi_{(1)}$ is to be selected. Procedures are indexed using the format "Procedure X.y.w" where w denotes the procedure considered for Problem X.y.

For example, Procedure M.a.1 will be the first procedure to be considered for the problem of selecting the multinomial category with largest probability.

CHAPTER B. CRITERIA FOR COMPARISON OF THE PROCEDURES

Procedures will be compared on the basis of

1. administrative simplicity
2. efficiency
3. nature of the prior probability statements, and
4. nature of the posterior probability statements.

These criteria will be discussed in detail in the succeeding sections. It does not appear feasible to obtain procedures which are optimal with respect to all four criteria simultaneously. However we shall develop procedures superior to existing procedures with respect to at least one of the above criteria.

Administrative simplicity

A procedure may be simple to administer in either of two ways. It may have

(a) a sampling rule which is simple for the experimenter to administer, or

(b) a decision rule which is simple for the statistician to evaluate.

Efficiency

A commonly accepted measure of efficiency is the average number of scalar observations (i.e. the sum of the number of observations taken from each population) required to make a selection (also referred to as the average sample number, or ASN). ASN will depend on the parametric configuration, the number of populations from which a selection is to be made, the level of PRCS which is to be maintained, and the

specification of the indifference region. When these quantities are fixed we may consider the relative efficiency of the two procedures (defined to be the ratio of their ASN).

Procedures proposed thus far in the literature require the experimenter to take an excessive number of observations. The primary reasons are that the procedures proposed in the literature require the experimenter to either

1. use a sampling rule which is inflexible (A sampling rule is considered inflexible to the extent to which the manner in which observations are taken must be determined in advance of the experiment.), or
2. achieve a higher level of PRCS than specified by the problem.

Clearly, these two causes are related since a complex sampling rule may make an accurate assessment of PRCS difficult. At present there appear to be three types of sampling rules which have been considered in the literature:

1. single sample
2. sequential nonscreening
3. sequential eliminating.

Single sample procedures probably require the most inflexible sampling plans since the number of observations to be taken from each population is completely determined in advance of the experiment. In the case where the best population is markedly superior to the others a selection may often be possible using fewer observations.

Sequential procedures may overcome the aforementioned difficulty in that sample size is determined on the basis of the observations. Sequential procedures are such that the experimenter alternates the operations of sampling observations and of constructing test statistics. Sampling conducted prior to testing for the rth time but after testing for the r-1st time is termed the rth stage of sampling, or more simply, the rth stage. By a sequential nonscreening procedure we mean a procedure which requires the experimenter to sample only k-dimensional vectors of observations (i.e.: one observation from each population) during each stage of sampling. Such sampling plans, which have been considered extensively by Bechhofer, Kiefer, and Sobel (6), are inflexible in that they require an equal number number of observations to be sampled from each population. This may be undesirable when one or more populations is markedly inferior to the others (for instance, when the parametric configuration is a GSC).

A sequential procedure employing elimination is a procedure in which the experimenter decides after each stage of sampling which populations are to be eliminated from future consideration in the selection procedure. Once a population is eliminated it is never sampled from again and it may not be selected as best. Procedures of this type have been studied prominently by Paulson (33), (34). Typically, a statistic T_{in} $(i=1,\ldots,k;$ $n=1, 2, \ldots)$ is computed after each stage n of sampling for each population i which has not been eliminated. The procedure defines a number T^* such that if

$$T_{max,n} - T_{min,n} > T^*$$

the population associated with $T_{min,n}$ is eliminated. A lower bound for PRCS when the parametric configuration is an SC is determined as

$$Pr(\bigcup_{i=2}^{k} A_i) \leq \sum_{i=2}^{k} Pr(A_i) = (k-1)Pr(A_2) \qquad (I.1)$$

where A_i is the event that $T_{in} - T_{1n}$ is ever greater than T^* (π_1 is assumed best). Empirical studies in this thesis and in O'Brien (30) indicate that the true error rate (1-PRCS) is often about half that computed using the approximation above.

In addition to relying on a conservative evaluation of PRCS, procedures employing elimination are based on a decision rule which is inefficient. Specifically, the selection of a population as best is <u>not</u> based on all the information collected. We illustrate with a hypothetical experiment in which π_2 is eliminated because

$$T_{1n} - T_{2n} > T^*$$

and π_1 is selected because

$$T_{1m} - T_{min,m} > T^* \quad .$$

At the time π_1 is selected, a total of n+m observations are available for the comparison of π_1 with π_2. However, the decision to select

π_1 over π_2 is based on only 2n observations. The failure to use the additional m-n observations represents an inefficient use of the information collected. This could be especially embarrassing if the latter m-n observations on π_1 should indicate that π_2 is superior to π_1.

The sampling rule is also inflexible in that once a population is eliminated it may not be sampled from again.

Nature of the prior and posterior probability statements

The single sample procedures, sequential nonscreening procedures, and sequential procedures employing elimination which have been proposed in the literature are all such that it may be proven analytically that PRCS $\geq$ P* whenever the parametric configuration lies outside the indifference region. This is a prior probability statement in the sense that it is appropriate prior to conducting the experiment.

Posterior probability statements (as defined here) are characterized by the fact that they are conditional on information which becomes available after the inception of the experiment. To date, selection procedures have been derived almost exclusively with the goal of enabling the experimenter to make certain prior probability statements. However, in many applications, posterior statements could be of greater interest. Consider a situation in which an experimenter wishes to select the most probable multinomial category, and suppose it is known that, $p_{(k)} \geq 3p_{(k-1)}$ (with k=3). Following the single sample procedure proposed by Bechhofer et al. (5), one may proceed by taking 30 observations and selecting the category with most outcomes as best, breaking ties using a random device. The prior statement "PRCS $\geq$ .90" is then justified. However, if

one obtains a sample with 10 outcomes from each category, the (posterior) statement of interest is "PRCS = $\frac{1}{3}$".

Similar situations may arise using sequential procedures. In selecting the best of three Poisson populations it may be known that $.99\theta_{(3)} = \theta_{(2)} = \theta_{(1)}$. Yet, using the sequential nonscreening procedure proposed by Bechhofer et. al. (6) a selection could be made after observing only one observation from each population regardless of how large PRCS was stated to be in the prior probability statement.

Even more awkward posterior probability statements are possible if one uses a sequential procedure employing elimination. For example, in using the procedure proposed by Paulson (33) for selecting the normal population with largest mean, it may turn out upon completion of the experiment that the population with largest observed sample mean was the first population eliminated, whereas the population actually selected had the second smallest observed sample mean. Still more awkward situations may be conceived when one incorporates a process of artificially constructed randomization into the experiment. Such a scheme has been proposed by Paulson (34) for selecting the Bernoulli population with smallest probability in order to justify certain prior probability statements. Using this procedure, a population may have the largest observed sample mean (i.e., largest estimate of probability) after each stage of the experiment and yet be selected as the population with smallest probability.

CHAPTER C. SUMMARY

Introduction

In this thesis sequential screening procedures for selecting the best of several populations will be developed. For purposes of this thesis, a screening procedure is defined to be a procedure in which the sampling rule permits the experimenter to sample from a proper subset of the populations during one or more stages of the experiment and the decision rule is such that all populations remain candidates for selection until a selection is actually made. We shall develop two types of screening procedures. An evaluation of the screening procedures which we shall consider in this thesis is given in the next four sections.

Administrative simplicity

The complexity of the decision rules employed by the screening procedures will vary. However, in all the procedures we consider the decision rule is well defined and computationally feasible. Perhaps the most useful aspect of the screening procedures which we shall develop is the flexibility of the sampling rule. The procedures are devised under the assumption that the experimenter specifies the manner in which he desires the observations to be taken. The test procedure then defines the decision rule to be used in confunction with this sampling rule. Two sampling rules which have been examined empirically and found to give satisfactory results are:

Sampling rule 1: Initiate the experiment by sampling one observation from each population. This is defined to be the first stage of sampling.

Upon concluding stage n $(n=1, 2, \ldots)$ screen the population associated with $L_{(t)n}$ $(t=1, \ldots, k-1)$ if

$$\sum_{j=t+1}^{k} \frac{L_{(j)n}}{L_{(t)n}} \geq \frac{k-1}{1-P^*} - t \ , \qquad \text{(I. 2)}$$

where the statistic L_{in} $(i=1, \ldots, k)$ is the likelihood function of either the original observations or the transformation of the original observations under the assumption that the parametric configuration is an SC with π_i best. Instructions for computing $L_{1n}, \ldots, L_{kn}$ are given in the specifications of the individual procedures. At stage $n+1$, sample one observation from each population which is not screened.

Sampling rule 2: Follow the instructions in sampling rule 1 but replace Relation I. 2 with Relation I. 3:

$$\sum_{j=t+1}^{k} \frac{L_{(j)n}}{L_{(t)n}} \geq t(1-2P^*)/(1-P^*)+k-1 \ . \qquad \text{(I. 3)}$$

It may be verified that sampling rule 2 screens more frequently since

$$\frac{k-1}{1-P^*} - t \geq t(1-2P^*)/(1-P^*) + k-1 \qquad \text{(I. 4)}$$

(Equality is obtained for $t = k-1$.)

Efficiency

It is demonstrated empirically that the screening procedures using sampling rules 1 and 2 provide a reduction in ASN relative to single

sample procedures, sequential nonscreening procedures, and sequential procedures employing elimination.

Nature of the prior probability statement

Unlike the procedures employing elimination, PRCS is evaluated exactly for the screening procedures proposed in Parts II through V for any sampling rule (ignoring the problem of "overshoot", common to all the sequential procedures and arising from the discrete nature of the sample path) when the parametric configuration is an SC. We are unable to justify analytically the statement that PRCS $\geq P^*$ for all parametric configurations lying outside the indifference region. However it will be shown analytically that PRCS $\geq P^*$ for all slippage configurations and some generalized slippage configurations. In an empirical study the proportion of correct selections exceed P^* for other configurations lying outside the indifference region. We are unable to prove that the screening procedures in Part VI maintain specified levels of PRCS.

Nature of the posterior probability statement

The sequential screening procedures proposed in this thesis should reduce the possibility of awkward posterior probability statements since all populations remain candidates for selection until the selection of a best population is made. In addition, the flexibility of the sampling rule enables the experimenter to resume sampling from any previously screened population should he feel that this is warranted. Only one procedure involves a decision rule based on artificially constructed randomization. This procedure has been devised in such a way that (except possibly in very small samples) the effect of such randomization is

negligible. Randomization may also be entirely eliminated by replacing the artificially created random variables by their expectation.

Additional procedures

Also developed in this thesis are inverse sampling procedures for selecting the best multinomial category and a single sample procedure for selecting the best negative binomial population. The relationship between these procedures and the single sample procedures for selecting the best gamma population are investigated. Finally we consider a sequential nonscreening procedure for selecting the best negative binomial population, procedures for modified selection problems which include a test for homogeneity (i.e., that all populations are identical), and a sequential procedure for the interval estimation of the shape parameter of the gamma distribution.

PART II. SELECTING THE BEST MULTINOMIAL CATEGORY

CHAPTER A. INTRODUCTION AND REVIEW OF THE LITERATURE

Introduction

It is assumed that one may observe a sequence of random variables $\{X_{1j}, \ldots, X_{kj}\}$ such that

$$\begin{aligned} &\Pr[(X_{1j}, \ldots, X_{kj}) = (x_{1j}, \ldots, x_{kj})] \\ &= \prod_{i=1}^{k} p_i^{x_i} \qquad \text{if } x_i=0,1 \text{ and } \Sigma x_i=1 , \\ &= 0 \qquad \text{otherwise} , \end{aligned} \tag{II. 1}$$

where $0 \leq p_i \leq 1$ and $\Sigma p_i = 1$. Category i is denoted by π_i. The probability associated with π_i is p_i. The ordered probabilities are denoted by $p_{(i)}$, $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(k)}$, and the corresponding categories by $\pi_{(i)}$. The event $\{X_{ij} = 1\}$ is referred to as event i (the dependence on j will be clear from the context). An alternative method of saying that $\{X_{ij} = 1\}$ is to say that an outcome was observed from π_i on trial j.

The problems considered in this part are listed below using the format described in Section 2 of Part I.

Problem M.a. Selecting the most probable multinomial category:

1) $f_{X_{1j}, \ldots, X_{kj}}(x_1, \ldots, x_k | p_1, \ldots, p_k) = \prod_{i=1}^{k} p_i^{x_i}$

2) $g(p_1, \ldots, p_k) = \frac{p_{(k)}}{p_{(k-1)}} \geq \theta^* \qquad (\theta^* > 1)$

3) best category: $\pi_{(k)}$

Problem M.b. Selecting the least probable multinomial category:

1) $f_{X_{1j}, \ldots, X_{kj}}(x_1, \ldots, x_k | p_1, \ldots, p_k) = \prod_{i=1}^{k} p_i^{x_i}$

2) $g(p_1, \ldots, p_k) = \frac{p_{(1)}}{p_{(2)}} \leq \theta^* \qquad (\theta^* < 1)$

3) best category: $\pi_{(1)}$

Review of the literature

In this section we shall describe single sample and sequential non-screening procedures which have been proposed in the literature for selecting the best multinomial category. We denote the number of times event i occurs in the first n trials by N_{in} $(i=1, \ldots, k;\ n=1, 2, \ldots)$.

The following single sample procedure for selecting the most probable category was suggested by Bechhofer, Elmaghraby, and Morse (5). While requiring a relatively large sample number, this procedure is recommended on the basis of its administrative simplicity.

Procedure M.a.1. A single sample procedure for selecting the most probable category: Take n observations and select the population associated with $\max_i \{N_{in}\}$ as best. Ties may be broken using a random device.

Tables in Bechhofer et al. (5) provide PRCS for selected values of n, k, and θ^*. In general the determination of n so that specified levels of PRCS are maintained appears very difficult. However, for n large Bechhofer et al give an approximate method based on the normal approximation to the multinomial distribution, together with references for suitable tables. This approximation may also be accomplished using tables in Gupta (18).

It is also suggested by Bechhofer et al. (5) that the single sample procedure for selecting the most probable category may be generalized to the problem of ranking all k categories. However the generalization is not clearly defined. Consider the problem of identifying the category associated with $p_{(1)}$ and suppose $p_{(k)}$ is very close to unity. In this case there may be a large probability of observing no outcomes from $\pi_1, \ldots, \pi_{k-1}$ if a procedure analogous to Procedure M.a.1 is used. In fact, given any $\epsilon > 0$ one may choose a sufficiently large $p_{(k)}$ so that PRCS $< \frac{1}{k-1} + \epsilon$. It appears that any procedure for selecting π_1 must require some minimum number of outcomes from k-1 of the categories before a selection can be made. As a result the possibility of using a single sample procedure appears unlikely. It follows that the possibility of using a single sample procedure for complete ranking of all k categories is also unlikely.

The following two procedures are discussed by Bechhofer et al. (6) and by Bechhofer and Sobel (9). These procedures, while relatively easy to administer, may provide a substantial savings in ASN relative to the single sample procedure.

Procedure M. a. 2. A sequential nonscreening procedure for selecting the most probable category: Take one observation at a time sequentially until for some n and some i

$$\sum_{\substack{j=1 \\ \neq i}}^{k} (\theta^*)^{N_{jn} - N_{in}} \leq \frac{P^*}{1-P^*} \quad . \tag{II. 2}$$

At this point, terminate the experiment and select π_i as best.

In the actual formulation of this procedure given by Bechhofer and Sobel (9) the probability of ties was anticipated and it was suggested that ties could be broken using a random device. However if after n trials

$$T_{in} \overset{\text{def.}}{=} \sum_{\substack{j=1 \\ \neq i}}^{k} (\theta^*)^{N_{jn} - N_{in}} > \frac{P^*}{1-P^*} \tag{II. 3}$$

and the n+1st outcome is from population i' ($i' \neq i$), then

$$T_{i, n+1} > T_{in} > \frac{P^*}{1-P^*} \tag{II. 4}$$

so that ties are impossible.

Procedure M. b. 2. A sequential nonscreening procedure for selecting the least probable category: Take one observation at a time until for some n and some i

$$\sum_{\substack{j=1 \\ \neq i}}^{k} (\theta^*)^{N_{in} - N_{jn}} \leq \frac{P^*}{1-P^*} \quad . \tag{II. 5}$$

At this point terminate the experiment and select π_i as best.

CHAPTER B. INVERSE SAMPLING PROCEDURES

Introduction

Procedure M. a. 3 is discussed by Cacoullos and Sobel (12). It was also developed independently by the author.

Procedure M. a. 3. An inverse sampling procedure for selecting the most probable category: "Sample one observation at a time sequentially. When for some i and some n

$$N_{in} = R \tag{II. 6}$$

stop and select π_i as best."

It is demonstrated by Cacoullos and Sobel that for this procedure any SC is an LFC. It is also shown that if the parametric configuration is an SC, then

$$PRCS = D_{k-1}(R, R; 1/\theta^*) \tag{II. 7}$$

and the average sample number

$$E(N) = \frac{R}{p}\left[1 - \left(\frac{\theta^*-1}{\theta^*}\right)D_{k-1}\left(R, R; \frac{1}{\theta^*}\right)\right.$$

$$\left. - \frac{1}{2\theta^* p}\, b\left(R; 2R, \frac{1}{\theta^*+1}\right)D_{k-2}\left(2R, R; \frac{1}{\theta^*+1}\right)\right] . \tag{II. 8}$$

where

$$D_m(M, N;a) = \frac{\Gamma[M + (m-1)N]}{\Gamma(M)[\Gamma(N)]^{m-1}} \int_a^\infty \cdots \int_a^\infty \frac{\prod_{i=1}^{m} (y_i^{N-1} dy_i)}{(1 + \sum_{i=1}^{k-1} y_i)^{M+(m-1)N}} \tag{II.9}$$

We shall present a derivation in the next section which yields formulae for PRCS, E(N), and Var(N) which are considerably easier to evaluate than the formulae given by Cacoullos and Sobel. In addition the derivation is obtained more simply than the derivation given by Cacoullos and Sobel by appealing to an equivalence between Procedure M.a.3 and the single sample procedure for selecting the best gamma population considered in Part II.

We also consider the following procedure:

Procedure M.b.3. An inverse sampling procedure for selecting the least probable category: "Sample one observation at a time until for some n

$$N_{(2)n} = R \ . \tag{II.10}$$

At this point stop and select the category associated with $N_{(1)n}$ as best."

$N_{(i)n}$ is defined to be the ordered N_{in} $(N_{(1)n} \leq \cdots \leq N_{(k)n})$, where N_{in} is the number of occurrences from category i after sampling n times.

Derivation of the inverse sampling procedures

In this section we prove that the inverse sampling procedures maintain specified levels of PRCS. The associated formulae for PRCS, E(N), and Var(N) are also derived. The following notation is adopted:

$$b(x;\ n, \rho) = \binom{n}{x} \rho^x (1-\rho)^{n-x} \qquad \text{(II. 11)}$$

$p_{R, N, K}$ = $\Pr[\text{Max}(X_1, \ldots, X_K) \leq R]$ assuming a multinomial sample of size N with K equiprobable categories. (II. 12)

$$p = \frac{1}{\theta^* + k - 1} \qquad \text{(II. 13)}$$

Theorem II. 1. (Feller (15) p. 99): Let A_i be the event R outcomes are observed from π_1 before π_i. Assuming all subscripts unequal define

$$p_i = \Pr\{A_i\}, \qquad p_{ij} = \Pr\{A_i A_j\} \text{ etc.} \qquad \text{(II. 14)}$$

$$S_1 = \sum_{i=1}^{K} p_i \qquad S_2 = \sum_{i, j>i}^{K} p_{ij} \text{ etc.} \qquad \text{(II. 15)}$$

Then $$\Pr\left[\bigcup_{i=2}^{K} A_i\right] = \sum_{j=1}^{K-1} S_j (-1)^{j-1} \ . \qquad \text{(II. 16)}$$

Theorem II. 2: For Problems M. a and M. b any SC is an LFC with respect to Procedures M. a. 3 and M. b. 3 respectively. (Proof is deferred to Chapter C of Part IV.)

Theorem II. 3: The PRCS when the parametric configuration is an SC is given by

$$PRCS = (\theta^*p) \sum_{N=R}^{R+(k-1)(R-1)} b[R-1;\ N-1,\ (\theta^*p)] \times p_{R-1,\ N-R,\ k-1} \qquad \text{(II. 17)}$$

for Procedure M. a. 3, and by

$$PRCS = 1 - \sum_{j=1}^{k-1} \binom{k-1}{j} (-1)^{j-1} \sum_{N=R}^{R+j(R-1)} (\theta^*p) b[R-1;\ N-1,\ \theta^*p] p_{R-1,\ N-R,\ j} \qquad \text{(II. 18)}$$

for Procedure M. b. 3. Relation II. 17 is obtained using the fact that

$$\text{PRCS} = S_{k-1}$$

$$= \sum_{N=R}^{R+(k-1)(R-1)} \Pr[R \text{ outcomes from } \pi_1 \text{ before } \pi_2, \ldots, \pi_k \text{ and selection is made on Nth trial} \mid \text{SC with } \pi_1 \text{ best}] \quad \text{(II. 19)}$$

Similarly, Relation II. 18 is obtained using the fact that

$$\text{PRCS} = 1 - \sum_{j=1}^{k-1} S_j (-1)^{j-1} \quad \text{(II. 20)}$$

where

$$S_j = \binom{k-1}{j} \sum_{N=R}^{R+j(R-1)} \Pr[\text{when the Rth outcome from } \pi_1 \text{ occurs fewer than R outcomes have occurred from } \pi_i \text{ for each } i=2, \ldots, j+1 \text{ and the combined number of outcomes from } \pi_1, \ldots, \pi_{j+1} \text{ equals N} \mid \text{an SC with } \pi_1 \text{ best}] \quad \text{(II. 21)}$$

These probabilities may be evaluated easily on a computer even for fairly large values of R and k using recurrence formulae for $b(x; n, p)$ and $p_{R,N,K}$. A useful relation for $p_{R,N,K}$ is obtained as follows:

$$P_{R,N,K} = \Pr\left[\bigcup_{j=1}^{K} [(X_1, \ldots, X_K) \leq R \text{ and event } j \text{ occurs on trial } N]\right]$$

$$= K \Pr[(X_1, \ldots, X_K) \leq R \text{ and event 1 occurs on trial } N]$$

$$= \sum_{j=0}^{R-1} b(j;\ N-1, \tfrac{1}{K}) P_{R,\ N-j-1,\ K-1} \cdot \qquad \text{(II. 22)}$$

Theorems II. 4a and II. 4b give exact expressions for ASN and Var(N) for Procedure M. a. 3. It is seen that if PRCS is to be evaluated using Theorem II. 3, ASN and Var(N) may be evaluated at little additional computational expense.

Theorem II. 4a: Assuming that Procedure M. a. 3 is used for selecting the most probable category and that the parametric configuration is the LFC with π_1 best,

$$E[f(N)] = \sum_{N=R}^{R+(k-1)(R-1)} f(N)$$

$$\cdot \{\Pr[\text{termination on trial } N \text{ and } \pi_1 \text{ selected}] + (k-1) \sum_{j=0}^{R-1} \Pr[\text{termination on trial } N \text{ and } \pi_2 \text{ selected}]\}$$

$$= \sum_{N=R}^{R+(k-1)(R-1)} f(N)\{\theta^* p\, b[R-1;\ N-1,\ \theta^* p]p_{R-1,\,N-R,\,k-1}$$

$$+ p(k-1) \sum_{j=0}^{R-1} b(j;\ N-1,\ \theta^* p)\, b(R-1;\ N-1-j,\ p)$$

$$\cdot\ p_{R-1,\,N-R-j,\,k-2}\} \qquad \text{(II. 23)}$$

Theorem II. 4b: Assuming Procedure M. a. 3 is used for selecting the most probable category and that $p_1 = \ldots = p_k$,

$$E[f(N)] = \sum_{N=R}^{R+(k-1)(R-1)} f(N) b(R-1, N-1, 1/k)$$

$$\cdot\ p_{R-1,\,N-R,\,k-1} \qquad \text{(II. 24)}$$

Evaluating PRCS using existing tables and approximate formulae when the parametric configuration is an SC

In this section we discuss the evaluation of PRCS for the inverse sampling procedures described in Section 1 using existing tables and approximations. An SC is assumed in all cases. The justification for these methods is obtained from Theorem IV. 2 in Part IV.

Tables giving exact levels of PRCS for the inverse sampling procedures may be found in Bechhofer and Sobel (7), Gupta (17), and Gupta and Sobel (22). To use the tables in (7) to evaluate PRCS for the inverse sampling procedure for selecting the multinomial category with largest

probability (Procedure M.a.3), one enters the column corresponding to the indifference constant θ^* as specified in Problem M.a and the row giving "d.f." equal to twice the stopping bound (2R) specified in Procedure M.a.3. Tables are given for k=2, 3; d.f. = 1, 2, . . . , 20; and $\theta^* = 1.0, 1.2, \ldots, 3.0$. To obtain PRCS for Procedure M.b.3 one uses tables in (7) for selecting the gamma population having largest scale parameter, entering the column corresponding to the reciprocal of the indifference constant $(1/\theta^*)$ as specified in Problem M.b and the row giving "d.f." equal to twice the stopping bound (2R) specified in Procedure M.b.3. Such tables are given for k = 2, 3, 4; d.f. = 1, 2, . . . , 20; and $1/\theta^* = 1.0, 1.2, \ldots, 3.0$.

Gupta's tables (17), may be used to evaluate the PRCS for the inverse sampling procedure for selecting the multinomial category with smallest probability. The value of "α" listed in (17) indicates the error rate (α = 1-P*), the index at the top of each column corresponds to the number of populations from which a selection is to be made (k), the index to the left of each row corresponds to twice the stopping bound as specified in Procedure M.b.3. Entries in the body of the table give values of the indifference constant θ^* as specified by Problem M.b. Tables are provided for $P^* = .75, .90, .95, .99$; $k = 2, \ldots, 10$; and $R = 1, \ldots, 25$.

As noted by Cacoullos and Sobel, tables given by Gupta and Sobel (22) may be used to evaluate PRCS for Procedure M.a.3. The construction of these tables is analouous to those given by Gupta (17).

We next consider approximate methods of evaluating the appropriate stopping bound R specified in Procedure M.a.3 to yield specified levels of PRCS. (The stopping bound for Procedure M.b.3 is determined

approximately to be the stopping bound required for Procedure M. a. 3 with an indifference constant equal to the reciprocal of the indifference constant specified in Problem M. b.

The first method we consider is based on the discussion by Bechhofer and Sobel (7) who rely on the fact that the logarithm of a gamma random variable has approximately a normal distribution.

The procedure is as follows: Enter the tables in Bechhofer (7) under the column corresponding to the value of "k" equal to the number of populations from which a selection is to be made for which "t" is equal to unity, and the row corresponding to the level of PRCS as specified by Problem M. a. Thus obtain "$\sqrt{N\lambda}$" from the body of the table. Recalling that θ^* is the indifference constant specified in Problem M. a, R is determined as the smallest integer which is no less than

$$(\sqrt{N\lambda}/\ell n\,\theta^*)^2 + \frac{1}{2} \ . \qquad \text{(II. 25)}$$

Comparisons given by Bechhofer and Sobel (7) between the approximate and exact sample numbers indicate that the approximate method is fairly accurate for values of R as small as 10. When $k > 10$, the approximation obtained from Dudewicz (14) by taking R as the smallest integer which is no less than

$$-\,4(\ell n\,\theta^*)^{-2}\ell n(1-P^*)$$

should be useful and especially accurate for P^* close to unity.

CHAPTER C. SEQUENTIAL SCREENING PROCEDURES

Introduction

The procedures that we have discussed thus far assumed that all observations were k-dimensional vectors. In this chapter, with a view towards future applications (see Parts III, IV, and V), we consider the possibility of screening. The screening procedures proposed in this chapter assume that all screening is such that for any two unscreened categories i and j

$$\Pr[\text{event } i]/\Pr[\text{event } j]$$

is constant for all possible screening configurations wherein categories i and j are not screened.

This screening criterion limits the applicability of the procedures to be proposed in this chapter. It is difficult to conceive of phenomena which one would normally associate with the multinomial distribution in which screening could be conducted in such a way that the screening criterion would be satisfied. However the screening criterion is satisfied in the applications considered in Parts III, IV, and V. As an example of a situation in which the screening criterion is not satisfied, consider the problem of determining the most popular of three political candidates. Suppose candidate one is a conservative and candidates two and three are liberals. Let E_i be the event candidate i is preferred. In this case,

$\Pr[E_1]/\Pr[E_2]$ may depend on whether or not E_3 is screened (included in the list of candidates).

The screening procedures proposed in this chapter are members of a large class possessing the property that PRCS may be evaluated exactly (ignoring "overshoot") for parametric configurations which lie on the boundary of the indifference region.

Statement of the procedures

We adopt the following notation:

for $j = 1, 2, \ldots$

$S_j = \{i :$ category i is unscreened during trial $j\}$

$C_j =$ cardinality of S_j

$L_{i0} = 1$

$$L_{ij} = L_{i,j-1} \times (C_j - 1 + \theta^*)^{-1} (\theta^*)^{X_{ij}}, \quad \text{if } i \in S_j$$

$$= L_{i,j-1} \times C_j^{-1}, \quad \text{if } i \notin S_j . \qquad \text{(II. 26)}$$

Procedure M.a.4. A sequential screening procedure for selecting the multinomial category with largest probability: "Sample one observation at a time sequentially in accordance with sampling rule 1 until after some trial n

$$L_{(k)n} / \sum_{i=1}^{n} L_{in} \geq P^* . \qquad \text{(II. 27)}$$

At this point stop and select the population associated with $L_{(k)n}$ as best. In the specifications of Procedure M. b. 4 it is assumed that either sampling rule 1 is used exclusively or sampling rule 2 is used exclusively.

Procedure M. b. 4. A sequential screening procedure for selecting the multinomial category with smallest probability: Sample in accordance with sampling rule 1 or sampling rule 2 until after some trial n

$$L_{(k)n} \Big/ \sum_{j=1}^{k} L_{in} \geq P^* \quad . \tag{II. 28}$$

At this point stop and select the population associated with $L_{(k)n}$ as best.

Derivation

The following theorem generalizes the discussion starting on page 260 of Bechhofer et al. (6) and that of O'Brien (30), where it was assumed that only k dimensional vectors of observations (i. e.: one observation from each population) are sampled. We assume that observations are taken according to some set of rules $\mathcal{K}$. For this set of rules the theorem defines a decision procedure for selecting the best population. The theorem states that for any parametric configuration which is a slippage configuration

$$PRCS \geq (1-\beta)P^*$$

where β is the probability that the procedure does not terminate. Conditions on $\mathcal{K}$ which permit analytic proof that $\beta = 0$ are given in

Theorem II. 6. Of course one desires the procedure to maintain specified levels of PRCS for any configuration outside the indifference region. This would follow if an SC were an LFC. The problem of demonstrating than an SC is also least favorable is as yet unsolved but is considered in the next section.

The following notation is used:

$\mathcal{K}$ the set of rules under which the experiment is conducted.

X_n a vector valued random variable composed of the first n (k dimensional) vectors of observations. Some components in each k dimensional vector may be empty. A non-empty component is associated with a sampled population. Realizable vectors are defined by $\mathcal{K}$.

χ_n the set of all X_n which may arise under $\mathcal{K}$.

β the probability that the procedure does not terminate.

D_{in} that part of χ_n for which population i is selected under $\mathcal{K}$.

$f_{in}(\cdot)$ the p.d.f. of the arguments when an SC in which population i is best is assumed.

Symmetry condition: Let Z^* and Z^{**} represent sample points in acceptance regions D_{in} and D_{jn} $(i \neq j)$ respectively, where $Z^* = (Z_1^*, \ldots, Z_k^*)$, $Z^{**} = (Z_1^{**}, \ldots, Z_k^{**})$. Z_i^* is an n-dimensional vector whose components (some of which may be empty) correspond to the observations obtained from π_i (if any) after sampling n times. (i.e.: Z_i^* is the vector created from the ith components in a sample path from χ_n.) Z_i^{**} is defined similarly. The regions $D_{1n}, \ldots, D_{kn}$ are defined to be symmetric if for any two regions D_{in} and D_{jn} and any sample point Z^* belonging to D_{in}, there corresponds a point Z^{**} belonging to D_{jn} such that

$$
\begin{aligned}
Z_s^* &= Z_s^{**} && s = 1, \ldots, k;\ \neq i \text{ or } j \\
&= Z_j^{**} && s = i \\
&= Z_i^{**} && s = j
\end{aligned}
\qquad \text{(II. 29)}
$$

and such that

$$
\begin{aligned}
f_{in}(Z^*) &= f_{jn}(Z^{**}) \\
f_{jn}(Z^*) &= f_{in}(Z^{**}) \\
f_{sn}(Z^*) &= f_{sn}(Z^{**}) \qquad s \neq i \text{ or } j.
\end{aligned}
\qquad \text{(II. 30)}
$$

Theorem II. 5:

Conditions: $\mathcal{K}$ is invariant under permutation of the indexing of the populations. The f_{in} are such that the regions D_{in} satisfy the symmetry condition. Observations are taken in accordance with the rules set forth in $\mathcal{K}$ until for some n and s,

$$
\frac{f_{sn}(X_n)}{\sum_{\substack{\ell=1 \\ \neq s}}^{k} f_{\ell n}(X_n)} \geq \frac{P^*}{1-P^*} \overset{\text{def.}}{=} A \qquad \text{(II. 31)}
$$

At this point the experiment is terminated and population s is selected as best. It is also assumed that the parametric configuration lies on the boundary of the indifference region.

Conclusion:

$$PRCS \geq P^*(1-\beta) \tag{II.32}$$

Proof: Relying on the symmetry of the regions D_{in} and the fact that

$$D_{in} \cap D_{i'n'} = \emptyset \text{ for } (i,n) \neq (i',n')$$

$$PRCS = \sum_{n=1}^{\infty} \int_{D_{ln}} f_{ln}(X_n)\, dX_n \tag{II.33}$$

$$\geq \sum_{n=1}^{\infty} A \int_{D_{ln}} \sum_{i=2}^{k} f_{in}(X_n)\, dX_n \tag{II.34}$$

$$= A \sum_{n=1}^{\infty} \sum_{i=2}^{k} \int_{D_{in}} f_{ln}(X_n)\, dX_n \tag{II.35}$$

$$= A(1-\beta-PRCS) \tag{II.36}$$

Therefore

$$\frac{1-P^*}{P^*} \geq \frac{1-\beta-PRCS}{PRCS} \qquad \text{(II.37)}$$

so that

$$P^* \leq PRCS/(1-\beta) \qquad \text{(II.38)}$$

which yields the desired result. Theorem II.6 states that if no screening is conducted beyond the Nth stage, where N is a finite integer, then $\beta = 0$. In practice (due to budgetary or other considerations) there is always a number N^* which serves as an upper limit for the number of observations which may be taken, so that one may simply choose $N \geq N^*$. Since in reality all sequential procedures are truncated procedures, the real concern of the experimenter is ASN and variance of the sample number.

Theorem II.6:

Suppose the screening rule in the procedure of Theorem II.5 is such that it defines a finite and integer valued random variable N, whereby after N stages no population is ever screened. Let the vectors $(X_{1j}, \ldots, X_{kj})$, $j = N+1, \ldots$ be independent and identically distributed. Then the procedure of Theorem II.5 terminates with probability one.

Proof: It follows from Chapter 3 Section 3 of Bechhofer et al. (6) regarding termination with probability one for sequential nonscreening procedures that for any integer N and $\epsilon_1, \ldots, \epsilon_k$ greater than zero.

$$\Pr[\text{for some } M > N \text{ and some } i,$$

$$\prod_{j=N+1}^{M} \frac{f_{tj}(X_{1j}, \ldots, X_{kj})}{f_{ij}(X_{1j}, \ldots, X_{kj})} < \epsilon_t \text{ for each } t=1, \ldots, k$$

$$(t \neq i)] = 1 \qquad \text{(II. 39)}$$

Let $$\epsilon_t = \frac{f_{iN}(X_N)}{f_{tN}(X_N)} \frac{(1-P^*)}{kP^*} \qquad \text{(II. 40)}$$

Since the probability that the procedure terminates is no less than the probability of termination given that no selection has been made during the first N trials,

$$\Pr[\text{termination}] \geq \Pr[\text{for some } M > N \text{ and some } i,$$

$$\sum_{\substack{t=1 \\ \neq i}}^{k} \frac{f_{tN}(X_N)}{f_{iN}(X_N)} \prod_{j=N+1}^{M} \frac{f_{tj}(X_{1j}, \ldots, X_{kj})}{f_{ij}(X_{1j}, \ldots, X_{kj})} \leq \frac{1-P^*}{P^*}] \qquad \text{(II. 41)}$$

$$\geq \Pr[\text{for some } M > N \text{ and some } i,$$

$$\frac{f_{tN}(X_N)}{f_{iN}(X_N)} \prod_{j=N+1}^{M} \frac{f_{tj}(X_{1j}, \ldots, X_{kj})}{f_{ij}(X_{1j}, \ldots, X_{kj})} \leq \frac{1-P^*}{kP^*}$$

$$\text{for each } t = 1, \ldots, k(t \neq i)] \qquad \text{(II. 42)}$$

$$\geq \Pr\Big[\text{for some } M > N \text{ and some } i$$

$$\prod_{j=N+1}^{M} \frac{f_{tj}(X_{1j}, \ldots, X_{kj})}{f_{ij}(X_{1j}, \ldots, X_{kj})} < \epsilon_t$$

$$\text{for each } t = 1, \ldots, k(t \neq i)\Big] \qquad \text{(II. 43)}$$

$$= 1 \ .$$

The problem of least favorable configuration

In this section we consider the determination of sampling rules for the selection of the most probable multinomial category. We desire rules such that any slippage configuration is an LFC if these rules are used in conjunction with the test procedure of Theorem II. 5 . We note first that this is trivially true if the sampling rule prohibits any screening. If there is no screening, the procedure of Theorem II. 5 is identical to the nonscreening procedure proposed by Bechhofer et al. (6). Careful inspection of the test statistic suggests that only populations which are observed to be inferior (i. e.: populations corresponding to relatively small likelihood functions) should be screened. Rules are desired which reduce ASN, ensure termination with probability one, and maintain specified levels of PRCS. These considerations led to the formation of the following list of desirable properties for screening rules:

(1) No category should ever be permanently screened.

(2) No more than k-2 categories should ever be screened during the same trial unless a selection is made.

(3) The desirability of screening the category associated with $L_{(t)n}$ increases as $L_{(j)n}/L_{(t)n}$ increases $(j=t+1, \ldots, k)$. Similarly if $t_1 > t_2$ the population associated with $L_{(t_1),n}$ should not be screened unless the population associated with $L_{(t_2),n}$ is also screened.

(4) If $L_{1n} = \ldots = L_{k-1,n} < L_{kn}$, $n=1, 2, \ldots$, then no category should be screened unless a selection is made, in which case categories $\pi_1, \ldots, \pi_{k-1}$ should all be screened.

It may be verified that sampling rules one and two possess properties 1-4 with a single exception. In the situation $L_{1n} = \ldots = L_{k-1,n} < L_{kn}$, $n=1, 2, \ldots$, screening is possible prior to a selection using sampling rule 2.

In the sections dealing with empirical results we have considered the empirical estimation of PRCS for parameter points lying outside, but near the boundary, of the indifference region. In each case the evidence is in agreement with the hypothesis that slippage configurations are least favorable. We next consider parameter points lying on the boundary of the parameter space. Clearly for Procedure M.a.4 PRCS equals one if $p_{(k)}$ equals one. Similarly, PRCS for Procedure M.b.4 equals one if $p_{(2)}$ equals $1/(k-1)$.

Theorem II.7a: If for $t=1, 2, \ldots,$ or $k-2$

$$p_1 = \ldots = p_t = 0$$

$$\theta^* p_{t+1} = \ldots = \theta^* p_{k-1} = p_k \,, \qquad \text{(II.44)}$$

then, using Procedure M.a.4 for selecting the most probable category,

$$PRCS \geq P^*(1-\beta) \qquad \text{(II.45)}$$

Proof: Since events $1, \ldots, t$ never occur, none of the categories $\pi_1, \ldots, \pi_t$ may be selected. If any of the categories $\pi_1, \ldots, \pi_t$ are not screened during any trial n, then none of the remaining categories will be screened during trial n. Let $Y_{in} = X_{t+i,n}$ $(i=1, \ldots, k-t;\ n=1, 2, \ldots)$ so that if S_n contains $\{t+1, \ldots, k\}$, then $(Y_{1n}, \ldots, Y_{k-t,n})$ has a multinomial distribution with parameter $(p'_1, \ldots, p'_{k-t})$ such that

$$p'_{k-t} = \theta^* p'_i \qquad i=1, \ldots, k-t-1 \qquad \text{(II.46)}$$

Let L'_{in} be the likelihood function of the $\{Y_{ij}\}$ computed using the formulae of Procedure M.a.4 with k replaced by $k-t$. Then it may be verified that

$$\frac{L_{in}'}{L_{ij}'} = \frac{L_{t+i,n}}{L_{t+j,n}} \qquad n=1, 2, \ldots \qquad \text{(II.47)}$$

Thus Procedure M.a.4 as it pertains to $\pi_{t+1}, \ldots, \pi_k$ may be rewritten "Screen the category associated with $L'_{(s)n}$ during trial $n+1$ if

$$\sum_{\ell=s+1}^{k-1} L'_{(\ell)n}/L'_{(s)n} \geq \frac{k-1}{1-P^*} - s \,. \qquad \text{(II.48)}$$

When for some trial n

$$\sum_{\ell=1}^{k-t-1} \frac{L'_{(\ell)n}}{L'_{(k-t)n}} \leq \frac{1-P^*}{P^*} - \sum_{\ell=1}^{t} \frac{L_{\ell n}}{L'_{(k-t)n}} \qquad \text{(II. 49)}$$

stop and select the category associated with $L'_{(k-t)n}$ as best." That this procedure has PRCS $\geq P^*(1-\beta)$ follows from Theorem III. 1.

Theorem II. 7b: If for $t=1, 2, \ldots,$ or $k-2$

$$p_1 = \theta^* p_2 = \cdots = \theta^* p_{k-t}$$

$$p_{k-t+1} = \cdots = p_k = p \qquad \text{(II. 50)}$$

then using Procedure M. b for selecting the least probable category,

$$\lim_{p \to 1/t} \text{PRCS} \geq P^*(1-\beta) \quad . \qquad \text{(II. 51)}$$

Proof: Let E represent the event (not to be confused with the multinomial events described earlier) that prior to making a selection an outcome from $\pi_1, \ldots, \pi_{k-t-1}$, or π_{k-t} is observed on a trial in which $\pi_{k-t+1}, \ldots, \pi_{k-1}$, or π_k has not been screened. Conditioning on E not occurring, one may follow the method of proof used in proving Theorem II. 7a to obtain

$$\text{PRCS} \geq P^*(1-\beta) - \Pr(E) \quad . \qquad \text{(II. 52)}$$

The conclusion of Theorem II. 7b is then obtained from the fact that

$$\lim_{p \to 1/t} \Pr(E) = 0 \ . \qquad \text{(II. 53)}$$

One would expect that a procedure which maintains specified levels of PRCS for parametric configurations lying on the boundary of the indifference region would maintain specified levels of PRCS for GSC's of the kind considered in Theorems II. 7a and II. 7b. However, such is not always the case. In Chapter B of Part IV we will examine a procedure for selecting the best gamma population for which GSC's analogous to those in Theorem II. 7a and II. 7b are less favorable than SC's.

Inspection of the test statistic

In this section we examine the test statistic of the procedure in Theorem II. 5. The conclusion drawn is that Property 3 of Section 4 is a desirable property for a sampling rule for maintaining specified levels of PRCS.

Define H_{i0} to be the hypothesis that the parametric configuration is an SC with π_i best. That is

$$H_{i0}: \ p_i = \theta^* p_j \qquad j=1, \ldots, k \ (j \neq i) \ . \qquad \text{(II. 54)}$$

At stage n suppose that $\pi_{n_1}, \ldots, \pi_{n_s}$ are observed. Define H_{in} to be the restriction of H_{i0} to $\pi_{n_1}, \ldots, \pi_{n_s}$ so that under H_{in}

$$p_{n_1} = \ldots = p_{n_s} = \frac{1}{s} \qquad \text{if } i \notin S_n \qquad \text{(II.55)}$$

$$p_i = \theta^* p_{n_j} \quad j=1,\ldots,s\ (j\neq i) \qquad \text{if } i \in S_n \ . \qquad \text{(II.56)}$$

We denote the parametric configuration associated with H_{in} by p_{in}^* and the parametric configuration corresponding to the true state of nature by p_n . Since the test statistic is a combination of likelihood functions in which the ith likelihood function is computed using the parametric configurations $(p_{i1}^*, \ldots, p_{in}^*)$ associated with $(H_{i1}, \ldots, H_{in})$, and since π_k is selected at trial n only if $L_{kn} > L_{in}$ $(i=1, 2, \ldots, k-1)$, it is desirable that H_{kn} be more consonant with the true parametric configuration than any of the other k-1 hypotheses $H_{1n}, \ldots, H_{k-1,n}$. That is the "distance" between p_{in}^* and p_n should be minimized by i=k . If "distance" is taken to be the sum of squared deviations (or almost any other reasonable measure of distance) distance will be minimized if $k \in S_n$ but not if $k \notin S_n$ and $p_{n_{(s)}} \geq \theta^* p_{n_{(s-1)}}$. Therefore a screening rule should be such that for any n , $\Pr[i \notin S_n | \pi_i \text{ best}]$ is a decreasing function of $\frac{p_{(k-1)}}{p_{(j)}}$ for each $j=1, \ldots, k-2$ and less than some constant when

$$p_{(k)} = \theta^* p_{(k-1)} = \ldots = \theta^* p_{(1)} \ . \qquad \text{(II.57)}$$

It is conjectured that there exist functions P_j and a constant c such that if for each $n=1, 2, \ldots$, $\Pr[i \notin S_n | \pi_i \text{ best}]$ decreases more rapidly as a

function in $\frac{p_{(k-1)}}{p_{(j)}}$ than does $P_j[\frac{p_{(k-1)}}{p_{(j)}}]$ and $Pr[i \notin S_n | SC \text{ with } \pi_i \text{ best}] < C$, then the SC is the LFC.

As noted by Bechhofer et al. (see p. 191 (6)) a screening rule which tends to observe less often those populations which are observed to be inferior may be desirable for reducing ASN as well as maintaining PRCS. This may be explained by considering the statements which are made about an individual population, π_i say, when a terminal decision is reached. If π_i is selected, the statement involving k-1 contrasts, "π_i is superior to π_j, $j=1,\ldots,k$; $j \neq i$" is made. On the other hand, if π_j $(j \neq i)$ is selected, the statement involving only one contrast, "π_i is inferior to π_j", is made. One would expect to require more information on π_i in making k-1 contrasts than would be the case for a single contrast.

We next consider an example of a sampling rule which does not possess Property 3, causing the procedure of Theorem II.5 to fail to maintain specified levels of PRCS when used to select the most probable category. Let m represent the category associated with the (multinomial) event which occurs on the first trial of the experiment. The sampling rule is to screen category m during trials $2,\ldots,N$, but to do no other screening in the course of the experiment. For a generalized slippage configuration having the form

$$p_k = M_1 p_{k-1}$$

$$p_{k-1} = M_2 p_{k-2} = \cdots = M_2 p_1$$

PRCS may be made arbitrarily close to zero by choosing N, M_1, and M_2 suitably large. (i.e.: by choosing M_1 and M_2 sufficiently large one ensures that N of the first N+1 outcomes are from π_{k-1}, despite the fact that π_k is best.)

CHAPTER D. EMPIRICAL COMPARISONS

Table II. 1 gives numerical comparisons between the screening procedure and the nonscreening procedure proposed by Bechhofer et al. (6) for the problem of selecting the least probably category. Table II. 2 gives similar comparisons for the problem of selecting the most probable category. Included in Table II. 1 is the single sample procedure proposed by Bechhofer et al. (5). The screening procedures provided a smaller ASN than its competitors in all cases. It appears from Table I. 1 that, as one might expect, the relative efficiency (R. E.) of the single sample procedure to the screening procedure (defined as the ASN of the screening procedure divided by the ASN of the single sample procedure) decreases as the differences among the p_i increase. The converse is true of the R. E. of the nonscreening procedure to the screening procedure. From Table II. 2 it appears that, for the problem of selecting the least probable category, sampling Rule 1 is the more efficient sampling rule for selecting from among 3 or 4 populations, but sampling Rule 2 is more efficient for more than 4 populations.

Table II. 1. Empirical comparisons of procedures for selecting the category with largest probability (100 samples, $P^* = .75$, $k=5$)

$\theta^* = 4/3$

$p_1 = \cdots = p_5$

	single sample	non screen	screen
Proportion of correct selections			
Average sample number	189.6	179.1	120.8
Standard deviation of the sample number		123.1	75.3

$\theta^* = 7/6$ $p_{(5)} = \theta^* p_{(i)}$ $i = 1, \ldots, 4$			$\theta^* = 7/6$ $p_{(5)} = (\theta^*)^{5-i} p_{(i)}$ $i = 1, \ldots, 5$		
single sample	non screen	screen	single sample	non screen	screen
	.64	.77		.85	.94
687.8	437.3	350.2	687.8	245.8	218.6
	164.5	193.8		137.3	128.2

Table II. 2. Empirical comparisons of procedures for selecting the category with smallest probability (100 samples, $P^* = .75$)

	k = 3		
	non screen	screen rule 2	screen rule 1
	$p_1 = \ldots = p_3$		$\theta^* = .75$
Average sample number	60.1	65.5	60.5
Standard deviation of the sample number	33.6	46.5	39.3
	$p_{(1)} = p_{(i)}\theta^*$ (i=2, 3)		$\theta^* = 6/7$
Proportion of correct selections	.79	.74	.80
Average sample number	179.4	163.1	164.4
Standard deviation of the sample number	102.1	109.7	97.3
	$p_{(1)} = p_{(i)}\theta^{*i-1}$ (i=2, 3)		$\theta^* = 6/7$
Proportion of correct selections	.83	.89	.85
Average sample number	146.3	141.2	128.2
Standard deviation of the sample number	104.3	106.3	73.5

k = 4			k = 5		
non screen	screen rule 2	screen rule 1	non screen	screen rule 2	screen rule 1
$p_1 = \ldots = p_k$			$\theta^* = .75$		
146.5	124.6	116.1	204.8	166.2	175.8
97.0	89.9	64.2	101.2	122.6	110.9
$p_{(1)} = p_{(i)}\theta^*$ (i=2, . . . , k)			$\theta^* = 6/7$		
.76	.74	.72	.70	.74	.80
325.8	304.1	279.2	450.0	372.8	386.0
163.2	189.4	148.1	158.0	183.7	187.6
$p_{(1)} = p_{(i)}\theta^{*i-1}$ (i=2, . . . , k)			$\theta^* = 6/7$		
.87	.87	.93	.78	.85	.87
227.7	177.7	186.8	343.6	237.2	241.9
139.7	95.5	106.1	145.5	111.8	119.3

PART III. SELECTING THE BEST INCOMPLETELY SPECIFIED POPULATION

In this part the following problems are considered:

Problem I.a. Selecting the population with largest mean:

1. $F_i(x) = 1 - [H(x)]^{\theta_i}$

2. $g(\theta_1, \ldots, \theta_k) = \frac{\theta_{(1)}}{\theta_{(2)}} \leq \theta^* \qquad (\theta^* < 1)$

3. best population: $\pi_{(1)}$

Problem I.b. Selecting the population with smallest mean:

1. $F_i(x) = 1 - [H(x)]^{\theta_i}$

2. $g(\theta_1, \ldots, \theta_k) = \frac{\theta_{(k-1)}}{\theta_{(k)}} \leq \theta^* \qquad (\theta^* < 1)$

3. best population: $\pi_{(k)}$

In both problems it is assumed that $H(x)$ is a continuous, non-negative function which may be unspecified. Since $F_i(x)$ is a distribution function, $H(x)$ must also be monotone decreasing with $\lim_{x \to \infty} H(x) = 0$ and $H(0) = 1$. The Weibull distribution and the Pareto distribution have this form with $H(x) = e^{-x^c}$, $x \geq 0$ and $H(x) = \frac{1}{x}$, $x \geq 1$ respectively.

The procedures of this part are obtained by creating multinomial type events and then appealing to the procedures of Part II. If $(X_{1j}, \ldots, X_{kj})$ is the vector of observations sampled during stage j, we define the vector $(Y_{1j}, \ldots, Y_{kj})$ by replacing the smallest component of $(X_{1j}, \ldots,$

X_{kj}) with unity and all other components with zero. Letting

$$p_i = \Pr[X_{ij} = \min(X_{1j}, \ldots, X_{kj})] , \qquad \text{(III. 1)}$$

it is seen that $(Y_{1j}, \ldots, Y_{kj})$ has a multinomial distribution with probabilities $p_1, \ldots, p_k$. Furthermore it is easily verified that for any choice of $s \leq k$ and $n_1, \ldots, n_s$

$$\Pr[X_{n_i j} \leq \min_{\ell}(X_{n_\ell j})] = \frac{\theta_{n_i}}{\sum_{\ell=1}^{s} \theta_{n_\ell}} ,$$

so that the screening criterion of Part II is satisfied.

Bechhofer (10) discusses the possibility of using procedures designed for selecting the most (least) probable multinomial category for the nonparametric problem of selecting the population having largest (smallest) probability of yielding the largest observation. If one has knowledge that the distributions involved have the form given in the specifications of Problems I. a and I. b, then the screening criterion discussed in Part II is satisfied and one may use the sequential screening procedures of Part II for selecting the best multinomial category. Table III. 1 gives numerical comparisons for the nonscreening of Bechhofer et al. (6) and the screening procedures of Part II for Problem I. a. (The samples used are the same as those used in obtaining Table I. 1.) Similar comparisons appear in Table III. 2 for the single sample procedure of Bechhofer, Elmaghraby, and Morse (5), the nonscreening procedure of Bechhofer,

Kiefer, and Sobel (6), and the sequential screening procedure for Problem I. b.

These tables indicate that the sequential screening procedures offer a substantial reduction in ASN relative to the sequential nonscreening and single sample procedures. The sequential screening procedures also yielded a substantial reduction in the variability of the sample number relative to the sequential nonscreening procedure. From the tables it appears that only screening rule 2 should be used for selecting the population with largest mean.

The large reduction in ASN achieved by the screening procedures is attributable to the fact that in using the screening procedure one is making use of the additional information contained in the specifications of Problems I. a and I. b relative to their nonparametric counterparts. Comparison of Tables III. 1 and III. 2 with Tables I. 1 and I. 2 respectively indicate that a further substantial reduction in the average sample number may be achieved if the function $H(x)$ is known.

Table III. 1. Empirical comparisons of procedures for selecting the population with largest mean (100 samples, $P^* = .75$)

	k = 3		
	non screen	screen 2	screen 1
	$\theta_1 = \ldots = \theta_k$		$\theta^* = .75$
Average sample number	180.2	148.9	153.0
Standard deviation of the sample number	100.8	103.7	98.5
	$\theta_{(1)} = \theta_{(i)}\theta^*$ $i=2,\ldots,k$		$\theta^* = 6/7$
Proportion of correct selections	.79	.74	.80
Average sample number	538.2	371.4	422.4
Standard deviation of the sample number	306.3	249.6	246.5
	$\theta_{(1)} = \theta_{(i)}\theta^{*i-1}$ $i=2,\ldots,k$		$\theta^* = 6/7$
Proportion of correct selections	.78	.89	.85
Average sample number	438.8	320.1	329.1
Standard deviation of the sample number	313.0	233.3	184.0

k = 4			k = 5		
non screen	screen 2	screen 1	non screen	screen 2	screen 1
$\theta_1 = \ldots = \theta_k$			$\theta^* = .75$		
585.8	331.3	365.9	1024.1	523.0	663.9
388.1	228.1	198.5	505.9	367.6	408.0
$\theta_{(1)} = \theta_{(i)}\theta^*$ i=2,...,k			$\theta^* = 6/7$		
.76	.74	.72	.70	.74	.80
1303.4	825.4	903.8	2250.2	1185.7	1511.4
652.9	495.5	459.8	758.1	571.8	739.0
$\theta_{(1)} = \theta_{(i)}\theta^{*i-1}$ i=1,...,k			$\theta^* = 6/7$		
.87	.87	.93	.83	.85	.87
910.9	480.7	575.9	1718.1	721.7	869.9
558.6	243.7	292.0	727.7	291.8	349.3

Table III. 2. Empirical comparisons of procedures for selecting the population with smallest mean (100 samples, $P^* = .75$, $k=5$)

$\theta^* = 4/3$

$\theta_1 = \ldots = \theta_5$

	single sample	non screen	screen 1
Proportion of correct selections			
Average sample number	948	895	476
Standard deviation of the sample number		616	284

$\theta^* = 7/6$ $\theta_{(k)} = \theta_{(i)}\theta^*$ $i = 1, \ldots, 4$			$\theta^* = 7/6$ $\theta_{(k)} = \theta_{(i)}\theta^{*i-1}$ $i = 1, \ldots, 5$		
single sample	non screen	screen 1	single sample	non screen	screen 1
	.64	.77		.85	.94
3439	2186	1410	3439	1229	825
	823	761		686	390

PART IV. SELECTING THE BEST GAMMA POPULATION

CHAPTER A. INTRODUCTION

The specifications for the problems of selecting the best gamma population are listed below.

Problem G.a:

1. $f_i(x) = \frac{x^{p-1} e^{-x/\theta_i}}{\Gamma(p)\,\theta_i^{p}}$, $p=1, 2, \ldots$ (known)

2. $g(\theta_1, \ldots, \theta_k) = \theta_{(k-1)}/\theta_{(k)} \leq \theta^*$ $\qquad (\theta^* < 1)$

3. Best population: $\pi_{(k)}$

Problem G.b:

1. $f_i(x) = \frac{x^{p-1} e^{-x/\theta_i}}{\Gamma(p)\,\theta_i^{p}}$, $p=1, 2, \ldots,$ (known)

2. $g(\theta_1, \ldots, \theta_k) = \theta_{(1)}/\theta_{(2)} \leq \theta^*$ $\qquad (\theta^* < 1)$

3. Best population: $\pi_{(1)}$.

Since the usual estimator of variance of a normally distributed random variable has a gamma distribution, there is a close relationship between selecting the best gamma population and selecting the normal population with smallest (largest) variance. The specifications appropriate for the latter problems are listed below.

Problem N. a:

1. $f(x;\ \mu_i,\ \sigma_i^2) = (2\pi\sigma_i^2)^{-\frac{1}{2}} \exp - \frac{(x-\mu_i)^2}{2\sigma_i^2}$

2. $g(\sigma_1^2, \ldots, \sigma_k^2) = \sigma^2_{(k-1)}/\sigma^2_{(k)} \leq \theta^* \qquad (\theta^* < 1)$

3. Best population: $\pi_{(k)}$

Problem N. b:

1. $f(x;\ \mu_i,\ \sigma_i^2) = (2\pi\sigma_i^2)^{-\frac{1}{2}} \exp - \frac{(x-\mu_i)^2}{2\sigma_i^2}$

2. $g(\sigma_1^2, \ldots, \sigma_k^2) = \sigma^2_{(1)}/\sigma^2_{(2)} \leq \theta^* \qquad (\theta^* < 1)$

3. Best population: $\pi_{(1)}$

Any procedure for selecting the best gamma population can be used for selecting the best normal population. (See p. 118 of Bechhofer et al. (6).) If one observes the sequence X_j $(j=1, 2, \ldots)$ where X_j are i.i.d. having a normal distribution with mean μ and variance σ^2 (i.e. $X_j \sim N(\mu, \sigma^2)$) with μ known, then the random variables $w_j = (X_j - \mu)^2$ are i.i.d. having a gamma distribution with shape parameter equal to $\frac{1}{2}$ and scale parameter equal to $2\sigma^2$ (i.e. $w_j \sim \Gamma(\frac{1}{2};\ 2\sigma^2)$). Thus procedures for selecting the best gamma population may be used for selecting the normal population with smallest (largest) variance when μ is known.

If μ is unknown one may appeal to the Helmert transformation as discussed in Bechhofer et al. (6). Specifically, define

$$a_{n,j} = 1/\sqrt{n(n+1)} \qquad j=1, \ldots, n$$

$$a_{n,n+1} = -n/\sqrt{n(n+1)}$$

$$a_{n,j} = 0 \qquad j=n+2, n+3, \ldots \qquad \text{(IV.1)}$$

For $n=1, 2, \ldots$, let

$$V_n = \sum_{j=1}^{n+1} a_{n_j} X_j$$

$$= \sqrt{n/(n+1)}\ (\overline{X}_{(n)} - X_{n+1}) \ . \qquad \text{(IV.2)}$$

This transformation is orthogonal. As a result, the random variables V_n are i.i.d. $N(0, \sigma^2)$ so that one may apply procedures for selecting the best gamma population to the random variables $V_1^2, V_2^2, \ldots$. Because procedures for selecting the best gamma population generally depend on the observations only through their sum, it is of interest to note that

$$\sum_{i=1}^{n} V_i^2 = \sum_{i=1}^{n} (X_i - \overline{X}_{(n)})^2 \overset{\text{def.}}{=} S_n^2 , \qquad \text{(IV.3)}$$

since, as may be verified algebraically,

$$S_{n+1}^2 - S_n^2 = V_{n+1}^2 \qquad \text{for } n=1, 2, \ldots . \qquad \text{(IV. 4)}$$

CHAPTER B. AN EQUIVALENCE BETWEEN THE PROBLEMS OF SELECTING THE BEST GAMMA POPULATION AND SELECTING THE BEST MULTINOMIAL CATEGORY

Introduction

As discussed by Bechhofer et al. (6) there is a close relationship between the problem of selecting the best multinomial category and the problem of selecting the best Poisson process. We will use the results of Bechhofer et al. (6) to show that any procedure for selecting the best multinomial category may be used to select the best gamma population. This conclusion will be used in Chapter C to provide a proof of the assertion that any SC is an LFC with respect to the inverse sampling procedures for selecting the best multinomial category and to justify the methods proposed in Part II for evaluating PRCS for the inverse sampling procedures. It will be used again to derive the sequential procedures proposed in Chapter D.

Selecting the best Poisson process

Consider k Poisson processes $\pi_1, \ldots, \pi_k$. Let $N_i(t)$ represent the number of occurrances in process i up to time t and let $X_i(t)$ be the wait from time t to the next occurrence in process i. If we let $F(n(t);\ \theta_i) = Pr[N_i(t) \leq n(t)]$ and denote the corresponding density function by $f(n(t);\ \theta_i)$, then the problem of selecting the best process may be specified as

Problem P.a. Selecting the Poisson process with largest intensity parameter:

1. Density function of the observations:

$$f(n(t);\ \theta_i) = e^{-\theta_i t}(\theta_i t)^{n(t)}/n(t)!$$

2. Indifference function: $g(\theta_1, \ldots, \theta_k) = \frac{\theta_{(k-1)}}{\theta_{(k)}} \leq \theta^* \qquad (\theta^* < 1)$

3. Best process: $\pi_{(k)}$

Problem P.b. Selecting the Poisson process with smallest intensity parameter:

1. Density function of the observations:

$$f(n(t);\ \theta_i) = e^{-\theta_i t}(\theta_i t)^{n(t)}/n(t)!$$

2. Indifference function: $g(\theta_1, \ldots, \theta_k) = \frac{\theta_{(1)}}{\theta_{(2)}} \leq \theta^* \qquad (\theta^* < 1)$

3. Best process: $\pi_{(1)}$

It is well known that the waiting time, $X_i(t)$, has an exponential distribution. Thus at any time t,

$$\Pr[\text{next occurrence is from } \pi_i]$$

$$= \int_0^{\infty} \frac{1}{\theta_i} e^{-x \sum_{j=1}^{k} \theta_j^{-1}} dx$$

$$= \theta_i^{-1} / \sum_{j=1}^{k} \theta_j^{-1}$$

$$= p_i \text{ , say .} \qquad \text{(IV.5)}$$

That is, the first arrival after time t occurring in process i corresponds to the multinomial event associated with an observation from category i. The multinomial probability is p_i. Furthermore, if θ^* is the indifference constant specified for the Poisson problem, then $\theta^* p_{(k)} \geq p_{(k-1)}$ if and only if $\theta^* \theta_{(2)} \geq \theta_{(1)}$. Thus any procedure for selecting the best multinomial category may be used to select the best Poisson process. That any procedure for selecting the best Poisson process may be used for selecting the best gamma population follows from the next theorem (given as a problem by Parzen (31) on page 143):

Theorem IV.1: Let $Y \sim \Gamma(p;\ \theta)$ and let $U_{(1)}, \ldots, U_{(p-1)}$ be the order statistics of a sample of size $p-1$ from a uniform $(0,\ Y)$ distribution. If $U_{(p)} = Y - U_{(p-1)}$, $Z_i = U_{(i)} - U_{(i-1)}$ $(i = 2, \ldots, p)$, and $Z_1 = U_{(1)}$,

then the random variables $Z_1, \ldots, Z_p$ are (unconditionally) independent and identically distributed, having an exponential distribution with parameter θ.

Proof:

$$f_{Z_1, \ldots, Z_p}(z_1, \ldots, z_p)$$

$$= f_{Z_1, \ldots, Z_p | Y=y}(z_1, \ldots, z_p) f_Y(y) \qquad \text{(IV.6)}$$

$$= f_{U_{(1)}, \ldots, U_{(p)} | Y=y}(u_1, \ldots, u_p) f_Y(y) \qquad \text{(IV.7)}$$

$$= [\Gamma(p) y^{1-p}] \{ \frac{y^{p-1} e^{-y/\theta}}{\Gamma(p) \theta^p} \} \qquad \text{(IV.8)}$$

$$= \prod_{i=1}^{p} \frac{1}{\theta} e^{-z_i/\theta}, \qquad 0 < z_i < \infty$$

$$= 0, \qquad \text{otherwise.} \qquad \text{(IV.9)}$$

Thus procedures for selecting the best gamma population may be derived from the procedures of Part II as follows: (1) Create exponential random variables from the original gamma observations by sampling uniform random variables in the manner indicated in Theorem IV.1. (2) Form Poisson processes with the exponential random variables thus obtained. (3) Form a sequence of multinomial random variables using the

method described earlier in this section. (4) Use the procedures of Part II on this sequence of multinomial random variables.

CHAPTER C. REVIEW OF THE LITERATURE

Single sample procedures

The single sample procedures described in this section were proposed by Bechhofer and Sobel (7). They are recommended on the basis that they are simple to administer.

Procedure G. a. 1. A single sample procedure for selecting the gamma population with largest scale parameter: "Take R observations from each population and select the population corresponding to $\max_i \sum_{j=1}^{R} X_{ij}$ as best."

Procedure G. b. 1. A single sample procedure for selecting the gamma population with smallest scale parameter: "Take R observations from each population and select the population corresponding to $\min_i \sum_{j=1}^{R} X_{ij}$ as best."

Tables for evaluating PRCS are given by Bechhofer and Sobel (7), Gupta and Sobel (22), and Gupta (17). A fairly accurate approximate method (obtained from the fact that the logarithm of a gamma random variable has approximately a normal distribution) is also given in (7). As will be verified, PRCS may also be obtained by using formulae for the inverse sampling procedures discussed in Part II. To evaluate PRCS for Procedure G. a. 1, one may evaluate PRCS for the inverse sampling procedure for selecting the least probable multinomial category assuming a boundary of pR. To evaluate PRCS for Procedure G. b. 1, one may evaluate PRCS for the inverse sampling procedure for selecting the most probable multinomial category assuming a boundary of pR and an indifference

constant equal to the reciprocal of that given in the original (gamma) formulation of the problem.

We now consider in detail the equivalence between the inverse sampling procedures of Part II and the single sample procedures of this section.

Theorem IV.2: The PRCS for Procedure G.a.1 with a sample size R from each population equals PRCS for Procedure M.b.3 using a stopping bound of pR with

$$p_i = \frac{1/\theta_i}{\sum_{j=1}^{k} 1/\theta_j}, \quad i=1,\ldots,k. \qquad \text{(IV.10)}$$

To demonstrate the equality we note first that PRCS using Procedure G.a.1 equals the probability that, in considering k Poisson processes with intensity parameter $\lambda_i = 1/\theta_i$, i=1,..., k, the pRth occurrence occurs in the process with smallest intensity parameter after the pRth occurrence in any of the other k-1 processes. From the previous chapter, this probability equals the probability that, in considering a multinomial population with probabilities

$$p_i = \frac{\lambda_i}{\Sigma\lambda_i} = \frac{1/\theta_i}{\Sigma\, 1/\theta_i}, \quad i=1,\ldots, k, \qquad \text{(IV.11)}$$

the category having smallest probability is the last category to yield pR outcomes. This is PRCS using Procedure M.b.3 with a stopping bound of pR.

We may now easily prove the assertion of Part II that, with respect to the inverse sampling procedure for selecting the least probable multinomial category, any slippage configuration is least favorable. It follows from Theorem IV. 2 that if an SC is least favorable with respect to the single sample procedure for selecting the best gamma population, then an SC is also least favorable with respect to the inverse sampling procedure for selecting the best multinomial category. Thus it is sufficient to show that any SC is an LFC with respect to the single sample procedure for selecting the gamma population with largest scale parameter. As indicated by Bechhofer and Sobel (7), the probability of correct selection for the single sample procedure for Problem G. a is obviously an increasing function of $\theta_{(i+1)}/\theta_{(i)}$ so that any SC is an LFC.

Sequential nonscreening procedures

None of the four procedures described in this section are recommended for use. Three of the four fail to maintain specified levels of PRCS. The question as to whether or not the fourth maintains specified levels of PRCS is as yet unanswered. However, it appears to be very insensitive when the parametric configuration is a GSC and hence is of little use regardless of its theoretical viability.

The following procedure was proposed by Bechhofer and Sobel (8) as a procedure for selecting the normal population having smallest variance. We shall show that if $k = 2$ and $\sigma_1^2 = \theta\sigma_2^2$, then

$$\lim_{\theta \to 0} \text{PRCS} = 0 \ . \qquad \text{(IV. 12)}$$

Essentially, the proof consists of showing that the procedure selects the population associated with the largest estimate of variance after the first stage of sampling.

Procedure N. b. 1:

Let $$\bar{x}_{im} = \sum_{j=1}^{m} x_{ij}/m \qquad i=1,\ldots,k \qquad \text{(IV. 13)}$$

$$s^2_{im} = \sum_{j=1}^{m} (x_{ij} - \bar{x}_{im})^2 \qquad i=1,\ldots,k \qquad \text{(IV. 14)}$$

$$R_{jim} = \frac{s^2_{jm}}{s^2_{im}} \qquad m=2,3,\ldots \qquad \text{(IV. 15)}$$

$$L_{im} = \left[\prod_{j=1}^{k} R_{jim}\right]^{\frac{m-3}{2}} \left[1+\theta^* \sum_{\substack{j=1 \\ \neq i}}^{k} R_{jim}\right]^{-\frac{k(m-1)}{2}} \qquad \text{(IV. 16)}$$

$$L_{(k)m} = \max_i (L_{im})$$

$$P_m = \frac{L_{(k)m}}{\sum_{i=1}^{k} L_{im}} \qquad \text{(IV. 17)}$$

"At the mth stage (m=2, 3, . . .) take the vector $(x_{1m}, \ldots, x_{km})$ and compute P_m . If $P_m \geq P^*$, stop and select the population associated with $L_{(k)m}$; if $P_m < P^*$, take the (m+1)st vector observation and compute P_{m+1}" .

It will be proven that Procedure N. b. 1 does not maintain levels of PRCS as set forth in Problem N. b. For the case $k=2$, Procedure N. b. 1 may be reformulated as follows:

$$L_{1m} = \left(\frac{s_{2m}^2}{s_{1m}^2}\right)^{\frac{m-3}{2}} / \left[1 + \theta^* \frac{s_{2m}^2}{s_{1m}^2}\right]^{m-1} \qquad \text{(IV. 18)}$$

$$L_{2m} = \left(\frac{s_{1m}^2}{s_{2m}^2}\right)^{\frac{m-3}{2}} / \left[1 + \theta^* \frac{s_{1m}^2}{s_{2m}^2}\right]^{m-1} \qquad \text{(IV. 19)}$$

$$P_m = \frac{L_{(2)m}}{L_{1m} + L_{2m}} \qquad m=2, 3, \ldots . \qquad \text{(IV. 20)}$$

Procedure: Continue sampling until $P_m \geq P^*$, at which point the population corresponding to $L_{(2)m}$ is selected as best. Equivalently, one may sample until

$$\frac{L_{1m}}{L_{(2)m}} + \frac{L_{2m}}{L_{(2)m}} \leq \frac{1}{P^*} \; . \qquad \text{(IV. 21)}$$

In this case, π_2 is selected if $\frac{L_{1m}}{L_{2m}} \leq \frac{1}{P^*} - 1$

or if

$$\left(\frac{s_{2m}^2}{s_{1m}^2}\right)^{m-3}\left[\frac{1+\theta^*\frac{s_{2m}^2}{s_{1m}^2}}{1+\theta^*\frac{s_{1m}^2}{s_{2m}^2}}\right]^{-m+1}$$

$$= \left(\frac{s_{2m}^2}{s_{1m}^2}\right)^{-2}\left[\frac{1+\theta^*\frac{s_{2m}^2}{s_{1m}^2}}{\frac{s_{2m}^2}{s_{1m}^2}+\theta^*}\right]^{-m+1} \leq \frac{1}{P^*} - 1 \quad \text{(IV. 22)}$$

It shall be shown that

$$\lim_{\theta \to 0} \Pr[CS|\sigma_1^2 = \theta\sigma_2^2] = 0 \ . \quad \text{(IV. 23)}$$

Let $F = \frac{s_{22}^2}{s_{12}^2}\theta$, so that $F \sim F(2, 2)$. π_2 is selected with $m=2$ if

$$\left(\frac{F}{\theta}\right)^{-2}\left[\frac{\theta+\theta^*F}{F+\theta^*\theta}\right]^{-1} \leq \frac{1}{P^*} - 1 \quad \text{(IV. 24)}$$

or if

$$\left(\frac{\theta}{F}\right)^{2}\left[\frac{F+\theta^*\theta}{\theta^*F}\right] \leq \frac{1}{P^*} - 1$$

or if

$$\frac{\theta^2}{\frac{1}{P^*} - 1} \leq \frac{F^3}{1+F}$$

or if

$$\frac{2\theta^2}{\frac{1}{P^*} - 1} < F^3 \qquad \text{(IV.25)}$$

(for θ sufficiently small). Since

$$\lim_{\theta \to 0} \Pr[F^3 \geq \theta] = 1 ,$$

$$\lim_{\theta \to 0} \Pr[\pi_2 \text{ selected with } m = 2 \mid \sigma_1^2 = \theta\sigma_2^2] = 1$$

so that

$$\lim_{\theta \to 0} \Pr[CS \mid \sigma_1^2 = \theta\sigma_2^2] = 0 . \qquad \text{(IV.26)}$$

That the analogous procedure for selecting the population with largest variance fails to maintain specified levels of PRCS may be demonstrated in similar fashion. It shall therefore not be discussed explicitly.

The next two procedures for selecting the best gamma population were proposed by O'Brien (30). Since Procedure N.b.1 is obtained from Procedure G.b.2 if the Jacobian in the derivation of the latter procedure is omitted, we conjecture that the two procedures listed below were the procedures which Bechhofer and Sobel desired to obtain in (8). These procedures are of interest in that they provide an example of a case in which some GSC's are less favorable than SC's.

Procedure G. a. 2. A sequential nonscreening procedure for selecting the gamma population with largest scale parameter: "Sample k-dimensional vectors of observations (i. e.: one observation from each population) until for some i

$$\sum_{\substack{t=1 \\ \neq i}}^{k} \left[\frac{X_{..} + (\theta^*-1)X_{i.}}{X_{..} + (\theta^*-1)X_{t.}}\right]^{pnk} \leq \frac{1-P^*}{P^*} \quad . \qquad \text{(IV. 27)}$$

At this point, terminate the experiment and select π_i as best."

Procedure G. b. 2. A sequential nonscreening procedure for selecting the gamma population with smallest scale parameter: "Sample k dimensional vectors of observations sequentially until for some i

$$\sum_{\substack{t=1 \\ \neq i}}^{k} \left[\frac{\theta^* X_{..} + (1-\theta^*)X_{i.}}{\theta^* X_{..} + (1-\theta^*)X_{t.}}\right]^{pnk} \leq \frac{1-P^*}{P^*} \quad . \qquad \text{(IV. 28)}$$

At this point terminate the experiment and select π_i as best."

We next show that Procedure G. a. 2 fails to maintain specified levels of PRCS. Let

$$\theta^* = \frac{1}{2}, \qquad X_{11} \sim \Gamma(1;2), \qquad X_{21} \sim \Gamma(1;1)$$

and

$$X_{i1} \sim \Gamma[1;\theta(k)] \qquad \text{(IV. 29)}$$

$i=3, \ldots, k$. Since for any $c > 0$.

$$\lim_{\theta(k) \to 0} \Pr[X_{i1} \leq c;\ i=3, \ldots, k] = 1 , \qquad \text{(IV.30)}$$

it follows that given any k, c, and $P(0 < P < 1)$ there is a number $\theta(k)$ such that

$$\Pr[X_i \leq \frac{c}{k},\ i=3, \ldots, k] \geq P . \qquad \text{(IV.31)}$$

For given k, π_2 is selected after the 1st observation if

$$T_2(k) = \{\frac{X_{..} + (-\frac{1}{2}) X_{21}}{X_{..} + (-\frac{1}{2}) X_{11}}\}^k + \sum_{t=3}^{k} \{\frac{X_{..} + (-\frac{1}{2}) X_{21}}{X_{..} + (-\frac{1}{2}) X_{t1}}\}^k \leq \frac{1-P^*}{P^*} \qquad \text{(IV.32)}$$

With probability no less than P, $X_{i1} \leq \frac{c}{k}$ for $i=3, \ldots, k$. In this case

$$T_2(k) < \{\frac{2X_{11} + X_{21} + 2c}{X_{11} + 2X_{21} + 2c}\}^k + \sum_{t=3}^{k} \{\frac{2X_{11} + X_{21} + 2c}{2X_{11} + 2X_{21} + 2c - \frac{c}{k}}\}^k \qquad \text{(IV.33)}$$

$$= \{\frac{2X_{11} + X_{21} + 2c}{X_{11} + 2X_{21} + 2c}\}^k + (k-3)\{\frac{2X_{11} + X_{21} + 2c}{2X_{11} + 2X_{21} + 2c - \frac{c}{k}}\}^k \qquad \text{(IV.34)}$$

Let $$D = X_{21} - X_{11} \quad (X_{21} = D + X_{11}) \qquad \text{(IV.35)}$$

then $$T_2(k) \leq \{\frac{3X_{11}+D+2c}{3X_{11}+2D+2c}\}^k$$

$$+ (k-3)\{\frac{3X_{11}+D+2c}{4X_{11}+2D+2c-\frac{c}{k}}\}^k \qquad \text{(IV.36)}$$

Thus if $D > 0$, then $\lim_{k \to \infty} T_2(k) = 0$. This means that given any $c > 0$, P^* and $P(0 < P, P^* < 1)$, there is a k (and associated $\theta(k)$) such that $X_{21} > X_{11}$ and $X_{i1} < \frac{c}{k}$ $(i=3, \ldots, k)$ guarantees that

$$T_2(k) \leq \frac{1-P^*}{P^*} , \qquad \text{(IV.37)}$$

where $Pr[X_{21} > X_{11}] = \frac{1}{3}$ and

$$Pr[X_{i1} < \frac{c}{k}, \quad i=3, \ldots, k] \geq P . \qquad \text{(IV.38)}$$

Therefore, arbitrarily choosing $P = \frac{3}{4}$, one may choose a value of k and corresponding parametric configuration as discussed above so that

$$Pr[\text{Incorrect Selection}]$$

$$\geq Pr[\pi_2 \text{ selected on first observation}] \qquad \text{(IV.39)}$$

$$\geq Pr[X_{21} > X_{11} \text{ and } X_{i1} < \frac{c}{k}, \ i=3, \ldots, k] \qquad \text{(IV.40)}$$

$$\geq \frac{1}{3} \cdot \frac{3}{4} \qquad \text{(IV.41)}$$

$$\geq \frac{1}{4} \ , \qquad \text{(IV.42)}$$

regardless of P^* , the required level of PRCS, specified by the experimenter.

CHAPTER D. NEW SEQUENTIAL PROCEDURES FOR SELECTING THE BEST GAMMA POPULATION

Introduction

The procedures of this chapter are recommended on the basis that they provide a savings in ASN relative to the single sample procedures. They are obtained by using the method discussed in Chapter B to transform the problem of selecting the best gamma population into one of selecting the best multinomial category. Procedures G. a. 4 and G. b. 4 are obtained from the sequential nonscreening procedures for selecting the best multinomial category. As a result, the proof that Procedures G. a. 4 and G. a. 5 maintain specified levels of PRCS follows immediately from the proof by Bechhofer et al. (6) that the sequential nonscreening procedures for selecting the best multinomial category maintain specified levels of PRCS. Since Procedures G. a. 5 and G. b. 5 are obtained from the screening procedures for selecting the best multinomial category, the proof that Procedures G. a. 5 and G. b. 5 maintain specified levels of PRCS is conditional on the assumption that SC's are least favorable. However, it follows from Theorems II. 7. a and II. 7. b that specified levels of PRCS are maintained for GSC's of the type considered in the previous chapter. In all the procedures of this chapter it is assumed that the experimenter samples as many times as desired during each stage. ASN will be minimized by testing after each observation is sampled.

Specification of the procedures

We adopt the following notation:

$X_{ir\ell}$ the ℓth observation from π_i taken during stage r

n_{ir} the number of observations drawn from π_i during stage r.

$$M_r = \min_i \{ET_{ir} + \sum_{\ell=1}^{n_{ir}} X_{ir\ell}\} \qquad \text{(IV. 43)}$$

m_r = the value of i which accomplishes the minimization in (IV. 43).

$$EO_{i1} = ET_{i1} = 0$$

for $i \neq m_r$, define

$$ET_{i,\, r+1} = ET_{ir} + \sum_{\ell=1}^{n_{ir}} X_{ir\ell} - M_r \qquad \text{(IV. 44)}$$

$\{Z_{ir}\}$ a sequence of independent random variables each having a binomial distribution with parameter a_{ir} and probability p_{ir}, defined further in the decision rule.

$$EO_{i,\, r+1} = Z_{ir} + 1 \qquad \text{(IV. 45)}$$

and define

$$ET_{m_r, r+1} = EO_{m_r, r+1} = 0 \qquad \text{(IV.46)}$$

$$N_{ir} = p \sum_{\ell=1}^{r} n_{i\ell} - EO_{i, r+1} \qquad \text{(IV.47)}$$

where p is the known shape parameter of the gamma populations.

Procedure G.a.4. A sequential procedure for selecting the gamma population largest scale parameter:

Sampling rule: "Initiate the experiment by taking one observation from each population. Thereafter, for stage $r=1, 2, \ldots,$ sample one observation at a time, always sampling from the population corresponding to

$$\min_i \{\sum_\ell X_{ir\ell} + ET_{ir}\} . \qquad \text{(IV.48)}$$

Decision Rule: for stage $r = 1, 2, \ldots$

Step 1: Compute M_r and $ET_{i, r+1}$, $i=1, \ldots, k$.

Step 2: Obtain Z_{ir} $(i=1, \ldots, k)$ by sampling from a binomial population with parameter a_{ir} and probability p_{ir}. If $n_{ir} > 0$, set $a_{ir} = p-1$ and $p_{ir} = ET_{i, r+1}/X_{irn_{ir}}$. If $n_{ir} = 0$, set $a_{ir} = EO_{ir} - 1$ and $p_{ir} = ET_{i, r+1}/ET_{ir}$.

Step 3: Evaluate $EO_{i, r+1}$ and N_{ir} $(i=1, \ldots, k)$.

Step 4: If

$$\sum_{j=1}^{k} \theta^{*N_{jr} - N_{(1)r}} \leq \frac{1}{P^*} \qquad \text{(IV.49)}$$

stop and select the population associated with $N_{(1)r}$ as best. Otherwise proceed to stage r+1 ."

Procedure G. b. 4. A sequential procedure for selecting the gamma population with smallest scale parameter: "Follow the specifications of Procedure G. a. 4 but replace Step 4 with

Step 4': If

$$\sum_{j=1}^{k} \theta^{*N_{(k)r} - N_{jr}} \leq \frac{1}{P^*} \qquad \text{(IV.50)}$$

stop and select the population associated with $N_{(k)r}$ as best. Otherwise proceed to stage r+1 ."

Of the procedures discussed in this thesis for selecting the best gamma population, the following two procedures provide the smallest ASN. The notation is the same as in Procedure G. a. 4 with the following additions and modifications

S_r = {i: π_i is not screened during stage r} r=1, 2, ...

C_r = the cardinality of S_r

$$M_r = \min_{i \in S_r} \{ET_{ir} + \sum_{\ell=1}^{n_{ir}} X_{ir\ell}\} \qquad \text{(IV.51)}$$

m_r = the value of i which achieves the minimization in IV. 51

for $i \neq m_r$,

$$ET_{i, r+1} = ET_{ir} + \sum_{\ell=1}^{n_{ir}} X_{ir\ell} - M_r \quad \text{if } i \in S_r \qquad \text{(IV.52)}$$

$$= ET_{ir} \quad \text{if } i \notin S_r \qquad \text{(IV.53)}$$

$$EO_{i, r+1} = Z_{ir} + 1 \quad \text{if } i \in S_r \qquad \text{(IV.54)}$$

$$= EO_{ir} \quad \text{if } i \notin S_r . \qquad \text{(IV.55)}$$

$$ET_{m_r, r+1} = EO_{m_r, r+1} = 0 \qquad \text{(IV.56)}$$

$$N_{ir} = p n_{ir} + EO_{ir} - EO_{i, r+1} \qquad \text{(IV.57)}$$

$$N_{.r} = \sum_{i=1}^{k} N_{ir} \qquad \text{(IV.58)}$$

$$L_{i0} = 1 \qquad \text{(IV.59)}$$

$$L_{ir} = L_{i,r-1}(\theta^*)^{N_{ir}}(C_r - 1 + \theta^*)^{-1} \quad \text{if } i \in S_r \qquad \text{(IV.60)}$$

$$= L_{i,r-1}\, C_r^{-1} \quad \text{if } i \notin S_r . \qquad \text{(IV.61)}$$

We consider the following screening rules:

Screening rule 1: Screen the population corresponding to $L_{(t)r}$ during stage $r+1$ if

$$\sum_{j=t+1}^{k} L_{(j)r}/L_{(t)r} \geq (k-1)/(1-P^*)-t . \qquad \text{(IV.62)}$$

Screening rule 2: Screen the population corresponding to $L_{(t)r}$ if

$$\sum_{j=t+1}^{k} L_{(j)r}/L_{(t)r} \geq \frac{t(1-2P^*)}{1-P^*} + k-1 . \qquad \text{(IV.63)}$$

Procedure G. a. 5. A sequential screening procedure for selecting the gamma population with largest scale parameter:

Sampling rule: "Initiate the experiment by taking one observation from each population. Thereafter, for stage $r=1, 2, \ldots$, sample one observation at a time always sampling from the population corresponding to

$$\min_{i \in S_r} \{\Sigma_{\ell} X_{ir\ell} + ET_{ir}\} \qquad \text{(IV.64)}$$

Decision rule: for stage $r=1, 2, \ldots$

Step 1: Compute M_r and $ET_{i, r+1}$, $i=1, \ldots, k$.

Step 2: For $i \in S_r$, obtain Z_{ir} by sampling from a binomial population with parameter a_{ir} and probability p_{ir}. If $n_{ir} > 0$ set $a_{ir} = p-1$ and $p_{ir} = ET_{ir}/X_{irn_{ir}}$. If $n_{ir} = 0$, set $a_{ir} = EO_{ir} - 1$ and $p_{ir} = ET_{i, r+1}/ET_{ir}$.

Step 3: Evaluate $EO_{i, r+1}$, N_{ir}, and L_{ir} $(i=1, \ldots, k)$.

Step 4: If

$$\frac{L_{(k)r}}{\sum_{\ell=1}^{k} L_{\ell r}} \geq P^* , \qquad \text{(IV.65)}$$

stop and select the population corresponding to $L_{(k)r}$ as best. Otherwise proceed to step 5.

Step 5: Determine S_{r+1} using screening rule 1 or 2 (it is assumed only one rule is used throughout the experiment) and proceed to the r+1st stage of sampling."

Procedure G.b.5. A sequential procedure for selecting the gamma population with smallest scale parameter: "Follow the specifications of Procedure G.a.5 but replace θ^* by its reciprocal. Only screening rule one should be used."

The theory associated with Procedures G. a. 5 and G. b. 5 assumes that, for $p \neq 1$, one samples Z_{ir} from a binomial population with parameter a_{ir} and probability p_{ir}. In practice one may prefer to replace Z_{ir} with its expectation: $E(Z_{ir}) = a_{ir}p_{ir}$. When $p=1$ no randomization is needed.

CHAPTER E. EMPIRICAL COMPARISONS

Relying on the equivalence discussed in Chapter B between the problems of selecting the best multinomial category and selecting the best gamma population we use the estimates of PRCS, ASN, and standard deviation of the ASN obtained for the screening procedures in Part II to obtain similar estimates for the screening procedures for selecting the best gamma population.

Table IV. 1 below compares the screening procedures with a single sample procedure proposed by Bechhofer and Sobel (7) for selecting the gamma population having largest scale parameter. Entries (based on 100 independent samples for 3, 4, and 5 populations with $P^* = .75$) give the number of correct selections, observed average sample number and the estimated standard deviation of the sample number. Table IV. 2 gives similar comparisons for the problem of selecting the population having smallest scale parameter for a universe of five populations. Parametric configurations considered are: equal means (EM), slippage configuration, and "staggered configuration". Since the screening procedure is invariant under permutation of the indexing of the populations, PRCS for the EM configuration is $1/k$. PRCS for the SC is no less than P^*. The only configuration considered here for which analytical results are unavailable is the staggered configuration. In this case the estimated PRCS is well above P^*.

The method used to determine the sample number for the single sample procedure is discussed by Bechhofer and Sobel (7). It appears that

the sequential screening procedures offer a substantial reduction in ASN in all cases.

Table IV. 1. Empirical comparisons of procedures for selecting the gamma population with largest scale parameter (100 samples, $P^* = .75$)

	k = 3		
	single sample	screen rule 2	screen rule 1
	$\theta_1 = \ldots = \theta_k$		$\theta^* = .75$
Average sample number	76.0	67.5	62.5
Standard deviation of the sample number		46.5	39.3
	$\theta_{(1)} = \theta_{(i)}\theta^*$ $(i=2,\ldots,k)$		$\theta^* = 6/7$
Proportion of correct selections		.74	.80
Average sample number	261.0	165.1	166.4
Standard deviation of the sample number		109.7	97.3
	$\theta_{(1)} = \theta_{(i)}\theta^{*i-1}$ $(i=2,\ldots,k)$		$\theta^* = 6/7$
Proportion of correct selections		.89	.85
Average sample number	261.0	143.2	130.2
Standard deviation of the sample number		106.3	73.5

k = 4			k = 5		
single sample	screen rule 2	screen rule 1	single sample	screen rule 2	screen rule 1
$\theta_1 = \ldots = \theta_k$			$\theta^* = .75$		
138.8	127.6	119.1	208.5	170.2	179.8
	89.9	64.2		122.6	110.9
$\theta_{(1)} = \theta_{(i)}\theta^*$ (i=2, ..., k)			$\theta^* = 6/7$		
	.74	.72		.74	.80
478.4	307.1	282.2	719.8	376.8	390.0
	189.4	148.1		183.7	187.6
$\theta_{(1)} = \theta_{(i)}\theta^{*i-1}$ (i=2, ..., k)			$\theta^* = 6/7$		
	.87	.93		.85	.87
478.4	180.7	189.8	719.8	241.2	245.9
	95.5	106.1		111.8	119.3

Table IV. 2. Empirical comparisons of procedures for selecting the gamma population with smallest scale parameter (100 samples, $P^* = .75$, $k=5$)

	$\theta^* = 4/3$ $\theta_1 = \ldots = \theta_5$		$\theta^* = 7/6$ $\theta_{(5)} = \theta^*\theta_{(i)}$ $i = 1, \ldots, 4$		$\theta^* = 7/6$ $\theta_{(5)} = (\theta^*)^{5-i}\theta_{(i)}$ $i = 1, \ldots, 5$	
	single sample	screen	single sample	screen	single sample	screen
Proportion of correct selections				.77		.94
Average sample number	208.45	124.79	719.75	354.18	719.75	222.56
Standard deviation of the sample number		75.3		193.8		128.2

PART V. SELECTING THE BEST NORMAL POPULATION

CHAPTER A. INTRODUCTION AND REVIEW OF THE LITERATURE

In this chapter we consider the problem of selecting the normal population with largest (smallest) mean assuming common known variance. The specifications for these problems are listed below, where we standardize by assuming unit variance.

Problem N. a. Selecting the normal population with largest mean:

1. $f_i(x) = (2\pi)^{-\frac{1}{2}} \exp - \frac{1}{2}(x-\mu_i)^2$

2. $g(\mu_1, \ldots, \mu_k) = \mu_{(k)} - \mu_{(k-1)} \geq \delta^*$ $\qquad (\delta^* > 0)$

3. Best population: $\pi_{(k)}$.

Problem N. b. Selecting the normal population with smallest mean:

1. $f_i(x) = (2\pi)^{-\frac{1}{2}} \exp - \frac{1}{2}(x-\mu_i)^2$

2. $g(\mu_1, \ldots, \mu_k) = \mu_{(2)} - \mu_{(1)} \geq \delta^*$ $\qquad (\delta^* > 0)$

3. Best population: $\pi_{(1)}$.

We state procedures explicitly only for Problem N. b. One may multiply the observations by minus one if the original problem of interest is Problem N. a. The following single sample procedure was proposed by

Bechhofer (4) and may be recommended on the basis of its administrative simplicity.

Procedure N. b. 1. A single sample procedure for selecting the normal population with smallest mean: "Sample n observations from each population and select the population with smallest observed mean as best."

The sequential nonscreening procedure proposed by Bechhofer, Kiefer, and Sobel (6) is simple to administer and provides a savings in ASN relative to the single sample procedure for most parametric configurations (a notable exception is the EM configuration). We state the procedure using the following notation:

$$X_{ij} = \text{the observation taken from } \pi_i \text{ during stage } j .$$

$$Y_{ij} = \sum_{s=1}^{j} X_{is}$$

$$Y_{(1)n} = \min (Y_{1n}, \ldots, Y_{kn}) .$$

Procedure N. b. 2. A sequential nonscreening procedure for selecting the normal population with smallest mean: "Sample k-dimensional vectors of observations sequentially until for some n

$$\sum_{j=1}^{k} \exp \delta^*(Y_{(1)n} - Y_{jn}) < 1/P^* \qquad (V.1)$$

At this point stop and select the population associated with $Y_{(1)n}$ as best."

We next consider a sequential procedure employing elimination. This procedure, which was proposed by Paulson (33), eliminates populations from candidacy for selection until only one population is left. At this point the experiment is terminated and the remaining population is selected as best.

As is true of all procedures employing elimination, Paulson's procedure may yield awkward posterior probability statements. It may also yield high ASN for k large or P* small. However it is relatively simple to administer and may provide a substantial reduction in ASN over the single sample procedure when P* is close to unity. The following additional notation is adopted:

λ = a constant specified by the experimenter $(0 < \lambda < \delta^*)$

$$a_\lambda = (\delta^* - \lambda)^{-1} \ell n((k-1)/(1-P^*)) .$$

Paulson recommends choosing $\lambda = \delta^*/4$.

Procedure N. b. 3. A sequential procedure employing elimination for selecting the normal population with smallest mean: "Initiate the experiment by sampling one observation from each population. This concludes the first stage of sampling. During the rth stage of sampling (r=2, 3, . . .), sample one observation from each population not yet eliminated. Population j is eliminated after the rth stage of sampling if

$$Y_{jr} - Y_{(1)r} > a_\lambda - r\lambda . \qquad (V.2)$$

When only one population remains, terminate the experiment and select the remaining population as best. "

CHAPTER B. A NEW PROCEDURE FOR SELECTING THE BEST NORMAL POPULATION

In this chapter we propose a procedure which has a flexible sampling rule. The procedure gives a smaller average sample number than any of the other procedures for selecting the best normal population considered in this thesis. It is based on the sequential screening procedure for selecting the most probable multinomial category. As a result, the proof that the procedure of this section maintains specified levels of PRCS is conditional on the assumption that an SC is an LFC. It follows from Theorem II. 7a that specified levels of PRCS are maintained for GSC's having the form

$$\delta^* + \mu_{(1)} = \mu_{(2)} = \cdots = \mu_{(k-t-1)}$$

$$\mu_{(k-t)} = \cdots = \mu_{(k)}$$

as $\mu_{(k)}$ approaches infinity.

Procedure N. b. 4. A sequential screening procedure for selecting the normal population with smallest mean: "For each observation, X , sampled from a normal population make the transformation

$$Y = \exp(.01X) . \qquad (V.3)$$

Use Procedure G. b. 5 for selecting the gamma population with smallest

scale parameter on the transformed observations, using a shape parameter of $p = 10,000$ and an indifference constant of

$$\theta^* = \exp(.01\delta^*) \ . \ " \qquad (V.4)$$

Procedure N. b. 4 is based on the fact that the logarithm of a gamma random variable has approximately a normal distribution. As discussed by Bechhofer and Sobel (7), if $Y \sim \Gamma(p, 1)$, then $\ell n\, Y \dot{\sim} N(a, b)$ where,

$$a = 1/2p + 1/12p^2 + \ell n\, p \qquad (V.5)$$

$$b = 2/(2p-1) \ .$$

(The notation $Y \sim N(a, b)$ indicates that Y has a normal distribution with mean a and variance b . $Y \sim \Gamma(a, b)$ indicates that Y has a gamma distribution with shape parameter a and scale parameter b . $\dot{\sim}$ indicates the relation is approximate.) As a result, if

$$Z = b^{-\frac{1}{2}}(\ell n\, Y - a) + \mu \qquad (V.6)$$

then $Z \dot{\sim} N(\mu, 1)$.

Working in the opposite direction, if $Z \sim N(\mu, 1)$ and

$$Y = \exp[b^{\frac{1}{2}}(Z-\mu) + a] \qquad (V.7)$$

then $Y \dot{\sim} \Gamma(p, 1)$ and

$$\exp(Zb^{\frac{1}{2}}) \dot{\sim} \Gamma[p;\ \exp(\mu b^{\frac{1}{2}} - a)] \qquad (V.8)$$

The discussion in Bechhofer and Sobel (7), indicates that the approximation is very accurate if $p \geq 20$. Here, p is chosen as 10,000.

CHAPTER C. COMPARISON OF THE PROCEDURES FOR SELECTING THE BEST NORMAL POPULATION

Empirical comparisons

In this section we compare estimates of PRCS, ASN, and standard deviation of the sample number obtained from suitable tables and Monte Carlo studies. Included in the comparisons are the single sample procedure of Bechhofer (4), the sequential nonscreening procedure of Bechhofer, Kiefer, and Sobel (6), the sequential procedure employing elimination proposed by Paulson (33), the sequential screening procedure proposed in this part, and a sequential screening procedure to be considered in Part VI (this latter procedure is referred to as Procedure N.b.5). ASN for the single sample procedure was obtained from tables in Bechhofer et al. (6). All entries for the sequential procedures are based on 100 independent samples unless otherwise indicated. Entries for the nonscreening procedure were obtained from Bechhofer et al. (6). Table V.1 indicates that both sequential procedures offer a substantial reduction in ASN relative to the other procedures when $P^* = .75$. Table V.2 indicates that Procedure N.b.4 offers a substantial reduction in ASN relative to the single sample and sequential nonscreening procedures, but only a slight reduction in ASN relative to the sequential procedure employing elimination when $P^* = .95$.

Table V. 1. Empirical comparisons of procedures for selecting the normal population with smallest (largest) mean

$P^* = .75 \qquad \delta^* = .2 \qquad k = 4$

$\mu_{(1)} = \mu_{(i)} - \delta^* \quad i=2, \ldots, k \qquad \mu_1 = \ldots = \mu_k$

Procedure	Estimate of PRCS	Average sample number	Standard deviation of the sample number	Average sample number	Standard deviation of the sample number
Single sample		282.96[+]		282.96[+]	
Nonscreening	.791[a]	217.74[a]	132.37[a]	290.79[b]	210.41[b]
Eliminating	.88	273.69	107.48	330.54	115.15
Screening (N. b. 4)	.73	165.50	103.30	210.41	130.7
Screening (N. b. 5)	.7275[d]	175.42[d]	97.95[d]	234.59[c]	166.55[c]

[+]Taken from tables of Bechhofer et. al. (6).

[a]Based on 800 samples.

[b]Based on 500 samples.

[c]Based on 200 samples.

[d]Based on 400 samples.

Table V.2. Empirical comparisons of procedures for selecting the normal population with smallest (largest) mean

$P^* = .95 \qquad \delta^* = .2 \qquad k = 2$

$\mu_{(1)} = \mu_{(i)} - \delta^* \quad i=2, \ldots, k \qquad \mu_1 = \ldots = \mu_k$

Procedure	Estimate of PRCS	Average sample number	Standard deviation of the sample number	Average sample number	Standard deviation of the sample number
Single sample		850.44[+]		850.44[+]	
Nonscreening	.955	500.94[a]	291.26[a]	1101.34[b]	718.48[b]
Eliminating	.99	449.22[c]	164.58[c]	742.92[d]	229.45[d]
Screening (N.b.4)	.95	414.04	205.20	732.45	426.80

[+]Taken from tables of Bechhofer et.al. (6).

[a]Based on 800 samples.

[b]Based on 500 samples.

[c]Based on 200 samples including 100 from Paulson (33).

[d]Based on 150 samples including 50 from Paulson (33).

Comparison of the nonscreening procedure with the procedure employing elimination

Let N_{NS} and N_E represent the ASN for the nonscreening procedure and the procedure employing elimination respectively. The following theorem is taken from Perng (35):

Theorem V.1: If $\mu_{(k)} - \mu_{(k-1)} = \delta^*$, then

$$\lim_{P^* \to 1} (N_{NS}/N_E) = k(\delta^* - \lambda)(\delta^*)^{-2}$$

$$\cdot \left[\sum_{i=1}^{k-2} (\lambda + \mu_{(k)} - \mu_{(i)})^{-1} + 2(\lambda + \delta^*)^{-1} \right]^{-1} . \qquad (V.9)$$

It follows from Theorem V.1 that if

$$\mu_1 = \cdots = \mu_{k-1} = \mu_k - \delta^* ,$$

then

$$\lim_{P^* \to 1} (N_{NS}/N_E) = [(\delta^*)^2 - \lambda^2](\delta^*)^{-2} < 1 .$$

Thus, for P^* sufficiently close to unity, the nonscreening procedure will give smaller ASN than the procedure employing elimination for SC's. Theorem V.1 also suggests that the procedure employing elimination will provide a reduction in ASN over the nonscreening procedure when the differences $\mu_{(k-1)} - \mu_{(i)}$ are large and P^* is close to one.

PART VI. SEQUENTIAL SCREENING PROCEDURES FOR SELECTING THE BEST KOOPMAN-DARMOIS POPULATION

CHAPTER A. INTRODUCTION

Bechhofer, Kiefer, and Sobel (6) have proposed a sequential non-screening procedure for selecting the best of several populations when all populations belong to the same one parameter Koopman-Darmois family. They assume that the p.d.f. associated with a random variable drawn from population i may be represented as

$$f(x, \theta_i) = \exp\{P(x)Q(\theta_i) + R(x) + T(\theta_i)\}$$

$$(i=1, 2, \ldots, k), \qquad \text{(VI. 1)}$$

for θ_i in an open interval (H), where the functions $P(x)$ and $R(x)$ do not depend on θ_i while the functions $Q(\theta_i)$ and $T(\theta_i)$ do not depend on x. The functions P, R, Q, and T are all assumed to be known, whereas the θ_i $(i=1, \ldots, k)$ are assumed unknown. For this distribution we may write the selection problem using the format of Chapter A of Part I:

Problem K-D.a. Selecting the Koopman Darmois population with largest parameter:

1. $f_i(x) = \exp\{P(x)Q(\theta_i) + R(x) + T(\theta_i)\}$

2. $g(\theta_1, \ldots, \theta_k) = Q(\theta_{(k)}) - Q(\theta_{(k-1)}) \geq \delta^* \qquad (\delta^* > 0)$

3. Best population: $\pi_{(k)}$

Problem K-D.b. Selecting the Koopman Darmois population with smallest probability:

1. $f_i(x) = \exp\{P(x)Q(\theta_i) + R(x) + T(\theta_i)\}$

2. $g(\theta_1, \ldots, \theta_k) = Q(\theta_{(2)}) - Q(\theta_{(1)}) \geq \delta^* \qquad (\delta > 0)$

3. Best population: $\pi_{(1)}$.

The requirement that the experimenter sample an equal number of observations from each population may result in a very low relative efficiency for the nonscreening procedure relative to a screening or elimination procedure. We illustrate this loss of efficiency for the case of selecting the best normal population.

We compare the relative efficiency of the nonscreening procedure proposed by Bechhofer et al. (6) to the screening procedure to be proposed in this part, Procedure N.b.5. (The same results will hold for any screening procedure, including the procedure given in Part V.) We consider GSC's of type t (t=1, 2, . . . , k-2) having the form

$$u_{(k-t)} = \cdots = u_{(2)} = u_{(1)} + \delta^*$$

$$u_{(k-t+1)} = \cdots = u_{(k)} = u_{(k-t)} + \delta' \qquad \text{(VI.2)}$$

and consider the limit as δ' approaches infinity. Let N_{NS} and N_S represent the ASN for the nonscreening and screening procedures

respectively for the problem of selecting the population with smallest mean where there are only k-t populations with

$$u_{(2)} = \cdots = u_{(k-t)} = u_{(1)} + \delta^* \qquad \text{(VI. 3)}$$

As δ' approaches infinity, ASN for the screening procedure converges to $N_S + t$, since only one observation will be taken from each of the populations $\pi_{(k-t+1)}, \ldots, \pi_{(k)}$. ASN for the nonscreening procedure converges to $(N_{NS})(\frac{k}{k-t})$. Therefore,

$$\lim_{\delta' \to \infty} \text{R.E.} = (1-f)\frac{N_S}{N_{NS}} + \frac{(1-f)t}{N_{NS}} \qquad \text{(VI. 4)}$$

where $f = t/k$ (i.e.: f is the number of inferior populations expressed as a fraction of k). The same expression for limiting relative efficiency holds if we consider configurations for which

$$u_{(k-t)} = \cdots = u_{(1)}$$

$$u_{(k-t+1)} = \cdots = u_{(k)} = u_{(k-t)} + \delta' \qquad \text{(VI. 5)}$$

as δ' approaches infinity. In this case N_S and N_{NS} are redefined as ASN for Problem N. b when there are k-t populations and

$$u_{(1)} = \cdots = u_{(k-t)} \qquad \text{(VI. 6)}$$

As f approaches unity, limiting R.E. approaches zero, illustrating the desirability of using a screening procedure when ASN is used as a measure of efficiency.

For GSC's of the type considered above, it is doubtful that an experimenter would continue sampling from all populations even if a nonscreening procedure had been adopted at the inception of the experiment. In this sense one would rarely use a true nonscreening procedure. Similarly, one would rarely use a procedure employing elimination in the strictest sense since one would not select as best a population which has been shown to be inferior. Thus one is led to consider modifications of both nonscreening procedures and procedures employing elimination.

In this part we propose a modification of the former.

Let $f^{(i)}_{x_j|S_j}(x, s)$ represent the conditional p.d.f. of the observations sampled during the jth stage given the observations sampled during the first j-1 stages, $\sum_{i \in S_j} P(x_i)$, and also given that the parametric configuration is an SC with π_i best. Define

$$L_{in} = \prod_{j=1}^{n} f^{(i)}_{x_j|S_j}(x, s) \quad . \tag{VI.7}$$

Let $\mathcal{K}$ be the sampling rule. The procedure proposed in this part is as follows:

Procedure K.D. A sequential screening procedure for selecting the best Koopman-Darmois population: "Sample in accordance with $\mathcal{K}$ until for some n

$$\frac{L_{(k)n}}{\sum_{j=1}^{k} L_{jn}} \geq P^* \quad . \tag{VI.8}$$

At this point discontinue sampling and select the population associated with $L_{(k)n}$ as best."

We note that, in fact, Procedure K.D. is a family of procedures for the Koopman Darmois family of densities.

In the next theorem we show that the test statistic given by Relation VI.8 depends on the unknown parameters only through the indifference constant θ^*, so that Procedure K.D. is always well defined.

Theorem VI.1: Let X_i $(i=1,\ldots,k)$ be independently distributed with p.d.f. given by VI.1. Suppose further that $Q(\theta_k) - Q(\theta_{k-1}) = \delta^*$ and $\theta_1 = \ldots = \theta_{k-1} = \theta$, and let $S = \sum_{i=1}^{k} P(X_i)$. Then the conditional distribution of $X_1, \ldots, X_k$ given S depends on the parameters (θ, θ_k) only through the indifference constant $\delta^* = Q(\theta_k) - Q(\theta)$.

Proof:

$$f_{X_1, \ldots, X_k}(x_1, \ldots, x_k)$$

$$= \exp\{S\,Q(\theta) + \delta^* P(x_k)$$

$$+ (k-1)T(\theta) + T(\theta_k) + \sum_{i=1}^{k} R(x_i)\} \qquad \text{(VI. 9)}$$

$$= c(x_1, \ldots, x_k;\ \delta^*)$$

$$\cdot \exp\{S\ Q(\theta) + (k-1)T(\theta) + T(\theta_k)\} \quad \text{(VI. 10)}$$

Using the factorization criterion (see p. 18-19 of Lehman (27)), Relation VI. 10 indicates that S is a sufficient statistic for (θ, θ_k), so that the conditional distribution of $X_1, \ldots, X_k$ given S does not depend on (θ, θ_k) (except through the indifference constant δ^*).

Procedure K. D. is identical to the nonscreening procedure of Bechhofer et al. (6) if no screening is permitted during the experiment. To demonstrate the equivalence we suppose that $\pi_{n_1}, \ldots, \pi_{n_r}$ are the unscreened populations during stage r and denote the corresponding observations by $(X_{n_1 r}, \ldots, X_{n_r r})$. Assuming an SC with π_i best, we define for $i=1, \ldots, k$, $f^{(i)}_{S_r}(s)$ to be the p.d.f. of $\sum_{j \in S_r} P(X_{jr})$ and define $f^{(i)}_{X_r}(x)$ to be the p.d.f. of $(X_{n_1 r}, \ldots, X_{n_r r})$. Recalling that $f^{(i)}_{X_r | S_r}(x, s)$ is defined as the conditional distribution of $(X_{n_1 r}, \ldots, X_{n_r r})$

given $\sum_{j=1}^{r} P(X_{n_j})$, we have that

$$f^{(i)}_{X_r|S_r}(x, s) = f^{(i)}_{X_r}(x)/f^{(i)}_{S_r}(s) \quad . \qquad \text{(VI. 11)}$$

If one does no screening $f^{(i)}_{S_r}(s)$ will not depend on $\underline{i}$. In this case, since Procedure K. D. is invariant under common change in scale of the functions $f^{(1)}_{X_r|S_r}, \ldots, f^{(k)}_{X_r|S_r}$ $(r=1, 2, \ldots)$ we may replace the definition of L_{in} given by Relation VI. with

$$L_{in} = \prod_{j=1}^{n} f^{(i)}_{X_j}(x) \quad . \qquad \text{(VI. 12)}$$

The procedure which is then obtained (assuming $\mathcal{K}$ calls for no screening) is identical to the nonscreening procedure proposed by Bechhofer et al. (6).

CHAPTER B. APPLICATIONS

Introduction

In this chapter we consider the specifications for the sequential screening procedure for selecting the best Koopman-Darmois population for the following distributions: gamma, normal, Poisson, Bernoulli, and negative binomial. In all cases the selection of the population with largest parameter is desired. In each case we shall specify

1. the density function of the observations: $f(x; \theta_i)$,
2. the indifference function: $Q(\theta_{(k)}) - Q(\theta_{(k-1)})$, and
3. the conditional density of the observations given the sum of the observations: $f_{X_j|S_j}(x, s)$.

These three items completely specify the problem and the sequential screening procedure. (For a listing of the functions P, Q, R, and T for each of the distributions considered in this chapter the reader is referred to page 67 of Bechhofer et al. (6).) As mentioned previously, Procedure K.D. is invariant under common change in scale of the functions

$$f^{(1)}_{X_j|S_j}, \ldots, f^{(k)}_{X_j|S_j}, \quad (j=1, 2, \ldots .)$$

Thus in order to keep the computations as simple as possible, we evaluate $f^{(i)}_{X_j|S_j}(x, s)$ only up to a multiplicative constant. We also point out that, for stages in which no screening is done, considerable computational simplification may often be achieved by appealing to this invariance property. The following notation is used

S_j = {i: π_i is not screened during stage j}

C_j = the cardinality of S_j

X_{ij} = the observation taken from π_i during stage j, for $i \in S_j$.

$X_{\cdot j} = \sum_{i \in S_j} X_{ij}$

$L_{i0} = 0 \qquad (i=1, \ldots, k)$

Selecting the best gamma population

We introduce the following notation:

$\chi^2(x, a)$ = the c.d.f. of a Chi-square variate with 2a degrees of freedom evaluated at x.

$$a_{1j} = \sum_{\ell=0}^{p-1} \binom{\ell+p(k-1)-1}{\ell} (-1)^{\ell} \cdot [\Gamma(p-1-\ell)]^{-1} X_{\cdot j}^{p-1-\ell} (\theta^*)^{-\ell-p(k-1)} \cdot \chi^2(2s\theta^*, \ell+p(k-1)) \qquad \text{(VI. 13)}$$

$$a_{2j} = \sum_{\ell=0}^{p-1} \binom{\ell+p(k-1)-1}{\ell} (-1)^{\ell} \cdot [\Gamma(p-1-\ell)]^{-1} X_{\cdot j}^{p-1-\ell} \chi^2(2X_{\cdot j}, \ell+p(k-1)). \qquad \text{(VI. 14)}$$

The specifications for selecting the best gamma population are:

1. $$f(x;\theta_i) = \frac{x^{p-1}e^{-x/\theta_i}}{\Gamma(p)\theta_i^p} \quad (p,\ \theta_i > 0)$$

2. $$\Omega(\theta_{(k)}) - \Omega(\theta_{(k-1)}) = \frac{1}{\theta_{(k-1)}} - \frac{1}{\theta_{(k)}}$$

3. $$f^{(i)}_{X_j|S_j}(x,\ s) = e^{\delta^* X_{ij}} a_{1j}^{-1} \quad \text{if } i \in S_j$$

$$= a_{2j}^{-1} \quad \text{if } i \notin S_j$$

To evaluate a_{1j} and a_{2j}, it is necessary for the experimenter to evaluate $\chi^2(2x,\ a)$ for various values of x and a. In cases where available tables of the χ^2 distribution prove inadequate, one may appeal to tables of the cumulative Poisson distribution, using the fact that

$$\chi^2(2x,\ a) = \Pr[Y \geq a] \tag{VI. 15}$$

where Y has a Poisson distribution with parameter x. Alternatively, for values of k and p such that $p(k-1) \geq 15$, the normal approximation

$$\chi^2(2x,\ a) = \Pr[Y \leq \sqrt{4x} - \sqrt{a-1}] \tag{VI. 16}$$

where $Y \sim N(0,\ 1)$ may prove useful.

The evaluation of $f^{(i)}_{X_j|S_j}(x,\ s)$ is based on the following theorem.

Theorem VI. 2: If $X_1, \ldots, X_k$ are independently distributed with the p.d.f. associated with X_i given by

$$f_i(x) = \frac{x^{p-1} e^{-x/\theta_i}}{\Gamma(p)\, \theta_i^p}, \quad x > 0, \tag{VI. 17}$$

and if

$$\frac{1}{\theta_1} = \ldots = \frac{1}{\theta_{k-1}} = \frac{1}{\theta_k} + \theta^* \tag{VI. 18}$$

then the conditional p.d.f. associated with $(X_1, \ldots, X_k)$ given $\sum_{i=1}^{k} X_i = s$ may be expressed

$$f_{X_1, \ldots, X_k | \Sigma X_i = s}(x_1, \ldots, x_k)$$

$$= \left(\prod_{i=1}^{k} x_i^p\right) e^{-\theta^*(x-x_k)} (r-1)$$

$$\cdot [\Gamma(r)]^{-k} \left\{ \sum_{\ell=0}^{p-1} \binom{\ell+p(k-1)-1}{\ell} (-1)^{\ell} \right.$$

$$[\Gamma(p-1-\ell)]^{-1} x_{\cdot j}^{p-1-\ell} (\theta^*)^{-\ell-p(k-1)}$$

$$\left. \cdot \chi^2(2s\theta^*, \ell+p(k-1)) \right\} . \tag{VI. 19}$$

Proof: $$f_{X_1, \ldots, X_k | \Sigma X_i = s}(x_1, \ldots, x_k)$$

$$= f_{X_1, \ldots, X_k}(x_1, \ldots, x_k) / f_{\Sigma X_i}(s) \quad . \qquad \text{(VI. 20)}$$

$$f_{\Sigma X_i}(s) = \int_0^s \{(s-x)^{p-1} e^{-\frac{s-x}{\theta_k}}$$

$$\cdot [\Gamma(p)]^{-1} \theta_k^{-p}\}$$

$$\cdot \{x^{p(k-1)-1} e^{-x/\theta_1} [\Gamma(p(k-1))]^{-1}$$

$$\cdot \theta_1^{p(k-1)}\} \, dx \qquad \text{(VI. 21)}$$

$$= A \int_0^s (s-x)^{p-1} x^{p(k-1)-1} e^{-x\theta^*} \, dx \qquad \text{(VI. 22)}$$

where

$$A = e^{-s/\theta_k} \{\Gamma(p) \Gamma[p(k-1)]$$

$$\cdot \theta_k^p \theta_1^{p(k-1)}\}^{-1} \quad . \qquad \text{(VI. 23)}$$

Thus, appealing to the binomial expansion,

$$f_{\Sigma X_i}(s) = A \cdot \int_0^s \sum_{\ell=0}^{p-1} \binom{p-1}{\ell} (-1)^{\ell} s^{p-1-\ell}$$

$$\cdot x^{\ell+p(k-1)-1} e^{-x\theta^*} dx \qquad \text{(VI. 24)}$$

$$= A \cdot \sum_{\ell=0}^{p-1} \binom{p-1}{\ell} (-1)^{\ell} s^{p-1-\ell}$$

$$\cdot (\theta^*)^{-\ell-p(k-1)} \Gamma[j+(k-1)p]$$

$$\cdot \int_0^s x^{\ell+p(k-1)-1} e^{-x\theta^*}$$

$$\cdot \{\Gamma[\ell+p(k-1)]\}^{-1} (\theta^*)^{-\ell-(k-1)p} dx \qquad \text{(VI. 25)}$$

$$= A \sum_{\ell=0}^{p-1} \binom{p-1}{\ell} (-1)^{\ell} s^{p-1-\ell} (\theta^*)^{-\ell-p(k-1)}$$

$$\cdot \Gamma[\ell+(k-1)p] \chi^2(2s\theta^*, \ell+(k-1)p) . \qquad \text{(VI. 26)}$$

The desired result is obtained by dividing $f_{X_1, \ldots, X_k}(x_1, \ldots, x_k)$ by the expression given in Relation VI. 26.

Selecting the best normal population

We define

$$\overline{X}_{.j} = X_{.j}/C_j \qquad \text{(VI. 27)}$$

the specifications for the problem of selecting the normal population with largest mean are listed below:

1. $$f(x, \theta_i) = \frac{1}{\sqrt{2\pi}} \exp - \frac{1}{2}(x-\theta_i)^2$$

2. $$\Omega(\theta_{(k)}) - \Omega(\theta_{(k-1)}) = \theta_{(k)} - \theta_{(k-1)}$$

3. $$f^{(i)}_{X_j|S_j}(x, s) = \exp \delta^* \{X_{ij} - \overline{X}_{.j} - \frac{\delta^*}{2}(1-C_j^{-1}) \quad \text{if } i \in S_j$$

$$= 1 \qquad \text{if } i \notin S_j \ .$$

The evaluation of $f^{(i)}_{X_j|S_j}(x, s)$ is based on the following theorem.

Theorem VI. 3: If $X_1, \ldots, X_k$ are independently distributed with the p.d.f. associated with X_i given by

$$f_i(x) = \frac{1}{\sqrt{2\pi}} \exp - \frac{1}{2}(x-\theta_i)^2 \qquad \text{(VI. 28)}$$

and if $$\theta_1 = \ldots = \theta_{k-1} = \theta_k - \delta^* \ , \qquad \text{(VI. 29)}$$

then the conditional p.d.f. associated with $(X_1, \ldots, X_k)$ given $\sum_{i=1}^{k} X_i = s$ may be expressed

$$f_{X_1, \ldots, X_k | \Sigma X_i = s}(x_1, \ldots, x_k)$$

$$= C[\exp \frac{1}{2}(\frac{s^2}{k} - \Sigma x_i^2)]$$

$$\cdot \exp \delta^*(x_k - \frac{s}{k} - \frac{1}{2}\delta^*(1 - \frac{1}{k})) , \qquad \text{(VI.30)}$$

where C is some constant which does not depend on $(x_1, \ldots, x_k, s, \delta^*, \theta_k)$.

Proof:

$$f_{X_1, \ldots, X_k}(x_1, \ldots, x_k) = C_1 \exp - \frac{1}{2} \sum_{i=1}^{k-1} (x_i - \theta_1)^2$$

$$- \frac{1}{2}(x_k - \theta_1 - \delta^*)^2 . \qquad \text{(VI.31)}$$

$$\sum_{i=1}^{k} X_i \sim N(k\theta_1 + \delta^*, k) \qquad \text{(VI.32)}$$

so that

$$f_{\Sigma X_i}(s) = C_2 \exp - \frac{1}{2k}(s - k\theta_1 - \delta^*)^2 . \qquad \text{(VI.33)}$$

As a result,

$$f_{X_1, \ldots, X_k | \Sigma X_i}(x_1, \ldots, x_k; s)$$

$$= f_{X_1, \ldots, X_k}(x_1, \ldots, x_k)/f_{\Sigma X_i}(s) \qquad \text{(VI. 33)}$$

$$= C_3 \exp - \frac{1}{2}\{(\sum_{i=1}^{k} x_i^2 - 2\theta_1 s + k\theta_1^2 + \delta^{*2}$$

$$+ 2\delta^* \theta_1 - 2\delta^* x_k) - \frac{1}{k}(s^2 + k^2\theta_1^2 + \delta^{*2}$$

$$- 2k s \theta_1 - 2s \delta^* + 2k \theta_1 \delta^*)\} \qquad \text{(VI. 34)}$$

$$= C_3[\exp \frac{1}{2}(\frac{s^2}{k} - \Sigma x_i^2)]$$

$$. \exp \delta^*(x_k - \frac{s}{k} - \frac{1}{2}\delta^*(1 - \frac{1}{k})) . \qquad \text{(VI. 35)}$$

Selecting the best Poisson population

1. $f(x, \theta_i) = e^{-\theta_i}(\theta_i)^x/x!$; $x=0, 1, 2, \ldots$

2. $Q(\theta_{(k)}) - Q(\theta_{(k-1)}) = \ell n\, \theta_{(k)} - \ell n\, \theta_{(k-1)}$

3. $f^{(i)}_{X_j | S_j}(x, s) = [C_j - 1 + 1/\theta^*]^{-X_{\cdot j}}$ if $i \in S_j$

$= (C_j)^{-X_{\cdot j}}$ if $i \notin S_j$

Specifying the indifference region in terms of $\ell n\, \theta_{(k)} - \ell n\, \theta_{(k-1)}$ is equivalent to specifying the indifference region in terms of $\theta_{(k)}/\theta_{(k-1)}$ since $\ell n\, \theta_{(k)} - \ell n\, \theta_{(k-1)} > \delta^*$ if and only if $\theta_{(k)}/\theta_{(k-1)} > e^{\delta^*}$.

The evaluation of $f^{(i)}_{X_j|S_j}(x, s)$ is based on the following theorem:

Theorem VI. 4: If $X_1, \ldots, X_k$ are independently distributed with the p.d.f. associated with X_i given by

$$f_i(x) = e^{-\theta_i}\, \theta_i^x / x! \quad , \tag{VI.36}$$

then the conditional p.d.f. associated with $(X_1, \ldots, X_k)$ given $\Sigma X_i = s$ may be expressed

$$f_{X_1, \ldots, X_k | \Sigma X_i = s}(x_1, \ldots, x_k) = (x_1^s, \ldots, x_k) \prod_{i=1}^{k} p_i^{x_i} \tag{VI.37}$$

where $p_i = \theta_i / \sum_j \theta_j$.

Proof:

$$f_{X_1, \ldots, X_k | \Sigma X_i = s}(x_1, \ldots, x_k) = f_{X_1, \ldots, X_k}(x_1, \ldots, x_k) / f_{\Sigma X_i}(s) \tag{VI.38}$$

$$= \prod_{i=1}^{k} [e^{-\theta_i} \theta_i^{x_i}/x_i!] \Big/ e^{-\Sigma\theta_j} (\Sigma\theta_j)^s/s! \qquad \text{(VI. 39)}$$

$$= \binom{s}{x_1, \ldots, x_k} \prod_{i=1}^{n} (\theta_i/\Sigma\theta_j)^{x_i} . \qquad \text{(VI. 40)}$$

Selecting the best Bernoulli population

1. $f(x, \theta_i) = \theta_i^x (1-\theta_i)^{1-x} \quad x = 0, 1$

2. $Q(\theta_{(k)}) - Q(\theta_{(k-1)}) = \ell n(\frac{\theta_{(k)}}{1-\theta_{(k)}}) - \ell n(\frac{\theta_{(k-1)}}{1-\theta_{(k-1)}})$

3. $f^{(i)}_{X_j|S_j}(x, s) = e^{\delta^* X_{ij}} [\binom{C_j - 1}{X_{.j} - 1} e^{\delta^*} + \binom{C_j - 1}{X_{.j}}]^{-1} \quad \text{if } i \in S_j$

$$= [\binom{C_j - 1}{X_{.j} - 1} + \binom{C_j - 1}{X_{.j}}]^{-1} \quad \text{if } i \notin S_j .$$

Specifying the indifference region in terms of

$$\ell n(\frac{\theta_{(k)}}{1-\theta_{(k)}} - \ell n(\frac{\theta_{(k-1)}}{1-\theta_{(k-1)}}) \qquad \text{(VI. 41)}$$

is equivalent to specifying the indifference region in terms of

$$\frac{\theta_{(k)}}{1-\theta_{(k)}} / \frac{\theta_{(k-1)}}{1-\theta_{(k-1)}} \qquad \text{(VI. 42)}$$

since

$$\ell n\left(\frac{\theta_{(k)}}{1-\theta_{(k)}}\right) - \ell n \frac{\theta_{(k-1)}}{1-\theta_{(k-1)}} \geq \delta^* \qquad \text{(VI. 43)}$$

if and only if

$$\frac{\theta_{(k)}}{1-\theta_{(k)}} \Big/ \frac{\theta_{(k-1)}}{1-\theta_{(k-1)}} \geq e^{\delta^*} . \qquad \text{(VI. 44)}$$

The evaluation of $f^{(i)}_{X_j|S_j}(x, s)$ is based on the following theorem:

Theorem VI. 5: If $X_1, \ldots, X_k$ are independently distributed with the p. d. f. associated with X_i given by

$$f_i(x) = \theta_i^x (1-\theta_i)^{1-x} \qquad \text{(VI. 45)}$$

and if

$$\theta_1 = \ldots = \theta_{k-1} = \theta$$

$$\frac{\theta_k}{1-\theta_k} \Big/ \frac{\theta}{1-\theta} = \theta^* , \qquad \text{(VI. 46)}$$

then the conditional p. d. f. associated with $X_1, \ldots, X_k$ given $\sum_{i=1}^{k} X_i = s$ may be expressed then

$$\Pr[X_i = x_i,\ i=1,\ldots,k \mid \sum_{i=1}^{k} X_i = s]$$

$$= (\theta^*)^{x_k}[\binom{k-1}{s-1}\theta^* + \binom{k-1}{s}]^{-1} \quad . \qquad \text{(VI. 47)}$$

Proof: $\Pr[X_i = x_i,\ i=1,\ldots,k]$

$$= \theta^{s-x_k}\,\theta_k^{x_k}(1-\theta)^{k-1-s+x_k}$$

$$\cdot\ (1-\theta_k)^{1-x_k} \qquad \text{(VI. 48)}$$

$$\Pr[\sum_{i=1}^{k} X_i = s] = \binom{k-1}{s-1}\theta^{s-1}(1-\theta)^{k-1}\theta_k$$

$$+ \binom{k-1}{s}\theta^{s}(1-\theta)^{k-1-s}(1-\theta_k) \qquad \text{(VI. 49)}$$

The conclusion of the theorem is obtained by evaluating

$$\frac{\Pr[X_i = x_i,\ i=1,\ldots,k]}{\Pr[\sum_{i=1}^{k} X_i = s]} \qquad \text{(VI. 50)}$$

Selecting the best negative binomial population

1. $$f(x, \theta_i) = \binom{r-1}{x} \theta_i^r (1-\theta_i)^x, \quad x=0, 1, 2, \ldots$$

2. $$Q(\theta_{(k)}) - Q(\theta_{(k-1)}) = \ell n(1-\theta_{(k)}) - \ell n(1-\theta_{(k-1)})$$

3. $$f^{(i)}_{X_j|S_j}(x, s) = e^{X_{ij}\delta^*} a_{1j}^{-1} \qquad \text{if } i \in S_j$$

$$= a_{2j}^{-1} \qquad \text{if } i \notin S_j$$

where

$$a_{1j} = \sum_{x=0}^{X_{.j}} \binom{r(k-1)+x-1}{x} \binom{r+X_{.j}-x-1}{X_{.j}-x} e^{\delta^*(X_{.j}-x)} \qquad \text{(VI.51)}$$

$$a_{2j} = \sum_{x=0}^{X_{.j}} \binom{r(k-1)+x-1}{x} \binom{r+X_{.j}-x-1}{X_{.j}-x} \qquad \text{(VI.52)}$$

Again, we note that specifying the indifference region to consist of parameter points $(\theta_1, \ldots, \theta_k)$ such that

$$\ell n(1-\theta_{(k)}) - \ell n(1-\theta_{(k-1)}) > \delta^* \qquad \text{(VI.53)}$$

is equivalent to specifying the indifference region to consist of points such that

$$\theta_{(k)}/\theta_{(k-1)} \geq e^{\delta^*} \overset{\text{def.}}{=} \theta^* \text{ , say} \qquad \text{(VI. 54)}$$

Note, since δ^* is negative, $\theta^* < 1$.

The computations required for the procedure of this section are relatively simple if $r=1$ (i.e. if we are selecting from among geometric populations). In this case,

$$a_{1j} = \{ \sum_{x=0}^{X_{.j}} \binom{k-2+x}{x} (\theta^*)^{X_{.j}-x} \}^{-1} \qquad \text{(VI. 55)}$$

$$a_{2j} = \{ \sum_{x=0}^{X_{.j}} \binom{k-2+x}{x} \}^{-1} \qquad \text{(VI. 56)}$$

The evaluation of a_{1j} may be further simplified by using the relation

$$\sum_{x=0}^{X_{.j}} \binom{C_j-2+x}{x} (\theta^*)^x = (1-\theta^*)^{-1} \sum_{x=C_j-2}^{n} \binom{n}{x} (1-\theta^*)^x (\theta^*)^{n-x} , \qquad \text{(VI. 57)}$$

where $n = X_{.j} + C_j - 2$. This relation enables the use of tables of the cumulative binomial distribution in evaluating a_{1j} .

The evaluation of $f^{(i)}_{X_j|S_j}(x, s)$ is obtained from the following theorem:

Theorem VI.6: If $X_1, \ldots, X_k$ are independently distributed with

$$\Pr[X_i = x] = \theta_i^r (1-\theta_i)^x$$

$$x=0, 1, 2, \ldots \qquad \text{(VI.58)}$$

and if

$$\theta_1 = \ldots = \theta_{k-1} = \theta$$

$$\theta_k > \theta^* \qquad \text{(VI.59)}$$

then,

$$\Pr[X_i = x_i,\ i=1, \ldots, k \,|\, \Sigma X_i = s]$$

$$= \left(\frac{1-\theta_k}{1-\theta}\right)^{x_k} \prod_{i=1}^{k} \binom{r+x_i-1}{x_i} \Big/ \sum_{x=0}^{s} \binom{r(k-1)+x-1}{x}\binom{r+s-x-1}{s-x}\left(\frac{1-\theta_k}{1-\theta}\right)^{s-x}. \qquad \text{(VI.60)}$$

Proof:

$$\Pr\left[\sum_{i=1}^{k} X_i = s\right] = \sum_{x=0}^{s} \Pr\left[\sum_{i=1}^{k-1} X_i = x\right] \cdot \Pr[X_k = s-x]$$

$$= \sum_{x=0}^{s} \left\{\binom{r(k-1)+x-1}{x}\theta^{r(k-1)}(1-\theta)^x\right\}$$

$$\cdot \left\{\binom{r+s-x-1}{s-x}\theta_k^r(1-\theta_k)^{s-x}\right\} \qquad \text{(VI.61)}$$

Also

$$\Pr[X_i = x_i,\ i=1,\ldots,k] = \prod_{i=1}^{k} \binom{r+x_i-1}{x_i} \theta_i^r (1-\theta_i)^{x_i} \qquad \text{(VI.62)}$$

As a result,

$$\Pr[X_i = x_i,\ i=1,\ldots,k \mid \sum_{i=1}^{k} X_i = s]$$

$$= \frac{(1-\theta)^{s-x_k}(1-\theta_k)^{x_k} \prod_{i=1}^{k} \binom{r+x_i-1}{x_i}}{\sum_{x=0}^{s} \binom{r(k-1)+x-1}{x}\binom{r+s-x-1}{s-x}(1-\theta)^x(1-\theta_k)^{s-x}} \qquad \text{(VI.63)}$$

$$= \frac{\left(\frac{1-\theta_k}{1-\theta}\right)^{x_k} \prod_{i=1}^{k} \binom{r+x_i-1}{x_i}}{\sum_{x=0}^{s} \binom{r(k-1)+x-1}{x}\binom{r+s-x-1}{s-x}\left(\frac{1-\theta_k}{1-\theta}\right)^{s-x}} \qquad \text{(VI.64)}$$

CHAPTER C. DERIVATION

Introduction

In general when applying Theorem II. 5 to distributions other than the multinomial distribution there are infinitely many parametric configurations corresponding to the hypothesis "the parametric configuration is an SC with π_i best." This is not cause for concern if one is willing to take an equal number of observations from each population: in this case the test statistic of Theorem II. 5 depends on the unknown parameters only through the indifference constant. Unfortunately, if one samples unequal numbers of observations from each population, then the test statistic of Theorem II. 5 will generally depend on the unknown parameters. This problem may be circumvented by using conditional p. d. f. 's in computing the test statistic, as is done in Procedure K. D.

In the following section it will be verified that if the sampling rule $\mathcal{K}$ does not depend on the observations and the parametric configuration is an SC, than PRCS $\geq$ P* for Procedure K. D. We do not, however, propose that one adopt a sampling rule which is independent of the observations. One reason is that such a rule may increase ASN. A more serious difficulty is that screening conducted independently of the observations may cause GSC's of the type considered in Theorems II. 7. a and II. 7. b to be considerably less favorable than SC's. We illustrate with Procedure K. D. for selecting the normal population with largest mean. Suppose the sampling rule calls for screening during stage one a single population chosen completely at random. If the parametric configuration

is given by

$$\theta_1 = \ldots = \theta_{k-2} = \theta$$

$$\theta_{k-1} = \theta_k - \delta^* \qquad (\delta^* > 0),$$

then

$$\lim_{\theta \to -\infty} \text{PRCS} < 1 - \frac{1}{k},$$

since (for θ sufficiently small) inspection of the test statistic indicates that screening π_k during stage 1 would be followed by selection of π_{k-1} upon completion of the first stage of sampling. Conditions on a sampling rule $\mathcal{K}$ which are sufficient to insure that specified levels of PRCS are maintained are given by the following:

1. there exists a sampling rule $\mathcal{K}'$ which does not depend on the observations and which is such that when the parametric configuration is an SC, PRCS under $\mathcal{K}'$ is not greater than PRCS under $\mathcal{K}$, and
2. an SC is an LFC.

These conditions are sufficient because (1) guarantees that PRCS $\geq P^*$ when $\mathcal{K}$ is used and the parametric configuration is an SC while (2) guarantees that PRCS is minimized outside the indifference region by an SC when $\mathcal{K}$ is used.

It appears from the numerical results listed in Table V. 1 that specified levels of PRCS will be maintained for Procedure K. D. when sampling rule 1 is used, but this has not been proven.

Derivation

In this section we show that if $\mathcal{K}$ is invariant under permutation of the indexing of the populations and does not depend on the observations, then Procedure K.D. is such that

$$PRCS \geq P^*(1-\beta) \qquad \text{(VI.65)}$$

where β is the probability that Procedure K.D. does not terminate. The following notation is used:

$\mathcal{K}$ the rules stating which populations are to be sampled during each stage n $(n=1, 2, \ldots)$.

X_n A vector valued random variable composed of the first n (k dimensional) vectors of observations. Some components in each k dimensional vector may be empty. A non-empty component is associated with a sampled population.

S_n the sum of the observations taken during stage n $(n=1, 2, \ldots)$.

s_n A realization of S_n

$\mathcal{B}_n$ the sample space associated with the collection of random variables $(S_1, S_2, \ldots, S_n)$, $(n=1, 2, \ldots, \infty)$.

s_n^* $s_n^* = (s_1, \ldots, s_n)$, a point in $\mathcal{B}_n$.

$\chi(s_n^*)$ the set of X_n which are realizable under $\mathcal{K}$, given that $S_1, \ldots, S_n = s_n^*$.

$D_i(s_n^*)$ That part of $\chi(s_n^*)$ in which π_i is selected.

$f_i(X_n \mid s_n^*)$ The p.d.f. associated with X_n given that $S_1, \ldots, S_n = s_n^*$ and that the parametric configuration is an SC with π_i best.

$\beta(s_\infty^*)$ The probability that the procedure of Theorem VI.7 does not terminate given $(S_1, S_2, \ldots) = s_\infty^*$.

β = $E(\beta(s^*_\infty))$, the unconditional probability that the procedure of Theorem VI. 7 does not terminate.

We also require a symmetry condition analogous to the symmetry condition referred to in Theorem II. 5. The definition is obtained by replacing χ_n, D_{in}, and $f_{in}(X_n)$ in the statement in Part II with $\chi(s^*_n)$, $D_i(s^*_n)$, and $f_i(X_n | s^*_n)$ respectively.

Theorem VI. 7:

Conditions: $\mathcal{K}$ is invariant under permutation of the indexing of the populations and does not depend on the observations. The $f_i(X_n | s^*_n)$ are such that the regions $D_i(s^*_n)$ satisfy the symmetry condition. Observations are taken in accordance with the rules set forth in $\mathcal{K}$ until for some n and t

$$\frac{f_t(X_n | s^*_n)}{\sum_{\substack{\ell=1 \\ \neq t}}^{k} f_\ell(X_n | s^*_n)} \geq \frac{P^*}{1-P^*} \overset{\text{def.}}{=} A . \qquad \text{(VI.66)}$$

At this point the experiment is terminated and the population t is selected as best. We also assume that the parametric configuration is given by a slippage configuration.

Conclusion: $PRCS \geq P^*(1-\beta)$.

Proof: Given that $(S_1, S_2, \ldots) = s^*_\infty$, it follows from Theorem II. 5 that

$$PRCS \geq P^*(1-\beta(s^*_\infty)) \; . \tag{VI.67}$$

Therefore, unconditionally,

$$PRCS \geq P^*[1-E(\beta(s^*_\infty))] \tag{VI.68}$$

$$= P^*(1-\beta) \; . \tag{VI.69}$$

PART VII. ADDITIONAL SELECTION PROBLEMS AND PROCEDURES

CHAPTER A. SELECTING THE NEGATIVE BINOMIAL POPULATION WITH LARGEST PROBABILITY

Introduction

In this chapter we propose procedures for selecting the negative binomial population with largest probability. Procedures for selecting the population with smallest probability may be easily derived by analogy. This problem was discussed in Part VI. The notation that we use for the negative binomial distribution is (as in Part VI):

$$\Pr[X = x] = \binom{r-1+x}{x}\theta^{r}(1-\theta)^{x-r}$$

$$x=0, 1, 2, \ldots; \quad r > 0. \quad \text{(VII. 1)}$$

In this part we consider definitions of the indifference region other than those considered in Part VI. The specifications of the indifference region which we consider here are

1. $$\theta_{(k)}/\theta_{(k-1)} \geq \theta^{*}, \quad \text{and} \quad \text{(VII. 2)}$$

2. $$\theta_{(k)} - \theta_{(k-1)} \geq \delta^{*}. \quad \text{(VII. 3)}$$

For the first case where the indifference region is specified as a ratio we propose the following single sample procedure, which is recommended on the basis of its administrative simplicity.

Procedure NB.a.2. A single sample procedure for selecting the negative binomial population with largest probability: "Take n observations from each population and select the population corresponding to the smallest observed sum as best. Ties may be broken using a random device."

For Procedure NB.a.2, n may be chosen to maintain specified levels of PRCS by using tables in Bechhofer and Sobel (7), Gupta and Sobel (22), and Gupta (17) for the single sample procedure for selecting the gamma population with smallest scale parameter (using the values of k, P^*, and θ^* as specified in the negative binomial problem).

For the second case, in which the indifference function is specified as a difference, we propose a sequential nonscreening procedure. Let

X_{ij} = the observation taken from π_i during the jth stage of sampling.

Z_{ij} = a random variable obtained by sampling from a gamma population with scale parameter $X_{ij}+1$ and shape parameter r, where r is the known parameter of the negative binomial populations.

$$Y_{ij} = \sum_{\ell=1}^{j} X_{i\ell}$$

$$W_{ij} = \sum_{\ell=1}^{j} Z_{i\ell}$$

Procedure NB.a.3. A sequential nonscreening procedure for selecting the negative binomial population with largest probability: "Sample k dimensional vectors of negative binomial observations sequentially. After each vector $(X_{1j}, \ldots, X_{kj})$ is sampled, sample Z_{ij} $(j=1, \ldots, k)$ from a gamma population with parameter r and scale parameter $X_{ij}+1$. When for some n

$$\sum_{\ell=1}^{k} \exp\{(W_{(1)n} - W_{\ell n})\theta^*\} \leq \frac{1}{P^*} \qquad \text{(VII.4)}$$

where $W_{(1)n}$ is the minimum over ℓ of $W_{\ell n}$ stop and select the population associated with $\min_i\{Y_{in}\}$ as best."

CHAPTER B. DERIVATION OF THE SINGLE SAMPLE AND SEQUENTIAL PROCEDURES

Introduction

The procedures of Chapter A are based on the following theorems. Theorem VII. 1 is due to Robbins and Pitman (39).

Theorem VII. 1: If Y has a negative binomial distribution with shape parameter r and probability θ and, for fixed Y, $Z \sim \Gamma(r+Y;\ 1)$. Then unconditionally $Z \sim \Gamma(r;\ 1/\theta)$.

The statement of Theorem VII. 2 (due to Bahadur and Goodman (2)) is taken from Lehman (27).

Theorem VII. 2: Let the distribution of the sufficient statistics $T = (T_1, \ldots, T_k)$ have density

$$h_\theta(t) = e(\theta) f_{\theta_1}(t_1) \ldots f_{\theta_k}(t_k) \qquad \text{(VII. 5)}$$

w. r. t. a σ-finite measure v which is invariant under the group G of all permutations of $(t_1, \ldots, t_k)$. For any permutation g of $(t_1, \ldots, t_k)$ define $\bar{g}$ and g^* as the same permutation of $(\theta_1, \ldots, \theta_k)$ and $(d_1, \ldots, d_k)$ (where d_i represents the decision to select π_i) respectively, and suppose that the loss function λ satisfies

1. $\lambda(\bar{g}_i\theta,\ g_i^* d_j) = \lambda(\theta,\ d_j)$ for any i, j, and θ, and (VII. 6)

2. $\theta_i < \theta_j = \lambda_i(\theta) \geq \lambda_j(\theta)$. (VII. 7)

Let $\sigma^{(0)}$ be the procedure which takes decision d_i when t_i is the unique largest among $(t_1, \ldots, t_k)$ and which takes decisions $d_{i_1}, \ldots, d_{i_r}$ each with probability $\frac{1}{r}$ if $(t_{i_1}, \ldots, t_{i_r})$ is the set of t values equal to $\max t_j$ (i.e.: which breaks ties at random). Then $\sigma^{(0)}$ uniformly minimizes the risk among all procedures based on t which are invariant under G .

To apply Theorem VII. 2 to the problem of selecting the best negative binomial population we define

$$X_{ij} = \text{the jth observation from } \pi_i \quad (j=1, \ldots, n; \; i=1, \ldots, k)$$

$$T_i = \sum_{j=1}^{n} X_{ij}$$

$$f_{\theta_i}(t_i) = \binom{nr+t_i-1}{t_i} \theta_i^{nr} (1-\theta_i)^{t_i} \qquad t_i=0, 1, 2, \ldots$$

$$e(\theta) = 1$$

$$v = \text{counting measure}$$

$$\lambda(\theta_i) = 1 \qquad \text{if } \theta_i \neq \theta_{(k)}$$

$$= 0 \qquad \text{if } \theta_i = \theta_{(k)} .$$

Derivation of the single sample procedure

It is desired to show that if one uses the tables for the single sample procedure for selecting the gamma population with smallest scale parameter (as proposed in Chapter A) to determine the sample number n to be taken from each population, specified levels of PRCS will be maintained. We note first that if one samples gamma random variables using the sampling scheme described in Theorem VII. 1 and then uses the single sample procedure (based on n observations from each population) to select the (artificially created) gamma population with smallest scale parameter, specified levels of PRCS will be maintained. Since there is a one to one correspondence between the parameters of the artificially created gamma populations and the parameters of the original negative binomial populations, specified levels of PRCS are maintained if one selects the negative binomial population corresponding to the selected gamma population as being the best negative binomial population. This method of selection is not recommended, but we note that specified levels of PRCS are maintained if it is used. That the single sample procedure of Chapter A maintains specified levels of PRCS follows from Theorem VII. 2 which states that, based on a fixed and equal sample size from each population, the single sample procedure of Chapter A maximizes the PRCS over all other (single sample) procedures which are invariant under permutation of the indexing of the populations.

We next demonstrate that the method of Chapter A which is based on gamma random variables maintains the same level of PRCS for the slippage configuration as the single sample gamma procedure upon which it

is based. We require the following notation:

$P_M(\theta, R)$ = PRCS for the inverse sampling procedure used to select the multinomial category with largest probability when the parametric configuration is given by

$$p_k = \theta p_i \qquad i=1,\ldots,k-1 \qquad (\theta > 1) \qquad \text{(VII. 8)}$$

and a stopping bound equal to R is used.

$P_G(\theta, R)$ = PRCS for the single sample procedure used to select the exponential population with smallest parameter when the parametric configuration is given by

$$\theta_k = \theta_i/\theta \qquad i=1,\ldots,k-1 \qquad (\theta < 1) \qquad \text{(VII. 9)}$$

and R observations are sampled from each population.

$P_{NB}(\theta, R)$ = PRCS for the single sample procedure used to select the negative binomial population with largest probability when the parametric configuration is given by

$$\theta_k = \theta_i/\theta \qquad i=1,\ldots,k-1 \qquad (\theta < 1) \qquad \text{(VII. 10)}$$

sampled from each population.

When the parametric configuration is given by VII. 10 we define

$$p = \theta_i \qquad (i=1,\ldots,k-1), \quad \text{and}$$
$$q = 1-p \qquad \text{(VII. 11)}$$

We note that due to the reproductive properties of the gamma and negative binomial distributions it suffices to consider the case of unit shape parameter when discussing single sample procedures for these populations. For example, sampling N observations from a negative binomial population with shape parameter equal to r is equivalent (as regards the single sample procedure) to sampling N. r observations from a geometric population (assuming N. r is integer valued). We consider the following selection procedures:

Procedure B. a. 1. An inverse sampling procedure for selecting the Bernoulli process with largest probability: "Sampling k dimensional vectors of (Bernoulli) observations sequentially, select the first process to yield R successes. Ties may be broken using a random device."

Procedure B. a. 2. An inverse sampling procedure for selecting the Bernoulli process with largest probability: "Proceed with the specifications of Procedure B. a. 1 except that only vectors in which exactly one success is observed are considered."

We define

$P_{B.1}(\theta, R)$ = PRCS when Procedure B.a.1 is used to select the Bernoulli process with largest probability when the parametric configuration is given by

$$\Pr[\text{success}] = p \qquad i=1, 2, \ldots, k-1$$

$$= \theta^* p \qquad i=k,\ \theta^* > 1 .$$

and a stopping bound equal to R is used.

$P_{B.2}(\theta, R)$ = PRCS for Procedure B.a.2 with the same stopping bound and parametric configuration.

Theorem VII.3: If the parametric configuration is an SC, with $\theta_k = \theta^* p$, $\theta_i = p$, $i \neq k$ and $\theta^* > 1$ then

$$P_M(\theta^*, R) = P_G(1/\theta^*, R) = \lim_{p \to 0} P_{NB}(1/\theta^*, R) . \quad \text{(VII.12)}$$

To prove Theorem VII.3 it will be useful to suppose that the geometric random variables arise from Bernoulli processes. That is, we suppose that the nth observation from $\pi_{(i)}$ represents the number of failures occurring after the n-1st success but prior to the nth success, where on any trial from $\pi_{(i)}$

$$\Pr[\text{success}] = p \qquad i=1, \ldots, k-1$$

$$= \theta^* p \qquad i=k \qquad (\theta^* > 1) . \quad \text{(VII.13)}$$

The proof of Theorem VII. 3 will consist of four steps. We shall show:

(1) $$P_M(\theta^*, R) = P_G(1/\theta^*, R)$$

(2) $$\lim_{p \to 0} P_{B.2}(1/\theta^*, R) = P_M(\theta^*, R)$$

(3) $$\lim_{p \to 0} P_{B.1}(1/\theta^*, R) = \lim_{p \to 0} P_{B.2}(1/\theta^*, R)$$

(4) $$P_{NB}(1/\theta^*, R) = P_{B.1}(1/\theta^*, R) (0 < p < 1)$$

where θ^* is assumed greater than unity in all cases.

Step 1: This result was established by Theorem IV. 2.

Step 2: Let N_{in} represent the number of trials required to yield the next success in process i after trial n ($i=1, \ldots, k$; $n=0, 1, 2, \ldots$). Then

$$\Pr[N_{kn} < \text{Min}(N_{1n}, \ldots, N_{k-1,n}) \mid N_{in} \neq N_{jn} \text{ for } i, j=1, \ldots, k \ (i \neq j)]$$

$$= \frac{\theta^* p q^{k-1}}{\theta^* p q^{k-1} + (k-1)p(1-\theta^* p)q^{k-2}} \qquad \text{(VII. 14)}$$

$$= \left(1 + \frac{k-1}{\theta^* q}(1-\theta^* p)\right)^{-1} . \qquad \text{(VII. 15)}$$

Also

$$\Pr[N_{tn} < \text{Min}\{N_{sn}: s \neq t \text{ or } k\} \mid N_{in} \neq N_{jn} \text{ for } i, j=1, \ldots, k \ (i \neq j)]$$

$$= \frac{pq^{k-2}(1-\theta^*p)}{\theta^*p\,q^{k-1} + (k-1)p(1-\theta^*p)q^{k-2}} \qquad \text{(VII. 16)}$$

$$= \left(\frac{\theta^*q}{1-\theta^*p} + k - 1\right)^{-1} . \qquad \text{(VII. 17)}$$

It follows that

$$P_{B.2}(1/\theta^*, R) = P_M\left(\frac{\theta^*q}{1-\theta^*p}, R\right) \qquad \text{(VII. 18)}$$

the conclusion of Step 2 follows from the fact that

$$\lim_{p \to 0} \frac{\theta^*q}{1-\theta^*p} = \theta^* \qquad \text{(VII. 19)}$$

Step 3: The proof of the conclusion of Step 3 consists of noting that (1) the sampling and selection rules for both Procedures B.a.1 and B.a.2 are identical for experiments in which two or more successes are never observed during the same stage of sampling, and (2) the probability that two or more successes are ever observed during the same stage of

sampling using Procedure B.a.1 or Procedure B.a.2 converges to zero as p approaches zero.

Step 4: The conclusion of Step 4 follows from the specification that the geometric random variables were created from Bernoulli processes.

Theorem VII.3 demonstrates the existence of a limiting slippage configuration such that the probability of correct selection for the negative binomial problem is the same as the PRCS for the gamma problem with slippage configuration. Hence, if the gamma procedure specifies the minimum sample number for a given PRCS this sample number is also a minimum for the negative binomial problem.

Derivation of the sequential procedure

The derivation of the sequential nonscreening procedure of Chapter A is analogous to the derivation of the single sample procedure. However instead of appealing to the single sample procedure for selecting the gamma population with smallest scale parameter one appeals to the sequential nonscreening procedure of Bechhofer et al. (6) for selecting the gamma population with smallest scale parameter. Let R_{NB} be the selection rule (based on negative binomial random variables) of Procedure NB.a.3, let R_G be the selection rule that choses the negative binomial on the basis of the artificially created gamma random variables, and let N be the number of stages required for termination of Procedure NB.a.3. (Note that N is a random variable and is a function of the gamma random variables). In the following theorem we assume the experimenter uses the sampling rule of Procedure NB.a.3.

Theorem VII.4: For any $n = 1, 2, \ldots$, given that $N = n$, PRCS using R_{NB} is greater than or equal to PRCS using Procedure R_G.

We first note that the theorem is almost obvious. Two situations are possible. i) The population with the smallest negative binomial sum also has the smallest sum for the artificially created gamma random variables. In this case the same decision is taken by both rules. ii) The artificial random variables for the population with smallest negative binomial sum exceeds at least one of the other sums of gamma random variables. In this case the decision based on R_G is contrary to that specified by Theorem VII. 2.

To prove Theorem VII. 4, we shall show that, if the theorem were not true for $N = n_0$ then the selection rule of the single sample procedure proposed in Chapter A would not maximize PRCS over all other single sample procedures (contradicting Theorem VII. 2). Consider the following single sample procedure for selecting the best negative binomial population.

Procedure NB. a. 2': "Sequentially sample k-dimensional vectors of observations one at a time. After each vector is sampled, sample a vector of gamma random variables in the manner indicated by Theorem VII. 1. After each vector is sampled check to see if Relation VII. 4 is satisfied. Continue until n_0 vectors of gamma random variables have been obtained. Let E be the event that Relation VII. 4 is satisfied upon (but not prior to) sampling the n_0th vector of gamma random variables. If E occurs select the negative binomial population corresponding to the (artificially created) gamma population having

smallest observed sum. If E does not occur select the negative binomial population from which the smallest sum (of negative binomial random variables) is obtained. Let

P_1' = PRCS using Procedure NB.a.2' given that E occurred

P_1 = PRCS using Procedure NB.a.2 given that E occurred

P_2 = PRCS (using either procedure) given that E did not occur

P = the probability that E occurs.

$PRCS_1$ = PRCS using Procedure Nb.2'

$PRCS_2$ = PRCS using Procedure Nb.2 .

Then,

$$PRCS_1 = P.P_1' + (1-P)P_2 \qquad \text{(VII.20)}$$

$$PRCS_2 = P.P_1 + (1-P)P_2 \ . \qquad \text{(VII.21)}$$

Under the assumption that Theorem VII.3 does not hold for $N = n_0$, $P_1' > P_1$ so that

$$PRCS_1 - PRCS_2 = P(P_1' - P_1) > 0 \qquad \text{(VII. 22)}$$

which contradicts Theorem VII. 2. This proves Theorem VII. 4, from which it follows immediately that Procedure NB. a. 3 maintains specified levels of PRCS.

Evaluation of the error involved in determining sample size

It would be desirable to evaluate analytically the extent to which PRCS is understated when tables appropriate for gamma populations are used to obtain the sample number for the single sample procedure for selecting the best negative binomial population. Unfortunately this would necessitate evaluating PRCS exactly, and it is this difficulty that led to using the approximate method. However, the expression for PRCS when $nr = 1$ is relatively uncomplicated. We consider first the problem of selecting the population with largest probability.

Theorem VII. 5. a: Assume that the single sample procedure based on n observations from each of k negative binomial populations is used to select the population with largest probability.

Also assume that the parametric configuration is of the form

$$\theta_i = p \qquad i=1, \ldots, k-1$$

$$\theta_k = \theta^* p \qquad (\theta^* > 1) \; . \qquad \text{(VII. 23)}$$

If $n.r = 1$,

$$PRCS = \frac{\theta^*(1-q^k)}{k-[1-(1-\theta^*p)q^{k-1}]} \qquad \text{(VII. 24)}$$

Proof:

$$PRCS = \sum_{n=0}^{\infty} \theta^* p(1-\theta^*)^n$$

$$\cdot \sum_{j=1}^{k} \frac{1}{j}\binom{k-1}{j-1} p^{j-1} q^{n(j-1)}$$

$$\cdot (q^{n+1})^{k-j} \qquad \text{(VII. 25)}$$

$$= \sum_{j=1}^{k} \theta^* p^j q^{k-j} \left(\frac{1}{j}\right)\binom{k-1}{j-1}$$

$$\cdot \sum_{n=0}^{\infty} [(1-\theta^*p)q^{k-1}]^n \qquad \text{(VII. 26)}$$

$$= \frac{\theta^*}{1-(1-\theta^*p)q^{k-1}} \sum_{j=1}^{k} \binom{k-1}{j-1} p^j q^{k-j} \frac{1}{j} \qquad \text{(VII. 27)}$$

$$= \frac{\theta^*}{k[1-(1-\theta^*p)q^{k-1}]} \sum_{j=1}^{k} \binom{k}{j} p^j q^{k-j} \qquad \text{(VII. 28)}$$

$$= \frac{\theta^*(1-q^k)}{k[1-(1-\theta^*p)q^{k-1}]} . \qquad \text{(VII. 29)}$$

The value of PRCS obtained from tables for selecting the best gamma population is given by

$$\hat{PRCS} = \frac{\theta^*}{k-1+\theta^*} \quad . \tag{VII.30}$$

Theorem VII.5.a indicates that the true PRCS is close to the estimated level of PRCS for SC's in which $p_{(k)}$ is close to zero. In Table VII.1 we compare the true with the estimated levels of PRCS for the situation n.r = 1 .

Table VII. 1. Comparison of true versus estimated PRCS for selecting the population with largest probability (n. r = 1)

k	θ*	p	PRCS Estimated	PRCS True
3	1.2	.001	.37500	.37503
3	1.2	.01		.37524
3	1.2	.1		.37744
3	1.2	.2		.38006
3	1.2	.4		.38583
3	1.6	.001	.44444	.44452
3	1.6	.01		.44519
3	1.6	.1		.45223
3	1.6	.2		.46081
3	1.6	.4		.48039
3	2.0	.001	.50000	.500131
3	2.0	.01		.50126
3	2.0	.1		.51326
3	2.0	.2		.52814
3	2.0	.4		.56322
6	1.2	.001	.19355	.19357
6	1.2	.01		.19370
6	1.2	.1		.19508
6	1.2	.2		.19651
6	1.2	.4		.19870
6	1.6	.001	.24242	.24245
6	1.6	.01		.24298
6	1.6	.1		.24792
6	1.6	.2		.25317
6	1.6	.4		.26155
6	2.0	.001	.28571	.28582
6	2.0	.01		.28674
6	2.0	.1		.29603
6	2.0	.2		.30614
6	2.0	.4		.32280

Table VII. 1. (Continued)

k	θ*	p	PRCS Estimated	PRCS True
9	1.2	.001	.13043	.13045
9	1.2	.01		.13055
9	1.2	.1		.13149
9	1.2	.2		.13231
9	1.2	.4		.13315
9	1.6	.001	.16667	.16671
9	1.6	.01		.16708
9	1.6	.1		.17059
9	1.6	.2		.17374
9	1.6	.4		.17706
9	2.0	.001	.20000	.20008
9	2.0	.01		.20080
9	2.0	.1		.20763
9	2.0	.2		.21393
9	2.0	.4		.22072

Theorem VII. 5. b: Assume that the single sample procedure based on n observations from each of k negative binomial populations is used to select the population with smallest probability.

Also assume that the parametric configuration is of the form

$$\theta_i = p \qquad i=2,\ldots,k$$

$$\theta_1 = \theta^* p \qquad (\theta^* < 1) . \qquad \text{(VII. 31)}$$

If n . r = 1 ,

$$PRCS = \frac{\theta^*}{k} \sum_{j=1}^{k} \binom{k}{j} (-1)^j (q^j - 1)$$

$$\cdot \left[1 - q^{j-1}(1-\theta^* p)\right]^{-1} , \qquad \text{(VII. 32)}$$

where $q = 1-p$.

Proof:

$$PRCS = \sum_{j=1}^{k} \sum_{n=0}^{\infty} \theta^* p^j (1-\theta^* p)^n q^{n(j-1)}$$

$$\cdot \frac{1}{j} \binom{k-1}{j-1} (1-q^n)^{k-j} \qquad \text{(VII. 33)}$$

$$= \sum_{n=0}^{\infty} \frac{\theta^*}{k} (1-\theta^* p)^n q^{-n}$$

$$\cdot \sum_{j=1}^{k} \binom{k}{j} (1-q^n)^{k-j} (pq^n)^j \qquad \text{(VII. 34)}$$

Using the fact that the probability generating function for a binomial distribution with parameters (n, p) is given by

$$P(s) = (q + ps)^n \qquad \text{(VII. 35)}$$

we obtain

$$PRCS = \sum_{n=0}^{\infty} \frac{\theta^*}{k} (1-\theta^* p)^n q^{-n}$$

$$\cdot [(1-q^n+pq^n)^k - (1-q^n)^k] \qquad \text{(VII.36)}$$

$$= \sum_{n=0}^{\infty} (\frac{\theta^*}{k})(1-\theta^* p)^n q^{-n}$$

$$\cdot [(1-q^{n+1})^k - (1-q^n)^k] \qquad \text{(VII.37)}$$

$$= \sum_{n=0}^{\infty} \frac{\theta^*}{k} (1-\theta^* p)^n q^{-n}$$

$$\cdot \sum_{j=0}^{k} (\frac{k}{j})[(-q^{n+1})^j - (-q^n)^j] \qquad \text{(VII.38)}$$

$$= \sum_{n=0}^{\infty} \frac{\theta^*}{k}(1-\theta^* p)^n q^{-n}$$

$$\cdot \sum_{j=1}^{k} (\frac{k}{j})(-1)^j (q^j)^n [q^j - 1] \qquad \text{(VII.39)}$$

$$= \frac{1}{k}\theta^* \sum_{j=1}^{k} (\frac{k}{j})(-1)^j (q^j - 1)$$

$$\cdot \sum_{n=0}^{\infty} (q^{j-1}(1-\theta^* p))^n \qquad \text{(VII.40)}$$

$$= \frac{\theta^*}{k} \sum_{j=1}^{k} \binom{k}{j} (-1)^j (q^j - 1)$$

$$\cdot \left[1 - q^{j-1}(1-\theta^* p)\right]^{-1} \qquad \text{(VII. 41)}$$

CHAPTER C. TESTING FOR HOMOGENEITY

Introduction

Thus far we have considered procedures for selecting the best of several populations such that specified levels of PRCS are maintained when the parametric configuration lies outside the indifference region. As discussed by Bechhofer et al. (6, pp. 51, 82), one may also wish to incorporate a test for homogeneity into the procedure. That is, one may wish to investigate a (k+1)st hypothesis that all populations are identical. While this problem has been discussed prominently in the literature, the discussion does not include procedures for which tables or computational formulae necessary for evaluating PRCS are available. In this chapter we consider the following problems of selecting the best multinomial population.

Problem M.a'. Selecting the multinomial category with largest probability: "Select the category associated with $p_{(k)}$ with PRCS $\geq P_1^*$ whenever $p_{(k)}/p_{(k-1)} \geq \theta^*$. If $p_1 = \ldots = p_k$, conclude that all populations are identical with probability no less than P_2^*."

Problem M.b'. Selecting the multinomial category with smallest probability: "Select the category associated with $p_{(1)}$ with PRCS $\geq P_1^*$ whenever $p_{(1)}/p_{(2)} \leq \theta^*$. If $p_1 = \ldots = p_k$, conclude that all populations are identical with probability no less than P_2^*." For these problems we propose the following two procedures:

Procedure M. a'. 1. A sequential procedure for selecting the multinomial category with largest probability: "Sampling sequentially, select the first category to yield R_1 outcomes with the provision that if no category has less than R_2 outcomes $(R_2 < R_1)$, the decision $p_1 = \ldots = p_k$ is made."

Procedure M. b'. 1. A sequential procedure for selecting the multinomial category with smallest probability: "Sampling sequentially, select the last category to yield R_1 outcomes with the provision that if no category has more than R_2 outcomes $(R_2 > R_1)$ the decision $p_1 = \ldots = p_k$ is made."

In view of the equivalence between the problems of selecting the best gamma population and selecting the best multinomial category (see Chapter B of Part IV), Procedures M. a'. 1 and M. b'. 1 may be modified for the problem of selecting the best gamma population when a test for slippage is desired. This result, which will not be pursued in this thesis, has interesting applications to the problem of testing homogeneity of variance.

Evaluation of the probability of a correct decision

Since the procedures proposed in the previous section are such that in some cases no population is selected as best, we adopt the expression "probability of a correct decision" and use the abbreviation PRCD in place of PRCS. In this section we provide computational formulae for evaluating PRCD when the parametric configuration is given by an SC or EM configuration. These formulae may be evaluated easily on a computer.

In addition to the notation introduced in Part II we define

X_{iN} = the number of outcomes from category i in N multinomial trials with K equiprobable categories.

$\{X_{(i)N}\}_{i=1}^{K}$ = the ordered $\{X_{iN}\}_{i=1}^{K}$, $X_{(1)N} \leq \cdots \leq X_{(K)N}$.

$$P_M(A, B, N, K) = Pr[X_{(1)N} \geq A;\ X_{(K)N} < B] \qquad \text{(VII. 42)}$$

$$P_m(A, B, N, K) = Pr[X_{(1)N} \leq A;\ X_{(K)N} < B] \qquad \text{(VII. 43)}$$

$$p_{R, N, K} = Pr[X_{(K)N} \leq R] \qquad \text{(VII. 44)}$$

Theorem VII. 1. a: The probability of a correct decision when using Procedure M. a'. 1 with parametric configuration

$$p_1 = \cdots = p_k \qquad \text{(VII. 45)}$$

is given by

$$PRCD = \sum_{N=R_1}^{R_1+(k-1)(R_1-1)} b(R_1-1;\ N-1,\ \tfrac{1}{k}) P_M(R_2, R_1, N-R_1, k-1) \ . \qquad \text{(VII. 46)}$$

The probability of a correct decision when the parametric configuration is an SC is given by

$$PRCD = \sum_{N=R_1}^{R_1+(k-1)(R_1-1)} \theta^* p\, b(R_1-1;\ N-1,\ \theta^* p)$$

$$\cdot\ P_m(R_2-1,\ R_1,\ N-R_1,\ k-1) \qquad \text{(VII.47)}$$

Relation VII.46 follows from the fact that

$$PRCD = \sum_{N=R_1}^{R_1+(k-1)(R_1-1)} k\ Pr[\pi_1 \text{ is selected on the Nth trial with}$$

$$\min(X_{2n}, \ldots, X_{kn}) \geq R_2] \qquad \text{(VII.48)}$$

Relation VII.47 is obtained using

$$PRCD = \sum_{N=R_1}^{R_1+(k-1)(R_1-1)} Pr[\pi_k \text{ is selected on the Nth trial with}$$

$$\min(X_1, \ldots, X_k) < R_2] \qquad \text{(VII.49)}$$

Theorem VII.1.b: The probability of a correct decision using Procedure M.b'.1 when the parametric configuration is an EM configuration is given by

$$\text{PRCD} = \sum_{N=kR_1}^{R_1+(k-1)R_2} b(R_1-1;\ N-1,\ 1/k)$$

$$\cdot\ P_M(R_1,\ R_2+1,\ N-R_1,\ k-1) \qquad \text{(VII.50)}$$

The probability of a correct decision when the parametric configuration is an SC is given by

$$\text{PRCD} = P(R_1) - \sum_{N=kR_1}^{R_1+(k-1)R_2} \theta^* p\ b(R_1-1;\ N-1,\ \theta^* p)$$

$$\cdot\ P_M(R_1,\ R_2+1,\ N-R_1,\ k-1) \qquad \text{(VII.51)}$$

where $P(R_1)$ = PRCS using the inverse sampling procedure of Part II with a stopping bound of R_1 to select the most probable category.

The derivation of Relation VII.50 is analogous to the derivation of Relation VII. 46. Relation VII. 51 is obtained using the fact that the probability of an incorrect decision may be expressed

$$1-\text{PRCD} = \Pr[\text{the least probable category is the last category to yield } R_1 \text{ outcomes and no category has more than } R_2 \text{ outcomes}] + \Pr[\text{the least probable category is not the last category to yield } R_1 \text{ outcomes}]. \qquad \text{(VII.52)}$$

Some recursive formulae which may be useful in the evaluation of PRCD are:

$$P_M(A, B, N, K) = K \Pr[\text{Nth outcome from } \pi_i;\ A \le (X_{2N}, \ldots, X_{KN}) < B \text{ and } A \le X_{1N} < B] \quad \text{(VII.53)}$$

$$= \sum_{j=A-1}^{B-2} b(j; N-1, 1/K)\, P_M(A, B, N-1-j, K-1) \quad \text{(VII.54)}$$

$$P_m(A, B, N, K) = K \Pr[\text{Nth outcome is from } \pi_1;$$

$$\text{Min}(X_{2N}, \ldots, X_{KN}) \le A;$$

$$\text{Max}(X_{2N}, \ldots, X_{KN}) < B;$$

$$A < X_{1N} < B]$$

$$+ K \Pr[\text{Nth outcome is from } \pi_1;$$

$$\text{Max}(X_{2N}, \ldots, X_{KN}) < B;$$

$$1 \le X_{1N} \le A] \quad \text{(VII.55)}$$

$$= \sum_{j=A}^{B-2} b(j;\, N-1,\, 1/K) P_m(A, B, N-1-j, K-1)$$

$$+ \sum_{j=0}^{A-1} b(j;\, N-1,\, 1/K) p_{B-1,\, N-1-j,\, K-1} \qquad \text{(VII.56)}$$

PART VIII. ESTIMATION OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION

The formulation given in Chapter B of Part IV for the problem of selecting the best gamma population assumed the populations had common, known, and integer valued shape parameter p. In view of the reproductive property of the gamma distribution, the assumption of integer valued shape parameter does not appear to be a serious restriction. However, the assumption of a common known shape parameter may require a preliminary sample to estimate p and test the assumption of homogeneity. Methods for the interval estimation of the shape parameter of a gamma distribution are presented in this part. They are recommended for use when $p \geq 2$.

CHAPTER A. AN APPROXIMATE SINGLE SAMPLE PROCEDURE

The only discussion in the literature regarding the interval estimation of the shape parameter of a gamma distribution appears to be Linhart (29) who proposes an approximate single sample procedure. The proposed estimate of p is

$$\hat{p}_n = \frac{1}{2}(\ell n \bar{x} - \overline{\ell n x})^{-1} + \frac{1}{6} \qquad \text{(VIII. 1)}$$

where

x_i is the ith observation $(i=1, \ldots, n)$

$$\bar{x} = \sum_{i=1}^{n} x_i/n \qquad \text{(VIII. 2)}$$

$$\overline{\ell n x} = \frac{1}{n} \sum_{i=1}^{n} \ell n x_i \ . \qquad \text{(VIII. 3)}$$

Linhart showed that $\hat{p}_n$ is the maximum likelihood estimate of p as follows:

The likelihood function of a sample of size n from a gamma population may be expressed

$$L = \frac{\prod_{i=1}^{n} x_i^{p-1} \, e^{-\Sigma x/\theta}}{[\Gamma(p)]^n \theta^{np}} \ , \qquad \text{(VIII. 4)}$$

so that

$$\frac{1}{n} \ell n\, L = (p-1)\, \overline{\ell n x} - \bar{x}/\theta - \ell n \Gamma(p) - p \ell n \theta \quad . \qquad \text{(VIII. 5)}$$

Therefore,

$$\frac{\partial(\frac{1}{n} \ell n\, L)}{\partial \theta}\Big|_{\substack{\theta = \hat{\theta} \\ p = \hat{p}}} = \frac{\bar{x}}{\hat{\theta}^2} - \frac{\hat{p}}{\hat{\theta}} \qquad \text{(VIII. 6)}$$

$$\frac{\partial(\frac{1}{n} \ell n\, L)}{\partial p}\Big|_{\substack{\theta = \hat{\theta} \\ p = \hat{p}}} = \overline{\ell n x} - \psi(\hat{p}) - \ell n \hat{\theta} \qquad \text{(VIII. 7)}$$

(where $\psi(p)$ is the Psi (Digamma) function). Thus the estimation equations for p and θ may be expressed

$$\ell n \hat{\theta} = \ell n \bar{x} - \ell n \hat{p} \qquad \text{(VIII. 8)}$$

$$\ell n \hat{\theta} = \overline{\ell n x} - \psi(\hat{p}) \, . \qquad \text{(VIII. 9)}$$

The estimate for p is obtained by solving the equation

$$\ell n \bar{x} - \overline{\ell n x} = \ell n \hat{p} - \psi(\hat{p}) \qquad \text{(VIII. 10)}$$

for $\hat{p}$. Using Relation 6. 3. 18 in Handbook of Mathematical Functions (1)

to approximate the Psi function:

$$\psi(\hat{p}) \doteq \ell n\,\hat{p} - (2\hat{p} - \tfrac{1}{3})^{-1} \qquad \text{(VIII. 11)}$$

we obtain VIII. 1 as the estimate for p.

Using

$$Z = \ell n\,\bar{x} - \overline{\ell n\,x} \qquad \text{(VIII. 12)}$$

as a pivotal quantity, Linhart obtains the approximate $100(1-\alpha)\%$ confidence interval for p as

$$p_U = C_L(Z,\ \alpha)\,\chi^2(1 - \tfrac{\alpha}{2},\ n-1)/2nZ \qquad \text{(VIII. 13)}$$

$$p_L = C_u(Z,\ \alpha)\,\chi^2(\tfrac{\alpha}{2},\ n-1)/2nZ \qquad \text{(VIII. 14)}$$

where

$$C_L(Z,\ \alpha) = \tfrac{1}{2}\{1+[1+4(n+1)Z/3\chi^2(1 - \tfrac{\alpha}{2},\ n-1)]^{\frac{1}{2}}\} \qquad \text{(VIII. 15)}$$

$$C_u(Z,\ \alpha) = \tfrac{1}{2}\{1+[1+4(n+1)Z/3\chi^2(\tfrac{\alpha}{2},\ n-1)]^{\frac{1}{2}}\} \qquad \text{(VIII. 16)}$$

Z is always non-negative since for any $n=1, 2, \ldots,$ the arithmetic mean of n positive numbers is always greater than or equal to the geometric mean (see Relations 3. 1. 11, 3. 1. 12, and 3. 2. 1 of

Handbook of Mathematical Functions). As a result, $\hat{p}_n$, p_L, and p_u are always non-negative.

CHAPTER B. A SEQUENTIAL PROCEDURE

Statement of the procedure

In this chapter we propose a sequential method for the interval estimation of the shape parameter of a gamma distribution.

The procedure is stated formally below, where we define

d a number between 0 and 1 specified by the experimenter. (Suggestions for choosing d will be given in a subsequent chapter.)

α $100(1-\alpha)\%$ is the confidence level specified by the experimenter

$$a_{nU} = (1-d)^{-1}\{\tfrac{1}{2}\ell n(\tfrac{1}{d}) + \tfrac{1}{n}[\tfrac{1}{2}\ell n\, d + \ell n\, \tfrac{2}{\alpha}]\} \qquad \text{(VIII. 17)}$$

$$a_{nL} = (1-d)^{-1}\{\tfrac{1}{2}\ell n(\tfrac{1}{d}) + \tfrac{1}{n}[\tfrac{1}{2}\ell n\, d - \ell n\, \tfrac{2}{\alpha}]\} \qquad \text{(VIII. 18)}$$

$$b_n = (1 - \frac{1}{n^2})(12d)^{-1} \qquad \text{(VIII. 19)}$$

$$C_{nU} = (2Z_n)^{-1}[a_{nU} + (a_{nU}^2 + 4b_n Z_n)^{1/2}] \qquad \text{(VIII. 20)}$$

$$C_{nL} = d(2Z_n)^{-1}[a_{nL} + (a_{nL}^2 + 4b_n Z_n)^{1/2}] \qquad \text{(VIII. 21)}$$

$$p_{nU} = \min(c_{1U}, \ldots, c_{nU}) \qquad \text{(VIII. 22)}$$

$$p_{nL} = \max(c_{1L}, \ldots, C_{nL}) \qquad \text{(VIII. 23)}$$

The estimation procedure is as follows: after each observation $X_n (n=1, 2, \ldots)$ compute the estimate $\hat{p}_n$ of p and the upper and lower confidence bounds p_{nL} and p_{nU}. It will be verified in the following section that the probability $p_{nL} < p < p_{nU}$ for all $n=1, 2, \ldots$ is at least $1-\alpha$.

Derivation of the sequential procedure

To derive the estimation procedure proposed in the preceding section let

$$T_n = \prod_{i=1}^{n} X_i / \left(\sum_{i=1}^{n} X_i\right)^n \qquad n=2, 3, \ldots \qquad \text{(VIII. 24)}$$

where X_i is the ith observation from a gamma population with shape parameter p and scale parameter θ. Bartholomew (3) has shown that $T_2, \ldots, T_n$ have joint p.d.f.

$$f^{(n)} \overset{\text{def.}}{=} f_{T_2, \ldots, T_n}(t_2, \ldots, t_n)$$

$$= \Gamma(np)\Gamma^{-n}(p) t_n^{p-1} g(t_2, \ldots, t_n) \qquad \text{(VIII. 25)}$$

where g is functionally independent of both p and θ. We define

$$f_d^{(n)} = \Gamma(npd)\Gamma^{-n}(pd)$$

$$\cdot\, t_n^{pd-1} g(t_2, \ldots, t_n), \qquad \text{(VIII. 26)}$$

so that

$$\frac{f_d^{(n)}}{f^{(n)}} = \frac{\Gamma(npd)}{\Gamma(np)}\left[\frac{\Gamma(p)}{\Gamma(pd)}\right]^n t_n^{pd-p} \quad . \qquad \text{(VIII. 27)}$$

Therefore

$$\ell n\, f_d^{(n)} - \ell n\, f^{(n)}$$

$$= \ell n\, \Gamma(npd) - \ell n\, \Gamma(np)$$

$$+ n[\ell n\, \Gamma(p) - \ell n\, \Gamma(pd)] + p(d-1)\ell n\, t_n \qquad \text{(VIII. 28)}$$

An approximation to $\ell n\, \Gamma(p)$ is given by Relation 6. 1. 41 of Handbook of Mathematical Functions (1):

$$\ell n\, \Gamma(p) = (p - \tfrac{1}{2})\ell n\, p - p + \tfrac{1}{2}\, \ell n(2\pi)$$

$$+ \frac{1}{12p} + R_n \qquad \text{(VIII. 29)}$$

Using approximation IV. 96 we have

$$n \ln \Gamma(p) - \ln \Gamma(pn)$$

$$= -\frac{n}{2} \ln p + \frac{1}{2} \ln np - np \ln n$$

$$+ \frac{n}{12p} - \frac{1}{12np} + R_{n1} - R_{n2} \qquad \text{(VIII. 30)}$$

$$= -\frac{n-1}{2} \ln p - (np - \frac{1}{2}) \ln n + \frac{n^2-1}{12np}$$

$$+ R_{n1} - R_{n2} \qquad \text{(VIII. 31)}$$

and $$n \ln \Gamma(pd) - \ln \Gamma(npd)$$

$$= -\frac{n-1}{2} \ln p - (npd - \frac{1}{2}) \ln n + \frac{n^2-1}{12npd}$$

$$- \frac{n-1}{2} \ln d + R_{n3} - R_{n4} \qquad \text{(VIII. 32)}$$

Therefore

$$\ln f_d^{(n)} - \ln f^{(n)}$$

$$= \frac{(n^2-1)}{12npd}(d-1) + \frac{n-1}{2} \ln d$$

$$- np(1-d) \ln n + p(d-1) \ln t_n$$

$$+ R_{n1} + R_{n3} - R_{n2} - R_{n4} \qquad \text{(VIII. 33)}$$

that is,

$$\ell n f_d^{(n)} - \ell n f^{(n)} = \frac{n^2-1}{12npd}(d-1) + \frac{n-1}{2}\ell n\, d$$

$$+ p(d-1)[n\,\ell n\, n + \ell n\, t_n]$$

$$+ R_n \qquad \text{(VIII. 34)}$$

where

$$R_n = R_{n1} + R_{n3} - R_{n2} - R_{n4} \qquad \text{(VIII. 35)}$$

Since

$$n\,\ell n\, n + \ell n\, t\, n$$

$$= \ell n\, n^n \prod_{i=1}^{n} x_i / \left(\sum_{i=1}^{n} x_i\right)^n \qquad \text{(VIII. 36)}$$

$$= \ell n \prod_{i=1}^{n} x_i / (\bar{x}_n)^n \qquad \text{(VIII. 37)}$$

$$= n(\overline{\ell n x} - \ell n \bar{x}) \qquad \text{(VIII. 38)}$$

$$= -\, n\, Z_n \qquad \text{(VIII. 39)}$$

we have

$$\ell n\, f_d^{(n)} - \ell n\, f^{(n)}$$

$$= n(1-d)\, p\, Z_n + \frac{n-1}{2}\, \ell n\, d$$

$$+ (12npd)^{-1}(1-d)(1-n^2) + R_n \qquad \text{(VIII. 40)}$$

At this point we appeal to the following theorem from p. 146 of Wald (46).

Theorem VIII. 1: Let $x_1, x_2, \ldots$ be a sequence of random variables, let $f_{1m}(x_1, \ldots, x_m)$ $(m=1, 2, \ldots)$ denote the joint density of $x_1, \ldots, x_m$ under a hypothesis H_1, and let $f_{0m}(x_1, \ldots, x_m)$ be the density function under the hypothesis H_0. Also let A be a constant between zero and one. Then under the hypothesis H_0, the probability that

$$\frac{f_{1m}(x_1, \ldots, x_m)}{f_{0m}(x_1, \ldots, x_m)} < \frac{1}{A} \qquad \text{(VIII. 41)}$$

will hold for all values of m is greater than or equal to $1-A$.

As an equivalent expression for the conclusion of Theorem VIII. 1 we write

$$\Pr[f_{1m}/f_{0m} \geq \frac{1}{A} \text{ for any } m=1, 2, \ldots] \leq A. \qquad \text{(VIII. 42)}$$

The proof of Theorem VIII. 1 is as follows. If H_0 is true,

$$\Pr[f_{1m}/f_{0m} \geq \frac{1}{A} \text{ for any } m=1, 2, \ldots]$$

$$= \sum_{m=0}^{\infty} \int_{S_m} f_{0m}\, dx \qquad \text{(VIII. 43)}$$

$$\leq \sum_{m=0}^{\infty} \int_{S_m} A\, f_{1m}\, dx \qquad \text{(VIII. 44)}$$

$$\leq A \qquad \text{(VIII. 45)}$$

where S_m is the set of m dimensional vectors $(x_1, \ldots, x_m)$ which are such that

$$\frac{f_{1m}(x_1, \ldots, x_m)}{f_{0m}(x_1, \ldots, x_m)} \geq \frac{1}{A} \qquad \text{(VIII. 46)}$$

and

$$\frac{f_{1n}(x_1, \ldots, x_n)}{f_{0n}(x_1, \ldots, x_n)} < \frac{1}{A} \qquad n=1, 2, \ldots, m-1 . \qquad \text{(VIII. 47)}$$

Using Theorem VIII. 1, we have from Relation VIII. 40 that for any d

$$\Pr[n(1-d)pZ_n + \frac{n-1}{2}\ell nd - \ell n \quad \frac{\alpha}{2} + R_n$$

$$+ (12\ npd)^{-1}(1-d)(1-n^2) \geq 0 \quad \text{for}$$

$$\text{any} \quad n=1, 2, \ldots] \leq \frac{\alpha}{2} \ . \qquad \text{(VIII. 48)}$$

Thus for $0 < d < 1$,

$$\Pr[p + \frac{1}{2}(1-\frac{1}{n})(\ell n\, d)Z_n^{-1}(1-d)^{-1}$$

$$- \frac{1}{n}(1-d)^{-1}Z_n^{-1}[\ell n(\frac{\alpha}{2}) - R_n]$$

$$+ Z_n^{-1}(12\ pd)^{-1}(\frac{1}{n^2} - 1) > 0 \text{ for any } n=1, 2, \ldots] \leq \frac{\alpha}{2} \ . \qquad \text{(VIII. 49)}$$

It follows that for $0 < d < 1$,

$$\Pr[p^2 - pZ_n^{-1}a_{nU} - Z_n^{-1}b_n \geq 0$$

$$\text{for any} \quad n=1, 2, \ldots] \leq \frac{\alpha}{2} \qquad \text{(VIII. 50)}$$

where

$$a_{nU} = (1-d)^{-1}\{\frac{1}{2}\ell n(\frac{1}{d}) + \frac{1}{n}[\ell n\, \frac{2}{\alpha} - R_n + \frac{1}{2}\ell n\, d\}$$

$$b_n = (12\,d)^{-1}\left(1 - \frac{1}{n^2}\right) . \qquad \text{(VIII. 51)}$$

(Note: the definition of a_{nU} given here differs from that given in the preceding section by a factor of R_n .)

We desire to determine those values of p for which the parabola

$$y = p^2 - p\,Z_n^{-1}\,a_{nU} - Z_n^{-1}\,b_n \qquad \text{(VIII. 52)}$$

is non-negative. Since the parabola achieves its minimum at

$$p = \frac{1}{2}\,Z_n^{-1}\,a_{nU} \qquad \text{(VIII. 53)}$$

and is negative when $p = 0$. The parabola is non-negative (when p is non-negative) only for values of p which are larger than the positive root of the equation

$$p^2 - p\,Z_n^{-1}a_{nU} - Z_n^{-1}b_n = 0 . \qquad \text{(VIII. 54)}$$

Thus for $p > 0$,

$$p^2 - p\,Z_n^{-1}a_{nU} - Z_n^{-1}b_n \geq 0 \qquad \text{(VIII. 55)}$$

if and only if

$$p \geq \frac{1}{2}\{a_{nU}Z_n^{-1} + (Z_n^{-2}a_{nU}^2 + 4b_nZ_n^{-1})^{\frac{1}{2}}\} . \qquad \text{(VIII. 56)}$$

As a result, we have from Relation VIII. 50 that for the upper confidence limit

$$\Pr[p \geq (2Z_n)^{-1}\{a_{nU} + (a_{nU}^2 + 4b_n Z_n)^{1/2}\}$$

$$\text{for any } n=1, 2, \ldots] \leq \frac{\alpha}{2} \quad . \tag{VIII. 57}$$

The derivation of the lower confidence limit is similar to the derivation of the upper limit. Appealing to Relation (VIII. 48) again, we have for $d' > 1$,

$$\Pr[p + \tfrac{1}{2}(1 - \tfrac{1}{n})(\ell n\, d')Z_n^{-1}(1-d')^{-1}$$

$$- \tfrac{1}{n}(1-d')^{-1}Z_n^{-1}[\ell n(\tfrac{\alpha}{2}) - R_n] \tag{VIII. 58}$$

$$- Z_n^{-1}(12\, pd')^{-1}(1 - \frac{1}{n^2}) \leq 0 \quad \text{for any } n=1, 2, \ldots]$$

$$\leq \frac{\alpha}{2} \ , \tag{VIII. 59}$$

so that

$$\Pr[p^2 + pZ_n^{-1}(1-d')^{-1}\{+ \tfrac{1}{2}\ell n\, d'$$

$$- \tfrac{1}{n}(\tfrac{1}{2}\ell n\, d' - R_n + \ell n\, \tfrac{\alpha}{2})\}$$

$$- Z_n^{-1}(12\,d')^{-1}\left(1 - \frac{1}{n^2}\right) \leq 0$$

$$\text{for any } n=1, 2, \ldots] \leq \frac{\alpha}{2} \ . \qquad \text{(VIII.60)}$$

Thus, taking $d' = 1/d$ $(0 < d < 1)$,

$$\Pr[p^2 - pZ_n^{-1}(1-d)^{-1}d\{\frac{1}{2}\ln\frac{1}{d}$$

$$- \frac{1}{n}(\ln\frac{\alpha}{2} - R_n - \frac{1}{2}\ln d)\} - \frac{1}{12}Z_n^{-1}\,d\left(1 - \frac{1}{n^2}\right) \leq 0$$

$$\text{for any } n=1, 2, \ldots] \leq \frac{\alpha}{2} \ . \qquad \text{(VIII.61)}$$

That is

$$\Pr[p^2 - pZ_n^{-1}da_{nL} - Z_n^{-1}db_n \leq 0$$

$$\text{for any } n=1, 2, \ldots] \leq \frac{\alpha}{2} \qquad \text{(VIII.62)}$$

where

$$a_{nL} = (1-d)^{-1}[\frac{1}{2}\ln\frac{1}{d} - \frac{1}{n}(\ln\frac{\alpha}{2} - \frac{1}{2}\ln d$$

$$- R_n)] \ . \qquad \text{(VIII.63)}$$

We desire to determine those non-negative values of p for which the parabola

$$y = p^2 - p Z_n^{-1} d a_{nL} - Z_n^{-1} d b_n \qquad \text{(VIII. 64)}$$

is non-positive. Since the minimum is achieved at

$$p = \frac{1}{2} Z_n^{-1} d a_{nL} \qquad \text{(VIII. 65)}$$

and the parabola is negative when $p=0$, the desired values of p are those which are greater than zero but less than the positive root of the equation

$$p^2 - p Z_n^{-1} d a_{nL} - Z_n^{-1} b_n = 0 . \qquad \text{(VIII. 66)}$$

Thus, for $p > 0$,

$$p^2 - p Z_n^{-1} d a_{nL} - Z_n^{-1} b_n \leq 0 \qquad \text{(VIII. 67)}$$

if and only if

$$p \leq \frac{d Z_n^{-1}}{2} [a_{nL} + (a_{nL}^2 + 4 b_n Z_n)^{1/2}] \qquad \text{(VIII. 68)}$$

As a result, we have from Relation VIII. 62 that for the lower confidence limit

$$\Pr\left[p \leq \frac{dZ_n^{-1}}{2} \left\{a_{nL} + (a_{nL}^2 + 4b_n Z_n)^{\frac{1}{2}}\right\}\right.$$

$$\left.\text{for any } n=1, 2, \ldots\right] \leq \frac{\alpha}{2} \ . \qquad \text{(VIII. 69)}$$

Ignoring the factor R_n, we obtain

$$\Pr[p_{nL} \leq p \leq p_{nU} \text{ for all } n=1, 2, \ldots]$$

$$= 1 - \Pr[p > p_{nU} \text{ or } p < p_{nL} \text{ for some}$$

$$n=1, 2, \ldots] \qquad \text{(VIII. 70)}$$

$$\geq 1 - \Pr[p > p_{nU} \text{ for some } n=1, 2, \ldots]$$

$$- \Pr[p < p_{nL} \text{ for some } n=1, 2, \ldots] \qquad \text{(VIII. 71)}$$

$$\geq 1-\alpha \ . \qquad \text{(VIII. 72)}$$

CHAPTER C. THE EFFECT OF ERROR IN EVALUATING THE LOG Γ FUNCTION

In this chapter we evaluate the effect on the sequential estimation procedure of replacing R_n (the error introduced by approximating the log Γ function) by zero. Recalling Relation VIII. 35

$$R_n = R_{n1} + R_{n3} - R_{n2} - R_{n4} \tag{VIII. 73}$$

we obtain from relation (6. 1. 42) of Handbook of Mathematical Functions

$$R_{n_1} = \sum_{m=2}^{\infty} \frac{n B_{2m}}{2m(2m-1)p^{2m-1}} \tag{VIII. 74}$$

$$R_{n_2} = \sum_{m=2}^{\infty} \frac{B_{2m}}{2m(2m-1)(np)^{2m-1}} \tag{VIII. 75}$$

$$R_{n_3} = \sum_{m=2}^{\infty} \frac{n B_{2m}}{2m(2m-1)(pd)^{2m-1}} \tag{VIII. 76}$$

$$R_{n_4} = \sum_{m=2}^{\infty} \frac{B_{2m}}{2m(2m-1)(npd)^{2m-1}} \cdot \tag{VIII. 77}$$

Thus,

$$R_n = \sum_{m=2}^{\infty} B_{2m}/[2m(2m-1)p^{2m-1}] \cdot [n(1+d^{-2m+1})-n^{-2m+1}(1+d^{-2m+1})] \cdot \tag{VIII. 78}$$

Therefore,

$$\frac{1}{n} R_n = \sum_{m=2}^{\infty} B_{2m}/[2m(2m-1)p^{2m-1}]$$

$$(1+d^{-2m+1})(1-n^{-2m}) \qquad \text{(VIII. 79)}$$

$$\leq \sum_{m=2}^{\infty} B_{2m}/[2m(2m-1)p^{2m-1}]$$

$$(1+d^{-2m+1}) \ . \qquad \text{(VIII. 80)}$$

Thus, from the discussion on p. 257 of Handbook of Mathematical Functions we have that R_n is positive and

$$|\frac{1}{n} R_n| \leq (360)^{-1}(p^{-3}+(pd)^{-3}) \ . \qquad \text{(VIII. 81)}$$

Since R_n is positive, the confidence intervals defined in the estimation procedure are wider than the intervals depending on R_n obtained in the derivation. Thus the proposed estimation procedure is conservative.

It is desired next to evaluate the extent of error involved in replacing R_n by zero in the definition of a_{nU} (denoted by a'_{nU}) given in the specification of the estimation procedure.

From Relation VIII. 81,

$$|a_{nU} - a'_{nU}| = |R_n/n(1-d)| \qquad \text{(VIII. 82)}$$

$$\leq (360)^{-1} p^{-3}(1+d^{-3})(1-d)^{-1} . \qquad \text{(VIII. 83)}$$

Thus, for the error expressed as a fraction of the true value (a_{nU}) we have

$$\frac{|a_{nU} - a'_{nU}|}{a_{nU}} = \frac{|\frac{1}{n} R_n|}{\frac{1}{2}\ell n(1/d) + \frac{1}{n}(\ell n(2/\alpha) - R_n + \frac{1}{2}\ell n\, d)} \qquad \text{(VIII. 84)}$$

$$= |(\tfrac{n-1}{2}\ell n(1/d) + \ell n(2/\alpha))R_n^{-1} - 1|^{-1} \qquad \text{(VIII. 85)}$$

$$= [1+(\tfrac{n-1}{2}\ell n(1/d) + \ell n(2/\alpha))R_n^{-1}]^{-1} \qquad \text{(VIII. 86)}$$

$$\leq [1+(\tfrac{n-1}{2n}\ell n(1/d) + \tfrac{1}{n}\ell n(2/\alpha)) \cdot 360p^3(1+d^{-3})^{-1}]^{-1} \qquad \text{(VIII. 87)}$$

$$= [1+(\tfrac{1}{2}\ell n(1/d) + \tfrac{1}{n}\ell n(2d^{1/2}/\alpha)) \cdot 360p^3(1+d^{-3})^{-1}]^{-1} \qquad \text{(VIII. 88)}$$

Assuming $d > \alpha^2/4$, we thus obtain

$$\frac{|a_{nU} - a'_{nU}|}{a_{nU}} \leq [1 + 180\,\ell n(1/d)p^3(1+d^{-3})^{-1}]^{-1} \quad . \text{(VIII. 89)}$$

Table VIII. 1 gives upper bounds on

$$\frac{|a_{nU} - a'_{nU}|}{a_{nU}} \times 100\% \quad . \qquad \text{(VIII. 90)}$$

Except for p=2, d=.99 .the approximation appears quite accurate. Relation VIII. 88 indicates that, for $d > \alpha^2/4$, the approximation is most accurrate for small values of n. For α small and d large, the upper bound on error given by Relation VIII. 88 may be substantially lower than that given by Relation VIII. 89.

Table VIII. 1. Percent error in evaluating a_{nU}.

d \ p	2	3	10
.9	1.54	.461	.012
.99	14.0	3.99	.112

CHAPTER D. DETERMINATION OF d

We assume that the experimenter is able to specify an integer N for which he would like to minimize $p_{NU} - p_{NL}$. In view of the obvious difficulties involved in attempting such a minimization, we shall instead assume that p is sufficiently large that we may ignore the term b_N (which arises from using a refinement of Sterlings approximation of the gamma function). We therefore consider the problem of minimizing $a_{NU} - d\,a_{NL}$.

$$a_{NU} - d\,a_{NL} = \frac{1}{2}(-\ell n\, d + \frac{1}{n}\ell n d) + (1-d)^{-1}(1+d)N^{-1}\ell n(2/\alpha) \qquad \text{(VIII. 91)}$$

$$= \frac{1}{2}\ell n d(N^{-1} - 1) + (1-d)^{-1}(1+d) \cdot N^{-1}\ell n(2/\alpha) \qquad \text{(VIII. 92)}$$

Thus, setting the derivative of $a_{NU} - d\,a_{NL}$ equal to zero, we obtain

$$(2d)^{-1} + 2N^{-1}\ell n(2/\alpha)(1-d)^{-2} = 0 \qquad \text{(VIII. 93)}$$

Therefore,

$$(1-d)^2(\frac{1}{N} - 1) + \frac{4d}{N} \ell n \frac{2}{\alpha} = 0 \qquad \text{(VIII. 94)}$$

$$- d^2(1 - \frac{1}{N}) + 2d(1 - \frac{1}{N}) - (1 - \frac{1}{N})$$

$$+ \frac{4d}{N} \ell n \frac{2}{\alpha} = 0 \ , \qquad \text{(VIII. 95)}$$

$$d^2 - 2d(1 + (N-1)^{-1}(2 \ell n \frac{2}{\alpha})) + 1 = 0 \ . \qquad \text{(VIII. 96)}$$

Solving for d, one obtains

$$d_{opt} = A_N - (A_N^2 - 1)^{\frac{1}{2}} \qquad \text{(VIII. 97)}$$

where

$$A_N = 1 + (N-1)^{-1}(2 \ell n \frac{2}{\alpha}) \qquad \text{(VIII. 98)}$$

Since A_N is greater than unity, $A_N^2 - 1$ is greater than zero. Thus the root in Relation IV. 135 is real. Since $A_N^2 - 1$ is less than A_N^2, d_{opt} is greater than zero. To show that d_{opt} is less than one, we first suppose the contrary. If $d_{opt} \geq 1$, then

$$A_N - (A_N^2 - 1)^{1/2} \geq 1 \ . \qquad \text{(VIII. 99)}$$

This implies that

$$(A_N - 1)^2 \geq A_N^2 - 1 \qquad \text{(VIII. 100)}$$

so that

$$-2A_N + 1 \geq -1 \qquad \text{(VIII. 101)}$$

or

$$A_N \leq 1 \qquad \text{(VIII. 102)}$$

which, by inspection of Relation (VIII. 98) is a contradiction. Thus, d_{opt} is less than one.

That d_{opt} is less than one may also be obtained from noting that $A_N > 1$, $d_{opt} = 1$ for $A_N = 1$, and d_{opt} is a decreasing function of A_N.

CHAPTER E. COMPARISON OF THE SINGLE SAMPLE AND SEQUENTIAL PROCEDURES

The choice of sampling plan, single sample versus sequential, may well be determined by non-statistical considerations. However, if the estimation of the shape parameter p is to be followed by a test (presumed to depend on p) on the scale parameter θ, then the following theorem due to Pitman (36) demonstrates that the sequential procedure has an important advantage over the single sample procedure.

Theorem VIII.2: Let X_i, $i=1, \ldots, n$ be i.i.d.r.v.'s having a gamma distribution. If $g(X_1, \ldots, X_n)$ is any function which is invariant under change in scale, then $g(X_1, \ldots, X_n)$ and $\sum_{i=1}^{n} X_i$ are statistically independent.

In view of the fact that $\hat{p}_n$ is invariant under change in scale and ΣX_i is a sufficient statistic for θ when p is known, Theorem VIII.2 suggests that one use the same observations for tests on θ as are used to estimate p. Specifically, if for p known the test on θ is a single sample procedure, one may estimate p sequentially until an estimate $\hat{p}_N$ is obtained for which the confidence interval is sufficiently small and the sample size is sufficiently large to permit the desired test on θ. In this manner the totality of observations collected are used for both estimation and testing. In comparison, using the single sample procedure, one cannot be sure of the width of the confidence interval for p upon completion of the estimation procedure, nor can one be sure that the

estimate of p will be such that additional sampling will not be required for the test on θ.

It is also desirable to compare the efficiency of the single sample and sequential procedures. One measure of efficiency is the average width of the confidence interval after sampling N observations. Unfortunately, due to the complex nature of the equations for the confidence bounds, it does not appear feasible to make comparisons based on average width. Therefore we shall compare the width of the confidence interval using the single sample procedure with an upper bound on the width of the confidence interval using the sequential procedure. We consider the case $N = 500$, $\alpha = .05$, $Z_{500} = .2$. For this configuration, $\hat{p}_{500} = 2.667$. The confidence interval obtained using the single sample procedure is $(2.345, 2.969)$. The values of $c_{500,U}$ and $c_{500,L}$ obtained using the sequential procedure are $(2.397, 3.108)$.

Computations for obtaining the single sample interval are

$$c_L(.2, .95) = 1.056$$

$$c_u(.2, .95) = 1.071$$

$$\chi^2(.975, 499) = 562.308$$

$$\chi^2(.025, 499) = 438.533$$

$$p_L = 2.345$$

$$p_U = 2.969 .$$

Computations for obtaining $(c_{500,L};\ c_{500,U})$ are

$$\ell n \frac{2}{\alpha} = \ell n\ 40 = 3.6889$$

$$A_N = 1.01476$$

$$d_{opt} = .84231 \qquad \ell n\ d = -.17162$$

$$1-d = .15769$$

$$a_{NU} = .58976 \qquad b_N = .0989$$

$$a_{NL} = .49654 \qquad 4b_N Z_N = .07912$$

$$C_{NU} = 3.108$$

$$C_{NL} = 2.397$$

REFERENCES

1. Abramowitz, M. and Stegun, I. A., ed. Handbook of Mathematical Functions. Washington, D. C., National Bureau of Standards. 1965.

2. Bahadur, R. R. and Goodman, L. A. Impartial decision rules and sufficient statistics. The Annals of Mathematical Statistics 23: 553-562. 1952.

3. Bartholomew, D. J. A sequential test for randomness of intervals. Journal of Royal Statistical Society 18: 95-102. 1956.

4. Bechhofer, R. E. A single sample multiple decision procedure for ranking means of normal populations with known variances. The Annals of Mathematical Statistics 25: 16-39. 1954.

5. Bechhofer, R. E., Elmaghraby, S. and Morse, N. A single sample multiple decision procedure for selecting the multinomial event which has the highest probability. The Annals of Mathematical Statistics 30: 102-119. 1959.

6. Bechhofer, R. E., Kiefer, J. and Sobel, M. Sequential Identification and Ranking Procedures. Chicago, Illinois, University of Chicago Press. 1968.

7. Bechhofer, R. E. and Sobel, M. A single sample multiple decision procedure for ranking variances of normal populations. The Annals of Mathematical Statistics 25: 273-289. 1954.

8. Bechhofer, R. E. and Sobel, M. A scale invariant sequential multiple decision procedure for selecting the population with smallest variance from k normal populations (preliminary report). The Annals of Mathematical Statistics 27: 219. 1956.

9. Bechhofer, R. E. and Sobel, M. A sequential multiple-decision procedure for selecting the multinomial event with the largest probability (abstract). The Annals of Mathematical Statistics 27: 861. 1956.

10. Bechhofer, R. E. and Sobel, M. Non-parametric multiple decision procedures for selecting that one of k populations which has the highest probability of yielding the largest observation (abstract). The Annals of Mathematical Statistics 29: 325. 1958.

11. Bliss, C. I. and Owen, R. G. Negative binomial distributions with a common k. Biometrika 45: 37. 1958.

12. Cacoullos, T. and Sobel, M. The effect of inverse sampling on ranking multinomial probabilities. University of Minnesota, Department of Statistics Technical Report 55. 1966.

13. Desu, M. M. A selection problem. The Annals of Mathematical Statistics 38: 956. 1967.

14. Dudewicz, E. J. An approximation to the sample size in a selection procedure of Bechhofer and its extension to selection from a multivariate normal population. The Annals of Mathematical Statistics 39: 301. 1968.

15. Feller, W. An Introduction to Probability Theory and Its Applications. Vol. 1. 3rd. ed. New York, N. Y. John Wiley and Sons. 1957.

16. General Electric Co. Tables of the Individual and Cumulative Terms of the Poisson Distribution. New York, N. Y., D. Van Nostrand. 1962.

17. Gupta, S. S. On a selection and ranking procedure for gamma populations. Ann. Inst. Statist. Math. Tokyo 14: 199-216. 1963.

18. Gupta, S. S. Probability integrals of multivariate normal and multivariate t. The Annals of Mathematical Statistics 34: 792-828. 1963.

19. Gupta, S. S. Selection and ranking procedures and order statistics for the binomial distribution. Stanford University Technical Report No. 66. 1966.

20. Gupta, S. S. and Nagel, K. On selection and ranking procedures and order statistics from the multinomial distribution. The Annals of Mathematical Statistics 39: 304. 1966.

21. Gupta, S. S. and Sobel, M. Selecting a subset containing the best of several binomial populations. In Olkin, I., ed. Contributions to Probability and Statistics. Chapter 10. Pp. 224-248. Stanford, Calif., Stanford University Press. 1960.

22. Gupta, S. S. and Sobel, M. On the smallest of several correlated F-statistics. Biometrika 49: 507. 1962.

23. Gupta, S. S. and Sobel, M. On selecting a subset containing the population with smallest variance. Biometrika 49: 495. 1962.

24. Hoel, D. G. and Mazumder, M. An extension of Paulson's selection procedure. The Annals of Mathematical Statistics 39: 2067. 1968.

25. Huyett, H. M. and Sobel, M. Selecting the best one of several binomial populations. The Bell System Technical Journal 36: 537-576. 1957.

26. Kesten, H. and Morse, N. A property of the multinomial distribution. The Annals of Mathematical Statistics 30: 120-127. 1959.

27. Lehman, E. L. On a theorem by Bahadur and Goodman. The Annals of Mathematical Statistics 37: 1-7. 1966.

28. Lehman, E. L. Testing Statistical Hypotheses. New York, N. Y. John Wiley and Sons. 1959.

29. Linhart, H. Approximate confidence limits for the coefficient of variation of gamma distributions. Biometrics 21: 733-737. 1965.

30. O'Brien, P. C. Selection Procedures Relating to Gamma and Weibull Populations. Unpublished M. S. thesis. Ames, Iowa, Library, Iowa State University. 1968.

31. Parzen, E. Stochastic Processes. San Francisco, California, Holden Day. 1962.

32. Patel, J. K. Selecting a subset containing the best one of several IFRA populations. The Annals of Mathematical Statistics 38: 957. 1967.

33. Paulson, E. A sequential procedure for selecting the population with the largest mean from k normal populations. The annals of Mathematical Statistics 35: 174-180. 1964.

34. Paulson, E. Sequential procedures for selecting the best one of several binomial populations. The Annals of Mathematical Statistics 38: 117-123. 1967.

35. Perng, S. K. A comparison of the asymptotic expected sample sizes of two sequential procedures for ranking problem. The Annals of Mathematical Statistics 40: 2198-2202. 1969.

36. Pitman, E. J. G. The closest estimates of statistical parameters. Proc. Cambridge. Philos. Soc. 33: 212-222. 1937.

37. Rand Corporation. A Million Random Digits with 10,000 Normal Deviates. Glencoe, Illinois, Free Press. 1955.

38. Robbins, H. Mixtures of distributions. The Annals of Mathematical Statistics 19: 360-369. 1948.

39. Robbins, H. and Pitman, E. Applications of the method of mixtures to quadratic forms in normal variates. The Annals of Mathematical Statistics 20: 552-560. 1949.

40. Sobel, M. Selecting the t populations with the largest α-percentiles. The Annals of Mathematical Statistics 37: 546. 1966.

41. Sobel, M. Sequential procedures for selecting the best exponential population. In Neyman, J., ed. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. Vol. 5. Pp. 99-110. Berkeley, Calif., University of California Press. 1956.

42. Sobel, M. Statistical techniques for reducing the experiment time in reliability studies. The Bell System Technical Journal 35: 179-202. 1956.

43. Sobel, M. and Risvi, M. H. Selecting a subset containing the population with the largest α-th percentile. The Annals of Mathematical Statistics 37: 315. 1966.

44. Tiecher, H. On the mixture of distributions. The Annals of Mathematical Statistics 31: 55-73. 1960.

45. Thomas, D. R. and O'Brien, P. C. Procedures for selecting the best reliability distribution. (Abstract) Technometrics 11: 220. 1969.

46. Wald, A. Sequential Analysis. New York, N.Y., John Wiley and Sons, Inc. 1947.

ACKNOWLEDGMENTS

I would like to thank my major professor, Dr. Wayne Fuller of Iowa State University, for his patience and guidance during the course of this research. I would also like to thank Miss Jan Bates for her careful typing of the manuscript under trying conditions. Last, but not least, thanks go to my wife Cathy for her patience, understanding and encouragement.