ON SELECTING A SUBSET WHICH INCLUDES THE $ t $ BEST OF $ k $ POPULATIONS : SCALE PARAMETER CASE

Ofosu, J. B.
Department of Mathematics, University of Cape Coast

https://doi.org/10.5109/13090

出版情報：統計数理研究. 16 (3/4), pp.7-18, 1975-03. Research Association of Statistical Sciences

ON SELECTING A SUBSET WHICH INCLUDES THE *t* BEST OF *k* POPULATIONS: SCALE PARAMETER CASE

By

J. B. Ofosu*

(Received October 20, 1973)

§0. Summary.

The problem of selecting a subset of fixed size s which includes the t best of k populations ($t \leqq s < k$), based on a pre-determined sample size n from each of the k populations, is studied as a multiple decision problem. It is assumed that the bestness of a population is characterized by its scale parameter; the best population being the one with the largest scale parameter, and so on. Exact small and large sample methods of finding n are given for the scale parameter problem for

(i) Gamma distributions with known (possibly unequal) shape parameters

(ii) Weibull distributions with known shape parameters.

Some tables computed by these methods are provided. A dual problem is also discussed.

§1. Introduction.

An important class of problem is concerned with selection and ranking of k populations. The selection and ranking may be defined in terms of a parameter of the population which may physically represent the mean, the variance, some quantile or a function of these quantiles. That the formulation in terms of ranking and selection is more realistic and meaningful than that of tests of homogeneity of the parameters is by now well recognized. For the general background and motivation for such problems the reader is referred to Bechhofer (1954), Bechhofer and Sobel (1954).

The formulation considered in this paper and in the paper by Mahamunulu (1967) is that of selecting a subset of specified size, from a given set of k populations, which contains the t $(< k)$ best populations. The t populations with the largest (or smallest) parameter values are usually defined as the t best populations.

In this paper we are interested in the scale parameters θ_i $(i=1, 2, \cdots, k)$ of the k populations. The goal is to select a subset of fixed size s which includes the populations with the t $(t \leqq s < k)$ largest θ-values, based on a pre-determined fixed sample size n from each population. A correct selection (CS) is defined as the selection of any subset of fixed size s which contains the populations with the t $(\leqq s)$ largest θ-values.

* Department of Mathematics, University of Cape Coast, Cape Coast, Ghana.

The main problem is formally described in section 2. In section 3 we define a procedure R for which the probability of a correct selection is at least P^* (specified) whenever the t best populations are at a distance of at least δ^* (specified) from the remaining $(k-t)$ populations. In section 4 we derive an expression for the infimum of the probability of a correct selection. Specific applications of the procedure are discussed in section 5. In sections 6 and 7 we discuss briefly the dual problem to that defined in section 2.

Our formulation is related to the paper by Desu and Sobel (1968) which deals with the problem, inverse to ours, of selecting a subset of the smallest possible fixed size s that will contain the t best of k populations $(t \leqq s \leqq k)$, for a given common sample size n. There are several aspects of our formulation which have already been described by Mahamunulu (1967) and which apply to the problems treated in this paper. One is that the procedure R can be regarded as an elimination or screening procedure. Another is that a confidence statement can be made after experimentation. A third is that if $\theta_i \geqq \theta_j$ the probability of including the population with parameter θ_i in the selected subset is not less than that of including the population with parameter θ_j.

It should be noted that procedure R could be applied to several other distributions in the Koopman-Darmois family. For the case in which the t best populations are defined to be the t populations with the smallest values of θ, the statistical problem is identical and all the results and tables of this paper apply with the obvious modifications.

§ 2. Formal statement of the problem.

Let X_{ij} $(j=1, 2, \cdots)$ denote observations from population π_i $(i=1, 2, \cdots, k)$; observations between and within populations are all independent. Let $F_{\theta_i}(x)=F(x/\theta_i)$ denote the cumulative distribution function (c. d. f.) of X_{ij} associated with π_i, which is known except for the value of the scale parameter θ_i. The ordered θ-values and associated vector are denoted respectively by

$$0<\theta_{[1]} \leqq \theta_{[2]} \leqq \cdots \leqq \theta_{[k]}<\infty\,; \qquad \underset{\sim}{\theta}=\{\theta_{[1]}, \theta_{[2]}, \cdots, \theta_{[k]}\}\,.$$

It is not known which population is associated with $\theta_{[i]}$ $(i=1, 2, \cdots, k)$. The t populations associated with $\theta_{[k-t+1]}, \cdots, \theta_{[k]}$ are defined as the t best populations. Following Bechhofer (1954) we partition the parameter space into a 'preference-zone' $\Omega(\delta^*)$ defined by

$$\Omega(\delta^*)=\{\underset{\sim}{\theta} : \theta_{[k-t+1]}/\theta_{[k-t]} \geqq \delta^*>1\} \tag{2.1}$$

and its complement, the 'indifference-zone'.

Our goal is to select a subset of fixed size s which includes the t $(\leqq s)$ best of the k populations on the basis of a pre-determined sample size n from each of the k populations.

We shall propose a procedure R for which the probability of a correct selection satisfies the condition

$$P[CS \mid R, \underset{\sim}{\theta}] \geqq P^*\,, \quad \binom{s}{t}\Big/\binom{k}{t}<P^*<1 \qquad \text{for all } \underset{\sim}{\theta} \in \Omega(\delta^*) \tag{2.2}$$

where P^* and δ^* are specified in advance by the experimenter.

§ 3. Procedure R.

The experimenter takes a pre-determined number $n=n(k, t, s, \delta^*, P^*)$ of independent observations from each population. A statistic Y is chosen depending on the given family $F_{\theta_i}(x)$ of c. d. f.'s, and the procedure R is based on the observations only through the k values of Y. Let $G_{\theta_i}(x)$ denote the c.d.f. of Y_i from population π_i $(i=1, 2, \cdots, k)$.

ASSUMPTION 1. Y is chosen so that the family of c. d. f.'s $\{G_\theta(x),\ 0<\theta<\infty\}$ is a stochastically increasing family of absolutely continuous c. d. f.'s, i. e ; $\theta<\theta' \Rightarrow G_{\theta'}(x) \leqq G_\theta(x)$ for all x.

Since θ is a scale parameter $G_\theta(x)$ can be expressed as

$$G_\theta(x)=G(x/\theta), \qquad G(0)=0. \tag{3.1}$$

Although it is not assumed that Y is sufficient for θ, it will be understood that if a sufficient statistic exists then Y will be chosen as a function of it.

The procedure R is defined as follows:

Procedure R: Select any subset of the k populations that gives rise to the s largest Y-values.

Once the common sample size n is specified, the procedure R is completely defined ; our problem will be that of determining this sample size so that the probability requirement (2.2) is satisfied.

§ 4. The infimum of $P[CS|R, \theta]$.

Let $Y_{(i)}$ denote the Y-value from the population with parameter $\theta_{[i]}$ $(i=1, 2, \cdots, k)$. It is easily seen that a correct selection occurs if and only if

$$\min_{k-t+1\leqq i\leqq k} Y_{(i)} > (s-t+1)\text{-th largest of } \{Y_{(1)}, \cdots, Y_{(k-t)}\}.$$

The $P[CS|R, \underline{\theta}]$ can therefore be represented as

$$\sum_{i=1}^{k-t} P[Y_{(i)}=(s-t+1)\text{-th largest of } \{Y_{(1)}, \cdots, Y_{(k-t)}\} \text{ and } Y_{(i)}<\min\{Y_{(k-t+1)}, \cdots, Y_{(k)}\}] \tag{4.1}$$

We want $P[CS|R, \theta]$ to satisfy inequality (2.2) and this is equivalent to requiring that the infimum over $\Omega(\delta^*)$ of $P[CS|R, \underline{\theta}]$ be greater than or equal to P^*.

Now, Assumption 1 satisfies the hypotheses of the first theorem of Mahamunulu (1967), hence $P[CS|R, \underline{\theta}]$ is a non-increasing function of $\theta_{[i]}$ $(i=1, 2, \cdots, k-t)$, and a non-decreasing function of $\theta_{[j]}$ $(j=k-t+1, \cdots, k)$. Thus the infimum over $\Omega(\delta^*)$ of $P[CS|R, \underline{\theta}]$ is the limit of $P[CS|R, \underline{\theta}]$ as $\theta_{[i]}\rightarrow\theta$, say, $(i=1, 2, \cdots, k-t)$, and $\theta_{[j]}\rightarrow\delta^*\theta$ $(j=k-t+1, \cdots, k)$. Since $G_\theta(y)$ is continuous in θ, this limit is, from (4.1), equal to

$$Q_S(k, t, s|\theta) = \frac{(k-t)!}{(s-t)!\,(k-s-1)!}\int_0^\infty \{1-G_{\delta^*\theta}(y)\}^t G_\theta^{k-s-1}(y)\{1-G_\theta(y)\}^{s-t} dG_\theta(y). \tag{4.2}$$

Using (3.1), expression (4.2) reduces to

$$Q_S(k, t, s) = \frac{(k-t)!}{(s-t)!\,(k-s-1)!}\int_0^\infty \{1-G(y/\delta^*)\}^t G^{k-s-1}(y)\{1-G(y)\}^{s-t} dG(y). \tag{4.3}$$

which does not depend on θ.

It should be noted that (4.3) can also be written as

$$tP[Y_{(i)}=\min\{Y_{(k-t+1)}, \cdots, Y_{(k)}\} \text{ and } Y_{(i)}>(s-t+1)\text{-th largest of } \{Y_{(1)}, \cdots, Y_{(k-t)}\}]$$
$$=t\sum_{r=0}^{s-t}\binom{k-t}{r}\int_0^{\infty} G^{k-t-r}(y\delta^*)\{1-G(y\delta^*)\}^r\{1-G(y)\}^{t-1}dG(y)\,. \tag{4.4}$$

The required sample size is the smallest value of n for which

$$Q_S(k, t, s)\geqq P^*\,. \tag{4.5}$$

The required sample size exists provided the left hand side of (4.5) tends to 1 as $n\to\infty$.

§ 5. Specific applications of the procedure.

The procedure R can be applied to several distributions in the Koopman-Darmois family. The gamma and Weibull distributions with unknown scale parameters and all other parameters known have frequency functions which are of the Koopman-Darmois form. In fact each of these distributions can be regarded as a different generalisation of the exponential distribution for a particular value of one of its parameters.

It is not difficult to think of a number of situations where the gamma and Weibull distributions are applicable. These include some problems in reliability, life testing, and fatigue testing. Because of the above properties, these two distributions are treated in this section.

5.1. Gamma populations with unknown scale parameters and known shape parameters.

For the gamma populations under consideration, the frequency function associated with population π_i is given by

$$\frac{1}{\theta_i}\left(\frac{x}{\theta_i}\right)^{a_i-1}\frac{1}{\Gamma(a_i)}\exp\left\{-\frac{x}{\theta_i}\right\} \qquad x>0 \tag{5.1}$$

where the a_i (>0) are the known shape parameters and the θ_i (>0) are the unknown scale parameters.

The k populations might be k different production precesses, such that the life-times of components taken from π_i are distributed according to the gamma distribution with p. d. f. given by (5.1). The mean item life for process π_i is $a_i\theta_i$. It follows that if the a_i are equal the problem of selecting the process with the largest mean life reduces to the problem of selecting the process with the largest value of the scale parameter.

For the problem of selecting a subset of fixed size s which includes the populations with the t ($t\leqq s<k$) largest θ-values, we take $Y_i=\bar{X}_i/a_i$ where $\bar{X}_i$ is the sample mean from π_i. Y_i is a sufficient statistic for the parameter θ_i.

5.1.1. Case A: Shape parameters known and equal.

If $a_i=a$ ($i=1, 2, \cdots, k$) we take n observations from each population. Hence,

$$G_{\theta_i}(y)=\int_0^y g_{\theta_i}(x)dx \tag{5.2}$$

where

$$g_{\theta_i}(x)=\frac{na}{\theta_i}\frac{1}{\Gamma(na)}\left(\frac{nax}{\theta_i}\right)^{na-1}\exp\{-nax/\theta_i\} \qquad x>0. \tag{5.3}$$

5.1.1.1. Exact expression for $Q_S(k, t, s)$.

It is readily seen that $G_{\theta_i}(y)$ satisfies Assumption 1. Thus using the above results, (4.3) takes the form

$$Q_S(k, t, s)=\frac{(k-t)!}{(s-t)!(k-s-1)!}\int_0^\infty\{1-H_r(y/\delta^*)\}^t H_r^{k-s-1}(y)\{1-H_r(y)\}^{s-t}dH_r(y) \tag{5.4}$$

where

$$H_r(y)=\int_0^y\frac{1}{\Gamma(r)}u^{r-1}e^{-u}du\,, \tag{5.5}$$

$$r=na\,. \tag{5.6}$$

$H_r(y)$ can be expressed as

$$1-\sum_{j=0}^{r-1}\frac{e^{-y}y_j}{j!} \qquad y>0, \qquad \text{if } r \text{ is an integer} \tag{5.7}$$

$$\left.\begin{array}{l} erf(y^{\frac{1}{2}}) \\ -\dfrac{y^{\frac{1}{2}}e^{-y}}{\sqrt{\pi}}\displaystyle\sum_{j=1}^{r-\frac{1}{2}}\frac{2^jy^{j-1}}{1\cdot3\cdots(2j-1)}\end{array}\right\} \quad y>0, \qquad \text{if } r \text{ is a half integer} \tag{5.8}$$

where

$$erf(x)=\frac{2}{\sqrt{\pi}}\int_0^y e^{-t^2}dt$$

and where the second term on the right hand side of (5.8) is equal to zero for $r=\frac{1}{2}$.

The required sample size is the smallest value of n for which the right hand side of (5.4) is greater than or equal to P^*.

The right hand side of (5.4) can be evaluated numerically by replacing the integral by a sum such as that given by Simpson's rule.

5.1.1.2. Approximation to $Q_S(k, t, s)$ based on asymptotic normality.

The evaluation of $Q_S(k, t, s)$ becomes increasingly tedious as r increases. We therefore investigate how large sample theory can be used to find very good approximations to $Q_S(k, t, s)$ even for relatively small r.

Using the fact (the details of the derivation are the same as in Bartlett and Kendall (1946)) that

$$\left(\frac{q-1}{2}\right)^{\frac{1}{2}}Z_i \qquad (i=1, 2, \cdots, k) \tag{5.9}$$

are independent and asymptotically $N(0, 1)$ as $q\to\infty$, where

$$Z_i=\log\{\bar{X}_i/(a\theta_i)\} \qquad (i=1, 2, \cdots, k) \tag{5.10}$$

and

$$q=2na=2r \tag{5.11}$$

we can obtain a normal approximation to $Q_S(k, t, s)$ as

$$Q_{SA}(k, t, s)=\frac{(k-t)!}{(s-t)!(k-s-1)!}\int_{-\infty}^{\infty}\Phi^t(z+h)\Phi^{k-s-1}(-z)\cdot\Phi^{s-t}(z)\phi(z)dz \tag{5.12}$$

where

$$h=\left(\frac{q-1}{2}\right)^{\frac{1}{2}}\log\delta^*=\left(\frac{2r-1}{2}\right)^{\frac{1}{2}}\log\delta^* \tag{5.13}$$

and Φ and ϕ refer to the cumulative distribution function and the density of the standard normal random variable.

For large values of q an approximate solution to the required sample size can be obtained by equating the right hand side of (5.12) to P^* and using (5.13) and (5.11) to solve for n. Values of h are tabulated in Table 1 of Desu and Sobel (1968) for $P^*=0.90, 0.95, 0.975, 0.99, 0.999$ and selected values of k, t, s $(t\leqq s<k)$, (our h corresponds to their λ).

It is of interest to investigate the magnitude of the error of approximation and how this error varies as a function of δ^*, and r for fixed k, t, s. Tables 1-4 give values of $Q_S(k, t, s)$, (the first figure in each cell) and $Q_{SA}(k, t, s)$, (the second figure in each cell), for selected values of k, t, s, δ^* and r.

It should be noted that $Q_{SA}(k, t, s)$ is fairly close to $Q_S(k, t, s)$ throughout the range of values computed and the normal approximation could therefore be used with very good results to find the required sample size.

The following examples show how the tables are to be used.

EXAMPLE 1. For $k=4$ gamma populations with unknown scale parameters and a common known shape parameter $a=2$, the goal is to select a subset of size $s=3$ which contains the population with the largest scale parameter.

We wish to attain a $P(CS)$ of least $P^*=0.999$ when $\theta_{[4]}/\theta_{[3]}\geqq\delta^*=3.0$. How many observations must be taken from each population?

Table 4 shows that $r=10$ satisfies the requirements.

Hence from (5.6) $na=2n=10$, giving $n=5$ observations from each population. It is interesting to note that use of the normal approximation gives $n=4$ observations from each population.

EXAMPLE 2. Given $k=8$ production processes π_i $(i=1, 2, \cdots, 8)$ such that the life times of components taken from process π_i are distributed according to the gamma distribution with an unknown scale parameter θ_i and a known shape parameter $a=\frac{1}{3}$ the problem is to select a subset of size $s=6$ which contains the $t=4$ processes with the largest mean lives. We wish to attain a $P(CS)$ of at least $P^*=0.990$ when $\theta_{[5]}/\theta_{[4]}\geqq\delta^*=2.0$. How many observations should be taken from each process?

As explained earlier in section 5.1, since the life time distributions of components from the processes have a common shape parameter, the processes with the largest mean lives correspond to those with the largest values of the scale parameters.

To find the required sample size we proceed as follows. We compute

$$\left(\frac{q-1}{2}\right)^{\frac{1}{2}}\log\delta^*=\left(\frac{2na-1}{2}\right)^{\frac{1}{2}}\log 2$$

and set it equal to 2.9565. (The number 2.9565 is obtained from Desu and Sobel (1968),

Table 1, column headed $P^*=0.990$, opposite $k=8$, $t=4$, $s=6$). Solving for n we find that 57 observations from each production process will meet the requirements.

5.1.2. Case B: Shape parameters known and unequal.

If the k gamma populaations have shape parameters $a_i=c_i a_0$ $(i=1,2,\cdots,k)$, where c_i and a_0 are known constants, then it may be desirable to choose the sample sizes n_i $(i=1,2,\cdots,k)$ subject to the restriction

$$n_1c_1=n_2c_2=\cdots=n_kc_k\,.$$

This choice is not most efficient; however it has an important practical advantage, namely, that the tables for Case A, above, then become applicable. We act as if the k populations have the common scale parameter a_0, which is the known constant referred to above; using the method of the previous section we find n_0, the number of observations from each population, where n_0 is not necessarily an integer.

We then set

$$n_ia_i=n_ic_ia_0=n_0a_0 \qquad (i=1,2,\cdots,k) \tag{5.14}$$

from which it follows that we choose n_i as the smallest positive integer for which

$$n_ic_i=n_0 \qquad (i=1,2,\cdots,k)\,. \tag{5.15}$$

Because of (5.14) it is easy to show that the n_i satisfy the requirements.

We briefly illustrate with the following example.

EXAMPLE 3. Given $k=5$ gamma populations with unknown scale parameters and known shape parameters $a_1=a_2=\frac{2}{3}$, $a_3=3$, $a_4=2$, $a_5=4$, the goal is to select a subset of size $s=3$ which contains the best $t=2$ populations. We wish to attain a $P(CS)$ of at least $P^*=0.95$ when $\theta_{[4]}/\theta_{[3]}\geqq\delta^*=1.5$. How many observations should be taken from each population?

We can take $a_0=2$, $c_1=c_2=\frac{1}{3}$, $c_3=\frac{3}{2}$, $c_4=1$, $c_5=2$. Using the method of section 5.1.1, we see that n_0 should be chosen to satisfy the equation

$$\left(\frac{2n_0a_0-1}{2}\right)^{\frac{1}{2}}\log 1.5=2.3321\,, \quad \text{where} \quad a_0=2\,.$$

(The number 2.3321 is obtained from Table 1 of Desu and Sobel (1968)). This gives $n_0=16.788$.

Using (5.15), we find that $n_1=n_2=50.36$, $n_3=11.19$, $n_4=16.79$, $n_5=8.39$.

Thus we select 51, 51, 12, 17, 9 observations from populations 1, 2, 3, 4, and 5 respectively. It should be pointed out that the n_i are independent of the value of a_0.

5.2. Weibull populations with unknown scale parameters all populations having a common known shape parameter.

For the Weibull populations under consideration, the frequency function associated with population π_i is

$$\frac{b}{\theta_i}\left(\frac{x}{\theta_i}\right)^{b-1}\exp\left\{-\left(\frac{x}{\theta_i}\right)^{b}\right\} \qquad x>0$$

where the θ_i (>0) are the unknown scale parameters and b is common known shape parameter.

We take Y_i as $\frac{1}{n}\sum_{j=1}^{n} X_{ij}^b$, where X_{ij} $(j=1, 2, \cdots, n)$ are independent observations from π_i $(i=1, 2, \cdots, k)$. The statistic Y_i is sufficient for the parameter θ_i.

Since $Z_{ij}=X_{ij}^b$ has the exponential distribution with parameter θ_i^b, both the c. d. f. and p. d. f. of Y_i can be obtained from (5.2) and (5.3) respectively by putting $a=1$ and replacing θ_i by θ_i^b.

5.2.1. Exact expression for $Q_S(k, t, s)$.

Using the above results it is easily verified from (4.3) that

$$Q_S(k, t, s) = \frac{(k-t)!}{(s-t)!\,(k-s-1)!}\int_0^{\infty} \{1-H_n(y/d^*)\}^t H_n^{k-s-1}(y)\cdot\{1-H_n(y)\}^{s-t} dH_n(y) \tag{5.16}$$

where $H_n(y)$ is given by (5.7) with r replaced by n, and

$$d^* = \delta^{*b}\,. \tag{5.17}$$

Expression (5.16) can be obtained from (5.4) by replacing δ^* and r by δ^{*b} and n respectively. It follows that values of $Q_S(k, t, s)$ can be obtained from Tables 1-5 by replacing δ^* and r by δ^{*b} and n respectively.

5.2.2. Approximation to $Q_S(k, t, s)$ based on asymptotic normaiity.

The normal approximation to $Q_S(k, t, s)$ is given by

$$Q_{SA}(k, t, s) = \frac{(k-t)!}{(s-t)!\,(k-s-1)!}\int_{-\infty}^{\infty} \Phi^t(y+c)\Phi^{k-s-1}(-y)\cdot\Phi^{s-t}(y)\phi(y)dy \tag{5.18}$$

where

$$c = \{(2n-1)/2\}^{\frac{1}{2}} b \log \delta^*\,. \tag{5.19}$$

Expressions (5.18) and (5.19) can be obtained from (5.12) and (5.13) by replacing δ^* and r by δ^{*b} and n respectively. They can also be obtained by using the fact (see Bartlett and Kendall (1946)) that

$$\left(\frac{2n-1}{2}\right)^{\frac{1}{2}} \log\left\{\frac{1}{n}\sum_{j=1}^{n}(X_{ij}/\theta_i)^b\right\} \qquad i=1, 2, \cdots, k \tag{5.20}$$

are independent and asymptotically $N(0, 1)$ as $n\to\infty$.

For large values of n an approximate value of the required sample size can be obtained by equating the right hand side of (5.18) to P^* and using (5.19) to solve for n. (see Example 2).

§ 6. A dual problem.

The dual problem to that defined in section 2 consists of selecting any s of the t $(\geqq s)$ best populations based on a pre-determined fixed sample size n from each population. In the dual problem a correct selection occurs if and only if

$$\max_{1\leqq i\leqq k-t} Y_{(i)} < \text{the } s\text{-th largest of } \{Y_{(k-t+1)}, \cdots, Y_{(k)}\}\,.$$

By similar methods to those of section 4 it can be shown that the infimum of $P[CS|R]$ over $\Omega(\delta^*)$ is given by

$$\hat{Q}_S(k,\,t,\,s)=\frac{t\,!}{(t-s)\,!\,(s-1)\,!}\int_0^\infty G^{k-t}(x\delta^*)G^{t-s}(x)\cdot\{1-G(x)\}^{s-1}dG(x)\,. \tag{6.1}$$

The required sample size is the smallest value of n for which

$$\hat{Q}_S(k,\,t,\,s)\geqq P^* \tag{6.2}$$

where $(s\leqq t<k)$ and $\binom{t}{s}\Big/\binom{k}{s}<P^*<1$.

It should be noted that the original problem and the dual problem coincide when $t=s$, the problem then reduces to that of selecting the t best populations without regard to order.

§ 7. Applications of the dual problem.

7.1. Weibull populations with unknown scale parameters and a common known shape parameter.

The assumptions are the same as those given in section 5.2.

It is easy to show that $\hat{Q}_S(k,t,s)$ is given by (6.1) with $G(y)$ and δ^* replaced by $H_n(y)$ and δ^{*b} respectively, where $H_n(y)$ is given by (5.7) with r replaced by n.

Using an argument similar to the one in section 5.2.2 for obtaining an approximation to $Q_S(k,t,s)$, we can obtain a normal approximation to $\hat{Q}_S(k,t,s)$ as

$$\hat{Q}_{SA}(k,\,t,\,s)=\frac{t\,!}{(t-s)\,!\,(s-1)\,!}\int_{-\infty}^{\infty}\Phi^{k-t}(y+c)\Phi^{t-s}(y)\cdot\Phi^{s-1}(y)\phi(y)dy \tag{7.1}$$

where c is given by (5.19).

It is readily seen from (5.18) and (7.1) that

$$\hat{Q}_{SA}(k,\,t,\,s)=Q_{SA}(k,\,k-t,\,k-s) \tag{7.2}$$

$$\hat{Q}_{SA}(k,\,t,\,t)=\hat{Q}_{SA}(k,\,k-t,\,k-t)=Q_{SA}(k,\,t,\,t)=Q_{SA}(k,\,k-t,\,k-t) \tag{7.3}$$

(see Lemma 8.1 of Desu and Sobel (1968) for similar results).

The above results make it possible for us to use Table 1 of Desu and Sobel (1968) for both the original and the dual problem.

As an illustration consider the following.

Given $k=7$ production processes π_i $(i=1,2,\cdots,7)$ such that the life times of components taken from process π_i are distributed according to the Weibull distribution with an unknown scale parameter θ_i and a known shape parameter $b=2$, the problem is to select any $s=3$ of the $t=5$ processes with the largest mean lives. We wish to attain a $P(CS)$ of at least $P^*=0.95$ when $\theta_{[3]}/\theta_{[2]}\geqq\delta^*=1.5$. How many observations must be taken from each process?

The mean item life for process π_i is $\theta_i\Gamma\left(1+\frac{1}{b}\right)$. Hence the t processes with the largest mean lives correspond to the t processes with the largest values of the scale parameters.

To find the required sample size we proceed as follows. We compute

$$c=\left(\frac{2n-1}{2}\right)^{\frac{1}{2}}b\log\delta^*=2\left(\frac{2n-1}{2}\right)^{\frac{1}{2}}\log 1.5$$

and set it equal to the c-root for the equation

$$\hat{Q}_{SA}(7,5,3)=P^{*}=0.95. \tag{7.4}$$

From (7.2) the c-root for equation (7.4) is equal to that for the equation $Q_{SA}(7,2,4)=0.95$, which, from Table 1 of Desu and Sobel (1968), is equal to 2.2054. Thus $2\left(\frac{2n-1}{2}\right)^{\frac{1}{2}}\log 1.5=2.2054$. Solving for n we find that 8 observations from each process will meet the requirements.

Corresponding expressions for $\hat{Q}_S(k,t,s)$ and $\hat{Q}_{SA}(k,t,s)$ can be derived for gamma populations with unknown scale parameters, all populations having a common known shape parameter.

Finally, we remark that the treatment in this paper could be applied to other problems, e. g.

(i) The scale parameter problem for the normal distribution with known and unknown mean,

(ii) Ranking uniform distributions with unknown upper end points and common lower end-point equal to zero.

The procedure R could also be applied to the case in which the bestness of a population is characterized by its location parameter; this will be considered elsewhere.

Acknowledgements

I should like to thank Professor G. B. Wetherill and Mr. C. M. Theobald for their assistance during the preparation of this paper.

References

[1] M.S. BARTLETT AND D.G. KENDALL, *The statistical analysis of variance-heterogeneity and the logarithmic transformation.* J. Roy. Statist. Soc. Suppl., **8** (1946), 128–138.

[2] R.E. BECHHOFER, *A single-sample multiple decision procedure for ranking means of normal populations with known variances.* Ann. Math. Stat., **25** (1954), 16–29.

[3] R.E. BECHHOFER AND M. SOBEL, *A single-sample mutiple decision procedure for ranking variances of normal populations.* Ann. Math. Stat., **29** (1954), 235–244.

[4] M.M. DESU AND M. SOBEL, *A fixed subset-size approach to the selection problem.* Biometrika, **55** (1968), 401–410.

[5] D.M. MAHAMUNULU, *Some fixed sample ranking and selection problems.* Ann. Math. Stat., **38** (1967), 1079–91.

Table 1-2. Values of $Q_S(k, t, s)$ (top) and $Q_{SA}(k, t, s)$ (bottom) for the scale parameter problem for gamma distributions with known shape parameters. Use (5.16)-(5.19) for the Weibull case, and (7.2)-(7.3) for the dual problem.

Table 1. $k=3,\ t=1,\ s=1$

r/δ^*	1.0	1.5	2.0	2.5	3.0
1	0.3333	0.4500	0.5333	0.5952	0.6428
	0.3333	0.4174	0.4796	0.5281	0.5675
2	0.3333	0.5050	0.6239	0.7065	0.7654
	0.3333	0.4815	0.5892	0.6684	0.7280
3	0.3333	0.5459	0.6873	0.7787	0.8388
	0.3333	0.5260	0.6611	0.7536	0.8176
4	0.3333	0.5795	0.7362	0.8301	0.8870
	0.3333	0.5619	0.7155	0.8129	0.8747
5	0.3333	0.6084	0.7754	0.8681	0.9198
	0.3333	0.5925	0.7588	0.8563	0.9127
6	0.3333	0.6339	0.8077	0.8969	0.9426
	0.3333	0.6194	0.7942	0.8887	0.9387
7	0.3333	0.6568	0.8346	0.9190	0.9587
	0.3333	0.6434	0.8236	0.9134	0.9567
8	0.3333	0.6775	0.8572	0.9361	0.9701
	0.3333	0.6651	0.8483	0.9323	0.9692
9	0.3333	0.6964	0.8764	0.9494	0.9784
	0.3333	0.6849	0.8692	0.9469	0.9781
10	0.3333	0.7138	0.8928	0.9599	0.9843
	0.3333	0.7031	0.8869	0.9583	0.9844
11	0.3333	0.7298	0.9069	0.9681	0.9885
	0.3333	0.7199	0.9021	0.9671	0.9888
12	0.3333	0.7447	0.9190	0.9746	0.9916
	0.3333	0.7355	0.9151	0.9741	0.9920
13	0.3333	0.7586	0.9294	0.9797	0.9939
	0.3333	0.7499	0.9263	0.9795	0.9942
14	0.3333	0.7715	0.9385	0.9838	0.9955
	0.3333	0.7634	0.9359	0.9838	0.9958
15	0.3333	0.7835	0.9463	0.9870	0.9967
	0.3333	0.7760	0.9442	0.9871	0.9970
16	0.3333	0.7948	0.9531	0.9896	0.9976
	0.3333	0.7878	0.9514	0.9898	0.9978
17	0.3333	0.8054	0.9590	0.9917	0.9982
	0.3333	0.7988	0.9577	0.9919	0.9984
18	0.3333	0.8154	0.9641	0.9933	0.9987
	0.3333	0.8092	0.9631	0.9936	0.9989

Table 2. $k=4,\ t=1,\ s=1$

r/δ^*	1.0	1.5	2.0	2.5	3.0
1	0.2500	0.3682	0.4572	0.5252	0.5786
	0.2500	0.3290	0.3903	0.4397	0.4808
2	0.2500	0.4234	0.5526	0.6460	0.7142
	0.2500	0.3923	0.5039	0.5904	0.6579
3	0.2500	0.4655	0.6218	0.7275	0.7991
	0.2500	0.4376	0.5822	0.6876	0.7636
4	0.2500	0.5007	0.6764	0.7873	0.8564
	0.2500	0.4750	0.6435	0.7580	0.8339
5	0.2500	0.5315	0.7213	0.8325	0.8965
	0.2500	0.5075	0.6936	0.8110	0.8823
6	0.2500	0.5590	0.7588	0.8674	0.9249
	0.2500	0.5364	0.7355	0.8516	0.9161
7	0.2500	0.5839	0.7905	0.3947	0.9454
	0.2500	0.5627	0.7709	0.3831	0.9400
8	0.2500	0.6067	0.8177	0.9161	0.9601
	0.2500	0.5867	0.8011	0.9077	0.9570
9	0.2500	0.6278	0.8410	0.9330	0.9708
	0.2500	0.6089	0.8270	0.9270	0.9691
10	0.2500	0.6473	0.8611	0.9464	0.9786
	0.2500	0.6294	0.8494	0.9421	0.9778
11	0.2500	0.6655	0.8785	0.9571	0.9843
	0.2500	0.6486	0.8687	0.9541	0.9840
12	0.2500	0.6825	0.8937	0.9656	0.9884
	0.2500	0.6665	0.8854	0.9635	0.9884
13	0.2500	0.6984	0.9068	0.9724	0.9915
	0.2500	0.6833	0.8999	0.9710	0.9916
14	0.2500	0.7133	0.9183	0.9778	0.9937
	0.2500	0.6990	0.9125	0.9770	0.9940
15	0.2500	0.7274	0.9283	0.9822	0.9954
	0.2500	0.7139	0.9235	0.9817	0.9956
16	0.2500	0.7407	0.9370	0.9857	0.9966
	0.2500	0.7278	0.9330	0.9854	0.9968
17	0.2500	0.7532	0.9447	0.9885	0.9975
	0.2500	0.7410	0.9414	0.9884	0.9977
18	0.2500	0.7650	0.9514	0.9907	0.9981
	0.2500	0.7535	0.9486	0.9907	0.9983

Table 3-4. Values of $Q_S(k, t, s)$ (top) and $Q_{SA}(k, t, s)$ (bottom) for the scale parameter problem for gamma distributions with known shape parameters. Use (5.16)-(5.19) for the Weibull case, and (7.2)-(7.3) for the dual problem.

Table 3. $k=4,\ t=1,\ s=2$

r/δ^*	1.0	1.5	2.0	2.5	3.0
1	0.5000 0.5000	0.6135 0.5942	0.6856 0.6581	0.7352 0.7049	0.7713 0.7407
2	0.5000 0.5000	0.6680 0.6601	0.7664 0.7598	0.8275 0.8245	0.8677 0.8682
3	0.5000 0.5000	0.7066 0.7029	0.8184 0.8188	0.8811 0.8857	0.9184 0.9255
4	0.5000 0.5000	0.7371 0.7358	0.8557 0.8594	0.9158 0.9228	0.9481 0.9561
5	0.5000 0.5000	0.7624 0.7626	0.8837 0.8891	0.9394 0.9468	0.9664 0.9736
6	0.5000 0.5000	0.7839 0.7853	0.9055 0.9116	0.9559 0.9629	0.9779 0.9838
7	0.5000 0.5000	0.8026 0.8049	0.9226 0.9290	0.9676 0.9738	0.9854 0.9900
8	0.5000 0.5000	0.8191 0.8220	0.9363 0.9426	0.9761 0.9814	0.9903 0.9938
9	0.5000 0.5000	0.8337 0.8370	0.9474 0.9534	0.9822 0.9867	0.9935 0.9961
10	0.5000 0.5000	0.8468 0.8505	0.9563 0.9620	0.9868 0.9905	0.9956 0.9975
11	0.5000 0.5000	0.8586 0.8625	0.9637 0.9690	0.9901 0.9932	0.9970 0.9984
12	0.5000 0.5000	0.8692 0.8734	0.9697 0.9746	0.9926 0.9951	0.9980 0.9990
13	0.5000 0.5000	0.8789 0.8832	0.9747 0.9791	0.9944 0.9964	0.9986 0.9994
14	0.5000 0.5000	0.8877 0.8921	0.9788 0.9828	0.9958 0.9974	0.9991 0.9996
15	0.5000 0.5000	0.8958 0.9002	0.9822 0.9858	0.9968 0.9981	0.9994 0.9997
16	0.5000 0.5000	0.9032 0.9077	0.9851 0.9883	0.9976 0.9986	0.9996 0.9998
17	0.5000 0.5000	0.9100 0.9144	0.9875 0.9903	0.9982 0.9990	0.9997 0.9999
18	0.5000 0.5000	0.9162 0.9206	0.9895 0.9920	0.9986 0.9993	0.9998 0.9999
19	0.5000 0.5000	0.9219 0.9263	0.9911 0.9933	0.9990 0.9995	0.9999 0.9999

Table 4. $k=4,\ t=1,\ s=3$

r/δ^*	1.0	1.5	2.0	2.5	3.0
1	0.7500 0.7500	0.7183 0.8178	0.7572 0.8582	0.7824 0.8851	0.8001 0.9042
2	0.7500 0.7500	0.8525 0.8594	0.9032 0.9139	0.9371 0.9439	0.9493 0.9617
3	0.7500 0.7500	0.8755 0.8840	0.9302 0.9414	0.9572 0.9683	0.9720 0.9818
4	0.7500 0.7500	0.8928 0.9017	0.9802 0.9583	0.9721 0.9810	0.9838 0.9907
5	0.7500 0.7500	0.9064 0.9153	0.9605 0.9695	0.9813 0.9882	0.9903 0.9951
6	0.7500 0.7500	0.9176 0.9262	0.9695 0.9773	0.9873 0.9925	0.9941 0.9073
7	0.7500 0.7500	0.9270 0.9352	0.9763 0.9829	0.9912 0.9952	9.9964 0.9985
8	0.7500 0.7500	0.9349 0.9428	0.9813 0.9870	0.9939 0.9968	0.9977 0.9992
9	0.7500 0.7500	0.9417 0.9492	0.9852 0.9900	0.9957 0.9979	0.9986 0.9995
10	0.7500 0.7500	0.9477 0.9548	0.9883 0.9923	0.9970 0.9986	0.9991 0.9997
11	0.7500 0.7500	0.9528 0.9595	0.9906 0.9940	0.9978 0.9991	0.9994 0.9998
12	0.7500 0.7500	0.9574 0.9637	0.9925 0.9953	0.9985 0.9994	0.9996 0.9999
13	0.7500 0.7500	0.9614 0.9674	0.9939 0.9963	0.9989 0.9996	0.9998 0.9999
14	0.7500 0.7500	0.9650 0.9706	0.9951 0.9971	0.9992 0.9997	0.9998 1.0000
15	0.7500 0.7500	0.9682 0.9735	0.9960 0.9977	0.9994 0.9998	0.9999 1.0000
16	0.7500 0.7500	0.9710 0.9760	0.9968 0.9982	0.9996 0.9999	0.9999 1.0000
17	0.7500 0.7500	0.9736 0.9782	0.9974 0.9986	0.9997 0.9999	1.0000 1.0000
18	0.7500 0.7500	0.9759 0.9803	0.9979 0.9989	0.9998 0.9999	1.0000 1.0000
19	0.7500 0.7500	0.9779 0.9821	0.9983 0.9991	0.9998 1.0000	1.0000 1.0000