Ghana J. Sci. **29-30**, 37-44

ON A PROCEDURE FOR SELECTION FROM GAMMA AND WEIB POPULATIONS

J. B. OFOSU
Department of Statistics, King Saud University, Saudi Arabia

Summary

Given k gamma populations with unknown scale parameters and known shape parameters, the problem is to find a procedure for selecting the populations with the t largest scale parameters. Exact and asymptotic methods for finding the sample sizes required to satisfy the δ^*, P^* requirement of Bechhofer (1954) are discussed. The results are applied to the problem of selecting the populations with the t largest scale parameters from k Weibull distributions with unknown scale parameters and known shape parameters.

Introduction

Let $\pi_1, \ldots\ldots, \pi_k$ denote k gamma populations, where the p.d.f. of population π_i may be written as

$$f(x,\theta_i) = \frac{1}{\theta_i}\left(\frac{x}{\theta_i}\right)^{a_i-1}\frac{1}{\Gamma(a_i)}\exp(-x/\theta_i) \qquad x>0. \tag{1.1}$$

It is assumed that the shape parameters $a_i(>0)$ are known but the scale parameters θ_i are unknown. The ordered θ-values and the associated vector are denoted respectively by

$$0<\theta_{[1]} \le \theta_{[2]} \le \ldots\ldots \le \theta_{[k]} < \infty$$

and

$$\theta = \{\theta_{[1]}, \theta_{[2]}, \ldots\ldots, \theta_{[k]}\}.$$

It is not known which population is associated with $\theta_{[i]}$ (i = 1, 2, ..., k). The t populations associated with $\theta_{[k-t+1]}, \ldots\ldots, \theta_{[k]}$ are defined as the t "best" populations.

Let Ω stand for the parameter space, which is the set of all admissible vectors θ. Following Bechhofer (1954), we partition the parameter space into a "preference-zone" $\Omega(\delta^*)$ defined by

$$\Omega(\delta^*) = \{\theta : \theta_{[k-t+1]}/\theta_{[k-t]} \ge \delta^* > 1\} \tag{1.2}$$

and its complement, the "indifference-zone". Our goal is to select the best t populations based on a pre-determined sample size n_i from population π_i (i = 1, 2,..........., k). In sections 2 and 3 we discuss a procedure R for which the probability of a correct selection (cs) satisfies the requirement

$$P(cs|R) \ge P^* \quad \text{for all } \theta \in \Omega(\delta^*) \tag{1.3}$$

where P^* $(1/\binom{k}{t} < P^* < 1)$ and δ^* are specified in advance by the experimenter. This requirement is called the (δ^*, P^*) requirement. In section 4 we apply procedure R to the problem of selecting the best t of k Weibull distributions.

The indifference-zone formulation considered in this paper has been applied to the normal distribution by several authors among whom are Bechhofer (1954), Bechhofer & Sobel (1954), Bechhofer, Dunnet & Sobel (1954), Ofosu (1973) Wetherill & Ofosu (1974), Dudewicz & Dalal (1975), Tong & Wetzell (1979), Hayre & Gittins (1981) and Driessen (1988).

The formulation has also been applied to other distributions by Sobel & Huyett (1957), Barr & Rizvi (1966), Alam (1971), Feigin & Weissman (1981), Giani (1986) and Mishra (1986). The formulation has not yet been applied to the gamma and Weibull distributions. These distributions have applications in reliability, life testing and fatigue testing. Each of these distributions can also be regarded as a different generalization of the exponential distribution for a particular value of one of its parameters. The results of this paper will, therefore, have applications in the above areas.

A minimax solution to the problem considered in this paper has been proposed by Ofosu (1972) for the case $t = 1$.

Experimental

The experimenter takes a pre-determined number $n_i = n_i(k, t, \delta^*, P^*)$ of independent observations from population π_i $(i = 1, 2, \ldots, k)$. Let X_i denote the mean of the n_i observations from π_i and let $Y_i = X_i/a_i$ $(i = 1, 2, \ldots, k)$. The statistic Y_i is sufficient for the parameter θ_i. Procedure R is based on the k Y-values and is defined as follows:

Procedure R

Select the populations which give rise to the t largest Y - values. Once the sample sizes $n_1, n_2, \ldots,$ n_k are specified, the procedure is completely defined. Our problem is to determine the sample sizes so that (1.3) is satisfied.

Probability of a correct selection and its infimum over $\Omega(\delta^*)$

Let $Y_{[i]}$ denote the Y-value from the population with parameter $\theta_{[i]}$ $(i = 1, 2, \ldots, k)$. A correct selection occurs if

$$\min(Y_{[k-t+1]}, \ldots, Y_{[k]}) > \max(Y_{[1]}, \ldots, Y_{[k-t]}).$$

Thus, the probability of a correct selection using procedure R is given by

$$P(cs|R) = P\{\min(Y_{[k-t+1]}, \ldots, Y_{[k]}) > \max(Y_{[1]}, \ldots, Y_{[k-t]})\}. \quad (3.1)$$

Now, the distribution function of $Y_{[i]}$ is given by

$$H(y \mid \theta_{[i]}) = \int_0^y h(x \mid \theta_{[i]})dx, \quad (3.2)$$

where

$$h(x \mid \theta_{[i]}) = \frac{v_i}{\theta_{[i]}} \frac{1}{\Gamma(v_i)} (v_i x/\theta_{[i]})^{v_i - 1} \exp(-v_i x / \theta_{[i]}) \; x > 0 \quad (3.3)$$

and $\quad v_i = n_i a_i.$ (3.4)

The family of distribution functions $\{H(y \mid \theta), 0 < \theta < \infty\}$ is a stochastically increasing family of aboslutely continuous distribution functions. Hence by the monotonicity results of the the first theorem of Mahamunulu (1967), the infimum over $\Omega(\delta^*)$ of $P(cs|R)$ occurs when

$$\theta_{[1]} = \theta_{[2]} = \ldots\ldots = \theta_{[k-t]} = \theta, \text{ say} \quad (3.5)$$

and

$$\theta_{[k-t+1]} = \ldots\ldots = \theta_{[k]} = \delta^*\theta \quad (3.6)$$

Case (a): Shape parameters known and equal

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

Let

$$Q_G(k, t) = \inf_{\theta \varepsilon \Omega(\delta^*)} P(cs \mid R) \tag{3.7}$$

Exact expression for Q_G (k, t). Using (3.5) and (3.6) in (3.1), we obtain the following two equivalent expressions for Q_G (k, t).

$$Q_G(k, t) = t \int_0^\infty G_V^{k-t}(x\delta^*)\ \{1 - G_V(x)\}^{t-1} g_V(x)\,dx \tag{3.8}$$

$$= (k - t) \int_0^\infty \{1 - G_V(x/\delta^*\}^t G_V^{k-t-1}(x)\, g_V(x)\,dx, \tag{3.9}$$

where

$$G_V(y) = \int_0^y g_V(x)\,dx \tag{3.10}$$

$$g_V(x) = \frac{1}{\Gamma(v)} x^{v-1} e^{-x} \qquad x > 0 \tag{3.11}$$

and

$$v = na \tag{3.12}$$

For given values of k, t, δ^* ($\delta^* > 1$) and P^* ($1/\binom{k}{t} < P^* < 1$) the required sample size is the smallest value of n for which $Q_G(k, t) \geq P^*$. The function Q_G (k,t) can be evaluated numerically. For integral values of v, G_V (x) can be expressed as

$$G_V(x) = 1 - \sum_{j=0}^{v-1} \frac{e^{-x} x^j}{j!} \qquad x > 0 \tag{3.13}$$

while for half integral values of v,

$$G_V(x) = \operatorname{erf}(\sqrt{x}) - \frac{\sqrt{x}\, e^{-x}}{\sqrt{\pi}} \sum_{j=1}^{v-\frac{1}{2}} \frac{2^j x^{j-1}}{1.3\ldots(2j-1)} \tag{3.14}$$

where

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \tag{3.15}$$

and the second term on the right hand side of (3.14) is equal to zero for $v = \frac{1}{2}$.

Table 1 gives values of Q_G (k, t) for k = 3 and t = 1.

Example 1

For k = 3 gamma populations with unknown scale parameters and a common known shape parameter a = ½ , the goal is to select the population with the largest scale parameter. We wish to attain a P(cs) of at least $P^* = 0.90$ when $\theta_{[3]}/\theta_{[2]} \geq \delta^* = 2$. How many observations must be taken from each population? Table 1 shows that v = 11 satisfies the requirement (1.3). Hence, from (3.12), na = = 11 , giving n = 22 observations from each population.

Asymptotic expression for Q_G (k,t)

The evaluation of Q_G (k, t) becomes increasingly tedious as v increases. The following theorem gives a normal approximation to Q_G(k, t).

Theorem 1

For large values of v, Q_G (k, t) can be approximated by

$$Q_{GA}(k, t) = (k - t) \int_{-\infty}^{\infty} \Phi^{t}(u + \lambda)\, \Phi^{k-t-1}(-u)\, \phi(u)\, du, \quad (3.16)$$

where $\lambda = \sqrt{\{(2v - 1)/2\}} \ln \delta^* = \sqrt{\{(2na - 1)/2\}} \ln \delta^*$ (3.17)

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

The proof is based on the result that $\phi \{(2v - 1)/2\} \ln \{\bar{x}_i / (a\theta)_i\}$ (i = 1, 2,....., k) are independent and asymptotically N (0,1) as $v \to \infty$. The method of proof is similar to that gien by Bartlett & Kendall (1946) and Ofosu (1975), and is, therefore, omitted.

For large values of v, an approximate sample size can be obtained by equating Q_{GA}(k, t) to P* and using (3.17) to solve for n. Equation (3.16) is a special case of Equation (5.1) of Desu & Sobel (1968). Values of λ can, therefore, be obtained from their Table 1 for P* = 0.900, 0.950, 0.975, 0.999 and for selected values of k and t.

It is of interest to investigate the magnitude of the error of approximation and how this error varies as a function of δ^* and v for fixed k and t. Table 1 gives values of Q_G(k, t) (the first figure in each cell) and Q_{GA}(k, t) (the second figure in each cell) for k=3 and t=1 and for selected values of v and δ^*

From Table 1, it appears that the normal approximation can be used with very good results to find the required sample size. For instance, in Example 1, both the exact and the asymptotic values for Q_G(k, t) give the same solution.

Example 2

Given k=7 production processes π_i (i = 1, 2,........,7) such that the life time 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 = 4, the problem is to select the t=3 processes with the largest mean lives. We wish to attain a P(cs) of at least P* = 0.975 when $\theta_{[5]} / \theta_{[4]} \geq \delta^* = 1.4$. How many observations should be taken from each process?

The mean item life for process π_i is $a\theta_i$. It follows that the problem of selecting the three processes with the largest mean lives is equivalent to that of selecting the processes associated with the three largest scale parameters. To find the required sample size, we proceed as follows. We compute

$$\sqrt{\{(2v - 1)/2)\}} \ln \delta^* = \sqrt{\{(2na - 1)/2)\}} \ln 1.4$$

and set it equal to 3.9679 (The number 3.9679 is obtained from Desu & Sobel (1968), Table 1, column headed P* = 0.975, opposite k=7, t=s=3). Solving for n we find that 35 observations from each production process will meet the requirements.

Case (b): Shape parameters known and unequal

We now consider the case where the shape parameters a_i are not equal. In this case, even-though $Y_{[1]}, \ldots, Y_{[k]}$ are independent, $Y_{[1]}, \ldots, Y_{[k-t]}$ are not always identically distributed when (3.5) holds and $Y_{[k-t+1]}, \ldots, Y_{[k]}$ are also not always identically distributed when (3.6) holds. This makes the determination of the infimum of P(cs|R) over $\Omega(\delta^*)$ very difficult. However, it can be seen from (3.3) and (3.4) that if $n_1, \ldots, n_k$ are chosen such that

$$n_1 a_1 = n_2 a_2 = \ldots = n_k a_k \quad (3.18)$$

then $Y_{[1]}, \ldots, Y_{[k-t]}$ are identically distributed when (3.5) holds and $Y_{[k-t+1]}, \ldots, Y_{[k]}$ are also identically distributed when (3.6) holds. This choice is not the most efficient. However, it has an important practical advantage, namely, that the results of case (a) then become applicable. We act as if the k populations have a common known scale parameter and obtain v_o, the value of v for which

$$Q_G(k, t) \geq P^*$$

where Q_G (k,t) is given by (3.8) or (3.9) when v is small and by (3.16) when v is large. We then choose n_i as the smallest positive integer for which

$$n_i a_i = v_0 \quad (i = 1, 2, ..., k). \tag{3.19}$$

We briefly illustrate with an example.

Example 3

Given k=5 gamma populations with unknown scale parameters and known shape parameters $a_1 = 2$, $a_2 = 3$, $a_3 = 4$, $a_4 = ½$ and $a_5 = 1$ the goal is to select the best t=2 populations. We wish to attain a P(cs) of at least P*=0.95 when $\theta_{[4]} / \theta_{[3]} \geq 1.5$. How many observations should be taken from each population?

Using section 3.1 with k = 5, t = 2 and δ* = 1.5 , we see that v_o should be chosen to satisfy the equation

$$\sqrt{\{(2v_0 - 1)/2\}} \ \text{In}\ 1.5 = 3.2805.$$

The number 3.2805 is obtained from Table 1 of Desu & Sobel (1968). This gives

$$v_o = 65.95.$$

Using this in (3.19), we find that n_1, n_2, n_3, n_4 and n_5 are given by

$$2n_1 = 3n_2 = 4n_3 = ½\, n_4 = n_5 = 65.95.$$

Thus, we select 33, 22, 17, 132 and 66 observations from populations π_1, π_2, π_3, π_4 and π_5, respectively.

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

For the Weibull populations under consideration, the p.d.f. associated with population π_i is

$$\frac{b}{\theta_i}\left(\frac{x}{\theta_i}\right)^{b-1} \exp\{-(x/\theta_i)^b\} \qquad x > 0 \tag{4.1}$$

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

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

Exact and asymptotic expressions for the infimum of P(cs|R) over Ω (δ)*

Since X^b_{ij} has the exponential distribution with parameter θ^b_i both the c.d.f. and p.d.f of $Y_{[i]}$ can be obtained from (3.2) and (3.3) by putting a=1 and replacing $\theta_{[i]}$ by $\theta^b_{[i]}$. The exact and the asymptotic expressions of the infimum of P(csR) over Ω (δ*) for the Weibull case can, therefore, be obtained from the corresponding expressions for the gamma case by replacing δ* and v by δ^{*b} and n, respectively. Thus, for given values of k, t, δ* and P*, the required sample size is the smallest value of n for which

$$Q_W(k, t) \geq P^*,$$

where $Q_W(k, t)$ is given by either of the following equivalent equations

$$Q_W(k, t) = t \int_0^\infty G_n^{k-t}(xd^*)\{1 - G_n(x)\}^{t-1} g_n(x)\, dx \quad (4.2)$$

$$= (k-t) \int_0^\infty \{1 - G_n(x/d^*)\}^t G_n^{k-t-1}(x) g_n(x) dx, \quad (4.3)$$

where $G_n(x)$ is given by (3.13) with v replaced by n and

$$d^* = \delta^{*b} \quad (4.4)$$

An asymptotic expression for $Q_W(k,t)$ is given by

$$Q_{WA}(k, t) = (k-t) \int_{-\infty}^{\infty} \phi^t(u+c)\, \phi^{k-t-1}(-u)\, \phi(u)\, du, \quad (4.5)$$

where

$$c = \sqrt{\{(2n - 1)/2\}}\ b\ \ln \delta^* \quad (4.6)$$

For large values of n, an approximate value of the required sample size can be obtained by equating $Q_{WA}(k,t)$ to P* and using (4.6) to solve for n. Table 1 of Desu & Sobel (1968) is again applicable.

Example 4

For k=6 production processes π_i (i=1,2,...,6) 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 the t=2 processes with the largest mean lives. We wish to attain a P(cs) of at least P* = 0.95 when $\theta_{[5]}/\theta_{[4]} \geq 1.3$. How many observations must be taken from each process?

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

To find the required sample size, we compute

$$c = \sqrt{\{(2n-1)/2\}}\ b \ln \delta^*$$
$$= 2\sqrt{\{(2n-1)/2\}} \ln 1.3$$

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

$$Q_{WA}(6,2) = 0.95,$$

which, from Table 1 of Desu & Sobel (1968), is equal to 3.4154. Thus, *n* is given by

$$2\sqrt{\{(2n-1)/2\}}\ \ln 1.3 = 3.4154.$$

Solving for *n*, we find that 43 observations from each process will meet the requirements.

References

ALAM, K. (1971) Selection from Poisson processes. *Ann. Inst. statist. Math.* **23**, 411-418.

BARR, D. R. & RIZVI, M. H. (1966) Ranking and selection problems of uniform distributions. *Trabajos Estadist.* **17**, 15-31.

BARTLETT, M. S. & KENDALL, D. G. (1946) The statistical analysis of variance-heterogeneity and the logarithmic trnasformation. *J. Roy. Statist. Soc. suppl.* **8**, 128-138.

BECHHOFER, R. E. (1954) A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. math. Statist.* **25**, 16-29.

BECHHOFER, R. E. & SOBEL, M. (1954) A single-sample multiple-decision procedure for ranking variances of normal populations. *Ann. math. Statist.* **25**, 273-289.

BECHHOFER, R. E., DUNNETT, C. W. & SOBEL, M, (1954) A two-sample multiple-decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika* **41**, 170-176.

TABLE 1

Values of Q_G (k, t) (top) and $Q_{G,I}$ (k, t) (bottom) for the scale parameter problem for gamma distributions with known shape parameters; k=3, t=1

ν/δ*	1.0	1.5	2.0	2.5	3.0
1	0.3333 0.3333	0.4500 0.4174	0.5333 0.4796	0.5952 0.5281	0.6428 0.5675
2	0.3333 0.3333	0.5050 0.4815	0.6239 0.5892	0.7065 0.6684	0.7654 0.7280
3	0.3333 0.3333	0.5459 0.5260	0.6873 0.6611	0.7787 0.7537	0.8388 0.8176
4	0.3333 0.3333	0.5795 0.5619	0.7362 0.7155	0.8301 0.8129	0.8870 0.8747
5	0.3333 0.3333	0.6084 0.5925	0.7754 0.7588	0.8681 0.8563	0.9198 0.9127
6	0.3333 0.3333	0.6339 0.6194	0.8077 0.7942	0.8969 0.8887	0.9426 0.9387
7	0.3333 0.3333	0.6568 0.6434	0.8346 0.8236	0.9190 0.9134	0.9587 0.9567
8	0.3333 0.3333	0.6775 0.6651	0.8572 0.8483	0.9361 0.9323	0.9701 0.9692
9	0.3333 0.3333	0.6964 0.6849	0.8764 0.8692	0.9494 0.9469	0.9784 0.9781
10	0.3333 0.3333	0.7138 0.7031	0.8928 0.8869	0.9599 0.9583	0.9843 0.9844
11	0.3333 0.3333	0.7298 0.7199	0.9069 0.9021	0.9681 0.9671	0.9885 0.9888
12	0.3333 0.3333	0.7447 0.7355	0.9190 0.9151	0.9746 0.9741	0.9916 0.9920
13	0.3333 0.3333	0.7586 0.7499	0.9294 0.9263	0.9797 0.9795	0.9939 0.9942
14	0.3333 0.3333	0.7715 0.7634	0.9385 0.9359	0.9838 0.9838	0.9955 0.9958
15	0.3333 0.3333	0.7835 0.7760	0.9463 0.9442	0.9870 0.9871	0.9967 0.9970
16	0.3333 0.3333	0.7948 0.7878	0.9531 0.9514	0.9896 0.9898	0.9976 0.9978
17	0.3333 0.3333	0.8054 0.7988	0.9590 0.9577	0.9917 0.9919	0.9982 0.9984
18	0.3333 0.3333	0.8154 0.8092	0.9641 0.9631	0.9933 0.9936	0.9987 0.9989

Desu, M. M. & Sobel, M. (1968) A fixed subset-size approach to a selection problem. *Biometrika* **55**, 401-410. Corrections and amendments: **63** (1976), 685.

Driessen, S. G. A. J. (1988) The probability of correct selection. *Biometrika*, **75**, 618-619.

Dudewicz, E. J. & Dalal, S. R. (1975) Allocation of observations in ranking and selection with unequal variances. *Sankhya*. Ser. B **37**, 28-78.

Feigin, P. D. & Weissman, I. (1981) On the indifference-zone approach to selection: a consistency result. *Analyzing Time Series* **33**, 471-474.

Giani, G. (1986) On a strengthening of the indifference-zone approach to a generalized selection goal. *Comm. Statist.* **15**, 3163-3171.

Hayre, L.S. & Gittins, J. C. (1981) A sequential selection of the larger of two normal means. *J. Am. statist. Ass.* **76**, 696-699.

Mahamunulu, D. M. (1967) Some fixed-sample ranking and selection problems. *Ann. math. Statist.* **38**, 1079-1091.

Mishra, S. N. (1986) Simultaneous selection of extreme population mean: Indifference-zone formulation. *Am. J. math. Manag. Sci.* **6**. 131-142.

Ofosu, J. B. (1972) A minimax procedure for selecting the population with the largest (smallest) scale parameter. *Calcutta Statist. Assoc. Bull.* **21**, 143-154.

Ofosu, J. B. (1973) A two-sample procedure for selecting the population with the largest mean from several normal populations with unknown variances. *Biometrika* **60**, 117-124.

Ofosu, J. B. (1975) A two-sample procedure for selecting the population with the largest mean from several normal populations with unknown variances. Correction. *Biometrika* **62**, 221.

Ofosu, J. B. (1975) On selecting a subset which includes the t best of k populations: scale parameter case. *Bull. math. Statist.* **16**, 7-18.

Raghavachari, M. & Starr, N. (1970) Selection problems for some terminal distributions. *Metron.* **28**, 185-197.

Sobel, M. & Huyett, M. J. (1957) Selecting the best one of several binomial populations. *Bell System Tech. J.* **36**, 537-576.

Tong, Y. L. & Wetzell, D. E. (1979) On the behaviour of the probability function for selecting the best normal populations. *Biometrika* **66**, 174-176.

Wetherill, G. B. & Ofosu, J. B. (1974) Selection of the best of k normal populations. *J. Roy. statist. Soc.* C **23**, 253-277.

Received 19 Aug 87; revised 10 Jul 89.