〈Research Paper〉

Journal of the Korean
Statistical Society
Vol. 16, No. 1, 1987

A Two-stage Selection Procedure for Exponential Populations[+]

Kyung Soo Han* and Woo-Chul Kim*

ABSTRACT

A two-stage selection procedure is considered in the case of exponential populations with common known scale parameter. The proposed procedure is designed following the lines of Tamhane and Bechhofer (1977). The design constants to implement the procedure are provided. Monte Carlo results show that the proposed procedure performs better than the single stage procedure by Raghavachari and Starr (1970) in terms of the expected total sample size.

1. Introduction

In statistical problems related to life time, the exponential distribution is often used as a model. Barr and Rizvi (1966), Desu and Sobel (1968), and Raghavachari and Starr (1970) considered the problem of selecting the exponential population with the largest location parameter out of several exponential populations in the case of common known scale parameter. Desu, Narula and Villarreal (1977), and Lee and Kim (1985) considered the same problem except that the common scale parameter is unknown.

The procedures in these articles have been designed only to control the probability of correctly selecting the "best" population following the indifference-zone approach of Bechhofer (1954). The purpose of this article is to devise a selection procedure, in the case of common known scale parameter, which is optimal in some sense.

* Department of Computer Science and Statistics, Seoul National University, Seoul 151, Korea.

[+] Research supported in part by MOE through the Research Institute for Basic Sciences, SNU, 1986~1987.

The proposed procedure is designed following the lines of Tamhane and Bechhofer (1977). The design constants to implement the procedure are provided in Section 3. Monte Carlo results in Section 4 reveal that the proposed procedure performs better than the single stage procedure by Raghavachari and Starr (1970).

2. A procedure and a design criterion

Let π_i denote the exponential population with the unknown location parameter $\theta_i (1 \leq i \leq k)$ and the common known scale parameter σ. The problem of interest is in identifying the population with the largest location parameter $\theta_{[k]} = \max_{1 \leq i \leq k} \theta_i$ where $\theta_{[1]} \leq \theta_{[2]} \leq \cdots \leq \theta_{[k]}$ denote the ordered θ_i's. We may remark that θ_i has the interpretation of the guaranteed life time in practice. Also we may assume that $\sigma = 1$.

Adopting the idea of Tamhane and Bechhofer (1977) for the normal means problem, we consider the following two-stage selection procedure $\mathscr{P}_2$.

At the first stage, we take n_1 independent observations $X_{ij} (j=1, \cdots, n_1)$ from each $\pi_i (i=1, \cdots, k)$ and compute $T_i^{(1)} = \min_{1 \leq j \leq n_1} X_{ij}$. Then we determine a subset S of $\{\pi_1, \cdots, \pi_k\}$ by

$$S = \{\pi_i \mid T_i^{(1)} \geq \max_{1 \leq j \leq k} T_j^{(1)} - d\}$$

where $d > 0$ is a design constant to be determined. If S has only one population, then stop further sampling and assert the population in S as the best. If S has more than one population, go to the second stage.

At the second stage, we take n_2 additional observations $Y_{ij} (j=1, \cdots, n_2)$ from each π_i in S and compute $T_i^{(2)} = \min_{1 \leq j \leq n_2} Y_{ij}$ for each π_i in S. Then we assert the population associated with the largest $T_i = \max_{\pi_j \in S} T_j$ as the best where $T_i = \min(T_i^{(1)}, T_i^{(2)})$.

In implementing this procedure, one would want to control the probability of correct selection (CS) as well as to minimize the necessary sample size. To be precise, with T denoting the total sample size, the criterion is to determine n_1, n_2, and $d > 0$ which

$$\text{minimizes } \sup_\theta E_\theta(T) \tag{2.1}$$

$$\text{subject to } \inf_{\theta \in \Omega(\delta^*)} P_\theta(CS) \geq P^* \tag{2.2}$$

where $\delta^* > 0$ and $P^* (1/k < P^* < 1)$ are pre-specified, and $\Omega(\delta^*)$ is the preference-zone defined by

$$\Omega(\delta^*)=\{\theta\,|\,\theta_{[k]}-\theta_{[k-1]}\geq\delta^*\}.$$

3. Design constants

To satisfy the probability requirement (2.2), we need the minimum probability of correct selection for the procedure in Section 2, which seems extremely complicate to find as can be seen in Tamhane and Bechhofer (1977). Thus we provide a lower bound on the probability of correct selection, which is easily computable.

Proposition 3.1 For the procedure $\mathscr{P}_2$, we have

$$\inf_{\theta\in\Omega(\delta^*)} P_\theta(CS)\geq P\{Z_k+\delta^*+d\geq Z_i,\ Z_k\wedge W_k+\delta^*\geq Z_i\wedge W_i,\ \ i=1,\cdots,k-1\} \qquad (3.1)$$

where n_1Z_i and $n_2W_i(i=1,\cdots,k)$ are independent standard exponential random variables and $a\wedge b=\min(a,b)$.

Proof. First note that

$$\begin{aligned}P_\theta(CS)&=P_\theta\{T^{(1)}_{[k]}\geq\max_{1\leq i\leq k-1} T^{(1)}_{[i]}-d,\ \ T_{[k]}=\max_{\pi^i\in S} T_i\}\\ &\geq P_\theta\{T^{(1)}_{[k]}\geq\max_{1\leq i\leq k-1} T^{(1)}_{[i]}-d,\ \ T_{[k]}\geq\max_{1\leq i\leq k-1} T_{[i]}\} \qquad (3.2)\end{aligned}$$

where $T^{(1)}_{[i]}$ and $T_{[i]}$ are associated with $\theta_{[i]}$ $(i=1,\cdots,k)$. Since the lower bound in (3.2) is non-increasing in $\theta_{[i]}$ $(i=1,\cdots,k-1)$, it is minimized for $\theta\in\Omega(\delta^*)$ when $\theta_{[i]}=\theta_{[k]}-\delta^*$ $(i=1,\cdots,k-1)$. Then the result follows by noting that $n_1(T^{(1)}_i-\theta_i)$ and $n_2(T^{(2)}_i-\theta_i)$ $(i=1,\cdots,k)$ are independent standard exponetial random variables. ▨

A simple calculation shows that the lower bound in (3.1) can be computed as follows:

$$\begin{aligned}&P\{Z_k+\delta^*+d\geq Z_i,\ \ Z_k\wedge W_k+\delta^*\geq Z_i\wedge W_i,\ \ i=1,\cdots,k-1\}\\ &=\int_0^\infty\int_0^\infty\{1-\alpha\beta e^{-n_1x}-\beta\gamma e^{-(n_1+n_2)x\wedge y}+\alpha\beta\gamma e^{-n_1x-n_2x\wedge y}\}^{k-1}n_1n_2e^{-n_1x-n_2y}dxdy\\ &=\sum_{a=0}^{k-1}\ \sum_{b=0}^{k-1-a}\ \sum_{c=0}^{k-1-a-b}\binom{k-1}{a,b,c}\ (-1)^{a+b}\alpha^{a+c}\beta^{a+b+c}\gamma^{b+c}\\ &\quad\times\frac{\log\beta+\dfrac{1}{a+c+1}\log\ \gamma}{(a+b+c+1)\log\beta+(b+c+1)\log\gamma} \qquad (3.3)\end{aligned}$$

where $\binom{k-1}{a,b,c}=(k-1)!/\{a!b!c!(k-1-a-b-c)!\}$ and

$$\alpha=e^{-n_1d},\ \ \beta=e^{-n_1\delta^*},\ \ \gamma=e^{-n_2\delta^*}.$$

Thus, by equating the expression in (3.3) to P^* we can satisfy the probability requirement (2.2).

In minimizing the expected total sample size, we note that the total sample size is given by

$$T=\begin{cases} n_1 k & \text{if } |S|=1 \\ n_1 k+n_2|S| & \text{if } |S|>1 \end{cases}$$

where $|S|$ denotes the number of populations in S. Arguments similar to those in Lee (1984) yield that the expected total sample size is maximized when all the parameters are equal. Thus, we have the next result.

Proposition 3.2 For the procedure $\mathscr{P}_2$, we have

$$\sup_{\theta} E_{\theta}(T)=n_1 k+n_2 k[P\{Z_k\geq \max_{1\leq j\leq k-1} Z_j-d\}-P\{Z_k\geq \max_{1\leq j\leq k-1} Z_j+d\}] \tag{3.4}$$

where $n_1 Z_i$ are independent standard exponential random variables.

A simple calculation shows that

$$\begin{aligned} &P\{Z_k\geq \max_{1\leq j\leq k-1} Z_j-d\}-P\{Z_k\geq \max_{1\leq j\leq k-1} Z_j+d\} \\ &=\int_0^{\infty}\left\{\int_0^{x+n_1 d} e^{-y}dy\right\}^{k-1} e^{-x}dx-\int_{n_1 d}^{\infty}\left\{\int_0^{x-n_1 d} e^{-y}\,dy\right\}^{k-1} e^{-x}dx \\ &=\frac{1}{k\alpha}\{1-(1-\alpha)^k\}-\frac{\alpha}{k} \end{aligned}$$

Therefore, the design criterion specified by (2.1) and (2.2) becomes the following; choose n_1, n_2 and $d>0$ or equivalently $\alpha=e^{-n_1 d}$, $\beta=e^{-n_1\delta^*}$, $\gamma=e^{-n_2\delta^*}$ which minimizes

$$kn_1+n_2\left\{\frac{1}{\alpha}\{1-(1-\alpha)^k\}-\alpha\right\} \tag{3.5}$$

subject to

$$P^*=\sum_{a=0}^{k-1}\sum_{b=0}^{k-1-a}\sum_{c=0}^{k-1-a-b}\binom{k-1}{a,b,c}(-1)^{a+b}\alpha^{a+c}\beta^{a+b+c}\gamma^{b+c} \times \frac{\log\beta+\frac{1}{a+c+1}\log\gamma}{(a+b+c+1)\log\beta+(b+c+1)\log\gamma} \tag{3.6}$$

Note that (3.5) can be represented as follows:

$$k\frac{1}{\delta^*}\log\frac{1}{\beta}+\left[\frac{1}{\alpha}\{1-(1-\alpha)^k\}-\alpha\right]\frac{1}{\delta^*}\log\frac{1}{\gamma}$$

Thus the optimization problem specified by (3.5) and (3.6) is independent on δ^*.

The optimization problem was solved by using the algorithm of Fiacco and McCormick (1968) for a given k and P^*. The solutions are given in Table I for selected values of k and P^*. Then the desired n_1, n_2 and d can be computed by

$$n_1=\left\{\frac{1}{\delta^*}\log\frac{1}{\beta}\right\},\quad n_2=\left\{\frac{1}{\delta^*}\log\frac{1}{\lambda}\right\},\quad d=\delta^*\frac{\log\alpha}{\log\beta}$$

where $\{a\}$ denotes the integer no less than a.

Table I Design constants (α, β, γ) for the procedure $\mathscr{P}_2$.

k	P^*	α	β	γ
3	0.90	0.1307	0.1101	0.9227
	0.95	0.2105	0.0560	0.8783
	0.99	0.3090	0.0114	0.8237
4	0.90	0.3186	0.0919	0.6232
	0.95	0.3489	0.0452	0.5997
	0.99	0.3821	0.0089	0.5737
5	0.90	0.3483	0.0801	0.4442
	0.95	0.3581	0.0392	0.4366
	0.99	0.3711	0.0077	0.4283
6	0.90	0.3375	0.0725	0.3435
	0.95	0.3419	0.0354	0.3409
	0.99	0.3476	0.0069	0.3384
7	0.90	0.3188	0.0674	0.2789
	0.95	0.3209	0.0328	0.2779
	0.99	0.3240	0.0064	0.2782
8	0.90	0.2999	0.0636	0.2339
	0.95	0.3009	0.0309	0.2343
	0.99	0.3027	0.0061	0.2351
9	0.90	0.2824	0.0606	0.2012
	0.95	0.2828	0.0295	0.2019
	0.99	0.2839	0.0058	0.2031
10	0.90	0.2667	0.0583	0.1761
	0.95	0.2670	0.0283	0.1770
	0.99	0.2676	0.0055	0.1779
11	0.90	0.2524	0.0563	0.1569
	0.95	0.2525	0.0273	0.1575
	0.99	0.2528	0.0053	0.1582
12	0.90	0.2399	0.0546	0.1408
	0.95	0.2398	0.0265	0.1414
	0.99	0.2399	0.0052	0.1422

The tabulated values may not be the exact optimal values, but the search for the optimal values has been done up to the accuracy of 10^{-6}. All the computations were done on VAX-11/780 at Seoul National University.

As an illustrative example for the usage of the Table I, suppose we have $k=4$ exponential populations with unknown location parameters. Further, suppose that we would like to select the best population correctly with minimum probability 90% whenever $\theta_{[4]}-\theta_{[3]}\geq 0.20$.

Thus, we have $k=4$, $\delta^*=0.20$, $P^*=0.90$. Hence from Table I, we find $\alpha=0.3186$, $\beta=0.0919$, $\gamma=0.6232$. Therefore, we have

$$n_1=\left\{\frac{1}{\delta^*}\log\frac{1}{\beta}\right\}=12,\quad n_2=\left\{\frac{1}{\delta^*}\log\frac{1}{\gamma}\right\}=3,\quad d=0.095.$$

Thus, we take 12 observations from each population and determine $S=\{\pi_i \mid T_i^{(1)}\geq \max T_j^{(1)}-0.095\}$. If there are more than one population in S, then we take 3 observations from those populations in S.

4. Comparison with a single stage procedure

Since the two-stage procedure $\mathscr{P}_2$ in Section 2 reduces to a single stage procedure in the extreme cases of $d=0$ or ∞, it is evident that the procedure $\mathscr{P}_2$ performs better than the single stage procedure $\mathscr{P}_1$, say, of Raghavachari and Starr (1970) in terms of the maximum expected total sample size. To get an insight how much the procedure $\mathscr{P}_2$ saves total number of samples relative to the procedure $\mathscr{P}_1$, we consider the "relative savings (RSAVE)" as follows;

$$\text{RSAVE}=\frac{E_\theta(T|\mathscr{P}_1)-E_\theta(T|\mathscr{P}_2)}{E_\theta(T|\mathscr{P}_1)}\times 100(\%) \tag{4.1}$$

Note that the total sample size of the single stage procedure $\mathscr{P}_1$ is a constant given by $T=kn$ where n is chosen to control the probability of correct selection, and can be found in Raghavachari and Starr (1970). Thus the relative savings in (4.1) is minimized at the equal means configuration (EMC) by Proposition 3.2. The minimum relative savings can be obtained as by-products of the optimization problem specified by (3.5) and (3.6), and these values are given in Table II for selected values of k and P^*.

Table II Minimum relative savings of $\mathscr{P}_2$ over $\mathscr{P}_1$ in percentage.

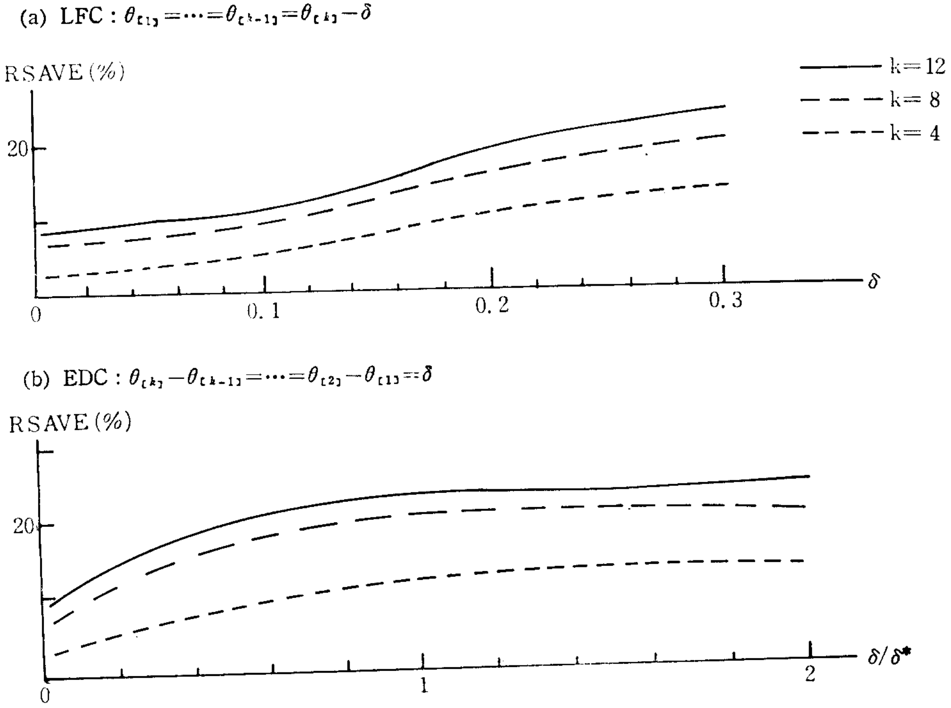
P^* \ k	3	4	5	6	7	8	9	10	11	12
0.90	0.02	0.77	2.18	3.68	5.10	6.41	7.61	8.71	9.72	10.65
0.95	0.05	0.76	1.93	3.15	4.33	5.44	6.46	7.40	8.28	9.08
0.99	0.08	0.66	1.48	2.36	3.22	4.03	4.80	5.51	6.18	6.81

We observe from Table II that the minimum relative savings of $\mathscr{P}_2$ over $\mathscr{P}_1$ are increasing in k and decreasing in P^* as can be expected.

To get a further insight into the performance of $\mathscr{P}_2$ relative to $\mathscr{P}_1$, we have carried out a Monte Carlo study to get the estimates of RSAVE in (4.1) at various parameter configurations. We have chosen two parameter configurations of frequent interest in selection and ranking area; the so-called least favorable configuration (LFC) where $\theta_{[1]}=\cdots=\theta_{[k-1]}=\theta_{[k]}-\delta$ and the equal distance configuration (EDC) where $\theta_{[i+1]}-\theta_{[i]}=\delta$ $(i=1,\cdots,k-1)$ for a fine grid of δ. Then Monte Carlo comparisons have been carried out for $k=4,8,12$, $P^*=0.9$ and $\delta^*=0.2$ with 1000 iterations.

The Monte Carlo results are summarized in Figure I. It can be observed from Figure I that the relative savings are increasing as the true best is more separated from the rest. Moreover, if δ^* can be specified without much error, it can be expected that the relative savings are over 10%. Thus, it can be concluded that the two-stage procedure $\mathscr{P}_2$ can be suggested when there are many populations to be compared and the perference-zone, i.e., δ^* can be specified without much error.

Figure I Relative savings at LFC and EDC configurations


(a) LFC : $\theta_{[1]}=\cdots=\theta_{[k-1]}=\theta_{[k]}-\delta$

(b) EDC : $\theta_{[k]}-\theta_{[k-1]}=\cdots=\theta_{[2]}-\theta_{[1]}=\delta$

References

(1) Barr, D.R. and Rizvi, M.H. (1966). An introduction to selection and ranking procedures. *Journal of the American Statistical Association*, Vol.61, 640～646.

(2) Bechhofer, R.E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variance. *The Annals of Mathematical Statistics*, Vol. 25, 16～39.

(3) Desu, M.M., Narula, S.C. and Villarreal, B. (1977). A two-stage procedure for selecting the best of *k* exponential distributions. *Communications in Statistics*, Ser. A6, 1231～1243.

(4) Desu, M.M. and Sobel, H. (1968). A fixed subset-size approach to a selection problem. *Biometrika*, Vol. 55, 401～410.

(5) Fiacco, A.V. and McCormick G.P. (1968). *Nonlinear Sequential Unconstrained Minimization Techniques*. John Wiley and Sons, Inc., New York, 1968.

(6) Lee, S.H. (1984). A study on two-stage selection procedures. Ph. D. Thesis. Dept. of Comp. Sc. and Statist., Seoul National Univ.

(7) Lee, S.H. and Kim, W.C. (1985). An elimination type two-stage selection procedure for exponential distributions. *Communications in Statistics-Theory and Methods*, A14, 2563～2571.

(8) Raghavachari, M. and Starr, N. (1970). Selection problems for some terminal distributions. *Metron*, Vol. 28, 185～197.

(9) Tamhane, A.C. and Bechhofer, R.E. (1977). A two-stage minimax procedure with screening for selecting the largest normal mean. *Communications in Statistics-Theory and Methods*, A6, 1003～1033.