ON SELECTION PROCEDURES FOR EXPONENTIAL DISTRIBUTIONS

Ofosu, J. B.
School of Mathematics, University of Bath

https://doi.org/10.5109/13078

出版情報：統計数理研究. 16 (1/2), pp.1-9, 1974-03. 統計科学研究会
バージョン：
権利関係：

ON SELECTION PROCEDURES FOR EXPONENTIAL DISTRIBUTIONS

By

J. B. OFOSU*

(Received October 24, 1972)

1. Introductson.

The problem of selecting a subset of k given populations which includes the best population has extensively been studied by several authors notably by Gupta ([1], [2]), Gupta and Sobel ([3]). The best population is usually defined as the one with the largest (or smallest) parameter value.

In this paper we are concerned with the location parameters θ_i $(i=1, 2, \cdots, k)$ of k given exponential populations with a common known scale parameter. The best population is defined as the one with the largest θ-value. Based on a common number of observations from each population, two procedures, R_1 and R_2 are defined such that each procedure selects a subset which is never empty, small in size and yet large enough to guarantee with pre-assigned probability that it includes the best population, regardless of the true unknown values of the θ_i's. Tables necessary for carrying out the procedures are given.

The main formulation is formally described in section 2. Procedures R_1 and R_2 are defined in sections 3 and 4 respectively. It should be pointed out that the expected size of the selected subset, using each procedure, is a random variable and can be regarded as a measure of the efficiency of the procedure. In section 5 we show, numerically, that in most situations, the expected size of the selected subset for R_1 is smaller than for R_2.

2. Formal statement of the problem.

Let $\pi_1, \pi_2, \cdots, \pi_k$ denote k given exponential populations with density functions

$$(2.1) \qquad \frac{1}{q} \exp\{-(x-\theta_i)/q\} \qquad x>\theta_i \ (i=1, 2, \cdots, k)$$

with a common known scale parameter $q(>0)$ and unknown location parameters θ_i. The population associated with the largest θ-value is defined as the best population. Our goal is to select a subset of the k populations which contains the best population. Any such selection will be called a correct selection (CS). The problem is to find a rule R such that for a pre-assigned probability $P^*(1/k<P^*<1)$,

* School of Mathematics, University of Bath, Bath, England.

(2.2) $$P(\mathrm{CS} \mid R) \geqq P^*$$

regardless of the unknown values of the θ_i's.

From each population π_i we take n observations and compute the minimum value Y_i $(i=1, 2, \cdots, k)$. $Y_1, \cdots, Y_k$ form a set of sufficient statistics for the problem and the rules R_1 and R_2, defined in sections 3 and 4 respectively, depend only on these statistics.

Let $Y_{(i)}$ denote the Y-value associated with population $\pi_{(i)}$ with parameter $\theta_{[i]}$, where $\theta_{[i]}$ is the i-th smallest θ_i. The c. d. f. and p. d. f. of $Y_{(i)}$ are respectively

(2.3) $$G_{\theta_{[i]}}(y) = 1 - \exp\{-n(y-\theta_{[i]})/q\} \qquad y > \theta_{[i]},$$

(2.4) $$g_{\theta_{[i]}}(y) = \frac{n}{q} \exp\{-n(y-\theta_{[i]})/q\} \qquad y > \theta_{[i]}.$$

3. Procedure R_1.

The procedure R_1 is defined as follows; "Include π_j in the selected subset iff

(3.1) $$Y_j \geqq \max_{1 \leqq i \leqq k} Y_i - dq/n \qquad (j=1, 2, \cdots, k)$$

where $d = d(k, P^*)$, a positive constant, is determined in advance so as to satisfy (2.2) regardless of the unknown values of the θ_i's."

3.1. The $P(\mathrm{CS} \mid R_1)$ and its infimum.

We now derive expressions for the $P(\mathrm{CS} \mid R_1)$ and its infimum over all points in the parameter space Ω, which is the set of all admissible vectors $\theta = (\theta_{[1]}, \cdots, \theta_{[k]})$.

The procedure R_1 yields a correct selection iff the event

$$Y_{(k)} \geqq \max_{1 \leqq i \leqq k} Y_{(i)} - dq/n \quad \text{occurs}.$$

Hence,

(3.2) $$\begin{aligned} P(\mathrm{CS} \mid R_1) &= P\{Y_{(i)} \leqq Y_{(k)} + dq/n \ (i=1, 2, \cdots, k-1)\} \\ &= \int_0^\infty \prod_{i=1}^{k-1} [1 - \exp\{-(x+d+n(\theta_{[k]}-\theta_{[i]})/q)\}] e^{-x} dx. \end{aligned}$$

It follows from (3.2) that

(3.3) $$\begin{aligned} \inf_{\Omega} P(\mathrm{CS} \mid R_1) &= \int_0^\infty [1 - \exp\{-(x+d)\}]^{k-1} e^{-x} dx \\ &= e^d \{1-(1-e^{-d})^k\}/k. \end{aligned}$$

Thus the constant d satisfying the P^* requirement (2.2) can be obtained by equating the last expression in (3.3) to P^*. This gives

$$d = -\log u$$

where

$$(1-u)^k + kP^*u - 1 = 0.$$

Table 1 gives values of d for $k = 2(1)20$ and $P^* = 0.85, 0.90, 0.95, 0.99$.

Table 1. This table gives the necessary d-value required for procedure R_1.

k/P^*	0.85	0.90	0.95	0.99
2	1.204	1.609	2.303	3.911
3	1.843	2.267	2.979	4.603
4	2.230	2.661	3.378	5.006
5	2.509	2.943	3.663	5.290
6	2.727	3.163	3.885	5.518
7	2.905	3.343	4.066	5.697
8	3.057	3.495	4.219	5.855
9	3.189	3.627	4.352	5.982
10	3.305	3.744	4.469	6.105
11	3.409	3.849	4.574	6.209
12	3.503	3.943	4.669	6.304
13	3.590	4.030	4.756	6.395
14	3.669	4.110	4.836	6.471
15	3.742	4.183	4.910	6.544
16	3.811	4.252	4.979	6.615
17	3.875	4.316	5.043	6.681
18	3.935	4.377	5.104	6.741
19	3.992	4.433	5.161	6.794
20	4.046	4.487	5.215	6.850

EXAMPLE 1. From each of $k=6$ exponential populations with a common scale parameter $q=3$, eight observations were taken. The observed minimum values based on the $n=8$ observations were 91.16, 106.20, 110.63, 126.81, 124.92, 80.81. If $P^*=0.99$, then from Table 1, $d=5.518$. Applying the procedure R_1 we find that the two populations that gave rise to the values 126.81, 124.92 are retained in the subset. At this point the experimenter can assert with confidence level 0.99 that one of these two populations has the largest θ-value among the 6 populations.

3.2. Expected size of the selected subset for R_1.

For the procedure R_1 the size S of the selected subset is a discrete random variable which can take on only integer values from 1 to k. The expression, $E(S\,|\,R_1)$, for the expected size of the selected subset using the procedure R_1, is

$$E(S|R_1)=\sum_{j=1}^{k} p\{\pi_{(j)} \text{ is selected} | R_1\} \tag{3.4}$$

where

$$\begin{aligned} P\{\pi_{(j)} \text{ is selected} \mid R_1\} &= P\{Y_{(j)} \geqq Y_{(i)} - dq/n \ (i=1,2,\cdots,k;\ i\neq j)\} \\ &= \int_{L_j}^{\infty} \prod_{\substack{i=1\\ i\neq j}}^{k} [1-\exp\{-(x+d+n(\theta_{[j]}-\theta_{[i]})/q)\}] e^{-x} dx \end{aligned} \tag{3.5}$$

and

$$L_j = \max[0, \{n(\theta_{[k]}-\theta_{[j]})/q - d\}]. \tag{3.6}$$

Consider the following configurations of the unknown parameters:

$$(3.7)\qquad \begin{cases} \theta_{[j]}=\theta & (j=1, 2, \cdots, k-1) \\ \theta_{[k]}=\theta+\delta q & \delta>0, \end{cases}$$

$$(3.8)\qquad \theta_{[j]}=\theta+(j-1)\delta q \qquad \delta>0 \qquad (i=1, 2, \cdots, k).$$

When (3.7) holds, (3.5) simplifies to

$$(3.9)\qquad \int_b^\infty [1-\exp\{-(x+d)\}]^{k-2}[1-\exp\{-(x+d-\delta n)\}]e^{-x}dx \quad \text{for } j=1, 2, \cdots, k-1,$$

$$(3.10)\qquad \int_0^\infty [1-\exp\{-(x+d+\delta n)\}]^{k-1}e^{-x}dx \quad \text{for } j=k$$
$$=e^s\{1-(1-e^s)^k\}/k$$

where

$$s=d+\delta n$$

and

$$b=\max\{0, (\delta n-d)\}.$$

Under (3.8) the probability (3.5) simplifies to

$$(3.11)\qquad \int_{r_j}^\infty \prod_{\substack{i=1\\ i\neq j}}^k [1-\exp\{-(x+d+\delta n(j-i))\}]e^{-x}dx \qquad (j=1, 2, \cdots, k)$$

where

$$r_j=\max[0, \{(k-j)\delta n-d\}].$$

The expected subset-size for R_2 under the configurations (3.7) and (3.8) of the population parameters has been calculated and are given in Table 3 and 4 for $P^*=0.95$ and for selected values of k and δn.

From (3.9) to (3.11) we can make the following two remarks!

REMARK 3.1. For fixed P^*, k, j $(j=1, 2, \cdots, k-1)$, the probability of selecting population $\pi_{(j)}$ decreases from P^* to zero as δn increases from 0 to ∞.

REMARK 3.2. For fixed P^*, the probability of selecting $\pi_{(k)}$ increases from P^* to 1 as δn increases from 0 to ∞.

3.3. Properties of the Procedure R_1.

The procedure R_1 has the following desirable properties:

(a) $E(S \mid R_1) \leqq kP^*$ for all points in the parameter space Ω.

We omit the proof; it is tedious algebraically and is available for similar situations, along with further discussion, in the references cited below.

(b) If $\theta_{[i]} \geqq \theta_{[j]}$, then $P\{\pi_{(i)}$ is selected $\mid R_1\} \geqq P\{\pi_{(j)}$ is selected $\mid R_1\}$.

The proof of this assertion follows from equation (3.5).

4. Procedure R_2.

This procedure is defined as follows. "Include π_j in the selected subset iff

(4.1) $$Y_j \geqq Z_j/(k-1)-cq/n$$

where

(4.2) $$Z_j=\sum_{\substack{i=1\\ i\neq j}}^{k} Y_i$$

and $c=c(k, P^*)$ is a positive constant which is determined so as to satisfy (2.2) regardless of the unknown values of the θ_i's."

4.1. The $P\{$selecting $\pi_{(j)}|R_2\}$.

The probability that population $\pi_{(j)}$, with parameter $\theta_{[j]}$, is selected using procedure R_2 can be expressed as

(4.3) $$P[Z_{(j)} \leqq (k-1)\{Y_{(j)}+c/w\}]$$

where

(4.4) $$w=n/q\,,$$

(4.5) $$Z_{(j)}=\sum_{\substack{i=1\\ i\neq j}}^{k} Y_{(i)}\,.$$

Now, the p. d. f. and c. d. f. of $\left(Z_{(j)}-\sum_{\substack{i=1\\ i\neq j}}^{k}\theta_{[i]}\right)$ are given respectively by

(4.6) $$f(z)=\frac{w(wz)^{k-2}e^{-wz}}{(k-2)\,!}\,,$$

(4.7) $$F(z)=H_{k-1}(wz)\,,$$

where

(4.8) $$H_{k-1}(x)=1-\sum_{r=0}^{k-2}\frac{e^{-x}x^r}{r\,!}\qquad x>0\,,$$

and w is given by (4.4).

Using (2.4) and (4.6)-(4.8), expression (4.3) simplifies to

(4.9) $$\int_0^\infty H_{k-1}(a+b_j)e^{-u}du$$

where

(4.10) $$a=(k-1)(u+c)$$

and

(4.11) $$b_j=w\Big\{(k-1)\theta_{[j]}-\sum_{\substack{i=1\\ i\neq j}}^{k}\theta_{[i]}\Big\}\,.$$

In particular, from (4.9),

(4.12) $$P(\mathrm{CS}\,|\,R_2)=P\{\pi_{(k)}\text{ is selected}\,|\,R_2\}=\int_0^\infty H_{k-1}(a+b_k)e^{-u}du$$

where a is given by (4.10) and

$$b_k=w\Big\{(k-1)\theta_{[k]}-\sum_{i=1}^{k-1}\theta_{[i]}\Big\}\,.$$

Clearly, $b_k \geqq 0$, hence

$$\inf_{\Omega} P(\mathrm{CS}|R_2) = \int_0^\infty H_{k-1}(a) e^{-u} du \,. \tag{4.13}$$

From (2.2) and (4.13) we see that the constant c to carry out the procedure R_2 can be obtained by solving

$$\int_0^\infty H_{k-1}\{(k-1)(u+c)\} e^{-u} du = P^* \qquad \text{for} \quad c \,. \tag{4.14}$$

Table 2 gives the values of c satisfying (4.14) for selected values of P^* and k.

Table 2. This table gives the necessary c-values required for procedure R_2.

k/P^*	0.85	0.90	0.95	0.99
3	1.067	1.336	1.774	2.736
4	1.001	1.216	1.558	2.283
5	0.962	1.147	1.437	2.037
6	0.937	1.103	1.359	1.879
7	0.919	1.071	1.303	1.769
8	0.905	1.048	1.262	1.687
9	0.895	1.030	1.230	1.622
10	0.887	1.015	1.204	1.571
15	0.864	0.971	1.125	1.413
20	0.853	0.949	1.084	1.330

We note that for $k=2$, the procedures R_1 and R_2 are equivalent, hence the c-values corresponding to $k=2$ can be obtained from Table 1.

It should be noted from (4.1) and (4.2) that the procedure R_2 can also be defined as follows: "Include π_j in the selected subset iff

$$Y_j \geqq \bar{Y} - c_1 q/n \tag{4.15}$$

where

$$\bar{Y} = 1/k \sum_{i=1}^{k} Y_i \,, \tag{4.16}$$

$$c_1 = c(k-1)/k \,, \tag{4.17}$$

and c is the root of equation (4.14)."

In applying the procedure R_2, it is easier to use equations (4.15)-(4.17) than equations (4.1) and (4.2).

EXAMPLE 2. Let us solve Example 1 by applying the procedure R_2. The mean $\bar{Y}$ of the 6 observations is 106.76. From Table 2, the value of c corresponding to $k=6$, $P^*=0.99$ is 1.879; thus from (4.17), $c_1 q/n = 0.59$. Hence, from (4.15), the selected subset contains all populations which gave rise to values greater than or equal to 106.17. Thus the populations which gave rise to the values 106.20, 110.63, 126.81, 124.92 are retained in the subset. The size of the selected subset is 4, compared with 2 when the procedure R_1 is used. The superiority of R_1 over R_2 is quite evident here.

4.2. Properties of the Procedure R_2.

The procedure R_2 possesses all the desirable properties that the procedure R_1 has and that are discussed in section 3.3. The proofs are omitted.

4.3. Expected size of the selected subset for R_2.

The expected subset-size using procedure R_2, $E(S|R_2)$, is given by

(4.18) $$E(S|R_2)=\sum_{j=1}^{k} P\{\pi_{(j)} \text{ is selected}|R_2\}$$

where $P\{\pi_{(j)}$ is selected$|R_2\}$ is given by (4.9)-(4.11).

We now evaluate $E(S|R_2)$ for the configurations (3.7) and (3.8) of the unknown parameters.

Case (i): For the configuration (3.7) of the $\theta_{[i]}$'s, we can express the $P\{\pi_{(j)}$ is selected$|R_2\}$ as

(4.19) $$\int_0^\infty H_{k-1}\{(k-1)(u+c)-\delta n\}\, e^{-u} du \qquad (j=1, 2, \cdots, k-1),$$

(4.20) $$\int_0^\infty H_{k-1}\{(k-1)(u+c+\delta n)\}\, e^{-u} du \qquad j=k.$$

It should be noted from (4.19) and (4.20) that Remarks 3.1 and 3.2 (see section 3.2) apply to the above probability. Values of $E(S|R_2)$ for the configuration (3.7) of the $\theta_{[i]}$'s are given in Table 3 for $P^*=0.95$ and for selected values of k and δn.

Case (ii). If the unknown parameters are given by (3.8), then from (4.9)-(4.11), we can write the $P\{\pi_{(j)}$ is selected$|R_2\}$ as

(4.21) $$\int_0^\infty H_{k-1}\{(k-1)(u+c)+kn\delta(2j-k-1)/2\}\, e^{-u} du \qquad (j=1, 2, \cdots, k).$$

From (4.21) we can make the following three remarks.

REMARK 4.1. For fixed P^*, k, j $(1\leqq j\leqq k;\ 2j<k+1)$, the probability of selecting $\pi_{(j)}$ decreases from P^* to 0 as δn increases from 0 to ∞.

REMARK 4.2. For fixed P^*, k, j $(1\leqq j\leqq k;\ 2j=k+1)$, the probability of selecting $\pi_{(j)}$ is equal to P^* for all values of δn.

REMARK 4.3. For fixed P^*, k, j $(1\leqq j\leqq k;\ 2j>k+1)$, the probability of selecting $\pi_{(j)}$ increases from P^* to 1 as δn increases from 0 to ∞. Thus for fixed P^*, k, $\delta(>0)$ $j(1\leqq j\leqq k,\ 2j>k+1)$ and for sufficiently large n, population $\pi_{(j)}$ is selected with probability 1.

Using (4.18) and (4.21) we have computed $E(S|R_2)$ and is given in Table 4 for $P^*=0.95$ and selected values of δn and k.

5. Comparison of $E(S|R_1)$ and $E(S|R_2)$.

Tables 3 and 4 compare $E(S|R_1)$ with $E(S|R_2)$ when the unknown parameters are given by (3.7) and (3.8) respectively. From the tables, we can make the following remarks.

Table 3. Comparison of $E(S|R_1)$ (top entry) with $E(S|R_2)$ (bottom entry); $\theta_{[j]}=\theta(j=1, 2, \cdots k-1)$, $\theta_{[k]}=\theta+\delta q(\delta>0)$, $P^*=0.95$.

$k\backslash\delta n$	1.50	2.50	3.50	4.50	5.50
3	2.72 2.68	2.35 2.36	1.59 1.91	1.22 1.55	1.08 1.34
4	3.67 3.59	3.30 3.28	2.30 2.83	1.48 2.34	1.18 1.96
5	4.62 4.52	4.25 4.19	3.24 3.75	1.86 3.23	1.32 2.75
10	9.37 9.17	9.00 8.80	7.99 8.34	5.24 7.81	2.59 7.22
15	14.12 13.85	13.76 13.46	12.75 12.98	9.99 12.44	4.82 11.83
20	18.88 18.56	18.50 18.14	17.50 17.64	14.74 17.08	7.96 16.48

Table 4. Comparison of $E(S|R_1)$ (top entry) with $E(S|R_2)$ (bottom entry); $\theta_{[j]}=\theta+(j-1)\delta q$ $(j=1, 2, \cdots, k)$ $(\delta>0)$, $P^*=0.95$.

$k\backslash\delta n$	1.5	2.5	3.5	4.5	5.5	∞
3	2.13 2.23	1.75 2.01	1.31 1.96	1.11 1.95	1.04 1.95	1.0 1.95
4	2.68 2.78	1.88 2.39	1.46 2.19	1.16 2.10	1.06 2.05	1.0 2.0
5	2.88 3.24	1.98 3.03	1.59 2.97	1.22 2.95	1.08 2.95	1.0 2.95
6	3.02 3.75	2.05 3.37	1.68 3.19	1.27 3.11	1.10 3.06	1.0 3.0

(i) Both $E(S|R_1)$ and $E(S|R_2)$ are monotonically decreasing with δn.

(ii) Both $E(S|R_1)$ and $E(S|R_2)$ decrease as the $\theta_{[j]}$'s become more unequal (compare Tables 3 and 4).

If we let $D(k, P^*, \delta n)=E(S|R_1)-E(S|R_2)$ then based on the behaviour of $E(S|R_1)$ and $E(S|R_2)$ in the range computed, the following additional properties would appear to hold:

(iii) For fixed k, P^*, there exists a value $d_0=d_0(k, P^*)$ of δn such that $D(k, P^*, \delta n)<0$ $\delta n>d_0$.

(iv) For fixed P^*, δn $(\delta n>d_0)$, $D(k, P^*, \delta n)$ decreases with k.

6. Discussion.

Although we have used special cases and relied heavily on numerical results, it seems that some important points emerge from this work.

The above results show that the expected size of the selected subset for the procedure R_1 is smaller than that for the procedure R_2 in most situations. This minimization of the expected size appears to be the most desirable criterion on which the selection of the optimum rule should be based. We therefore conclude that R_1 will be superior to R_2 in almost all situations and even for those rare situations when R_2 is superior, R_1 will be only slightly inferior. The present study will suggest that R_1 should be preferred to R_2 in all practical situations.

7. Acknowledgement.

The author would like to thank Mr. C. M. Theobald for some helpful discussions and suggestions which have helped to improve the results of this paper.

References

[1] S. S. GUPTA, *On a selection and ranking procedure for gamma populations.* Ann. Inst. Statist. Math. **14** (1963), 199-216.

[2] S. S. GUPTA, *On some multiple decision rules.* Technometrics, **7** (1965), 225-245.

[3] S. S. GUPTA and M. SOBEL, *On selecting a subset containing the population with the smallest variance.* Biometrika, **49** (1962), 495-507.