DISTRIBUTION OF THE RATIO OF THE LOGARITHM OF ANY ONE OF THE RANGES OF SAMPLES FROM A RECTANGULAR POPULATION TO THE SUM OF THE LOGARITHMS OF EACH OF THEM

By D. N. Lal and D. Mishra

Patna University

1. Introduction

In an earlier paper [1] read at the Baroda (1955) session of the Indian Science Congress the distribution of the product of the ranges of K independent random samples from a continuous rectangular population was obtained. In the present paper the distribution of the ratio of the logarithm of any one of the ranges to the sum of their logarithms has been found. In both these papers the distributions obtained are valid for unequal sample-sizes such that the difference between no two of them is less than two. Besides, these distributions are valid only for the ranges lying between 0 and 1. Thus the ranges will have to be brought down to less than 1 by a suitable change of origin and scale, which can always be done, before applying these distributions to any problem. These can be of use in such situations where we can get samples at various intervals of time, the only restriction being that the population should be rectangular.

2. We consider K independent random samples from the population

$$f(x) = \begin{cases} 1 & \text{for } 0 \leqslant x \leqslant 1 \\ 0 & \text{elsewhere.} \end{cases}$$

of sizes $n_1, n_2, \cdots n_k$ no two of which are equal or differ by one. The distribution of

$$W = x_1, x_2, \cdots x_k,$$

where $x_1, x_2, \cdots x_k$ are the ranges of the K samples, has been obtained in [1] and is given by

$$\left(\prod_{i=1}^{k} n_i^{(2)}\right)\left\{\sum_{i=1}^{k} \frac{W^{n_i-2}}{\prod\limits_{i\neq j=1}^{k}(n_i - n_j)^{(2)}} - \sum_{i=1}^{k} \frac{W^{n_i-1}}{\prod\limits_{i\neq j=1}^{k}(n_i - n_j + 1)^{(2)}}\right\} dW$$

$$0 \leqslant W \leqslant 1 \qquad (2)$$

The distribution of the ranges of random samples of size n from the population (1) is

$$n\,(n-1)\,x^{n-2}\,(1-x)\,dx \tag{3}$$

Thus if x_l be the range of a sample of size n_l and W be the product of the ranges $x_1, x_2 \cdots x_{l-1}\, x_{l+1}, \cdots x_k$ of the remaining $(K-1)$ samples, then the joint distribution of W and x_l is given by

$$\left(\prod_{i=1}^{k} n_i^{(2)}\right)\left\{\sum_{i\neq l=1}^{k}\frac{W^{n_i-2}}{\prod\limits_{i\neq l\neq j=1}^{k}(n_i-n_j)^{(2)}}-\sum_{i\neq l=1}^{k}\frac{W^{n_i-1}}{\prod\limits_{i\neq j\neq l=1}^{k}(n_i-n_j+1)^{(2)}}\right\} \times x_l^{n_l-2}\,(1-x_l)\,dx_l\,dW \tag{4}$$

Let us put in (4)

$$W=e^{-z} \text{ and } x_l=e^{-y_l}$$

then (4) becomes

$$\left(\prod_{i=1}^{k} n_i^{(2)}\right)\left\{\sum_{i\neq l=1}^{k}\frac{e^{-z(n_i-1)}}{\prod\limits_{i\neq j\neq l=1}^{k}(n_i-n_j)^{(2)}}-\sum_{i\neq l=1}^{k}\frac{e^{-zn_i}}{\prod\limits_{i\neq j\neq l=1}^{k}(n_i-n_j+1)^{(2)}}\right\} \times e^{y_l(n_l-1)}\,(1-e^{-y_l})\,dy_l\,dz.$$

$$0\leqslant y_l\leqslant\infty \text{ and } 0\leqslant z\leqslant\infty \tag{5}$$

Put $y_l = zu_l$ so that

$$u_l=\frac{y_l}{z}=\frac{\log_e x_l}{\sum\limits_{i\neq l=1}^{k}\log_e x_i}$$

With these substitutions (5) takes the form

$$\left(\prod_{i=1}^{k} n_i^{(2)}\right)\left\{\sum_{i\neq l=1}^{k}\frac{e^{-z\{n_i-1+u_l(n_l-1)\}}}{\prod\limits_{i\neq j\neq l=1}^{k}(n_i-n_j)^{(2)}}\right.$$

$$-\sum_{i\neq l=1}^{k}\frac{e^{-z\{n_i+u_l(n_l-1)\}}}{\prod\limits_{i\neq j\neq l=1}^{k}(n_i-n_j+1)^{(2)}}$$

$$-\sum_{i\neq l=1}^{k}\frac{e^{-z\{n_i+n_lu_l-1\}}}{\prod\limits_{i\neq j\neq l=1}^{k}(n_i-n_j)^{(2)}}$$

$$+\sum_{i\neq l=1}^{k} \frac{e^{-z(n_i+n_l u_l)}}{\prod\limits_{i\neq j\neq l=1}^{k} (n_i - n_j + 1)^{(2)}} \Bigg\} zdzdu_l \tag{6}$$

3. To obtain the distribution of u_l we integrate (6) with respect to z from 0 to ∞. The distribution of u_l is then found to be

$$\left(\prod_{i=1}^{k} n_i^{(2)}\right)\Bigg\{ \sum_{i\neq l=1}^{k} \frac{1}{\prod\limits_{i\neq j\neq l=1}^{k} (n_i - n_j)^{(2)} \{n_i - 1 + u_l (n_l - 1)\}^2}$$

$$-\sum_{i\neq l=1}^{k} \frac{1}{\prod\limits_{i\neq j\neq l=1}^{k} (n_i - n_j + 1)^{(2)} \{n_i + u_l (n_l - 1)\}^2}$$

$$-\sum_{i\neq l=1}^{k} \frac{1}{\prod\limits_{i\neq j\neq l=1}^{k} (n_i - n_j)^{(2)} \{n_i - 1 + n_l u_l\}^2}$$

$$+\sum_{i\neq l=1}^{k} \frac{1}{\prod\limits_{i\neq j\neq l=1}^{k} (n_i - n_j + 1)^{(2)} \{n_i + n_l u_l\}^{(2)}} \Bigg\} du_l$$

$$0 \leqslant u_l \leqslant \infty \tag{7}$$

This is the distribution of the ratio of the logarithm of any one of the ranges to the sum of the logarithms of the remaining $(k - 1)$ ranges. In order to find the distribution of the ratio of the logarithm of any one of the ranges to the sum of the logarithms of all of them let us put in (7)

$$w_l = \frac{u_l}{u_l + 1} = \frac{\log x_l}{\sum\limits_{i=1}^{k} \log x_i}$$

then

$$u_l = \frac{w_l}{1 - w_l}$$

and

$$du_l = \frac{dw_l}{(1 - w_l)^2}$$

and (7) becomes

$$f(w_l)\,dw_l = \left(\prod_{i=1}^{k} n_i^{(2)}\right)$$

$$\times\left\{\sum_{i\neq l=1}^{k} \frac{1}{\prod\limits_{i\neq j\neq l=1}^{k}(n_i - n_j)^{(2)}\,\{n_i - 1 + w_l(n_l - n_i)\}^2}\right.$$

$$-\sum_{i\neq l=1}^{k} \frac{1}{\prod\limits_{i\neq j\neq l=1}^{k}(n_i - n_j + 1)^{(2)}\,\{n_i + w_l(n_l - n_i - 1)\}^2}$$

$$-\sum_{i\neq l=1}^{k} \frac{1}{\prod\limits_{i\neq j\neq l=1}^{k}(n_i - n_j)^{(2)}\,\{n_i - 1 + w_l(n_l - n_i + 1)\}^2}$$

$$\left.+\sum_{i\neq l=1}^{k} \frac{1}{\prod\limits_{i\neq j\neq l=1}^{k}(n_i - n_j + 1)^{(2)}\,\{n_i - w_l(n_l - n_i)\}^2}\right\} dw_l$$

$$0 \leqslant w_l \leqslant 1 \qquad (8)$$

4. The corresponding formula for the case where the sample sizes are the same and equal to n is given by

$$c\left[\left\{a_0\frac{1}{(n-1)^2} - a_1\frac{2!\,(1-w_l)}{(n-1)^3} + a_2\frac{3!\,(1-w_l)^2}{(n-1)^4} - \cdots\cdots\cdots\right.\right.$$

$$\left. + (-1)^{k-2}a_{k-2}\frac{(k-1)!\,(1-w_l)^{k-2}}{(n-1)^k}\right\} - \left\{a_0\frac{1}{(n-w_l)^2} + \cdots\right.$$

$$\left. + a_1\frac{2!\,(1-w_l)}{(n-w_l)^3} + \cdots + a_{k-2}\frac{(k-1)!\,(1-w_l)^{k-2}}{(n-w_l)^k}\right\}$$

$$-\left\{a_0\frac{1}{(n+w_l-1)^2} - a_1\frac{2!\,(1-w_l)}{(n+w_l-1)^3} + \cdots\right.$$

$$\left. + (-1)^{k-2}a_{k-2}\frac{(k-1)!\,(1-w_l)^{k-2}}{(n+w_l-1)^k}\right\}$$

$$+\left\{a_0\frac{1}{n_2} + a_1\frac{2!\,(1-w_l)}{n^3} + \cdots\right.$$

$$\left.\left. + a_{k-2}\frac{(k-1)!\,(1-w_l)^{k-2}}{n^k}\right\}\right] dw_l \qquad (9)$$

Where

$$c = (-1)^{k-2} n^k (n-1)^k$$

$$a_{k-2} = \frac{1}{(k-2)!}$$

$$a_{k-3} = \frac{k-1}{(k-3)!}$$

....

$$a_{k-r} = \frac{(k+r-3)!}{(k-2)!\,(r-1)!\,(k-r-1)!}$$

....

$$a_0 = \frac{(2k-4)!}{[(k-2)!]^2}$$

This result (9) has been derived on the same lines as (8) above with the help of the result in [3].

5. This distribution can be used to test the homogeneity with regard to variation in samples from a rectangular population. The probability that at least the largest of these ratios will exceed a preassigned value may be obtained by using the following well-known theorem in probability.

If $P(w_i)$ is the probability that the ith ratio exceeds a preassigned value w, $P(w_i w_j)$ the probability that the ith and jth ratios both exceed w and so on, then the probability that at least the largest of the ratios exceeds w is given by

$$\Sigma\, P(w_i) - \Sigma\, P(w_i w_j) + \Sigma\, P(w_i w_j w_k) - \ldots\ldots \tag{10}$$

where summations extend over all combinations of subscripts taken one, two, three......all at a time.

To find the above probability it is necessary to obtain the joint distributions of two, three......all of w_i's. But as w_i's are not independent since $\sum_{i=1}^{k} w_i = 1$ the joint distributions cannot be directly obtained.

Let

$$y_1 = w_1 = \log x_1 / \sum_{i=1}^{k} \log x_i$$

$$y_2 = \frac{w_2}{1 - w_1} = \log x_2 / \sum_{i=2}^{k} \log x_i$$

$$y_3 = \frac{w_3}{1 - w_1 - w_2} = \log x_3 / \sum_{i=3}^{k} \log x_i$$

and so on.

y_i's are now obviously independently distributed as w_i's. Thus the joint distributions of w_i's may be obtained from those of y_i's. For example since Y_1 and Y_2 are independently distributed .

$$\begin{aligned} dF(y_1 y_2) &= f(y_1) f(y_2)\, dy_1\, dy_2 \\ &= f(w_1)\, dw_1\, f[w_2/(1 - w_1)]\, dw_2/(1 - w_1)^2 \end{aligned}$$

where $f(w_1)$ and $f[w_2/(1 - w_1)]$ follow from (8). The joint distribution of w_1 and w_2 is therefore given by

$$n_1^{(2)} \left\{ \left(\prod_{i=2}^{k} n_i^{(2)} \right) \right\}^2 \left[\left\{ \sum_{i=2}^{k} \frac{1}{\prod\limits_{i \neq j=2}^{k} (n_i - n_j)^{(2)} \{w_1 (n_1 - n_i) + n_i - 1\}^2} \right. \right.$$

$$- \sum_{i=2}^{k} \frac{1}{\prod\limits_{i \neq j=2}^{k} (n_i - n_j + 1)^{(2)} \{w_1 (n_1 - 1 - n_i) + n_i\}^2}$$

$$- \sum_{i=2}^{k} \frac{1}{\prod\limits_{i \neq j=2}^{k} (n_i - n_j)^{(2)} \{w_1 (n_1 - n_i + 1) + n_i - 1\}^2}$$

$$\left. + \sum_{i=2}^{k} \frac{1}{\prod\limits_{i \neq j=2}^{k} (n_i - n_j + 1)^{(2)} \{w_1 (n_1 - n_i) + n_i\}^2} \right\}$$

$$\left\{ \sum_{i=3}^{k} \frac{1}{\prod\limits_{i \neq j=3}^{k} (n_i - n_j)^{(2)} \{w_2 (n_2 - n_i) + (n_i - 1)(1 - w_1)\}^2} \right.$$

$$- \sum_{i=3}^{k} \frac{1}{\prod\limits_{i \neq j=3}^{k} (n_i - n_j + 1)^{(2)} \{w_2 (n_2 - 1 - n_i) + n_i (1 - w_1)\}^2}$$

$$-\sum_{i=3}^{k} \frac{1}{\prod\limits_{i\neq j=3}^{k}(n_i - n_j)^{(2)}\{w_2(n_2+1-n_i)+(n_i-1)(1-w_1)\}^2}$$

$$+\sum_{i=3}^{k} \frac{1}{\prod\limits_{i\neq j=3}^{k}(n_i - n_j + 1)^{(2)}\{w_2(n_2-n_i)+n_i(1-w_1)\}^2}\Bigg\}\Bigg]$$

$$\times\, dw_1 dw_2 \quad (11)$$

The joint distributions of any two, three, ... of w_i's can be obtained in the same manner. Similar results can be obtained for the case of samples of equal size.

6. The different terms in (10) can be calculated as follows:—

$$P(w_i) = \int_w^1 f(w_i)\, dw_i$$

$$P(w_i w_i) = \begin{cases} 0 \text{ if } w > \frac{1}{2} \\ \int_w^{1-w} dw_i \int_w^{1-w_i} f(w_i w_j)\, dw_j \text{ if } w \leqslant \frac{1}{2}. \end{cases}$$

$$P(w_i w_j w_l) = \begin{cases} 0 \text{ if } w > \frac{1}{3} \\ \int_w^{1-2w} dw_i \int_w^{1-w_i-w} dw_j \int_w^{1-w_i-w_j} (w_i w_j w_l)\, dw_l \text{ if } w \leqslant \frac{1}{3}. \end{cases}$$

and so on.

7.. To find the value of w corresponding to a given probability will be difficult and will involve heavy calculations. The very first term in (10) will involve equations of $4(K-1)$th degree, the second term equations of $8(K-1)$th degree and so on. The calculation of probability corresponding to a given value of w will be comparatively easy for $w > \frac{1}{2}$, but again for $w \leqslant \frac{1}{2}$ there will be heavy calculations. Probabilities have been calculated for the following set of values:

$$n_1 = 3,\ n_2 = 5,\ n_3 = 8 \text{ and } K = 3$$

and tabulated below for values of w differing by ·1.

$W =$ ·1,	·2,	·3,		·4,	·5,
$P =$ 1,	1,	1,		·7262946*	·6064524*
$W =$ ·6,		·7,		·8,	·9,
$P =$ ·4313434,		·2086090,		·0739872,	·0086890,

* These probabilities (for w = ·4 and ·5) have been calculated on the assumption of the independence of the ratios and since the ratios are not independent $P(w_i w_j)$ must be less than $P(w_i)\, P(w_j)$ for $P(w_i) > P(w_i/w_j)$.

These are therefore not the actual probabilities which must be more than these values.

Tables of P for different values of w can be prepared for ready reference.

SUMMARY

In order to test the homogeneity of the samples from a rectangular population it has been suggested that the ranges of the samples can be utilised for the purpose, the usual method of comparing variances being not available for this type of population. The calculations necessary for the purpose unfortunately turn out to be rather heavy.

REFERENCES

1. Lal, D. N. .. "A note on the distribution of the product of ranges in samples from rectangular population" (unpublished), 1955.

2. Cochran, W. G. .. "The distribution of the largest of a set of estimated variances as a fraction of their total," *Ann. Eug.*, 1941–42, **11**, 48–52.

3. Paul, R. Rider .. "The distribution of the product of ranges in samples from a rectangular population," *J.A.S.A.*, 1953, **48**, 546–49.