LIMITING DISTRIBUTIONS OF HOMOGENEOUS FUNCTIONS OF SAMPLE SPACINGS[1]

By Lionel Weiss

Cornell University

1. Summary. Suppose $T_1, T_2, \cdots, T_n$ are the lengths of n subintervals into which the interval [0, 1] is broken by $(n - 1)$ independent chance variables, each with a uniform distribution on [0, 1]. Moran [1], Kimball [2], and Darling [3] have shown that if r is a positive number, then the asymptotic distribution of $T_1^r + T_2^r + \cdots + T_n^r$ is normal. It is the purpose of this note to extend this result in two directions: more general functions of $T_1, \cdots, T_n$ are handled, and the joint distribution of several such functions is discussed. The proof is short and very simple.

2. Notation and assumptions. As already indicated, $T_1, T_2, \cdots, T_n$ are the n subintervals into which the unit interval is randomly broken. $U_1, U_2, \cdots, U_n$ are independent chance variables, each with the density function e^{-u} for $u \geqq 0$, zero for $u < 0$. $S_n = U_1 + U_2 + \cdots + U_n$. $V_i = U_i/S_n$ for $i = 1, \cdots, n$. It is known (and is very easily verified) that S_n is distributed independently of $(V_1, V_2, \cdots, V_n)$, and that the joint distribution of

$$(V_1, V_2, \cdots, V_n)$$

is exactly the same as the joint distribution of $T_1, T_2, \cdots, T_n$.

We are given k sequences of functions:

$$\{G_{1,n}(U_1, U_2, \cdots, U_n)\}, \cdots, \{G_{k,n}(U_1, U_2, \cdots, U_n)\},$$

$n = 1, 2, \cdots$. These functions are assumed to satisfy the following conditions:

(1) $G_{i,n}(U_1, \cdots, U_n)$ is homogeneous of order r_i for all n, r_i a positive quantity;

(2) the joint distribution of

$$\frac{G_{1,n}(U_1, \cdots, U_n) - A_1 n}{B_1\sqrt{n}}, \cdots, \frac{G_{k,n}(U_1, \cdots, U_n) - A_k n}{B_k\sqrt{n}}$$

approaches a k-variate normal distribution with zero means and covariance matrix C, say, as n increases. $A_1, \cdots, A_k$ and $B_1, \cdots, B_k$ are positive constants. (The results hold for any values of $A_1, \cdots, A_k$. The assumption that they are positive is merely a convenience.)

We denote the element of C in row i and column j by c_{ij}.

3. The asymptotic distribution of $G_{1,n}(T_1, \cdots, T_n), \cdots, G_{k,n}(T_1, \cdots, T_n)$.

Theorem. *Under the assumptions of Sec. 2, the joint distribution of*

$$\frac{n^{r_1} G_{1,n}(T_1, \cdots, T_n) - A_1 n}{B_1\sqrt{n}}, \cdots, \frac{n^{r_k} G_{k,n}(T_1, \cdots, T_n) - A_k n}{B_k\sqrt{n}}$$

Received July 8, 1957.

[1] Research supported by the Office of Naval Research.

approaches a k-variate normal distribution with zero means and covariance matrix

$$\left\{c_{ij} - \frac{r_r r_j A_i A_j}{B_i B_j}\right\}$$

as n increases.

PROOF. By assumption, the distribution of the k-dimensional vector $\bar{V}(n)$ whose ith element is

$$\frac{G_{i,n}(U_1, \cdots, U_n) - A_i n}{B_i\sqrt{n}}$$

approaches the k-variate normal distribution with zero means and covariance matrix C. We rewrite the ith term of $\bar{V}(n)$ as

$$\frac{G_{i,n}(U_1, \cdots, U_n) - S_n^{r_i} A_i n^{1-r_i} + S_n^{r_i} A_i n^{1-r_i} - A_i n.}{B_i\sqrt{n}}$$

Now S_n/n converges stochastically to one as n increases; therefore the distribution of the k-dimensional vector $\bar{V}'(n)$ whose ith element is

$$\frac{G_{i,n}(U_1, \cdots, U_n) - S_n^{r_i} A_i n^{1-r_i} + S_n^{r_i} A_i n^{1-r_i} - A_i n}{\left(\frac{S_n}{n}\right)^{r_i} B_i\sqrt{n}}$$

approaches the k-variate normal distribution with zero means and covariance matrix C. $\bar{V}'(n)$ may be written as the sum of two vectors, $\bar{V}_1(n)$ and $\bar{V}_2(n)$, whose ith elements are respectively

$$\frac{n^{r_i} G_{i,n}(V_1, \cdots, V_n) - A_i n}{B_i\sqrt{n}}$$

and

$$\frac{A_i n - n^{r_i+1} A_i S_n^{-r_i}}{B_i\sqrt{n}}$$

We note that $\bar{V}_1(n)$ and $\bar{V}_2(n)$ are distributed independently of each other.

Next we examine the distribution function, say $F_n(x_1, \cdots, x_k)$, of $\bar{V}_2(n)$.

$$F_n(x_1, \cdots, x_k) = \Pr\left[\frac{A_i n - n^{r_i-1} A_i S_n^{-r_i}}{B_i\sqrt{n}} \leqq x_i ; i = 1, \cdots, k\right]$$

$$= \Pr\left[\frac{S_n - n}{\sqrt{n}} \leqq \sqrt{n}\left\{\left(\frac{A_i n}{A_i n - \sqrt{n}\, B_i x_i}\right)^{\frac{1}{r_i}} - 1\right\}, \quad i = 1, \cdots k\right].$$

As n increases, the distribution of $(S_n - n)/\sqrt{n}$ approaches the standard normal distribution, by the univariate central-limit theorem. And for any fixed x_i,

$$\sqrt{n}\left\{\left(\frac{A_i n}{A_i n - \sqrt{n}\, B_i x_i}\right)^{\frac{1}{r_i}} - 1\right\} \to \frac{B_i x_i}{r_i A_i}$$

as n increases. Thus, if Z denotes a chance variable with a standard normal distribution, $F_n(x_1, \cdots, x_k)$ approaches

$$\Pr\left[\frac{r_i A_i Z}{B_i} \leqq x_i \,;\, i = 1, \cdots, k\right]$$

for each vector $(x_1, \cdots, x_k)$.

Next, we denote by $\rho_{1,n}(t_1, \cdots, t_k)$ the characteristic function of $\bar{V}_1(n)$, by $\rho_{2,n}(t_1, \cdots, t_k)$ the characteristic function of $\bar{V}_2(n)$, and by $\rho_n(t_1, \cdots, t_k)$ the characteristic function of $\bar{V}'(n)$.

We have $\rho_n(t_1, \cdots, t_k) = \rho_{1,n}(t_1, \cdots, t_k)\cdot\rho_{2,n}(t_1, \cdots, t_k)$, or

$$\rho_{1,n}(t_1, \cdots, t_k) = \frac{\rho_n(t_1, \cdots, t_k)}{\rho_{2,n}(t_1, \cdots t_k)}.$$

As n increases,

$$\rho_n(t_1, \cdots, t_k) \to \exp\left\{-\frac{1}{2}\sum\sum_{i,j=1}^{k} c_{ij} t_i t_j\right\}$$

and

$$\rho_{2,n}(t_1, \cdots, t_k) \to \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{k} \frac{t_j r_j A_j}{B_j}\right]^2\right\}.$$

Therefore, as n increases,

$$\rho_{1,n}(t_i, \cdot \cdot, t_k) \to \exp\left\{-\frac{1}{2}\sum\sum_{i,j=1}^{k} t_i t_j \left[c_{ij} - \frac{r_i r_j A_i A_j}{B_i B_j}\right]\right\}.$$

This proves the theorem.

REFERENCES

[1] P. Moran, "The random division of an interval," *J. Roy. Stat. Soc., Suppl.*, Vol. 9 (1947), pp. 92–98.

[2] B. F. Kimball, "Some basic theorems for developing tests of fit for the case of non-parametric probability distribution functions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 540–548.

[3] D. A. Darling, "On a class of problems related to the random division of an interval," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 239–253.