A THEOREM ON THE CORRELATION COEFFICIENT FOR SAMPLES OF THREE WHEN THE VARIABLES ARE INDEPENDENT

By C. Chandra Sekar[1]

All India Institute of Hygiene and Public Health, Calcutta

In this note the following theorem will be established:

Theorem. *If (x_i, y_i) for $i = 1, 2$ and 3 denote three pairs of random values of two independent continuous stochastic variables x and y, r, their correlation coefficient, is given by*

$$r = \frac{1}{3 s_x s_y} \sum_{i=1}^{3} (x_i - \bar{x})(y_i - \bar{y}), \tag{1}$$

where

$$\bar{x} = \frac{1}{3}\sum_{i=1}^{3} x_i, \qquad \bar{y} = \frac{1}{3}\sum_{i=1}^{3} y_i,$$
$$s_x^2 = \frac{1}{3}\sum_{i=1}^{3}(x_i - \bar{x})^2, \qquad s_y^2 = \frac{1}{3}\sum_{i=1}^{3}(y_i - \bar{y})^2, \tag{2}$$

and $P(a \le r \le b)$ denotes the probability of r taking values in the range $a \le r \le b$, then

$$P\left(-1 \le r \le -\frac{1}{2}\right) = P\left(-\frac{1}{2} \le r \le \frac{1}{2}\right) = P\left(\frac{1}{2} \le r \le 1\right) = \frac{1}{3}. \tag{3}$$

Proof. If τ_i is defined by

$$\tau_i = \frac{x_i - \bar{x}}{s_x}, \qquad i = 1, 2, 3, \tag{4}$$

it is readily seen that the three values of τ are connected by the two relations

$$\sum_{i=1}^{3} \tau_i = 0, \qquad \sum_{i=1}^{3} \tau_i^2 = 3. \tag{5}$$

Similar conditions exist between the three t_i's defined by

$$t_i = \frac{y_i - \bar{y}}{s_y}, \qquad i = 1, 2, 3. \tag{6}$$

The set (τ_1, τ_2, τ_3) can be considered as the Cartesian coordinates of a point in three dimensional space. The conditions (5) restrict the point to a circle. The set (t_1, t_2, t_3) defined by (6) represents a point on the same circle. The correlation coefficient, r, defined in (1) and also given by

$$r = \tfrac{1}{3}\sum_{i=1}^{3} \tau_i t_i \tag{7}$$

[1] On loan to Population Division, United Nations.

can be regarded as the cosine of the angle θ between the lines joining (τ_1, τ_2, τ_3) and (t_1, t_2, t_3) respectively to the centre of the above-mentioned circle.

The relationships between the τ_i's given by (5) make it necessary for one value of the τ_i's to occur in each of the three non-overlapping intervals $-\sqrt{2}$ to $-\frac{1}{\sqrt{2}}$; $-\frac{1}{\sqrt{2}}$ to $\frac{1}{\sqrt{2}}$ and $\frac{1}{\sqrt{2}}$ to $\sqrt{2}$. Exactly the same conditions hold for the t_i's.[2]

The 6 permutations of τ_1, τ_2, τ_3 in these three intervals correspond to a subdivision of the circle on which the point (τ_1, τ_2, τ_3) lies into 6 equal arcs of 60° each. Every point on any one of these arcs can be shown to correspond, one to one, to the position of τ_i in any one of the intervals; also proceeding along the circle, points on three alternate arcs correspond to the positions of τ_i as it takes on values from the highest to the lowest in this interval and points on the other three correspond to the positions of τ_i as it moves from the lowest to the highest value.

It is clear that when adjacent arcs are combined in pairs dividing the circle into 3 equal arcs of 120°, the probability density function of (τ_1, τ_2, τ_3) is the same on the 3 arcs and is symmetric on each. At any three points on the circle which divide it into three arcs of 120°, the probability density function of (τ_1, τ_2, τ_3) is therefore the same. The same conditions hold for (t_1, t_2, t_3).

It therefore follows that

$$(8) \quad \begin{aligned} P\left(-\frac{\pi}{3} < \theta \leq \frac{\pi}{3}\right) &= P\left(-\frac{2\pi}{3} < \theta \leq -\frac{\pi}{3} \text{ or } \frac{\pi}{3} < \theta \leq \frac{2\pi}{3}\right) \\ &= P\left(-\pi < \theta \leq -\frac{2\pi}{3} \text{ or } \frac{2\pi}{3} < \theta \leq \pi\right). \end{aligned}$$

CORRECTION TO "THE DISTRIBUTION OF EXTREME VALUES IN SAMPLES WHOSE MEMBERS ARE SUBJECT TO A MARKOFF CHAIN CONDITION"

By Benjamin Epstein

Wayne University

In the paper mentioned in the title (*Annals of Math. Stat.*, Vol. 20 (1949), pp. 590–594) I claim to have proved a number of results dealing with the distribution of extreme values in samples of size n drawn at equally spaced intervals from a stationary Markoff process. As Professor W. Feller has kindly pointed

[2] This property has been utilised by the author and S. C. Bhoumik to obtain distributions of the correlation coefficient for samples of three, under varying assumptions regarding the distributions of independent variables x and y. The distribution of τ_i or t_i is also of help in working out the distribution of Fisher's g_1 for samples of three. For the distribution of g_1 for samples of three from continuous rectangular distribution, refer to C. Chandra Sekar in *Current Science*, Vol. 13 (1944), pp. 10–11.