ON THE FREQUENCY DISTRIBUTION OF CERTAIN RATIOS

By H. L. Rietz

University of Iowa

Considerable interest in the distribution of ratios, $t = y/x$, has no doubt been suggested by important applications. For example, we may mention the opsonic index in bacteriology, the ratio of systolic to diastolic blood pressure in physiology, and ratios such as link relatives or certain index numbers in economics.

In 1910, Karl Pearson[1] gave certain properties of the distribution of ratios by means of approximate formulas for moments up to order four in terms of means, variances, product moments, and coefficients of variability of x and y. The resulting formulas did not give, with sufficient accuracy, the constants of the distribution of the opsonic index for the purpose of Dr. Greenwood to whom Pearson attributed the derivation of the formulas for the special case in which x and y are uncorrelated. Pearson next adopted the plan of tabulating the reciprocals, say $x' = \frac{1}{x}$, and then finding the constants of the distribution of the product yx' in the case in which x' and y are uncorrelated. He then obtained satisfactory results in illustrative examples.

In 1929, C. C. Craig[2] obtained the semi-invariants of y/x in terms of moments of x and y, and then expressed the moments in terms of the semi-invariants of the distribution function, $f(x, y)$, of x and y. By this means, he was able to deal with the case in which x and y are normally correlated under suitable conditions. Craig found it desirable to restrict the distribution of x in such a way that the probability of a zero value of x is an infinitesimal of sufficiently high order that a certain integral exists. This limitation seems to imply in applications to actual data that no zero values of x are to occur. This suggests that we deal with the cases of x at or near zero with considerable care.

By starting with the assumption that the values of x and y are a set of normally distributed pairs of values with correlation coefficient r, and by considering the quotient $z = \frac{b + y}{a + x}$, a and b being constants, R. C. Geary,[3] in a paper published in 1930, found an algebraic function, $u = f(z)$, of fairly simple form with the property that u is nearly normally distributed with arithmetic mean zero and standard deviation unity provided that $a + x$ is unlikely to

[1] On the constants of index distributions, Biometrika, Vol. 7 (1910), pp. 531–546.

[2] The frequency function of y/x, Annals of Mathematics, Vol. 30 (1928–29), pp. 471–486.

[3] The frequency distribution of the quotient of two normal variates, J. Royal Statistical Society, Vol. XCIII (1930), pp. 442–7.

have negative values. Here we have again a suggestion to exercise special care in the case of quotients with the divisor near zero or negative.

In 1932, Fieller[4] obtained in explicit form the approximate distribution of $t = y/x$ where values (x, y) are drawn from the bivariate normal distribution

$$\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\, e^{-\frac{1}{2}\frac{1}{1-r^2}\left\{\frac{(x-\bar{x})^2}{\sigma_x^2}+\frac{(y-\bar{y})^2}{\sigma_y^2}-2r\frac{(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y}\right\}}$$

under the condition that $\bar{x}$ is large compared with σ_x.

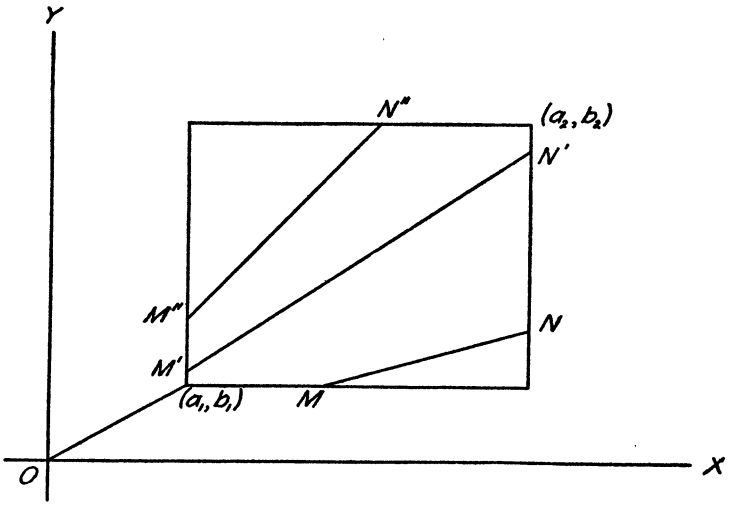

FIG. 1

Very recently Kullback[5] found the distribution law of the quotient, $t = y/x$, where x and y are drawn from Pearson Type III parent populations given by

$$f_1(x) = \frac{e^{-x}x^{p-1}}{\Gamma(p)}; \qquad f_2(y) = \frac{e^{-y}y^{q-1}}{\Gamma(q)}, \qquad 0 \leqq x \leqq \infty, \quad 0 \leqq y \leqq \infty.$$

It is fairly easy to see, in a general way, that the distribution of $t = y/x$ depends very much on the location of the origin as well as on the parent distribution from which x and y are drawn. This fact will be fairly obvious from the present paper whose main purpose is to give clear geometrical descriptions of the distributions of ratios, $t = y/x$, for each of several cases in which (x, y) are points taken at random from certain simple geometrical figures conveniently located with respect to the origin.

In accord with the suggestions to be cautious when the divisor is near zero or negative, we consider first the very simple case of ratios $t = y/x$ obtained

[4] E. C. Fieller, The distribution of the index in a normal bivariate population, Biometrika, Vol. 24 (1932), pp. 428–440.

[5] Solomon Kullback, Annals of Mathematical Statistics, Vol. VII (1936), pp. 51–53.

from points uniformly distributed over a rectangle such as is shown in Fig. 1 with sides parallel to coordinate axes and $a_1 > 0$, $b_1 > 0$. As indicated on Fig. 1, we assume for simplicity that the coordinates of the points are positive and $a_1 \leqq x \leqq a_2$, $b_1 \leqq y \leqq b_2$.

Case I. When $\frac{b_1}{a_1} \leqq \frac{b_2}{a_2}$, Fig. 1.

Let $k\,dx\,dy$ be the probability that a point (x, y) taken at random in the rectangle will fall into $dxdy$ where k is a constant. Then

$$k\int_{b_1}^{b_2}\int_{a_1}^{a_2} dxdy = k(a_2 - a_1)(b_2 - b_1) = 1,$$

and

$$k = \frac{1}{(a_2 - a_1)(b_2 - b_1)}.$$

Transform the element $k\,dxdy$ into one with variables t, and x by making

$$x = x,$$

$$y = tx.$$

The Jacobian is $|x| = x$.

The new element is $k\,x\,dxdt$ and is to be integrated over the range on x for an assigned t in order to get the probability, to within infinitesimals of higher order, that a random t falls into an assigned dt. By assigning t any value such that $\frac{b_1}{a_2} \leqq t \leqq \frac{b_1}{a_1}$, say t is the slope of MN, (Fig. 1), we have

$$(1) \qquad k\int_{\frac{b_1}{t}}^{a_2} xdxdt = \frac{k}{2}\left(a_2^2 - \frac{b_1^2}{t^2}\right)dt$$

the limits of integration being indicated by the ends of the line MN.

When the assigned t is such that $\frac{b_1}{a_1} \leqq t \leqq \frac{b_2}{a_2}$, say t is the slope of the line $M'N'$, we have

$$(2) \qquad k\int_{a_1}^{a_2} x\,dx\,dt = \frac{k}{2}(a_2^2 - a_1^2)\,dt$$

When the assigned t is such that $\frac{b_2}{a_2} \leqq t \leqq \frac{b_2}{a_1}$, say it is the slope of $M''N''$, we have

$$(3) \qquad k\int_{a_1}^{\frac{b_2}{t}} x\,dx\,dt = \frac{k}{2}\left(\frac{b_2^2}{t^2} - a_1^2\right)dt$$

Thus, from (1), (2), (3), when as in Fig. 1, $\frac{b_1}{a_1} \leqq \frac{b_2}{a_2}$, the frequency function of t is given by

$$F(t) = \frac{k}{2}\left(a_2^2 - \frac{b_1^2}{t^2}\right) \quad \text{when} \quad \frac{b_1}{a_2} \leqq t \leqq \frac{b_1}{a_1}, \tag{4}$$

$$F(t) = \frac{k}{2}(a_2^2 - a_1^2) \quad \text{when} \quad \frac{b_1}{a_1} \leqq t \leqq \frac{b_2}{a_2}, \tag{5}$$

$$F(t) = \frac{k}{2}\left(\frac{b_2^2}{t^2} - a_1^2\right) \quad \text{when} \quad \frac{b_2}{a_2} \leqq t \leqq \frac{b_2}{a_1}. \tag{6}$$

See Fig. 2 for the general form of the frequency curve $F(t)$ when $\frac{b_1}{a_1} < \frac{b_2}{a_2}$ with the segment from $t = \frac{b_1}{a_1}$ to $\frac{b_2}{a_2}$ a horizontal straight line and with discontinuities in the first derivatives of $F(t)$ at $t = \frac{b_1}{a_1}$ and $t = \frac{b_2}{a_2}$.

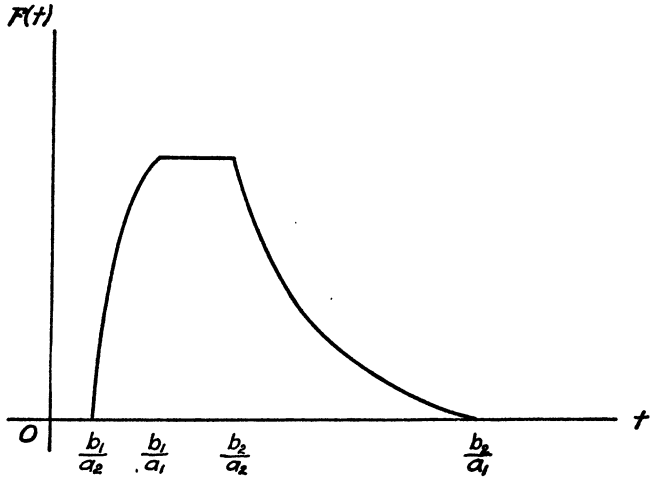

FIG. 2

When $a_1 \to 0$, and $b_1 = 0$, the frequency curve approaches

$$F(t) = \frac{a_2}{2b_2} \quad \text{when} \quad 0 \leqq t \leqq \frac{b_2}{a_2} \tag{7}$$

$$F(t) = \frac{b_2}{2a_2t^2} \quad \text{when} \quad t \geqq \frac{b_2}{a_2}. \tag{8}$$

It may be noted that the curve given by making $a_1 = 0$ and $b_1 = 0$ extends to infinity, and that the first and second moments about the origin are each infinite.

Case II. When $\frac{b_1}{a_1} > \frac{b_2}{a_2}$.

If the rectangle in Fig. 1 were moved upward keeping its sides parallel to the x and y axes until $\frac{b_1}{a_1} > \frac{b_2}{a_2}$, we would obtain

$$F(t) = \frac{k}{2}\left(a_2^2 - \frac{b_1^2}{t^2}\right) \quad \text{if} \quad \frac{b_1}{a_2} \leqq t \leqq \frac{b_2}{a_2}, \tag{9}$$

$$F(t) = \frac{k}{2t^2}(b_2^2 - b_1^2) \quad \text{if} \quad \frac{b_2}{a_2} \leqq t \leqq \frac{b_1}{a_1}, \tag{10}$$

$$F(t) = \frac{k}{2}\left(\frac{b_2^2}{t^2} - a_1^2\right) \quad \text{if} \quad \frac{b_1}{a_1} \leqq t \leqq \frac{b_2}{a_1}. \tag{11}$$

By comparing (5) and (10), it may be observed that $F(t)$ of the middle segment of the distribution curve differs much in Case II from its corresponding constant value in Case I.

By moving the rectangle of Fig. 1 downward, keeping its sides parallel to the x and y axes until b_1 is negative, we easily find further forms of the distribution curve $F(t)$.

To consider the distribution of the ratio $t = y/x$ for another very simple type of distribution of x and y, suppose we have given the distribution function

$$f(x, y) = k\, e^{-\frac{x}{a} - \frac{y}{b}}, \begin{pmatrix} x \geqq c > 0, \, y \text{ non-negative} \\ a > c, \, b > 0 \end{pmatrix} \tag{12}$$

where $\int_0^\infty \int_c^\infty f(x, y)\, dx\, dy = 1$. Then

$$k = \frac{e^{c/a}}{ab}.$$

In this case,

$$\begin{aligned} F(t) &= \frac{e^{c/a}}{ab} \int_c^\infty x\, e^{-\frac{x}{a} - \frac{xt}{b}}\, dx \\ &= \frac{1}{b + at}\left(c + \frac{ab}{b + at}\right) e^{-\frac{ct}{b}}, \end{aligned} \tag{13}$$

a monotone decreasing function from $t = 0$ to $t = \infty$.

With $c = 0$ as a limiting value, we obtain

$$F(t) = \frac{ab}{(b + at)^2}, \tag{14}$$

a distribution curve with the mean value of t at infinity.

If we should similarly consider

$$f(x, y) = \frac{2}{\pi\sigma_x\sigma_y} e^{-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}} \quad (x \text{ and } y \text{ non-negative}) \tag{15}$$

we easily obtain

$$F(t) = \frac{1}{2\pi\sigma_x\sigma_y\left(\frac{1}{\sigma_x^2} + \frac{t^2}{\sigma_y^2}\right)} \tag{16}$$

as the distribution function.

Although the difficulties[6] of the problem of the distribution of the ratio y/x when x and y are normally correlated have been overcome[7] to a considerable

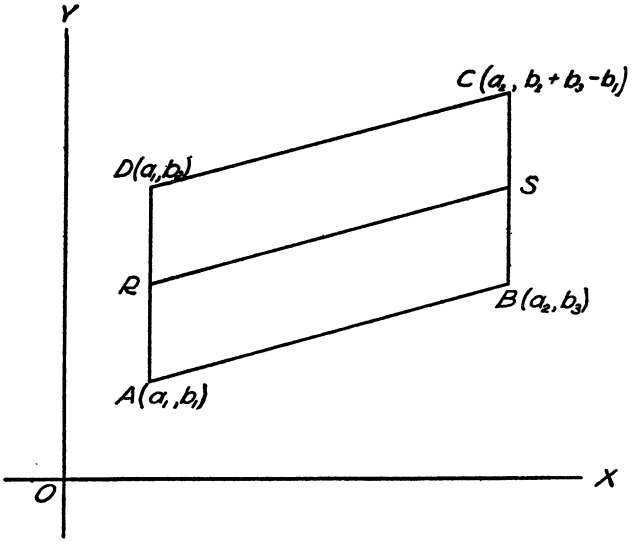

FIG. 3

extent, still the examination of some very simple geometric cases of non-normal but linear correlation may not be without some interest. Such a case will now be considered.

For one very simple case in which x and y are correlated, suppose we are given a set of points (x, y) uniformly distributed over the parallelogram $ABCD$ (Fig. 3) with sides AD and BC parallel to the y-axis so that the regression of y on x is linear as shown by the line RS.

The equation of RS is

$$y = m(x - a_1) + \frac{b_1 + b_2}{2}. \tag{17}$$

[6] Loc. cit., Pearson, p. 531.

[7] Loc. cit., C. C. Craig, R. C. Geary, E. C. Fieller.

Then although x_i and y_i are correlated, x_i and

$$y_i' = y_i - m(x_i - a_i) - \frac{b_1 + b_2}{2}$$

are uncorrelated. Let us consider the distribution of the ratio $t' = \frac{y_i'}{x_i}$.

Consider the element of frequency $kdxdy'$, where

$$k(b_2 - b_1)(a_2 - a_1) = 1. \tag{18}$$

Change variables to x and t' by the transformation

$$x = x,$$
$$y' = t'x.$$

Then the element of frequency becomes

$$kx\,dx\,dt'. \tag{19}$$

Next integrate (19) with respect to x under the restriction that t' is assigned. Three cases occur:

(a) When $-\frac{b_2 - b_1}{2a_2} \leqq t' \leqq \frac{b_2 - b_1}{2a_2}$, we obtain by integration of (19) for the element of relative frequency of t' in dt',

$$k\int_{a_1}^{a_2} x\,dx\,dt' = \frac{k}{2}(a_2^2 - a_1^2)\,dt'. \tag{20}$$

(b) When $t' \geqq \frac{b_2 - b_1}{2a_2}$, we obtain

$$k\int_{a_1}^{\frac{b_2-b_1}{2t'}} x\,dx\,dt' = \frac{k}{2}\left[\frac{(b_2 - b_1)^2}{4t'^2} - a_1^2\right]dt' \tag{21}$$

(c) When $t' \leqq -\frac{b_2 - b_1}{2a_2}$, we similarly obtain

$$k\int_{a_1}^{-\frac{b_2-b_1}{2t'}} x\,dx\,dt' = \frac{k}{2}\left[\frac{(b_2 - b_1)^2}{4t'^2} - a_1^2\right]dt' \tag{22}$$

From (18), (19), (20), (21) and (22), the frequency function of t' is given by

$$F(t') = \frac{a_2 + a_1}{2(b_2 - b_1)} \quad \text{when} \quad -\frac{b_2 - b_1}{2a_2} \leqq t' \leqq \frac{b_2 - b_1}{2a_2}; \tag{23}$$

$$F(t') = \frac{1}{2(b_2 - b_1)(a_2 - a_1)}\left[\frac{(b_2 - b_1)^2}{4t'^2} - a_1^2\right], \tag{24}$$

where the range of t' is subject to either the inequalities,

$$\frac{b_2 - b_1}{2a_2} \leqq t' \leqq \frac{b_2 - b_1}{2a_1}, \quad \text{or} \quad -\frac{b_2 - b_1}{2a_1} \leqq t' \leqq -\frac{b_2 - b_1}{2a_2}.$$

See Fig. 4 for the general form of the $F(t')$ frequency curve.
If we make $a_1 = 0$, the curve becomes infinite in range. If we make not only $a_1 = 0$, but $(b_1 + b_2)/2 = 0$, we have, in place of (17),

$$y = mx.$$

In this limiting situation, if we make $a_2 = a$ and $\frac{b_2 - b_1}{2} = b$,

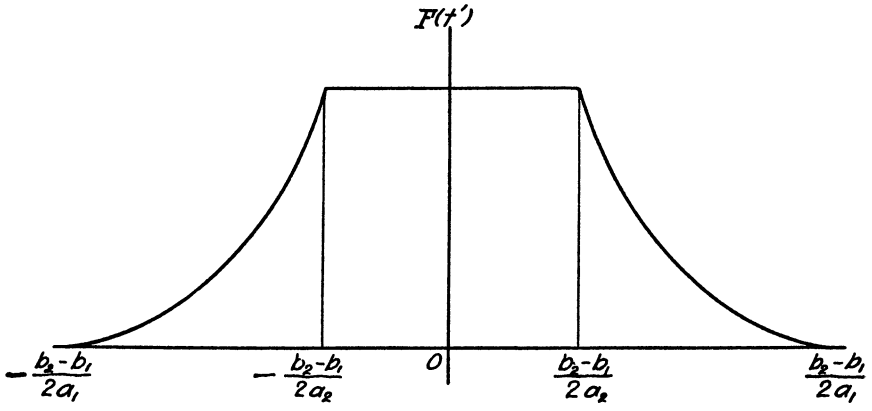

FIG. 4

(23) becomes

$$F(t') = \frac{a}{4b}, \quad \text{for} \quad -\frac{b}{a} \leqq t' \leqq \frac{b}{a}, \tag{25}$$

and (24) becomes

$$F(t') = \frac{b}{4\,at'^2} \quad \text{for} \quad t' \geqq \frac{b}{a} \quad \text{and for} \quad t' \leqq -\frac{b}{a}. \tag{26}$$

Then we have $y' = y - mx$

and

$$t' = \frac{y'}{x} = t - m.$$

Further, if t' is distributed in accord with a frequency function, $F(t')$, the distribution of $t = t' + m$ with m constant is given by

$$F(t - m).$$

Hence, the probability that a random value t will fall into a range t to $t + dt$ is given to within infinitesimals of higher order by

$$\frac{a}{4b}\,dt \quad \text{when} \quad m - \frac{b}{a} \leqq t \leqq m + \frac{b}{a}, \tag{27}$$

and by

$$\frac{b\,dt}{4a(t-m)^2} \quad \text{when} \quad t \geqq m + \frac{b}{a} \text{ and } t \leqq m - \frac{b}{a}. \tag{28}$$

With the frequency curve given by (27) and (28) we may note that the variance of t becomes infinite.

Without taking the space to continue illustrations, it is fairly obvious that a wide diversity of form can be given to the frequency function of the quotients $t = y/x$ by relatively simple changes in the location of a sample parent population with reference to the origin.