CHAPTER XI

MULTIVARIATE STATISTICS. CORRELATION

A. Measures of Correlation in Two Dimensions (Sections 1-3)

1. Correlation

1.1. *The problem.* In previous chapters of this book, various distribution parameters have been studied, the simplest of which were mean value and variance, then moments of higher order, etc. A specific need to characterize certain distribution properties arises in the case of distributions of more than one variable. We first take up the case of two variates x, y and designate by $P(x, y)$ the c.d.f., referring either to the probability (chance) of $\xi \leqslant x$, $\eta \leqslant y$, or, in the case of a sample distribution, to the relative number of individuals for which the observed ξ- and η-values do not exceed x and y, respectively. We write $p(x, y)$ for either the probability (probability density) or the frequency (frequency density) as the case may be.[1] We thus have

$$\iint dP(x, y) = 1 \quad \text{and either} \quad \iint p(x, y)\, dx\, dy = 1, \quad \text{or} \quad \sum_{x,y} p(x, y) = 1\,. \tag{1}$$

In Chapter III we introduced the *marginal distributions* that can be derived from a multivariate distribution. Assume that $p(x, y)$ is the probability of the combination x, y occurring in a collective with discrete two-dimensional label values. Then, the probability of getting an x combined with *any* y-value is

$$p_1(x) = \sum_y p(x, y). \tag{2}$$

[1] In the first three sections of this chapter we have no need to distinguish between frequencies and probabilities (sample distributions and theoretical distributions) and therefore use the notation $p(x, y)$, $P(x, y)$ indiscriminately for both types of distributions. The corresponding parameters (mean values, variances, etc.) are then denoted by Latin letters (a, b, s_{11}, s_{12}, s_{22}, etc.)

The summation has to be extended over all y-values with x kept constant. This distribution $p_1(x)$ and also the analogous distribution of y

$$p_2(y) = \sum_x p(x, y) \tag{2'}$$

are called the marginal distributions belonging to $p(x, y)$. Summing Eq. (2) over x [and similarly, Eq. (2') over y] and using (1) we find $\Sigma_x p_1(x) = 1$, $\Sigma_y p_2(y) = 1$.

In the case of a sample of size n, the sampling distribution might be given in the form of a rectangular table

$$\begin{array}{cccc} n_{11} & n_{12} & \cdots & n_{1l} \\ n_{21} & n_{22} & \cdots & n_{2l} \\ \cdot & \cdot & \cdots & \\ n_{k1} & n_{k2} & \cdots & n_{kl}\,, \end{array}$$

where n_{ij} is the number of individuals with the label value x_i, y_j ($i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, l$). Here, the sum of the kl values n_{ij} is n, the distribution consists of the kl quotients n_{ij}/n, and the two marginal distributions are the quotients of the sums of columns or rows in the table divided by n.

We now study two special kinds of two-dimensional distributions. The first is *the case where $p(x, y)$ is the product of a function of x times a function of y*: $p(x, y) = f(x)g(y)$. The corresponding marginal distributions according to (2) and (2') are

$$p_1(x) = f(x) \sum_y g(y) \qquad \text{and} \qquad p_2(y) = g(y) \sum_x f(x). \tag{3}$$

From (1) we derive

$$\sum_{x,y} p(x, y) = \sum_x f(x) \sum_y g(y) = 1$$

and, thus, from (3), by multiplying the two equations

$$p_1(x)p_2(y) = f(x)g(y) = p(x, y). \tag{3'}$$

If a distribution $p(x, y)$ is the product of a function $f(x)$ of x and a function $g(y)$ of y, it is also the product of the two marginal distributions $p_1(x)$ and $p_2(y)$. If $f(x)$ and $g(y)$ are distributions, and therefore $\Sigma_x f(x) = 1$, $\Sigma_y g(y) = 1$, then $f(x) = p_1(x)$, $g(y) = p_2(y)$.

Since the formula (3′) coincides with the multiplication rule of two independent collectives with distributions $p_1(x)$ and $p_2(y)$, the case now under consideration, i.e., a distribution $p(x, y)$ of the form $f(x)g(y)$, is called the case of two *stochastically independent* variables.

The type of distribution given by (3′) is one extreme case of a distribution in two variables. At the opposite extreme (the counterpart of two stochastically independent variables) is the case where the choice of one variable uniquely determines the second, that is, the probability $p(x, y)$ is different from zero only for such couples x, y for which y is a univalued function of x and vice versa. In other words, there exists only one non-vanishing p-value for each x and one for each y. For those couples x, y for which p is different from zero, one has then $p(x, y) = p_1(x) = p_2(y)$. This is called the case of two *completely dependent* variables. In all other circumstances, when $p(x, y)$ is neither a product $p_1(x)p_2(y)$ nor such that $p_1(x) = p_2(y) = p(x, y)$, we speak of two (*stochastically*) *correlated variables.*

Example. If on n days we measure the intensity of rainfall per day at two places A and B by weighing the quantity x of water accumulated during the day in some open receptable placed at A, and the quantity y accumulated in another vessel at B, we can obtain all three cases. If A is in the immediate neighborhood of B, x and y will be (almost) completely dependent, i.e., to each observed value of x a certain y will correspond and vice versa. If A and B are on opposite sides of the earth, we may expect x and y to be stochastically independent. If, however, A and B are separated by a distance of a few hundred miles, neither of the extreme cases will present itself, and we shall observe a certain correlation. The desideratum is to have a distribution parameter—a function of $p(x, y)$—which measures the "degree" of correlation.

A suitable correlation measure will have to fulfill the following conditions:

(a) It is zero if and only if $p(x, y) = p_1(x)p_2(y)$
(b) It has the absolute value 1 if and only if $p_1(x) = p_2(y) = p(x, y)$
(c) Its absolute value is between 0 and 1 in all other cases.

1.2. *The correlation coefficient.* The quantity usually called the *correlation coefficient* which we shall now discuss does not fulfill these conditions exactly. It measures the extent of *linear* dependency between x and y only, reaching ± 1 when there is a linear relation of the form $\gamma x + \delta y + \epsilon = 0$ between x and y.

As in Chapters III and VIII, we call a and b the two mean values and

s_{11}, s_{12}, s_{22} the three variance components of $p(x, y)$; for example, if x and y are discrete[2]

$$a = \sum_{x,y} xp(x, y), \qquad b = \sum_{x,y} yp(x, y) \tag{4}$$

$$s_{11} = \sum_{x,y} (x - a)^2 p(x, y), \qquad s_{12} = \sum_{x,y} (x - a)(y - b)p(x, y),$$

$$s_{22} = \sum_{x,y} (y - b)^2 p(x, y). \tag{5}$$

We have called $s_{11} = s_1^2$ and $s_{22} = s_2^2$ the variance of x or of y; for s_{12} the term covariance (of x and y) is used: $s_{12} = \mathrm{Cov}[x, y]$. Note that we also have

$$a = \sum_{x} xp_1(x), \qquad b = \sum_{y} yp_2(y) \tag{4'}$$

$$s_{11} = \sum_{x} (x - a)^2 p_1(x), \qquad s_{22} = \sum_{y} (y - b)^2 p_2(y). \tag{5'}$$

The reader should note that the present notation (where one also sets $s_{11} = s_1^2$, $s_{22} = s_2^2$) is adapted to describing moments of second order by a *variance matrix* in any number x_i, $i = 1, 2, \ldots, k$ of variables. We then write

$$s_{ij} = \int (x_i - a_i)(x_j - a_j)\, dP(x_i, x_j).$$

However, if moments of higher order are used, this notation is not adequate. Suppose there are two variabes x, y; then $m_{rs} = \int (x - a)^r (y - b)^s\, dP(x, y)$, (or M_{rs}) is an adequate notation for central moments (with a, b as mean values) of order r in x and order s in y. Then $s_{11} = m_{20}$, $s_{12} = m_{11}$, $s_{22} = m_{02}$. Right here the notation s_{ik} is practical.

The *correlation coefficient r is defined by*

$$r = \frac{s_{12}}{|\sqrt{s_{11}s_{22}}|}. \tag{6}$$

It can be seen at once that r fulfills condition (c) in a certain way. In fact, by Schwarz's inequality $s_{11}s_{22} \geqslant s_{12}^2$. Thus, it is proved that $|r| \leqslant 1$ for all $p(x, y)$.

We can also easily see that r vanishes in the case (3') of stochastic independence. Here we have indeed

$$s_{12} = \sum_{x} (x - a)p_1(x) \cdot \sum_{y} (y - b)p_2(y) = 0, \tag{7}$$

[2] More generally, we could use Stieltjes integrals in what follows.

since each of these two sums is zero by virtue of the first two Eqs. (4'). *The converse is not true.* We call two random variables *uncorrelated* if $r = 0$. Two uncorrelated variables need to be independent. Let $p(x, y) = e^{-\sqrt{x^2+y^2}}/2\pi\sqrt{x^2+y^2}$. Here $p(x, y)$ is constant on every circle $x^2 + y^2 = c^2$. Therefore the center of gravity is $a = b = 0$, and also $s_{12} = 0$, and $r = 0$. Nevertheless, $p(x, y)$ is not of the form $f(x)g(y)$.

Let us finally assume that $p(x, y)$ is different from zero only if x and y are related by the linear equation

$$\gamma x + \delta y + \epsilon = 0 \qquad (\gamma, \delta \neq 0). \tag{8}$$

If we multiply by $p(x, y)$ and sum over all x and y, we obtain

$$\gamma a + \delta b + \epsilon = 0 \tag{8'}$$

and subtracting this from (8)

$$\gamma(x - a) + \delta(y - b) = 0 \quad \text{or} \quad \gamma^2(x - a)^2 = \delta^2(y - b)^2. \tag{8''}$$

If this is used in computing s_{11}, s_{12}, and s_{22} from (5), we find

$$\gamma^2 s_{11} = \delta^2 s_{22}, \qquad -\gamma\delta s_{12} = \gamma^2 s_{11} = \delta^2 s_{22}$$

and, therefore,

$$r = \frac{-\gamma s_{11}/\delta}{|\sqrt{s_{11}\gamma^2 s_{11}/\delta^2}|} = \begin{cases} -1 & \text{for } \gamma/\delta > 0 \\ +1 & \text{for } \gamma/\delta < 0. \end{cases} \tag{9}$$

The correlation coefficient r defined by (6) *ranges from* -1 to $+1$; *it vanishes in the case of stochastically independent variables*; *it has the absolute value* 1 *if x and y are linearly connected, with $r = 1$ if x and y simultaneously increase and $r = -1$ if y decreases with increasing x.*

One often speaks of *positive correlation* when $r > 0$ and of *negative correlation* in the case $r < 0$.

The correlation coefficient r fails to fulfill the requirement (b) if non-linear relations between x and y are admitted. We show this by the following example. Assume that $p(x, y)$ is different from zero only if $xy = 1$ and that $p(x, y) = p(y, x)$. Then, we have

$$a = \sum_{x,y} xp(x, y) = \sum_{x,y} yp(y, x) = b;$$

$$s_{11} = \sum_{x,y} x^2 p(x, y) - a^2 = \sum_{x,y} y^2 p(x, y) - b^2 = s_{22};$$

$$s_{12} = \sum_{x,y} xyp(x, y) - ab = 1 - a^2$$

and thus,

$$r = \frac{1 - a^2}{\sum_{x,y} x^2 p(x, y) - a^2}.$$

For a given $a = \sum_{x,y} xp(x, y)$ the sum of $x^2 p(x, y)$ can have a value much greater than 1 and thus r may become as small as desired. To make this quite concrete call p_1 the probability at $x = y = 1$. The point with abscissa n (> 1) and the symmetric point with abscissa $1/n$ may each have the probability $p_n = 1/(n - 1)^2$ and we assume $p_1 + 2p_n = 1$. Then by an elementary computation $a = 1 + 1/n = b$, $s_{11} = s_{22} = 1$, $r = 1 - a^2 = -2/n - 1/n^2$ whose absolute value is obviously as small as desired. *In the case of non-linear complete dependence of x and y, the correlation coefficient r can have any small absolute value.* Its use as a measure of correlation must, therefore, be restricted to such problems where relations of at least approximately linear type only can be expected.

In Section 3, we will study a correlation measure which is not subject to this restriction, the so-called *contingency coefficient*. Other measures of correlation will also be mentioned there.

Finally we compute r for two important particular cases.

(a) For a *normal distribution* in two variables

$$p(x, y) = \text{const. } e^{-\frac{1}{2}Q}, \qquad Q = a_{11}x^2 + 2a_{12}xy + a_{22}y^2 \tag{10}$$

we have found in Chapter VIII, Section 8:

$$s_{11} : s_{12} : s_{22} = a_{22} : -a_{12} : a_{11}. \tag{10'}$$

Hence from the definition of r

$$r = \frac{-a_{12}}{\sqrt{a_{11}a_{22}}} \tag{11}$$

This case is often referred to as a *normal correlation.*

(b) We compute r for the case when the population of the sample under consideration is a *double alternative.* Assume that x as well as y can take only the values 0 and 1. The four $p(x, y)$-values may then be written as p_{00}, p_{10}, p_{01}, p_{11}. In this case we have (using the shift-of-origin rule)

$$p_{10} + p_{11} = a = s_{11} + a^2, \qquad p_{01} + p_{11} = b = s_{22} + b^2, \qquad p_{11} = s_{12} + ab$$

and from a simple computation, considering that $p_{00} + p_{10} + p_{01} + p_{11} = 1$,

$$s_{11} = (p_{10} + p_{11})(p_{00} + p_{01}), \quad s_{22} = (p_{01} + p_{11})(p_{00} + p_{10}), \quad s_{12} = p_{11}p_{00} - p_{01}p_{10},$$

and

$$r = \frac{p_{11}p_{00} - p_{10}p_{01}}{\sqrt{(p_{10} + p_{11})(p_{01} + p_{11})(p_{00} + p_{01})(p_{00} + p_{10})}}. \tag{12}$$

This case is often called a *Bernoullian correlation.* It is seen that r^2 assumes its maximum, 1, if either $p_{01} = p_{10} = 0$ $(r = +1)$, or $p_{00} = p_{11} = 0$ $(r = -1)$.

Problem 1. Compute the correlation coefficient for the following two densities:

$$(a)\ p(x, y) = \tfrac{3}{2}(x^2 + y^2), \qquad 0 \leqslant x, y \leqslant 1$$

$$(b)\ p(x, y) = \tfrac{1}{2}(x + y)e^{-x-y}, \qquad x, y \geqslant 0.$$

Problem 2. Three numbered (unbiased) coins are simultaneously tossed. Call x the number of heads appearing on the first two coins and y the number of heads on the second and third coins. Set up the scheme of the 9 probability values $p(x, y)$ for $x, y = 0, 1, 2$ and compute the correlation coefficient.

Problem 3. Answer the same question, if, instead of three coins, three unbiased dice are used and x, y are the sums of points counted on the first and second and on the second and third dice, respectively.

Problem 4. If $(2n - 1)$ equal coins (head probability $= q$) are tossed and $p(x, y)$ is the probability that x and y are the numbers of heads on the first n and the last n coins, respectively, show that with $p_\nu(x) = \binom{\nu}{x} q^x p^{\nu-x}$, $p(x, y) = (1 - q)p_{n-1}(x)p_{n-1}(y) + qp_{n-1}(x - 1)p_{n-1}(y - 1)$. Compute the correlation coefficient and show that it decreases indefinitely with increasing n.

Problem 5. If n observations of a two-dimensional variate lead to the results $x_1, y_1, \ldots, x_n, y_n$ prove that, with

$$a = \frac{1}{n}\sum_{\nu=1}^{n} x_\nu,\ b = \frac{1}{n}\sum_{\nu=1}^{n} y_\nu, \qquad r = \frac{\sum_{\nu=1}^{n}(x_\nu - a)(y_\nu - b)}{\sqrt{\sum_{\nu=1}^{n}(x_\nu - a)^2 \cdot \sum_{\nu=1}^{n}(y_\nu - b)^2}}.$$

Problem 6. The following result has been achieved in experimenting with a certain antitoxin: Out of 239 patients treated with the serum, 9 patients died, and out of 244 left without serum, 29 died. Compute the correlation coefficient between treatment and recovery.

Problem 7. The height of 330 young fir trees and the length of their topmost branches are given below. Compute the coefficient of correlation.

Height (cm)	Length of tops (cm)						
	1–4	5–8	9–12	13–16	17–20	21–24	25–28
1–10	5	3					
11–20	7	15	1	2			
21–30	2	21	17	2			
31–40		15	37	20	3		
41–50		1	31	26	4		
51–60		2	8	27	12	1	
61–70			2	9	24	4	2
71–80			1	4	7	6	2
81–90				1		3	
91–100						1	2

2. *Regression Lines*

2.1. *"General" regression lines.* Another approach to the study of multivariate distributions consists in computing the so-called *regression curves.* Given a distribution function $p(x, y)$—probabilities or relative frequencies or probability densities—we defined in (2) and (2′) the two marginal distributions $p_1(x)$ and $p_2(y)$. The conditional probability of x given y equals $p(x, y)/p_2(y)$. We write

$$p_1(x \mid y) = \frac{p(x, y)}{\sum_x p(x, y)} = \frac{p(x, y)}{p_2(y)}, \tag{13}$$

and in the same way

$$p_2(y \mid x) = \frac{p(x, y)}{\sum_y p(x, y)} = \frac{p(x, y)}{p_1(x)}, \tag{13′}$$

where y is the distribution variable and x the parameter on which this distribution of y, for a given x, depends (see Chapter III, Section 7).

If x and y are stochastically independent, that is, if $p(x, y) = p_1(x)p_2(y)$, it is seen from (13) and (13′) that

$$p_1(x \mid y) = p_1(x), \qquad p_2(y \mid x) = p_2(y). \tag{14}$$

In general, the distribution $p_2(y \mid x)$ will vary with the value of x

and so will the mean value of this distribution (see Chapter III, Section 8.2) which we may call $\bar{y}(x)$:

$$\bar{y}(x) = \sum_y y p_2(y \mid x) = \frac{\Sigma_y y p(x, y)}{p_1(x)}. \tag{15}$$

On each line parallel to the y-axis, the centroid of the "masses" on this line has the ordinate $\bar{y}(x)$. The locus of these centroids is one regression curve. Another one is defined in the same way by

$$\bar{x}(y) = \sum_x x p_1(x \mid y) = \frac{\Sigma_x x p(x, y)}{p_2(y)}. \tag{15'}$$

The lines which give $\bar{x}$ as a function of y, and $\bar{y}$ as a function of x, in the original x, y coordinates, are called *regression lines. The regression lines in the case of stochastic independence are straight lines, one parallel to the x-axis and the other parallel to the y-axis.* The point of intersection is obviously the mean value of $p(x, y)$:

$$\bar{y}(x) = \sum_y y p_2(y) = \sum_{x,y} y p(x, y) = b, \qquad \bar{x}(y) = \sum_x x p_1(x) = \sum_{x,y} x p(x, y) = a. \tag{16}$$

In the other extreme case, where x and y are completely dependent, the mean $\bar{y}(x)$ for a given x must coincide with the one y-value allowed for this x. The analogous statement being true for $\bar{x}(y)$, we have: *If x, y are completely dependent on each other, both regression lines coincide with the one curve which represents the corresponding x- and y-values.*

The regression lines (15) and (15') have an important minimum property. If x and y are not completely dependent on each other, we may ask for a function $y = g(x)$ of x which gives the best possible estimation of the other variable y. Interpreting the term "best possible" in the sense of the principle of least squares we wish to determine $g(x)$ so as to minimize the expression

$$\sum_x \sum_y [y - g(x)]^2 p(x, y) = \sum_x p_1(x) \sum_y [y - g(x)]^2 p_2(y \mid x).$$

From the minimum property of the variance, we see that, for each x, $\Sigma_y [y - g(x)]^2 p_2(y \mid x)$ becomes a minimum when $g(x) = \bar{y}(x)$, the mean value of $p_2(y \mid x)$. Thus, the *minimum of $\Sigma_x \Sigma_y [y - g(x)]^2 p(x, y)$ among all possible functions $g(x)$ is attained for $g(x) = \bar{y}(x)$. In the same way $\Sigma_x \Sigma_y [x - h(y)]^2 p(x, y)$ attains its minimum for $h(y) = \bar{x}(y)$.*

The above results regarding the two extreme cases show how the

stochastic relationship between x and y is reflected by the regression lines.

For the *normal distribution* (10) we find from the results in Section 9.3 of Chapter VIII or by direct computation:

$$\bar{y}(x) = -\frac{a_{12}}{a_{22}}x \qquad \text{and} \qquad \bar{x}(y) = -\frac{a_{12}}{a_{11}}y. \tag{17}$$

The regression curves in the case of a normal distribution are straight lines passing through the center (mean value) with slopes (relative to the x- and y-directions, respectively) equal to $-a_{12}/a_{22}$ *and* $-a_{12}/a_{11}$.

2.2. *Linear regression lines.* The above result suggests that it may be useful to study, in a more general way, the case of *straight regression lines.* Assume that

$$(L_1) \quad \bar{x}(y) = \beta_1 y + \gamma_1 , \qquad (L_2) \quad \bar{y}(x) = \beta_2 x + \gamma_2 \tag{18}$$

with β_1 , γ_1 , β_2 , γ_2 constant.[1] According to the definitions (15) and (2), the second of these equations is equivalent to

$$\sum_y yp(x, y) = \beta_2 x \sum_y p(x, y) + \gamma_2 \sum_y p(x, y). \tag{19}$$

Summing over x we obtain

$$b = \beta_2 a + \gamma_2 . \tag{19'}$$

This expresses the fact that the regression line, if it is a straight line, must pass through the center a, b. If we multiply both sides of (19) by x, and then sum over x we obtain

$$s_{12} + ab = \beta_2(s_{11} + a^2) + \gamma_2 a. \tag{20}$$

Subtracting Eq. (19′) multiplied by a from this equation, we find

$$s_{12} = \beta_2 s_{11} \qquad \text{or} \qquad \beta_2 = \frac{s_{12}}{s_{11}} = r\sqrt{\frac{s_{22}}{s_{11}}} = r\frac{s_2}{s_1} . \tag{21}$$

[1] Greek letters are used here to avoid confusion with components a, b of the mean value; the Greek letters do not mean "theoretical parameters" in contrast to sample parameters.

Thus, the straight regression line of y on x, the second of Eqs. (18), has the equation

$$(L_2) \qquad (y - b) = r \frac{s_2}{s_1} (x - a). \tag{18'}$$

In the same way the assumption $\bar{x}(y) = \beta_1 y + \gamma_1$ for (L_1) leads to the result

$$s_{12} = \beta_1 s_{22}, \qquad \text{or} \qquad \beta_1 = \frac{s_{12}}{s_{22}} = r \frac{s_1}{s_2}. \tag{21'}$$

Thus (L_1) has the equation

$$(L_1) \qquad (x - a) = r \frac{s_1}{s_2} (y - b). \tag{18''}$$

The two factors β_1 and β_2 in (18') and (18'') are called *regression coefficients*. Hence, from (21) and (21'), the correlation coefficient is the geometrical mean of the two regression coefficients.

If the variables are standardized, i.e., $a = b = 0$, $s_{11} = s_{22} = 1$, then $s_{12} = r$ and the equations of the regression lines are

$$(L_1) \quad x = ry, \qquad (L_2) \quad y = rx.$$

The following is an immediate consequence of our definitions of linear regression. Let us compute the variance of $y - \bar{y}$. Using integrals and assuming, without loss of generality, $a = b = 0$ we have

$$\operatorname{Var}(y - \bar{y}) = \operatorname{Var}\left(y - \frac{s_{12}}{s_{11}} x\right) = \int\int \left(y - \frac{s_{12}}{s_{11}} x\right)^2 dP(x, y)$$

$$= s_{22} - 2 \frac{s_{12}}{s_{11}} s_{12} + \frac{s_{12}^2}{s_{11}^2} s_{11} = s_{22} - \frac{s_{12}^2}{s_{11}} = s_{22}(1 - r^2).$$

Thus

$$\begin{aligned} \operatorname{Var}(y - \bar{y}) &= (1 - r^2) \operatorname{Var}(y) \\ \operatorname{Var}(x - \bar{x}) &= (1 - r^2) \operatorname{Var}(x). \end{aligned} \tag{22}$$

Thus r^2 equals that proportion of the variance of either variable which is removable by linear regression on the other.

The angle θ between L_1 and L_2 reflects in a certain way the degree of correlation between x and y: In the case of stochastic independence, $s_{12} = 0$, both slopes (the first with respect to y, the second with respect

to x) are zero, thus $\theta = 90°$. In the case of complete dependence, we already know that the regression lines must coincide, thus, the angle is zero. In the general case, we have (see Fig. 50) $\theta = 90 - \theta_1 - \theta_2$ and since $\tan \theta_2 = \beta_2 = s_{12}/s_{11}$, and $\tan (90 - \theta_1) = 1/\beta_1 = s_{22}/s_{12}$, it follows that

$$\begin{aligned} \tan \theta &= \frac{s_{22}/s_{12} - s_{12}/s_{11}}{1 + s_{12}s_{22}/s_{12}s_{11}} = \frac{s_{11}s_{22} - s_{12}^2}{s_{12}(s_{11} + s_{22})} \\ &= \frac{1 - r^2}{s_{12}(1/s_{11} + 1/s_{22})} = \frac{1 - r^2}{\beta_1 + \beta_2}. \end{aligned} \tag{23}$$

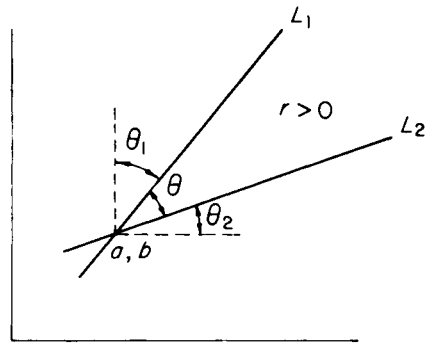

FIG. 50. Linear regression lines.

In the case of positive correlation ($s_{12} > 0$), both L_1 and L_2 lie in the first and third quadrant and θ in the range 0° to 90°. In the negative case ($s_{12} < 0$), both lines are in the second and fourth quadrant and θ is between 90° and 180°.

The above considerations hold provided that s_{11} and s_{22} are finite and not zero.

For distributions $p(x, y)$ for which the regression lines are not linear, the two straight lines passing through the center with slopes β_1, β_2, that is, the lines (18′), (18″) (with $s_{11} > 0$, $s_{22} > 0$) form, in a definite sense, the *best linear approximations to the correct regression lines.* In fact, if a function $\bar{y}(x)$ is to be approximated by a linear expression $\beta x + \gamma$, one may choose β and γ such that the sum of $\lambda(\bar{y} - \beta x - \gamma)^2$ over all x-values is minimized with some positive "weights" λ. If we decide to take $\lambda = p_1(x)$ as the weight, we have to minimize

$$\sum_x p_1(x)[\bar{y} - \beta x - \gamma]^2. \tag{24}$$

The derivatives of this sum with respect to β and γ supply the two conditions

$$\sum_x xp_1(x)(\bar{y} - \beta x - \gamma) = \sum_{x,y} xyp(x, y) - \beta \sum_{x,y} x^2 p(x, y) - \gamma \sum_{x,y} xp(x, y) = 0,$$

$$\sum_x p_1(x)(\bar{y} - \beta x - \gamma) = \sum_{x,y} yp(x, y) - \beta \sum_{x,y} xp(x, y) - \gamma \sum_{x,y} p(x, y) = 0.$$

These two equations are identical with Eqs. (20) and (19), except that now β, γ stand in place of β_2 and γ_2. Therefore, (24) is a minimum if $\beta = \beta_2$ and $\gamma = \gamma_2$, and this means that L_2 is, in this sense, the "best" linear substitute for the regression line $\bar{y}(x)$. In the same way, it is seen that L_1 approximates $\bar{x}(y)$.

We note that minimizing expression (24) is the same as minimizing

$$\sum_x \sum_y (y - \bar{y})^2 p(x, y) = \sum_x \sum_y (y - \beta x - \gamma)^2 p(x, y)$$

$$= s_{22} + b^2 + \beta^2(s_{11} + a^2) - 2\beta(s_{12} + ab) - 2b\gamma + 2a\gamma\beta + \gamma^2 \quad (25)$$

with respect to β, γ. The above double sum may be considered as the weighted mean of the square of the vertical distance $(y - \bar{y})^2 = (y - \beta x - \gamma)^2$ of a mass particle p with coordinates x, y from the straight line $\bar{y} = \beta x + \gamma$. This mean becomes a minimum for the regression line L_2; analogous results hold for L_1. In another way these results are expressed in Eqs. (22).

Notwithstanding these optimum properties, the "approximation" of general regression lines by linear regression lines is not reliable. In the example of the hyperbola, p. 571 the two general regression lines coincide with each other and with the equilateral hyperbola: $x\bar{y} = 1$, $\bar{x}y = 1$. The linear regression lines are the lines L_2: $y - a = r(x - a)$ and L_1: $x - a = r(y - a)$ for which $\tan \theta = (1 - r^2)/2r$. If r becomes very small—as we saw is possible—the angle between these two "approximation lines to one and the same curve" approaches 90°! The fact remains that the correlation coefficient relates to linear dependence only; i.e., to the case where the general regression lines are straight.

We have treated the regression lines symmetrically. This may seem artificial in applications where one might want to "predict" y as a function of x.

2.3. *Comment on regression lines of higher order.* We introduce this problem in the following way: n observations x_ν, y_ν have been made; it is assumed that they lie on a true (theoretical) curve which is a parabola of order m:

$$\bar{y} = \gamma + \beta_1 x + \cdots + \beta_m x^m, \quad (26)$$

where $m < n$. We wish to find the $m + 1$ constants $\gamma, \beta_1, \ldots, \beta_m$ in such a way that the sum of the squares of the $y_\nu - \bar{y}_\nu$ is a minimum, $\nu = 1, 2, \ldots, n$. Such a problem is not necessarily linked to probability considerations. Often x is an independent variable not subject to chance, for example time, and y a variate which on one hand changes with x according to some conjectured law and, on the other hand, depends on chance. This problem has no built-in symmetry between x and y.

We wish to determine the coefficients in (26) in such a way that with

$$u_\nu = y_\nu - \bar{y}_\nu = y_\nu - \gamma - \beta_1 x_\nu - \beta_2 x_\nu^2 - \cdots - \beta_m x_\nu^m \tag{27}$$

we have

$$\sum_{\nu=1}^{n} u_\nu^2 = \text{minimum.} \tag{28}$$

In computing the minimum, the so-called Gaussian symbols are generally used:

$$[aa] = \sum_{\nu=1}^{n} a_\nu^2, \qquad [ab] = \sum_{\nu=1}^{n} a_\nu b_\nu, \qquad \text{etc.,} \qquad [a] = \sum_{\nu=1}^{n} a_\nu. \tag{29}$$

For $m = 1$ we obtain formulas equivalent to (19′) and (20), which, in the present notation, read

$$\begin{aligned} \gamma n + \beta_1[x] &= [y] \\ \gamma[x] + \beta_1[xx] &= [xy]. \end{aligned}$$

For regression of order m the equations are

$$\begin{aligned} \gamma n + \beta_1[x] + \cdots + \beta_m[x^m] &= [y] \\ \gamma[x] + \beta_1[x^2] + \cdots + \beta_m[x^{m+1}] &= [xy] \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ \gamma[x^m] + \beta_1[x^{m+1}] + \cdots + \beta_m[x^{2m}] &= [x^m y]. \end{aligned} \tag{30}$$

The matrix is symmetric and positive definite, since we are solving a minimum problem. The conditions of positive definiteness lead to the previously established inequalities for the moments (Chapter VIII, Section 3). If x is the time, the (linear or higher order) regression $\bar{y} = \gamma + \beta_1 x + \cdots + \beta_m x^m$ is often called the *trend*.

Problem 8. Find the general regression lines and their linear approximations in the cases of Problem 1.

Problem 9. Prove that the two angles θ_1, θ_2 which the linear regression lines form with their respective axes always have the same sign and that $|\theta_1 + \theta_2| \leqslant 90°$.

Problem 10. Set up and solve the problem of Eqs. (30) for $m = 2$.

3. Other Measures of Correlation

3.1. *Contingency coefficient and similar measures.* A great many measures of correlation have been defined in the literature. None, however,

has been explored as thoroughly as r. We will first study the *contingency coefficient*.

Let us assume that $p(x, y)$ is a discrete distribution and that x as well as y can take k different values. Then the definition of Pearson's contingency measure is

$$f^2 = \frac{1}{k-1} \sum_{x,y} \frac{[p(x, y) - p_1(x)p_2(y)]^2}{p_1(x)p_2(y)}. \tag{31}$$

Since, in the case of stochastically independent variables, each term in the numerator vanishes, it is clear that $f^2 = 0$ in this case. In general, developing the square we find

$$f^2 = \frac{1}{k-1} \left[\sum_{x,y} \frac{p^2(x, y)}{p_1(x)p_2(y)} - 2 \sum_{x,y} p(x, y) + \sum_{x,y} p_1(x)p_2(y) \right]$$

$$= \frac{1}{k-1} \left[\sum_{x,y} \frac{p^2(x, y)}{p_1(x)p_2(y)} - 1 \right]. \tag{32}$$

In the case of complete dependence, $p(x, y) = p_1(x) = p_2(y)$, each term in the last sum is 1, the number of terms is k; thus, $f^2 = (k-1)/(k-1) = 1$ To complete the proof that f^2 fulfills the requirements (a) to (c) of p. 568. we have to show that f^2 always lies in the interval zero to one and takes the value zero and one *only* in the extreme cases.

Since $p_1(x) \geqslant p(x, y)$, we have, suppressing the factor $p(x, y)/p_1(x)$ to the right in (32),

$$f^2 \leqslant \frac{1}{k-1} \left[\sum_{x,y} \frac{p(x,y)}{p_2(y)} - 1 \right] = \frac{1}{k-1} \left[\sum_{y} \frac{p_2(y)}{p_2(y)} - 1 \right] = \frac{1}{k-1} (k-1) = 1. \tag{33}$$

On the other hand, the equality sign in (33) holds only if for *all* pairs x, y the omitted quotient $p(x, y)/p_1(x)$ equals 1, that is, if $p(x, y) = p_1(x)$, which is the case only when x and y are completely dependent. Note that we also have $f^2 = 1$ in the case of an arithmetic distribution like that of Fig. 51a, while $f^2 < 1$ in Fig. 51b, for example. Finally—since a sum of squares vanishes only if each square is zero—it follows from (31) that $f^2 = 0$ only if $p(x, y) = p_1(x)p_2(y)$ for *all* x, y. The result can be stated: *Pearson's contingency coefficient fulfills the three requirements* (a) *to* (*c*) *exactly*. Note also that f^2 does not depend on the numbering of the coordinates. If we write $x' = x'(x)$ and $y' = y'(y)$, calling, for example, the coordinates 1, 4, ..., k^2 instead of 1, 2, ..., k the same f^2 is obtained. This is, of course, in contrast to r^2.

The drawback is that definition (31) does not apply to continuous

distributions (since then k would become infinite) and fails even in the discontinuous case, if x and y take a different number of values.

As an application consider the result (12), p. 572. Each of the four sums in the denominator is one of the four values of $p_1(x)$ and $p_2(y)$. By elementary computation *it is found that* f^2 *is exactly the square of* (12).

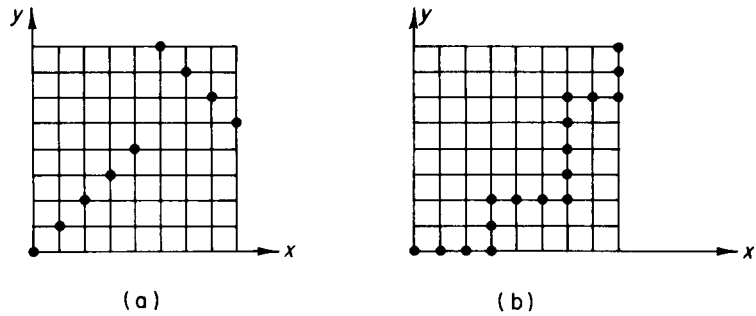

FIG. 51. Two discrete two-dimensional distributions with (a) $f^2 = 1$ and (b) $f^2 < 1$.

We mention briefly a measure which in a way stands between r and f^2 and is again due to K. Pearson. With the notation of (15) and (15′) we write

$$g_2^2 = \frac{1}{s_{22}} \sum_x p_1(x)[\bar{y}(x) - b]^2 = \frac{1}{s_{22}} \left[\sum_x [\bar{y}(x)]^2 p_1(x) - b^2 \right]. \tag{34}$$

The reader will easily verify that $g_2^2 = 1$ if there is complete dependence (then the expression in brackets reduces to s_{22}) and $g_2^2 = 0$ in the case of independence. We prove that $0 \leqslant g_2^2 \leqslant 1$. We may set $b = 0$; then we have only to show that for any x

$$[\bar{y}(x)]^2 p_1(x) \leqslant \sum_y y^2 p(x, y). \tag{35}$$

Dividing by $p_1(x)$ and substituting the value of $\bar{y}(x)$ from (15) we obtain from (35) with $p(x, y)/p_1(x) = p_2(y \mid x)$

$$\left[\sum_y y p_2(y \mid x) \right]^2 \leqslant \sum_y y^2 p_2(y \mid x):$$

this, however, is true by the shift of origin rule since the right-hand side is the moment of second order of $p_2(y \mid x)$ and the left-hand side is the square of its mean value. Of course g_1^2 may be defined in an analogous way and we state:

If we define g_2^2 *the correlation of* y *over* x, *and in the same way the correlation* g_1^2 *of* x *over* y:

$$g_2^2 = \frac{1}{s_{22}} \sum_x p_1(x)[\bar{y}(x) - b]^2, \qquad g_1^2 = \frac{1}{s_{11}} \sum_y p_2(y)[\bar{x}(y) - a]^2, \tag{34′}$$

then both g_1^2 *and* g_2^2 *are between zero and one and they reach these bounds in the cases of independence and complete dependence.*

3.2. *A correlation measure based on the* c.d.f. We mention finally a very general measure of correlation[1] which, like the ω^2-measure of Chapter IX, is based on the distribution function $P(x, y)$ *and is valid for any* $P(x, y)$. At any point x, y the following four "masses" are defined, the mass $P(x, y)$ "to the left and below" hence in $-\infty < \xi \leqslant x$, $-\infty < \eta \leqslant y$; we set $P(x, y) = A(x, y)$; next $D(x, y)$, the mass "to the right and above," hence, $x < \xi < \infty$, $y < \eta < \infty$, $C(x, y)$ the one "to the right and below": $x < \xi < \infty$; $-\infty < \eta \leqslant y$ and $B(x, y)$: $-\infty < \xi \leqslant x$; $y < \eta < \infty$. We have $P(x, \infty) = P_1(x) = A + B$, $P(\infty, y) = P_2(y) = A + C$ and

$$P - P_1P_2 = A - (A + B)(A + C) = A(1 - B - C - A) - BC = AD - BC$$

If we then form the expression

$$F = \frac{\iint (AD - BC)\, dx\, dy}{\iint (AD + BC)\, dx\, dy}. \tag{36}$$

it can be shown that F is always between -1 and $+1$, that $F = +1$ ($F = -1$) if and only if the whole distribution is concentrated on a non-decreasing (on a non-increasing) curve, and that $F = 0$ in the case of independence, i.e., if $P(x, y) = P_1(x)P_2(y)$. This is, however, not the only case; $F = 0$ can also happen in certain cases of symmetry. Consider, for example $p(0, 0) = \frac{1}{9}$, $p(1, 2) = p(2, 1) = \frac{4}{9}$, $p(0, 0) + p(1, 2) + p(2, 1) = 1$. In this case $r = F = 0$, but $f^2 = 1$.

We compute F for the Bernoulli correlation and for normal correlation. For the first problem we have probabilities p_{00}, p_{01}, p_{10}, p_{11} at the points 00, 01, 10, 11. We divide the whole plane into two parts: I consists of the inside of the unit square including the two sides which meet at the origin; II consists of all the remainder of the plane. In II everywhere $AD = 0$, $BC = 0$, in I everywhere $A = p_{00}$, $D = p_{11}$; hence $\iint AD\, dx\, dy = p_{00}p_{11}$ and similarly $\iint BC\, dx\, dy = p_{10}p_{01}$. Therefore, *for the Bernoulli correlation*

$$F = \frac{p_{00}p_{11} - p_{10}p_{01}}{p_{00}p_{11} + p_{01}p_{10}}. \tag{37}$$

The computation of F for the normal distribution is a straightforward

[1] Following a suggestion of v. Mises, it has been defined and studied by H. GEIRINGER, "Korrelationsmessung auf Grund der Summenfunktion." *Z. Angew. Math. Mech.* 13 (1933), p. 121.

integration problem, for the details of which we refer to the paper quoted on p. 582; the result is, with $D = a_{11}a_{22} - a_{12}^2$,

$$F = -\frac{\pi a_{12}}{2\left(\sqrt{D} + a_{12} \arctan \frac{a_{12}}{\sqrt{D}}\right)}. \tag{37'}$$

If $a_{12} = 0$, $F = 0$ (independence); if $D = 0$ (complete dependence) $F = -\pi a_{12}/2a_{12} \cdot \pi/2 = -1$. It is easily seen that $|F| \leqslant 1$ for all a_{12} and D.

We have seen that the various measures of correlation may lead to very different results for the same material. If we want to compare the degree of correlation in two different situations we must use the same measure. It is hardly possible to consider one or the other measure as "better." One should, however, know what can be expected in each case.[2]

Problem 11. Compute f^2 for problems 2 and 3.

Problem 12. Consider a problem similar to 4; but now $n + 1$ coins are tossed, $p(x, y)$ is the probability to obtain x heads in the first n tosses and y heads in the last n tosses. Prove that

$$p(x, x) = q^2 p_{n-1}(x - 1) + p^2 p_{n-1}(x), \qquad x = 0, \ldots, n$$

$$p(x, x + 1) = p(x + 1, x) = pq p_{n-1}(x), \qquad x = 0, \ldots, n - 1$$

$$r = \frac{n - 1}{n}, \qquad f^2 = \frac{1}{n^2}\left(\tfrac{1}{3}n^2 - \tfrac{1}{2}n + \tfrac{1}{6}\right).$$

Discuss these results.

Problem 13. Generalizing Problem 4 the coin is tossed $2n - m$ times ($m \leqslant n$); $p(x, y)$ is the probability of x heads in the first n and y heads in the last n throws. Show that

$$r = \frac{m}{n}, \qquad f^2 = \frac{1}{n}\sum_{z=1}^{m} \binom{m}{z}^2 \Big/ \binom{n}{z}^2.$$

Problem 14. Consider the two-dimensional symmetric Bernoulli distribution ($p_{00} = p_1$, $p_{10} = p_{01} = q$, $p_{11} = 0$)

$$p(x, y) = \binom{n}{x}\binom{n - x}{y} q^{x+y} p^{n-x-y}$$

[2] Cf. H. GEIRINGER, "Korrelationsmodelle." *Z. Angew. Math. Mech.* **14** (1934), pp. 19-35.

Show that $r = \frac{-q}{1-q}$, $f^2 = \frac{r^2}{n}\frac{1-r^{2n}}{1-r^2}$. Discuss these results.

Problem 15. Consider the "double alternative" with $p_{00} = p$, $p_{10} = p_{01} = q$, $p_{11} = s$, and the probability $p_n(x, y)$ to obtain in n experiments the sum x for the first coordinate, y for the second coordinate, $x, y = 0, 1, \ldots, n$.

(a) Show that

$$p_n(x, y) = \sum_{z=0}^{n} \binom{n}{z}\binom{n}{x-z}\binom{n}{y-z} s^z q^{x-z} q^{y-z} p^{n-x-y+z}.$$

(b) Write the scheme of 16 probabilities for $n = 3$.

(c) Show in two ways that $r = \frac{sp - q^2}{(q+p)(q+s)}$ independent of n, and $f^2 = \frac{r^2}{n}\frac{1-r^{2n}}{1-r^2}$. Discuss.

B. Distribution of the Correlation Coefficient (Sections 4 and 5)

4. Asymptotic Expectation and Variance of the Correlation Coefficient

4.1. *The problem.* In Sections 1–3 we were concerned with what one might call the descriptive statistics of two-dimensional distributions. We introduced and discussed certain measures of correlation valid for empirical as well as for probability distributions.

Now, we turn once more to the theoretical aspect of the problem where this distinction is basic. We have to do with two different kinds of distribution, one being the continuously or discretely distributed theoretical distribution, the other the empirical distribution consisting of a finite number n of real quantities forming a sample. In the latter case the distribution is necessarily a discontinuous one.

As we know from earlier discussions, in one and the same statistical problem both kinds of distributions occur. We have to consider the empirical distribution of a sample of n drawn from a population which has a certain theoretical (i.e., probability) distribution. As usual in theoretical statistics, we employ Latin or Greek letters to denote quantities referring to the appropriate distribution. Thus, if r is the sample correlation coefficient, ρ will denote the correlation coefficient of the probability distribution in the collective from which the sample is drawn. In the same way a, b, and α, β and likewise s_{11}, s_{12}, s_{22} and σ_{11}, σ_{12}, σ_{22} will be used.

Assume a discontinuous case where the label variate under consideration can take only kl discrete values x_ι , y_κ , the subscripts ranging from 1 to k and 1 to l, respectively. The sample then will be determined by the kl integers $n_{\iota\kappa}$ indicating how many individuals have the label value x_ι , y_κ . The quantities that have previously been called $p(x, y)$ are now the ratios $n_{\iota\kappa}/n$ and will be denoted shortly by $p_{\iota\kappa}$.[1] We restate the definitions

$$a = \sum_{\iota,\kappa} x_\iota p_{\iota\kappa} , \qquad b = \sum_{\iota,\kappa} y_\kappa p_{\iota\kappa} ,$$

$$s_{11} = \sum_{\iota,\kappa} (x_\iota - a)^2 p_{\iota\kappa} = \sum_{\iota,\kappa} x_\iota^2 p_{\iota\kappa} - \Big(\sum_{\iota,\kappa} x_\iota p_{\iota\kappa}\Big)^2 \tag{38}$$

$$s_{12} = \sum_{\iota,\kappa} (x_\iota - a)(y_\kappa - b) p_{\iota\kappa} = \sum_{\iota,\kappa} x_\iota y_\kappa p_{\iota\kappa} - \Big(\sum_{\iota,\kappa} x_\iota p_{\iota\kappa}\Big)\Big(\sum_{\iota,\kappa} y_\kappa p_{\iota\kappa}\Big), \qquad \text{etc.}$$

$$r = \frac{s_{12}}{|\sqrt{s_{11} s_{22}}|} . \tag{38'}$$

All summations here and in what follows, if not otherwise indicated, are extended over all kl combinations of $\iota = 1, 2, \ldots, k$, $\kappa = 1, \ldots, l$.

In the collective from which the sample is taken each point x_ι , y_κ has a certain probability which will be called $\pi_{\iota\kappa}$. Without loss of generality we can assume that the mean value (the center) has been chosen as origin of the coordinate system. Then, the statements analogous to (38) are

$$\alpha = \sum_{\iota,\kappa} x_\iota \pi_{\iota\kappa} = 0, \qquad \beta = \sum_{\iota,\kappa} y_\kappa \pi_{\iota\kappa} = 0, \qquad \sigma_{11} = \sum_{\iota,\kappa} (x_\iota - \alpha)^2 \pi_{\iota\kappa} = \sum_{\iota,\kappa} x_\iota^2 \pi_{\iota\kappa},$$

$$\sigma_{12} = \sum_{\iota,\kappa} (x_\iota - \alpha)(y_\kappa - \beta)\pi_{\iota\kappa} = \sum_{\iota,\kappa} x_\iota y_\kappa \pi_{\iota\kappa} , \qquad \sigma_{22} = \cdots \tag{39}$$

$$\rho = \frac{\sigma_{12}}{|\sqrt{\sigma_{11}\sigma_{22}}|} .$$

For later use we also introduce the moments of fourth order of the theoretical distribution,

$$\tau_4 = \sum_{\iota,\kappa} x_\iota^4 \pi_{\iota\kappa} , \qquad \tau_3 = \sum_{\iota,\kappa} x_\iota^3 y_\kappa \pi_{\iota\kappa} , \qquad \tau_2 = \sum_{\iota,\kappa} x_\iota^2 y_\kappa^2 \pi_{\iota\kappa} ,$$

$$\tau_1 = \sum_{\iota,\kappa} x_\iota y_\kappa^3 \pi_{\iota\kappa} , \qquad \tau_0 = \sum_{\iota,\kappa} y_\kappa^4 \pi_{\iota\kappa} . \tag{40}$$

[1] In Chapters VII and X we denoted relative frequencies by r_i; in this chapter, we cannot use r_{ik} since r stands for the correlation coefficient and $r_{\iota\kappa}$ is used in Section 6.2 for the partial correlation coefficients. Therefore we use p_ι , $p_{\iota\kappa}$ for relative frequencies and π_ι , $\pi_{\iota\kappa}$ for the corresponding probabilities.

Drawing a sample of n from the collective under consideration means as always that a group of n successive independent trials is considered as one element of a new collective in which any group of n points x_ι, y_κ has a certain probability expressed by the respective product of the $\pi_{\iota\kappa}$-values. With respect to this distribution any quantity that depends on the outcome of the n successive trials has a certain distribution. Thus we may ask for the *distribution of the (sample) correlation coefficient* r which is entirely determined by the kl values $\pi_{\iota\kappa}$ and the definition of r. At present, we content ourselves with a more restricted question asking only for the *mean value* (expectation) and the *variance* of the distribution of r. We shall even restrict our investigation to the case *where n is a large number*, and we wish to find the *asymptotic values* of $E[r]$ and $\mathrm{Var}[r]$ as n tends toward infinity. It will be seen that these values can be expressed in terms of the three σ's and five τ's defined in (39) and (40).

In order not to lose the thread of the argument we will start by stating and using two general formulas [see Chapter X, Section 6.4, Eqs. (107)], the derivation of which we postpone to the end of this chapter. Assume that f is any bounded function of k relative frequencies $p_1, p_2, \ldots, p_k$ in a sample of n taken from a collective with the probabilities $\pi_1, \pi_2, \ldots, \pi_k$ (of course in a discontinuous problem we may denote the number of attributes by k as well as by kl). The function $f(p_1, \ldots, p_k)$ is continuous and has derivatives up to order two which are continuous at least in the neighborhood of the point $p_1 = \pi_1, \ldots, p_k = \pi_k$. Let f_ι be the partial derivative of f with respect to p_ι at the point $p_1 = \pi_1, p_2 = \pi_2, \ldots, p_k = \pi_k$. Then, it will be shown in the last section that

$$\lim_{n\to\infty} E[f(p_1, p_2, \ldots, p_k)] = f(\pi_1, \pi_2, \ldots, \pi_k), \tag{41}$$

$$\begin{aligned}\lim_{n\to\infty} n\,\mathrm{Var}[f(p_1, p_2, \ldots, p_k)] &= \sum_{\kappa=1}^{k} f_\kappa^2 \pi_\kappa - \Big(\sum_{\kappa=1}^{k} f_\kappa \pi_\kappa\Big)^2 \\ &= \sum_{\kappa=1}^{k} (f_\kappa - \bar{f})^2 \pi_\kappa, \qquad \text{where} \qquad \bar{f} = \sum_{\kappa=1}^{k} f_\kappa \pi_\kappa .\end{aligned} \tag{42}$$

Also for two functions $f(p_1, \ldots, p_k)$ and $g(p_1, \ldots, p_k)$ the following formula holds with analogous notation

$$\lim_{n\to\infty} n\,\mathrm{Cov}[f, g] = \sum_{\kappa=1}^{k} (f_\kappa - \bar{f})(g_\kappa - \bar{g})\pi_\kappa = \sum_{\kappa=1}^{k} f_\kappa g_\kappa \pi_\kappa - \bar{f}\bar{g} . \tag{42'}$$

Before starting any computations we make the following comment (see footnote 9 Ch. X, p. 558). The mean and variance of the limit distribution of a statistic and

the limit of mean and variance of this statistic need not always be the same. The variance or mean of the limit distribution are usually called *asymptotic variance* and *asymptotic mean*. In Chapter X, Section 6.4 where we applied formulas (41) and (42) for the first time, we checked that for the m.l. estimate the results *were* the same. This in fact is also true for r. But it can happen that mean and (or) variance of a statistic, for any n and for $n \to \infty$, become infinite, while mean and variance of the asymptotic distribution of the identically normed statistic exist. A simple example (Cramér) is $1/m_2$; its mean and variance cannot be finite for a discrete population since there is a finite probability that $m_2 = 0$. Nevertheless $1/m_2$ is asymptotically normally distributed with mean value $1/\mu_2$, variance $(\mu_4 - \mu_2^2)/n\mu_2^4$ as will follow from results in Chapter XII.

Before applying (41), (42) to the correlation coefficient r we consider as an illustration a simpler computation, namely, the asymptotic expectation and variance of a central moment m_ν:

$$m_\nu = f(p_1, p_2, \ldots, p_k) = \sum_{\iota=1}^{k} (x_\iota - a)^\nu p_\iota, \qquad a = \sum_{\iota=1}^{k} x_\iota p_\iota$$

$$\frac{\partial f}{\partial p_\iota} = (x_\iota - a)^\nu - \nu x_\iota \sum_{\kappa=1}^{k} p_\kappa (x_\kappa - a)^{\nu-1} = (x_\iota - a)^\nu - \nu x_\iota m_{\nu-1}. \tag{43}$$

If we pass now from the m_ν to the μ_ν and from a to α we may assume without loss of generality that $\alpha = 0$ and obtain

$$f_\iota = (x_\iota - \alpha)^\nu - \nu x_\iota \mu_{\nu-1} = x_\iota^\nu - \nu x_\iota \mu_{\nu-1}, \qquad \sum_\iota f_\iota \pi_\iota = \mu_\nu - \nu\alpha\mu_{\nu-1} = \mu_\nu$$

$$f_\iota^2 = x_\iota^{2\nu} + \nu^2 x_\iota^2 \mu_{\nu-1}^2 - 2\nu x_\iota^{\nu+1}\mu_{\nu-1}, \qquad \sum_\iota f_\iota^2 \pi_\iota = \mu_{2\nu} + \nu^2\sigma^2\mu_{\nu-1} - 2\nu\mu_{\nu+1}\mu_{\nu-1}$$

$$\lim_{n\to\infty} n \operatorname{Var} m_\nu = \sum_\iota f_\iota^2 \pi_\iota - \Big(\sum_\iota f_\iota \pi_\iota\Big)^2 = \mu_{2\nu} + \nu^2\sigma^2\mu_{\nu-1} - 2\nu\mu_{\nu+1}\mu_{\nu-1} - \mu_\nu^2.$$

The reader should carry out the computation for $\alpha \neq 0$ and verify that four terms containing α cancel.

We turn now to the computations for the correlation coefficient.

4.2. *Expectation and variance of* r. As previously said, it makes no difference if instead of k we have kl label values and if double subscripts ι, κ are attached to the values of sample frequencies and probabilities. Therefore, by Eq. (41)

$$\lim_{n\to\infty} E[r] = \rho \qquad \text{or} \qquad E(r) \sim \rho. \tag{44}$$

The second equation (44) should be read: *the expectation of r, for large n, is approximately equal to ρ*. This result is certainly not surprising.

Now, we have to compute the partial derivatives of r with respect to each $p_{\iota\kappa}$. From Eq. (38′) it follows, for any p,

$$\frac{\partial r}{\partial p} = \frac{s_{12}}{|\sqrt{s_{11}s_{22}}|}\left[\frac{1}{s_{12}}\frac{\partial s_{12}}{\partial p} - \frac{1}{2}\frac{1}{s_{11}}\frac{\partial s_{11}}{\partial p} - \frac{1}{2}\frac{1}{s_{22}}\frac{\partial s_{22}}{\partial p}\right]. \tag{45}$$

Next, the derivative of s_{12} with respect to $p_{\iota\kappa}$, from the second definition of s_{12} in (38) is seen to be

$$\frac{\partial s_{12}}{\partial p_{\iota\kappa}} = x_\iota y_\kappa - x_\iota \sum_{\iota,\kappa} y_\kappa p_{\iota\kappa} - y_\kappa \sum_{\iota,\kappa} x_\iota p_{\iota\kappa}.$$

If, in this expression, the $p_{\iota\kappa}$ are replaced by $\pi_{\iota\kappa}$ the two sums on the right-hand side vanish according to (39), so that $\partial s_{12}/\partial p_{\iota\kappa}$ reduces to the product $x_\iota y_\kappa$. In exactly the same way the derivatives of s_{11} and s_{22} are found, and we have

$$\frac{\partial s_{12}}{\partial p_{\iota\kappa}} = x\, y_\kappa, \qquad \frac{\partial s_{11}}{\partial p_{\iota\kappa}} = x_\iota^2, \qquad \frac{\partial s_{22}}{\partial p_{\iota\kappa}} = y_\kappa^2 \qquad (\text{at } p_{\iota\kappa} = \pi_{\iota\kappa}).$$

If these expressions are inserted into (45) and, at the same time $p_{\iota\kappa}$ is everywhere replaced by $\pi_{\iota\kappa}$ (i.e., the s by the σ), we find for the derivative of r with respect to $p_{\iota\kappa}$, taken at the point $p_{\iota\kappa} = \pi_{\iota\kappa}$:

$$f_{\iota\kappa} = \frac{\sigma_{12}}{|\sqrt{\sigma_{11}\sigma_{22}}|}\left[\frac{x_\iota y_\kappa}{\sigma_{12}} - \frac{1}{2}\frac{x_\iota^2}{\sigma_{11}} - \frac{1}{2}\frac{y_\kappa^2}{\sigma_{22}}\right]. \tag{46}$$

These $f_{\iota\kappa}$ have now to be introduced on the right-hand side of (42) for the quantities denoted there by f_κ. We compute, first, the sum of all products $f_{\iota\kappa}\pi_{\iota\kappa}$. Since the sum of $x_\iota y_\kappa \pi_{\iota\kappa}$ equals σ_{12}, etc., we have

$$\sum_{\iota,\kappa} f_{\iota\kappa}\pi_{\iota\kappa} = \frac{\sigma_{12}}{|\sqrt{\sigma_{11}\sigma_{22}}|}\left[\frac{\sigma_{12}}{\sigma_{12}} - \frac{1}{2}\frac{\sigma_{11}}{\sigma_{11}} - \frac{1}{2}\frac{\sigma_{22}}{\sigma_{22}}\right] = 0. \tag{47}$$

Thus, the quadratic term in (42) vanishes and the formula for the variance of r reduces to

$$\mathrm{Var}[r] \sim \frac{1}{n}\sum_{\iota,\kappa} f_{\iota\kappa}^2 \pi_{\iota\kappa}. \tag{48}$$

In squaring the expression in the brackets of (46) we get six terms, the first of which is $x_\iota^2 y_\kappa^2/\sigma_{12}^2$. Now, according to (40), the sum of the pro-

ducts $x_\iota^2 y_\kappa^2 \pi_{\iota\kappa}$ is the fourth order moment τ_2 and in the same way each of the other five terms leads to one of the τ. The result is, from (48)

$$\mathrm{Var}[r] \sim \frac{1}{n}\frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}}\left[\frac{\tau_4}{4\sigma_{11}^2} + \frac{\tau_0}{4\sigma_{22}^2} + \frac{\tau_2}{2\sigma_{11}\sigma_{22}} + \frac{\tau_2}{\sigma_{12}^2} - \frac{\tau_3}{\sigma_{11}\sigma_{12}} - \frac{\tau_1}{\sigma_{12}\sigma_{22}}\right]. \tag{49}$$

Thus the variance of r is given asymptotically by Eq. (49). This together with (44) solves our problem.[2] The practical value of the general formula (49) is not too great since we would have to estimate all the moments.

To show how the formulas (44) and (49) can be used, we take as a first example *the case where x and y are independent variates*, that is, $\pi_{\iota\kappa}$ is the product of two factors ϕ_ι and ψ_κ. Then, the (unessential) assumptions $\alpha = \beta = 0$ require that $\Sigma_\iota x_\iota\phi_\iota = \Sigma_\kappa y_\kappa\psi_\kappa = 0$ and, on account of the independence: $\sigma_{12} = 0$, $\rho = 0$ and from (39) and (40)

$$\tau_2 = \sum_\iota x_\iota^2\phi_\iota \sum_\kappa y_\kappa^2\psi_\kappa = \sigma_{11}\sigma_{22}\,.$$

The only term in (49) in which the vanishing factor σ_{12} before the bracket cancels out is τ_2/σ_{12}^2. It gives $(1/n)$ times $\tau_2/\sigma_{11}\sigma_{22} = 1$. Thus, the right-hand side of (49) reduces to $1/n$ and we have the result $\mathrm{Var}(r) \sim 1/n$, or

$$E[r] = 0 \pm \frac{1}{\sqrt{n}} \qquad \text{for } x, y \text{ mutually independent.} \tag{50}$$

If, e.g., in a sample of $n = 100$ a sample correlation coefficient r of the order of magnitude ± 0.1 (or smaller) has been observed, this result does not contradict the assumption that the two variables are stochastically independent. By further analysis (see Section 5.2 of this chapter and Chapter XII)[3] it can also be shown that the distribution of r (even in the general case and for discrete as well as continuous parent distribution) approaches a normal distribution as n increases, with asymptotic mean value equal to ρ and asymptotic variance given by (49). This leads beyond (50) to the statement that, for large n, the 50 % limits for r if x and y are stochastically independent are $\pm 0.674/\sqrt{n}$.

As a second example consider the Bernoullian or four-point correlation ($k = 2$) for which the value of r was given in (12) and ρ can be written as

$$\rho = \frac{\delta}{\sqrt{\gamma_1\gamma_2\gamma_3\gamma_4}} \quad \text{with} \quad \delta = \begin{vmatrix} \pi_{00} & \pi_{10} \\ \pi_{01} & \pi_{11} \end{vmatrix}, \quad \begin{matrix} \gamma_1 = \pi_{00} + \pi_{10} \\ \gamma_3 = \pi_{10} + \pi_{11} \end{matrix}, \quad \begin{matrix} \gamma_2 = \pi_{01} + \pi_{11} \\ \gamma_4 = \pi_{01} + \pi_{00} \end{matrix}. \tag{51}$$

[2] It can be shown (see Section 8) that (49) is true for any n with an error of $O(n^{-3/2})$.

[3] See Chapter XII, p. 615, footnote 1, (A) and (C).

Using this notation, a rather cumbersone computation leads to the formula

$$\mathrm{Var}[r] \sim \frac{1}{n}\Big[1 + 5\rho^2 - \frac{3}{4}\rho^2\Big(\frac{1}{\gamma_1\gamma_3} + \frac{1}{\gamma_2\gamma_4}\Big) + \Big(1 + \frac{\rho^2}{2}\Big)\frac{\delta(\gamma_4 - \gamma_2)(\gamma_1 - \gamma_3)}{\gamma_1\gamma_2\gamma_3\gamma_4}\Big]. \tag{52}$$

If only the observed values are given, one might consider the p_{00}, p_{10}, p_{01}, p_{11} as approximations to the π_{00}, π_{10}, π_{01}, π_{11} and thus obtain an estimate for the right-hand side of (52).

In the argument that led to (49) the number kl plays no role. In fact, Eq. (49) remains valid for a continuous distribution, provided that the moments of fourth order τ_0, τ_1, ..., τ_4 exist.

We will close this section by computing $\mathrm{Var}[r]$ for the normal correlation r as given in (11). The variance components of the bivariate normal distribution have been given in Chapter VIII, Section 8 as $\sigma_{11} = a_{22}/D$, $\sigma_{12} = -a_{12}/D$, $\sigma_{22} = a_{11}/D$, where D is the determinant $a_{11}a_{22} - a_{12}^2$ of the quadric. By direct computation one can find without difficulty the moments of fourth order. Using $\sigma_{11} = {\sigma_1}^2$, $\sigma_{22} = {\sigma_2}^2$ we have

$$\tau_4 = \frac{3a_{22}^2}{D^2} = 3{\sigma_1}^4, \quad \tau_3 = \frac{-3a_{22}a_{12}}{D^2} = 3\rho{\sigma_1}^3\sigma_2, \quad \tau_1 = \frac{-3a_{11}a_{12}}{D^2} = 3\rho\sigma_1{\sigma_2}^3,$$

$$\tau_2 = \frac{a_{11}a_{22} + 2a_{12}^2}{D^2} = (1 + 2\rho^2){\sigma_1}^2{\sigma_2}^2, \quad \tau_0 = \frac{3a_{11}^2}{D^2} = 3{\sigma_2}^4. \tag{53}$$

By inserting these expressions into (49) and rearranging, we obtain

$$\mathrm{Var}[r] \sim \frac{1}{n}\frac{D^2}{a_{11}^2a_{22}^2} = \frac{1}{n}\frac{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)^2}{\sigma_{11}^2\sigma_{22}^2} = \frac{1}{n}(1 - \rho^2)^2. \tag{54}$$

Thus approximately for a normal correlation

$$E[r] = \rho \pm \frac{1}{\sqrt{n}}(1 - \rho^2). \tag{54'}$$

This coincides with (50) if $\rho = 0$. Note that in (54') the factor multiplying $1/\sqrt{n}$ is smaller than 1; in contrast to (49), the variance depends here only on ρ^2.

Problem 16. Is the result of Problem 6 consistent with independence?

Problem 17. Let $f(p_1, p_2, \ldots, p_k) = {p_1}^2 + {p_2}^2 + \cdots + {p_k}^2$ (not all p_κ equal); find $E[f]$ and $\mathrm{Var}[f]$.

Problem 18. Consider a function $H[g_1(p_1, p_2, \ldots, p_k), g_2(p_1, p_2, \ldots, p_k)]$, bounded for all p_κ where $g_1(p_1, p_2, \ldots, p_k)$, $g_2(p_1, p_2, \ldots, p_k)$ satisfy

the conditions for the formulas (41) and (42); denote $g_i(\pi_1, \pi_2, \ldots, \pi_k) = \gamma_i$, $i = 1, 2$; let H be continuous around (γ_1, γ_2) and have continuous first and second order derivatives with respect to g_1, g_2 at γ_1, γ_2. Denoting $(\partial H/\partial g_i)_{\pi_1, \ldots, \pi_k} = H_i$, $i = 1, 2$, prove that

$$\lim_{n\to\infty} n \operatorname{Var}(H) = \sigma_{11}(g_1)H_1^2 + 2\sigma_{12}(g_1, g_2)H_1H_2 + \sigma_{22}(g_2)H_2^2,$$

where $\sigma_{11}(g) = E[g^2] - (E[g])^2 = \Sigma g_\kappa^2 \pi_\kappa - (\Sigma g_\kappa \pi_\kappa)^2$, etc., as in (42) and (42′).

Problem 19. (a) Prove that r being a natural number and m_k a central moment

$$n \operatorname{Var}[m_k^r] \sim r^2 \mu_k^{2(r-1)}[\mu_{2k} - 2k\mu_{k-1}\mu_{k+1} + k^2\sigma^2\mu_{k-1}^2 - \mu_k^2].$$

(b) Prove that for the two-dimensional central moment m_{kl} (of order k in x, and of order l in y):

$$n \operatorname{Var}[m_{kl}] \sim \mu_{2k,2l} + k^2\mu_{20}\mu_{k-1,l}^2 + l^2\mu_{02}\mu_{k,l-1}^2 - 2k\mu_{k-1,l}\mu_{k+1,l} - 2l\mu_{k,l-1}\mu_{k,l+1} + 2kl\mu_{11}\mu_{k,l-1}\mu_{k-1,l} - \mu_{kl}^2.$$

Problem 20. Generalize the statement of Problem 18 to $H(g_1, g_2, \ldots, g_m)$. Take in particular $g_1 = a$, $g_2 = m_\nu$, $g_3 = m_\rho$.

Problem 21. Using the formulas given in Problem 18, prove that for the skewness $\alpha = m_3/s^3$ we have in the above approximation with δ the theoretical value of d:

$$E[d] = \delta;$$

$$\operatorname{Var}[d] = \frac{1}{n} \frac{4\mu_2^2\mu_6 - 12\mu_2\mu_3\mu_5 - 24\mu_2^3\mu_4 + 9\mu_3^2\mu_4 + 35\mu_2^2\mu_3^2 + 36\mu_2^5}{4\mu_2^5}.$$

5. The Distribution of r in Normal Samples

5.1. *Joint distribution of* $\bar{x}$, $\bar{y}$, s_1, s_2, r. Assume that x and y are normally distributed. Without loss of generality we may take $\alpha = \beta = 0$, $\sigma_{11} = \sigma_{22} = 1$ and [writing $p(x, y)$ for the parent distribution]

$$p(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right) \qquad (55)$$

The joint probability density of n sample values x_1y_1 , ..., x_ny_n from this bivariate population is then, with the summations going from 1 to n:

$$p(x_1 , y_1 , \ldots, x_n , y_n) =$$

$$\frac{1}{(2\pi)^n(1 - \rho^2)^{n/2}} \exp\Big(- \frac{1}{2(1 - \rho^2)} \Big[\sum x_\nu^2 - 2\rho \sum x_\nu y_\nu + \sum y_\nu^2\Big]\Big). \quad (56)$$

We transform the exponent in this expression to have

$$\sum_1^n x_\nu^2 - 2\rho \sum x_\nu y_\nu + \sum y_\nu^2 = \sum (x_\nu - \bar{x})^2 + n\bar{x}^2 - 2\rho \sum (x_\nu - \bar{x})(y_\nu - \bar{y})$$

$$- 2\rho n\bar{x}\bar{y} + \sum (y_\nu - \bar{y})^2 + n\bar{y}^2 = n[\bar{x}^2 - 2\rho\bar{x}\bar{y} + \bar{y}^2]$$

$$+ n[s_{11} - 2\rho r\sqrt{s_{11}s_{22}} + s_{22}]. \quad (57)$$

It is thus seen that the exponent in (56) is expressed in terms of the five statistics $\bar{x}$, $\bar{y}$, s_{11} , s_{22} , r and the corresponding theoretical parameters which are here 0, 0, 1, 1, ρ and in general α, β, σ_{11} , σ_{22} , ρ.

We want to find the joint distribution of these five statistics. We write $s_{11} = s_1^2$, $s_{22} = s_2^2$. We imagine two sample spaces of n dimensions each, one for x and one for y. The sample point may vary in the x-space and in the y-space but not independently. Let P be the point $(x_1 , \ldots, x_n)$ and Q the point $(y_1 , \ldots, y_n)$ in the respective sample spaces and O_1 , O_2 the points with all n coordinates equal to $\bar{x}$ and to $\bar{y}$, respectively. Then

$$r = \frac{\Sigma[(x_\nu - \bar{x})(y_\nu - \bar{y})]}{[\Sigma(x_\nu - \bar{x})^2 \cdot \Sigma(y_\nu - \bar{y})^2]^{1/2}}$$

equals $\cos\theta$ where θ is the "angle" between O_1P and O_2Q.

We will find an expression for the five-dimensional volume element of r, $\bar{x}$, $\bar{y}$, s_1 , and s_2 . We begin with the result of Chapter VIII, Section 6.4, that in y-space the point Q varies on the surface of an $(n - 1)$-dimensional hypersphere, H_{n-1} , whose area is proportional to $(s_2\sqrt{n})^{n-2}$. The corresponding statement holds for P in x-space alone. Now in ordes to account for the correlation, we assume fixed O_1 , O_2 , P(i.e., $\bar{x}$, $\bar{y}$, s_1) and r. The point Q then varies such that O_2Q generates a hyper-cone of half-angle θ and axis parallel to O_1P. With given s_2, Q will range over the surface of a hypersphere H_{n-2} , formed by the intersection of H_{n-1} and the cone; the surface area of H_{n-2} is proportional to $(s_2\sqrt{n}\sin\theta)^{n-3} = (s_2\sqrt{n}\sqrt{1 - r^2})^{n-3}$. Allowing θ to vary by $d\theta$, we obtain a zone of thickness $s_2\sqrt{n}\,d\theta = s_2\sqrt{n}/(\sqrt{1 - r^2})\,dr$ and thus Q varies in an element proportional to $(s_2\sqrt{n}\sqrt{1 - r^2})^{n-3}(s_2\sqrt{n}/\sqrt{1 - r^2})\,dr = s_2^{n-2}\,n^{(n-2)/2}(1 - r^2)^{(n-4)/2}\,dr$. The complete element in y-space is pro-

portional to this expression multiplied by $ds_2\, d\bar{y}$ (since both are orthogonal to the coordinates of H_{n-1}). Now considering also the element in x-space, proportional to $(s_1\sqrt{n})^{n-2}\, ds_1\, d\bar{x}$, we have for the five-dimensional volume element, suppressing factors in n,

$$\begin{aligned} dV_5 &= (s_1^{n-2}\, ds_1\, d\bar{x})(s_2^{n-2}\, ds_2\, d\bar{y}(1 - r^2)^{(n-4)/2}\, dr) \\ &= s_1^{n-2}s_2^{n-2}\, ds_1\, ds_2(1 - r^2)^{(n-4)/2}\, dr\, d\bar{x}\, d\bar{y}, \end{aligned} \tag{58}$$

and the joint probability element of $\bar{x}$, $\bar{y}$, s_1, s_2, r is proportional to

$$\exp\Big(-\frac{n}{2(1-\rho^2)}(\bar{x}^2 - 2\rho\bar{x}\bar{y} + \bar{y}^2) + n(s_1^2 - 2\rho r s_1 s_2 + s_2^2)\Big)\, dV_5\,. \tag{58'}$$

It is seen that this expression factors into

$$\Big(\exp\Big[-\frac{n}{2(1-\rho^2)}(\bar{x}^2 - 2\rho\bar{x}\bar{y} + \bar{y}^2)\Big]\, d\bar{x}\, d\bar{y}\Big)$$
$$\cdot\Big(\exp\Big[-\frac{n}{2(1-\rho^2)}(s_1^2 - 2\rho r s_1 s_2 + s_2^2)\Big]\, s_1^{n-2}s_2^{n-2}(1 - r^2)^{(n-4)/2}\, ds_1\, ds_2\, dr\Big)\,.$$

The joint distribution of the two means $\bar{x}$, $\bar{y}$ is thus independent of that of s_1, s_2, r. As expected, the two means are normally distributed and the correlation coefficient between $\bar{x}$ and $\bar{y}$ is again ρ as in the given population (55). For the second distribution, the missing constant can be computed by the condition that the total probability be one (see details in M. G. Kendall, "Advanced Theory of Statistics" I, 1943, p. 341). Introducing α, β, σ_1^2, σ_2^2, instead of 0, 0, 1, 1, we obtain R. A. Fisher's result:

In sampling from a bivariate normal population, the random variables $\bar{x}$, $\bar{y}$ are independent of s_1, s_2, r. For the former we have

$$p_n(\bar{x}, \bar{y}) =$$
$$\frac{1}{2\pi\sigma_1\sigma_2}\sqrt{\frac{n}{1-\rho^2}}\exp\Big(-\frac{n}{2(1-\rho^2)}\Big[\frac{(\bar{x}-\alpha)^2}{\sigma_1^2} - 2\rho\frac{(\bar{x}-\alpha)(\bar{y}-\beta)}{\sigma_1\sigma_2} + \frac{(\bar{y}-\beta)^2}{\sigma_2^2}\Big]\Big)\,. \tag{59}$$

Hence the variance matrix Σ of the parent population is replaced by Σ/n.

The joint density of s_1, s_2 and r equals

$$\begin{aligned} q_n(s_1, s_2, r) &= \frac{n^{n-1}}{\pi\sigma_1^{n-1}\sigma_2^{n-1}(1-\rho^2)^{(n-1)/2}\Gamma(n-2)} \\ \cdot\exp\Big(-\frac{n}{2(1-\rho^2)}\Big[\frac{s_1^2}{\sigma_1^2} - \frac{2\rho r s_1 s_2}{\sigma_1\sigma_2} + \frac{s_2^2}{\sigma_2^2}\Big]\Big)\, &s_1^{n-2}s_2^{n-2}(1-r^2)^{(n-4)/2} \end{aligned} \tag{60}$$

Here $s_1 > 0$, $s_2 > 0$, $r^2 < 1$. Outside these limits $q_n = 0$.

5.2. *Distribution of r.* From (60) we may now compute the (marginal distribution $f_n(r)$ of r by integrating over s_1 and s_2 from 0 to ∞. This may be done by expanding the factor $\exp\left(\frac{n\rho}{1-\rho^2}\cdot\frac{s_1s_2}{\sigma_1\sigma_2}r\right)$ into a power series; then the integration can be performed and the result is, for $-1 < r < 1$,

$$f_n(r) = \frac{1}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}(1-\rho^2)^{(n-1)/2}(1-r^2)^{(n-4)/2} \sum_{\nu=0}^{\infty}\Gamma^2\left(\frac{n+\nu-1}{2}\right)\frac{(2\rho r)^\nu}{\nu!}. \tag{61}$$

It is most remarkable that ρ is the only parameter appearing in (61).

Another way which avoids the infinite series (cf. Kendall, *loc. cit.*, p. 342) is to introduce the new variables

$$\zeta = \frac{s_1s_2}{\sigma_1\sigma_2}, \qquad z = \log\frac{s_1}{s_2}\Big/\frac{\sigma_1}{\sigma_2}, \qquad r = r. \tag{62}$$

The Jacobian of the transformation is $-2/\sigma_1\sigma_2$. In these variables the exponential function in (60) becomes

$$\exp\left[\frac{-n}{2(1-\rho^2)}(\zeta e^z - 2\rho r\zeta + \zeta e^{-z})\right] = \exp\left[\frac{-n}{1-\rho^2}\zeta(\cosh z - \rho r)\right],$$

and the joint density $q(\zeta, z, r)\,d\zeta\,dz\,dr$:

$$\frac{n^{n-1}}{\pi(1-\rho^2)^{(n-1)/2}\Gamma(n-2)}\exp\left[\frac{-n}{1-\rho^2}\zeta(\cosh z - \rho r)\right]\zeta^{n-2}(1-r^2)^{(n-4)/2}\,d\zeta\,dz\,dr$$

Then $q(\zeta, z, r)\,d\zeta\,dz\,dr$ can be integrated with respect to ζ and we obatin

$$q\,dz\,dr = \frac{(1-\rho^2)^{(n-1)/2}\Gamma(n-1)}{\pi\Gamma(n-2)}\frac{(1-r^2)^{(n-4)/2}}{(\cosh z - \rho r)^{n-1}}\,dz\,dr.$$

Putting $-\rho r = \cos\theta$, $\theta = \arccos(-\rho r) = \pi/2 + \arcsin\rho r$, we note that

$$\int_0^\infty \frac{dz}{\cos\theta + \cosh z} = \frac{\theta}{\sin\theta} = \frac{1}{\sqrt{1-r^2\rho^2}}\left(\frac{\pi}{2} + \arcsin\rho r\right).$$

With the help of this relation we perform the integration over z in $q\,dz\,dr$ and the density of r results in the form

$$f_n(r) = \frac{(1-\rho^2)^{(n-1)/2}}{\pi\Gamma(n-2)}(1-r^2)^{(n-4)/2}\frac{d^{n-2}}{d(-\cos\theta)^{n-2}}\left(\frac{\theta}{\sin\theta}\right)$$
$$= \frac{1}{\pi\Gamma(n-2)}(1-\rho^2)^{(n-1)/2}(1-r^2)^{(n-4)/2}\frac{d^{n-2}}{d(\rho r)^{n-2}}\left[\frac{\arccos(-\rho r)}{(1-\rho^2r^2)^{1/2}}\right]. \tag{63}$$

This is another form of Fisher's result (61).

In the important case $\rho = 0$ the power series in (61) reduces to its first term $\Gamma^2\left(\frac{n-1}{2}\right)$ and $f_n(r)$ becomes

$$f_n(r) = \frac{1}{\sqrt{\pi}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)}(1-r^2)^{(n-4)/2}, \qquad \rho = 0; \tag{64}$$

by substituting

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2} \tag{65}$$

this gives the density, $n > 2$:

$$\frac{1}{\sqrt{n-2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}\right)}\frac{1}{\left(1+\frac{t^2}{n-2}\right)^{(n-1)/2}}. \tag{64'}$$

This is a Student distribution for $n-2$ degrees of freedom. Hence, a table of Student's distribution can be used *to test independence in samples from a bivariate normal distribution.* If t_p denotes the value of t such that the probability of $|t| > t_p$ equals $p\,\%$, then with $p\,\%$ probability $|r| > t_p/\sqrt{t_p^2+n-2}$ [as seen from (65)].

The general distribution (63) or (61) has been widely studied. Tabulation is cumbersome since for each n we must now consider various values of ρ. Extensive tables have been published by Miss David.[1]

If $n = 2$, the density $f_n(r) = y$ becomes

$$y = \frac{\sqrt{1-\rho^2}\arccos(-\rho r)}{\pi(1-r^2)\sqrt{1-\rho^2r^2}} = \text{const}\,\frac{\theta}{(1-r^2)\sin\theta}, \tag{66}$$

[1] F. N. David, *Tables of the Correlation Coefficient.* London, 1938.

where $-\rho r = \cos\theta$ has been used. For $r = \pm 1$, this density tends to ∞: the distribution is U-shaped.

For $n = 3$

$$y = \frac{\text{const}}{\sin^2\theta}\,[1 - \theta\cot\theta]\,\frac{1}{\sqrt{1-r^2}}, \tag{66'}$$

a U-shaped distribution with $y = \infty$ for $r = \pm 1$.

For $n = 4$ we obtain

$$y = \frac{\text{const}}{\sin^3\theta}\,[\theta - 3\cot\theta + 3\theta\cot^2\theta],$$

This reduces to $y = \frac{1}{2}$ if $\rho = 0$. For other ρ-values y increases from a minimum at $r = -1$ to a (finite) maximum at $r = +1$.

For $n > 4$ the curves have one maximum; they are more skewed for larger $|\rho|$; they tend to normality for large n.

5.3. *Fisher's transformation.* The distribution of r is transformed into an approximately normal distribution by means of

$$\begin{aligned} z &= \frac{1}{2}\log\frac{1+r}{1-r}, \qquad r = \tanh z \\ \zeta &= \frac{1}{2}\log\frac{1+\rho}{1-\rho}, \qquad \rho = \tanh\zeta \end{aligned} \tag{67}$$

With $x = z - \zeta$ the distribution $f_n(r)$ of (64) becomes constant $\cdot \exp\left(-\frac{n-1}{2}x^2\right)$ multiplied by a series in powers of x and negative powers of n. Fisher has shown that even for moderate values of n, x is approximately normally distributed with first and second moments:

$$E[z - \zeta] = \frac{\rho}{2(n-1)} \tag{68}$$

$$\operatorname{Var}[z - \zeta] = \frac{1}{n-1} + \frac{4-\rho^2}{2(n-1)^2}. \tag{69}$$

Terms of the order n^{-2} have been neglected in the derivation of (68) and terms of the order n^{-3} have been neglected in (69). For small values of ρ the right-hand side of (69) is approximated by

$$\frac{1}{n-1} + \frac{2}{(n-1)^2} \sim \frac{1}{n-3}.$$

Hence approximately: z *is normally distributed with*

$$E[z] = \zeta + \frac{\rho}{2(n-1)} \tag{68'}$$

$$\mathrm{Var}[z] = \frac{1}{n-3}. \tag{69'}$$

(David, *loc. cit.*, states in the introduction to her tables that (68') and (69') are adequate for $n > 50$.) *Of course, all this has been proved only for a normal parent population.*

5.4. *Distribution of the regression coefficient.*[2] We return to Eq. (60), the joint density of s_1, s_2, r and substitute $b_2 = rs_2/s_1$; we want to find the distribution of b_2. The joint density of s_1, s_2, b_2 is proportional to

$$\exp\Big(-\frac{n}{2(1-\rho^2)}\Big[\frac{s_1^2}{\sigma_1^2} - \frac{2\rho s_1^2 b_2}{\sigma_1\sigma_2} + \frac{s_2^2}{\sigma_2^2}\Big]\Big) s_1^{n-1} s_2^{n-3}\Big(1 - \frac{s_1^2 b_2^2}{s_2^2}\Big)^{(n-4)/2}. \tag{70}$$

Integration with respect to s_2 (over all values such that $s_2^2 > s_1^2 b_2^2$) gives the joint distribution of s_1 and b_2 and a further integration over all s_1 gives the distribution of b_2, proportional to

$$\Big(1 - 2\rho\frac{\sigma_1}{\sigma_2} b_2 + \frac{\sigma_1^2}{\sigma_2^2} b_2^2\Big)^{-n/2};$$

if the constant is evaluated we obtain *for the density of* b_2:

$$g_n(b_2) = \frac{\Gamma\big(\frac{n}{2}\big)\sigma_1}{\sqrt{\pi}\,\Gamma\big(\frac{n-1}{2}\big)\sigma_2} \frac{(1-\rho^2)^{(n-1)/2}}{\Big(1 - 2b_2\rho\frac{\sigma_1}{\sigma_2} + b_2^2\frac{\sigma_1^2}{\sigma_2^2}\Big)^{n/2}}$$

$$= \frac{\Gamma\big(\frac{n}{2}\big)\sigma_2^{n-1}(1-\rho^2)^{(n-1)/2}}{\sqrt{\pi}\,\Gamma\big(\frac{n-1}{2}\big)\sigma_1^{n-1}} \cdot \Big[\frac{\sigma_2^2}{\sigma_1^2}(1-\rho^2) + \Big(b_2 - \rho\frac{\sigma_2}{\sigma_1}\Big)^2\Big]^{-n/2}. \tag{71}$$

This distribution contains all three theoretical parameters σ_1, σ_2, and ρ.

[2] On p. 575 the regression coefficient $s_{12}/s_{11} = rs_2/s_1$ was denoted by β_2; there we did not distinguish between sample distributions and theoretical distributions. Here we write $rs_2/s_1 = b_2$, and $\rho\sigma_2/\sigma_1 = \beta_2$.

It is symmetric about $\beta_2 = \rho\sigma_2/\sigma_1$; it tends to a normal distribution as $n \to \infty$. *The distribution of the new variable*

$$t = \frac{\sigma_1\sqrt{n-1}}{\sigma_2\sqrt{1-\rho^2}}(b_2 - \beta_2) \tag{72}$$

is a Student distribution with $n - 1$ *degrees of freedom.* Compared to the distribution of r, the disadvantage of (71) or (72) is that the results contain all three theoretical values σ_1, σ_2, ρ. If, however, instead of (72) the distribution of the statistic

$$t' = \frac{s_1\sqrt{n-2}}{s_2\sqrt{1-r^2}}(b_2 - \beta_2) \tag{73}$$

is considered *Bartlett*[3] *found that* t' *has a Student distribution with* $n - 2$ *degrees of freedom.* This is a more useful result than (71) and (72).

Problem 22. Give the distribution of the regression coefficient b_1.

Problem 23. Prove that the joint distribution of s_{11}, s_{12}, s_{22}, with $D = s_{11}s_{22} - s_{12}^2$, $\Delta = \sigma_{11}\sigma_{22} - \sigma_{12}^2$, is given by

$$p_n(s_{11}, s_{12}, s_{22}) = \frac{n^{n-1}}{4\pi\Gamma(n-2)}\frac{D^{(n-4)/2}}{\Delta^{(n-1)/2}}\exp\Big[-\frac{n}{2(1-\rho^2)}\Big(\frac{s_{11}}{\sigma_{11}} - \frac{2\rho^2 s_{12}}{\sigma_{12}} + \frac{s_{22}}{\sigma_{22}}\Big)\Big].$$

Problem 24. In the 10 years from 1899 to 1908, the influx of water into a reservoir in Lund, Sweden, measured in 1000 m³ and the amount of rain measured in mm was observed, and the correlation 0.705 found. Is this a result against independence on the 1 % level?

Problem 25. In 234 observations of the thickness of the stem and the length of the longest flower-petal of a certain plant, $r = 0.83$ was found. Is this compatible with independence? Use also the z-transformation (67) and discuss your result.

Problem 26. Consider Problem 6, where $n = 483$, $r = 0.16$. Is this a significant deviation from independence on the 1 % level?

Problem 27. Show that in samples from a normal bivariate population, the variance of $b_2 = r\,s_2/s_1$ is exactly equal to

$$\mathrm{Var}(b_2) = \frac{1}{n-3}\frac{\sigma_2^2}{\sigma_1^2}(1 - \rho^2).$$

[3] M. S. BARTLETT, "On the theory of statistical regression." *Proc. Roy. Soc. Edinburgh* **53** (1933), p. 260.

C. Generalizations to k Variables (Sections 6 and 7)

6. Regression and Correlation in k Variables

6.1. *Regression surfaces.* Various problems arise if a distribution in more than two variables is to be classified with respect to its degree of correlation. Let $p(x_1, x_2, ..., x_k)$ be the distribution referring either to probabilities or to a sample repartition, in the first case either continuous (a density) or discontinuous. For the sake of simplicity we assume that the origin of the k-dimensional coordinate system coincides with the mean value

$$\sum_{x_1}\sum_{x_2}\cdots\sum_{x_k} x_\kappa p(x_1, x_2, ..., x_k) = 0, \qquad \kappa = 1, 2, ..., k.$$

The first concept that can be taken over without difficulty is the definition of the regression lines of Section 2, which now become *regression surfaces.* Their definition is [compare Eqs. (15) and (15′)].

$$\begin{aligned} \bar{x}_1 &= \Big[\sum_{x_1} x_1 p(x_1, x_2, ..., x_k)\Big] \Big/ \Big[\sum_{x_1} p(x_1, x_2, ..., x_k)\Big] \\ \bar{x}_2 &= \Big[\sum_{x_2} x_2 p(x_1, x_2, ..., x_k)\Big] \Big/ \Big[\sum_{x_2} p(x_1, x_2, ..., x_k)\Big], \qquad \text{etc.} \end{aligned} \tag{74}$$

If the hypothesis is introduced that each of these functions of $k - 1$ variables is *linear* (and homogeneous) we write

$$\begin{aligned} \bar{x}_1 &= \quad b_{12}x_2 + b_{13}x_3 + \cdots + b_{1,k-1}x_{k-1} + b_{1,k}x_k \\ \bar{x}_2 &= b_{21}x_1 \quad + b_{23}x_3 + \cdots + b_{2,k-1}x_{k-1} + b_{2,k}x_k \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \bar{x}_k &= b_{k1}x_1 + b_{k2}x_2 + \quad \cdots\cdots \quad + b_{k,k-1}x_{k-1}\,. \end{aligned}$$

Each of these is the equation of a hyperplane: the *regression plane* of x_1 on $x_2, x_3, ..., x_k$, etc. The regresssion between two variables, say x_1 on x_2 is measured by the regression coefficient b_{12}. Since, however, x_1 and x_2 are considered together with $k - 2$ other variables $x_3, x_4, ..., x_k$ it is often useful to show this in the notation. In a notation introduced by Yule, each b has—in case of k variates—k subscripts, *two primary* subscripts, the first and the second: the first is that of the variable to the left of the regression equation, the second that of the variable to which the coefficient belongs; the $k - 2$ remaining subscripts, separated from the primary subscripts by a dot, are *secondary* subscripts; the arrange-

ment of the secondary subscripts does not matter, while that of the primary ones does. A *regression coefficient*, correlation coefficient, etc., is called *of order* ν if there are ν secondary subscripts. Therefore in (74′) instead of b_{12} we may write $b_{12.34\ldots k}$ and similarly $b_{k-1,k} = b_{k-1,k.123\ldots k-2}$, etc. If there is no possibility of misunderstanding, the secondary subscripts are omitted as in (74′). The $b_{\iota\kappa}$ in (74′) are all of order $(k-2)$. The β_1, β_2 in Eqs. (18) which, would in the present notation be denoted by b_{12}, b_{21} are of order zero.

6.2. *Regression coefficients. Correlation matrix.* To determine the $b_{\iota\kappa}$ of (74′) we introduce in the first of these equations the expression for $\bar{x}_1$ from (74) and multiply by the denominator. We then multiply both sides of that equation by x_i (where, successively, $i = 2, 3, \ldots, k$) and each time sum over all combinations of $x_1, x_2, \ldots, x_k$. As the variance components $s_{\iota\kappa}$ are defined by

$$s_{\iota\kappa} = \sum \cdots \sum x_\iota x_\kappa p(x_1, x_2, \ldots, x_k) \tag{75}$$

this procedure leads to $(k-1)$ equations for the $b_{12}, b_{13}, \ldots, b_{1k}$

$$\begin{aligned}
s_{12} &= b_{12}s_{22} + b_{13}s_{32} + \cdots + b_{1k}s_{k2} \\
s_{13} &= b_{12}s_{23} + b_{13}s_{33} + \cdots + b_{1k}s_{k3} \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
s_{1k} &= b_{12}s_{2k} + b_{13}s_{3k} + \cdots + b_{1k}s_{kk}\,.
\end{aligned} \tag{76}$$

If we call $D_{\iota\kappa}$ the cofactor of $s_{\iota\kappa}$ in the determinant D

$$D = \begin{vmatrix} s_{11} & s_{12} \cdots s_{1k} \\ s_{21} & s_{22} \cdots s_{2k} \\ \cdots\cdots\cdots \\ s_{k1} & s_{k2} \cdots s_{kk} \end{vmatrix}, \tag{77}$$

the solution of (76) reads

$$b_{12} = -\frac{D_{12}}{D_{11}}, \qquad b_{13} = -\frac{D_{13}}{D_{11}}, \ldots, b_{1k} = -\frac{D_{1k}}{D_{11}}. \tag{78}$$

In the same way, starting from the second, third, ... of Eqs. (74) we arrive at the general expression for the partial regression coefficients.

$$b_{\iota\kappa} = -\frac{D_{\iota\kappa}}{D_{\iota\iota}}, \qquad \iota \neq \kappa = 1, 2, \ldots, k, \tag{78′}$$

and $b_{\iota\kappa} = b_{\iota\kappa.12\ldots k}$, where the $k-2$ secondary subscripts do not contain ι or κ. In the case $k = 2$, this reduces to $b_{12} = s_{12}/s_{22}$, and $b_{21} = s_{12}/s_{11}$ in accordance with the results (21) and (21′) in Section 2. Just as in the case $k = 2$, it can be shown that if the actual regression functions as defined in (74) are not linear, *Eqs.* (74′) *with the values of* $b_{\iota\kappa}$ *given in* (78′) *are the "best" linear approximations in the sense described* in Section 2.

The *matrix of the second order moments* $s_{\iota\kappa}$ whose determinant is D may be called S; then $|S| = D$. If we put

$$s_{\iota\iota} = s_\iota^2, \qquad s_{\iota\kappa} = s_\iota s_\kappa r_{\iota\kappa},$$

and accordingly $r_{\iota\iota} = 1$, we obtain the *correlation matrix*

$$R = \begin{pmatrix} r_{11} & r_{12} \cdots r_{1k} \\ r_{21} & r_{22} \cdots r_{2k} \\ \cdots\cdots\cdots \\ r_{k1} & r_{k2} \cdots r_{kk} \end{pmatrix}. \tag{79}$$

For simplicity we shall use R for the determinant of the matrix (79) too. The symmetric matrices S and R are non-negative, i.e., the corresponding quadratic forms are $\geqq 0$. If the equality sign holds, then the rank of S is less than k and we call the distribution to which S belongs *singular*. A singular k-dimensional distribution reduces, in appropriate coordinates, to an h-dimensional distribution, where $h < k$. For a singular distribution some of the $D_{\iota\iota}$ may vanish and some regression coefficients become infinite or indeterminate. The reader may think it over in terms of a Gauss distribution with $k = 2$ or $k = 3$. If the distribution $p(x_1, x_2, \ldots, x_k)$ is not singular, the matrices S and R are positive definite. *Then, in* (78′), $D_{\iota\iota} > 0$ *and the solution* (78) *is unique.* If the variables are completely uncorrelated, all $D_{\iota\kappa} = 0$, all $b_{\iota\kappa} = 0$.

6.3. *Partial and multiple correlation coefficients.* In further analogy to the two-dimensional case we can now introduce the *partial correlation coefficient*

$$r_{\iota\kappa\cdot} = \frac{-D_{\iota\kappa}}{|\sqrt{D_{\iota\iota}D_{\kappa\kappa}}|}. \tag{80}$$

Here the point in $r_{\iota\kappa\cdot}$ means that $k-2$ secondary indices might be written. We have, for example, for $n = 3$

$$r_{12\cdot 3} = \frac{r_{12} - r_{13}r_{23}}{[(1 - r_{13}^2)(1 - r_{23}^2)]^{1/2}}. \tag{80′}$$

In contrast to the "total" correlation coefficients $r_{\iota\kappa}$ the $r_{\iota\kappa\cdot}$ are "partial" correlation coefficients.

It can easily be seen that the quantities $r_{\iota\kappa\cdot}$ fulfill the following conditions: (1) If x_ι and x_κ are stochastically independent in the sense that $p(x_1, x_2, \ldots, x_k)$ is a product of two factors one of which does not depend on x_ι while the other does not depend on x_κ (and if the regression functions are linear), then $r_{\iota\kappa\cdot} = 0$. (2) In the case of complete linear dependence, that is, if any value of x_ι is linked to a finite value of x_κ (and if the regression functions are linear) then $|r_{\iota\kappa\cdot}| = 1$. In this case the matrix of the $s_{\iota\kappa}$ is singular; the distribution is no longer k-dimensional. (3) Whatever the distribution $p(x_1, x_2, \ldots, x_k)$ may be, the absolute value of $r_{\iota\kappa\cdot}$ cannot exceed 1. The last point follows from the fact that since the matrix of the $s_{\iota\kappa}$ is non-negative definite, the same is true of the matrix of the cofactors $D_{\iota\kappa}$ and, therefore, the inequality $D_{\iota\iota}D_{\kappa\kappa} \geqslant D_{\iota\kappa}^2$ holds for any pair of subscripts ι, κ.

We arrive at the definition (80) in the following way also. If the variables whose correlation is measured—say, x_1 and x_2—are considered in conjunction with the $(k-2)$ other variables $x_3, x_4, \ldots, x_k$ we may consider the correlation between x_1 and x_2 to be partly due to the influence of the other variables. To make this more explicit, the so-called *residuals* are introduced.

We consider first the case $k = 3$. It may be that the correlation between x and y is due (partly, at least) to the fact that both x and y are correlated to a third variable z. We try to eliminate the correlation with z by replacing x by $x' = x - \lambda z$ and y by $y' = y - \mu z$ such that x' as well as y' are no longer correlated to z, that is: $s_{x'z} = 0$, $s_{y'z} = 0$.[1] By the definitions of the $s_{\iota\kappa}$ and of x', y' these are equivalent to

$$s_{xz} - \lambda s_{zz} = 0 \qquad \text{and} \qquad s_{yz} - \mu s_{zz} = 0$$

or

$$\lambda = \frac{s_{xz}}{s_{zz}} = r_{xz}\frac{s_x}{s_z}, \qquad \mu = \frac{s_{yz}}{s_{zz}} = r_{yz}\frac{s_y}{s_z}. \tag{81}$$

We see that λ, μ are regression coefficients of order zero from x on z and from y on z, respectively. The correlation coefficient between x' and y' is as far as linear correlation is concerned—now free of the influence of z. We compute the correlation between x' and y':

$$r_{x'y'} = \frac{s_{x'y'}}{s_{x'}s_{y'}} = \frac{s_{x-\lambda z,\, y-\mu z}}{s_{x-\lambda z}s_{y-\mu z}}. \tag{82}$$

[1] In German the suggestive term "*bereinigter*" *Korrelationskoeffizient* is used.

If we substitute for λ and μ their values from (81) we obtain

$$s_{x-\lambda z,y-\mu z} = s_{xy} - \lambda s_{yz} - \mu s_{xz} + \lambda\mu s_{zz} = s_x s_y(r_{xy} - r_{xz}r_{yz}) \tag{83}$$

$$\begin{aligned} s_{x-\lambda z} &= s_{xx} - 2\lambda s_{xz} + \lambda^2 s_{zz} = s_x{}^2(1 - r_{xz}^2) \\ s_{y-\mu z}^2 &= s_{yy} - 2\mu s_{yz} + \mu^2 s_{zz} = s_y{}^2(1 - r_{yz}^2)\,. \end{aligned} \tag{84}$$

Substituting into (82) we find for $r_{x'y'} = r_{xy.z}$

$$r_{xy\cdot z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{1 - r_{xz}^2}\sqrt{1 - r_{yz}^2}} = -\frac{D_{12}}{\sqrt{D_{11}D_{22}}} = -\frac{R_{11}}{\sqrt{R_{11}R_{22}}}, \tag{85}$$

as in (80) and (80′), with the dependence on z now put in evidence. We may call $r_{xy\cdot z}$ the *partial correlation coefficient of x and y with respect to z.* The expressions (84) are called residual variances.

Geometrically, we may visualize x and y as two points in n-space corresponding to the results of n observations x_1, x_2, ..., x_n and y_1, y_2, ..., y_n and z a third point with z_1, z_2, ..., z_n. We take $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ as vectors: r_{xy} is the cosine of the angle between $\mathbf{x}$ and $\mathbf{y}$. Next $\mathbf{x}$ is resolved into a component parallel to $\mathbf{z}$ and one perpendicular to $\mathbf{z}$ in the $(n - 1)$-space normal to $\mathbf{z}$; this component is $\mathbf{x}'$. The same is then done for $\mathbf{y}$ and $\mathbf{y}'$ is obtained. The correlation coefficient $r_{xy\cdot z}$ is the cosine of the angle between $\mathbf{x}'$ and $\mathbf{y}'$ in the linear $(n - 1)$-dimensional space normal to $\mathbf{z}$. This remark will be used in the study of the distribution of partial correlation coefficients (Section 7).

We wrote the above formulas for empirical variances and correlations. The analogous definitions for theoretical parameters can be derived in the same way by computing expectations. We wish to determine λ and μ so that

$$\sigma_{x'z} = E[x'z] = E[xz - \lambda z^2] = 0, \qquad E[y'z] = E[yz - \mu z^2] = 0.$$

Hence

$$\lambda = \frac{E[xz]}{E[z^2]} = \frac{\sigma_{xz}}{\sigma_z{}^2}, \qquad \mu = \frac{\sigma_{yz}}{\sigma_z{}^2}.$$

$$\rho_{x'y'} = \rho_{xy\cdot z} = \frac{E[(x - \lambda z)(y - \mu z)]}{\sigma_{x'}\sigma_{y'}} = \frac{(\rho_{xy} - \rho_{xz}\rho_{yz})\sigma_x\sigma_y}{\sigma_{x'}\sigma_{y'}},$$

$$\sigma_{x'}^2 = E[(x - \lambda z)^2] = (1 - \rho_{xz}^2)\sigma_x{}^2, \qquad \sigma_{y'}^2 = (1 - \rho_{yz}^2)\sigma_y{}^2,$$

$$\rho_{xy\cdot z} = \frac{\rho_{xy} - \rho_{xz}\rho_{yz}}{\sqrt{1 - \rho_{xz}^2}\sqrt{1 - \rho_{yz}^2}} = \frac{-\Delta_{12}}{\sqrt{\Delta_{11}\Delta_{22}}}. \tag{85′}$$

The $\Delta_{\iota\kappa}$ are the theoretical equivalents of the $D_{\iota\kappa}$.

From our derivation of (85) it follows that $r_{xy\cdot z}$ is not changed if x and y are replaced by $x'' = x - az$ and $y'' = y - bz$ where a and b are quite arbitrary; in fact *these* linear expressions $x' = x - \lambda z$, $y' = y - \mu z$ which have no correlation with z will always be the same no matter whether we start with x, y, or with x'', y''. Hence, in discussing properties of $r_{xy\cdot z}$, we can assume from the start that the x, y have no correlation with z; hence we can assume, without loss of generality, that $\rho_{xz} = \rho_{yz} = 0$.

We now define residuals or residual variates (variances and correlations) for general k:

$$x_{1\cdot 23\ldots k} = x_1 - \beta_{12}x_2 - \cdots - \beta_{1k}x_k \tag{86}$$

is the *residual* (first residual variate) *of order* $k - 1$. [Note that for $k = 3$, this does not correspond to $x' = x_{1\cdot 3} = x - \lambda z$ since in (86) we also substract a term in x_2.] The $\beta_{12}, \ldots, \beta_{1k}$ correspond to the regression coefficients defined in (78) and which more explicitly read $\beta_{12\cdot 34\ldots k}$, $\beta_{13\cdot 24\ldots k}$, etc. *The residual* $x_{1\cdot 23\ldots k}$ *is that part of* x_1 *which remains after subtraction of the best linear estimate of* x_1. *Its expectation is zero*

$$E[x_{1\cdot 23\ldots k}] = 0,$$

since the mean value vanishes. *The residual is uncorrelated to any of the subtracted variables.* Take, for example $E[x_{1\cdot 23\ldots k}x_2]$.

$$\begin{aligned} &E[(x_1 - \beta_{12}x_2 - \beta_{13}x_3 - \cdots - \beta_{1k}x_k)x_2] \\ &= E[x_1x_2] - \beta_{12}E[x_2^2] - \beta_{13}E[x_2x_3] - \cdots - \beta_{1k}E[x_2x_k] \\ &= \sigma_{12} - \beta_{12}\sigma_2^2 - \beta_{13}\sigma_{23} - \cdots - \beta_{1k}\sigma_{2k} = 0 \end{aligned}$$

on account of the first Eq. (76) (which holds, of course, for the theoretical values as well).

Consider $E[x_{1\cdot 23\ldots k}x_1]$. Using the formulas analogous to (78) we find

$$E[x_{1\cdot 23\ldots k}x_1] = \sigma_{11}\frac{\Delta_{11}}{\Delta_{11}} + \sigma_{12}\frac{\Delta_{12}}{\Delta_{11}} + \cdots + \sigma_{1k}\frac{\Delta_{1k}}{\Delta_{11}} = \frac{\Delta}{\Delta_{11}}.$$

Thus: the residual variance

$$\sigma^2_{1\cdot 23\ldots k} = E[x^2_{1\cdot 23\ldots k}] = \frac{\Delta}{\Delta_{11}} = \sigma_1^2\frac{P}{P_{11}}, \tag{87}$$

where P and $P_{\iota\kappa}$ are the theoretical counterparts of R and $R_{\iota\kappa}$. Formula (87) holds similarly for any other primary subscript.

Next we wish to generalize the approach (p. 602) which led to (85). We put

$$\begin{aligned} x_1' &= x_{1\cdot 34\ldots k} = x_1 - \beta_{13\cdot 45\ldots k}x_3 - \beta_{14\cdot 35\ldots k}x_4 - \cdots - \beta_{1k\cdot 34\ldots k-1}x_k \\ x_2' &= x_{2\cdot 34\ldots k} = x_2 - \beta_{23\cdot 45\ldots k}x_3 - \beta_{24\cdot 35\ldots k}x_4 - \cdots - \beta_{2k\cdot 34\ldots k-1}x_k \,. \end{aligned} \tag{88}$$

Here the β are regression coefficients of order $k-3$; the λ and μ in the case $k=3$ were of order zero. There is no correlation between x_1' and any of the subtracted variables and similarly for x_2'. In fact, denoting any of the subscripts 3, 4, ..., k by i and the present coefficients briefly by their principal subscripts [although they are different from those in (86)] the $k-2$ equations

$$\sigma_{1i} - \beta_{13}\sigma_{3i} - \beta_{14}\sigma_{4i} - \cdots - \beta_{1k}\sigma_{ki} = 0, \qquad i = 3, 4, \ldots, k \tag{89}$$

have one and only one solution on account of the positive definiteness of the matrix of variances. The β are quotients of minors of order $k-2$

$$\beta_{1i} = -\frac{\Delta_{22\cdot 1i}}{\Delta_{22\cdot 11}}, \qquad i = 3, 4, \ldots, k$$

where $k-3$ secondary subscripts in β are omitted and $\Delta_{22\cdot 1i}$ is the cofactor of σ_{1i} in Δ_{22}. The same holds for the second set of $k-2$ equations.

We wish to find the correlation coefficient of x_1', x_2',[2] a partial correlation coefficient:

$$\rho_{12\cdot 34\ldots k} = \frac{E[x_{1\cdot 34\ldots k}x_{2\cdot 34\ldots k}]}{\sqrt{E[x^2_{1\cdot 34\ldots k}]\cdot E[x^2_{2\cdot 34\ldots k}]}}\,. \tag{90}$$

We see immediately

$$E[x_{1\cdot 34\ldots k}x_{2\cdot 34\ldots k}] = E[x_{1\cdot 34\ldots k}x_2] = E[x_{2\cdot 34\ldots k}x_1], \tag{91}$$

and after a brief computation, with $\Delta_{jj,\iota\kappa}$ the cofactor of $\sigma_{\iota\kappa}$ in Δ_{jj}:

$$\begin{aligned} E[x_1x_{2\cdot 34\ldots k}] &= -\frac{\Delta_{12}}{\Delta_{11\cdot 22}} \\ E[x^2_{1\cdot 34\ldots k}] = E[x_1x_{1\cdot 34\ldots k}] &= \frac{\Delta_{22}}{\Delta_{11\cdot 22}} \\ E[x^2_{2\cdot 34\ldots k}] = E[x_2x_{2\cdot 34\ldots k}] &= \frac{\Delta_{11}}{\Delta_{11\cdot 22}} \end{aligned}$$

and, hence

$$\rho_{12\cdot 34\ldots k} = -\frac{\Delta_{12}}{\sqrt{\Delta_{11}\Delta_{22}}} = -\frac{P_{12}}{\sqrt{P_{11}P_{22}}}\,, \tag{90'}$$

as in (80). A similar formula holds for the correlation of any two residual variates. We see in this way that any residual correlation or residual variance can be expressed in terms of the $\sigma_{\iota\kappa}, \rho_{\iota\kappa}$ (or $s_{\iota\kappa}, r_{\iota\kappa}$) of order zero.

---

[2] If in Chapter VIII (78), we put $X^{(1)} = \binom{x_1}{x_2}$, define $S_{\iota\kappa}$, $\iota, \kappa = 1, 2$ as in (78′) of Chapter VIII, and introduce as in Sect. 9, Chapter VIII the matrix

$$S_{11} - S_{12}S_{22}^{-1}S_{21}$$

we may denote the components of this matrix by $\sigma_{11\cdot 3\ldots k}$, $\sigma_{12\cdot 3\ldots k}$, etc.; then the right-hand side of (90) (this chapter) takes the form

$$\frac{\sigma_{12\cdot 3\ldots k}}{\sqrt{\sigma_{11\cdot 3\ldots k}}\sqrt{\sigma_{22\cdot 3\ldots k}}}$$

We do not further follow up the relation to the matrix notation of Chapter VIII.

Often it is convenient to use recurrence relations as, for example,

$$\rho_{12\cdot 34\ldots k} = \frac{\rho_{12\cdot 34\ldots k-1} - \rho_{1k\cdot 34\ldots k-1}\rho_{2k\cdot 34\ldots k-1}}{\sqrt{(1-\rho^2_{1k\cdot 34\ldots k-1})(1-\rho^2_{2k\cdot 34\ldots k-1})}}, \tag{92}$$

which expresses a correlation coefficient in terms of coefficients of lower order. If we divide the secondary subscripts, here 3, 4, ..., k into two parts p and q [in (92), q stands for k; p for 3, 4, ..., $k-1$] Eq. (92) takes the form

$$\rho_{12\cdot pq} = \frac{\rho_{12\cdot p} - \rho_{1q\cdot p}\rho_{2q\cdot p}}{\sqrt{(1-\rho^2_{1q\cdot p})(1-\rho^2_{2q\cdot p})}}. \tag{92'}$$

There are many partial regressions, correlations and variances; they are all expressible in terms of those of zero order.

We define finally the *multiple correlation coefficient*. We write Eq. (86) as

$$x_1' = x_{1\cdot 23\ldots k} = x_1 - \bar{x}_1 \tag{86'}$$

where $\bar{x}_1 = \beta_{12}x_2 + \beta_{13}x_3 + \cdots + \beta_{1k}x_k$ with $(\beta_{12} = \beta_{12\cdot 34\ldots k})$ is the best linear estimate of x_1. *The correlation coefficient between x_1 and $\bar{x}_1$ is called the multiple correlation coefficient.*

$$\rho_{1(23\ldots k)} = \frac{E[x_1\bar{x}_1]}{\sqrt{E[x_1^2]E[\bar{x}_1^2]}}. \tag{93}$$

Then

$$E[x_1\bar{x}_1] = E[x_1(x_1 - x_1')] = \sigma_{11} - \frac{\Delta}{\Delta_{11}}$$

$$E[\bar{x}_1^2] = E[x_1^2 - 2x_1x_1' + x_1'^2] = \sigma_{11} - 2\frac{\Delta}{\Delta_{11}} + \frac{\Delta}{\Delta_{11}}$$

$$E[x_1^2] = \sigma_{11}$$

$$\rho_{1(23\cdots k)} = \sqrt{1 - \frac{\Delta}{\sigma_{11}\Delta_{11}}} = \sqrt{1 - \frac{P}{P_{11}}}. \tag{93'}$$

For a non-negative quadratic form the inequality $\Delta \leqslant \sigma_{11}\Delta_{11}$ holds. Therefore, $E[x_1\bar{x}_1] \geqslant 0$ and

$$0 \leqslant \rho^2_{1(23\cdots k)} \leqslant 1. \tag{93''}$$

It can be shown without difficulty that the multiple correlation coefficient

has properties similar to the ordinary ρ (to which it reduces for $k = 2$). Note that $\rho_{1(23\cdots k)} = 0$ if and only if $\rho_{12} = \rho_{13} = \cdots = \rho_{1k} = 0$.

Problem 28. Study the triple correlation between x, y, z in the following example. In a group of 10 successive trials of an alternative with the event probability q, we call x the number of events in the first six trials, z the number of events in the last six, and y the number of events that occur in trials numbered 3 to 8.

Problem 29. Prove the last statement of this section regarding $\rho_{1(23\cdots k)} = 0$.

7. Remarks on the Distribution of Correlation Measures from a k-Dimensional Normal Population

7.1. *Partial correlation coefficients.* At the end of Chapter VIII we considered the multivariate normal distribution and its variances and covariances. Introducing the (theoretical) correlation matrix P (we agreed to use P also for its determinant) we have for the normal density

$$p(x_1, x_2, \ldots, x_k) = \frac{1}{(2\pi)^{k/2}\sqrt{P}\,\sigma_1\sigma_2\cdots\sigma_k} \exp\Big(-\frac{1}{2P}\sum_{\iota,\kappa=1}^{k} P_{\iota\kappa}\frac{x_\iota x_\kappa}{\sigma_\iota\sigma_\kappa}\Big). \quad (94)$$

We sketch an essentially geometrical derivation of the distribution of partial correlation coefficients. Since every x_ι is normally distributed every linear function of the x_ι is also, as found in Chapter VIII, Section 9. In particular, any residual (88) is normally distributed. For $k = 3$, the residuals $x' = x_{1\cdot 3}$ and $y' = x_{2\cdot 3}$ are bivariately normally distributed with correlation $\rho_{12\cdot 3}$; therefore $\rho_{12\cdot 3}$ is distributed as the correlation coefficient between two normal variates.

Remember now the geometrical interpretation, p. 603. The origin O is at the mean. We have in n-space three vectors: $\mathbf{OP} = \mathbf{x}$, $\mathbf{OQ} = \mathbf{y}$, $\mathbf{OR} = \mathbf{z}$, each with n components. The ordinary correlation coefficient $r_{12} = r$ is the cosine of the angle POQ, while $r_{12\cdot 3}$ is the cosine of the angle between the vectors $\mathbf{x}'$ and $\mathbf{y}'$ where we have resolved $\mathbf{OP}$ into a component parallel to $\mathbf{z}$ and into one, $\mathbf{x}'$ in the $(n-1)$-space perpendicular to $\mathbf{z}$. Therefore the distribution of $r_{12\cdot 3}$, the cosine of the angle in the space normal to $\mathbf{OR}$, must be of the same form as that of r_{12} except that $n - 1$ now takes the place of n. Hence for samples from a normal population, the distribution of the partial correlation coefficient of first order from n sets of k observations is the same as that of a correlation of zero order from $n - 1$ sets of k observations. Likewise, *the distri-*

bution of the partial correlation coefficient $r_{12\ldots s\cdot s+1\ldots k}$ *of order* $k - s$ *from n sets of observations is the same as that of a correlation of order zero from* $n - (k - s)$ *sets of observations.* Note that in these considerations no restriction regarding the correlation in the multidimensional normal parent population was made.

An elegant proof regarding the distribution of $r_{12\cdot 3}$ is in van der Waerden [28], p. 306. He assumes that $\rho_{12} = 0$ in the normal parent distribution. From the remark on p. 604 we know that we may assume $\rho_{13} = \rho_{23} = 0$; hence x, y z are independent and normally distributed. The final result, which agrees, of course with the above geometric conclusion is: *If we sample* x_ν , y_ν , z_ν , $\nu = 1, 2, \ldots, n$ *from an uncorrelated normal parent distribution* $p(x, y, z)$, *the distribution of* $r_{xy\cdot z}$ *is the same as that of* r *when* x_i , y_i , $i = 1, 2, \ldots, n - 1$ *are sampled from an uncorrelated normal bivariate parent distribution,* $p(x, y)$.

7.2. *Wishart distribution.* We finally turn to the generalization of the first results of Section 5, p. 593 ff where we considered a bivariate normal parent population from which a sample of n was drawn, x_1y_1 , x_2y_2 , ..., x_ny_n , and we computed the simultaneous distribution of $\bar{x}$, $\bar{y}$, s_1 , s_2 , r. Our first result was that the two distributions, that of $\bar{x}$, $\bar{y}$ and that of s_1 , s_2 , r are completely independent [see Eqs. (77) and (78)]. Since our aim was to find the distribution of r we computed the distribution of s_1 , s_2 , r, rather than that of s_{11} , s_{22} , s_{12} . However, for the generalization from $k = 2$ to general k it is preferable to have instead of (60) the density $f_n(s_{11}$, s_{22} , $s_{12})$. Then

$$\begin{aligned} f_n(s_{11}, s_{22}, s_{12})\, ds_{11}\, ds_{22}\, ds_{12} &= f_n(s_1^2, s_2^2, rs_1s_2) \cdot 2s_1\, ds_1 \cdot 2s_2\, ds_2 s_1 s_2\, dr \\ &= 4s_1^2 s_2^2 f_n(s_1^2, s_2^2, rs_1s_2)\, ds_1\, ds_2\, dr \\ &= q_n(s_1, s_2, r)\, ds_1\, ds_2\, dr\,, \end{aligned}$$

where $q_n(s_1$, s_2 , $r)$ is the density in (60). We obtain, therefore,

$$f_n(s_{11}, s_{22}, s_{12}) =$$

$$\frac{n^{n-1} s_1^{n-4} s_2^{n-4} (1 - r^2)^{(n-4)/2}}{4\pi\sigma_1^{n-1}\sigma_2^{n-1}(1 - \rho^2)^{(n-1)/2}\Gamma(n - 2)} \exp\Big(-\frac{n}{2(1 - \rho^2)}\Big[\frac{s_1^2}{\sigma_1^2} - \frac{2\rho s_{12}}{\sigma_1\sigma_2} + \frac{s_2^2}{\sigma_2^2}\Big]\Big)\,. \tag{95}$$

Here, we may introduce (see Problem 23) the determinants D and Δ of the matrices of the $s_{\iota\kappa}$ and $\sigma_{\iota\kappa}$, respectively, and obtain

$$f_n(s_{11}, s_{22}, s_{12}) = \frac{n^{n-1}}{4\pi\Gamma(n - 2)} \frac{D^{(n-4)/2}}{\Delta^{(n-1)/2}} \exp\Big(-\frac{n}{2\Delta}(s_{11}\sigma_{22} - 2s_{12}\sigma_{12} + s_{22}\sigma_{11})\Big). \tag{96}$$

In the case of general k we start with a k-variate normal parent distribution and a sample consisting of kn values $x_{11}, x_{12}, \ldots, x_{1k}, x_{21}, x_{22}, \ldots, x_{2k}, \ldots, x_{n1}, x_{n2}, \ldots, x_{nk}$. From this we compute the averages $\sum_{\nu=1}^{n} x_{\nu 1} = n\bar{x}_1, \ldots, \sum_{\nu=1}^{n} x_{\nu k} = n\bar{x}_k$. The k^2 components $s_{\iota\kappa}$ of the variance, where $s_{\iota\kappa} = s_{\kappa\iota}$, are as always $\sum_{\nu=1}^{n} (x_{\nu\iota} - \bar{x}_\iota)(x_{\nu\kappa} - \bar{x}_\kappa) = ns_{\iota\kappa} = nr_{\iota\kappa}s_\iota s_k$, and $s_{\iota\iota} = s_\iota^2$.

The first result, that of *the independence of the distributions of moments of first order and of moments of second order* generalizes easily and we have (see p. 593):

The distribution of the $\bar{x}_\kappa$, $\kappa = 1, 2, \ldots, k$ is normal with the mean values α_κ, $\kappa = 1, 2, \ldots, k$ of the normal parent distribution and the matrix of variances $n^{-1}\Delta = n^{-1}\sigma_1\sigma_2 \ldots \sigma_k P$.

It remains to generalize the result (96′) from $k = 2$ to general k. This computation has been carried out by Wishart[1] and we shall not reproduce it here since (except for the value of the constant) the result is a straightforward generalization of (96′). We remember that we denote in D the minors of order $k - 1$ by $D_{\iota\kappa}$ and their theoretical counterparts by $\Delta_{\iota\kappa}$. *If the $x_{\nu\kappa}$, $\kappa = 1, \ldots, k$, $\nu = 1, \ldots, n$ are sampled from a k-variate normal distribution, the joint distribution of the $\frac{1}{2}k(k+1)$ variables $s_{\iota\kappa}$ has the density $f_n(s_{11}, \ldots, s_{kk})$ given by*

$$f_n(s_{11}, \ldots, s_{kk}) = C_k \frac{D^{(n-k-2)/2}}{\Delta^{(n-1)/2}} \cdot \exp\Big(-\frac{n}{2\Delta} \cdot \sum_{\iota,\kappa} \Delta_{\iota\kappa} s_{\iota\kappa}\Big) \tag{97}$$

where

$$C_k = \Big(\frac{n}{2}\Big)^{\frac{k(n-1)}{2}} \frac{1}{\pi^{k(k-1)/4}\,\Gamma\Big(\frac{n-1}{2}\Big)\,\Gamma\Big(\frac{n-2}{2}\Big) \cdots \Gamma\Big(\frac{n-k}{2}\Big)}. \tag{97′}$$

For $k = 2$ we use the formula $\Gamma(p)\Gamma(p + \frac{1}{2}) \cdot 2^{2p-1} = \sqrt{\pi}\Gamma(2p)$ with $p = (n-2)/2$ and obtain $C_2 = n^{n-1}/4\pi\Gamma(n-2)$ as in (96′). The distribution (97) is the *Wishart distribution*. It forms the starting point of further work belonging to multivariate statistical analysis. We shall, however, not go on with these considerations.

[1] J. WISHART, "The generalized product moment distribution in samples from a normal multivariate distribution." *Biometrica* **20A** (1928), pp. 32-52 and a proof based on characteristic functions, in J. WISHART and M. S. BARTLETT, "The generalized product moment distribution." *Proc. Cambridge Phil. Soc.* **29** (1933), pp. 260-270.

D. First Comments on Statistical Functions (Section 8)

8. Asymptotic Expectation and Variance of Statistical Functions

8.1. *Proof of formulas* (41) *and* (42). In this concluding section a proof for formulas (41) and (42) will be given. They concern the asymptotic value of expectation and variance for functions that depend on the distribution of the outcome of n trials. Such functions may conveniently be called *statistical functions.*

Consider a collective K_1 with k label values $a_1, a_2, \ldots, a_k$ and probabilities $\pi_1, \pi_2, \ldots, \pi_k$ the sum of which is 1. If n successive elements of K_1 are considered as one element of a new collective this has an n-dimensional label space with k^n different label points. We mix all label points with the same "repartition of coordinates," that is, those points for which n_1 of the n coordinates equal $a_1, \ldots, n_k$ equal a_k. After this mixing operation we have a collective K_n with a $(k-1)$-dimensional label, the components of which are either $n_1, n_2, \ldots, n_{k-1}$ or the quotients $p_1 = n_1/n$, $p_2 = n_2/n, \ldots, p_{k-1} = n_{k-1}/n$. Any one of these variables can take one of the values $0, 1/n, 2/n, \ldots, (n-1)/n, 1$, with the restriction that the sum be not greater than 1. Any function of $p_1, p_2, \ldots, p_{k-1}$ can also be written as a function of $p_1, p_2, \ldots, p_k$ where $p_1 + p_2 + \cdots + p_k = 1$, and any such function must have a definite expectation $E[f]$, a variance $\mathrm{Var}[f]$ distribution $P(x) = \Pr\{f \leqslant x\}$, all determined by the distribution in K_n.

If f is a twice differentiable function of the $p_1, p_2, \ldots, p_k$ we can use the Taylor formula

$$f(p_1, p_2, \ldots, p_k) = f(\pi_1, \pi_2, \ldots, \pi_k) + \sum_{1=\kappa}^{k} (p_\kappa - \pi_\kappa) \frac{\partial f}{\partial p_\kappa}(\pi_1, \ldots, \pi_k) + R_2, \quad (98)$$

and, therefore,

$$E[f] = f(\pi_1, \pi_2, \ldots, \pi_k) + \sum_{\kappa=1}^{k} \frac{\partial f}{\partial p_\kappa}(\pi_1, \ldots, \pi_k) E[p_\kappa - \pi_\kappa] + E[R_2]. \quad (99)$$

It is easily seen and has been explained in previous instances that the expectation of $p_\kappa - \pi_\kappa$ vanishes. We have for the variance

$$E[(p_1 - \pi_1)^2] = \frac{1}{n^2} E[(n_1 - n\pi_1)^2] = \frac{1}{n} \pi_1(1 - \pi_1) \leqslant \frac{1}{4n}, \quad (100)$$

according to Eq. (27) in Chapter IV. From (99) we have then

$$| E[f] - f(\pi_1, \pi_2, \ldots, \pi_k) | = | E[R_2] | \leqslant E[| R_2 |], \quad (101)$$

and since the Taylor formula supplies

$$| R_2 | \leqslant \tfrac{1}{2} M_2 \sum_{\iota,\kappa}^{1\cdots k} (p_\iota - \pi_\iota)(p_\kappa - \pi_\kappa), \tag{102}$$

where M_2 is an upper bound for the various second derivatives of f:

$$| E[f] - f(\pi_1, \pi_2, \ldots, \pi_k) | \leqslant \tfrac{1}{2} M_2 \sum_{\iota,\kappa} E[| p_\iota - \pi_\iota | \, | p_\kappa - \pi_\kappa |]. \tag{103}$$

Out of the k^2 terms in this sum those with $\iota = \kappa$ are given by Eq. (100) and for the others we use the formula $| 2ab | \leqslant a^2 + b^2$, which supplies

$$E[| ab |] \leqslant \tfrac{1}{2} E[a^2] + \tfrac{1}{2} E[b^2]. \tag{104}$$

This inequality shows that

$$E[| p_\iota - \pi_\iota | \, | p_\kappa - \pi_\kappa |] \leqslant \frac{1}{2n} \pi_\iota(1 - \pi_\iota) + \frac{1}{2n} \pi_\kappa(1 - \pi_\kappa) \leqslant \frac{1}{4n}. \tag{104'}$$

Thus, the sum in (103) consists of a finite number of terms each of which is, in absolute value, surpassed by $\frac{1}{4}n$. Herewith Eq. (41) is proved.

We proceed in the same way to derive the asymptotic expression (42), here extending the Taylor development up to the second order term. Setting

$$F(p_1, p_2, \ldots, p_k) = [f(p_1, p_2, \ldots, p_k) - f(\pi_1, \pi_2, \ldots, \pi_k)]^2 \tag{105}$$

we have (if we first differentiate and then put $\pi_\iota = p_\iota$)

$$F(\pi_1, \pi_2, \ldots, \pi_k) = 0, \qquad \frac{\partial F}{\partial p_\iota}(\pi_1, \pi_2, \ldots, \pi_k) = 0, \qquad \frac{\partial^2 F}{\partial p_\iota \partial p_k}(\pi_1, \pi_2, \ldots, \pi_k)$$

$$= 2 \frac{\partial f}{\partial p_\iota}(\pi_1 \ldots, \pi_k) \frac{\partial f}{\partial p_\kappa}(\pi_1 \ldots, \pi_k),$$

and, therefore

$$F(p_1, p_2, \ldots, p_k) = \sum_{\iota,\kappa}^{1\cdots k} f_\iota f_\kappa (p_\iota - \pi_\iota)(p_\kappa - \pi_\kappa) + R_3. \tag{106}$$

where the abbreviation f_ι for the first derivative is used as in Section 4.1. It follows that

$$\mathrm{Var}[f] = E[F] = \sum_{\iota,\kappa}^{1\cdots k} f_\iota f_\kappa E[(p_\iota - \pi_\iota)(p_\kappa - \pi_\kappa)] + E[R_3]. \tag{107}$$

The expectation of the products with $\iota = \kappa$ was seen in (100) to equal $\pi_\iota(1 - \pi_\iota)/n$. The expectation of a term with $\iota \neq \kappa$ has been computed in Chapter IX, Section 4.2.[1] The magnitudes $n_\kappa - nq_\kappa = \delta_\kappa$ are in the present notation (with $n_\kappa/n = p_\kappa$ and π_κ instead of q_κ) $\delta_\kappa = n(p_\kappa - \pi_\kappa)$, and from $E[\delta_1\delta_2] = -nq_1q_2$ follows

$$E[(p_\iota - \pi_\iota)(p_\kappa - \pi_\kappa)] = -\frac{1}{n}\pi_\iota\pi_\kappa . \tag{108}$$

From (99) and (108) we find the sum in (107):

$$\begin{aligned}\sum_{\iota,\kappa}^{1\cdots k} f_\iota f_\kappa E[(p_\iota - \pi_\iota)(p_\kappa - \pi_\kappa)] &= \frac{1}{n}\sum_{\kappa=1}^{k} f_\kappa^2\pi_\kappa - \frac{1}{n}\sum_{\iota,\kappa}^{1\cdots k} f_\iota f_\kappa\pi_\iota\pi_\kappa \\ &= \frac{1}{n}\sum_{\kappa=1}^{k} f_\kappa^2\pi_\kappa - \frac{1}{n}\Big(\sum_{\kappa=1}^{k} f_\kappa\pi_\kappa\Big)^2 .\end{aligned} \tag{109}$$

Therefore, our formula (42) will be proved if we show that $nE[R_3]$ tends to zero with increasing n.

Calling M_3 an upper bound for the third derivatives of F, the remainder term in the Taylor formula leads to

$$E[|\, R_3 \,|] \leqslant \tfrac{1}{6}M_3 \sum_{\iota,\kappa,\lambda}^{1\cdots k} E[|\, (p_\iota - \pi_\iota)(p_\kappa - \pi_\kappa)(p_\lambda - \pi_\lambda) \,|]. \tag{110}$$

In order to find an upper bound for the expectations of the absolute values of the triple products we note that the inequality

$$E[|\, f \,|] \leqslant \sqrt{E[f^2]}. \tag{111}$$

holds whatever f is. Thus,

$$E[|\, (p_1 - \pi_1)(p_2 - \pi_2)(p_3 - \pi_3) \,|\,] \leqslant \sqrt{E[(p_1 - \pi_1)^2(p_2 - \pi_2)^2(p_3 - \pi_3)^2]} . \tag{112}$$

The expectation on the right-hand side can be computed by means of the method of generating functions.

We need however only an estimate of the expectation of expressions

[1] Now we need the correct value of $E[(p_\iota - \pi_\iota)(p_\kappa - \pi_\kappa)]$, not an estimate only as in Eq. (104′).

like $\delta_1^2\delta_2^2\delta_3^2$, or $\delta_1^4\delta_2^2$, etc., where $\delta_1 = n(p_1 - \pi_1)$, etc. We know from our study of the moments of the Bernoulli distribution (Chapter IV, Section 3) that these expectations are polynomials in n; in the present case of moments of order 6 the polynomials are of order 3 in n. Hence the expectation of $(p_1 - \pi_1)^2(p_2 - \pi_2)^2(p_3 - \pi_3)^2$ is of the order $n^3/n^6 = n^{-3}$, and the same holds, of course, for the other terms which appear as squares of the right-hand side in (110). Therefore, the order of the expectations in (110) is, according to (111), not higher than $n^{-3/2}$. Multiplied by n this gives an expression with the factor $n^{-1/2}$, that is, vanishing with increasing n. Thus, the proof has been completed.

8.2. *Further Remarks on statistical functions.* Equations (41) and (42) mark a first step in the *theory of statistical functions*, that is, of functions depending on the repartition of n random results. One important generalization of our formulas is this: If n collectives with *different* distributions are combined, so that $\pi_1, \pi_2, \ldots, \pi_k$ take different values for each of them, the analogous equations hold:

$$\lim_{n\to\infty} E[f(p_1, p_2, \ldots, p_k)] = f(\bar{\pi}_1, \bar{\pi}_2, \ldots, \bar{\pi}_k) \tag{113}$$

$$\lim_{n\to\infty} n \operatorname{Var}[f(p_1, p_2, \ldots, p_k)] = \sum_{\kappa=1}^{k} f_\kappa^2 \bar{\pi}_\kappa - \sum_{\nu,\kappa}^{1\cdots k} f_\iota f_\kappa \overline{\pi_\iota \pi_\kappa}, \tag{114}$$

where $\bar{\pi}_\kappa$ and $\overline{\pi_\iota\pi_\kappa}$ denote the arithmetical means of the n values of π_κ and $\pi_\iota\pi_\kappa$, respectively. Furthermore, the asymptotic expressions for all moments of higher order than two can be derived in the same way. The main results in this respect are: (a) The moments of order $2m$ and of order $(2m + 1)$ behave asymptotically as finite expressions divided by the mth power of n; (b) These finite expressions depend on the two sets of means $\bar{\pi}_\kappa$ and $\overline{\pi_\iota\pi_\kappa}$ only ($\iota, \kappa = 1, 2, \ldots, k$). Once the complete sequence of moments is known, it is possible in principle to draw conclusions on the asymptotic behavior of the distribution of $f(p_1, p_2, \ldots, p_k)$.

The simplest case of a statistical function is the *linear* case

$$f = a_1p_1 + a_2p_2 + \cdots + a_kp_k .$$

It has been seen in Chapter VI that the distribution of the arithmetical means of n observations tends, under very light restrictions for the a_κ and π_κ, toward the normal or Gaussian distribution as n increases. This classical theorem (central limit theorem) does not represent adequately the actual situation. The fact is that "in general" the distribution

of any "differentiable statistical function," whether linear or not, tends toward normality with increasing n. Thus, for instance, moments of any order of a set of n independent chance variables, and also functions of those moments like the Lexis quotient or Student's t or the correlation coefficient, etc., are normally distributed for infinite n. It is this general limit theorem which is tacitly presupposed if we judge the distribution of a random variable from its mean value and variance. Note that Pearson's X^2 is an exceptional case, just as v. Mises' ω^2. The asymptotic distribution of X^2 was computed in Chapter IX, Section 4.4. that of ω^2 in Chapter IX, Section 7.2. The theory of statistical functions allows us to deal in a general way also with such cases that do not lead to normal distributions.

In the present section we spoke only of discrete distributions with finite point probabilities π_κ. The argument, however, can be extended to the continuous case also. The quantity f becomes then a function of the repartition $S_n(x)$ of the set of observations and f_κ is replaced by a more complicated concept, the derivative of a "functional" (*fonction de ligne*, after Volterra). The study of those functions defined in the "space of distribution functions" is one of the basic elements on which further progress in the mathematical theory of statistics depends. An introduction to the theory of statistical functions will be the subject of our last chapter.