ON CORRELATION SURFACES OF SUMS WITH A CERTAIN NUMBER OF RANDOM ELEMENTS IN COMMON*

By

CARL H. FISCHER

Introduction. The study of correlation due to a common factor has been a more or less familiar one in the literature of mathematical statistics. Kapteyn,[1] in an exposition of the Pearsonian coefficient of correlation, considered the correlation between two sums of normally distributed variables, the sums having k random elements in common. In 1920, Rietz[2] devised urn schemata which yield sums with common items involved in such a way that the correlation and regression properties can be dealt by a priori methods. In a later paper, Rietz[3] considered two variables, each the sum of two random drawings of elements from a continuous rectangular distribution, with one of the elements in common. Here, the emphasis was placed principally upon the description of the correlation surface. Some other aspects and extensions of this problem were brought out by Karl Pearson[4] in an editorial discussion of Rietz's paper.

In the literature, the theory of correlation has been discussed principally in connection with its applications. One of the objects of some of the above-mentioned papers is the establishment of a closer connection between correlation theory and abstract probability theory. Such a connection would give a more precise

*Presented to the American Mathematical Society, Dec. 28, 1931.

[1]J. C. Kapteyn, "Definition of the Correlation-Coefficient," Monthly Notices of the Royal Astronomical Society, Vol. 72 (1912), pp. 518-525.

[2]H. L. Rietz, "Urn Schemata as a Basis for the Development of Correlation Theory," Annals of Mathematics, Vol. 21 (1920), pp. 306-322.

[3]H. L. Rietz, "A Simple Non-Normal Correlation Surface," Biometrika, Vol. 24 (1932), pp. 288-291.

[4]Karl Pearson, "Professor Rietz's Problem," (Editorial), Biometrika, Vol. 24 (1932), pp. 290-291.

meaning to correlation and would tend to make the study of correlation theory more attractive to mathematicians. With this aim in view, the present paper is concerned with correlation among sums having common elements, extending and generalizing the preceding papers in several ways.

We shall assume our drawings made from a continuous universe characterized by a rather arbitrary law of distribution. We shall define n sums, each of an arbitrary number of elements, formed in such a manner that any two consecutive sums have elements in common, and inquire into the correlation between any two of these sums. The equations of the correlation surfaces will be expressed in terms of iterated integrals, the regression of each variable on the other will be shown to be linear, and the equations of the regression lines will be obtained. The coefficient of correlation may then be computed from the slopes of these lines.

Throughout this paper we shall understand a probability function, $f(t)$, to be, for all values of t on a range R, a single-valued, real-valued, non-negative, continuous function of t. It is then Riemann integrable on R, and we shall require that $\int_R f(t)dt = 1$. We define the probability that a value of t, drawn at random from the range R, lie in the interval (a, b), a and b in R and $b>a$, to be $\int_a^b f(t)dt$. We may then say that $f(t)dt$ is, to within infinitesimals of higher order, the probability that a value of t drawn at random lies in the interval $(t, t+\Delta t)$. Bachelier[5] has classified probabilities into those of the first, second, and third kinds, and Craig[6] has extended this to probability functions, according as R is the range $(-\infty,\infty)$, $(0,\infty)$, and $(0, a)$, respectively. We shall find it convenient to adopt this classification.

[5] L. Bachelier, "Calcul des Probabilities." (1912), p. 155.

[6] Allen T. Craig, "On the Distribution of Certain Statistics," American Journal of Mathematics, Vol. 54 (1932), pp. 353-366.

I. Sums of elements drawn from a universe characterized by a probability function of the first kind.

1. The correlation between two sums having random elements in common. Let $f(t)$, a probability function of the first kind, characterize the distribution of the variable t. Let the principal variable x_1 be defined as the sum of n_1 independent values of t drawn at random. Further, let the principal variable x_2 be defined as the sum of k_{12} random values of the n_1 values of t composing x_1 and of $n_2 - k_{12}$ independent random values of t taken directly from the universe characterized by $f(t)$.

Theorem I. *Given the sums x_1 and x_2 as defined above, with k_{12} random elements in common.*

a) *The marginal distributions of x_1 and x_2 are given, respectively, by*

$$G_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_{11}) f(t_{12}) \cdots f(t_{1,n_1-1}) f(x_1 - t_{11} - \cdots - t_{1,n_1-1})\, dt_{1,n_1-1} \cdots dt_{11}, \tag{1.11}$$

and

$$\begin{aligned} G_2(x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} & f(t_{11}) \cdots f(t_{1k_{12}}) f(t_{2,k_{12}+1}) \cdots f(t_{2,n_2-1}) \\ & \times f(x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1} - \cdots - t_{2,n_2-1})\, dt_{2,n_2-1} \cdots dt_{2,k_{12}+1}\, dt_{1k_{12}} \cdots dt_{11}. \end{aligned} \tag{1.12}$$

b) *The correlation surface, $w = F(x_1, x_2)$, or the simultaneous law of distribution of x_1 and x_2, is given by*

$$\begin{aligned} F(x_1, x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} & f(t_{11}) \cdots f(t_{1,n_1-1}) f(x_1 - t_{11} - \cdots - t_{1,n_1-1}) f(t_{2,k_{12}+1}) \cdots f(t_{2,n_2-1}) \\ & \times f(x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1} - \cdots - t_{2,n_2-1})\, dt_{2,n_2-1} \cdots dt_{2,k_{12}+1}\, dt_{1k_{12}} \cdots dt_{11}. \end{aligned} \tag{1.2}$$

c) *The regression curves of x_2 on x_1 and of x_1 on x_2 are linear, and are given, respectively, by the following equations:*

$$\bar{x}_2 = \frac{k_{12} x_1}{n_1} + (n_2 - k_{12}) M, \tag{1.31}$$

and

$$\bar{x}_1 = \frac{k_{12} x_2}{n_2} + (n_1 - k_{12}) M, \tag{1.32}$$

where $$M = \int_{-\infty}^{\infty} t f(t)\, dt.$$

Hence, the coefficient of correlation between x_1 *and* x_2 *is*

$$r_{x_1 x_2} = \frac{k_{12}}{(n_1 n_2)^{1/2}}.$$

Proof. The proof for the expressions for the marginal distributions of x_1 and x_2 are given by Craig[7] and need not be repeated here. The correlation surface $w = F(x_1, x_2)$ is derived by a simple extension of the same method to two independent variables.

The regression curve of x_2 on x_1 is the locus of the ordinate of the centroid $\bar{x}_2$ of a section of the surface for any given x_1. Thus

$$\bar{x}_2 = \frac{\left[\int_{-\infty}^{\infty} x_2 F(x_1, x_2)\, dx_2\right]}{\left[\int_{-\infty}^{\infty} F(x_1, x_2)\, dx_2 \quad\right]}. \tag{1.4}$$

It will be convenient in what follows to use an abbreviated notation by letting

$$\theta(x_1, t_{11}, \cdots t_{1,n_1-1}) = f(t_{11}) \cdots f(t_{1,n_1-1}) f(x_1 - t_{11} - \cdots - t_{1,n_1-1}), \tag{1.5}$$

which is merely the integrand of the marginal distribution of x_1. Where no ambiguity can result, θ_{x_1} will be used in place of

$\theta(x_1, t_{11}, \cdots t_{1,n_1-1})$. Then $F(x_1, x_2)$ may be written

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \theta(x_1, t_{11}, \cdots t_{1,n_1-1}) \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1} - \cdots - t_{2,n_2-1})$$

$$\times\, dt_{2,n_2-1} \cdots dt_{2,k_{12}+1}\, dt_{1,n_1-1} \cdots dt_{11}.$$

Now let $v = x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1} - \cdots - t_{2,n_2-1}$. Changing the variable

[7]Allen T. Craig, loc. cit., pp. 355-356.

from x_2 to v, (1.4) becomes

$$(1.6)\quad \bar{x}_2 = \Big\{ \Big[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_{11} + \cdots + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_{1k_{12}} + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_{2,k_{12}+1} + \cdots$$

$$+ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_{2,n_2-1} + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} v \Big] \theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v)\, dt_{2,n_2-1} \cdots dt_{11}\, dv \Big\}$$

$$\Big/ \Big\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v)\, dt_{2,n_2-1} \cdots dt_{11}\, dv \Big\}.$$

It will be noted that the terms in the numerator fall into two groups: those terms containing the factors t_{1i}, $(i = 1, 2, \cdots k_{12})$, and those terms containing the factors v or t_{2j}, $(j = k_{12}+1, k_{12}+2, \cdots n_2-1)$. Further, since the order of integration here is immaterial, the equality of the k_{12} integrals of the first group follows readily. Similarly, the equality of the $n_2 - k_{12}$ integrals of the second group follows. The expression (1.6) may then be written

$$(1.7)\quad \bar{x}_2 = \Big\{ k_{12} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_{11} \theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v)\, dv\, dt_{2,n_2-1} \cdots dt_{11}$$

$$+ (n_2 - k_{12}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} v \theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v)\, dv\, d_{2,n_2-1} \cdots dt_{11} \Big\}$$

$$\Big/ \Big\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v)\, dv\, dt_{2,n_2-1} \cdots dt_{11} \Big\}$$

In (1.7), it is clear that the integrations with respect to each t_{2j} may be effected immediately, making use of $\int_{-\infty}^{\infty} f(v)\, dv = 1$ In the first term of the numerator and in the denominator the variable v may likewise be integrated out. The denominator is now equal to (1.11), the marginal distribution function of x_1. In the second term of the numerator, $v \cdot f(v)$ is independent of the remaining factors, and $\int_{-\infty}^{\infty} v f(v)\, dv$ is a constant which we shall denote by M. This second term of the numerator is now equal to $(n_2 - k_{12})M$ times the marginal distribution function of x_1

Hence, we have now reduced the expression (1.7) for $\bar{x}_2$ to the following form:

$$(1.8) \qquad \bar{x}_2 = k_{12} I_{n_1} + (n_2 - k_{12}) M,$$

where $$I_{n_1} = \frac{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} t_{11}\,\theta(x_1, t_{11}, \cdots t_{1,n_1-1})\,dt_{1,n_1-1}\cdots dt_{11}}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \theta(x_1, t_{11}, \cdots t_{1,n_1-1})\,dt_{1,n_1-1}\cdots dt_{11}}.$$

To evaluate I_n, let $t_{11} = x_1 - u - t_{12} - \cdots - t_{1,n_1-1}$.

Then
$$I_{n_1} = \frac{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1\,\theta(x_1, u, t_{12}, \cdots t_{1,n_1-1})\,dt_{1,n_1-1}\cdots dt_{12}\,du}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \theta(x_1, u, t_{12}, \cdots t_{1,n_1-1})\,dt_{1,n_1-1}\cdots dt_{12}\,du}$$

$$- \frac{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} u\,\theta(x_1, u, t_{12}, \cdots t_{1,n_1-1})\,dt_{1,n_1-1}\cdots dt_{12}\,du}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \theta(x_1, u, t_{12}, \cdots t_{1,n_1-1})\,dt_{1,n_1-1}\cdots dt_{12}\,du}$$

$$- \sum_{j=2}^{n_1-1} \frac{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} t_{1j}\,\theta(x_1, u, t_{12}, \cdots t_{1,n_1-1})\,dt_{1,n_1-1}\cdots dt_{12}\,du}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \theta(x_1, u, t_{12}, \cdots t_{1,n_1-1})\,dt_{1,n_1-1}\cdots dt_{12}\,du}.$$

The first term in the above expression for I_{n_1} is equal to x_1. Each of the remaining $n_1 - 1$ terms is equal to I_{n_1}. Hence

$$I_{n_1} = x_1 - (n_1 - 1) I_{n_1},$$

and

$$I_{n_1} = \frac{x_1}{n_1}.$$

From (1.8) and (1.9), we have

$$\bar{x}_2 = \frac{k_{12} x_1}{n_1} + (n_2 - k_{12}) M.$$

In exactly the same manner, we may show that

$$\bar{x}_1 = \frac{k_{12} x_2}{n_2} + (n_1 - k_{12}) M.$$

Making use of the fact that in the case of linear regression the square of the correlation coefficient is equal to the product of the slopes of the two lines of regression, we obtain

$$r_{x_1 x_2} = \frac{k_{12}}{(n_1 n_2)^{\frac{1}{2}}},$$

which completes the proof of the theorem.

Corollary. If x and y are each the sum of n independent random values of a variable t from a universe characterized by $f(t)$, and have k of these values in common, the coefficient of correlation between x and y is equal to the ratio of the number of values of t held in common to the total number composing each principal variable. Thus, $r_{xy} = \frac{k}{n}$.

This corollary of Theorem I was proved by Kapteyn[8] for the special case of a normal parent distribution of the variable t.

Illustration. As a simple illustration of the application of the foregoing theorem, let us consider the case where $x_1 = t_{11} + t_{12}$, $x_2 = t_{11} + t_{22}$ with t_{11}, t_{12}, t_{22}, as independent random drawings of t from the Gaussian distribution,

$$f(t) = (2\pi)^{-\frac{1}{2}} e^{-\frac{t^2}{2}}$$

From (1.11), the marginal distribution of x_1 is

$$G_1(x_1) = (4\pi)^{-\frac{1}{2}} e^{-\frac{x_1^2}{4}}.$$

Similarly, the marginal distribution of x_2 is

$$G_2(x_2) = (4\pi)^{-\frac{1}{2}} e^{-\frac{x_2^2}{4}}$$

The correlation surface, $w = F(x_1, x_2)$, obtained by applying (1.2), is

$$F(x_1, x_2) = \frac{e^{-\frac{(x_1^2 - x_1 x_2 + x_2^2)}{3}}}{(2\pi \cdot 3^{\frac{1}{2}})},$$

a normal correlation surface with $r_{x_1 x_2} = \frac{1}{2}$.

2. The correlation among three sums. We now proceed to extend the preceding theorem to more than two sums. Let us define a third sum, or principal variable, x_3, as the sum of k_{23}

[8] J. C. Kapteyn, loc. cit.

elements taken at random from the n_2 values of t composing x_2 plus the sum of n_3-k_{23} independent random values of t drawn from the parent population. It is apparent, then, that the marginal distributions of x_1, x_2, and x_3, and the correlation surfaces $F_1(x_1,x_2)$ and $F_2(x_2,x_3)$ will be formed exactly as were those of x_1 and x_2 in Theorem I. From this theorem, we are at once in a position to write the equations of the lines of regression and the coefficients of correlation for these surfaces. The surface $w = F(x_1,x_3)$ remains to be investigated, as does the four-dimensional surface, $v=\psi(x_1,x_2,x_3)$, which may be obtained in almost the same manner.

Theorem II. *Given $f(t)$ and x_1, x_2, x_3, as defined above. Let θ_{x_1}. be defined as in* (1.5). *Let*

$$\phi(x,t_{11},\cdots t_{1,k_{23}-g},\ t_{2,k_{23}-g+1},\cdots t_{2k_{23}},\ t_{3,k_{23}+1},\cdots t_{3,n_3-1})= f(t_{2,k_{23}-g+1})\cdots f(t_{2k_{23}})f(t_{3,k_{23}+1})\cdots f(t_{3,n_3-1})$$
$$\times f(x-t_{11}-\cdots-t_{1,k_{23}-g}-t_{2,k_{23}-g+1}-\cdots-t_{2k_{23}}-t_{3,k_{23}+1}-\cdots-t_{3,n_3-1}).$$

If $f(t)$ is a probability function of the first kind, then the expression for the simultaneous distribution of x_1 and x_3 is

$$F(x_1,x_3)=\sum_{g=0}^{k_{23}}\left\{\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\theta_{x_1}\phi(x_3,t_{11},\cdots t_{1,k_{23}-g},\right.$$
$$t_{2,k_{23}-g+1},\cdots t_{2k_{23}},\ t_{3,k_{23}+1},\cdots t_{3,n_3-1})dt_{3,n_3-1}\cdots dt_{3,k_{23}+1}$$
$$\left.\times\, dt_{2k_{23}}\cdots dt_{2,k_{23}-g+1}\,dt_{1,n_1-1}\cdots dt_{11}\right\}\Big/\binom{n_2}{k_{23}}, \tag{2.1}$$

where by $\binom{c}{d}$ is understood the number of combinations of c items taken d at a time.

Proof. Let us temporarily require that $k_{12} \geq k_{23}$. We shall show later that this restriction may be removed. The probability that x_1 and x_3 as defined contain $k_{23}-g$, $(g=0,1,2,\cdots k_{23})$, elements in common is $\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}\Big/\binom{n_2}{k_{23}}$.

The probability of the occurrence of any given pair of values (x_1, x_3), that is, the probability of a point falling into a given rectangle, $(x_1, x_1+\Delta x_1, x_3, x_3+\Delta x_3)$, is the sum of the probabilities of all of the mutually exclusive ways in which it can occur. Each of the terms in (2.1) multiplied by $dx_1 dx_3$ consists of the integral, (derived by the method of Theorem I), which is the probability, to within infinitesimals of higher order, of the occurrence of a given pair, (x_1, x_3), with a specified number of values of t in common, times a coefficient which is equal to the probability of the occurrence of this specified number of values of t in common. Each of the terms as a whole, then, is the probability that the given (x_1, x_3) will occur with a specified number of values of t in common. Hence, the expression (2.1), being the sum of the probabilities of all of the mutually exclusive ways in which x_1 and x_3 can fall within the desired rectangle, is the probability that this will occur. This establishes the theorem when $k_{12} \geqq k_{23}$.

If $k_{12} < k_{23}$, then the maximum number of values of t which x_1 and x_3 can have in common is k_{12}. The expression for $F'(x_1, x_3)$ in this case, then, consists of the sum of all of the terms of (2.1) beginning with the term where x_1 and x_3 have k_{12} values of t in common and continuing to include the term derived from the case where they have no values of t in common. Equation (2.1), however, in its present form may be considered as a correct formal expression for the correlation surface even when $k_{12} < k_{23}$, since in this case all of the coefficients of the terms where x_1 and x_3 are to have more than k_{12} values of t in common are zero. This follows from the definition $\binom{c}{d} = 0$ if $c < d$. Thus

$$\binom{k_{12}}{k_{23}} = \binom{k_{12}}{k_{23}-1} = \cdots = \binom{k_{12}}{k_{12}+1} = 0 \text{ if } k_{12} < k_{23}.$$

Hence, we may now remove the restriction that $k_{12} \geqq k_{23}$. This establishes the theorem.

We are now in a position to write down the surface

$$v = \psi(x_1, x_2, x_3).$$

It is given by the following expression, where, by $t_{2,k_{23}-g+1}, \cdots t_{2,k_{23}}$ is meant any g values of the t_{2j} :

$$\psi(x_1, x_2, x_3) = \sum_{g=0}^{k_{23}} \left\{ \binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g} \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \theta_{x_1} f(t_{2,k_{12}+1})\cdots f(t_{2,n_2-1}) \right.$$

$$\times f(x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1} - \cdots - t_{2,n_2-1}) f(t_{3,k_{23}+1})\cdots f(t_{3n_3-1})$$

$$\times f(x_3 - t_{11} - \cdots - t_{1,k_{23}-g} - t_{2,k_{23}-g+1} - \cdots - t_{2,k_{23}} - t_{3,k_{23}+1} - \cdots - t_{3,n_3-1})$$

$$\left. dt_{3,n_3-1} \cdots dt_{3,k_{23}+1}\, dt_{2,n_2-1} \cdots dt_{2,k_{12}+1}\, dt_{1,n_1-1} \cdots dt_{11} \right\}.$$

Theorem III. *The regression curves of x_3 on x_1 and of x_1 on x_3 for the correlation surface $w = F(x_1, x_3)$, defined in Theorem II, are linear and are given, respectively, by the following equations:*

$$\bar{x}_3 = \frac{k_{12} k_{23} x_1}{n_1 n_2} + \frac{(n_2 n_3 - k_{12} k_{23})M}{n_2}, \tag{2.21}$$

and

$$\bar{x}_1 = \frac{k_{12} k_{23} x_3}{n_2 n_3} + \frac{(n_1 n_2 - k_{12} k_{23})M}{n_2}, \tag{2.22}$$

where M is defined as in Theorem I. Further, the coefficient of correlation between x_1 and x_3 is

$$r_{x_1 x_3} = \frac{k_{12} k_{23}}{n_2 (n_1 n_3)^{\frac{1}{2}}} = r_{x_1 x_2} r_{x_2 x_3}. \tag{2.3}$$

Proof. As in the proof of Theorem I, we set up the expression for the locus of the ordinate of the centroid of a section of the surface for a fixed x_1. We have

$$\bar{x}_3 = \frac{\int_{-\infty}^{\infty} x_3 F(x_1, x_3)\, dx_3}{\int_{-\infty}^{\infty} F(x_1, x_3)\, dx_3}$$

where $F(x_1, x_3)$ is given by (2.1). From the definition of a marginal distribution, we know that $\int_{-\infty}^{\infty} F(x_1, x_3)\,dx_3$ reduces to (1.11), the marginal distribution of x_1. Let us now write the expression for $\bar{x}_3$ as the sum of $k_{23}+1$ fractions. Thus

$$(2.4)\quad \bar{x}_3 = \sum_{g=0}^{k_{23}} \binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}\Big\{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_3\,\theta_{x_1}$$

$$\times\,\phi(x_3, t_{11}, \cdots t_{1,k_{23}-g}, t_{2,k_{23}-g+1}, \cdots t_{2k_{23}}, t_{3,k_{23}+1}, \cdots t_{3,n_3-1})$$

$$\times\, dt_{3,n_3-1}\cdots dt_{3,k_{23}+1}\, dt_{2k_{23}}\cdots dt_{2,k_{23}-g+1}\, dt_{1,n_1-1}\cdots dt_{11}\Big\}$$

$$\Big/\Big\{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\theta_{x_1}\,dt_{1,n_1-1}\cdots dt_{11}\binom{n_2}{k_{23}}\Big\}.$$

Hereafter, we shall call an expression of the form

$$\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}\Big/\binom{n_2}{k_{23}}$$

a "probability coefficient." Then (2.4) is the sum of products, each of which is a probability coefficient times an expression which is equivalent to the expression for $\bar{x}_3$ for the simple case where x_3 would be derived directly from x_1 by the drawing of $k_{23}-g$ values of t from x_1. These latter expressions, by the application of Theorem I, may each be written in the same form as (1.3). Hence, (2.4) has been reduced to

$$(2.5)\quad \bar{x}_3 = x_1\Big[\frac{1}{n_1}\sum_{g=0}^{k_{23}-1}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}(k_{23}-g)\Big]\Big/\binom{n_2}{k_{23}}$$

$$+M\Big[\sum_{g=0}^{k_{23}}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}(n_3-k_{23}+g)\Big]\Big/\binom{n_2}{k_{23}}$$

$$=\frac{x_1 k_{12}}{n_1}\Big[\sum_{g=0}^{k_{23}-1}\binom{k_{12}-1}{k_{23}-g-1}\binom{n_2-k_{12}}{g}\Big]\Big/\binom{n_2}{k_{23}}$$

$$+M\left[(n_3-k_{23})\sum_{g=0}^{k_{23}}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}\right.$$

$$\left.+\sum_{g=0}^{k_{23}}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}g\right]\Big/\binom{n_2}{k_{23}}.$$

By the use of a well-known theorem of combinatory analysis,[9] we have that

$$\frac{1}{n_1}\sum_{g=0}^{k_{23}-1}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}(k_{23}-g)\Big/\binom{n_2}{k_{23}}=\frac{k_{12}}{n_1}\binom{n_2-1}{k_{23}-1}\Big/\binom{n_2}{k_{23}}=\frac{k_{12}k_{23}}{n_1n_2},$$

and

$$(n_3-k_{23})\sum_{g=0}^{k_{23}}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}\Big/\binom{n_2}{k_{23}}=\binom{n_2}{k_{23}}(n_3-k_{23})\Big/\binom{n_2}{k_{23}}=(n_3-k_{23}).$$

Moreover,

$$\sum_{g=0}^{k_{23}}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g}g\Big/\binom{n_2}{k_{23}}=(n_2-k_{12})\sum_{g=1}^{k_{23}}\binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}-1}{g-1}\Big/\binom{n_2}{k_{23}},$$

which reduces to

$$(n_2-k_{12})\binom{n_2-1}{k_{23}-1}\Big/\binom{n_2}{k_{23}}=(n_2-k_{12})k_{23}/n_2$$

by the same theorem of combinatory analysis.

Hence, (2.5) becomes

$$\bar{x}_3=\frac{k_{12}k_{23}x_1}{n_1n_2}+\frac{(n_2n_3-k_{12}k_{23})M}{n_2}.$$

In exactly the same manner, we may show that

$$\bar{x}_1=\frac{k_{12}k_{23}x_3}{n_2n_3}+\frac{(n_1n_2-k_{12}k_{23})M}{n_2}.$$

We then obtain the coefficient of correlation from the slopes of these lines. It is

$$r_{x_1x_3}=\frac{k_{12}k_{23}}{n_2(n_1n_3)^{1/2}}=r_{x_1x_2}\,r_{x_2x_3}.$$

This completes the proof of the theorem, since

$$r_{x_1x_2}=\frac{k_{12}}{(n_1n_2)^{1/2}}\qquad r_{x_2x_3}=\frac{k_{23}}{(n_2n_3)^{1/2}}.$$

3. The correlation among p sums. We now extend our discussion to p principal variables, forming each successive one

[9] E. Netto, "Lehrbuch der Combinatorik," (1901), pp. 12-13.

in the same manner in which x_2 and x_3 were formed above; that is, x_i, $(i=2,3\cdots\cdots p)$, is equal to the sum of $k_{i-1,\,i}$ random drawings of t from the constituent values of t forming x_{i-1}, plus the sum of $n_i - k_{i-1,\,i}$ independent random drawings of t directly from the universe characterized by $f(t)$. The correlation surface, $w = F(x_1, x_p)$, can at once be written in the same manner as the surface considered in Theorem II. That is, each term of the expression for $F(x_1, x_p)$, multiplied by $dx_1\, dx_p$, consists of an iterated integral which represents the probability, to within infinitesimals of higher order, of the occurrence of a given pair, (x_1, x_p), with a specified number of values of t in common, times a probability coefficient which represents the probability of the occurrence of this specified number of values of t in common. This same method may be employed in writing the correlation surface for any pair of principal variables. The expressions for the probability coefficients, however, become increasingly complex as the number of ways in which the two principal variables can have $0, 1, 2, \cdots$ values of t in common increases.

The following theorem can be proved by mathematical induction. The proof is not difficult, though tedious, and on that account will not be presented here.

Theorem IV. *If $f(t)$ is a probability function of the first kind, and $F(x_1, x_p)$ is the simultaneous law of distribution of x_1 and x_p, then the regression of x_1 on x_p and of x_p on x_1 are linear and are given, respectively, by the following equations:*

$$(3.1)\quad \bar{x}_p = \frac{k_{12}\,k_{23}\cdots\cdot k_{p-1,\,p}}{n_1\,n_2\cdots n_{p-1}}\,x_1 + \frac{n_2\,n_3\cdots n_p - k_{12}\,k_{23}\cdots k_{p-1,\,p}}{n_2\,n_3\cdots n_{p-1}}\,M,$$

$$(3.2)\quad \bar{x}_1 = \frac{k_{12}\,k_{23}\cdots\cdot k_{p-1,\,p}}{n_2\,n_3\cdots n_p}\,x_p + \frac{n_1\,n_2\cdots n_{p-1} - k_{12}\,k_{23}\cdots k_{p-1,\,p}}{n_2\,n_3\cdots\cdot n_{p-1}}\,M.$$

Further, the coefficient of correlation between x_1 and x_p is

$$(3.3)\qquad r_{x_1 x_p} = \frac{k_{12}\,k_{23}\cdots\cdot k_{p-1,\,p}}{n_2\,n_3\cdots n_{p-1}(n_1 n_p)^{\frac{1}{2}}} = r_{x_1 x_2}\cdot r_{x_2 x_3}\cdots r_{x_{p-1} x_p}.$$

II. Sums of elements drawn from a universe characterized by a probability function of the second kind.

4. The correlations between two sums. Let $f(t)$, a probability function of the second kind, characterize the distribution of the variable t. Let the principal variable x_1 be defined as the sum of n_1 independent values of t drawn at random. Further, let the principal variable x_2 be defined as the sum of k_{12} random values of the n_1 values of t composing x_1 and of n_2-k_{12} independent random values of t taken directly from the universe characterized by $f(t)$.

Theorem V. *Given the sums x_1 and x_2 as defined above with k_{12} random elements in common.*

a) *The marginal distributions of x_1 and x_3 are given, respectively, by*

$$(4.11)\quad G_1(x_1) = \int_0^{x_1}\int_0^{x_1-t_{11}}\cdots\int_0^{x_1-t_{11}-\cdots-t_{1,n_1-2}} f(t_{11})\cdots f(t_{1,n_1-1})$$

$$\times f(x_1-t_{11}-\cdots-t_{1,n_1-1})\,dt_{1,n_1-1}\cdots dt_{11},$$

and

$$(4.12)\quad G_2(x_2) = \int_0^{x_2}\int_0^{x_2-t_{11}}\cdots\int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}-\cdots-t_{2,n_2-2}}$$

$$\times f(t_{11})\cdots f(t_{1k_{12}})f(t_{2,k_{12}+1})\cdots f(t_{2,n_2-1})$$

$$\times f(x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}-\cdots-t_{2,n_2-1})\,dt_{2,n_2-1}\cdots dt_{2,k_{12}+1}\,dt_{1k_{12}}\cdots dt_{11}.$$

b) *The correlation surface, $w=F(x_1,x_2)$, which is in two distinct parts joined along the plane $x_1 - x_2 = 0$, is given by*

(4.2a)

$$F_1(x_1,x_2) = \int_0^{x_2}\int_0^{x_2-t_{11}}\cdots\int_0^{x_2-t_{11}-\cdots-t_{1,k_{12}-1}}\int_0^{x_1-t_{11}-\cdots-t_{1,k_{12}-1}-t_{1k_{12}}}$$

$$\int_0^{x_1-t_{11}-\cdots-t_{1,n_1-2}}\int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}}\int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}}$$

$$\int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}-\cdots-t_{2,n_2-2}} \theta(x_1,t_{11},\cdots t_{1,n_1-1})$$

$$\times \phi(x_2, t_{11}, \cdots t_{1k_{12}}, t_{2,k_{12}+1}, \cdots t_{2,n_2-1}) dt_{2,n_2-1} \cdots$$

$$\times\, dt_{2,k_{12}+1}\, dt_{1,n_1-1} \cdots dt_{11},$$

$$(x_2 \leqq x_1 < \infty);$$

(4.2b)

$$F_2(x_1, x_2) = \int_0^{x_1}\int_0^{x_1-t_{11}} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,n_1-2}} \int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}}$$

$$\int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}} \cdots \int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}-\cdots-t_{2,n_2-2}}$$

$$\times\, \theta(x_1, t_{11}, \cdots t_{1,n_1-1})\, \phi(x_2, t_{11}, \cdots t_{1k_{12}}, t_{2,k_{12}+1}, \cdots t_{2,n_2-1})$$

$$\times\, dt_{2,n_2-1} \cdots dt_{2,k_{12}+1}\, dt_{1,n_1-1} \cdots dt_{11},$$

$$(x_1 \leqq x_2 < \infty).$$

c) *The regression curves of x_2 on x_1 and of x_1 on x_2 are linear and are given, respectively, by the following equations:*

(1.31) $$\bar{x}_2 = \frac{k_{12}\, x_1}{n_1} + (n_2 - k_{12})M,$$

and

(1.32) $$\bar{x}_1 = \frac{k_{12}\, x_2}{n_2} + (n_1 - k_{12})M,$$

where $$M = \int_0^{\infty} t\, f(t)\, dt.$$

Hence, the cofficient of correlation between x_1 and x_2 is

$$r_{x_1 x_2} = \frac{k_{12}}{(n_1 n_2)^{\frac{1}{2}}}.$$

Proof. The proof for the marginal distributions of x_1 and of x_2 are given by Craig[10] and need not be repeated here. The expressions for the correlation surface are derived by a simple extension of the same method to two independent variables. The

[10]Allen T. Craig, loc. cit., p. 356.

limits of integration may be easily verified.

As in the proof of Theorem I, the regression of x_2 on x_1 is given by the locus of the ordinate of the centroid of the section of the surface for a given x_1. However, as the surface here is in two distinct, but connected, parts, we have two terms in both numerator and denominator. The expression for $\bar{x}_2$ is

$$
(4.3) \qquad \bar{x}_2 = \frac{\int_0^{x_1} x_2 F_1(x_1, x_2)\,dx_2 + \int_{x_1}^{\infty} x_2 F_2(x_1, x_2)\,dx_2}{\int_0^{x_1} F_1(x_1, x_2)\,dx_2 + \int_{x_1}^{\infty} F_2(x_1, x_2)\,dx_2},
$$

where $F_1(x_1, x_2)$ and $F_2(x_1, x_2)$ are defined by (4.2a) and (4.2b), respectively.

In the paragraphs immediately following, we shall be concerned principally with interchanging the order of integration, with the accompanying changes in the limits. It will be convenient to write the differential immediately following its respective integral sign. Consider the first term of the numerator. Successive interchanging the order of integration between integration with respect to x_2 and with respect to $t_{11}, t_{12}, \dots t_{1k_{12}}$ respectively, and making the appropriate changes in the limits, we get, writing Φ_{x_2} for $\Phi(x_2, t_{11}, \cdots t_{1k_{12}}, t_{2,k_{12}+1}, \cdots t_{2,n_2-1})$

$$
(4.4) \quad \int_0^{x_1} dt_{11} \int_0^{x_1 - t_{11}} dt_{12} \cdots \int_0^{x_1 - t_{11} - t_{12} - \cdots - t_{1,k_{12}-1}} dt_{1k_{12}} \int_{t_{11}+t_{12}+\cdots+t_{1k_{12}}}^{x_1} dx_2
$$

$$
\int_0^{x_1 - t_{11} - \cdots - t_{1k_{12}}} dt_{1,k_{12}+1} \cdots \int_0^{x_1 - t_{11} - \cdots - t_{1,n_1-2}} dt_{1,n_1-1} \int_0^{x_2 - t_{11} - \cdots - t_{1k_{12}}} dt_{2,k_{12}+1}
$$

$$
\int_0^{x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1}} dt_{2,k_{12}+2} \cdots \int_0^{x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1} - \cdots - t_{2,n_2-2}} dt_{2,n_2-1}
$$

$$
\times\, x_2\, \theta_{x_1} \Phi_{x_2}.
$$

Now consider the second term of the numerator of (4.3). As the limits are constants with respect to the variables of integration

$x_2, t_{11}, \ldots . t_{1k_{12}}$, we may interchange the order of integration successively until we have

(4.5)

$$\int_0^{x_1} dt_{11} \int_0^{x_1-t_{11}} dt_{12} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,k_{12}-1}} dt_{1k_{12}} \int_{x_1}^{\infty} dx_2$$
$$\int_0^{x_1-t_{11}-\cdots-t_{1k_{12}}} dt_{1,k_{12}+1} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,n_1-2}} dt_{1,n_1-1}$$
$$\int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}} dt_{2,k_{12}+1} \int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}} dt_{2,k_{12}+2} \cdots$$
$$\int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}-\cdots-t_{2,n_2-2}} dt_{2,n_2-1}\; x_2\, \theta_{x_1}\, \phi_{x_2}\,.$$

We may now combine the first and second terms, (4.4) and (4.5), getting

$$\int_0^{x_1} dt_{11} \int_0^{x_1-t_{11}} dt_{12} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,k_{12}-1}} dt_{1k_{12}} \int_{t_{11}+t_{12}+\cdots+t_{1k_{12}}}^{\infty} dx_2$$
$$\int_0^{x_1-t_{11}-\cdots-t_{1k_{12}}} dt_{1,k_{12}+1} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,n_1-2}} dt_{1,n_1-1}$$
$$\int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}} dt_{2,k_{12}+1} \int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}} dt_{2,k_{12}+2} \cdots$$
$$\int_0^{x_2-t_{11}-\cdots-t_{1k_{12}}-t_{2,k_{12}+1}-\cdots-t_{2,n_2-2}} dt_{2,n_2-1}\; x_2\, \theta_{x_1}\, \phi_{x_2}\,.$$

As the limits of integration are constant with respect to the variables x_2 and $t_{1,k_{12}+1}, \ldots t_{1,n_1-1}$, we may at once interchange successively the orders of integration with respect to x_2 and with respect to $t_{2,k_{12}+1}, t_{2,k_{12}+2}, \ldots\ldots . t_{2,n_2-1}$, respectively, making the proper changes in the limits. We then have

(4.6)

$$\int_0^{x_1} dt_{11} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,n_1-2}} dt_{1,n_1-1} \int_0^{\infty} dt_{2,k_{12}+1} \cdots$$
$$\int_0^{\infty} dt_{2,n_2-1} \int_{t_{11}+\cdots+t_{1k_{12}}+t_{2,k_{12}+1}+\cdots+t_{2,n_2-1}}^{\infty} dx_2\;\; x_2\, \theta_{x_1}\, \phi_{x_2}\,.$$

The denominator of (4.3) may be reduced to this same form except for the absence of the factor x_2 in the integrand.

Let us make the transformation

$$v = x_2 - t_{11} - \cdots - t_{1k_{12}} - t_{2,k_{12}+1} - \cdots - t_{2,n_2-1},$$

as was done in the proof of Theorem I. The limits $t_{11} + \cdots + t_{2n_2-1}$ to ∞ on x_2 now become 0 to ∞ on v. We have now reduced (4.3) to the following form:

(4.7)

$$\bar{x}_2 = \Bigg\{\Bigg[\int_0^{x_1}\cdots\int_0^{\infty} t_{11} + \int_0^{x_1}\cdots\int_0^{\infty} t_{12} + \cdots\cdots$$

$$\int_0^{x_1}\cdots\int_0^{\infty} t_{1k_{12}} + \int_0^{x_1}\cdots\int_0^{\infty} t_{2,k_{12}+1} + \cdots + \int_0^{x_1}\cdots\int_0^{\infty} t_{2,n_2-1}$$

$$+ \int_0^{x_1}\cdots\int_0^{\infty} v\Bigg]\theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v)\, dv\, dt_{2,n_2-1} \cdots$$

$$dt_{2,k_{12}+1}\, dt_{1,n_1-1} \cdots dt_{11}\Bigg\} \Bigg/$$

$$\left\{\int_0^{x_1}\cdots\int_0^{\infty} \theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v)\, dv\, dt_{2,n_2-1} \cdots dt_{2,k_{12}+1}\, dt_{1,n_1-1} \cdots dt_{11}\right\}.$$

The denominator reduces at once to $G(x_1)$ in (4.11). As in the proof of Theorem I directly following equation (1.6), it will be noted that the terms of the numerator fall into two groups: those k_{12} terms containing the factor t_{1i}, $(i = 1, 2, \ldots\ldots\, k_{12})$, and the $n_2 - k_{12}$ terms containing the factor v or t_{2j}, $(j = k_{12}+1, \ldots\ldots n_2-1)$. As the limits of integration with respect to each of these letter variables are 0 and ∞, and since complete interchangeability of the order of integration is then permissible, it is readily seen that any two of these $n_2 - k_{12}$ terms are equivalent. The sum of the entire group, then, may be written

(4.8)

$$(n_2 - k_{12}) \int_0^{x_1} dt_{11} \cdots \int_0^{x_1 - t_{11} - \cdots - t_{1,n_1-2}} dt_{1,n_1-1} \int_0^{\infty} dt_{2,k_{12}+1} \cdots$$

$$\int_0^{\infty} dt_{2,n_2-1} \int_0^{\infty} dv\; v\theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j})\, f(v).$$

In (4.8), it is clear that the integrations with respect to each t_{2j} may be effected immediately by making use of the hypothesis that $\int_0^\infty f(t)\,dt = 1$. This leaves $v f(v)\,\theta_{x_1}$ remaining as the integrand. The $\int_0^\infty v f(v)\,dv$ is a constant which we shall designate by M. Removing this constant from under the integral signs leaves us merely tne expression for the marginal distribution of x_1 times $M(n_2-k_{12})$. We then have

$$(4.9)\quad \bar{x}_2 = (n_2-k_{12})M + \sum_{i=1}^{k_{12}} \frac{\int_0^{x_1} dt_{11} \cdots \int_0^{\infty} dv\, t_{1i}\, \theta_{x_1} \prod_{j=k_{12}+1}^{n_2-1} f(t_{2j}) f(v)}{\int_0^{x_1} dt_{11} \cdots \int_0^{x_1 - t_{11} - \cdots - t_{1,n_1-2}} dt_{1,n_1-1}\, \theta_{x_1}}.$$

That each term in the summation in the right member of (4.9) is equal to any other term in the summation, follows from the complete interchangeability of the order of integration of any two consecutive variables, provided a corresponding interchange between these two variables is likewise carried out in the limits of integration. By successive interchanges of variables we may put the original t_{11} , t_{12} , ... $t_{1k_{12}}$ in any order we choose. Hence, the sum of the last k_{12} terms of (4.9) may be written as k_{12} times any one of them. For definiteness, select the one containing the factor t_{11} in the integrand of the numerator. We may now integrate out all of the t_{2j} , $(j = k_{12}+1, \ldots n_2-1)$ and the v exactly as before. Equation (4.9) then becomes

$$\bar{x}_2 = (n_2-k_{12})M + k_{12} \frac{\int_0^{x_1} dt_{12} \cdots \int_0^{x_1 - t_{12} - \cdots - t_{1,n_1-1}} dt_{11}\, t_{11}\, \theta_{x_1}}{\int_0^{x_1} dt_{12} \cdots \int_0^{x_1 - t_{12} - \cdots - t_{1,n_1-1}} dt_{11}\, \theta_{x_1}},$$

or $\bar{x}_2 = (n_2-k_{12})M + k_{12} I_{n_1}$

It is not difficult to show that $I_{n_1} = \frac{x_1}{n_1}$ Hence, we have

$$\bar{x}_2 = \frac{k_{12}\, x_1}{n_1} + (n_2-k_{12})M.$$

In exactly the same manner, we may show that

$$\bar{x}_1 = \frac{k_{12}\, x_2}{n_2} + (n_1-k_{12})M.$$

The coefficient of correlation between x_1 and x_2 is

$$r_{x_1 x_2} = \frac{k_{12}}{(n_1 n_2)^{1/2}} ,$$

which completes the proof of the theorem.

Illustration. Consider the two sums, $x_1 = t_{11} + t_{12}$, and $x_2 = t_{11} + t_{22}$, with t_{11}, t_{12}, t_{22}, as random drawings of t from the distribution characterized by the function $f(t) = e^{-t}$ for t on the range 0 to ∞. From (4.1i), the marginal distribution of x_1 is

$$G_1(x_1) = x_1 e^{-x_1}.$$

Similarly, the marginal distribution of x_2 is

$$G_2(x_2) = x_2 e^{-x_2}.$$

The correlation surface, obtained by applying (4.2a) and (4.2b), is

$$F_1(x_1, x_2) = e^{-x_1}(1 - e^{-x_2}), \quad (0 \leq x_2 \leq x_1);$$

and

$$F_2(x_1, x_2) = e^{-x_2}(1 - e^{-x_1}), \quad (x_1 \leq x_2 < \infty).$$

5. The correlation among more than two sums. We shall state, without proof, the following theorems.

Theorem VI. *Given a probability function,* $f(t)$, *of the second kind, and three principal variables,* x_1, x_2, x_3, *defined as for Theorem II. Then the correlation surface* $w = F(x_1, x_3)$ *is given by*

(5.1a)

$$F_1(x_1, x_3) = \frac{1}{\binom{n_2}{k_{23}}} \sum_{g=0}^{k_{23}} \binom{k_{12}}{k_{23}-g}\binom{n_2-k_{12}}{g} \int_0^{x_3} dt_{11} \cdots \int_0^{x_3 - t_{11} - \cdots - t_{1,k_{23}-g-1}} dt_{1,k_{23}-g}$$

$$\int_0^{x_1 - t_{11} - \cdots - t_{1,k_{23}-g}} dt_{1,k_{23}-g+1} \cdots \int_0^{x_1 - t_{11} - \cdots - t_{1,n_1-2}} dt_{1,n_1-1}$$

$$\int_0^{x_3 - t_{11} - \cdots - t_{1,k_{23}-g}} dt_{2,k_{23}-g+1} \cdots$$

$$\int_0^{x_3 - t_{11} - \cdots - t_{1,k_{23}-g} - t_{2,k_{23}-g+1} - \cdots - t_{2,k_{23}-1}} dt_{2k_{23}}$$

$$\int_0^{x_3-t_{11}-\cdots-t_{1,k_{23}-g}-t_{2,k_{23}-g+1}-\cdots-t_{2k_{23}}} dt_{3,k_{23}+1}\cdots$$

$$\int_0^{x_3-t_{11}-\cdots-t_{3,n_3-2}} dt_{3,n_3-1}\,\theta_{x_1}\,\phi(x_3, t_{11}, \cdots t_{1,k_{23}-g},$$

$$t_{2,k_{23}-g+1}, \cdots t_{2k_{23}}, t_{3,k_{23}+1}, \cdots t_{3,n_3-1}),$$

$$(x_3 \leqq x_1 < \infty);$$

and

(5.1b)

$$F_2(x_1,x_3) = \frac{1}{\binom{n_2}{k_{23}}} \sum_{g=0}^{k_{23}} \binom{k_{12}}{k_{23}-g} \binom{n_2-k_{12}}{g} \int_0^{x_1} dt_{11} \cdots \int_0^{x_1-t_{11}-\cdots-t_{1,n_1-2}} dt_{1,n_1-1}$$

$$\int_0^{x_3-t_{11}-\cdots-t_{1,k_{23}-g}} dt_{2,k_{23}-g+1}\cdots$$

$$\int_0^{x_3-t_{11}-\cdots-t_{1,k_{23}-g}-t_{2,k_{23}-g+1}-\cdots-t_{2,k_{23}-1}} dt_{2k_{23}}$$

$$\int_0^{x_3-t_{11}-\cdots-t_{2,k_{23}-1}-t_{2,k_{23}}} dt_{3,k_{23}+1}\cdots$$

$$\int_0^{x_3-t_{11}-\cdots-t_{1,k_{23}-g}-t_{2,k_{23}-g+1}-\cdots-t_{2k_{23}}-t_{3,k_{23}+1}-\cdots-t_{3,n_3-2}} dt_{3,n_3-1}$$

$$\theta_{x_1}\,\phi(x_3, t_{11}, \cdots t_{1,k_{23}-g}, t_{2,k_{23}-g+1}, \cdots t_{2k_{23}}, t_{3,k_{23}+1}, \cdots t_{3,n_3-1})$$

$$(x_1 \leqq x_3 < \infty).$$

Theorem VII. *The regression curves of x_3 on x_1 and of x_1 on x_3 of the correlation surface in Theorem VI are linear and are given, respectively, by the following equations:*

(2.21)
$$\bar{x}_3 = \frac{k_{12}k_{23}x_1}{n_1n_2} + \frac{(n_2n_3-k_{12}k_{23})M}{n_2},$$

and

(2.22)
$$\bar{x}_1 = \frac{k_{12}k_{23}x_3}{n_2n_3} + \frac{(n_1n_2-k_{12}k_{23})M}{n_2},$$

where M is defined as in Theorem V. Further, the coefficient of correlation between x_1 and x_3 is

$$(2.3) \qquad r_{x_1x_3} = \frac{k_{12}k_{23}}{n_2(n_1 n_3)^{1/2}} = r_{x_1x_2} r_{x_2x_3}.$$

Theorem VIII. The statement of this theorem differs from that of Theorem IV only in that $f(t)$ is now to be a probability function of the second kind.

III. Sums of elements drawn from a universe characterized by a probability function of the third kind.

6. The correlation between two sums. We shall now consider principal variables defined as the sums of values of t drawn from a universe characterized by $f(t)$, a probability function of the third kind, defined on the range 0 to a, and with

$$\int_0^a f(t)\,dt = 1.$$

The correlation surfaces are not developed with the same degree of generality as were those in the preceding pages because of the tediousness of the labor involved and the complexity of the correlation surface, which may consist of many sections joined together. Thus, if x is the sum of m values of t and y the sum of n, all drawn from a universe characterized by a probability function of the third kind, the correlation surface, $w = F(x,y)$, consists of $2(mn-1)$ sections, each having its own equation. Hence, only the case where x and y each consist of the sum of two values of t, with one of these held in common, will be considered here.

Theorem IX. *Let $f(t)$, a probability function of the third kind, characterize the distribution of a variable t. Let the principal variables x and y be defined by the relations $x = t_{11} + t_{12}$, $y = t_{11} + t_{22}$, where t_{11}, t_{12}, t_{22}, are independent random drawings of t from the universe.*

a.) *The marginal distributions of x and of y are given by*

(6.11)
$$G_1(x) = \int_0^x f(t)f(x-t)dt, \qquad (0 \leqq x \leqq a);$$
$$= \int_{x-a}^a f(t)f(x-t)dt, \qquad (a \leqq x \leqq 2a);$$

and

(6.12)
$$G_2(y) = \int_0^y f(t)f(y-t)dt, \qquad (0 \leqq y \leqq a);$$
$$= \int_{y-a}^a f(t)f(y-t)dt, \qquad (a \leqq y \leqq 2a).$$

b) *The correlation surface,* $w = F(x,y)$ *, is given by*

(6.2)
$$F(x,y) = \int_0^y f(t)f(x-t)f(y-t)dt, \quad (0 \leqq y \leqq x \leqq a);$$
$$= \int_0^x f(t)f(x-t)f(y-t)dt, \quad (0 \leqq x \leqq y \leqq a);$$
$$= \int_{y-a}^x f(t)f(x-t)f(y-t)dt, \quad (a \leqq y \leqq x+a \leqq 2a);$$
$$= \int_{x-a}^y f(t)f(x-t)f(y-t)dt, \quad (0 \leqq x-a \leqq y \leqq a);$$
$$= \int_{x-a}^a f(t)f(x-t)f(y-t)dt, \quad (a \leqq y \leqq x \leqq 2a);$$
$$= \int_{y-a}^a f(t)f(x-t)f(y-t)dt, \quad (a \leqq x \leqq y \leqq 2a).$$

In a) and b) above, the subscripts have been omitted from the t_{ii}

c) *The regression curves of* y *on* x *and of* x *on* y *are linear and are given, respectively, by the following equations:*

(6.31)
$$\bar{y} = \frac{x}{2} + M,$$

(6.32) *and*
$$\bar{x} = \frac{y}{2} + M,$$

where
$$M = \int_0^a t f(t)dt.$$

Hence, the coefficient of correlation between x *and* y *is* $\frac{1}{2}$.

This theorem is a direct generalization of Rietz's paper in Biometrika cited in the introduction to this paper. The proof may be supplied by the reader.

Illustration. Let us consider the rectangular distribution given by $f(t)=\frac{1}{a}$, for t on the range 0 to a, and a to 0. This is the parent distribution in Rietz's case when $a=1$. From (6.11), the marginal distribution of x is

$$G_1(x)=\frac{x}{a^2}, \qquad (0\leqq x\leqq a);$$
$$=\frac{(2a-x)}{a^2}, \qquad (a\leqq x\leqq 2a).$$

Similarly, the marginal distribution of y is

$$G_2(y)=\frac{y}{a^2}, \qquad (0\leqq y\leqq a);$$
$$=\frac{(2a-y)}{a^2}, \qquad (a\leqq y\leqq 2a).$$

The application of (6.2) yields

$$F(x,y)=\frac{y}{a^3}, \qquad (0\leqq y\leqq x\leqq a);$$
$$=\frac{x}{a^3}, \qquad (0\leqq x\leqq y\leqq a);$$
$$=\frac{(x-y+a)}{a^3}, \qquad (a\leqq y\leqq x+a\leqq 2a);$$
$$=\frac{(y-x+a)}{a^3}, \qquad (0\leqq x-a\leqq y\leqq a);$$
$$=\frac{(2a-x)}{a^3}, \qquad (a\leqq y\leqq x\leqq 2a);$$
$$=\frac{(2a-y)}{a^3}, \qquad (a\leqq x\leqq y\leqq 2a).$$

These results, obtained directly by the use of Theorem IX, agree with those obtained by Rietz in the above-mentioned paper.

Carl H. Fischer.