MOMENTS AND DISTRIBUTIONS OF ESTIMATES OF POPULATION PARAMETERS FROM FRAGMENTARY SAMPLES*

S. S. WILKS**

CONTENTS:

---

*Presented to the American Mathematical Society, March 25, 1932.
**National Research Fellow in Mathematics.

I. Introduction.

It frequently happens that all of the individuals of a sample of statistical data from a multivariate population are not observed or classified with respect to all of the variates. If a sample be represented in matrix form by allowing the rows to represent the individuals and the columns to represent the variates, then the matrix of the type of sample with which we are concerned is incomplete in that some of the elements are not present. As an example of a fragmentary sample of this nature, we may consider a series of measurements taken from certain parts of a group of human skeletons from some archeological find, in which some of the parts under consideration are missing from some of the skeletons. Again, we find such a class of samples in the social sciences and government statistics arising from incompletely answered questionnaires.

In dealing with fragmentary samples, it is important to have at hand techniques which will enable the investigator to extract as much information as possible from the data. This is especially true if the data are unique or expensive. An important problem in this connection is that of estimating the population parameters from the sample.

In this paper it is the purpose of the author to investigate incomplete samples from a normal bivariate population. To be more specific, samples are considered from a normal bivariate population of x and y, in which s of the items are observed with respect to x and y, m with respect to x only and n with respect to y only. In the first part of the paper we shall consider various sets of simultaneous maximum likelihood estimates of the population parameters and the limiting forms of their sampling variances and covariances in large samples. In the second part we shall consider other less efficient, but simpler systems of estimates.

II. Simultaneous Estimation by the Method of Maximum Likelihood.

Let a sample ω of N individuals be drawn from the population of the two variates x and y whose distribution is given by

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

where a and b are the means, σ_x and σ_y the standard deviations and r the correlation of x and y in the population. Let ω_{xy} be the set of s individuals of this sample observed with respect to x and y, ω_x the set of m items observed with respect to x only and ω_y the remaining n items observed with respect to y only. To avoid trivial results we shall assume that s is not zero. Furthermore, we shall let ξ and η be the variances and ζ the convariance[1], $\bar{x}$ and $\bar{y}$ the means of x and y in ω_{xy}. The variance and mean of x in ω_x will be denoted by u and $\bar{x}_1$ respectively, and similarly, the variance and mean of y in ω_y will be v and $\bar{y}_1$. The joint distribution of $\bar{x}$, $\bar{y}$, $\bar{x}_1$, $\bar{y}_1$, u, v, ξ, η and ζ can be written from several well known independent distributions as

$$(1)\qquad F=K(\sigma_x)^{-s-m}(\sigma_y)^{-s-n}(1-r^2)^{-\frac{s}{2}}\, e^{-\frac{s}{2(1-r^2)}\left[\frac{\xi+(\bar{x}-a)^2}{\sigma_x^2}+\frac{\eta+(\bar{y}-b)}{\sigma_y^2}-2r\frac{\zeta+(\bar{x}-a)(\bar{y}-b)}{\sigma_x\sigma_y}\right]}\, e^{-\frac{m}{2\sigma_x^2}\left[u+(\bar{x}_1-a)^2\right]-\frac{n}{2\sigma_y^2}\left[v+(\bar{y}_1-b)^2\right]}$$

$$u^{\frac{m-3}{2}}\, v^{\frac{n-3}{2}}\,(\xi\eta-\zeta^2)^{\frac{s-4}{2}},$$

where

$$K=\frac{(m)^{\frac{m}{2}}(n)^{\frac{n}{2}}\, s^{s}\, 2^{-\frac{m+n+2s}{2}}}{(\pi)^{\frac{1}{2}}\,\Gamma\!\left(\frac{m-1}{2}\right)\Gamma\!\left(\frac{n-1}{2}\right)\Gamma\!\left(\frac{s-1}{2}\right)\Gamma\!\left(\frac{s-2}{2}\right)}$$

[1] If, for two sets of variates, $t_{11}, t_{12}, \cdots t_{1n}$; $t_{21}, t_{22}, \cdots t_{2n}$ we let $\bar{t}_j=\frac{1}{n}\sum_{i=1}^{n} t_{ji}$ and $v_{jk}=\frac{1}{n}\sum_{i=1}^{n}(t_{ji}-\bar{t}_j)(t_{ki}-\bar{t}_k)$, $(j,k=1,2)$. then $\bar{t}_1$ and $\bar{t}_2$ are the means, v_{11} and v_{22} are the variances and v_{12} is the covariance of the two sets of t's.

The likelihood of ω when ω is specified in terms of the foregoing statistics is given by (1). We shall use this expression of the likelihood to obtain approximations for the maximum likelihood estimates of the population parameters σ_x, σ_y, r, a and b. Following Fisher[2], we shall take the logarithm of (1) and denote it by L. For convenience, we shall, once for all, set up the following set of first derivatives,

$$(e_1)\ \frac{\partial L}{\partial a} = \frac{s}{\sigma_x}\left[\frac{(\bar{x}-a)}{\sigma_x(1-r^2)} + \frac{\alpha(\bar{x}_1-a)}{\sigma_x} - \frac{r(\bar{y}-b)}{\sigma_y(1-r^2)}\right]$$

$$(e_2)\ \frac{\partial L}{\partial b} = \frac{s}{\sigma_y}\left[\frac{(\bar{y}-b)}{\sigma_y(1-r^2)} + \frac{\beta(\bar{y}_1-b}{\sigma_y} - \frac{r(\bar{x}-a)}{\sigma_x(1-r^2)}\right]$$

$$(2)\ (e_3)\ \frac{\partial L}{\partial \sigma_x} = \frac{s}{\sigma_x}\left[-(1+\alpha) + \frac{1}{1-r^2}\left(\frac{\bar{\xi}}{\sigma_x^2} - \frac{r\bar{\zeta}}{\sigma_x\sigma_y}\right) + \frac{\alpha\bar{u}}{\sigma_x^2}\right]$$

$$(e_4)\ \frac{\partial L}{\partial \sigma_y} = \frac{s}{\sigma_y}\left[-(1+\beta) + \frac{1}{1-r^2}\left(\frac{\bar{\eta}}{\sigma_y^2} - \frac{r\bar{\zeta}}{\sigma_x\sigma_y}\right) + \frac{\beta\bar{v}}{\sigma_y^2}\right]$$

$$(e_5)\ \frac{\partial L}{\partial r} = \frac{s}{1-r^2}\left[r + \frac{\bar{\zeta}}{\sigma_x\sigma_y} - \frac{r}{1-r^2}\left(\frac{\bar{\xi}}{\sigma_x^2} + \frac{\bar{\eta}}{\sigma_y^2} - \frac{2r\bar{\zeta}}{\sigma_x\sigma_y}\right)\right],$$

where $\bar{\xi} = \xi + (\bar{x}-a)^2$, $\bar{\eta} = \eta + (\bar{y}-b)^2$, $\bar{u} = u + (\bar{x}_1-a)^2$

$$\bar{v} = v + (\bar{y}_1-b)^2,\quad \bar{\zeta} = \zeta + (\bar{x}-a)(\bar{y}-b),\quad \alpha = \frac{m}{s},\ \beta = \frac{n}{s}.$$

In order to consider the limiting form of the sampling variances and covariances of the maximum likelihood estimates, we shall need the matrix of mathematical expectations of the second derivatives of L with respect to the five population parameters. This matrix of expected values turns out to be,

[2] R. A. Fisher, The Mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London, Vol. 222 (1922), pp. 309-368.

(3)

	$\frac{\partial L}{\partial z_1}$	$\frac{\partial L}{\partial z_2}$	$\frac{\partial L}{\partial z_3}$	$\frac{\partial L}{\partial z_4}$	$\frac{\partial L}{\partial z_5}$
$\frac{\partial L}{\partial z_1}$	$\frac{s}{\sigma_x^2}\left(\frac{1}{1-r^2}+\alpha\right)$	$-\frac{sr}{\sigma_x\sigma_y(1-r^2)}$	0	0	0
$\frac{\partial L}{\partial z_2}$	$-\frac{sr}{\sigma_x\sigma_y(1-r^2)}$	$\frac{s}{\sigma_y^2}\left(\frac{1}{1-r^2}+\beta\right)$	0	0	0
$\frac{\partial L}{\partial z_3}$	0	0	$\frac{s[2\alpha(1-r^2)+(2-r^2)]}{\sigma_x^2(1-r^2)}$	$-\frac{sr^2}{\sigma_x\sigma_y(1-r^2)}$	$-\frac{sr}{\sigma_x(1-r^2)}$
$\frac{\partial L}{\partial z_4}$	0	0	$-\frac{sr^2}{\sigma_x\sigma_y(1-r^2)}$	$\frac{s[2\beta(1-r^2)+(2-r^2)]}{\sigma_y^2(1-r^2)}$	$-\frac{sr}{\sigma_y(1-r^2)}$
$\frac{\partial L}{\partial z_5}$	0	0	$\frac{sr}{\sigma_x(1-r^2)}$	$-\frac{sr}{\sigma_y(1-r^2)}$	$\frac{s(1+r^2)}{(1-r^2)^2}$

where the entry in the i-th row and j-th column is $-E\left(\frac{\partial^2 L}{\partial z_i\,\partial z_j}\right)$ where z_i is identical with a, b, σ_x, σ_y, and r as i takes the values 1, 2, . . 5 respectively. Again, α and β denote the ratios $\frac{m}{s}$ and $\frac{n}{s}$ which we shall consider constant as $s \to \infty$.

The maximum likelihood estimates of any number of the five population parameters for given values of the remaining parameters are to be found by setting the corresponding first derivatives in (2) equal to zero and solving the resulting equations simultaneously. In most of the cases of practical interest, the solutions must be reached by approximation. In this paper we shall consider the following cases:-

1. Estimation of a and b for given estimates of σ_x, σ_y

and r .

2. Estimation of σ_x and σ_y for given estimates of a , b and r .
3. Estimation of σ_x, σ_y and r for given values of a and b .

Before proceeding with the maximum likelihood estimates of these parameters we shall consider the notion of the efficiency of a set of statistics designed to estimate a set of population parameters.

1. Joint efficiency of a set of estimates.

In order to attach an economic value to a sample and its individuals, Fisher[3] has defined the reciprocal of the variance of a maximum likelihood statistic w of a sample from a univariate population as the amount of information contained in the sample relative to the population value of w . For large samples, in which the distribution of w tends to normality, this quantity is a constant multiple of the number of items in the sample. The amount of information contributed by each member of the sample can be found by dividing by the number in the sample.

We can extend the idea of amount of information relative to a system of population parameters contained in a sample by considering the reciprocal of the determinant of the limiting values, in large samples, of the variances and covariances of the maximum likelihood estimates of this system of parameters. This extended definition also holds for systems of parameters estimated from multivariate populations.

The reason for adopting this determinant as the extension of the idea of the amount of information relative to the set of parameters under consideration, is apparent when we note that the square root of its reciprocal enters as a multiplier in the asymptotic normal distribution of the maximum likelihood estimates of the parameters in the same way that the square root of the reciprocal

[3] R. A. Fisher, Statistical methods for research workers, third edition, Oliver and Boyd (1930) pp. 266-270.

of the sampling variance of the maximum likelihood estimate of a single parameter enters as a multiplier in its asymptotic normal distribution.

Fisher[4] has shown that for large samples, the maximum likelihood estimate of a population parameter is distributed with smaller variance than any other statistic designed to estimate the same parameter. In the case of a set of parameters, the determinant of the matrix of limiting values, in large samples, of the variances and covariances of the maximum likelihood estimates of the parameters is smaller than that for any other estimates of the same set of parameters.

To prove this, let us consider a set $\{p_i\}$, $(i=1,2,\cdots n)$ of population parameters, and let the set $\{t_i\}$ be their maximum likelihood estimates, whose sampling distribution for large samples is

$$\frac{\sqrt{H}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i,j=1}^{n} h_{ij}(t_i-p_i)(t_j-p_j)}$$

where $H = |h_{ij}|$, where $h_{ij} = -E\left(\frac{\partial^2 L}{\partial p_i \partial p_j}\right)$ and L is the logarithm of the likelihood of the sample. H is the reciprocal of the matrix of variances and covariances and covariances of the t's. Let the set $\{u_i\}$ be any set of estimates of $\{p_i\}$ in which at least one u is not a maximum likelihood estimate, and let the asymptotic normal distribution of the u's be.

$$\frac{\sqrt{K}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i,j=1}^{n} k_{ij}(u_i-p_i)(u_j-p_j)}$$

where $K = |k_{ij}|$, which is the reciprocal of the determinant of the matrix of variances and covariances of $\{u_i\}$. Now, our problem is equivalent to that of showing that $H > K$. Suppose there is at least one set of estimates of $\{p_i\}$ containing at least

[4]R. A. Fisher, The theory of statistical estimation, Proceedings of the Cambridge Philosophical Society, Vol. 22 (1925) pp. 700-725.

one estimate which is not a maximum likelihood value, such that the reciprocal of the determinant of its variances and covariances is greater than or equal to H. Let this set be $\{u_i\}$. Then, by hypothesis, $K \geq H$.

Let T be any linear transformation $u_i - p_i = \sum_{\alpha=1}^{n} a_{i\alpha} x_\alpha$ of pure rotation of the axes representing $u_i - p_i$ $(i = 1, 2, \cdots n)$ about the point $p_1, p_2, \cdots p_n$ as origin, which will reduce $\sum_{i,j=1}^{n} k_{ij}(u_i - p_i)(u_j - p_j)$ to a sum of squares, $\sum_{\alpha=1}^{n} \bar{k}_\alpha x_\alpha^2$. Here we have $\bar{k}_\alpha = \sum_{i,j=1}^{n} k_{ij} a_{i\alpha} a_{j\alpha}$ and $x_\alpha = \sum_{i=1}^{n} b_{i\alpha}(u_i - p_i)$, where $b_{i\alpha}$ is the cofactor of $a_{i\alpha}$ in $|a_{ij}|$. Then $\bar{k}_\alpha$ is the reciprocal of the variance of the variable $\bar{u}_\alpha = \sum_{i=1}^{n} b_{i\alpha} u_i$ about its mean value $\bar{p}_\alpha = \sum_{i=1}^{n} b_{i\alpha} p_i$, and $\prod_{\alpha=1}^{n} \bar{k}_\alpha = K$, since the determinant $|a_{ij}|$ of T is unity. But $\bar{u}_\alpha$ is not the maximum likelihood estimate of $\bar{p}_\alpha$, since at least one of the u's is not a maximum likelihood value. As a matter of fact $\sum_{i=1}^{n} b_{i\alpha} t_i = \bar{t}_\alpha$, say, is the maximum likelihood value of $\bar{p}_\alpha$, for

$$\frac{\partial L}{\partial \bar{p}_\alpha} = \sum_{i=1}^{n} \frac{\partial L}{\partial p_i} \frac{\partial p_i}{\partial \bar{p}_\alpha} = \sum_{i=1}^{n} a_{i\alpha} \frac{\partial L}{\partial p_i}$$

vanishes only for $p_i = t_i$, that is, for $\bar{p}_\alpha = \sum_{i=1}^{n} b_{i\alpha} t_i$; (provided we assume that $\frac{\partial L}{\partial p_i} = 0 (i = 1, 2, \cdots n)$ has the unique solution $p_i = t_i$).

It follows from Fisher's[5] proof for the case of one variable, that the reciprocal q_α of the variance of $\bar{t}_\alpha$ is greater than $\bar{k}_\alpha$. Hence, $\prod_{\alpha=1}^{n} q_\alpha > \prod_{\alpha=1}^{n} \bar{k}_\alpha$. We note however, that the maximum likelihood estimates $\{\bar{t}_\alpha\}$ are not independent, for their distribution is

$$\frac{|a_{ij}| \sqrt{H}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{\alpha,\beta=1}^{n} \bar{h}_{\alpha\beta} (\bar{t}_\alpha - \bar{p}_\alpha)(\bar{t}_\beta - \bar{p}_\beta)}$$

where $\bar{h}_{\alpha\beta} = \sum_{i,j=1}^{n} h_{ij} a_{i\alpha} a_{j\beta}$, which is not necessarily zero for $\alpha \neq \beta$. The effect of this non-independence is to introduce a term $\frac{1}{R}$ as a multiplier of $\prod_{\alpha=1}^{n} q_\alpha$, where R is the determinant of correlations among the $\bar{t}$'s, and is less than unity. Hence, $\prod_{\alpha=1}^{n} \frac{q_\alpha}{R} > \prod_{\alpha=1}^{n} q_\alpha >$

[5] R. A. Fisher, loc. cit.

$\prod_{\alpha=1}^{n} \bar{k}_\alpha = K$. It is well known from the theory of quadratic forms that the matrix $\|\bar{h}_{\alpha\beta}\|$ is found as the product $\|\bar{a}_{ij}\| \cdot \| h_{ij}\| \cdot \|a_{ij}\|$, where $\|\bar{a}_{ij}\|$ is $\|a_{ij}\|$ with its rows and columns interchanged. Since the determinant $|a_{ij}|$ is unity, it is clear that $|\bar{h}_{\alpha\beta}|$, which is equal to $\prod_{\alpha=1}^{n} \frac{q_\alpha}{\bar{k}}$, has the value $|h_{ij}|$ which is H by definition. Therefore we have $H>K$, which contradicts the hypothesis that $K \geq H$. Hence, we must have $K<H$.

Thus, the proposition is proved that the reciprocal of the determinant of variances and covariances of the maximum likelihood estimates $\{t_i\}$ is smaller than that of any other set of estimates, all of which are not likelihood values.

We are now provided with a means of measuring the joint efficiency of a set of estimates in utilizing information in the sample relevant to the population parameters estimated by the set. We shall take as a measure of this efficiency the ratio of the reciprocal of the determinant of its variances and covariances to that of the set of maximum likelihood estimates of the same parameters. This quantity is less than unity, as we have just proved. The efficiency of $\{u_i\}$ is therefore

(4)
$$Eff = \frac{K}{H}.$$

2. Simultaneous estimation of a and b.

We shall suppose that satisfactory estimates have been obtained for σ_x, σ_y and r. If they are to be taken from the sample ω, we can take σ_x^2 as the variance of the x's in ω_{xy} and ω_x, σ_y^2 as that of the y's in ω_{xy} and ω_y and r from ω_{xy}. In any case our problem is to find the optimum values of a and b for given values of σ_x, σ_y and r. These values of a and b are found as the solution of the equations obtained by setting (e_1) and (e_2) in (2) equal to zero. Accordingly, we find,

$$\hat{a} = \frac{1}{\Delta\sigma_y}\left[\frac{(1+\beta)\bar{x}}{\sigma_x(1-r^2)} + \frac{\alpha\bar{x}_1}{\sigma_x}\left(\frac{1}{1-r^2}+\beta\right) + \frac{\beta r}{\sigma_x(1-r^2)}(\bar{y}_1-\bar{y})\right]$$
(5)
$$\hat{b} = \frac{1}{\Delta\sigma_x}\left[\frac{(1+\alpha)\bar{y}}{\sigma_y(1-r^2)} + \frac{\beta\bar{y}_1}{\sigma_y}\left(\frac{1}{1-r^2}+\alpha\right) + \frac{\alpha r}{\sigma_y(1-r^2)}(\bar{x}_1-\bar{x})\right]$$

where $\Delta = \frac{1}{\sigma_x\sigma_y(1-r^2)}\left[1+\alpha+\beta+\alpha\beta(1-r^2)\right]$.

The matrix of the variances and covariances of $\hat{a}$ and $\hat{b}$ in samples is obtained by taking the reciprocal form of the two way principal minor in the upper left corner of the matrix (2)[6]. Thus we find,

(6)
$$\left\| \begin{matrix} \frac{\sigma_x^2[1+\beta(1-r^2)]}{sD} & \frac{r\sigma_x\sigma_y}{sD} \\ \frac{r\sigma_x\sigma_y}{sD} & \frac{\sigma_y^2[1+\alpha(1-r^2)]}{sD} \end{matrix} \right\|$$

where $D = 1+\alpha+\beta+\alpha\beta(1-r^2)$

We note from (6) that the variance of $\hat{a}$ is

$$\sigma_{\hat{a}}^2 = \frac{\sigma_x^2[1+\beta(1-r^2)]}{s[1+\alpha+\beta+\alpha\beta(1-r^2)]},$$

and a similar expression holds for $\sigma_{\hat{b}}^2$. The correlation coefficient of $\hat{a}$ and $\hat{b}$ is

$$r_{\hat{a}\hat{b}} = \frac{r}{\{[1+\beta(1-r^2)][1+\alpha(1-r^2)]\}^{\frac{1}{2}}}$$

[6]See Karl Pearson, On the influence of natural selection on the variability and correlation of organs, Philosophical Transactions of the Royal Society of London, series A, vol. 200 (1900), pp. 3-10. Here Pearson gives a method of obtaining the variances and co-variances of the variates in a normal multivariate probability function.

From the definitions in section 1, we find that the amount of information in ω relative to a and b is the reciprocal of the determinant of (5). That is,

$$(7) \qquad A(m,n,s) = \frac{s^2 + s(m+n) + mn(1-r^2)}{\sigma_x^2 \sigma_y^2 (1-r^2)} \; .$$

From (7) we can find the relative amounts of information contributed by members of ω_{xy}, ω_x and ω_y by means of differences. For given values of n and s, we have as the information contributed to ω by an $m+1st$ individual of ω_x,

$$(7a) \quad A_{\omega_x}(m+1) = A(m+1,n,s) - A(m,n,s) = \frac{s + n(1-r^2)}{\sigma_x^2 \sigma_y^2 (1-r^2)} ,$$

which is independent of m. A similar expression holds for the $n+1st$ member of ω_y.

For the $s+1st$ member of ω_{xy}, for given values of m and n, we get

$$(7b) \qquad A_{\omega_{xy}}(s+1) = \frac{m+n+2s+1}{\sigma_x^2 \sigma_y^2 (1-r^2)} \; .$$

It is clear that an additional member to ω_{xy} is more informative than one to each of ω_x and ω_y by an amount $\frac{(m+n+1)r^2}{\sigma_x^2 \sigma_y^2 (1-r^2)}$, or, considering the ratio rather than the difference, we have

$$(7c) \quad \frac{A(m+1,\, n+1,\, s) - A(m,n,s)}{A_{\omega_{xy}}(s+1)} = 1 - \frac{r^2(m+n+1)}{2s + m+n+1} \; .$$

We find that the amount of information introduced by ω_x and ω_y is $A(m,n,s) - A(0,0,s)$, which is $\frac{s(m+n) + mn(1-r^2)}{\sigma_x^2 \sigma_y^2 (1-r^2)}$ and its ratio to the total information (7) is

$$1 - \frac{s^2}{s^2 + s(m+n) + mn(1-r^2)} \; .$$

3. Simultaneous estimation of σ_x and σ_y.

If we suppose that r is given as well as a and b, we can find the optimum value of σ_x^2 and σ_y^2 by solving the equations obtained by setting (e_3) and (e_4) in (2) equal to zero. Accordingly, we get,

$$(8)\quad \hat{\sigma}_x^2 = \frac{2E(EF-G^2)}{2EF(1+\alpha)-G^2(\alpha-\beta)+\sqrt{4EFG^2(1+\alpha)(1+\beta)+G^4(\alpha-\beta)^2}}$$

$$\hat{\sigma}_y^2 = \frac{}{2EF(1+\beta)-G^2(\beta-\alpha)+\sqrt{4EFG^2(1+\alpha)(1+\beta)+G^4(\alpha-\beta)^2}}$$

where $E = \alpha\bar{u} + \frac{\bar{\xi}}{1-r^2}$, $F = \beta\bar{v} + \frac{\bar{\eta}}{1-r^2}$ and $G = \frac{r\bar{\zeta}}{1-r^2}$.

The variances and covariances of $\hat{\sigma}_x$ and $\hat{\sigma}_y$ are given by the reciprocal form of the matrix obtained by striking out the last row and column from the third order principal minor in the lower right corner of (3). For the variance of $\hat{\sigma}_x$, we find,

$$\sigma_{\hat{\sigma}_x}^2 = \frac{\sigma_x^2[2(1+\beta)-r^2(2\beta+1)]}{2s[2(1+\alpha)(1+\beta)-r^2(\alpha+\beta+\alpha\beta)]}.$$

A similar expression exists for $\sigma_{\hat{\sigma}_y}^2$. The amount of information yielded by ω relative to σ_x and σ_y under these conditions is

$$(9)\quad A'(m,n,s) = \frac{4(m+s)(n+s)-2r^2[sm+sn+2mn]}{\sigma_x^2\sigma_y^2(1-r^2)}.$$

From (9), we find the differences corresponding to (7,a,b,c) to be

$$(9a)\quad A'_{\omega_x}(m+1) = \frac{4(s+n)-2r^2(s+2n)}{\sigma_x^2\sigma_y^2(1-r^2)},$$

$$(9b)\quad A'_{\omega_{xy}}(s+1) = \frac{8s+4(m+n+1)-2r^2(m+n)}{\sigma_x^2\sigma_y^2(1-r^2)},$$

$$(9c)\quad \frac{A'(m+1,n+1,s)-A(m,n,s)}{A'_{\omega_{xy}}(s+1)} = 1-\frac{r^2(m+n+2s+2)}{4s+2(m+n+1)-r^2(m+n)}.$$

4. Simultaneous estimation of σ_x, σ_y and r.

Let us suppose a and b to be satisfactorily estimated. For large samples, a and b can be estimated from the sets of x's and y's obtained by pooling ω_{xy}, ω_x and ω_y. Whatever estimates we may choose for a and b, our problem is to solve the equations obtained by setting (e_3), (e_4) and (e_5) in (2) equal to zero, for σ_x, σ_y and r.

If we denote the quantities in the brackets of (e_3), (e_4) and (e_5) by f, g and h respectively, then we are to solve the equations $f = g = h = 0$ for σ_x, σ_y and r. The method of elimination seems to be of little value in solving these equations. Then we shall use the extended form of Newton's approximation method and find an approximate solution. Considering nothing higher than the first order terms of Taylor's expansion of f, g and h we have (letting $\sigma_x = x$, $\sigma_y = y$, $r = z$),

$$(10)\quad \left\{\begin{array}{l} f_1 + (x - x_1) f_{x_1} + (y - y_1) f_{y_1} + (z - z_1) f_{z_1} = 0 \\ g_1 + (x - x_1) g_{x_1} + (y - y_1) g_{y_1} + (z - z_1) g_{z_1} = 0 \\ h_1 + (x - x_1) h_{x_1} + (y - y_1) h_{y_1} + (z - z_1) h_{z_1} = 0 \end{array}\right\}$$

where $f_1 = f(x_1, y_1, z_1)$, $f_{x_1} = \dfrac{\partial f(x_1, y_1, z_1)}{\partial x}$ and so on.

We shall take for the initial point, $x_1 = \sqrt{\xi}$, $y_1 = \sqrt{\eta}$ and $z_1 = \dfrac{\zeta}{\sqrt{\xi\eta}}$, which are, for $m = n = 0$ and $a = \bar{x}$, $b = \bar{y}$, the maximum likelihood estimates of σ_x, σ_y and r from ω.

Solving equations (10) for x, y and z by Cramer's rule we find for the first approximation beyond the initial values,

$$(11)\quad \begin{aligned} \tilde{\sigma}_x &= \sqrt{\xi}\left[1 + \frac{\alpha\left(\frac{\bar{u} - \bar{\xi}}{\bar{\xi}}\right)\left(1 + \frac{\beta\bar{v}}{\bar{\eta}}(1 - \rho^4)\right) + \rho^2\beta\left(\frac{\bar{v} - \bar{\eta}}{\bar{\eta}}\right)}{2D}\right] \\ \tilde{\sigma}_y &= \sqrt{\eta}\left[1 + \frac{\beta\left(\frac{\bar{v} - \bar{\eta}}{\bar{\eta}}\right)\left(1 + \frac{\alpha\bar{u}}{\bar{\eta}}(1 - \rho^4)\right) + \rho^2\alpha\left(\frac{\bar{u} - \bar{\xi}}{\bar{\xi}}\right)}{2D}\right] \\ \tilde{r} &= \rho\left[1 + (1 - \rho^2)^2\, \frac{\alpha\left(\frac{\bar{u} - \bar{\xi}}{\bar{\xi}}\right)\left(\frac{1}{1 - \rho^2} + \frac{\beta\bar{v}}{\bar{\eta}}\right) + \beta\left(\frac{\bar{v} - \bar{\eta}}{\bar{\eta}}\right)\left(\frac{1}{1 - \rho^2} + \frac{\alpha\bar{u}}{\bar{\xi}}\right)}{2D}\right] \end{aligned}$$

where $D = 1 + \left(\frac{\beta\bar{v}}{\bar{\eta}} + \frac{\alpha\bar{u}}{\bar{\xi}}\right) + \frac{\alpha\beta\bar{u}\bar{v}}{\bar{\xi}\bar{\eta}}(1-\rho^4)$ and $\rho = \frac{\bar{s}}{\sqrt{\bar{\xi}\bar{\eta}}}$. By using the point whose coordinates are given by (11) in place of the initial point in (10), we find a second approximation point, and continuing the process we get a sequence of points. Such a sequence would raise questions of convergence which will not be considered in this paper. However, it can be shown without much difficulty that the likelihood of the point whose coordinates are given by (11) is greater than that of the initial point for variations of $\bar{u}$ and $\bar{v}$ about $\bar{\xi}$ and $\bar{\eta}$ respectively, and for $\bar{u} = \bar{\xi}$ and $\bar{v} = \bar{\eta}$ the likelihoods are equal. Indeed the problem is equivalent to showing that the ratio of the likelihood (1) with the values $\sqrt{\bar{\xi}}$, $\sqrt{\bar{\eta}}$ and $\frac{\bar{s}}{\sqrt{\bar{\xi}\bar{\eta}}}$ for σ_x, σ_y and r to the likelihood with the values given by (11) for σ_x, σ_y and r has a maximum of unity for variations of $\bar{u}$ and $\bar{v}$ about $\bar{\xi}$ and $\bar{\eta}$. This can be readily done by examining, in the ordinary manner for maxima and minima, the first and second derivatives with respect to $\bar{u}$ and $\bar{v}$ of the ratio of these likelihoods.

The matrix of limiting values of the sampling variances and covariances of the maximum likelihood estimates of σ_x, σ_y and r can be obtained by taking the reciprocal form of the third order principal minor in the lower right hand corner of (3). This reciprocal matrix is,

(12)

$$\left\| \begin{matrix} \frac{\sigma_x^2(1+\beta(1-r^4))}{2sE} & \frac{r^2\sigma_x\sigma_y(1-r^2)}{2sE} & \frac{r\sigma_x(1-r^2)(1+\beta(1-r^2))}{2sE} \\ \frac{r^2\sigma_x\sigma_y(1-r^2)}{2sE} & \frac{\sigma_y^2(1+\alpha(1-r^4))}{2sE} & \frac{r\sigma_y(1-r^2)(1+\alpha(1-r^2))}{2sE} \\ \frac{r\sigma_x(1-r^2)(1+\beta(1-r^2))}{2sE} & \frac{r\sigma_y(1-r^2)(1+\alpha(1-r^2))}{2sE} & \frac{(1-r^2)\left[1+(\alpha+\beta)(1-\frac{r^2}{2})+\alpha\beta(1-r^4)\right]}{2sE} \end{matrix} \right\|$$

where $E = 1 + \alpha + \beta + \alpha\beta(1-r^4)$.

The amount of information in ω relative to σ_x, σ_y and r is the reciprocal of the determinant of (12). Denoting this quantity by $B(m, n, s)$, we have,

$$(13)\quad B(m,n,s)=\frac{4}{\sigma_x^2\sigma_y^2(1-r^2)^3}\left[s^3+s^2(m+n)+mns(1-r^4)\right].$$

Proceeding as we did with (7a), (7b) and (7c), we find the following incremental contributions:

$$(13a)\quad B_{\omega_x}(m+1)=\frac{4}{\sigma_x^2\sigma_y^2(1-r^2)^3}\left[s^2+sn(1-r^4)\right]$$

$$(13b)\quad B_{\omega_{xy}}(s+1)=\frac{4}{\sigma_x^2\sigma_y^2(1-r^2)^3}\left[3s^2+3s+1+(2s+1)(m+n)+mn(1-r^4)\right].$$

$$(13c)\quad B(m+1,n+1,s)-B(m,n,s)=\frac{4}{\sigma_x^2\sigma_y^2(1-r^2)^3}\left[2s^2+s(m+n+1)(1-r^4)\right].$$

$$(13d)\quad B(m,n,s)-B(0,0,s)=\frac{4}{\sigma_x^2\sigma_y^2(1-r^2)^3}\left[s^2(m+n)+mns(1-r^4)\right].$$

We note that the $s+1st$ member of ω_{xy} is much more important than an additional item to each of ω_x and ω_y when σ_x, σ_y and r are considered, than when a and b are considered. The amount of information contributed relative to r by ω_x and ω_y can be found by differencing the reciprocal of the element in the lower right corner of (12), with respect to m and n. If we call this reciprocal $K_{re}(m,n,s)$, we have as the ratio of the contribution of information by ω_x and ω_y, to the total information in ω regarding r,

$$\frac{K_{re}(m,n,s)-K_{re}(0,0,s)}{K_{re}(m,n,s)}=\frac{\frac{r^2}{2}s(m+n)+rmn(1-r^2)}{s^2+s(m+n)+mn(1-r^4)}.$$

In a similar manner, we can find the contributions of the various parts ω_{xy}, ω_x and ω_y to the information relative to any one of the parameters σ_x, σ_y and r and we can find their effects upon the covariances of the maximum likelihood estimates by considering the non-diagonal elements of (12). We find that the information afforded by ω_y relative to σ_x expressed in terms of the total amount of information in ω_{xy}, ω_x and ω_y regarding σ_x is

$$\frac{r^4 sn}{s^2+s(m+n)+mn(1-r^4)}.$$

We remark without going further, that, by considering the five equations obtained by equating each of the expressions in (2) to zero, we can find approximations for the maximum likelihood estimates of a, b, σ_x, σ_y and r by the foregoing method. Since the process is straightforward, though somewhat cumbersome in that it involves fifth order determinants, we shall not consider it here.

III. Systems of independent estimates.

We have seen that the problem of finding the maximum likelihood estimates of a, b, σ_x, σ_y and r from the sample leads to expressions which are not very simple, especially from the point of view of practical application. However, the variances and covariances of these estimates were found to be relatively simple. In view of the difficulties connected with the foregoing maximum likelihood estimates, we shall devote the remainder of this paper to a consideration of the moments, distributions and efficiencies of simpler systems of estimates.

If we are interested in the means of the x's in ω apart from any contribution of the y's, the optimum value of a is $\bar{x}_0 = \frac{\bar{x}+\alpha\bar{x}_1}{1+\alpha}$. Similarly, for the means of the y's, we have $\bar{y}_0 = \frac{\bar{y}+\beta\bar{y}_1}{1+\beta}$. The best estimates of the variances σ_x^2, and σ_y^2 under these conditions are,

$$\xi_0 = \frac{1}{N_1}\left[s\xi + mu + s(\bar{x}-\bar{x}_0)^2 + m(\bar{x}_1-\bar{x}_0)^2\right]$$

$$\eta_0 = \frac{1}{N_2}\left[s\eta + nv + s(\bar{y}-\bar{y}_0)^2 + n(\bar{y}_1-\bar{y}_0)^2\right],$$

where $N_1 = s+m$, and $N_2 = s+n$.

For the covariance we shall take the product moment of the deviations of the x's and y's in ω_{xy} from $\bar{x}_0$ and $\bar{y}_0$ respectively. That is,

$$\zeta_0 = \frac{1}{s}\sum_{i=1}^{s}(x_i-\bar{x}_0)(y_i-\bar{y}_0) = \zeta + (\bar{x}-\bar{x}_0)(\bar{y}-\bar{y}_0).$$

From these values, we can take as the estimate of the correlation coefficient,

$$r_0 = \frac{\zeta_0}{\sqrt{\xi_0\eta_0}}.$$

1. Distribution of $\bar{x}_0$ and $\bar{y}_0$

The variances and correlation of $\bar{x}_0$ and $\bar{y}_0$ can be found from

$$\frac{\sqrt{mns}}{(2\pi)^2\sigma_x^2\sigma_y^2\sqrt{1-r^2}}\, e^{-\frac{s}{2(1-r^2)}\left[\frac{(\bar{x}-a)^2}{\sigma_x^2}+\frac{(\bar{y}-b)^2}{\sigma_y^2}-\frac{2r(\bar{x}-a)(\bar{y}-b)}{\sigma_x\sigma_y}\right]-\frac{m}{2\sigma_x^2}(\bar{x}_1-a)^2-\frac{n}{2\sigma_y^2}(\bar{y}_1-b)^2}$$

by making the substitution $\bar{x}_0 = \frac{\bar{x}+\alpha\bar{x}_1}{1+\alpha}$, $\bar{y}_0 = \frac{\bar{y}+\beta\bar{y}_1}{1+\beta}$, and using determinant analysis[7] on the symmetric matrix of the resulting quadratic exponential.

The variances of $\bar{x}_0$ and $\bar{y}_0$ are found to be $\frac{\sigma_x^2}{N_1}$ and $\frac{\sigma_y^2}{N_2}$ respectively, and the correlation between $\bar{x}_0$ and $\bar{y}_0$ is $\frac{rs}{\sqrt{N_1N_2}}$. The exact distribution of $\bar{x}_0$ and $\bar{y}_0$ is normal.

The amount of information relative to a and b furnished by $\bar{x}_0$ and $\bar{y}_0$ is, according to our definition,

$$\frac{(s+m)(s+n)}{\sigma_x^2\sigma_y^2\left[1-\frac{r^2s^2}{(m+s)(n+s)}\right]}. \tag{14}$$

The efficiency of x_0 and y_0 is, therefore, the ratio of (14) to (7), which is

$$\frac{(m+s)^2(n+s)^2(1-r^2)}{[(m+s)(n+s)-mnr^2][(m+s)(n+s)-s^2r^2]}.$$

[7]Karl Pearson, loc. cit.

2. Characteristic function of ξ_0, η_0 and ζ_0.

The characteristic function or generating function of the moments of ξ_0, η_0 and ζ_0, which we shall denote by $\phi(\gamma,\delta,\varepsilon)$, is defined as the mathematical expectation of $e^{\gamma\xi_0+\delta\eta_0+\varepsilon\zeta_0}$. Since ξ_0, η_0 and ζ_0 are expressible in terms of $\bar{x}$, $\bar{y}$, $\bar{x}_1$, $\bar{y}_1$, u, v, ξ, η and ζ, whose distribution is given by (1), then clearly, we can write,

$$(15) \qquad \phi(\gamma,\delta,\varepsilon) = \int e^{\gamma\xi_0+\delta\eta_0+\varepsilon\zeta_0}\, F\, dV,$$

Where F is given by (1) and dV is the product of the differentials of the variables in F, and the integration in taken over all possible values of the variables.

The integral (15) can be broken into the product of a constant by a quadruple integral, a triple integral and two single integrals. The quadruple integral is of the form

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\sum\limits_{i,j=1}^{4} b_{ij}(t_i-c_i)(t_j-c_j)}\, dt_1\, dt_2\, dt_3\, dt_4$$

which has the value[8] $\dfrac{\pi^2}{\sqrt{\Delta}}$, where Δ is the determinant $|b_{ij}|$ $(i,j=1,2,3,4)$ and $b_{ij}=b_{ji}$. The triple integral is of the form

$$\int_0^{\infty}\int_0^{\infty}\int_{-\sqrt{xy}}^{\sqrt{xy}} e^{-(c_{11}x+c_{22}y+c_{12}z)}\,(xy-z^2)^{\frac{s-4}{2}}\, dz\, dx\, dy$$

which has the value[9]

$$\frac{\sqrt{\pi}\,\Gamma\left(\frac{s-1}{2}\right)\Gamma\left(\frac{s-2}{2}\right)}{\begin{vmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{vmatrix}^{\frac{s-1}{2}}}$$

[8]Karl Pearson, loc. cit.

[9]See V. Romanovsky, On the moments of the standard deviations and of the correlation coefficient in samples from a normal population, Metron, vol. 5, no. 4 (1925) pp. 3-46.

Each of the single integrals is of the well known form

$$\int_0^\infty t^{K-1} e^{-c^2 t}\, dt$$

which has the value $\frac{\Gamma(K)}{C^{2K}}$

Using the above results for the integrals into which (15) resolves itself, we get,

(16)
$$\phi(\gamma,\delta,\epsilon) = A^{\frac{m}{2}} B^{\frac{n}{2}} (\bar{A}\bar{B} - C^2)^{\frac{s-1}{2}} \left(A - \frac{\gamma}{N_1}\right)^{-\frac{m-1}{2}} \left(B - \frac{\delta}{N_2}\right)^{-\frac{n-1}{2}}$$

$$\times \left[\left(\bar{A} - \frac{\gamma}{N_1}\right)\left(\bar{B} - \frac{\delta}{N_2}\right) - \left(C + \frac{\epsilon}{2s}\right)^2\right]^{-\frac{s-1}{2}} \left[\left(A - \frac{\gamma}{N_1}\right)\left(B - \frac{\delta}{N_2}\right)\right.$$

$$\left. + \frac{\epsilon^2}{4s^2}\left(\frac{r^2 m^2 n^2}{N_1^2 N_2^2} - \frac{mn}{N_1 N_2}\right) - \frac{\epsilon\, r m n}{2 s \sigma_x \sigma_y N_1 N_2} - \frac{\gamma\delta\, m n r^2}{N_1^2 N_2^2}\right]^{-\frac{1}{2}},$$

where $A = \frac{1}{2\sigma_x^2}$, $B = \frac{1}{2\sigma_y^2}$, $\bar{A} = \frac{1}{2\sigma_x^2(1-r^2)}$, $\bar{B} = \frac{1}{2\sigma_y^2(1-r^2)}$.

$$C = \frac{r}{2\sigma_x\sigma_y(1-r^2)}.$$

If we write $M(h,k,l) = \frac{\partial^h}{\partial\gamma^h}\frac{\partial^k}{\partial\delta^k}\frac{\partial^l}{\partial\epsilon^l}\,\phi(\gamma,\delta,\epsilon)\Big|_{\gamma=\delta=\epsilon=0}$

we find the following expressions for the first few moments of ξ_0, η_0 and ζ_0,

$$M(1,0,0) = \frac{N_1-1}{N_1}\sigma_x^2, \quad M(0,1,0) = \frac{N_2-1}{N_2}\sigma_y^2$$

$$M(0,0,1) = r\sigma_x\sigma_y\left(\frac{s-1}{s} + \frac{mn}{sN_1N_2}\right)$$

$$M(1,1,0) = \frac{\sigma_x^2\sigma_y^2}{N_1N_2}\left[(N_1-1)(N_2-1) + 2r^2\left(\frac{mn}{N_1N_2} + s - 1\right)\right]$$

(17) $$M(1,0,1) = \frac{r\sigma_x^3\sigma_y}{N_1 s}(N_1+1)\left(\frac{mn}{N_1N_2} + s - 1\right)$$

$$M(0,1,1)=\frac{r\sigma_x\sigma_y^3}{N_2 s}(N_2+1)\left(\frac{mn}{N_1N_2}+s-1\right)$$

$$M(2,0,0)=\frac{\sigma_x^4(N_1^2-1)}{N_1^2},\quad M(0,2,0)=\frac{\sigma_y^4(N_2^2-1)}{N_2^2},$$

$$M(0,0,2)=\frac{\sigma_x^2\sigma_y^2}{s^2}\left[(1+r^2)(s-1)+r^2\frac{m^2n^2}{N_1^2N_2^2}+\frac{mn}{N_1N_2}+r^2\left(\frac{mn}{N_1N_2}+s-1\right)^2\right]$$

$$M(1,1,1)=\frac{r\sigma_x^3\sigma_y^3}{N_1N_2}\left[\frac{4(s-1)(1+r^2)}{s}+\frac{4mn}{N_1N_2}\left(r^2\frac{mn}{N_1N_2}+1\right)\right.$$

$$\left.+\frac{2r^2}{s}\left(\frac{mn}{N_1N_2}+s-1\right)^2+\left(\frac{mn}{sN_1N_2}+\frac{s-1}{s}\right)\left(N_1N_2+N_1+N_2-3\right)\right].$$

If the sample ω is fairly large, we can neglect the contributions of the means $\bar{x}$, $\bar{y}$, $\bar{x}_1$, and $\bar{y}_1$ to ξ_0, η_0, ζ_0 and consider as satisfactory estimates of the variances and covariance,

$$\bar{\xi}_0=\frac{\xi+\alpha u}{1+\alpha},\qquad \bar{\eta}_0=\frac{\eta+\beta v}{1+\beta},\qquad \bar{\zeta}_0=\zeta\ .$$

3. Characteristic function and sampling distribution of $\bar{\xi}_0$, $\bar{\eta}_0$ and $\bar{\zeta}_0$.

It is clear that $\bar{\xi}_0$, $\bar{\eta}_0$ and $\bar{\zeta}_0$ are obtained from ξ_0, η_0 and ζ_0 by dropping the terms involving the means $\bar{x}$, $\bar{y}$, $\bar{x}_1$ and $\bar{y}_1$. The characteristic function $\bar{\phi}(\gamma,\delta,\epsilon)$ of these statistics can be obtained from (15) by replacing ξ_0, η_0 and ζ_0 by $\bar{\xi}_0$, $\bar{\eta}_0$ and $\bar{\zeta}_0$ and integrating. The integral in this case will not involve the quadruple integral, but only the triple integral and the two single integrals. Accordingly, we find

$$\bar{\phi}(\gamma,\delta,\epsilon)=A^{\frac{m-1}{2}}B^{\frac{n-1}{2}}(\bar{A}\bar{B}-C^2)^{\frac{s-1}{2}}\left(A-\frac{\gamma}{N_1}\right)^{-\frac{m-1}{2}}$$

$$\times\left(B-\frac{\delta}{N_2}\right)^{-\frac{n-1}{2}}\left[\left(\bar{A}-\frac{\gamma}{N_1}\right)\left(\bar{B}-\frac{\delta}{N_2}\right)-\left(C+\frac{\epsilon}{2s}\right)^2\right]^{-\frac{s-1}{2}}$$

which is somewhat simpler than (16).

The first few moments of $\bar{\xi}_0$, $\bar{\eta}_0$ and $\bar{\zeta}_0$ evaluated from $\bar{\phi}(\gamma,\delta,\epsilon)$ are (using the notation of (17),

$$M(1,0,0)=\frac{\sigma_x^2(N_1-2)}{N_1},\quad M(0,1,0)=\frac{\sigma_y^2(N_2-2)}{N_2},$$

$$M(0,0,1)=\frac{s-1}{s}r\sigma_x\sigma_y$$

$$M(1,1,0)=\frac{\sigma_x^2\sigma_y^2}{N_1N_2}\left[2r^2(s-1)+(N_1-2)(N_2-2)\right]$$

$$M(1,0,1)=\frac{r\sigma_x^3\sigma_y(s-1)}{s},\quad M(0,1,1)=\frac{r\sigma_x\sigma_y^3(s-1)}{s}$$

$$(18)\quad M(2,0,0)=\frac{\sigma_x^4(N_1-2)}{N_1},\qquad M(0,2,0)=\frac{\sigma_y^4(N_2-2)}{N_2}$$

$$M(0,0,2)=\frac{\sigma_x^2\sigma_y^2(s-1)(1+r^2s)}{s},$$

$$M(1,1,1)=\frac{(s-1)r\sigma_x^3\sigma_y^3}{N_1N_2s}\left[2(N_1+N_2-2)+2(s+1)r^2+(N_1-2)(N_2-2)\right].$$

In order to find the exact sampling distribution of $\bar{\xi}_0$, $\bar{\eta}_0$ and $\bar{\zeta}_0$, it is more convenient to consider the statistics $\xi_1=\frac{N_1}{s}\bar{\xi}_0$, $\eta_0=\frac{N_2}{s}\bar{\eta}_0$ and $\zeta_1=\bar{\zeta}_0$. The characteristic function of these statistics is found from $\bar{\varphi}(\gamma,\delta,\varepsilon)$ by replacing $\frac{N_1}{s}\gamma$ by γ_1, $\frac{N_2}{s}\delta$ by δ_1, and ε by ε_1. Thus, we have

$$(19)\quad \varphi_1(\gamma_1,\delta_1,\varepsilon_1)=A_1^aB_1^b(\bar{A}_1\bar{B}_1-C_1^2)^c(A_1-\gamma_1)^{-a}(B_1-\delta_1)^{-b}\left[(\bar{A}_1-\gamma_1)(\bar{B}_1-\delta_1)-(C_1+\tfrac{\varepsilon_1}{2})^2\right]^{-c}$$

where A_1, B_1, $\bar{A}_1$, $\bar{B}_1$ and C_1 are the constants A, B, $\bar{A}$, $\bar{B}$ and C each multiplied by s, and $a=\frac{m-1}{2}$, $b=\frac{n-1}{2}$ and $c=\frac{s-1}{2}$. The distribution $f(\xi_1,\eta_1,\zeta_1)$ of ξ_1, η_1 and ζ_1 is then the solution of the integral equation,

$$(20)\quad \int_0^\infty\int_0^\infty\int_{-\sqrt{\xi_1\eta_1}}^{\sqrt{\xi_1\eta_1}} e^{\gamma_1\xi_1+\delta_1\eta_1+\varepsilon_1\zeta_1}f(\xi_1,\eta_1,\zeta_1)\,d\xi_1\,d\eta_1\,d\zeta_1=\varphi_1(\gamma_1,\delta_1,\varepsilon_1).$$

We note from (19) that the factor $(A_1-\gamma_1\ )^{-a}$ can be written as $(\bar{A}_1-\gamma_1)^{-a}\left(1-\frac{r^2\bar{A}_1}{\bar{A}_1-\gamma_1}\right)^{-a}$ and likewise with respect to $(B_1-\delta_1)^{-b}$. For sufficiently small values of γ_1 and δ_1 these terms can be represented by series expansions. It will be convenient to rearrange the product of these two series in a power series in r^2. Expanding and arranging in this manner we get

$$(21)\quad \varphi_1(\gamma_1,\delta_1,\varepsilon_1)=\frac{A_1^a B_1^b(\bar{A}_1\bar{B}_1-C_1^2)^c}{\Gamma(a)\Gamma(b)}(\bar{A}_1-\gamma_1)^{-a}(\bar{B}_1-\delta_1)^{-b}\left[(\bar{A}_1-\gamma_1)(\bar{B}_1-\delta_1)-\left(C_1+\frac{\varepsilon_1}{2}\right)^2\right]^{-c}$$

$$\times\sum_{i=0}^{\infty}\frac{r^{2i}}{i!}\sum_{j=0}^{i}\binom{i}{j}\Gamma(a+i-j)\Gamma(b+j)\left(\frac{\bar{A}_1}{\bar{A}_1-\gamma_1}\right)^{i-j}\left(\frac{\bar{B}_1}{\bar{B}_1-\delta_1}\right)^{j}.$$

Each term of this expansion is of the form

$$(22)\quad \varphi_k(\gamma_1,\delta_1,\varepsilon_1)=G_k(\bar{A}_1-\gamma_1)^{-a_k}(\bar{B}_1-\delta_1)^{-b_k}\left[(\bar{A}_1-\gamma_1)(\bar{B}_1-\delta_1)-\left(C_1+\frac{\varepsilon_1}{2}\right)^2\right]^{-c}$$

where G_k is a constant independent of γ_1, δ_1 and ε_1.

We are now in position to find $f(\xi_1,\eta_1,\zeta_1\ )$ as a series of terms $f_k(\xi_1,\eta_1,\zeta_1)$ whose form is given as the solution of

$$(23)\quad \int_0^{\infty}\int_0^{\infty}\int_{-\sqrt{\xi_1\eta_1}}^{\sqrt{\xi_1\eta_1}} e^{\gamma_1\xi_1+\delta_1\eta_1+\varepsilon_1\zeta_1} f_k(\xi_1,\eta_1,\zeta_1)\,d\xi_1\,d\eta_1\,d\zeta_1=\varphi_k(\gamma_1,\delta_1,\varepsilon_1)$$

The integral equation (23) can be solved by the methods used by Romanovsky[10]. Following Romanovsky we find that

$$(24)\quad f_k(\xi_1,\eta_1,\zeta_1)=G_k e^{-\bar{A}_1\xi_1-\bar{B}_1\eta_1+2C_1\zeta_1}\,\xi_1^{a_k+c-\frac{3}{2}}\,\eta_1^{b_k+c-\frac{3}{2}}\,\omega\left(\frac{\zeta_1}{\sqrt{\xi_1\eta_1}}\right)$$

where $\omega\left(\frac{\zeta_1}{\sqrt{\xi_1\eta_1}}\right)$ is an even function of $\frac{\zeta_1}{\sqrt{\xi_1\eta_1}}=t$, say, which satisfies the condition

[10] V. Romanovsky, loc. cit.

$$(25)\quad \int_{-1}^{+1} t^{2g}\omega(t)dt = \frac{\Gamma(g+\frac{1}{2})\Gamma(c+g)}{\Gamma(a_k+c+g)\Gamma(b_k+c+g)\Gamma(\frac{1}{2})\Gamma(c)} = M_g$$

for $g = 0, 1, 2, \ldots$ and $\omega(t)$ is independent of g.

To solve (25) we observe that the right side can be written as

$$(26)\quad M_g = H\int_0^1\int_0^1 u^{g-\frac{1}{2}}(1-u)^{a_k+c-\frac{3}{2}} v^{c+g-1}(1-v)^{b_k-1} du\,dv,$$

where $H = \frac{1}{\Gamma(\frac{1}{2})\Gamma(c)\Gamma(b_k)\Gamma(a_k+c-\frac{1}{2})}$. The g-th moment of t^2 is now identical with the g-th moment of the product uv. Since ω(t) is even, we have,

$$(27)\quad \int_0^1 t^{2g}\omega(t)dt = \tfrac{1}{2}M_g .$$

Setting $v = \frac{t^2}{u}$, $dv = \frac{2t}{u}dt$ in (26) we find

$$(28)\quad \omega(t) = H\int_{t^2}^1 (1-\tfrac{t^2}{u})^{a_k+c-\frac{3}{2}} u^{c-\frac{3}{2}}(1-u)^{b_k-1} du .$$

Making the transformation $\frac{1-u}{1-t^2} = \theta$, we finally obtain,

$$(29)\quad \omega(t) = \frac{(1-t^2)^{a_k+b_k+c-\frac{3}{2}}}{\Gamma(\frac{1}{2})\Gamma(c)\Gamma(a_k+b_k+c-\frac{1}{2})} F\left[a_k, b_k, a_k+b_k+c-\tfrac{1}{2}, 1-t^2\right],$$

where the F function is the ordinary hypergeometric series. Using this form with t replaced by $\frac{\zeta_1}{\sqrt{\xi_1\eta_1}}$ in (24), we have $f_k(\xi_1, \eta_1, \zeta_1)$ fully determined.

The complete solution $f'(\xi_1, \eta_1, \zeta_1)$ of (20) can be found by summing all of the expressions of the form (24) whose characteristic functions appear in the sum (21). Without much difficulty we can sum this series by expressing the coefficients as beta functions and interchanging the order of summation and integration. Accordingly, we can express $f(\xi_1, \eta_1, \varsigma_1)$ in closed form as

$$(30)\quad f(\xi_1,\eta_1,\varsigma_1)=\bar{K}e^{-\bar{A}_1\xi_1-\bar{B}_1\eta_1+2C_1\varsigma_1}\xi_1^{a+c-\frac{3}{2}}\eta_1^{b+c-\frac{3}{2}}\left(1-\frac{\varsigma_1^2}{\sqrt{\xi_1\eta_1}}\right)^{a+b+c-\frac{3}{2}}$$

$$\times\int_0^1\int_0^1(1-x)^{a+c-\frac{3}{2}}x^{b-1}\left[1-\left(1-\frac{\varsigma_1^2}{\xi_1\eta_1}\right)x\right]^{-a}y^{a-1}(1-y)^{c-\frac{3}{2}}$$

$$\times e^{r^2\left(1-\frac{\varsigma_1^2}{\xi_1\eta_1}\right)\left(\eta_1 x_1\bar{B}_1+\frac{\xi_1(1-x)y\bar{A}_1}{1-\left(1-\frac{\varsigma_1^2}{\xi_1\eta_1}\right)x}\right)}dx\,dy,$$

where $\bar{K}=\dfrac{A_1^a B_1^b(\bar{A}_1\bar{B}_1-C_1^2)^c}{\Gamma(\frac{1}{2})\Gamma(a)\Gamma(b)\Gamma(c)\Gamma(c-\frac{1}{2})}$

The distribution of $\bar{\xi}_o$, $\bar{\eta}_o$ and $\bar{\varsigma}_o$ can be found by the change of variables $\xi_1 = \frac{N_1}{s}\bar{\xi}_o$, $\eta_1 = \frac{N_2}{s}\bar{\eta}_o$, $\varsigma_1 = \varsigma_o$. It is clear that our estimate $t_o = \frac{\bar{\varsigma}_o}{\sqrt{\bar{\xi}_o\bar{\eta}_o}}$ of the correlation coefficient can range in value from $-\sqrt{\frac{N_1N_2}{s^2}}$ to $\sqrt{\frac{N_1N_2}{s^2}}$.

4. Moments of $\bar{\xi}_o$, $\bar{\eta}_o$ and $\bar{\varsigma}_o$ when $r=0$.

The general product moment $M(h,k,t)=E(\bar{\xi}_o^h\bar{\eta}_o^k\bar{\varsigma}_o^t)$ obtained from (30) for $r\neq 0$ is extremely unmanageable and impractical, since it is a generalized hypergeometric series expressed by a quadruple summation. However, for $r=0, M(h,k,t)$ is quite simple. Indeed, for this case, we have,

$$(31)\quad M(h,k,t)=\frac{\bar{K}\Gamma(a)\Gamma(c-\frac{1}{2})s^{h+k}}{\Gamma(a+c-\frac{1}{2})N_1^hN_2^k}\int_0^\infty\int_0^\infty\int_{-\sqrt{\xi_1\eta_1}}^{\sqrt{\xi_1\eta_1}}\int_0^1 e^{-A_1\xi_1-B_1\eta_1}$$

$$\times\xi_1^{a+c+h-\frac{3}{2}}\eta_1^{b+c+k-\frac{3}{2}}\varsigma_1^t\left(1-\frac{\varsigma_1^2}{\xi_1\eta_1}\right)^{a+b+c-\frac{3}{2}}$$

$$\times(1-\theta)^{a+c-\frac{3}{2}}\theta^{b-1}\left[1-\left(1-\frac{\varsigma_1^2}{\xi_1\eta_1}\right)\theta\right]^{-a}d\theta\,d\varsigma_1\,d\xi_1\,d\eta_1\,.$$

$M(h, k, t)$ exists for all positive values of h and k and for all positive integral values of t. Since the integrand is an even function of ξ_1 it follows that $M(h, k, t)=0$ for t an odd integer. If we let $t=2\nu$, set $\xi_1=t\sqrt{\xi_1\eta_1}$ in (31) and make use of (25), we find,

$$M(h,k,2\nu)=\frac{s^{h+k}A_1^{-h-\nu}B_1^{-k-\nu}}{N_1^h N_2^k} \tag{32}$$
$$\times\frac{\Gamma(b+c+k+\nu)\Gamma(a+c+h+\nu)\Gamma(\nu+\frac{1}{2})\Gamma(c+\nu)}{\Gamma(b+c+\nu)\Gamma(a+c+\nu)\Gamma(c)\Gamma(\frac{1}{2})}.$$

The 2ν-th moment of the correlation coefficient can be found from (32) by letting $h=k=-\nu$. Thus,

$$M_{2\nu}(r_0)=\left(\frac{N_1N_2}{s^2}\right)^{\nu}\frac{\Gamma(\nu+\frac{1}{2})\Gamma(c+\nu)\Gamma(a+c)\Gamma(b+c)}{\Gamma(a+c+\nu)\Gamma(b+c+\nu)\Gamma(c)\Gamma(\frac{1}{2})}.$$

The variance of r_0 is $\sigma_{r_0}^2=\frac{(s-1)N_1N_2}{s^2(N_1-2)(N_2-2)}$, which does not differ appreciably from $\frac{1}{s-1}$, which is the sampling variance of r when it is computed from ω_{xy}. The distribution of r_0 is found by setting $t=\frac{s}{\sqrt{N_1N_2}}r_0$ in (29) and multiplying by $\Gamma(a+c)$ $\Gamma(b+c)$.

The 2ν-th moments of the regression coefficient of y on x, $p=\frac{\bar{\xi}_0}{\bar{\xi}_0}$ say, is

$$M_{2\nu}(p)=M(-2\nu,0,2\nu)=\left(\frac{N_1}{s}\right)^{2\nu}\left(\frac{\sigma_y}{\sigma_x}\right)^{2\nu}\frac{\Gamma(a+c-\nu)\Gamma(\nu+\frac{1}{2})\Gamma(c+\nu)}{\Gamma(a+c+\nu)\Gamma(c)\Gamma(\frac{1}{2})},$$

and the variance is

$$\sigma_p^2=\frac{N_1^2(s-1)}{(N_1-2)(N_1-4)s^2},\frac{\sigma_y^2}{\sigma_x^2}.$$

This differs very little from the variance $\frac{1}{s-3}\cdot\frac{\sigma_y^2}{\sigma_x^2}$ of the regression coefficient using only the data from ω_{xy}.

Slightly more accurate estimates can be obtained for σ_x^2, σ_y^2 and $r\sigma_x\sigma_y$ by multiplying $\bar{\xi}_0$, $\bar{\eta}_0$ and $\bar{\zeta}_0$ by $\frac{N_1}{N_1-2}$, $\frac{N_2}{N_2-2}$ and $\frac{s}{s-1}$ respectively. These corrected estimates will have their mathematical expectations identical with σ_x^2, σ_y^2 and $r\sigma_x\sigma_y$, as will be seen from $M(1,0,0)$, $M(0,1,0)$ and $M(0,0,1)$ in (18). In this case, the general moment $M(h,k,2\nu)$ will be identical with (32) multiplied by

$$\left(\frac{N_1}{N_1-2}\right)^h\left(\frac{N_2}{N_2-2}\right)^k\left(\frac{s}{s-1}\right)^{2\nu}$$

The variance for r_0 in this case is $\frac{1}{s-1}$, and that for the regression coefficient is

$$\frac{(N_1-2)\sigma_y^2}{(s-1)(N_1-4)\sigma_x^2} .$$

5. Variances and covariances of $\sqrt{\bar{\xi}_0}$, $\sqrt{\bar{\eta}_0}$ and r_0 in large samples.

As we have seen in the last section, the product moments of $\bar{\xi}_0$, $\bar{\eta}_0$ and $\bar{\zeta}_0$ evaluated from (30) are too complicated to be of much practical value, and there is not much hope from this source of finding the sampling variance of the estimate r_0 of the correlation coefficient. The moments and variances of $\sqrt{\bar{\xi}_0}$ and $\sqrt{\bar{\eta}_0}$ taken separately are well known results. In fact, for large samples, the variances are $\frac{\sigma_x^2}{2N_1}$ and $\frac{\sigma_y^2}{2N_2}$ respectively. The variance of r_0 is not so immediately obtained. We shall find its limiting form for large samples from the normal form approached by the distribution of $\sqrt{\bar{\xi}_0}$, $\sqrt{\bar{\eta}_0}$ and r_0 as m, n and s approach ∞ in constant ratios $\frac{m}{\alpha} = \alpha$, $\frac{n}{\alpha} = \beta$.

For convenience let $\sqrt{\bar{\xi}_0} = \theta$, $\sqrt{\bar{\eta}_0} = \varphi$ and $r_0 = t$. Then we have

$$(33) \quad \theta^2 = \frac{\xi + \alpha u}{1+\alpha}, \; \varphi^2 = \frac{\eta + \beta v}{1+\beta}, \; t = \frac{\zeta}{\theta\varphi} .$$

If we integrate (1) with respect to $\bar{x}$, $\bar{y}$, $\bar{x}_1$, and $\bar{y}_1$ and perform the following transformations on the remaining part of the distribution,

$$\xi = \theta^2(1+\alpha) - \alpha u \qquad d\xi = 2\theta(1+\alpha)d\theta$$

$$\eta = \varphi^2(1+\beta) - \beta v \qquad d\eta = 2\varphi(1+\beta)d\varphi$$

$$\zeta = t\theta\varphi \qquad d\zeta = \theta\varphi dt ,$$

we can write it in the form,

$$(34) \qquad F(u,v,\theta,\phi,t)\left[f(u,v,\theta,\phi,t)\right]^{s},$$

where

$$F(u,v,\theta,\phi,t) = 4C(1+\alpha)(1+\beta)(1-r^2)^{\frac{s}{2}}\sigma_x^{s+m}\sigma_y^{s+n}e^{-\frac{m+n+2s}{2}}$$

$$\times u^{-\frac{3}{2}}v^{-\frac{3}{2}}\theta^2\phi^2\left[(\overline{1+\alpha}\,\theta^2+\alpha u)(\overline{1+\beta}\,\phi^2-\beta v)-t^2\theta^2\phi^2\right]^{-2}$$

and

$$C = \frac{s^{s-1}m^{\frac{m-1}{2}}n^{\frac{n-1}{2}}\sigma_x^{-s-m+2}\sigma_y^{-s-n+2}(1-r^2)^{-\frac{s-1}{2}}}{2^{\frac{m+n+2s-4}{2}}\Gamma(\frac{1}{2})\Gamma(\frac{s-1}{2})\Gamma(\frac{s-2}{2})\Gamma(\frac{m-1}{2})\Gamma(\frac{n-1}{2})}$$

and

$$f(u,v,\theta,\phi,t) = (1-r^2)^{-\frac{1}{2}}\sigma_x^{-1-\alpha}\sigma_y^{-1-\beta}e^{\frac{\alpha+\beta+2}{2}}u^{\frac{\alpha}{2}}v^{\frac{\beta}{2}}$$

$$\times\left[(\overline{1+\alpha}\,\theta^2-\alpha u)(\overline{1+\beta}\,\phi^2-\beta v)-t^2\theta^2\phi^2\right]^{\frac{1}{2}}$$

$$\times e^{-\frac{1}{2(1-r^2)}\left[\frac{\overline{1+\alpha}\,\theta^2-r^2\alpha u}{\sigma_x^2}+\frac{\overline{1+\beta}\,\phi^2-r^2\beta v}{\sigma_y^2}-\frac{2rt\theta\phi}{\sigma_x\sigma_y}\right]}.$$

If (34) were integrated with respect to u and v, we would get the distribution of $\theta=\sqrt{\frac{s}{N_1}}\,\xi_1$, $\phi=\sqrt{\frac{s}{N_2}}\,\eta_1$ and $t=\frac{\zeta_1}{\theta\phi}$ where the distribution of ξ_1, η_1 and ζ_1 is given by (30). The problem of finding the asymptotic normal form of the distribution of θ, ϕ and t from (30) seems extremely complicated. However, we can find this asymptotic form by first finding the limiting normal form of the distribution of u, v, θ, ϕ, and t from (34) and then integrating with respect to u and v.

The limiting normal form of (34) can be found by methods developed by von Mises[11] in a paper which appeared in 1919. In fact, $f(u,v,\theta,\phi,t)$ satisfies all of the conditions of the generalization of his first theorem to functions of more than one variable. In particular, the first order partial derivatives vanish and the

[11] R. von Mises, Fundamentalsätze der Wahrscheinlichkeitsrechnung, Mathematische Zeitschrift, Bd. 4 (1919) S. 14-18.

determinant and all of its principal minors of the negative of the Hessian are positive at the point P whose coordinates are $u=\sigma_x^2$, $v=\sigma_y^2$, $\theta=\sigma_x$, $\phi=\sigma_y$ and $t=r$. Furthermore, f is identically zero outside the region of possible values of u, v, θ, ϕ, and t.

The matrix of the negative of the second derivatives at P is

$$
\begin{array}{c|c|c|c|c|c|}
 & \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} & \frac{\partial f}{\partial \theta} & \frac{\partial f}{\partial \phi} & \frac{\partial f}{\partial t} \\
\hline
\frac{\partial f}{\partial u} & \frac{\alpha}{2\sigma_x^4}\left(\frac{\alpha}{\rho^4}+1\right) & \frac{r^2\alpha\beta}{2\sigma_x^2\sigma_y^2\rho^4} & -\frac{\alpha}{\sigma_x^3\rho^2}\left(1+\frac{\alpha}{\rho^2}\right) & -\frac{\alpha\beta r^2}{\sigma_x^2\sigma_y\rho^4} & \frac{\alpha r}{\sigma_x^2\rho^4} \\
\hline
\frac{\partial f}{\partial v} & \frac{r^2\alpha\beta}{2\sigma_x^2\sigma_y^2\rho^4} & \frac{\beta}{2\sigma_y^4}\left(\frac{\beta}{\rho^4}+1\right) & -\frac{\alpha\beta r^2}{\sigma_x\sigma_y^2\rho^4} & -\frac{\beta}{\sigma_y^3\rho^2}\left(1+\frac{\beta}{\rho^2}\right) & \frac{\beta r}{\sigma_y^2\rho^4} \\
\hline
\frac{\partial f}{\partial \theta} & -\frac{\alpha}{\sigma_x^3\rho^2}\left(1+\frac{\alpha}{\rho^2}\right) & -\frac{\alpha\beta r^2}{\sigma_x\sigma_y^2\rho^4} & \frac{2}{\sigma_x^2\rho^2}\left(\frac{r^2}{2}+\frac{(\rho^2+\alpha)^2}{\rho^2}\right) & -\frac{2r^2}{\sigma_x\sigma_y\rho^2}\left(\frac{1}{2}-\frac{\alpha\beta}{\rho^2}\right) & -\frac{2r}{\sigma_x\rho^2}\left(\frac{1}{2}+\frac{\alpha}{\rho^2}\right) \\
\hline
\frac{\partial f}{\partial \phi} & -\frac{\alpha\beta r^2}{\sigma_x^2\sigma_y\rho^4} & -\frac{\beta}{\sigma_y^3\rho^2}\left(1+\frac{\beta}{\rho^2}\right) & -\frac{2r^2}{\sigma_x\sigma_y\rho^2}\left(\frac{1}{2}-\frac{\alpha\beta}{\rho^2}\right) & \frac{2}{\sigma_y^2\rho^2}\left(\frac{r^2}{2}+\frac{(\rho^2+\beta)^2}{\rho^2}\right) & -\frac{2r}{\sigma_y\rho^2}\left(\frac{1}{2}+\frac{\beta}{\rho^2}\right) \\
\hline
\frac{\partial f}{\partial t} & \frac{\alpha r}{\sigma_x^2\rho^4} & \frac{\beta r}{\sigma_y^2\rho^4} & -\frac{2r}{\sigma_x\rho^2}\left(\frac{1}{2}+\frac{\alpha}{\rho^2}\right) & -\frac{2r}{\sigma_y\rho^2}\left(\frac{1}{2}+\frac{\beta}{\rho^2}\right) & \frac{1+r^2}{\rho^4} \\
\hline
\end{array}
$$

Now it follows at once from von Mises' theorem that

$$
(36)\quad F(u,v,\theta,\phi,t)\left[f(u,v,\theta,\phi,t)\right]^s \sim F(\sigma_x^2,\sigma_y^2,\sigma_x,\sigma_y,r)\,e^{-\frac{s}{2}\sum_{i,j=1}^{5} h_{ij}x_i x_j}
$$

where $x_1=u-\sigma_x^2$, $x_2=v-\sigma_y^2$, $x_3=\theta-\sigma_x$, $x_4=\phi-\sigma_y$ and $x_5=t-r$, and h_{ij} is the element in the i-th row and j-th column of the matrix (35). Now,

$$
F(\sigma_x^2,\sigma_y^2,\sigma_x,\sigma_y,r)=\frac{s^{\frac{5}{2}}\sqrt{\alpha\beta}\,(1+\alpha)(1+\beta)}{(2\pi)^{\frac{5}{2}}\sigma_x^3\sigma_y^3(1-r^2)^{\frac{3}{2}}}
$$

which is equal to $\left(\frac{s}{2\pi}\right)^{\frac{5}{2}}\sqrt{h}$, where h is the determinant $|h_{ij}|$.

The variables in which we are primarily interested are θ, ϕ and t. The matrix of variances and covariances of θ, ϕ and t is formed by taking the third order matrix in the lower right corner of the reciprocal form of $\|h_{ij}\|$. This matrix turns out to be

(37)

	θ	ϕ	t
θ	$\frac{\sigma_x^2}{2s(1+\alpha)}$	$\frac{r^2\sigma_x\sigma_y}{2s(1+\alpha)(1+\beta)}$	$\frac{r\sigma_x(1+\beta-r^2)}{2s(1+\alpha)(1+\beta)}$
ϕ	$\frac{r^2\sigma_x\sigma_y}{2s(1+\alpha)(1+\beta)}$	$\frac{\sigma_y^2}{2s(1+\beta)}$	$\frac{r\sigma_y(1+\alpha-r^2)}{2s(1+\alpha)(1+\beta)}$
t	$\frac{r\sigma_x(1+\beta-r^2)}{2s(1+\alpha)(1+\beta)}$	$\frac{r\sigma_y(1+\alpha-r^2)}{2s(1+\alpha)(1+\beta)}$	$\frac{(1+\alpha)(1+\beta)-r^2\left(\frac{\alpha+\beta}{2}-\alpha\beta+2\right)+r^4}{s(1+\alpha)(1+\beta)}$

The determinant of (37) is

$$(38)\qquad \frac{\sigma_x^2\sigma_y^2\left[(1-r^2)^3+(\alpha+\beta)(1-r^2)+\alpha\beta(1+r^2)\right]}{4s^3(1+\alpha)^2(1+\beta)^2}\;.$$

The variance of r_0 is given by the element in the lower right corner of (37). It can be readily shown that this variance is greater than $\frac{(1-r^2)^2}{s}$, the variance of the estimate of the correlation coefficient from ω_{xy} only—a rather surprising result.

The efficiency of θ, ϕ and t taken jointly is the ratio of the reciprocal of the determinant of (37) to B(m,n,s) in (13). That is,

$$(39)\; Eff(\theta,\phi,t)=\frac{(1-r^2)^3(1+\alpha)^2(1+\beta)^2}{\left[(1+\alpha)(1+\beta)-\alpha\beta r^4\right]\left[(1-r^2)^3+(\alpha+\beta)(1-r^2)+\alpha\beta(1+r^2)\right]},$$

which is less than unity except for the cases $r=0$ and $\alpha=\beta=0$.

6. Efficiency of the system θ, ϕ and $\frac{\zeta}{\sqrt{\xi\eta}}$.

If we use $\frac{\zeta}{\sqrt{\xi\eta}}=r_1$, say, which is the maximum likelihood estimate of r from ω_{xy}, instead of r_0 in section 5, and use the foregoing analysis of von Mises, we find the following matrix of variances and covariances for the asymptotic normal distribution of θ, ϕ and r_1 :

(40)

	θ	ϕ	r_1
θ	$\dfrac{\sigma_x^2}{2s(1+\alpha)}$	$\dfrac{r^2\sigma_x\sigma_y}{2s(1+\alpha)(1+\beta)}$	$\dfrac{r\sigma_x(1-r^2)}{2s(1+\alpha)}$
ϕ	$\dfrac{r^2\sigma_x\sigma_y}{2s(1+\alpha)(1+\beta)}$	$\dfrac{\sigma_y^2}{2s(1+\beta)}$	$\dfrac{r\sigma_y(1-r^2)}{2s(1+\beta)}$
r_1	$\dfrac{r\sigma_x(1-r^2)}{2s(1+\alpha)}$	$\dfrac{r\sigma_y(1-r^2)}{2s(1+\beta)}$	$\dfrac{(1-r^2)^2}{s}$

The determinant of this matrix is

$$(41)\qquad \frac{\sigma_x^2\sigma_y^2(1-r^2)^2\left[(1+\alpha)(1+\beta)-\frac{r^2}{2}(\alpha+\beta+2)\right]}{4s^2(1+\alpha)^2(1+\beta)^2},$$

whose reciprocal provides us with the amount of information relative to σ_x, σ_y and r yielded by the estimates θ, ϕ and r_1. The efficiency of this system of estimates is given by the ratio of the reciprocal of (41) to (13), that is,

$$Eff(\theta,\phi,r_1)=\frac{(1-r^2)(1+\alpha)^2(1+\beta)^2}{\left[(1+\alpha)(1+\beta)-r^4\alpha\beta\right]\left[(1+\alpha)(1+\beta)-\frac{r^2}{2}(\alpha+\beta+2)\right]}.$$

By comparing the systems θ, ϕ, r_0 and θ, ϕ, r_1, we actually find the latter to be more efficient, since

$$\frac{(1-r^2)^2\left[(1+\alpha)(1+\beta)-\frac{r^2}{2}(\alpha+\beta+2)\right]}{\left[(1-r^2)^3+(\alpha+\beta)(1-r^2)+\alpha\beta(1+r^2)\right]}\leq 1,$$

which is the ratio of the reciprocal of (38) to that of (41). The equality holds ony when $r=0$ or $\alpha=\beta=0$.

The distribution $f(z,w,t)$ of $\bar{\xi}_0 = z$, $\bar{\eta}_0 = w$ and $r_1 = t$ can be readily found from the distribution of ξ, η, ζ, u and v, which is included in (1), by making the following sets of transformations in succession,

(a) $\zeta = t\sqrt{\xi\eta}$, $\qquad d\zeta = \sqrt{\xi\eta}\,dt$,

(b) $$\begin{cases} \xi = \overline{1+\alpha}\,z - \alpha u & d\xi = \overline{1+\alpha}\,dz \\ \eta = \overline{1+\beta}\,z - \beta v & d\eta = \overline{1+\beta}\,dw \end{cases}$$

(c) $$\begin{cases} u = \frac{1+\alpha}{\alpha} z(1-\theta) & du = -\frac{1+\alpha}{\alpha} z\,d\theta \\ v = \frac{1+\beta}{\beta} w(1-\phi) & dv = -\frac{1+\beta}{\beta} w\,d\phi. \end{cases}$$

The result can be expressed in closed form as the definite integral

$$(42)\quad f(z,w,t) = K(1-t^2)^{\frac{s-4}{2}} z^{\frac{s+m-4}{2}} w^{\frac{s+n-4}{2}}$$
$$\times \int_0^1\int_0^1 e^{-g}\,\theta^{\frac{s-3}{2}}(1-\theta)^{\frac{m-3}{2}}\phi^{\frac{s-3}{2}}(1-\phi)^{\frac{n-3}{2}}\,d\theta\,d\phi.$$

where
$$K = \frac{\left(\frac{1}{2}\right)^{\frac{N_1+N_2-4}{2}} N_1^{\frac{N_1-2}{2}} N_2^{\frac{N_2-2}{2}} \sigma_x^{-N_1+2}\sigma_y^{-N_2+2}(1-r^2)^{-\frac{s-1}{2}}}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{s-1}{2}\right)\Gamma\left(\frac{s-2}{2}\right)\Gamma\left(\frac{m-1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)},$$

and

$$g = \frac{1}{2(1-r^2)}\left[\frac{N_1[1-r^2(1-\theta)]z}{\sigma_x^2} + \frac{N_2[1-r^2(1-\phi)]w}{\sigma_y^2} - \frac{2rt\sqrt{N_1N_2\theta\phi zw}}{\sigma_x\sigma_y}\right]$$

When $r=0$, (42) breaks into the product of three well known functions, two of which represent the distributions of the variances in samples having $s+m-2$ and $s+n-2$ degrees of freedom, and the third which is the distribution of the correlation coefficient in samples of s items from a normal population in which the correlation is zero.

IV. Summary.

Samples are considered from a bivariate normal population of x and y in which all of the members are not observed with respect to both x and y. Such a sample is broken into three parts ω_{xy}, ω_x and ω_y, where ω_{xy} is the set of s members observed with respect to both x and y, ω_x the set of m members observed with respect to x only and ω_y the remaining items observed with respect to y only.

Maximum likelihood estimates are found for the following sets of conditions:

(a) For given values of σ_x, σ_y and r, optimum estimates are found for the means a and b.

(b) For given values of a, b and r, optimum estimates are found for σ_x and σ_y.

(c) For given values of a and b, approximations are found for the optimum estimates of σ_x, σ_y and r.

Other sets of estimates considered are:

(1) Means a and b estimated independently from the x's and the y's respectively, of the sample ω.

(2) Maximum likelihood estimates of σ_x from ω_{xy} and ω_x and σ_y from ω_{xy} and ω_y, each estimated independently of the other. The estimate of $r\sigma_x\sigma_y$ is taken as the covariance from ω_{xy}. The characteristic function of these estimates is found.

(3) Estimates of σ_x and σ_y taken as the square root of the weighted averages of the variances from ω_{xy} and ω_x, and from ω_{xy} and ω_y respectively, with the estimate of r taken as the ratio of the covariance of ω_{xy} to the product of these estimates of the standard deviations.

(4) Estimates of σ_x and σ_y the same as in (3), with r estimated entirely from ω_{xy}.

The exact forms of the sampling distributions of the systems in (3) and (4) are found, as well as the asymptotic normal forms approached by these exact distributions as the size of the sample ω increases, subject to the condition that $\frac{m}{s}=\alpha$ and $\frac{n}{s}=\beta$ are constant. The limiting value of the variance of the estimate of r in (4) was found to be less than that of r in (3).

We have defined the amount of information available in a sample relative to any set of population parameters as the reciprocal of the determinant of the matrix of the limiting values, for large samples, of the variances and covariances of the maximum likelihood estimates of these parameters. It is shown that this determinant is smaller than that obtained from the asymptotic normal form approached by any other set of estimates of the same set of parameters. The amount of information relative to the parameters utilized by any other set of estimates is the reciprocal of the determinant of the matrix of the limiting values of the variances and covariances of this set of estimates. The measure of the efficiency of any set of estimates is taken as the ratio of the amount of information yielded by this set to the amount yielded by the maximum likelihood estimates. The efficiency thus defined was found for each of the sets of estimates (1), (3) and (4). It was found that the set (4) is more efficient than set (3).

S. S. Wilks