On the Effects of Non-normality on the Distribution of the Sample Product-moment Correlation Coefficient

CHARLES J. KOWALSKI

University of Michigan

SUMMARY

Samples from non-normal bivariate distributions are simulated and the densities of the sample product-moment correlation coefficient, r, estimated and compared with the corresponding normal theory densities. The results are contrasted with the literature on the subject and an attempt is made to reconcile some of the earlier conflicting conclusions regarding the robustness of the distribution of r.

Keywords: NON-NORMALITY; DISTRIBUTION OF CORRELATION COEFFICIENT; ROBUSTNESS; TRANSFORMATIONS; DENSITY ESTIMATION

1. INTRODUCTION

GIVEN a sample $(x_1, y_1), \ldots, (x_N, y_N)$ from a bivariate normal distribution, the sample product-moment correlation coefficient is given by

$$r = \sum_1^N (x_i - \bar{x})(y_i - \bar{y}) \Big/ \left\{ \sum_1^N (x_i - \bar{x})^2 \sum_1 (y_i - \bar{y})^2 \right\}^{\frac{1}{2}}. \tag{1}$$

Fisher (1915) obtained the exact sampling distribution of r for samples from a bivariate normal distribution, showing that the density of r is

$$f_N(r \mid \rho) = \frac{2^{N-3}(1-\rho^2)^{(N-1)/2}(1-r^2)^{(N-4)/2}}{\Gamma(N-2)\,\pi} \sum_{j=0}^{\infty} \frac{(2r)^j}{j!} \Gamma^2\{(N+j-1)/2\} \tag{2}$$

for samples of size N. In the case $\rho = 0$, (2) reduces to

$$f_N(r \mid \rho = 0) = \frac{\Gamma\{(N-1)/2\}}{\Gamma\{(N-2)/2\}\sqrt{\pi}} (1-r^2)^{(N-4)/2}. \tag{3}$$

If the observations are not normal, tests based on (2) may be misleading for (at least) two reasons. Firstly, the distribution of r may then differ from its normal-theory form and, secondly, we may be in a situation in which ρ is a poor measure of association. Frechét (1959) and Smith (1959) gave good discussions of this second point. The first point, that of the robustness of (2) with respect to non-normality, is considered in the following section.

2. THE EFFECTS OF NON-NORMALITY ON THE DISTRIBUTION OF r

The result (2) is valid only when r is computed from a sample from a bivariate normal distribution. The importance of studying the effects of non-normality was widely recognized and, shortly after the paper by Fisher (1915), a number of investigators attempted to determine how closely the observed distribution of r, in non-normal samples of known form, compared with the corresponding normal-theory distribution. Most of these studies were based on a Monte Carlo approach to the problem, but in several cases the exact sampling distribution of r was derived and compared with (2).

A review of the literature revealed an approximately equal dichotomy of opinion. For every study indicating the robustness of the distribution of r, one could cite another claiming to show just the opposite. In one case (Rider, 1932) there arose quite a difference of opinion on the implications of a given set of experimental results. Before discussing the reconciliation of these apparently contradictory views we outline the results of several of these studies. First we discuss the studies which support the robustness of the distribution of r and then the negative results in this connection.

2.1. *A Historical Survey*

E. S. Pearson (1929), after studying samples of sizes 20 and 30 from two "considerably non-normal distributions" (each a mixture of bivariate normal distributions) with respective correlation coefficients of 0·5346 and 0·4626, concluded that "the normal bivariate surface can be mutilated and distorted to a remarkable degree without affecting the frequency distribution of r". Two years later, Pearson (1931) considered the distribution of r in samples from independent co-ordinate variables and again reported good agreement: "The values of σ_r are in these cases in remarkably close agreement with the normal theory values." He admitted that when $\rho \neq 0$ the situation is more complicated, but claimed his earlier (1929) study showed that "even in this case the distribution of r is remarkably stable". In 1932, Pearson considered the case where the independent co-ordinate variables were not identically distributed. Summarizing the results of these studies, Pearson (1931) stated that "these results emphasize the insensitivity of the distribution of r to change in the population form".

Dunlap (1931) studied samples of size $N = 52$ (a card-shuffling experiment) and found "generally good agreement" with the normal-theory values for means and variances as well as for the correlation coefficient. Hey (1928) sampled from four non-normal populations and studied the correlation coefficient, regression coefficients and the ratio of two independent estimates of variance for each population. On the basis of his results Hey concluded "considerable non-normality in the original distribution will not affect the distributions of correlation and regression coefficients". Quensel derived the density of r in samples from a Gram–Charlier distribution and found "good agreement" with normal theory, even for small values of N, but his results are valid only when $\rho = 0$, as was noted by Gayen (1951). Nair (1941) considered 433 values of r computed from independent exponential variables and found the distribution of r "fits normal theory fairly well". He found similar results for 260 values of r from another population of independent variables. Cheriyan (1945) continued Nair's work, extending his investigation to the case when ρ was large. He used samples from χ^2 distributions to construct three populations with correlation coefficients of 0·50, 0·75 and 0·89. He considered the distribution of r in some 200 samples of size $N = 5$ from each of these populations. He found close agreement of σ_r with its normal theory value for certain values of r, but noted that as ρ increased, the agreement decreased. Rider (1932) studied the distribution of r in samples from "rectangular" and "triangular" populations. He fitted Pearson-type curves to the observed distributions of r and applied χ^2 tests to evaluate the agreement with normal theory. He agreed with Pearson, concluding that "the distribution of r is quite insensitive to parental non-normality", despite the fact that most people (including the editor of the journal in which the paper appeared) would interpret his Tables V and VI as showing quite the opposite.

Despite the number of studies pointing to the robustness of (2) with respect to non-normality, some people noticed that the distribution of r in samples from certain populations did not attain the promised agreement with normal theory. This was especially evident when $|\rho|$ was near unity. Haldane (1949) agreed with Pearson as long as $|\rho|$ was small and as long as the "mutilations" were confined to skewness. He gave an example showing that when $|\rho|$ is large a slight change in kurtosis may have a large effect on the variance of r, even for large values of N. Gayen (1951) derived the distribution of r in samples from bivariate Edgeworth surfaces. He indicated that when $\rho = 0$ the effect of non-normality is not serious, even for samples as small as $N = 11$, but less agreement with normal theory is in evidence when $\rho \neq 0$.

Baker (1930) considered an example where non-normality led to a significant value of r when the normal-theory test for independence was used, but where there was good reason to believe that the variables were independent. He presented a detailed description of his data, including a scatter diagram and histograms of the marginals, and argued that this inconsistency could be attributed to the non-normal character of the marginal distributions. His analysis was reviewed by Kowalski and Tarter (1969).

Chesire, *et al.* (1932) sampled from two "triangular" distributions and noted that the variance of r was always less (and often considerably less) than the normal-theory value. They concluded that "there exist real differences between the distributions of r from these two populations and those appropriate for the normal case". They also discussed the inadequacy of ρ as a measure of association in the event of non-linear regression and tied this new evidence to the implications of Pearson's earlier studies.

More recently, Farlie (1960) examined the performance of some correlation coefficients in samples from a general class of bivariate distributions. This class is defined by the distribution function

$$H(x,y) = F(x)\,G(y)\,[1+\alpha\,A\{F(x)\}B\{G(y)\}], \tag{4}$$

where F and G are the marginal distribution functions, $|\alpha| \leqslant 1$ is a parameter measuring association, A and B are bounded functions such that $A(1) = \mathrm{B}(1) = 0$ and the derivatives $d(FA)/dF$, $d(GB)/dG$ are bounded. He showed that r provides an efficient test for independence whenever A and B have the form

$$A\{F(x)\} = \int_{-\infty}^{x} x dF \Big/ \int_{-\infty}^{x} dF, \quad B\{G(y)\} = \int_{-\infty}^{y} y dG \Big/ \int_{-\infty}^{y} dG \tag{5}$$

but for other choices of the "disturbing functions" A and B other coefficients of association are more efficient. For example, if $A = 1-F$ and $B = 1-G$, Spearman's rank correlation coefficient and Kendall's τ are more efficient. This case was first studied by Morgenstern (1956) and Gumbel considered the Morgenstern system for various marginal distributions $F = G$: normal (Gumbel, 1958); exponential (Gumbel, 1960); logistic (Gumbel, 1961). Kowalski (1968) considered other choices of $F = G$ and, prompted by Pearson's (1932) study, situations in which $F \neq \mathrm{G}$.

2.2. *Outline of the Methods Used to Re-examine this Problem*

Everyone seems to agree that the distribution of r is quite robust to non-normality when $\rho = 0$, but there is good evidence that this becomes less stable with increasing values of $|\rho|$, especially when kurtosis is in evidence. It is the variance of r which is most vulnerable to the effects of non-normality and this variance may be either larger or smaller than the normal-theory value, depending on the type of non-normality under consideration.

The amount of disagreement exhibited above may be reduced by the observation that the distribution of r is in fact quite robust to certain types of non-normality and a realization of the difficulties (c.f., Teicheroew, 1965) inherent in simulating non-normal distributions before the advent of electronic computers. Another important factor was the difficulty of estimating the distribution of r. This was done either by simple histograms or by fitting a Pearson-type curve to the observed distribution of r. Only in rare instances could the distribution of r be derived explicitly and hence it was necessary to rely on Monte Carlo experiments for the bulk of the theory at a time when this approach required massive quantities of data to get a reliable estimate of the density of r.

This reliance on a Monte Carlo approach to the problem is often still necessary today, but now simulation presents few real difficulties and the art of density estimation has been perfected to the extent that reliable estimates of the density of r are readily available. Thus we propose to study the robustness of the distribution of r using modern simulation and density estimation techniques. Good discussions of bivariate simulation are given by Mardia (1967) and Tocher (1963). Further references may be found in Shubick (1960) and Teicheroew (1965).

2.3. *Density Estimation*

Recently there have appeared a number of papers concerned with the estimation of the density function of an absolutely continuous distribution on the real line. A discussion of these procedures has been given by Kowalski (1968) and for reasons given there the author feels that the "Fourier approach" to density estimation provides accurate estimators of the densities considered in this paper. Other estimators may, of course, be used but in this paper we employ only the Fourier estimator. This method of estimation was developed by Kronmal and Tarter (1968). The form of the estimator of the density f is

$$\hat{f}_N(r) = \frac{1}{2(b-a)} + \frac{1}{2}\sum_{k=1}^{m}\left\{\bar{c}_k \cos\left(k\pi\frac{r-a}{b-a}\right) + \bar{s}_k \sin\left(k\pi\frac{r-a}{b-a}\right)\right\}, \tag{6}$$

where $\bar{c}_k$, $\bar{s}_k$ are the sample trigonometric moments, [a, b] is the (finite) interval over which the density estimate is desired and m, the number of terms in the summation in (6), is determined by a stopping rule designed to optimize the fit of (6) with respect to the minimization of the mean-integrated-square-error of the estimator. For details see Tarter *et al.* (1967) and Kronmal and Tarter (1968).

We use the estimator (6) to estimate the density of r in samples from a variety of non-normal (X, Y) distributions. $\hat{f}(r)$ is reserved to denote the Fourier estimator (6) based on 100 samples of size 30 from some non-normal distribution (X, Y). We begin by reviewing Pearson's results for mixtures of bivariate normal distributions.

3. The Distribution of r from a Mixture of Bivariate Normal Distributions

It is clear that Pearson's conclusions concerning the insensitivity of the distribution of r must be confined to the situation he considered, but one might hope that the study provides evidence that the distribution of r is robust with respect to the class of mutilations which may be introduced by mixing.

The (X, Y) densities obtained by mixing bivariate normal densities have the structure $h(x,y) = w_1\phi_1(x,y)+\ldots+w_k\phi_k(x,y)$ where the w_i are probability weights and the ϕ_i are (standard) bivariate normal densities with respective correlation coefficients ρ_i. The correlation coefficient of (X, Y) is

$$\rho_{XY} = \sum_{i=1}^{k} w_i \rho_i. \tag{7}$$

Pearson considered $k = 3$ in his investigations and the mixtures he studied did lead to distributions of r (as estimated by (6)) which agreed quite well with (2), but it is an easy task to find mixtures for which this agreement is far from satisfactory. Fig. 1 shows $f_{30}(r\,|\,\rho = 0)$ and the estimated densities of r in mixtures where $k = 2$, $w_1 = w_2 = \frac{1}{2}$, $\rho_1 = -\rho_2$ and it is easily seen that the goodness of fit deteriorates rapidly with increasing values of $|\rho_1|$. Fig. 2 illustrates the estimated densities of r for several other mixtures which have $\rho_{XY} = 0$, but not $\rho_1 = -\rho_2$. Fig. 3 and 4 illustrate the situation for several values of $\rho_{XY} \neq 0$.

The goodness of fit of these estimated densities is a function of $(w_1, w_2, \rho_1, \rho_2)$. The choice of these parameters induces a distortion of the normal surface which may or may not grossly affect the distribution of r. In particular, Fig. 1 shows that even when $\rho_{XY} = 0$ the density of r may depart considerably from its normal-theory form. Fig. 5 shows the sort of mutilation of the surface mixing can produce. The contour map depicted there departs considerably from the equiprobable ellipses of the normal model.

It appears, then, that Pearson's (1929) claim should be modified. The distribution of r in samples from mixtures of bivariate normal distributions may depart considerably from the corresponding normal density even when $\rho_{XY} = 0$ and even for large sample sizes.

4. The Distribution of r from Other Systems of Bivariate Distributions

Fig. 6 shows the estimated density of r in samples of size 30 from the exponential distribution defined by the distribution function

$$H(x,y) = 1-\exp(-x)-\exp(-y)+\exp(-x-y-\delta xy), \tag{8}$$

where $0 \leqq \delta \leqq 1$ is a parameter measuring association. Gumbel (1960) studied the relationship between ρ_{XY} and δ and provided a graph from which one can read off values of ρ_{XY} for given values of δ. For Fig. 6 we have put $\delta = 0{\cdot}3$; this corresponds to $\rho_{XY} = -0{\cdot}2$ so $\hat{f}(r)$ is compared with the normal-theory density $f_{30}(r\,|\,\rho = -0{\cdot}2)$.

Fig. 7 is the analogous treatment of the exponential distribution in the Morgenstern system, that is (4) with $A = 1-F$, $B = 1-G$, where F, G are exponential distribution functions. Here $\alpha = 1$, which corresponds to $\rho_{XY} = \frac{1}{4}$ by the relation $\rho_{XY} = \alpha/4$. Farlie (1960) showed that r was less efficient than Spearman's ρ and Kendall's τ in this case, even though the estimated density of r agrees quite well with normal theory.

Fig. 8 compares $\hat{f}(r)$ and $f_{30}(r \mid \rho = 0{\cdot}75)$ for samples of size 30 from Cheriyan's (1945) bivariate χ^2 distribution with $\rho = 0{\cdot}75$. Cheriyan claimed that he found good agreement with normal theory for values of r in the interval $(0{\cdot}1, 0{\cdot}8)$ and samples as small as 5, but this finding is not substantiated by Fig. 8.

Fig. 9 compares the estimated density of r in samples from the standard semi-logarithmic surface (Yuan, 1933) with the corresponding normal theory density when $\rho_{XY} = 0{\cdot}57$. Fig. 10 is the analogous treatment of the bivariate lognormal distribution with $\rho_{XY} = 0{\cdot}65$. These distributions are examples of distributions generated from the bivariate normal by the method of translation (Johnson, 1949). If ρ is the correlation coefficient of the normal variables it is easy to show that for semilogarithmic (X, Y), $\rho_{XY} = \rho/(e-1)^{\frac{1}{2}}$ and for bivariate lognormal (X, Y),

$$\rho_{XY} = \{\exp(\rho - 1)\}/(e-1),$$

where e is the base of the natural logarithms.

The graphs given here are only a representative portion of the work done on this problem. Other (X, Y) systems have been considered as well as other sample sizes (N) and values of ρ_{XY}. The conclusion reached by several of the earlier investigators

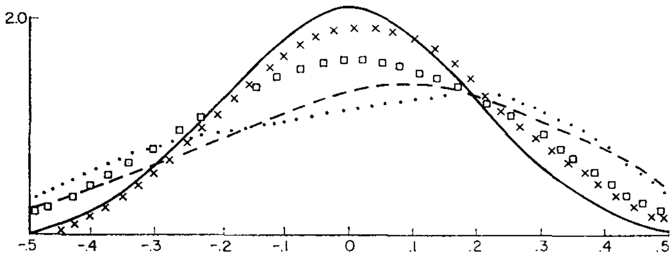

FIG. 1

——— $f_{30}(r \mid \rho = 0)$.
× × × × $\hat{f}_{30}(r)$ for mixture $w_1 = w_2 = \frac{1}{2}$, $\rho_1 = -\rho_2 = 0{\cdot}25$.
□□□□ $\hat{f}_{30}(r)$ for mixture $w_1 = w_2 = \frac{1}{2}$, $\rho_1 = -\rho_2 = 0{\cdot}50$.
– – – – $\hat{f}_{30}(r)$ for mixture $w_1 = w_2 = \frac{1}{2}$, $\rho_1 = -\rho_2 = 0{\cdot}75$.
.... $\hat{f}_{30}(r)$ for mixture $w_1 = w_2 = \frac{1}{2}$, $\rho_1 = -\rho_2 = 0{\cdot}90$.

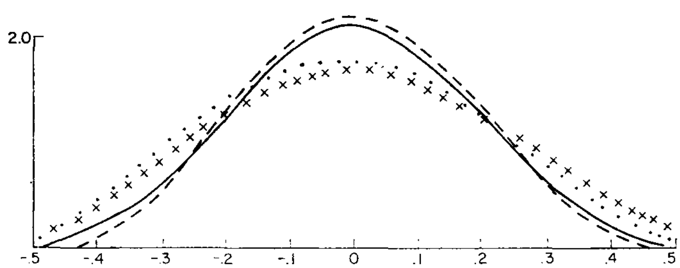

FIG. 2

– – – – $\hat{f}_{30}(r)$ for mixture $w_1 = 0{\cdot}431$, $w_2 = 0{\cdot}569$, $\rho_1 = 0{\cdot}333$, $\rho_2 = -0{\cdot}25$.
——— $f_{30}(r \mid \rho = 0)$.
.... $\hat{f}_{30}(r)$ for mixture $w_1 = 0{\cdot}25$, $w_2 = 0{\cdot}75$, $\rho_1 = 0{\cdot}75$, $\rho_2 = -0{\cdot}25$.
× × × × $\hat{f}_{30}(r)$ for mixture $w_1 = 0{\cdot}333$, $w_2 = 0{\cdot}667$, $\rho_1 = -0{\cdot}50$, $\rho_2 = 0{\cdot}25$.

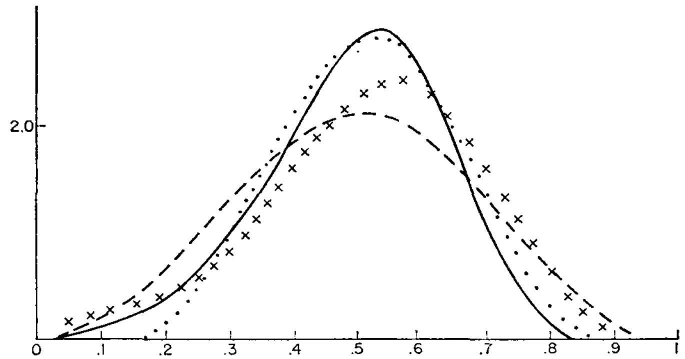

FIG. 3

———— $f_{30}(r \mid \rho = 0{\cdot}5)$.

.... $\hat{f}_{30}(r)$ for mixture $w_1 = 0{\cdot}7$, $w_2 = 0{\cdot}3$, $\rho_1 = 0{\cdot}4$, $\rho_2 = 0{\cdot}733$.

× × × × $\hat{f}_{30}(r)$ for mixture $w_1 = 0{\cdot}5$, $w_2 = 0{\cdot}5$, $\rho_1 = 0{\cdot}9$, $\rho_2 = 0{\cdot}1$.

– – – – $\hat{f}_{30}(r)$ for mixture $w_1 = 0{\cdot}9$, $w_2 = 0{\cdot}1$, $\rho_1 = 0{\cdot}6$, $\rho_2 = -0{\cdot}4$.

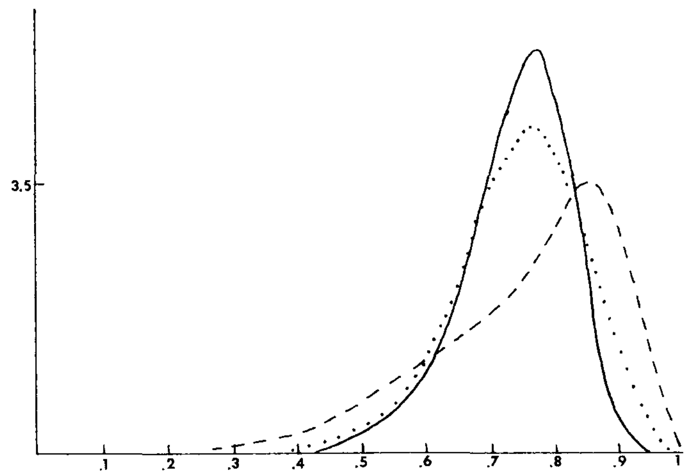

FIG. 4

———— $f_{30}(r \mid \rho = 0{\cdot}75)$.

.... $\hat{f}(r)$ for mixture $w_1 = 0{\cdot}5$, $w_2 = 0{\cdot}5$, $\rho_1 = 0{\cdot}9$, $\rho_2 = 0{\cdot}6$.

– – – – $\hat{f}(r)$ for mixture $w_1 = 0{\cdot}9$, $w_2 = 0{\cdot}1$, $\rho_1 = 0{\cdot}9$, $\rho_2 = -0{\cdot}6$.

FIGS. 1–4. Comparisons of the densities $\hat{f}(r)$ of the sample product-moment correlation coefficient for various mixtures of two bivariate normal densities, estimated from 100 samples of size 30, with the corresponding normal-theory density $f_{30}(r \mid \rho)$ given by equation (2) and having the same value of the correlation coefficient ρ_{XY}. The components of the mixtures have correlation coefficients ρ_1, ρ_2 and probability weights w_1, w_2.

of this problem that the distribution of r in samples from non-normal distributions may differ substantially from the normal-based density (2), especially for large values of $|\rho|$, has been verified and extended. It has been further noted that even when good agreement between the densities is in evidence, for example in Fig. 7, other coefficients of association may be more efficient than r. The departure from normal theory is less serious when $\rho = 0$ but there exist situations in which $\hat{f}_N(r)$ differs substantially from $f_N(r \mid \rho = 0)$. Fig. 11 illustrates that even if X and Y are independent the distribution of r may be skewed and differ considerably from $f_N(r \mid \rho)$.

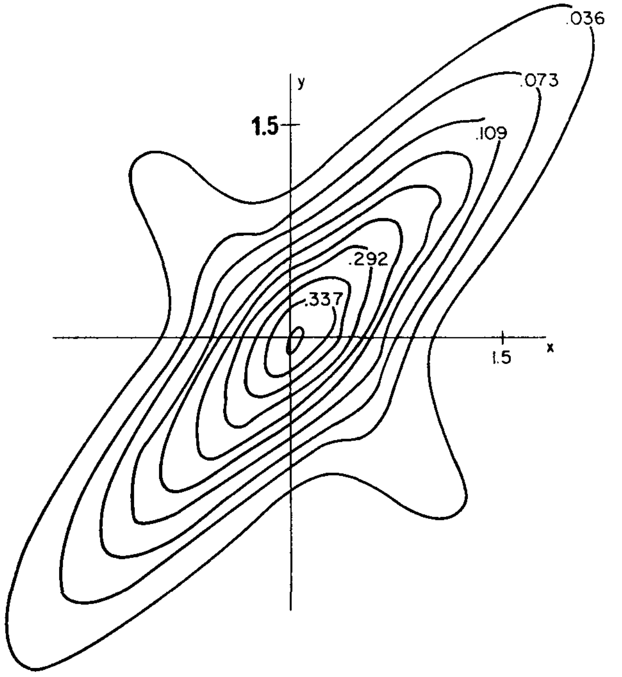

Fig. 5. Contour map of the mixture of two bivariate normal densities with correlation coefficients 0·9, −0·9 and probability weights 0·8, 0·2.

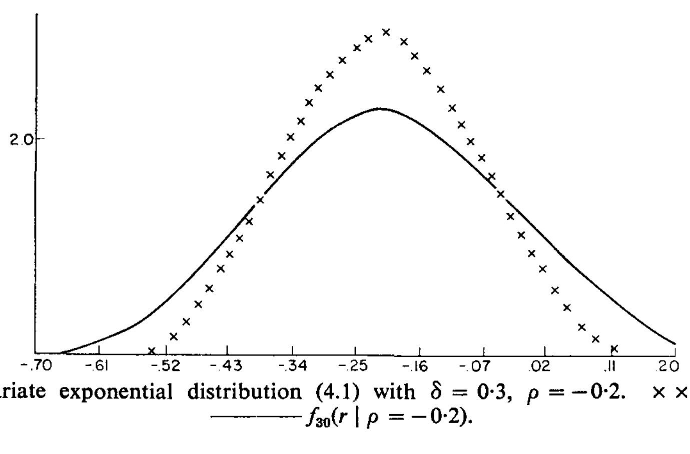

Fig. 6. Bivariate exponential distribution (4.1) with $\delta = 0{\cdot}3$, $\rho = -0{\cdot}2$. × × × × $\hat{f}_{30}(r_{XY})$.
——— $f_{30}(r \mid \rho = -0{\cdot}2)$.

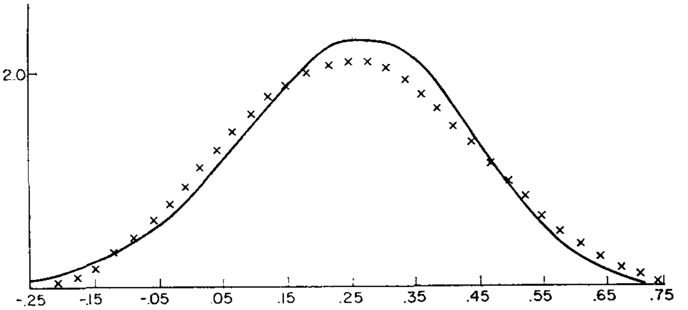

FIG. 7. Bivariate exponential distribution (2.1) with $\alpha = 1$, $\rho = \alpha/4 = 0{\cdot}25$. ——— $f_{30}(r \mid \rho = 0{\cdot}25)$. × × × × $\hat{f}_{30}(r_{XY})$.

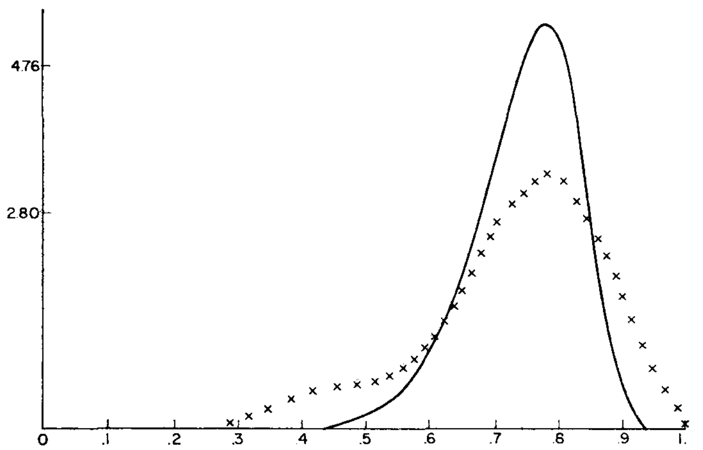

FIG. 8. Cheriyan's bivariate χ^2 distribution with $\rho = 0{\cdot}75$. ——— $f_{30}(r \mid \rho = 0{\cdot}75)$. × × × × $\hat{f}_{30}(r_{XY})$.

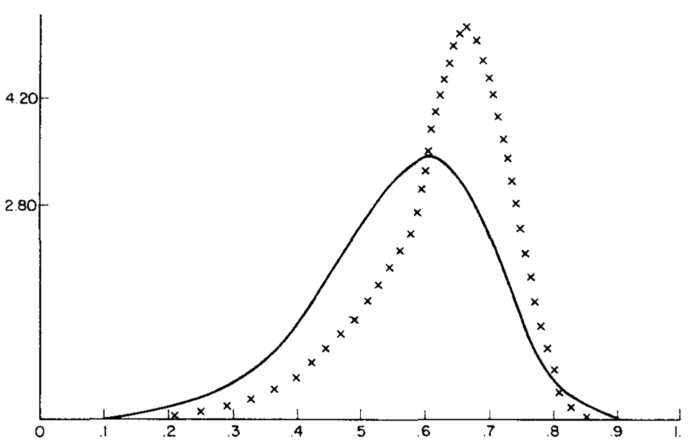

FIG. 9. The semilogarithmic correlation surface with $\rho = 0{\cdot}75$ and $\rho_{XY} = 0{\cdot}57$. ——— $f_{30}(r \mid \rho = 0{\cdot}57)$. × × × × $\hat{f}_{30}(r_{XY})$.

This figure is based on (X, Y) independent exponential variables (compare Nair, 1941, who found better agreement in this case), and similar results have been noted for other independent systems. The dashed line in the graph, corresponding to $\hat{f}_{30}(r_{ZW})$, is discussed in the following section.

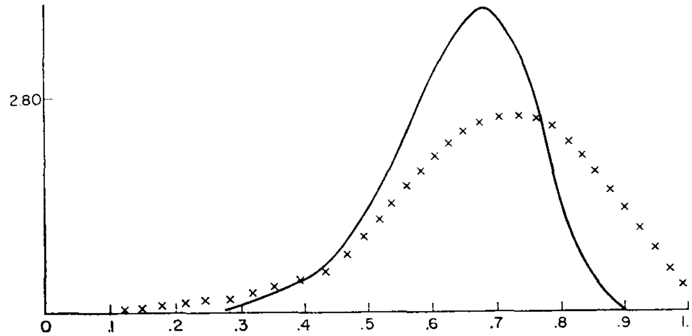

FIG. 10. The standard bivariate lognormal distribution with $\rho = 0{\cdot}75$ and $\rho_{XY} = 0{\cdot}65$. ——— $f_{30}(r \mid \rho = 0{\cdot}65)$. × × × × $\hat{f}_{30}(r_{XY})$.

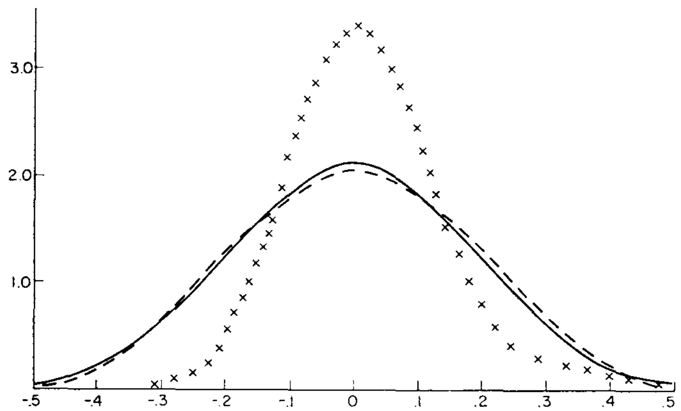

FIG. 11. Independent exponential variables. ——— $f_{30}(r \mid \rho = 0)$. × × × × $\hat{f}_{30}(r_{XY})$. – – – – $\hat{f}_{30}(r_{ZW})$.

FIGS. 6–11. Comparisons of the densities $\hat{f}(r_{XY})$ of the sample product-moment correlation coefficient from various systems of bivariate (X, Y) distributions estimated from 100 samples of size 30, with the corresponding normal-theory density $f_{30}(r \mid \rho)$ given by equation (2) and having the same value of the correlation coefficient ρ_{XY}. FIG. 11 also shows the density $\hat{f}_{ZW}$ obtained after transforming the marginal distributions of (X, Y) to normality.

5. Some Methodological Suggestions

If $(X_1, Y_1), \ldots, (X_N, Y_N)$ are normal, a test of $H_0\colon \rho = 0$ is a test for independence of X and Y. There is little argument over the propriety of basing this test on r when (X, Y) is normal, but there is less agreement when normality is suspect. One possible approach is suggested by Fieller and Pearson (1961) who showed that for any bivariate distribution which is roughly normal under the transformation of its marginal distributions to normality the Fisher–Yates (1938) rank correlation coefficient, r_F, can be used, when testing $H_0\colon \rho = 0$, as though it were the product-moment correlation, r, of N pairs of independently distributed normal variables. Carrying this approach one step further the author (Kowalski, 1968, 1970) and Kowalski and Tarter (1969) investigated the use of co-ordinate transformations to normality as a prelude to the application of normal-based correlation analyses. Given non-normal (X, Y), the technique consists of making the co-ordinate transformations

$$Z = \Phi^{-1}\{\hat{F}(X)\}, \quad W = \Phi^{-1}\{\hat{G}(Y)\}, \tag{9}$$

where Φ^{-1} is the inverse of the standard normal distribution function and $\hat{F}$, $\hat{G}$ are the Fourier estimators of the marginal distribution functions, and then using normal theory to analyse the correlation structure of (Z, W). This method is based on the assumption that co-ordinate transformations to normality produce bivariate distributions which are "approximately normal". Otherwise stated, the assumption is that normal correlation analyses are robust with respect to the kinds of non-normality possible when the marginals are constrained to be normal. The author (Kowalski, 1968) assembled a list of non-normal bivariate distributions with normal marginals and contrasted their correlation structures with the corresponding normal models, compared (Kowalski, 1970) the performance of tests for bivariate normality before and after applying the transformations (9) and showed (Kowalski and Tarter, 1969) that the normal-based test for independence is generally more powerful if based on r_{ZW} than if r_{XY} is used. It seems clear as a result of these studies that (Z, W) is generally "more bivariate normal" than (X, Y) and that *if normal correlation theory is going to be used*, it is more properly based on r_{ZW} than on r_{XY}. An example of the agreement of r_{ZW} with normal theory is illustrated in Fig. 11: the dashed line is the estimated density of r_{ZW}. For a wide class of (X, Y) distributions, r_{ZW} agrees more closely with normal theory than does r_{XY} and this agreement is exhibited for a wide range of values of ρ. This stands in opposition to what is known about the distribution of r_F; while r_F does agree quite well with normal theory when $\rho = 0$, the agreement is less satisfactory when $\rho \neq 0$ (e.g. $E(r_F) < E(r)$ in this case). Despite these advantages there do exist examples of non-normal distributions with normal marginals whose correlation structures differ substantially from that of the bivariate normal distribution. Indeed, mixtures of normal distributions have normal marginals. See also Vaswani (1947, 1950).

In the event, the examples considered in this paper serve to warn against the indiscriminate use of normal correlation analyses. There may remain arguments as to which alternative analysis should be applied, but it is hoped that there will no longer be any doubt that alternative analyses are needed.

6. Conclusions

The literature on the subject of the robustness of the distribution of r with respect to non-normality of the observations was reviewed in the light of new experimental results based on modern simulation and density-estimation techniques. The general conclusion is that the distribution of r may be quite sensitive to non-normality and that normal correlation analyses should be limited to situations in which (X, Y) is (at least very nearly) normal. The distribution of r need not agree well with normal theory when $\rho = 0$ and even if the distribution of r is close to the normal theory density alternative analyses may be more efficient.

References

Baker, G. A. (1930). The significance of the product moment correlation coefficient with special reference to the character of the marginal distributions. *J. Amer. Statist. Ass.*, **25**, 387–396.

CHERIYAN, K. C. (1945). Distributions of certain frequency constants in samples from non-normal populations. *Sankhyā*, **7**, 159–166.

CHESIRE, L., OLDIS, E. and PEARSON, E. S. (1932). Further experiments on the sampling distribution of the correlation coefficient. *J. Amer. Statist. Ass.*, **27**, 121–128.

DUNLAP, H. F. (1931). An empirical determination of the distributions of means, standard deviations, and correlation coefficients drawn from rectangular populations. *Ann. Math. Statist.*, **2**, 66–81.

FARLIE, D. J. G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, **47**, 307–323.

FIELLER, E. C. and PEARSON, E. S. (1961). Tests for rank correlation coefficients II. *Biometrika*, **48**, 29–40.

FISHER, R. A. (1915). The frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, **10**, 507–521.

FISHER, R. A. and YATES, R. (1938). *Statistical Tables for Biological, Agricultural and Medical Research*. Edinburgh: Oliver and Boyd.

FRECHÉT, M. (1959). A note on simple correlation. *Math. Mag.*, **32**, 265–268.

GAYEN, A. K. (1951). The frequency distribution of the product moment correlation coefficient in random samples of any size drawn from non-normal universes. *Biometrika*, **38**, 219–247.

GUMBEL, E. J. (1958). Distributions à plusiers variables dont les marges sont données. *Comptes Rendus de l'Academie des Sciences, Paris*, **246**, 2712.

—— (1960). Bivariate exponential distributions. *J. Amer. Statist. Ass.*, **55**, 698–707.

—— (1961). Bivariate logistic distributions. *J. Amer. Statist. Ass.*, **56**, 335–349.

HALDANE, J. B. S. (1949). A note on non-normal correlation. *Biometrika*, **36**, 467–468.

HEY, G. B. (1928). A new method of experimental sampling illustrated on certain non-normal populations. *Biometrika*, **30**, 68–80.

JOHNSON, N. L. (1949). Bivariate distributions based on simple translation systems. *Biometrika*, **36**, 297–304.

KOWALSKI, C. J. (1968). *Coordinate Transformations to Normality in the Theory of Correlation.* Doctoral Dissertation, University of Michigan.

—— (1970). The performance of some rough tests for bivariate normality before and after coordinate transformations to normality. *Technometrics*, **12**, 517–544.

KOWALSKI, C. J. and TARTER, M. E. (1969). Coordinate transformations to normality and the power of normal tests for independence. *Biometrika*, **56**, 139–148.

KRONMAL, R. and TARTER, M. E. (1968). Fourier methods for the estimation of densities and distribution functions. *J. Amer. Statist. Ass.*, **63**, 925–952.

MARDIA, K. V. (1967). Some contributions to contingency-type bivariate distributions. *Biometrika*, **54**, 235–249.

MORGENSTERN, D. (1956). Einfache Beispiele Zweidimensionalter Verteilungen. *Mitt. Math. Statist.*, **8**, 234–235.

NAIR, A. N. K. (1941). Distribution of Student's *t* and the correlation coefficient in samples from non-normal populations. *Sankhyā*, **5**, 383–400.

PEARSON, E. S. (1929). Some notes on sampling tests with two variables. *Biometrika*, **21**, 337–360,

—— (1931). The test of significance for the correlation coefficient. *J. Amer. Statist. Ass.*, **26**. 128–134.

RIDER, P. R. (1932). The distribution of the correlation coefficient in small samples. *Biometrika*, **24**, 382–403.

SHUBICK, M. (1960). Bibliography on simulation, gamming, artificial intelligence and allied topics. *J. Amer. Statist. Ass.*, **55**, 736–751.

SMITH, C. D. (1959). Some further notes on the theory of correlation. *Math. Mag.*, **32**, 269–270.

TARTER, M. E., HOLCOMB, R. L. and KRONMAL, R. A. (1967). A description of new computer methods for estimating the population density. *Proc. Ass. Comp. Mach.*, Thompson Book Co., **22**, 511–519.

TEICHEROEW, D. (1965). A history of distribution sampling prior to the era of the computer and its relevance to simulation. *J. Amer. Statist. Ass.*, **60**, 27–49.

TOCHER, K. D. (1963). *The Art of Simulation.* London: English Universities Press.

VASWANI, S. P. (1947). A pitfall in correlation theory. *Nature, Lond.*, **160**, 405–406.

—— (1950). Assumptions underlying the use of the tetrachoric correlation coefficient. *Sankhyā*, **10**, 269–276.

YUAN, P. (1933). On the logarithmic frequency distribution and the semilogarithmic correlation surface. *Ann. Math. Statist.*, **4**, 30–74.