The Annals of Statistics
1980, Vol. 8, No. 2, 264–278

ESTIMATION OF THE CORRELATION COEFFICIENT FROM A BROKEN RANDOM SAMPLE[1]

BY MORRIS H. DEGROOT AND PREM K. GOEL

Carnegie-Mellon University and Purdue University

Inference about the correlation coefficient ρ in a bivariate normal distribution is considered when observations from the distribution are available only in the form of a broken random sample. In other words, a random sample of n pairs is drawn from the distribution but the observed data are only the first components of the n pairs and, separately, some unknown permutation of the second components of the n pairs. Under these conditions, the estimation of ρ is, as Samuel Johnson put it, "like a dog's walking on his hinder legs. It is not done well; but you are surprised to find it done at all." We study the maximum likelihood estimation of ρ and present some effective procedures for estimating the sign of ρ.

1. Introduction. Suppose that a random sample of size n is drawn from a bivariate normal distribution with pdf $f(x, u)$. Suppose also, however, that before the sample values can be recorded, each observation vector (x_i, u_i) gets broken into the two separate components x_i and u_i, and these observations are available only in the form $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$, where $y_1, \cdots, y_n$ is some unknown permutation of $u_1, \cdots, u_n$. Since the pairings in the original sample are not known, the observed values are called a *broken random sample* from the given bivariate population [DeGroot, Feder, and Goel (1971)]. It is assumed for simplicity, and without loss of generality, that all of the observed values are distinct. If the correlation coefficient ρ is known, then the problem of optimally repairing the observed values so as to reproduce as many of the vectors from the original sample as possible was considered by DeGroot, Feder, and Goel (1971). Related matching problems have been considered by Chew (1973), Goel (1975), and DeGroot and Goel (1976). In this paper, we shall primarily consider the problem of making inferences about ρ when it is unknown, and we shall only secondarily consider the problem of making inferences about the unknown pairings in the original sample. It is well known that ρ can be estimated reasonably well if the original paired sample values are known. For example, the maximum likelihood estimator (MLE) of ρ, based on the unbroken sample, is the sample correlation coefficient. Here, we shall investigate the question of whether there is any information about ρ in the

Received March 1976; revised February 1979.

[1]This research was supported in part by the National Science Foundation under Grant SOC77-07548 at Carnegie-Mellon University.

AMS 1970 *subject classifications.* Primary 62F10, 62H20; secondary 62C99.

Key words and phrases. Correlation coefficient, estimation, broken random sample, bivariate normal distribution, likelihood function, Fisher information.

broken random sample. This question is answered, at least partly, in the affirmative in the next four sections.

This problem is related to the problem of information in the marginal totals of a contingency table, which has been studied by Good (1976), Plackett (1977), Berkson (1978), and others. The basic difference between the two problems is that they are dealing with categorical data and we are dealing with continuous data.

In Section 2, the MLE's of the means, the variances, the correlation coefficient ρ, and the unknown pairing are derived. In particular, it is shown that the MLE's of the means and the variances are the same as for an unbroken sample, and the MLE of ρ is either the largest or the smallest possible sample correlation coefficient that can be obtained among all possible pairings of the values in the broken sample. In Section 3, we consider the problem of determining the MLE's of the means, the variances, and ρ from a likelihood function that has been summed over all possible pairings of the values in the broken sample. Various properties of the likelihood equation for ρ and its solutions are derived. Some typical plots of the likelihood function are presented. In Section 4 we consider the problem of estimating the sign of ρ and present three simple decision rules. The results of a Monte Carlo study show the effectiveness of these rules. In Section 5, the Fisher information matrix for a broken random sample is studied. It is shown that the information about ρ in a broken random sample of size n is at least as large as the information in a random sample of size 1 from a bivariate normal distribution with known means and variances. However, at $\rho = 0$, these two numbers are equal.

2. Estimation of the unknown parameters. We shall assume that the random variable (X, U) has a bivariate normal distribution with $E(X) = \mu_1$, $E(U) = \mu_2$, $\mathrm{Var}(X) = \sigma_1^2$, $\mathrm{Var}(U) = \sigma_2^2$, and $\mathrm{Corr}(X, U) = \rho$, and that all the parameters μ_1, μ_2, σ_1^2, σ_2^2, and ρ are unknown. Furthermore, we shall assume that the observed values in a broken random sample from this distribution are $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$. Let Φ be the set of all permutations $\phi = (\phi(1), \cdots, \phi(n))$ of the integers $1, 2, \cdots, n$. Pairing the observed values in the broken random sample according to the permutation ϕ means pairing x_i with $y_{\phi(i)}$ for $i = 1, \cdots, n$.

We begin by determining the MLE's of the unknown parameters μ_1, μ_2, σ_1^2, σ_2^2, and ρ and the unknown permutation ϕ which specifies the pairings in the original sample. For given values $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$, the log-likelihood function of $\phi, \rho, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ is

$$(2.1)\quad L(\phi, \rho, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2 | \mathbf{x}, \mathbf{y}) = -\frac{n}{2}\log(1-\rho^2) - \frac{n}{2}\log\sigma_1^2 - \frac{n}{2}\log\sigma_2^2$$
$$-\frac{1}{2(1-\rho)^2}\left\{\Sigma_1^n\frac{(x_i-\mu_1)^2}{\sigma_1^2} + \Sigma_1^n\frac{(y_i-\mu_2)^2}{\sigma_2^2} - 2\rho\Sigma_1^n\frac{(x_i-\mu_1)}{\sigma_1}\frac{(y_{\phi(i)}-\mu_2)}{\sigma_2}\right\}.$$

A constant term not involving the parameters has been omitted from (2.1).

For any fixed ϕ, it is known that based on the sample paired according to ϕ, the log-likelihood function (2.1) is maximized when $\hat{\mu}_1 = \bar{x}$, $\hat{\mu}_2 = \bar{y}$, $\hat{\sigma}_1^2 = s_x^2$, $\hat{\sigma}_2^2 = s_y^2$, and $\hat{\rho} = r_\phi$ where r_ϕ denotes the usual sample correlation coefficient,

$$r_\phi = \frac{1}{n}\Sigma_1^n(x_i - \bar{x})y_{\phi(i)}/(\hat{\sigma}_1\hat{\sigma}_2). \tag{2.2}$$

Then the log-likelihood function (2.1) evaluated at these points is

$$\hat{L}(\phi) = L(\phi, r_\phi, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2|\mathbf{x}, \mathbf{y}) = -\frac{n}{2}\log(1 - r_\phi^2) - n. \tag{2.3}$$

We must determine the maximum of this function over all permutations ϕ.

We shall let

$$r_{\min} = \min_{\phi \in \Phi} r_\phi \quad \text{and} \quad r_{\max} = \max_{\phi \in \Phi} r_\phi \tag{2.4}$$

and let $x_{(1)} < \cdots < x_{(n)}$ and $y_{(1)} < \cdots < y_{(n)}$ denote the order statistics of the x_i's and the y_i's. It is well known [Hardy, Littlewood, and Pólya (1967)] that $r_{\max}$ is attained by the permutation ϕ^0 such that $x_{(i)}$ is paired with $y_{(i)}$ for $i = 1, \cdots, n$, and that $r_{\min}$ is attained by the permutation ϕ_0 such that $x_{(i)}$ is paired with $y_{(n+1-i)}$ for $i = 1, \cdots, n$. It follows from Theorem C, page 61, of Hájek and Šidák (1967) that $\Sigma_{\phi \in \Phi} r_\phi = 0$. Since the value of r_ϕ is not the same for every permutation $\phi \in \Phi$, $r_{\min} < 0$ and $r_{\max} > 0$.

It now follows from (2.3) and this discussion that the MLE's $\hat{\rho}$ and $\hat{\phi}$ are given by

$$\begin{aligned} \hat{\rho} &= r_{\max} \quad \text{and} \quad \hat{\phi} = \phi^0 \quad \text{if} \quad r_{\max} > |r_{\min}|, \\ \hat{\rho} &= r_{\min} \quad \text{and} \quad \hat{\phi} = \phi_0 \quad \text{if} \quad r_{\max} < |r_{\min}|. \end{aligned} \tag{2.5}$$

It should be noted that $\hat{\rho}$ is not a reasonable estimator of ρ in the sense that $\hat{\rho}$ is always equal to either the maximum or the minimum possible sample correlation coefficient that can be calculated from the broken sample. In the next section, we shall consider a useful modification of the likelihood function. We shall complete this section by considering briefly problems in which the values of some of the parameters are known.

Suppose that the values of μ_1 and μ_2 are known and the values of σ_1^2, σ_2^2, and ρ are unknown. Without loss of generality, we shall assume that $\mu_1 = \mu_2 = 0$. For $\phi \in \Phi$, let

$$r_\phi^* = (\Sigma_1^n x_i y_{\phi(i)})/[(\Sigma_1^n x_i^2)(\Sigma_1^n y_i^2)]^{\frac{1}{2}}, \tag{2.6}$$

and let

$$r_{\min}^* = \min_{\phi \in \Phi} r_\phi^* \text{ and } r_{\max}^* = \max_{\phi \in \Phi} r_\phi^*. \tag{2.7}$$

Then it can be shown that the MLE's $\hat{\rho}$ and $\hat{\phi}$ are given by

$$\begin{aligned} \hat{\rho} &= r_{\max}^* \quad \text{and} \quad \hat{\phi} = \phi^0 \quad \text{if} \quad r_{\max}^* > |r_{\min}^*|, \\ \hat{\rho} &= r_{\min}^* \quad \text{and} \quad \hat{\phi} = \phi_0 \quad \text{if} \quad r_{\max}^* < |r_{\min}^*|. \end{aligned} \tag{2.8}$$

In this problem it is not necessarily true that $r_{\max}^* > 0$ and $r_{\min}^* < 0$. However, it can be seen from (2.8) that if $r_{\max}^* < 0$, then $\hat{\rho} = r_{\min}^*$, and if $r_{\min}^* > 0$, then $\hat{\rho} = r_{\max}^*$.

When the values of μ_1, μ_2, σ_1^2, and σ_2^2 are known, there is no simple expression for $\hat{\rho}$. Suppose without loss of generality that $\mu_1 = \mu_2 = 0$ and $\sigma_1^2 = \sigma_2^2 = 1$. By maximizing the function L given in (2.1) over $\phi \in \Phi$ for any fixed value of ρ, and then differentiating the result with respect to ρ, it can be shown that $\hat{\rho}$ must be a solution of the following equation:

$$n\rho(1 - \rho^2) - \rho\Sigma_1^n(x_i^2 + y_i^2) + (1 + \rho^2)\Sigma_1^n x_i y_{\phi(i)} = 0, \tag{2.9}$$

where $\phi = \phi^0$ for $\rho > 0$ and $\phi = \phi_0$ for $\rho < 0$. We shall not consider this problem further.

3. The integrated likelihood function. For each $\phi \in \Phi$, let $g(\mathbf{x}, \mathbf{y} | \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho, \phi)$ denote the joint pdf of the observations $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$ when they are paired according to the permutation ϕ. We shall now assume that the true unknown permutation can be assigned a prior distribution. In particular, we assume that the $n!$ permutations in Φ are equally likely a priori to have generated the original sample. Therefore, the (marginal) joint pdf of the observations $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$ in the broken sample, for given values of the parameters μ_1, μ_2, σ_1^2, σ_2^2, and ρ, is the average

$$\frac{1}{n!}\Sigma_{\phi \in \Phi} g(\mathbf{x}, \mathbf{y} | \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho, \phi). \tag{3.1}$$

When (3.1) is regarded as a function of the parameters μ_1, μ_2, σ_1^2, σ_2^2, and ρ, for given values of $\mathbf{x}$ and $\mathbf{y}$, it is called the *integrated likelihood function*. (See, e.g., Kalbfleisch and Sprott (1970). In our paper, since ϕ is discrete, the integrated likelihood function might better be called the summed likelihood function.) Since the integrated likelihood function (3.1) no longer involves the nuisance parameter ϕ, this function would seem to be appropriate for problems in which we are interested solely in making inferences about the remaining parameters.

Some statisticians would regard ϕ as missing data rather than as a parameter. In these circumstances, it is appropriate to use a method such as the EM algorithm of Dempster, Laird and Rubin (1977) for obtaining MLE's from incomplete data. The estimates that we shall obtain by using (3.1) are the ones that would be obtained from the EM algorithm.

Each term in the sum in (3.1) is the product of n bivariate normal pdf's for x_i and $y_{\phi(i)}$. It can be shown, therefore, that the log-likelihood function has the following form:

$$\begin{aligned} \mathcal{L}(\rho, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2 | \mathbf{x}, \mathbf{y}) = &-\frac{n}{2}\log\left[(1 - \rho^2)\sigma_1^2\sigma_2^2\right] \\ &- \frac{1}{2(1 - \rho^2)}\Sigma_1^n\left[\left(\frac{x_i - \mu_1}{\sigma_1}\right)^2 + \left(\frac{y_i - \mu_2}{\sigma^2}\right)^2\right] \\ &+ \log \Sigma_{\phi \in \Phi} \exp\left[\frac{\rho}{1 - \rho^2}\Sigma_1^n\left(\frac{x_i - \mu_1}{\sigma_1}\right)\left(\frac{y_{\phi(i)} - \mu_2}{\sigma_2}\right)\right]. \end{aligned} \tag{3.2}$$

A constant term has been omitted from (3.2).

It can be shown by direct differentiation that the MLE's $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\sigma}_1^2$, and $\hat{\sigma}_2^2$ are simply the sample means $\bar{x}$ and $\bar{y}$ and the sample variances s_x^2 and s_y^2, just as they would be for an unbroken random sample. The concentrated log-likelihood function (i.e., the log-likelihood function evaluated at $\mu_1 = \hat{\mu}_1$, $\mu_2 = \hat{\mu}_2$, $\sigma_1^2 = \hat{\sigma}_1^2$, and $\sigma_2^2 = \hat{\sigma}_2^2$) is

$$\hat{\mathcal{L}}(\rho) = \mathcal{L}(\rho, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2 | \mathbf{x}, \mathbf{y}) = -\frac{n}{2}\log(1-\rho^2) - \frac{n}{1-\rho^2} + \log \Sigma_{\phi \in \Phi} \exp\left(\frac{n\rho r_\phi}{1-\rho^2}\right), \tag{3.3}$$

where r_ϕ is defined by (2.2). Terms not involving ρ have been dropped from (3.3). The MLE of ρ will be a solution of the equation

$$\frac{d\hat{\mathcal{L}}(\rho)}{d\rho} = 0. \tag{3.4}$$

We shall rewrite this equation in a particularly useful form.

For $\phi \in \Phi$ and any given value of $\rho\,(|\rho| < 1)$, let

$$\xi(\phi|\rho) = \frac{\exp\left(\frac{n\rho r_\phi}{1-\rho^2}\right)}{\Sigma_{\psi \in \Phi} \exp\left(\frac{n\rho r_\psi}{1-\rho^2}\right)}. \tag{3.5}$$

Thus, for each given value of ρ, $\xi(\phi|\rho)$ represents a probability distribution over the $n!$ permutations in Φ. It can now be found by differentiating (3.3) that equation (3.4) can be written in the following form:

$$\Sigma_{\phi \in \Phi} r_\phi \xi(\phi|\rho) = \rho. \tag{3.6}$$

Furthermore, for any value of ρ which satisfies equation (3.6), it can be shown that the relation $d^2\hat{\mathcal{L}}(\rho)/d\rho^2 < 0$ will be satisfied if and only if

$$\Sigma_{\phi \in \Phi} r_\phi^2 \xi(\phi|\rho) < \rho^2 + \frac{(1-\rho^2)^2}{n(1+\rho^2)}. \tag{3.7}$$

Hence, the MLE $\hat{\rho}$ will be a value of ρ that satisfies (3.6) and (3.7). The value of $\hat{\rho}$ can be determined by an iterative computation, but if n is moderately large, the cost of this computation may be large. However, a few important properties of the local maxima and minima of $\hat{\mathcal{L}}(\rho)$ can be derived from (3.6) and (3.7).

THEOREM 1. (i) *The function* $\hat{\mathcal{L}}(\rho)$ *is increasing for* $\rho \leqslant r_{\min}$ *and decreasing for* $\rho \geqslant r_{\max}$.

(ii) *The value* $\rho = 0$ *is a local minimum of* $\hat{\mathcal{L}}(\rho)$.

(iii) *There exists at least one local maximum of* $\hat{\mathcal{L}}(\rho)$ *in each of the intervals* $r_{\min} < \rho < 0$ *and* $0 < \rho < r_{\max}$.

PROOF. It is easily verified that $d\hat{\mathfrak{L}}(\rho)/d\rho \gtrless 0$ according as $\Sigma_{\phi\in\Phi}r_\phi\xi(\phi|\rho) \gtrless \rho$. Since $\Sigma_{\phi\in\Phi}r_\phi\xi(\phi|\rho)$ is a weighted average of the $n!$ values of r_ϕ, and these values are not all equal, it follows that

$$r_{\min} < \Sigma_{\phi\in\Phi}r_\phi\xi(\phi|\rho) < r_{\max}.$$

Hence, $d\hat{\mathfrak{L}}(\rho)/d\rho > 0$ for $\rho \leqslant r_{\min}$ and $d\hat{\mathfrak{L}}(\rho)/d\rho < 0$ for $\rho \geqslant r_{\max}$, which proves part (i) of the theorem.

To prove part (ii), note that $\xi(\phi|0) = 1/(n!)$ for each $\phi \in \Phi$ and $\Sigma_{\phi\in\Phi}r_\phi = 0$. Hence, the value $\rho = 0$ satisfies (3.6). Furthermore, it follows from (3.7) that the value of $\rho = 0$ will be a local minimum if $(1/n!)\Sigma_{\phi\in\Phi}r_\phi^2 > 1/n$. However, by Theorem C, page 61 of Hájek and Šidák (1967), $(1/n!)\Sigma_{\phi\in\Phi}r_\phi^2 = 1/(n-1) > 1/n$. Thus, part (ii) is proved.

Together, parts (i) and (ii) of the theorem imply part (iii). □

The positive and negative values of ρ which are local maxima will typically not be equal in absolute value, since the $n!$ values of r_ϕ will not be symmetric with respect to 0 unless either the x's or the y's are perfectly symmetric. Based on extensive simulation, it is our belief that there is exactly one positive and one negative local maximum. It should be noted that the use of $\hat{\mathfrak{L}}(\rho)$ as a log-likelihood function for ρ alone has some undesirable features. For example, values of ρ near 0 are more "likely" than $\rho = 0$ for every sample.

A Monte Carlo study of the likelihood functions for both an unbroken sample and a broken sample for $n = 5$ was performed. Three functions of ρ were studied under various conditions: (1) the likelihood function based on an unbroken sample when the means and variances of the underlying distribution are assumed known; (2) the integrated likelihood function (3.1) based on a broken sample when the means and variances are known; and (3) the concentrated likelihood function, given by (3.1) evaluated at $\mu_i = \hat{\mu}_i$ and $\sigma_i^2 = \hat{\sigma}_i^2$ for $i = 1, 2$. These functions were plotted for several different samples drawn from underlying distributions with various true values of ρ.

For many samples, the likelihood function for the broken sample looked very similar to the one for the unbroken sample, but the concentrated likelihood function was very much different from these two. This situation is illustrated in Figure 1, where these functions are sketched for particular samples of size $n = 5$ from distributions in which the true ρ is 0, .5, and .9.

For other samples, however, the likelihood function and the concentrated likelihood function for the broken sample looked very similar and the likelihood function for the unbroken sample was very much different from these two. This situation is illustrated in Figure 2.

When the means and variances are known, the likelihood function for ρ based on a broken sample will often be similar to what it would have been based on the unbroken sample. For such data, the MLE of ρ based on the broken sample will be reasonable. However, for many samples, these two likelihood functions will not be

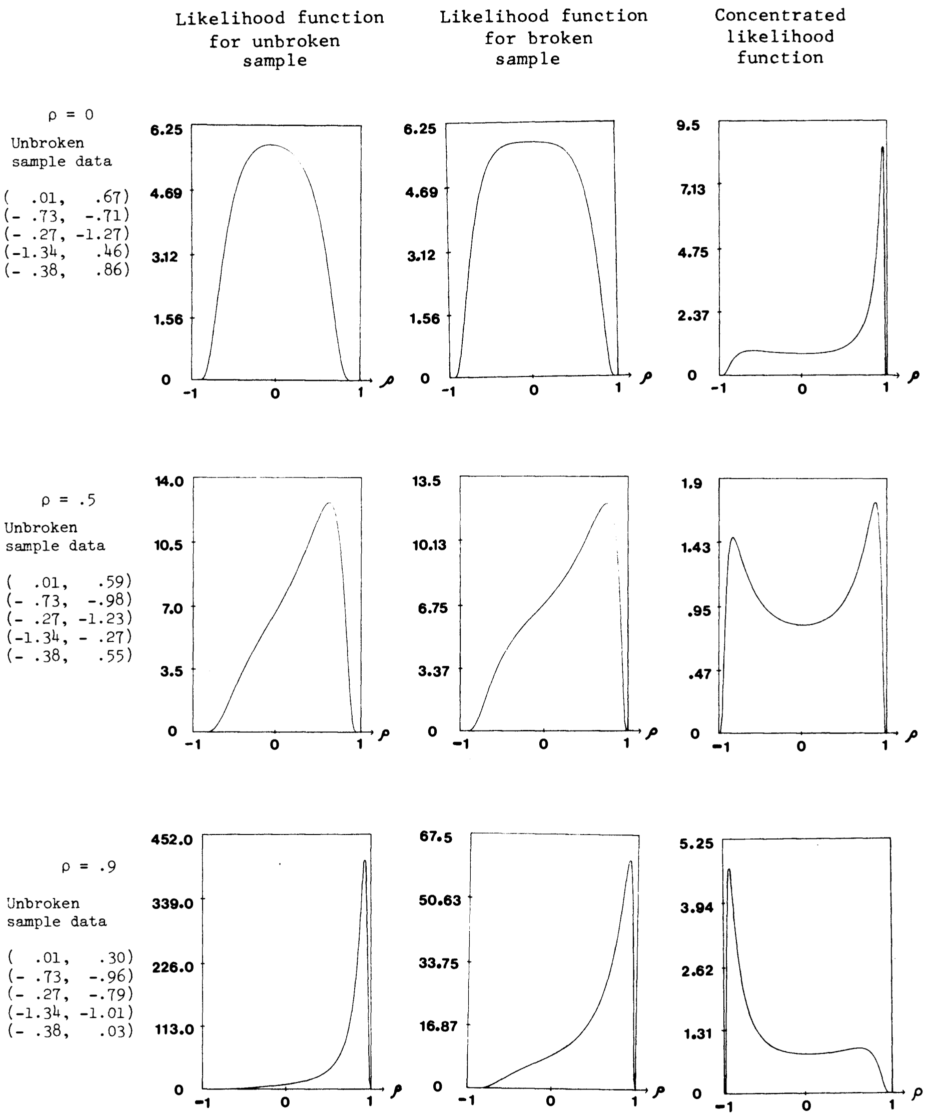

FIG. 1.

similar. The remark by Plackett (1977) in regard to the marginal totals of a 2×2 contingency table that "the standard procedures for making statements about an unknown parameter are found to be inconclusive" also holds for this problem in the sense that this procedure will be good for some broken samples and poor for others. Unfortunately, we do not know of any methods by which one can determine which type of sample has been obtained.

When the means and variances are unknown, the maxima of the concentrated likelihood function for ρ based on a broken sample will typically occur in the extremes of the interval $-1 < \rho < 1$, as illustrated in Figures 1 and 2. Thus, in this

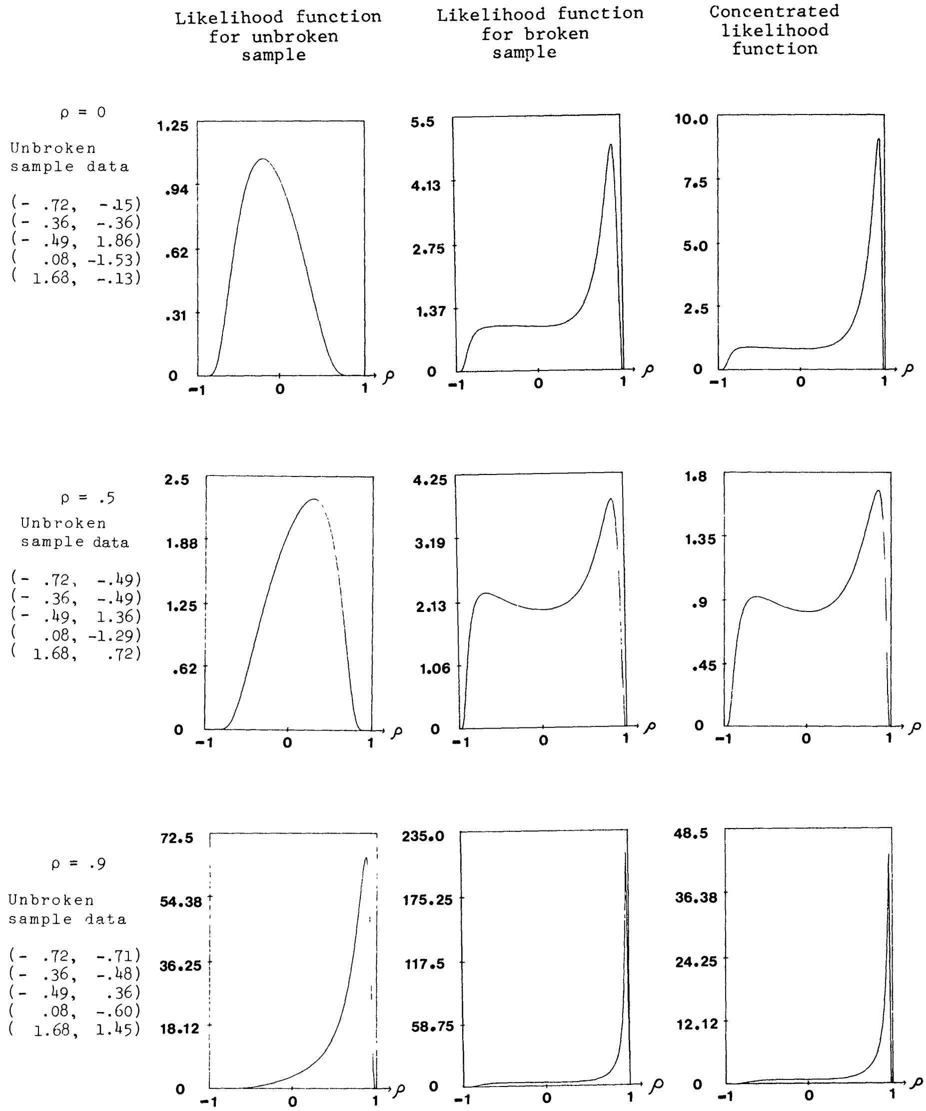

FIG. 2.

case, the MLE of ρ will not be a good estimator. Other estimators of ρ can be constructed based on the fact that the joint distribution of the $2n$ standardized residuals $(x_i - \bar{x})/s_x$ and $(y_i - \bar{y})/s_y$ $(i = 1, 2, \cdots, n)$ depends only on the parameter ρ. Bayes estimators might also be constructed. Although all these estimators will be rather weak estimators of the magnitude of ρ, we shall show in the next section that a broken sample does contain useful information about the sign of ρ.

4. Inference about the sign of ρ. In many practical problems, the experimenter is interested in determining the sign of ρ. As pointed out in DeGroot, Feder and

Goel (1971), the only information about ρ needed for obtaining the maximum likelihood matching is its sign. An important question is whether or not there is any information about the sign of ρ in the broken random sample. We will try to answer this question in this section.

Since both x and y are marginally normally distributed, it follows that, for large samples, the data for each of the two variables will be close to symmetric and hence the likelihood function (3.3) will be approximately a symmetric function of ρ. Therefore, one might intuitively conclude that the broken random sample has no information about the sign of ρ. However, we will show that this is not the case. In fact, we will present three simple decision rules that give the correct sign of ρ, with high probability, when the true value of $|\rho|$ is large.

The first decision rule is motivated by the MLE's $\hat{\rho}$ and $\hat{\phi}$ in (2.5) and the fact that $\hat{\phi} = \phi^0$ if $r_{\max} \geqslant |r_{\min}|$ and $\hat{\phi} = \phi_0$ if $r_{\max} \leqslant |r_{\min}|$. It was shown in DeGroot, Feder and Goel (1971) that if $\rho > 0$, the MLE of ϕ is ϕ^0, and if $\rho < 0$, the MLE of ϕ is ϕ_0. Hence our first decision rule for determining $\operatorname{sgn}(\rho)$ is:

Rule R_1. If $r_{\max} \geqslant |r_{\min}|$, estimate $\operatorname{sgn}(\rho)$ to be $+1$, and if $r_{\max} < |r_{\min}|$, estimate $\operatorname{sgn}(\rho)$ to be -1.

The other two decision rules are based on the reasoning that if $\rho > 0$, then we expect the skewness characteristics of the two samples to be similar. Therefore, we consider simple measures of skewness for the x and the y samples and declare $\operatorname{sgn}(\rho) = +1$ if both of these measures have the same sign. These rules can be described as follows.

Rule R_2. Let $N(\mathbf{x})$ = number of x_i's $\geqslant \frac{1}{2}(x_{(1)} + x_{(n)})$ and $N(\mathbf{y})$ = number of y_i's $\geqslant \frac{1}{2}(y_{(1)} + y_{(n)})$. If both $N(\mathbf{x})$ and $N(\mathbf{y})$ are at least $[(n+1)/2]$ or both are less than $[(n+1)/2]$, then estimate $\operatorname{sgn}(\rho)$ to be $+1$. Otherwise, estimate $\operatorname{sgn}(\rho)$ to be -1.

The third decision rule is based on the behavior characteristic of the third central moments of the x and the y samples. It is described as follows.

Rule R_3. Estimate $\operatorname{sgn}(\rho)$ to be $+1$ if $\Sigma_1^n(x_i - \bar{x})^3 \cdot \Sigma_1^n(y_i - \bar{y})^3 > 0$. Otherwise, estimate $\operatorname{sgn}(\rho)$ to be -1.

In order to find the probability of correct decision for each of these decision rules, a Monte Carlo study was carried out, in which 15,000 samples of size 5, 11, 15, 19, and 25 were drawn from the bivariate normal populations with means zero, variances one, and $\rho = 0, .25, .5, .75, .9$ using the acceptance-rejection algorithm and the random number generation package RVP at Purdue University; see Rubin (1976). In addition, to investigate the behavior of these decision rules for large n, 3,000 samples of size 100 and 500 were drawn using the IMSL subroutine GGNMP which uses Marsaglia's algorithm. For $n = 100$ and 500, the decision rule R_2 was not investigated because its performance for smaller sample sizes was consistently inferior to R_1 and R_3. Table 1 provides estimated values of the probability, P^*, of correct decision regarding $\operatorname{sgn}(\rho)$. For each sample size, the first row corresponds to

TABLE 1

Proportion of samples in which sgn(ρ) *is estimated to be* +1.

$n \setminus \rho$	0	.25	.5	.75	.9
	.4962	.5015	.5221	.5930	.7244
5*	.4958	.5021	.5259	.5880	.6952
	.4960	.5011	.5233	.5922	.7021
	.5041	.5035	.5229	.6077	.7601
11*	.5032	.5031	.5221	.5982	.7093
	.5003	.5029	.5238	.6045	.7253
	.5028	.5019	.5314	.6193	.7795
15*	.5033	.5040	.5280	.6084	.7182
	.5003	.5064	.5339	.6169	.7386
	.5021	.5048	.5307	.6237	.7875
19*	.4971	.5061	.5287	.6151	.7232
	.5035	.5096	.5332	.6221	.7495
	.5078	.5089	.5347	.6279	.7909
25*	.5033	.5047	.5291	.6106	.7216
	.5055	.5072	.5351	.6261	.7477
	.4960	.5103	.5377	.6530	.8360
100**	—	—	—	—	—
	.4957	.4983	.5293	.6277	.7533
	.4980	.5000	.5573	.6717	.8517
500**	—	—	—	—	—
	.5090	.5127	.5553	.6483	.7610

*based on 15,000 runs
**based on 3,000 runs

the decision rule R_1, the second row corresponds to the decision rule R_2 and the third row corresponds to the decision rule R_3. For $\rho = 0$, each of these rules has probability .5 of estimating sgn(ρ) to be +1, and it can be seen from Table 1 that all of the corresponding empirical values are within 1.35 σ of .5.

It is clear from Table 1 that for large $|\rho|$, the value of P^* is generally largest for the decision rule R_1 and smallest for the decision rule R_2, and that the procedures R_1 and R_3 perform reasonably well in estimating sgn(ρ). Furthermore, for any fixed $n > 2$, it can be shown that $P^* \to 1$ as $|\rho| \to 1$ for all three decision rules. In this sense, these rules are consistent. However, we do not believe that there is any decision rule for which $P^* \to 1$ as $n \to \infty$ for all values of ρ. For the rule R_3, we can obtain the asymptotic value of P^*.

THEOREM 2. *Let* $X_1, \cdots, X_n$ *and* $Y_1, \cdots, Y_n$ *be a broken random sample from a bivariate normal distribution with correlation* ρ. *Then*

$$\text{(4.1)} \qquad \lim_{n\to\infty} \Pr\left[\Sigma_1^n (X_i - \bar{X})^3 \cdot \Sigma_1^n (Y_i - \bar{Y})^3 > 0\right] = \tfrac{1}{2} + \frac{1}{\pi}\sin^{-1}\rho^3.$$

PROOF. The joint limiting distribution of $m_{3,X} = 1/n\,\Sigma_1^n(X_i - \bar{X})^3$ and $m_{3,Y} = 1/n\,\Sigma_1^n(Y_i - \bar{Y})^3$ is bivariate normal with means 0. Furthermore, the correlation of $m_{3,X}$ and $m_{3,Y}$ is ρ^3; see Kendall and Stuart (1963) page 323. Therefore, the left-hand side of (4.1) is equal to the probability that a two-dimensional random vector having a bivariate normal distribution with mean vector $\mathbf{0}$ and correlation ρ^3 takes values in the first or third quadrant. This probability is $\frac{1}{2} + (1/\pi)\sin^{-1}\rho^3$; see Kendall and Stuart (1963) page 351. □

It follows from Theorem 2 that for the rule R_3, the limiting value of P^* as $n \to \infty$ is given by the right-hand side of (4.1). Some values of this function are given in Table 2.

TABLE 2
Limiting values of P^* for the decision rule R_3.

$\lvert\rho\rvert$	.250	.500	.750	.800	.900	.950	.983	.996	1
P^*	.505	.540	.639	.671	.760	.828	.900	.950	1

Thus, no matter how large the sample size, the broken random sample contains some information about sgn(ρ). Furthermore, one observes that the empirical values for $n = 500$ given in Table 1 and the limiting values of P^* given in Table 2 are well within sampling errors, and even for $n = 25$, they are not too different.

At this time, we do not have any asymptotic result for the rule R_1. However, based on the above discussion and the empirical values of P^* for R_1 and R_3, we conjecture that the limiting value of P^* for the rule R_1 is greater than that for the rule R_3.

5. Fisher information. We shall now obtain the Fisher information matrix for the parameters μ_1, μ_2, σ_1, σ_2, and ρ in a bivariate normal distribution based on a broken random sample from that distribution. It is well known that the Fisher information matrix based on an unbroken random sample of n pairs is as follows:

$$(5.1)\qquad \mathbf{I} = \frac{n}{1-\rho^2}\left[\begin{array}{cc|ccc} \mu_1 & \mu_2 & \sigma_1 & \sigma_2 & \rho \\ 1/\sigma_1^2 & -\rho/(\sigma_1\sigma_2) & 0 & 0 & 0 \\ & 1/\sigma_2^2 & 0 & 0 & 0 \\ \hline & & \dfrac{2-\rho^2}{\sigma_1^2} & \dfrac{-\rho^2}{\sigma_1\sigma_2} & \dfrac{-\rho}{\sigma_1} \\ & & & \dfrac{2-\rho^2}{\sigma_2^2} & \dfrac{-\rho}{\sigma_2} \\ & & & & \dfrac{1+\rho^2}{1-\rho^2} \end{array}\right]$$

It can be proved that for a broken random sample, the first two rows and first two columns of the information matrix, pertaining to the parameters μ_1 and μ_2, will remain the same as in (5.1). However, the 3×3 submatrix on the lower diagonal, pertaining to the parameters ρ, σ_1, and σ_2, will change. We shall derive only one of these terms. The other terms can be obtained by a similar argument. In particular, we shall determine the value of

$$(5.2) \qquad i_n(\rho) = E\left[-\frac{\partial^2 L(\rho, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2 | \mathbf{X}, \mathbf{Y})}{\partial \rho^2} \right],$$

where $\mathbf{X} = (X_1, \cdots, X_n)$ and $\mathbf{Y} = (Y_1, \cdots, Y_n)$ are the random variables in a broken random sample, and the log-likelihood function L is defined by equation (3.2).

For any variables $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$, let

$$(5.3) \qquad \beta_\phi = \frac{1}{\sigma_1 \sigma_2} \Sigma_i (X_i - \mu_1)(Y_{\phi(i)} - \mu_2)$$

and

$$(5.4) \qquad \alpha_\phi = \frac{\exp\left(\dfrac{\rho \beta_\phi}{1 - \rho^2} \right)}{\Sigma_{\psi \in \Phi} \exp\left(\dfrac{\rho \beta_\psi}{1 - \rho^2} \right)}.$$

Then it can be shown from equation (3.2) that

$$(5.5) \qquad i_n(\rho) = -\frac{n(1 + \rho^2)}{(1 - \rho^2)^2} + \frac{2n(1 + 3\rho^2)}{(1 - \rho^2)^3} - \frac{2\rho(3 + \rho^2)}{(1 - \rho^2)^3} E(\Sigma_{\phi \in \Phi} \beta_\phi \alpha_\phi)$$
$$- \frac{(1 + \rho^2)^2}{(1 - \rho^2)^4} E\left[\Sigma_{\phi \in \Phi} \beta_\phi^2 \alpha_\phi - (\Sigma_{\phi \in \Phi} \beta_\phi \alpha_\phi)^2 \right],$$

where the expectations in (5.5) are calculated under the assumption that the n pairs of random variables $(X_1, Y_1), \cdots, (X_n, Y_n)$ form a random sample from a bivariate normal distribution with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation ρ. Furthermore, it can be seen that the values of the expectations in (5.5) will remain unchanged if it is assumed that $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$ in this bivariate normal distribution and in the definitions (5.3) and (5.4) of β_ϕ and α_ϕ.

It now follows that

$$(5.6) \qquad E(\Sigma_{\phi \in \Phi} \beta_\phi \alpha_\phi) = E(\Sigma_i X_i Y_i) = n\rho.$$

Let

$$(5.7) \qquad \gamma = \frac{1}{n} E\left[\Sigma_{\phi \in \Phi} \beta_\phi^2 \alpha_\phi - (\Sigma_{\phi \in \Phi} \beta_\phi \alpha_\phi)^2 \right].$$

Then from (5.5), (5.6) and (5.7), we find that

$$i_n(\rho) = \frac{n(1+\rho^2)}{(1-\rho^2)^2}\left[1 - \frac{(1+\rho^2)}{(1-\rho^2)^2}\gamma\right]. \tag{5.8}$$

For any given values of $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_n$, it can be seen from equation (5.4) that $\alpha_\phi > 0$ for $\phi \in \Phi$ and $\Sigma_{\phi\in\Phi}\,\alpha_\phi = 1$. Hence, it is clear from (5.7) that $\gamma > 0$. Also, by evaluating $E(\Sigma_{\phi\in\Phi}\,\beta_\phi^2\alpha_\phi)$, we find that

$$\gamma = 1 + (n+1)\rho^2 - \frac{1}{n}E\left[(\Sigma_{\phi\in\Phi}\,\beta_\phi\alpha_\phi)^2\right]. \tag{5.9}$$

Now let

$$\gamma_1 = \frac{(1+\rho^2)\gamma}{(1-\rho^2)^2} \quad \text{and} \quad \gamma_2 = \frac{\rho^2\gamma}{1-\rho^2}. \tag{5.10}$$

Then, by using arguments similar to those just given we can show that the 3×3 submatrix on the lower diagonal of the Fisher information matrix based on a broken random sample is:

$$\mathbf{I}_{22} = \frac{n}{1-\rho^2}\begin{bmatrix} \sigma_1 & \sigma_2 & \rho \\ \frac{1}{\sigma_1^2}(2-\rho^2-\gamma_2) & \frac{1}{\sigma_1\sigma_2}(-\rho-\gamma_2) & -\frac{\rho}{\sigma_1}(1-\gamma_1) \\ & \frac{1}{\sigma_2^2}(2-\rho^2-\gamma_2) & -\frac{\rho}{\sigma_2}(1-\gamma_1) \\ & & \frac{1+\rho^2}{1-\rho^2}(1-\gamma_1) \end{bmatrix}. \tag{5.11}$$

It should be emphasized that we do not have n independent and identically distributed bivariate observations in this problem. Therefore, the conditions for the validity of the standard asymptotic properties of MLE's are not satisfied and $\mathbf{I}_{22}^{-1}$ is not necessarily the asymptotic covariance matrix of $\hat{\sigma}_1$, $\hat{\sigma}_2$, and $\hat{\rho}$. Nevertheless, the Fisher information matrix is useful as a measure of information in statistical inference because of its relationship to Kullback-Leibler information (see, e.g., Barndorff-Nielsen (1978), page 189).

Furthermore, the matrix $\mathbf{I}_{22}^{-1}$ for a broken sample and the corresponding matrix for an unbroken sample have some interesting features in common. Since the MLE's $\hat{\sigma}_1$ and $\hat{\sigma}_2$ based on a broken sample are the same as they would be for an unbroken sample, the asymptotic covariance matrix of $\hat{\sigma}_1$ and $\hat{\sigma}_2$ must be the same in both cases. Although we cannot interpret $\mathbf{I}_{22}^{-1}$ as an asymptotic covariance matrix, it can be shown that the 2×2 submatrix on the upper diagonal of $\mathbf{I}_{22}^{-1}$ is this asymptotic covariance matrix. Also, the third diagonal element of $\mathbf{I}_{22}^{-1}$ is larger than the asymptotic variance of the MLE of ρ for an unbroken random sample.

It can be shown that $i_n(\rho)$, as given by (5.8), is actually the Fisher information about ρ in a broken random sample when μ_1, μ_2, σ_1 and σ_2 are known. We shall conclude the paper with two simple properties of $i_n(\rho)$.

THEOREM 3. *$i_n(0) = 1$ for $n = 1, 2, \cdots$.*

PROOF. For $\rho = 0$, $\alpha_\phi = 1/n!$ for every $\phi \in \Phi$. Therefore,

$$\Sigma_{\phi \in \Phi} \beta_\phi \alpha_\phi = \frac{1}{n!} \Sigma_{\phi \in \Phi} \beta_\phi = \frac{1}{n} (\Sigma_i x_i)(\Sigma_i y_i). \tag{5.12}$$

Also, for $\rho = 0$,

$$E\left\{\left[\frac{1}{n}(\Sigma_i X_i)(\Sigma_i Y_i)\right]^2\right\} = 1. \tag{5.13}$$

The theorem now follows from (5.9) and (5.8). □

THEOREM 4. *For $-1 < \rho < 1$ and $n = 1, 2, \cdots$,*

$$i_n(\rho) \geqslant i_1(\rho) = \frac{1 + \rho^2}{(1 - \rho^2)^2}. \tag{5.14}$$

PROOF. Since $i_n(\rho)$ is the same as the Fisher information about ρ, when the means and variances are known, we can assume in this context that ρ is the only unknown parameter. Consider the statistic $(\Sigma_i X_i, \Sigma_i Y_i)$. The value of this statistic can be calculated from the values in the broken random sample. However, the Fisher information about ρ in this statistic is the same as $i_1(\rho)$, the Fisher information about ρ in a single pair of observations from a bivariate normal distribution. Since the Fisher information $i_n(\rho)$ in the entire broken random sample must be at least as large as the Fisher information in any statistic calculated from the broken sample, the inequality in (5.14) must be satisfied. The value of $i_1(\rho)$, given in (5.14), can easily be obtained. □

Theorem 4 shows that the information in a broken sample of size n is at least as great as that in a single bivariate observation. The exact behavior of $i_n(\rho)$ as $n \to \infty$ for a fixed value of $\rho \neq 0$ is not known to us.

Acknowledgments. We are indebted to Joseph B. Kadane, William F. Eddy, the referees and the Associate Editor for their helpful comments.

REFERENCES

BARNDORFF-NIELSEN, O. (1978). *Information, and Exponential Families in Statistical Theory.* Wiley, New York.

BERKSON, JOSEPH (1978). Do the marginal totals of the 2 × 2 table contain relevant information respecting the table proportions? *J. Stat. Plann. Inference* **2** 43–44.

CHEW, MILTON C., JR. (1973). On pairing observations from a distribution with monotone likelihood ratio. *Ann. Statist.* **1** 433–445.

DEGROOT, MORRIS H., FEDER, PAUL I. and GOEL, PREM K. (1971). Matchmaking. *Ann. Math. Statist.* **42** 578–593.

DeGroot, Morris H. and Goel, Prem K. (1976). The matching problem for multivariate normal data. *Sankhyā Ser. B* **38** 14–29.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

Goel, Prem K. (1975). On repairing observations in a broken random sample. *Ann. Statist.* **3** 1364–1369.

Good, I. J. (1976). On the application of symmetric Dirichlet distributions and their mixtures to contingency tables. *Ann. Statist.* **4** 1159–1189.

Hájek, Jaroslav, and Sidák, Zbyněk (1967). *Theory of Rank Tests.* Academic Press, New York.

Hardy, G. H., Littlewood, J. E. and Pólya, G. (1967). *Inequalities, 2nd ed.* Cambridge Univ. Press.

Kalbfleisch, John D. and Sprott, D. A. (1970). Applications of likelihood methods to models involving large numbers of parameters (with discussion). *J. Roy. Statist. Soc. Ser. B* **32** 175–208.

Kendall, M. G. and Stuart A. (1963). *The Advanced Theory of Statistics, Vol. 1, 2nd ed.* Hafner, New York.

Plackett, R. L. (1977). The marginal totals of a 2 × 2 table. *Biometrika* **64** 37–42.

Rubin, Herman (1976). Some fast methods of generating random variables with preassigned distributions I—general acceptance-rejection procedures. Mimeo Ser. #466, Depart. Statist., Purdue Univ.

Department of Statistics
Carnegie-Mellon University
Schenley Park
Pittsburgh, Pennsylvania 15213

Department of statistics
Purdue University
West Lafayette, Indiana 47907