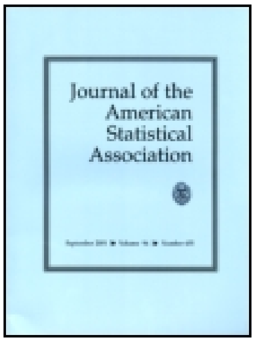



Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20


# Ordinal Measures of Association


William H. Kruskal



**To cite this article:** William H. Kruskal (1958) Ordinal Measures of Association, Journal of the American Statistical Association, 53:284, 814-861, DOI: 10.1080/01621459.1958.10501481

**To link to this article:** https://doi.org/10.1080/01621459.1958.10501481

# ORDINAL MEASURES OF ASSOCIATION*

WILLIAM H. KRUSKAL
*University of Chicago*

Ordinally invariant, i.e., rank, measures of association for bivariate populations are discussed, with emphasis on the probabilistic and operational interpretations of their population values. The three measures considered at length are the quadrant measure, Kendall's *tau*, and Spearman's *rho*. Relationships between these measures are discussed, as are connections between these measures and certain measures of association for cross classifications. Sampling theory is surveyed with special attention to the motivation for sample values of the measures. The historical development of ordinal measures of association is outlined.

CONTENTS



## 1. INTRODUCTION

What is meant by the degree of association or dependence between two random variables with a joint distribution? For example, what is meant by the degree of association between scores on two intelligence tests with respect to the population of seventh grade students in the United States today? Again, what is meant by the degree of association between 1955 income from wages and age among English wage earners?

Obviously the above questions do not have unique answers. There are infinitely many possible measures of association, and it sometimes seems that

* This research was supported in part by the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government. A large part of the work leading to this paper was done at the Department of Statistics, University of California, Berkeley.

I would like to thank the following persons for their suggestions and criticisms: R. Féron (University of Paris), M. Fréchet (University of Paris), C. Gini (University of Rome), H. D. Griffin (Lincoln, Nebraska), W. Hoeffding (University of North Carolina), H. S. Konijn (University of Sydney), H. Messinger (University of California), J. R. Rosenblatt (National Bureau of Standards), T. Salvemini (University of Rome), I. R. Savage (University of Minnesota), A. Stuart (London School of Economics and Political Science), R. M. Sundrum (University of Rangoon).

almost as many have been proposed at one time or another. On the other hand, it has been argued that, except in special cases, it is fatuous to attempt to represent the degree of association of a bivariate population by a single number. For example, see Guldberg [38].

The major purpose of the following remarks is to discuss the *probabilistic* or *operational interpretation* of several well-known measures of association, particularly those that are ordinally invariant in a sense to be defined. Although discussion of sample analogs and their distributions will be included, emphasis will be on the interpretation of measures of association supposing the population known. For it seems desirable to decide what we would mean by association if we knew the population of interest completely, before turning to the more complex questions of making inferences about measures of association from a sample. There is little point in estimating a population characteristic if the meaning of that characteristic is not clear.

The approach of this paper differs from the standard textbook approach, which begins with a sample, defines some intuitively reasonable sample measure of association, investigates its distribution, and—sometimes as an afterthought, but more often not at all—finally asks about the underlying population quantity.

Concern with the population measure is important for the applied statistician in other contexts than that of point estimation. For example, in testing the null hypothesis of independence via a sample measure of association, the kinds of dependence of most interest as alternative hypotheses should guide us in the choice of the appropriate test statistic, and understanding of the population characteristic estimated by the test statistic is often of great help.

Of course, if a strong structural assumption, typically that of bivariate normality, is made, the situation is very different. The ordinary correlation coefficient or its close relatives are natural measures of association in the bivariate normal case, and they have quite clear-cut interpretations. In this paper, however, I shall be concerned primarily with measures of association appropriate where strong structural assumptions are not present. That is, I shall be concerned with nonparametric measures of association.

It is important to recognize that the question, "Which single measure of association should I use?," is often unimportant. There may be no reason why two or more measures should not be used; the point I stress is that, whichever ones are used, they should have clear-cut population interpretations.

Much of the paper assumes no component-wise ties, or continuity of the marginals of the bivariate populations of interest. But the case of discrete populations is also considered. Although the paper is largely self-contained, it would be desirable for the reader to have some acquaintance with M. G. Kendall's monograph [51] and with the subject of measures of association in cross classifications as presented by Goodman and Kruskal [35] and [36].

To recapitulate: my purpose is to discuss operational interpretations for measures of association of bivariate populations, first supposing the populations known and only later supposing them unknown, with inferences to be made by sampling. The measures to be discussed will nearly all be ordinally invariant and will not presume detailed structure such as bivariate normality.

While the purpose of the paper is to provide a unified exposition of material already largely known, some methods and results may be novel.

The structure of the paper is as follows. Section 2 discusses as background the standard correlation ratio and correlation coefficient. Section 3 presents quadrant association and serves as a motivation for the sequence. Section 4 presents Kendall's $\tau$. Sections 5 and 6 suggest two approaches to Spearman's $\rho_S$. Sections 7 and 8 discuss inequalities between the three measures. Sections 10, 11, 12, and 13 consider estimation of the measures and questions of distribution for the estimators. Section 14 compares the measures. Sections 15 and 16 extend the measures to general distributions, considering in particular wholly discrete cases (cross classifications), and relate the cross classification analogs to the no-ties estimators. Section 17 presents historical material, including a discussion of some little-known early papers by Lipps and Deuchler, and in passing mentions some related topics not covered earlier.

## 2. THE CLASSICAL SECOND-MOMENT MEASURES

The ordinary (Pearson, product-moment) correlation coefficient, and the correlation ratios, retain some interest even when an assumption of normality is absent. They do not form part of the central subject matter of this paper, but they should be mentioned briefly for the sake of completeness and contrast.

Suppose, then, that we are given a specific bivariate distribution, expressed in terms of the pair of random variables $(X, Y)$. A classical measure of association is the correlation ratio of $Y$ on $X$, $\eta_{YX}$. This quantity may be defined in two different, but formally equivalent, ways. The first is

$$\eta_{YX}^2 = \frac{\text{Var } Y - E[\text{Var }(Y|X)]}{\text{Var } Y}, \tag{2.1}$$

i.e., $\eta_{YX}^2$ is the average relative reduction in the variance of $Y$, if we take $Y$ conditionally on $X$ given, over the unconditional or marginal variance of $Y$. The expectation in the numerator of (2.1) is with respect to $X$. The vertical stroke is that of conditional probability; I assume that the conditional distributions are defined. (For a detailed discussion of this point see Féron [30].)

Second,

$$\eta_{YX}^2 = \frac{\text{Var } [E(Y|X)]}{\text{Var } Y}, \tag{2.2}$$

or the variance of the conditional expectation of $Y$ given $X$, relative to the unconditional variance of $Y$. The numerator of the above expression is $E[E(Y|X)-EY]^2$, the expectation of the squared deviation of $E(Y|X)$ around the expected value of $Y$.

The possible values of $\eta_{YX}^2$ range from 0 to 1 inclusive. It is zero if and only if the conditional expectation of $Y$ given $X$ is the same for all values of $X$ (strictly, almost all). It is unity if and only if the conditional distribution of $Y$ given $X$ is concentrated on a single point which may depend on $X$, i.e. if and only if $Y$ is a function of $X$.

The remarks of the above paragraph apply equally to $\eta_{YX}^2$ and to its positive square root, $\eta_{YX}$, the correlation ratio itself. One may also switch coordinates and define $\eta_{XY}$ analogously. Naturally, all the indicated moments are assumed to exist, and Var $X$ and Var $Y$ are assumed positive to avoid rather trivial special cases.

The other classical measure of association is the correlation coefficient, $\rho_{XY}$, which may be defined as Cov $(X, Y)/\sqrt{\text{Var } X \cdot \text{Var } Y}$. Interpretations of $\rho_{XY}^2$, parallel to the above interpretations for the correlation ratios, may be given as follows. Let $\alpha+\beta X$ be the best linear estimate of $Y$ given $X$ in the least-square sense, i.e., $E[Y-a-bX]^2$ is minimized by $a=\alpha$ and $b=\beta$. Then $\beta=\text{Cov }(X, Y)/\text{Var } X$ and $\alpha=EY-\beta EX$.[1] It follows that

$$\rho_{XY}^2 = \frac{\text{Var } Y - E[Y - \alpha - \beta X]^2}{\text{Var } Y}, \tag{2.3}$$

and that

$$\rho_{XY}^2 = \frac{\text{Var }(\alpha + \beta X)}{\text{Var } Y} = \frac{E[\alpha + \beta X - EY]^2}{\text{Var } Y} \tag{2.4}$$

The first of these equations, (2.3), says that $\rho_{XY}^2$ is the average relative reduction in the squared deviation of $Y$ from its "best" linear estimate, relative to the marginal variance of $Y$. The second, (2.4), says that $\rho_{XY}^2$ is the variance of the "best" estimate of $Y$ based on $X$, relative to the variance of $Y$. A restatement is that $\rho_{XY}^2$ is the average squared deviation of the "best" fitting straight line from the over-all average of $Y$, all relative to Var $Y$. Since $\rho_{XY}$ is symmetrical in the two components, analogous relations with $X$ and $Y$ interchanged also hold. $\rho_{XY}$ can take values from $-1$ to 1 inclusive. It is equal to $\pm 1$ if and only if the joint distribution of $X$ and $Y$ is wholly concentrated on a straight line. It is equal to zero if $X$ and $Y$ are independent, but the converse does not in general hold.

It was, I believe, first pointed out by Maurice Fréchet [31] and [32] that $\eta_{YX}^2=\rho^2_{Y,E(Y|X)}$ and that

$$\rho_{XY} = \eta_{YX}\rho_{X,E(Y|X)} = \eta_{XY}\rho_{Y,E(X|Y)}.$$

These relations follow quite directly from the above definitions, and it then is immediately clear that $\rho_{XY}^2 \leqq \eta_{YX}^2$ and $\rho_{XY}^2 \leqq \eta_{XY}^2$, with equality holding if and only if the corresponding regression is linear.

An important distinction is that the $\eta$'s are invariant under rearrangement of the values of the 'independent' variable, while $\rho$ is not thus invariant. More precisely, if $Z=\psi(X)$, where $\psi$ is one to one and measurable, then $\eta_{YX}=\eta_{YZ}$. However the correlation coefficient between $X$ and $Y$ may be very different than that between $Z$ and $Y$. A more obvious, but still important, distinction is that the correlation ratios are asymmetrical measures of association, while the correlation coefficient is symmetric. All further measures of association to be discussed here will be symmetric.

The above interpretations of $\rho$ and the $\eta$'s are in terms of expected or average squared deviations. But probabilities are more basic than moments, and the two bear only weak relations to each other unless strong structural assumptions

[1] The same $\alpha$ and $\beta$ would result from minimization of $E[E(Y|X)-a-bX]^2$.

are made or unless a number of moments are given. Hence, in my opinion, the above interpretations of the classical measures may not be suitable if we want to apply our measures to general distributions. Exception could be taken to this dogma in certain cases when loss functions proportional to squared deviations exist.

Another kind of general interpretation for the correlation coefficient may be stated in terms of common components. It stems, to the best of my knowledge, from a 1912 article by J. C. Kapteyn [49]. Suppose that the structure of $X$ and $Y$ is such that we may write

$$X = U_1 + U_2 + \cdots + U_m + V_1 + \cdots + V_n$$
$$Y = U_1 + U_2 + \cdots + U_m + W_1 + \cdots + W_n$$

where the $U$'s, $V$'s, and $W$'s are all mutually uncorrelated and with the same variance. Then $\rho_{XY}$ is easily seen to be $m/(m+n)$, or the proportion of common components between $X$ and $Y$. One or two variations on this theme have been discussed more recently. This kind of interpretation seems to be useful only when it makes substantive sense to think of $X$ and $Y$ as having the above kind of overlapping additive structure, with the $U$'s, $V$'s, and $W$'s corresponding to quantities of substantive interest.

There is an enormous literature on the correlation coefficient and the correlation ratios. Much of this literature discusses interpretation of these measures of association only slightly or not at all. When interpretations are given, they are nearly always the same as those given above in terms of expected squared deviations. Uncritical use of the $\eta$'s and $\rho$ has been rightly criticized by many writers, and from many viewpoints. In this connection, it is particularly worth mentioning an investigation of the use and misuse of $\rho$ that was carried out under the leadership of M. Fréchet and sponsored by the International Statistical Institute. This investigation culminated in a 1935 article [33] containing comments by many eminent statisticians.

Some measures of association reflect aspects of *concordance* (greater values of $X$ go with greater values of $Y$), while other measures reflect aspects of *connection* that do not take the sense or direction into account. For example, $\rho$ is a measure of concordance while $\eta$ is a measure of connection. This distinction between connection and concordance, although perhaps difficult to make precise, is a useful one to bear in mind. It has been strongly stressed by Corrado Gini in his many writings on the subject. The ordinal measures of association that we shall now discuss in detail all reflect aspects of concordance.

## 3. QUADRANT ASSOCIATION AND THE QUANTITY $\varrho$

Perhaps the simplest measure of association between two random variables is one directly related to the sum of the probabilities in the first and third quadrants of some natural Cartesian coordinate system. If we call the pair of variables $(X, Y)$, and if we let $(x_0, y_0)$ be some fixed value of $(X, Y)$, then the quadrant measures of association are based on

$$\Pr\{(X > x_0 \text{ and } Y > y_0) \text{ or } (X < x_0 \text{ and } Y < y_0)\}, \tag{3.1}$$

where "Pr" means: probability of. The above quantity may conveniently be written

$$\Pr\{(X - x_0)(Y - y_0) > 0\}; \tag{3.2}$$

it is simply the probability that the deviations of $X$ and $Y$ from $x_0$ and $y_0$ respectively have the same signs, i.e., that $(X, Y)$ lies in the first or third quadrants around $(x_0, y_0)$. Note that our assumption of continuity implies that it is immaterial whether strong or weak inequalities are used in the above expressions.

What should $x_0$ and $y_0$ be? Any choice is somewhat arbitrary, but a rather natural one is to take $x_0$ as the median of $X$, Med $X$, and to take $y_0$ as the median of $Y$, Med $Y$. If these medians are not uniquely defined, any median value may be used, since nonuniqueness of definition simply means that there is some interval for $X$ (say) with zero probability, and such that $X$ lies to the left of the interval with probability $\frac{1}{2}$ and to the right with probability $\frac{1}{2}$. Hence any point in the interval, chosen as median, will give the same numerical value for (3.1) and (3.2).

Using the medians as $x_0$ and $y_0$ we may consider

$$\sigma_s = \Pr\{(X - \text{Med } X)(Y - \text{Med } Y) > 0\}, \tag{3.3}$$

or the probability that the deviations of $X$ and $Y$ from their respective medians have the same sign ($\sigma$ for sign; $s$ for same). It is clear that $\sigma_s$ takes values between 0 and 1 inclusive. It is 1 if and only if $X-\text{Med } X$ and $Y-\text{Med } Y$ are positive or negative together with probability one. It is zero if and only if $X-\text{Med } X$ and $Y-\text{Med } Y$ have different signs with probability one. If $X$ and $Y$ are independent, $\sigma_s$ is equal to $\frac{1}{2}$ (but the converse need not be true).

We may also consider

$$\sigma_d = \Pr\{(X - \text{Med } X)(Y - \text{Med } Y) < 0\}, \tag{3.4}$$

or the probability that $(X, Y)$ lies in the second or fourth quadrant of a coordinate system with origin at (Med $X$, Med $Y$). Equivalently, $\sigma_d$ is just the probability that $X-\text{Med } X$ and $Y-\text{Med } Y$ have different signs. Clearly $\sigma_s+\sigma_d=1$.

Using both $\sigma_s$ and $\sigma_d$, a natural and more symmetric quantity is the quadrant measure

$$\varphi = \sigma_s - \sigma_d = 2\sigma_s - 1, \tag{3.5}$$

the *difference* between the probabilities of same and different signs for the deviations of $X$ and $Y$ from their medians. If and only if $X$ and $Y$ always deviate from their medians in the same direction, $\varphi$ is 1; if and only if they always deviate in opposite directions, $\varphi$ is $-1$. If $X$ and $Y$ are independent $\varphi$ is zero (but the converse need not be true).

(Typographical note: I shall use Greek letters for population quantities, and corresponding Latin letters for their sample analogues. The letter "$q$" is so natural a symbol for (3.5) that I think it psychologically mandatory to use it, yet the standard Greek alphabet contains no "$q$". Older Greek, however, does have a "$q$," the letter *koppa*, which is here used and printed thus: ϙ. Although *koppa* was superseded in standard classical Greek by *kappa*, it remained in the Greek numbering system as a symbol for 90. It appears in several variant forms, some resembling the capital Latin "$G$" and some the lower case Latin "$q$".)

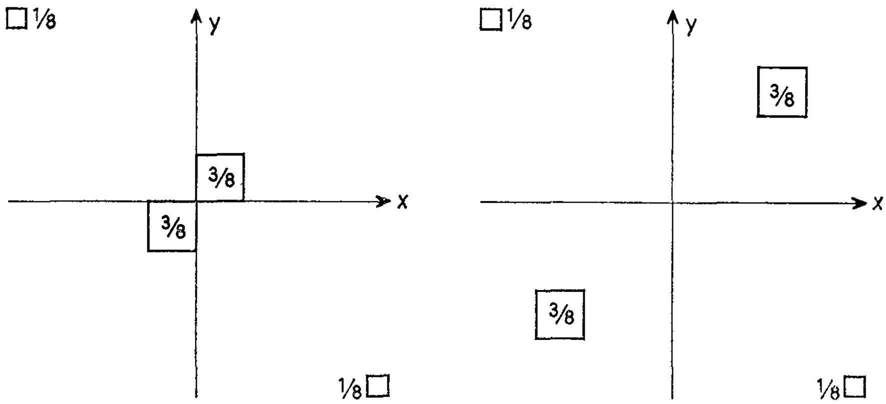
Thus ϙ provides a measure of association that lies in the customary range $-1$ to 1, takes the value 0 in the case of independence, and takes its extreme values in well-defined extreme cases. But one should not interpret intermediate numerical values of ϙ in the light of preconceptions about numerical values of other quantities that lie in the same range, for example the ordinary (product-moment) correlation coefficient. The interpretation of ϙ is precisely that of the difference between two probabilities as stated above. In the bivariate normal case there is a simple relation between the correlation coefficient, $\rho$, and ϙ: $\rho = \sin [(\pi/2)$ ϙ$]$. See, e.g., Cramér [12, p. 290]. Another name for ϙ is "the coefficient of medial correlation"; see Quenouille [71, Chapter 3]. It may also be noted that ϙ is the ordinary correlation coefficient between $\mathrm{sgn}(X - \mathrm{Med}\ X)$ and $\mathrm{sgn}(Y - \mathrm{Med}\ Y)$.

The meaning of ϙ may be expressed in terms of the following example: suppose that we are concerned with the association between two "intelligence" tests with respect to a given population. Scores on the two tests correspond to the random variables $X$ and $Y$. If we hypothetically choose an individual at random from the population of interest, then ϙ is the probability that that individual's two intelligence scores will both deviate from the respective population medians in the same direction, minus the probability that they will deviate in different directions.

Another interpretation is the following: suppose that you are a commodity speculator, betting in effect on the price of wheat in December on the basis of its price in September. Suppose, further, that when the price of wheat is above its long-run September median you bet that it will also be above its long-run December median for that year; whereas if it is below its September median, you bet that it will also be below its December median. Finally, in this grossly simplified market, suppose that each year you either win or lose \$1000 depending on whether your bet on the December median turned out correctly or not. Then your expected or average income is \$1000 ϙ.

Notice that ϙ, unlike the correlation coefficient, remains unchanged by monotone functional transformations of the coordinates: if, instead of $X$ and $Y$, we consider $f(X)$ and $g(Y)$, where $f$ and $g$ are both monotone strictly increasing (or both strictly decreasing) then ϙ is unchanged. If one of $f$, $g$ is strictly increasing and the other strictly decreasing, then the value of ϙ simply has its sign switched. Thus ϙ is an *ordinal* (i.e., ordinally invariant) measure of association; the same will be true of the other measures to be next considered.

The desirability of using ordinal measures for nonparametric work has been defended by some writers (see, for example, Wolfowitz [93, p. 104] and Hoeffding [40 and 41]), and most, but not all, statistical procedures that fall under the loose rubric of nonparametric analysis are invariant under wide classes of monotone transformations. (An important class of exceptions is the family of tests, first discussed by R. A. Fisher, based upon permutations of the observed sample). The consequences of using an ordinal measure of association between two random variables might however be anti-intuitive in some contexts. For example, consider the two distributions described graphically by Figure 821.

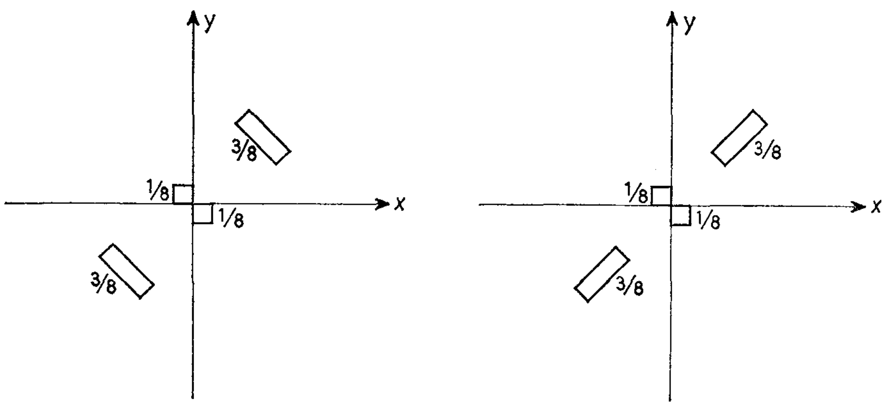



Distribution 1 Distribution 2

FIG. 821

(In Figure 821 the probability mass is thought of as smoothly spread over the indicated squares, so that they have the probabilities given by the attached fractions.) Both distributions would give rise to a value of $\frac{1}{2}$ for the quadrant measure $\rho$, although, if a metric were relevant, distribution 1 might seem intuitively to exhibit sharper association. Ordinal measures do have the advantage that they may be estimated from a sample in which numerical observations are lacking and one has only available the joint marginal orderings, or ranks.

It is obvious that the quadrant measure $\rho$ is not only ordinally invariant, but is also invariant under *any* transformations maintaining quadrant probabilities. For example, the two distributions described graphically by Figure 821a would both give rise to $\frac{1}{2}$ as the value of $\rho$, yet destribution 4 certainly seems to exhibit higher intuitive association than does distribution 3.




Distribution 3 Distribution 4

FIG. 821a

The measures of association next to be described may be doubly motivated by (1) the above lack of sensitivity of $\rho$ and (2) arbitrariness in the choice of $x_0$ and $y_0$ required for quadrant measures of association.

### 4. CONCORDANCE OF TWO "OBSERVATIONS" AND THE QUANTITY $\tau$

Perhaps the most natural way to avoid arbitrariness in the choice of $x_0$ and $y_0$ is to *average* $\Pr\{(X-x_0)(Y-y_0)>0\}$ over *all* $(x_0, y_0)$, with weights given by the joint distribution of $(X, Y)$ itself. Or perhaps we might average $\Pr\{(X-x_0)(Y-y_0)>0\}-\Pr\{(X-x_0)(Y-y_0)<0\}$. This kind of averaging amounts to considering the probability

$$\begin{aligned}\Pi_c &= \Pr\{(X_1 - X_2)(Y_1 - Y_2) > 0\} \\ &= \Pr\{X_1 > X_2 \text{ and } Y_1 > Y_2\} + \Pr\{X_1 < X_2 \text{ and } Y_1 < Y_2\}\end{aligned} \tag{4.1}$$

and its complement

$$\Pi_d = \Pr\{(X_1 - X_2)(Y_1 - Y_2) < 0\}. \tag{4.2}$$

Here $(X_1, Y_1)$ and $(X_2, Y_2)$ are taken as two independent bivariate random variables, each with the bivariate distribution under consideration. In other words, $\Pi_c$ is the probability that two hypothetical bivariate observations on the distribution of interest are *concordant*, in the sense that the two $x$ coordinates differ with the same sign as the two $y$ coordinates. $\Pi_d$ has a similar meaning but for *discordance:* different signs for the two differences.[2] Thus $\Pi_c$ is just $\sigma_s$ evaluated, not for $(X, Y)$, but for the *difference* between $(X, Y)$ and an identically distributed but independent bivariate random variable, i.e., evaluated for the distribution of $(X_1-X_2, Y_1-Y_2)$. This distribution will of course have both medians equal to zero.

To avoid possible confusion, note that the two observations mentioned above are not two observations of a sample from which we might want to estimate a measure of association, but rather are hypothetical observations about which we are entitled to think apart from any real sampling situation.

A convenient measure of association based on $\Pi_c$ and $\Pi_d$ is

$$\tau = \Pi_c - \Pi_d = 2\Pi_c - 1 = 1 - 2\Pi_d, \tag{4.3}$$

the difference between the probabilities of concordance and discordance for two observations on the distribution of interest. $\tau$ has, therefore, a direct and simple operational meaning. We also see that $\tau$ is $\rho$ for the distribution of $(X_1-X_2, Y_1-Y_2)$, or the correlation coefficient between the signs of $(X_1-X_2)$ and $(Y_1-Y_2)$. For this reason $\tau$ has sometimes been called the difference sign correlation. Hoeffding has called it [44] the difference sign covariance (since $\text{Var}[\text{sgn}(X_1-X_2)]=\text{Var}[\text{sgn}(Y_1-Y_2)]=1$.)

Several authors have independently proposed $\tau$, or its sample analogue, as a measure of association; the basic notion seems to derive from G. T. Fechner's

---

[2] These definitions may be restated as follows: $(X_1, Y_1)$ and $(X_2, Y_2)$ are concordant [discordant] as the line segment joining the two points has positive [negative] slope. It is immediate that concordance and discordance are ordinally invariant. The possibilities of zero or infinite slope may be neglected here, since marginal ties have, by assumption, zero probability. Later on the question of ties will be discussed.

work in 1897, and to have been first discussed in some detail by G. F. Lipps around 1905. The most recent independent proposal of $\tau$ is that of M. G. Kendall in 1938. Kendall gave a very thorough discussion of $\tau$ and its associated sampling theory; the measure is sometimes called Kendall's $\tau$. More detailed historical remarks about this and the other measures appear in section 17 at the end of this paper.

If the bivariate distribution is normal, $\tau$ is related to the standard (Pearson product-moment) correlation coefficient, $\rho$, by the formula: $\rho = \sin[(\pi/2)\ \tau]$. This is a direct consequence of the analogous formula for $\rho$, and the observation that the correlation coefficient between $X_1 - X_2$ and $Y_1 - Y_2$ is the same as that between $X_1$ and $Y_1$.

From its definition, $\tau$ is ordinally invariant. It lies between $-1$ and 1 inclusive, taking $\pm 1$ as its value if and only if all the probability mass lies on the graph of an increasing or decreasing function respectively. If $X$ and $Y$ are independent, $\tau = 0$, but the converse is not in general true.

A rewording of the interpretation of $\tau$ is the following. Suppose that observations $(X_1, Y_1)$ and $(X_2, Y_2)$ are drawn but that only $X_1$ and $X_2$ are revealed to us at first by some sort of mechanical device. Suppose further that we agree to play a game wherein we predict $Y_1 < Y_2$ when $X_1 < X_2$, and $Y_1 > Y_2$ when $X_1 > X_2$. If our prediction turns out to be correct we win one dollar; if wrong, we lose one dollar. After prediction, the mechanical device reveals $Y_1$ and $Y_2$ and the pay-off is made. Our expected gain in dollars is $\tau$.

We may consider another example. Suppose that we are interested in the degree of association between two "intelligence" tests for some very large specified population. Think of taking two individuals at random from the population and comparing their scores on the two tests. $\Pi_c$ is the probability that the more "intelligent" according to one test is also the more "intelligent" according to the other. $\Pi_d$ is the probability that the orderings differ. And $\tau$ is just the difference of these two probabilities. For this description to correspond perfectly with our prior more abstract discussion, it must be supposed that the population is infinitely large and that there is zero probability of ties.

### 5. ANOTHER METHOD OF AVERAGING AND THE QUANTITY $\rho_S$

In the prior section we averaged $\Pr\{(X - x_0)\ (Y - y_0) > 0\}$ over all values of $(x_0, y_0)$ weighted according to the distribution of $(X, Y)$ itself. However, it might be reasonable to average with respect to the *marginal* distributions, taken independently. In other words, we might consider $\Pr\{(X - X')(Y - Y'') > 0\}$ where $X'$ has the distribution of $X$, $Y''$ has the distribution of $Y$, *and* where the pair $(X, Y)$ and the two single variates, $X'$ and $Y''$, are all three independent.

To rephrase this slightly, suppose that we take *three* hypothetical independent observations from the bivariate distribution of interest: $(X_1, Y_1)$, $(X_2, Y_2)$, and $(X_3, Y_3)$. Then consider the probability of concordance between $(X_1, Y_2)$ and the "crossed observation" $(X_2, Y_3)$,

$$\iota_c = \Pr\{(X_1 - X_2)(Y_1 - Y_3) > 0\}. \tag{5.1}$$

(Clearly we could have used $(X_3, Y_2)$ instead of $(X_2, Y_3)$ without affecting the above quantity.) Similarly consider the probability of discordance between $(X_1, Y_2)$ and $(X_2, Y_3)$, $\iota_d$, and subtract it from the probability of concordance to obtain

$$\begin{aligned} \iota_c - \iota_d &= \Pr\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - \Pr\{(X_1 - X_2)(Y_1 - Y_3) < 0\} \\ &= 2\Pr\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - 1 = 2\iota_c - 1. \end{aligned} \tag{5.2}$$

This quantity thus has a direct interpretation, although not quite so direct as $\tau$. Under independence between $X$ and $Y$ it is zero, but not conversely in general. However, its minimum and maximum values are $-\frac{1}{3}$ and $+\frac{1}{3}$. These are taken on if and only if $Y$ is a strictly monotone function of $X$, decreasing for the minimum and increasing for the maximum.

To see this, note that all the quantities considered are ordinally invariant; hence we may perform probability integral transformations on the component random variables without affecting any of the above quantities. This means that we replace each $X_i$ by $X_i^* = F(X_i)$, and each $Y_i$ by $Y_i^* = G(Y_i)$, where $F$ is the cumulative distribution function of $X$, and $G$ that of $Y$. When we do this, $X_i^*$ and $Y_i^*$ have marginal distributions uniform between 0 and 1. Then

$$\begin{aligned} \iota_c &= \Pr\{(X_1 - X_2)(Y_1 - Y_3) > 0\} = \Pr\{(X_1^* - X_2^*)(Y_1^* - Y_3^*) > 0\} \\ &= E[\Pr\{(X_1^* - X_2^*)(Y_1^* - Y_3^*) > 0 \mid (X_1^*, Y_1^*)\}] \\ &= E[X_1^*Y_1^* + (1 - X_1^*)(1 - Y_1^*)] \\ &= 2E[X_1^*Y_1^*] = 2\,\mathrm{Cov}(X_1^*, Y_1^*) + 1/2 \\ &= (2/12)\ (\text{corr. coeff. bet. } X_1^*, Y_1^*) + 1/2. \end{aligned} \tag{5.3}$$

Since the correlation coefficient can take all values between $-1$ and 1, and only those values, we see that $\iota_c$ can take all values between $\frac{1}{3}$ and $\frac{2}{3}$. Hence $2\iota_c - 1$ can take all values between $-\frac{1}{3}$ and $\frac{1}{3}$ and only those values. (The correlation coefficient between $X_i^*$ and $Y_i^*$ has been called by K. Pearson the *grade correlation coefficient* between $X$ and $Y$.)

In the above manipulations, the transition from the second to the third line follows from the fact that $(X_2^*, Y_3^*)$ has the uniform distribution over the unit square. The symbol "Cov" means: covariance of. Further, we make use of the facts that the mean and variance of the uniform distribution over the unit interval are $\frac{1}{2}$ and $\frac{1}{12}$ respectively. The vertical stroke in the second line is that of conditional probability.

In order to shift our measure to a scale running from $-1$ to 1 in the conventional way, it is convenient to multiply by 3, thus obtaining finally

$$\rho_S = 3(\iota_c - \iota_d) = 6\iota_c - 3 = 3 - 6\iota_d. \tag{5.4}$$

This is the population analogue—in one natural sense at least—of the so-called Spearman rank correlation coefficient, whence the subscript $S$. $\rho_S$ is equal to the grade correlation coefficient between $X$ and $Y$.

Clearly $\rho_S$ is ordinally invariant. It takes the value $\pm 1$ just when all the probability mass is on the graph of an increasing or decreasing function respectively. If $X$ and $Y$ are independent, $\rho_S$ is zero, but not conversely in general.

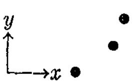
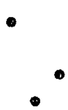
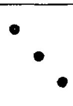
### 6. ANOTHER APPROACH TO $\rho_S$

Another point of view towards $\rho_S$ may be adopted that in some ways seems more natural. In particular, it obviates introduction of the arbitrary factor 3 as in section 5. Suppose that we consider, as before, three hypothetical independent observations on $(X, Y)$ and ask for the probability that *at least* one of the three is concordant with *both* the other two. Let us call this quantity $\omega_c$, and call the corresponding quantity for discordance $\omega_d$. Clearly $\omega_c + \omega_d = 1$. What is the relation between $\omega_c$ and $\iota_c$?

The three bivariate observations may occur in six different patterns, viewing them without regard to which is numbered observation 1, which 2, and which 3. These patterns are given in Table 825.

TABLE 825

THE SIX DIFFERENT PATTERNS OBTAINABLE WHEN DRAWING THREE OBSERVATIONS (WITHOUT TIES) FROM A BIVARIATE POPULATION

| Pattern picture | Description | $y$ permutation | In chance event whose probability is $\omega_c$? | How many of six equally likely assignments of names lead to chance event $(X_1-X_2)(Y_1-Y_3)>0$? |
|---|---|---|---|---|
|  | All conc. | 123 | Yes | 4 |
|  | Lower left conc. with two others, but two others not conc. | 132 | Yes | 4 |
|  | Upper right conc. with two others, but two others not conc. | 213 | Yes | 4 |
|  | Upper left disc. with two others, and two others conc. | 312 | No | 2 |
|  | Lower right disc. with two others, and two others conc. | 231 | No | 2 |
|  | All disc. | 321 | No | 2 |

The first column shows the picture, and the second column gives its verbal description. The third column gives a convenient description of the pattern in terms of the permutations of the integers 1, 2, and 3; let "1" be assigned to the dot with smallest $y$ coordinate, "2" to the dot with next smallest, and "3" to the dot with greatest $y$ coordinate. Then write down in order of $x$ coordinate magnitude the three integers. For example, in the second line, the dot with smallest $x$ coordinate is numbered 1, since it has smallest $y$ coordinate; thus "1" comes first. The dot with middle $x$ coordinate is numbered 3 since it has the greatest $y$ coordinate, etc. This kind of description has proved quite useful.

The fourth column simply says whether or not each pattern falls into the chance event: at least one $(X_i, Y_i)$ concordant with the other two. If we look at the number of inversions of neighboring integers necessary to achieve the primary ordering 1, 2, 3 for the $y$ permutations of the third column, we see that a pattern falls into the above chance event just when its $y$ permutation requires 0 or 1 inversions to achieve primary ordering. A pattern *fails* to fall in the above chance event just when its $y$ permutation requires 3 or 4 inversions to achieve the primary ordering.

Now for each pattern, there are six actual orderings of the three observations, for there are six ways of numbering the dots "1," "2," and "3." Given the pattern, each of these six ways is equally likely, since the observations are independent and identically distributed. The last column of Table 825 shows how many of these six ways lead to satisfaction of the chance event $(X_1-X_2)(Y_1-Y_3)>0$. For example, on the first line, either of the two numberings in which the lower left point is $(X_1, Y_1)$ satisfies this chance event. Also either of the two numberings in which the upper right point is $(X_1, Y_1)$ satisfies it. But if the middle point is $(X_1, Y_1)$, then $(X_1, Y_1)$ (and $(X_2, Y_3)$ are *discordant*. Hence $2+2+0=4$ appears in the last column.

It follows that

$$\iota_c = \Pr\{(X_1 - X_2)(Y_1 - Y_3) > 0\} = \frac{4}{6}\,\omega_c + \frac{2}{6}\,\omega_d$$

$$= \frac{2}{3}\,\omega_c + \frac{1}{3}\,(1 - \omega_c) \tag{6.1}$$

$$= \frac{1}{3}\,(\omega_c + 1).$$

Quite analogously,

$$\Pr\{(X_1 - X_2)(Y_1 - Y_3) < 0\} = \frac{1}{3}(\omega_d + 1) \tag{6.2}$$

so that

$$\rho_S = \omega_c - \omega_d. \tag{6.3}$$

This is perhaps the most natural definition of $\rho_S$: the difference between the probabilities (among three observations) that (a) at least one will be concordant with the other two and (b) at least one will be discordant with the other two.

In the case of a normal bivariate distribution, $\rho_S$ and $\rho$ (the correlation coefficient) are related by the equation

$$\rho = 2 \sin \left[ \frac{\pi}{6} \rho_S \right]. \tag{6.4}$$

This is easily derived as follows. If $(X, Y)$ is a hypothetical observation from our normal distribution of interest and $(X', Y')$ is an independent observation from the bivariate normal distribution with the same marginals as $(X, Y)$ but with independent coordinates, then

$$\rho_S = 3\Pr\{(X - X')(Y - Y') > 0\} - 3\Pr\{(X - X')(Y - Y') < 0\}. \tag{6.5}$$

Without loss of generality, let $X$, $Y$, $X'$, and $Y'$ have zero means and unit variances. The correlation coefficient between $X$ and $Y$ is $\rho$; that between $X'$ and $Y'$ is zero. Then $X-X'$ and $Y-Y'$ have a bivariate normal distribution with zero means, variances 2, and correlation coefficient $\rho/2$. It follows from our earlier discussion of the quadrant measure, applied to the distribution of $(X-X', Y-Y')$, that

$$\begin{aligned} \Pr\{(X - X')(Y - Y') > 0\} - \Pr\{(X - X')(Y - Y') < 0\} \\ = \frac{2}{\pi} \arcsin (\rho/2). \end{aligned} \tag{6.6}$$

From this, (6.4) follows immediately.

Interpretations of $\rho_S$, similar to those given at the end of sections 3 and 4 for $\rho$ and $\tau$, present no difficulty of statement, but are not so directly intuitive, since they involve three hypothetical observations, rather than one or two.

## 7. RELATIONS BETWEEN $\tau$ AND $\rho_S$

The probabilities of the six patterns of Table 825 must of course sum to unity; but there are other relations between them that give rise to inequalities between $\tau$ and $\rho_S$.

The first of these was pointed out by H. E. Daniels [14]; it is

$$-1 \leqq 3\tau - 2\rho_S \leqq 1. \tag{7.1}$$

A direct proof follows. Denote by $p_{123}$, $p_{123}$, etc., the probabilities of the six patterns of Table 825. Then

$$\begin{aligned} \Pi_c &= p_{123} + \tfrac{2}{3}(p_{132} + p_{213}) + \tfrac{1}{3}(p_{312} + p_{231}) \\ \omega_c &= p_{123} + p_{132} + p_{213}. \end{aligned} \tag{7.2}$$

The second part of (7.2) is immediate from Table 825; the first part follows by noting that the probability of concordance between $(X_1, Y_1)$ and $(X_2, Y_2)$, given some one of the six patterns, is just one-third the number of concordant pairs out of the three possible unordered pairs of points in the pattern. From (7.2) it follows that $3\Pi_c - 2\omega_c = p_{123} + p_{312} + p_{231}$, whence $0 \leqq 3\Pi_c - 2\omega_c \leqq 1$, whence Daniels' inequality by substituting $\Pi_c=(1+\tau)/2$ and $\omega_c=(1+\rho_S)/2$.

It is clear from the above that the right (left) side of Daniels' inequality is achieved only when $p_{123}+p_{312}+p_{231}=1$ (0).

The second inequality, or strictly pair of inequalities, was demonstrated by Durbin and Stuart [25]; it is

$$\begin{aligned} 1+\rho_S &\geqq (1+\tau)^2/2 \\ 1-\rho_S &\geqq (1-\tau)^2/2. \end{aligned} \tag{7.3}$$

By making the substitutions mentioned at the end of the next-to-last paragraph, it is easily seen that (7.3) is equivalent to

$$\begin{aligned} \omega_c &\geqq \Pi_c^2 \\ \omega_d &\geqq \Pi_d^2. \end{aligned} \tag{7.4}$$

A brief proof of the first part of (7.4) follows; the second part may be thrown back on the first by noting that $\omega_c$ and $\Pi_c$ for the distribution of $(X, -Y)$ are the same respectively, as $\omega_d$ and $\Pi_d$ for the distribution of $(X, Y)$.

Consider, then, three independent observations on the distribution of $(X, Y)$: $(X_i, Y_i)$, $i=1, 2, 3$. Define the random variable

$$Z = Z(X_1, Y_1) = \Pr\{(X_2, Y_2) \text{ concordant with } (X_1, Y_1) \mid (X_1, Y_1)\} \tag{7.5}$$

where the stroke is that of conditional probability. $Z$ is the random version of (3.1). Since Var $Z\geqq 0$, we have $E(Z^2)\geqq(EZ)^2$ or

$$\begin{aligned} E[\Pr\{(X_2, Y_2) &\text{ concordant with } (X_1, Y_1) \mid (X_1, Y_1)\} \\ &\Pr\{(X_3, Y_3) \text{ concordant with } (X_1, Y_1) \mid (X_1, Y_1)\}] \geqq (EZ)^2. \end{aligned} \tag{7.6}$$

Now $EZ=\Pi_c$, and the left-hand side of (7.6), which we may call $\Pi_{cc}$, is the probability that of three independent numbered observations, the second and third are both concordant with the first. Hence

$$\Pi_{cc} \geqq \Pi_c^2. \tag{7.7}$$

Next, observe that

$$\omega_c = p_{123} + p_{132} + p_{213} \geqq p_{123} + \tfrac{1}{3}p_{132} + \tfrac{1}{3}p_{213} = \Pi_{cc}. \tag{7.8}$$

The only novelty in (7.8) is the right-most equality; this follows, as did (7.2), from an immediate examination of the conditional probabilities of the chance event whose probability is $\Pi_{cc}$, given each of the six patterns of Table 825. Finally, putting (7.7) and (7.8) together, we have the desired first part of (7.4).

Note that equality of this first part of (7.4), i.e., equality of $\omega_c$ and $\Pi_c^2$ requires that $Z(X, Y)=\Pi_c$ identically, except for a set of $(X, Y)$ values having zero probability, and also that $p_{132}=p_{213}=0$.

It is interesting to exhibit the restrictions of (7.1) and (7.3) graphically. Figure 829 is a graph, in the $(\tau, \rho_S)$ plane, showing values of $(\tau, \rho_S)$ that are precluded by (7.1), (7.3), or both.

From the following figure, or by the corresponding algebraic manipulations, we see that (7.1) and (7.3) together give

$$-1 + (1+\tau)^2/2 \leqq \rho_S \leqq (1+3\tau)/2, \qquad \tau \leqq 0 \tag{7.9n}$$

$$(-1+3\tau)/2 \leqq \rho_S \leqq 1 - (1-\tau)^2/2, \qquad \tau \geqq 0 \tag{7.9p}$$

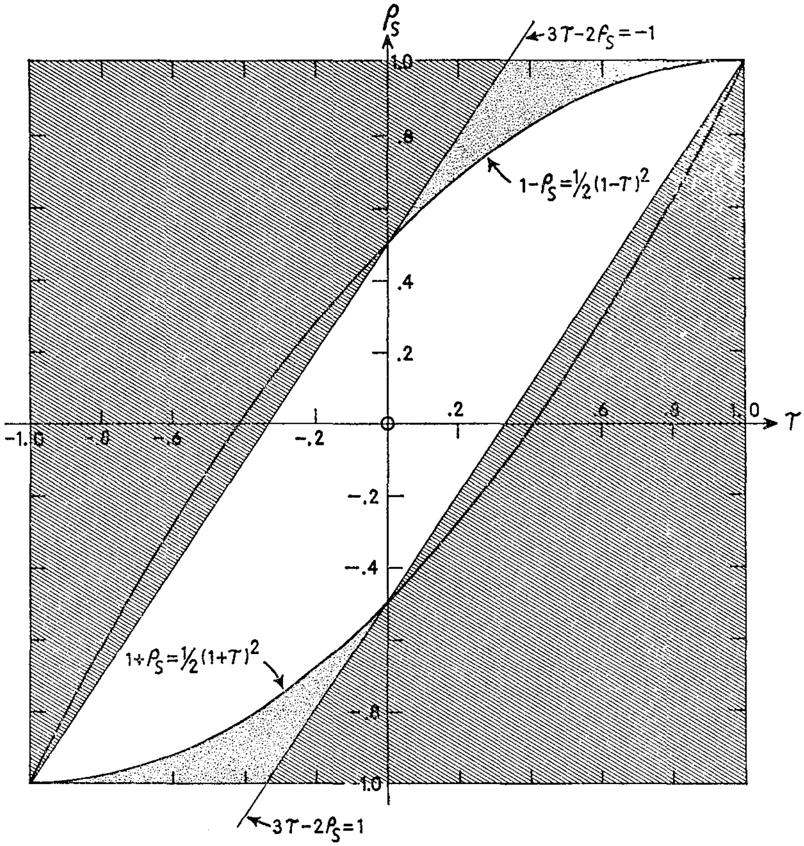



FIG. 829. Restrictions on $\tau$ and $\rho_S$ from inequalities (7.1) and (7.3).

Line shading: impossible values from (7.1).
Dot shading: impossible values in addition from (7.3).

Finally, a few comments as to the extent to which these bounds may be achieved, i.e., as to whether the inequalities of (7.9) are the best possible.

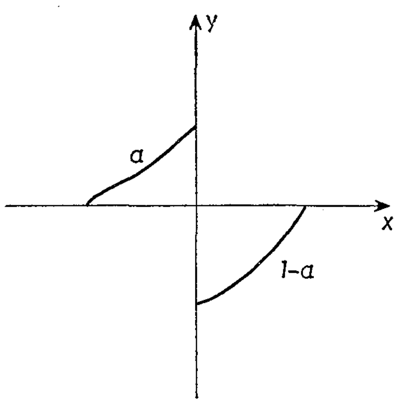



FIG. 829a

Consider distributions of the kind indicated by Figure 829a. Here all the mass is concentrated continuously on two monotone curves as shown, the upper left one having probability $a$ in toto, and the lower right one having probability $1-a$. $0 \leqq a \leqq 1$. For such a distribution one easily computes

$$\Pi_c = a^2 + (1-a)^2, \qquad \omega_c = a^3 + (1-a)^3,$$
$$\tau = 4a^2 - 4a + 1, \qquad \rho_S = 6a^2 - 6a + 1,$$

whence

$$\frac{1-\tau}{2} = \frac{1-\rho_S}{3},$$

or

$$\rho_S = (-1 + 3\tau)/2.$$

Since $0 \leqq a \leqq 1$, $\tau$ (in this case) runs from 0 to 1 inclusive. Hence, for $\tau \geqq 0$, the left side of (7.9p) may be achieved. Similarly, considering distributions of the form of Figure 830, for $\tau \leqq 0$, the right side of (7.9n) may be achieved. In short,

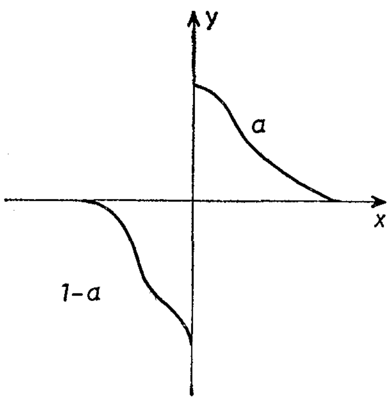



FIG. 830

the straight line boundaries of the unshaded region in Figure 829 are the best possible.

As to the parabolic boundaries, the situation seems unclear. They are achieved for special values of $\tau$, namely values of $\tau$ of form $\pm(1-(2/m))$, $m=1, 2, 3, \cdots$. Consider, for given integral $m$, a distribution of the form of Figure 831, in which the probability mass is entirely on $m$ monotone decreasing curve segments as shown, the segments themselves falling in a sequence of concordant rectangles. (In Figure 831 the dashed lines are guides to the eye only.) Each segment has probability $1/m$ spread on it in any continuous manner. Clearly

$$\Pi_c = 1 - \Pi_d = 1 - m/m^2 = (m-1)/m$$
$$\omega_c = 1 - \omega_d = 1 - m/m^3 = (m^2-1)/m^2$$

so that

$$\tau = 1 - (2/m), \qquad \rho_S = 1 - (2/m^2)$$

and

$$1 - \rho_S = 2/m^2 = (1 - \tau)^2/2.$$

Thus, for $\tau=0, \frac{1}{3}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{5}{7}, \cdots, 1-(2/m), \cdots$, the right side of (7.9p) may be achieved. A similar argument shows that for $\tau=0, -\frac{1}{3}, -\frac{1}{2}, \cdots,$

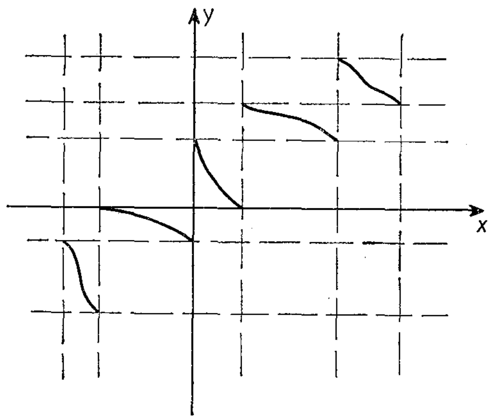



FIG. 831. (Case $m=5$).

$-1+(2/m), \cdots$, the left side of (7.9n) may be achieved. So far as I know, the question of best bounds for $\rho_S$, when $\tau$ has other values than the above, is unresolved.

### 8. RELATIONS BETWEEN $\varrho$ AND $\tau$, AND BETWEEN $\varrho$ AND $\rho_S$

Suppose that $\varrho$ has some given value; how large and how small can $\tau$ be? It is easy to see that, if $\varrho$ is given, the first and third quadrants around (Med $X$, Med $Y$) must each have probability $(1+\varrho)/4$, while the second and fourth quadrants must each have probability $(1-\varrho)/4$. In order to make $\Pi_c$ (whence $\tau$) as large as possible we clearly want the mass concentrated on monotone increasing curves within each quadrant. A further argument, based on the two "observations" entering into the definition of $\Pi_c$, when they fall in different quadrants, shows that $\Pi_c$ is maximum when the joint distribution has the general appearance of Figure 832.

For such a distribution $\Pi_d=2[(1-\varrho)/4]^2=(1-\varrho)^2/8$ so that $\tau=1-(1-\varrho)^2/4$. A similar argument shows that the minimum value of $\tau$ is $(1+\varrho)^2/4-1$. Hence we have the inequality

$$(1 + \varrho)^2/4 - 1 \leqq \tau \leqq 1 - (1 - \varrho)^2/4 \tag{8.1}$$

which is best possible. Figure 832a shows the relationship between $\tau$ and $\kappa$ expressed by (8.1).

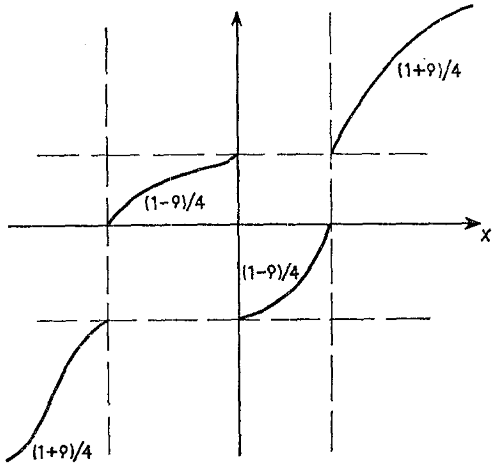



FIGURE 832. Form of distribution that maximizes $\Pi_c$ for given $\rho$. Dashed lines are guides to eye.

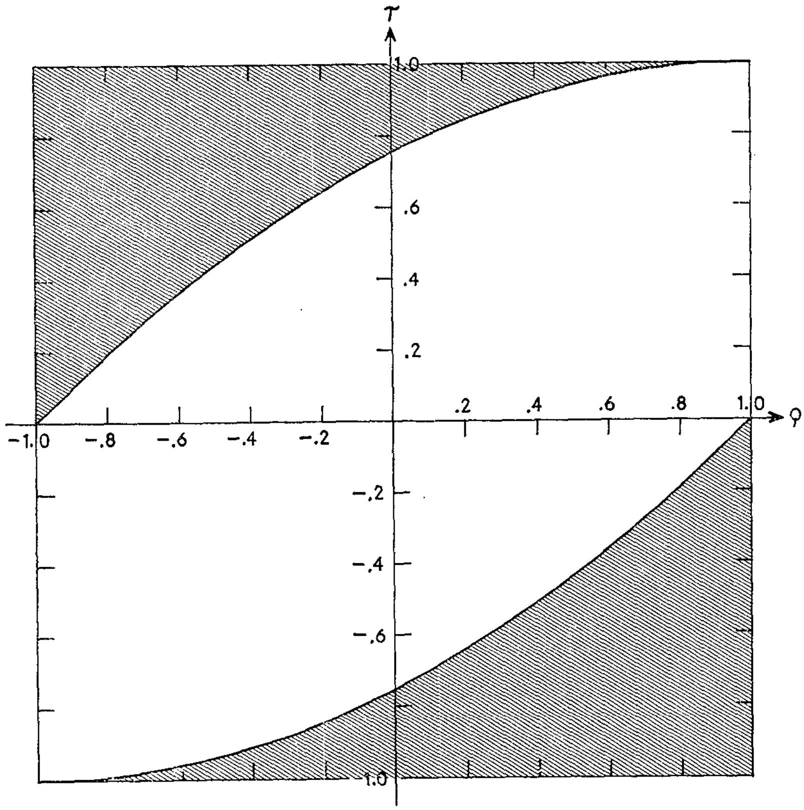



FIG. 832a. Inequalities between $\rho$ and $\tau$. Shaded area excluded.

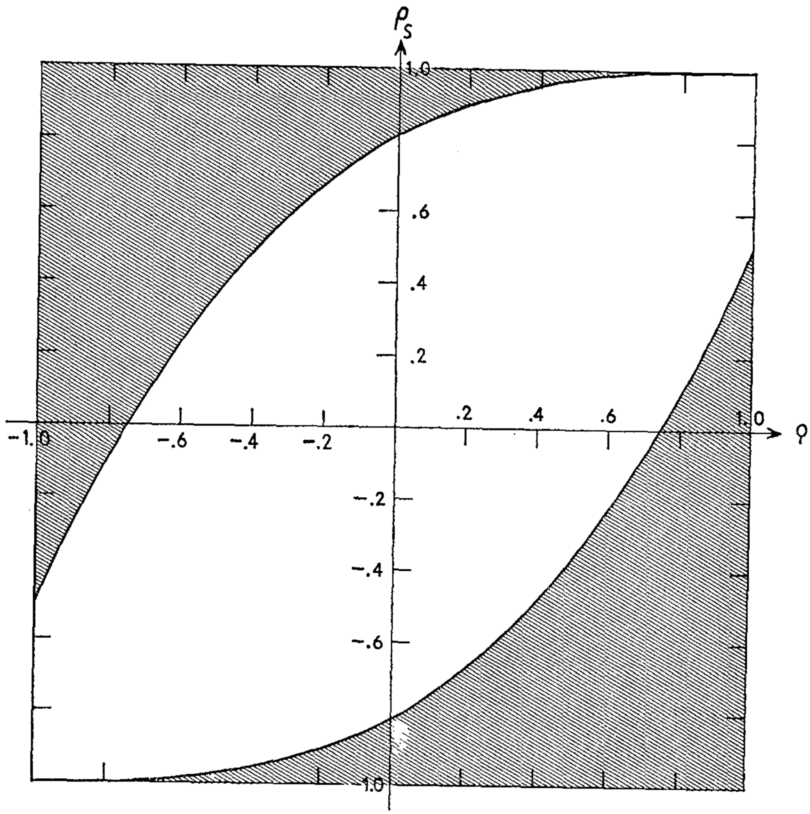



FIG. 833. Inequalities between $\varrho$ and $\rho_S$. Shaded areas excluded.

Now for similar work as to $\varrho$ and $\rho_S$. A little reflection shows that a distribution like that of Figure 832 maximizes $\rho_S$ for given $\varrho$. For such a distribution

$$\omega_d = 6[(1 - \varrho)^3/64]$$

so that

$$\rho_S = 1 - \frac{3}{16}(1 - \varrho)^3.$$

If the complementary computation be made, we find that

$$\frac{3}{16}(1 + \varrho)^3 - 1 \leqq \rho_S \leqq 1 - \frac{3}{16}(1 - \varrho)^3 \tag{8.2}$$

gives the best possible inequality between $\varrho$ and $\rho_S$. Figure 833 exhibits it.

## 9. BIVARIATE NORMAL DISTRIBUTIONS AND A ONE-PARAMETER NON-NORMAL FAMILY

It may be worth recapitulating the values of $\varrho$, $\tau$, and $\rho_S$ for bivariate normal distributions. They depend only on the correlation coefficient, $\rho$, and we have shown that

$$\varrho = \tau = \frac{2}{\pi} \sin^{-1} \rho \qquad \text{(Sections 3 and 4)}$$

$$\rho_S = \frac{6}{\pi} \sin^{-1} (\rho/2) \qquad (6.4)$$

Graphs of these two functions appear in Figure 834 for non-negative values of $\rho$. For negative values of $\rho$, only the signs of $\varrho$, $\tau$, and $\rho_S$ are changed.

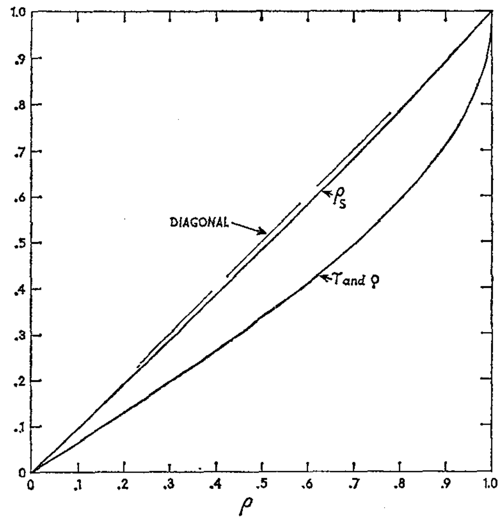



FIG. 834. Relations between $\rho$, $\varrho$, $\tau$, and $\rho_S$ for bivariate normal distributions.

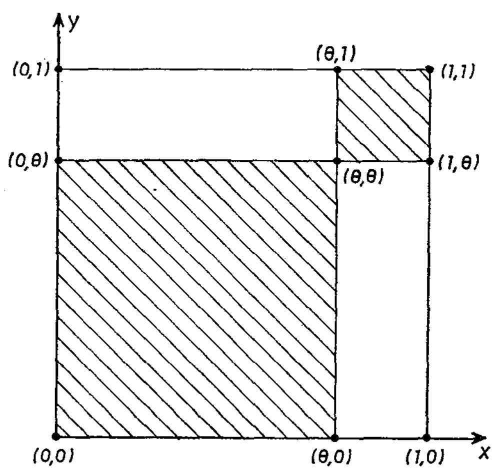



FIG. 834a. Distribution of kind discussed in section 9 Density function is $1/[\theta^2+(1-\theta)^2]$ in shaded area, and zero elsewhere.

Further insight into the nature of $\rho$, $\tau$, and $\rho_S$ may be obtained by computing their values for the members of a simple one-parameter family of distributions. As an illustration consider the one-parameter family of distributions within the unit square for which the probability mass is uniformly spread within the two squares $((0, 0), (\theta, 0), (\theta, \theta), (0, \theta))$ and $((\theta, \theta), (1, \theta), (1, 1), (\theta, 1))$. Here $0 \leqq \theta \leqq 1$. Such a distribution is pictured in Figure 834a. If we let $A(\theta) = \theta^2 + (1-\theta)^2$ then it is easily computed that the marginal density function of both $X$ and $Y$ is

$$\begin{array}{ll} \theta/A(\theta) & 0 \leqq x \text{ (or } y) \leqq \theta \\ (1-\theta)/A(\theta) & \theta \leqq x \text{ (or } y) \leqq 1 \\ 0 & \text{elsewhere} \end{array} \tag{9.1}$$

Further Med $X =$ Med $Y = 1 - [A(\theta)]/[2(1-\theta)]$ when $\theta \leqq \frac{1}{2}$ and $A(\theta)/(2\theta)$ when $\theta \geqq \frac{1}{2}$. It is then readily computed that

$$\rho = \begin{cases} \theta^2/(1-\theta)^2 & \theta \leqq \frac{1}{2} \\ (1-\theta^2)/\theta^2 & \theta \geqq \frac{1}{2} \end{cases} \tag{9.2}$$

$$\tau = 2\theta^2(1-\theta)^2/A^2(\theta) \tag{9.3}$$

$$\rho_S = 3\theta^2(1-\theta)^2/A^2(\theta) = \tfrac{3}{2}\tau. \tag{9.4}$$

Graphs of these functions (all symmetrical about $\theta = \frac{1}{2}$) are given in Fig. 835.

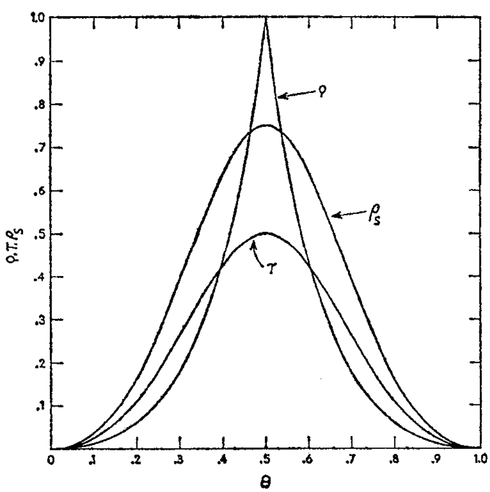



FIG. 835. Graphs of $\rho$, $\tau$, and $\rho_S$ as functions of $\theta$ for family of distributions discussed in section 9.

## 10. ESTIMATION OF THE QUADRANT MEASURE

Suppose now that we have a random sample of $n$, $(X_1, Y_1)$, $(X_2, Y_2)$, $\cdots$, $(X_n, Y_n)$, from a bivariate distribution with continuous marginals, and that we wish to estimate the quadrant measure $\rho$ for this distribution. The natural procedure is to find the sample medians and to let the estimator be

$$
\begin{aligned}
q = \frac{1}{n} & \text{ [No. } (X_i, Y_i)\text{'s in first and third quadrants around sample medians]} \\
& - \frac{1}{n} \text{ [No. } (X_i, Y_i)\text{'s in second and fourth quadrants around sample medians]}
\end{aligned} \tag{10.1}
$$

If $n$ is even, (10.1) defines a sample statistic unambiguously. If $n$ is odd, some slight modification is needed, since one or two sample points will have their $x$ or $y$ values exactly equal to the corresponding medians.

Blomqvist [10] has made the definition precise and has discussed the distribution of $q$. His method of making (10.1) precise is the following:

> If $n$ is even, let $n_1$ and $n_2$ be the integers in the two square brackets of (10.1) respectively.
>
> If $n$ is odd, and $(X_m, Y_m)$ furnishes the sample median for both $x$ and $y$, neglect $(X_m, Y_m)$ in counting the quadrant numbers, $n_1$ and $n_2$.
>
> If $n$ is odd and the sample medians for $x$ and $y$ correspond to two *different* $(X_i, Y_i)$'s, count these two as one point only, and assign them to the quadrant touched by both points for purposes of computing $n_1$ and $n_2$.

A simple sketch for $n=3$ will illustrate the above definition of $n_1$ and $n_2$. Then $q=(n_1-n_2)/(n_1+n_2)$. (Note on terminology: Blomqvist uses "$q$" for our "$\rho$" and "$q'$" for our "$q$.")

Blomqvist showed that, under mild regularity conditions,

$$\sqrt{n}\,\frac{(q-\rho)}{\sqrt{1-\rho^2}} \tag{10.2}$$

is asymptotically unit-normal. From this result, approximate confidence intervals for $\rho$, or tests of the null hypothesis that $\rho$ is some specific value, may readily be obtained. Even more simply, but probably less accurately,

$$\sqrt{n}\,\frac{q-\rho}{\sqrt{1-q^2}} \tag{10.3}$$

is approximately unit-normal for large $n$, thus permitting very simple approximate tests and confidence interval procedures.

An exact test that $X$ and $Y$ are independent is readily obtained on the basis of $q$. Blomqvist provided tables for its use, and showed that, in the bivariate normal case, its asymptotic efficiency is 41 per cent of the test based on the correlation coefficient. Konijn [53] corrects Blomquist's statement of regularity conditions and discusses asymptotic power.

## 11. ESTIMATION OF $\tau$

The natural estimator of $\Pi_c$ from a random sample of $n$, $(X_1, Y_1), \cdots, (X_n, Y_n)$, is simply the relative frequency of concordant pairs of observations. As before, we assume continuous marginals.

Let us denote this estimator by

$$P_c = \frac{\text{Number concordant } (X_i, Y_i), (X_j, Y_j), \quad i \neq j}{n(n-1)} = \frac{\text{Number concordant } (X_i, Y_i), (X_j, Y_j), \quad i < j}{\frac{1}{2}n(n-1)} \tag{11.1}$$

where the second form uses the fact that a pair of bivariate observations need only be considered in one order.

It is sometimes useful to present $P_c$ in terms of a score, as follows. For each unordered pair of observations, $((X_i, Y_i), (X_j, Y_j),)$ with $i \neq j$, score 1 or 0 according as concordance or discordance obtains. Then $P_c$ = sum of scores divided by $n(n-1)$.

An equivalent and often convenient rewording is readily given in terms of ranks. Replace the $X_i$'s by the numbers $1, 2, \cdots, n$ (lowest $X_i$ replaced by 1, next lowest $X_i$ by 2, and so on). Similarly replace the $Y_i$'s by their ranks. If we compute $P_c$ with $(X_i, Y_i)$ replaced by (rank $X_i$, rank $Y_i$), $P_c$ is unchanged, since two observations are concordant if and only if they are concordant after their components are replaced by ranks.

We may write down, in order, the rank of that $Y_i$ whose $X_i$ has rank 1, followed by the rank of that $Y_i$ whose $X_i$ has rank 2, etc. A table of the following form is obtained:

| | | | | | | |
|---|---|---|---|---|---|---|
| $X_i$ ranks | 1 | 2 | 3 | 4 | $\cdots$ | $n$ |
| Corresponding $Y_i$ ranks | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $\cdots$ | $S_n$ |

where the $S_i$'s form a permutation of $1, 2, \cdots, n$. Consider now concordances among pairs of observed ranks one of whose members is $(1, S_1)$. This number will range from 0 (if $S_1 = n$) to $n-1$ (if $S_1 = 1$). We may consider it as obtained by scoring a 1 or a 0 for each of the pairs considered, according as concordance or discordance holds, and then summing the scores. Having finished with pairs of observations that include $(1, S_1)$, we turn to those that include $(2, S_2)$ and *ex*clude $(1, S_1)$. Scoring as before, we obtain a sum ranging from 0 to $n-2$. Continue in this way, and add up all the scores. We obtain a number ranging from 0 (when $S_i = n-i$) to $(n-1)+(n-2)+(n-3)+ \cdots 1 = \frac{1}{2}n(n-1)$ (when $S_i = i$). Dividing the total score by $\frac{1}{2}n(n-1)$ we obtain $P_c$, a number ranging from 0 to 1. The computing procedure described above is often the simplest. Note that the ranks do not enter intrinsically at all, since the same operation may be performed on the ordered observations. However, the use of ranks simplifies the comparisons.

Similarly, one may score for *dis*cordance and obtain $P_d$, the natural estimator of $\Pi_d$. Then $t$, the estimator of $\tau$, is $P_c - P_d$, or, equivalently, $2P_c - 1 = 1 - 2P_d$. Again, equivalently, we may score each of the $\frac{1}{2}n(n-1)$ *un*ordered pairs 1 or $-1$ according as concordance or discordance obtains. To obtain $t$, divide the total resulting score by $\frac{1}{2}n(n-1)$.

A clever graphical method of computing $t$ has been suggested by S. D.

Holmes [75, Appendix B]. I am indebted to Harold D. Griffin for calling this to my attention. Griffin discusses the method [37].

It is interesting to note that $\binom{n}{2}P_d$ is just the minimum number of inversions of neighboring $S_i$'s necessary to permute the $S_i$'s into the order $1, 2, \cdots, n$. For (a) the inversion of a discordant pair of neighbor $S_i$'s decreases $\binom{n}{2}P_d$ by 1, and also decreases the number of necessary subsequent inversions by 1. And (b) the greatest value of $\binom{n}{2}P_d$ and the greatest number of necessary inversions (both occurring when $S_i = n-i$) are $\frac{1}{2}n(n-1)$.

The exact distribution of $t$ in the case of independence may be computed recursively; the most extensive tabulation is that of Kaarsemaker and van Wijngaarden [48]. The distribution of $t$ in the general bivariate normal case has been investigated by E. F. Fieller, H. O. Hartley, and E. S. Pearson [30a], using empirical sampling methods. In the same article the inverse hyperbolic tangent function is considered as a normalizing and variance-stabilizing transformation.

It is immediate from its definition that $Et = \tau$. The variance of $t$ is not hard to compute in terms of a quantity already introduced in section 7,

$$\Pi_{cc} = \Pr\{\text{of three independent observations, the second and third are concordant with the first}\}. \tag{11.2}$$

One finds (see, e.g., Hoeffding [43]),

$$\operatorname{Var} t = \frac{8}{n(n-1)}\Pi_c(1-\Pi_c) + 16\frac{1}{n}\frac{n-2}{n-1}(\Pi_{cc} - \Pi_c^2). \tag{11.3}$$

Note that as $n \to \infty$, this quantity times $n$ has the limit $16(\Pi_{cc} - \Pi_c^2)$. It may be shown (Hoeffding [43] and [44]) that, unless $\Pi_{cc} = \Pi_c^2$,

$$\frac{\sqrt{n}(t-\tau)}{4\sqrt{P_{cc} - P_c^2}} \to N(0, 1) \tag{11.4}$$

in distribution, where $P_{cc}$ is the sample analogue of $\Pi_{cc}$ just as $P_c$ is that of $\Pi_c$. The meaning of (11.4) is that the probability that the quantity on the left lies in any fixed interval has as its limit ($n \to \infty$) the probability of that interval under the unit-normal distribution.

The condition $\Pi_{cc} \neq \Pi_c^2$ means (see section 7) that the probability that observation 2 is concordant with observation 1, given observation 1, is *not* essentially constant. It will be satisfied except for rather unusual distributions such as those used as examples in section 7. If a genuine density function exists, the condition will be satisfied.

An upper bound for Var $t$, suggesting "conservative" simple approximate tests and confidence intervals, has been given by Daniels and Kendall [15]: Var $t \leq 2(1-\tau^2)/n$. This inequality follows by direct substitution from the relationship $2\Pi_{cc} \leq \Pi_c + \Pi_c^2$. And this relationship, in turn, may readily be demonstrated, for the no-ties random sample situation that we are discussing, as follows. Consider four independent observations on the distribution of interest: $(X_i, Y_i)$, $i = 1, 2, 3, 4$. Let $W_{ij} = 1$ or 0 according as $(X_i, Y_i)$ is concordant with $(X_j, Y_j)$ or discordant. Then $W_{ij} = W_{ji}$, $EW_{ij} = \Pi_c$, Var

$W_{ij}=\Pi_c-\Pi_c^2$, and Cov $(W_{ij}, W_{ik})=\Pi_{cc}-\Pi_c^2$ (for $j\neq k$) by direct computation. Next compute

$$\text{Var}\ [2(W_{12}+W_{34})-(W_{13}+W_{14}+W_{23}+W_{24})]\geqq 0$$

to find that $\Pi_c+\Pi_c^2-2\Pi_{cc}\geqq 0$, the desired result.

To use $t$ as an approximate test of independence we note that, under independence, $\Pi_{cc}=5/18$ and $\Pi_c=1/2$. Hence $\Pi_{cc}-\Pi_c^2=1/36$ and $\sqrt{n}\ 3t/2$ is approximately unit-normal for large $n$. Apparently, a more accurate test is obtained by using the exact variance of $\sqrt{n}t$ under the hypothesis of independence, $2(2n+5)/[9(n-1)]$; that is by taking

$$\sqrt{n}\sqrt{\frac{9(n-1)}{2(2n+5)}}\,t$$

as approximately unit-normal under the null hypothesis.

The asymptotic efficiency (Pitman sense) of this test in the normal case, as contrasted with that based on the correlation coefficient, is $9/\pi^2\cong.91$.

A variant of $t$ has been suggested by Whitfield [90] as appropriate for cases where the distribution is known to be symmetric, in the sense that $(X, Y)$ has the same distribution as $(Y, X)$. This situation is analogous to one-way model II analysis of variance with two observations per group, when viewed from the intra-class correlation standpoint. The essential point is that $(X_i, Y_i)$ and $(Y_i, X_i)$ are equally good observations on the distribution of $(X, Y)$. Whitfield suggests, in effect, that all the $2^n$ possible values of $t$ be averaged as an estimate of $\tau$, and he provides a shorter way of carrying out the computation that does not require the evaluation of many $t$'s. In addition he discusses and tabulates for $n=6(2)20$ the distribution of the resulting average $t$ under the hypothesis of independence.

## 12. ESTIMATION OF $\rho_S$

Perhaps the simplest estimator of $\rho_S$ may be motivated by using the definition of $\rho_S$ in form

$$\begin{aligned}\rho_S &= 6\Pr\{(X_1, Y_1), (X_2, Y_3) \text{ concordant}\} - 3\\ &= 6\iota_c - 3.\end{aligned}$$

In one sense the natural estimator of the joint distribution of $(X_1, Y_1)$ is that discrete distribution putting probability mass $1/n$ on each $(X_i, Y_i)$ of the sample actually obtained. Similarly, one natural estimator of the distribution of $(X_2, Y_3)$, that is the bivariate distribution with independent coordinates and the same marginals as $(X_1, Y_1)$, is that discrete distribution putting mass $1/n^2$ on each of the $n^2$ points $(X_j, Y_k)$ formed from the sample that is actually obtained.

If we adopt this viewpoint, a fairly reasonable estimator of $\iota_c=\Pr\ \{(X_1, Y_1), (X_2, Y_3) \text{ concordant}\}$ is obtained by computing the corresponding quantity for the two discrete distributions described above. But here we very definitely do not have continuous marginals, and there are several ways one might proceed. I choose one variation that leads to the conventional sample value of Spearman's rank correlation coefficient.

Consider any one of the actual observations, $(X_i, Y_i)$, and obtain from this a partial estimate of $\iota_c = \Pr\{X_1, Y_1), (X_2, Y_3) \text{ conc.}\}$ as follows.

There are $n^2-1$ points of form $(X_j, Y_k)$, *excluding* the point that we are working with itself. If $R_{X_i}$ is the rank of $X_i$ among the $X$'s and $R_{Y_i}$ the rank of $Y_i$ among the $Y$'s, then, out of the $n^2-1$ points $(X_j, Y_k)$, exactly $(R_{X_i}-1)\times(R_{Y_i}-1)$ will lie below and to the left of $(X_i, Y_i)$. Similarly exactly $(n-R_{X_i})\times(n-R_{Y_i})$ will lie above and to the right of $(X_i, Y_i)$. These relations are easy to picture; Figure 840 shows a typical sample with $n=8$. The heavy dots are the

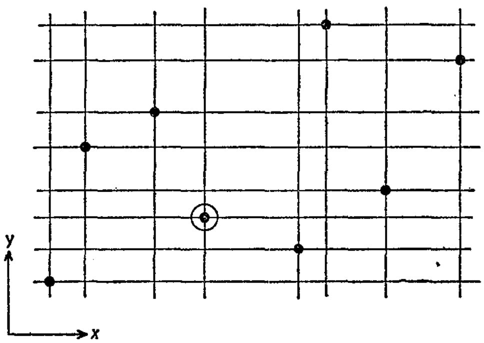



FIG. 840

observed points, and through them horizontal and vertical lines parallel to the axes have been drawn. The heavy circled dot is $(X_i, Y_i)$. In the pictured case $R_{X_i}=4$ and $R_{Y_i}=3$. There are $(4-1)\ (3-1)=6$ intersections below and to the left of the circled dot; there are $(8-4)\ (8-3)=20$ intersections above and to the right of the circled dot.

The total number of intersections concordant with $(X_i, Y_i)$ is, therefore, in general

$$\begin{aligned}(R_{X_i}-1)(R_{Y_i}-1)&+(n-R_{X_i})(n-R_{Y_i})\\ &= 2R_{X_i}R_{Y_i}-(n+1)(R_{X_i}+R_{Y_i})+n^2+1.\end{aligned}\tag{12.1}$$

But this count is unfair, for there are $2(n-1)$ intersections *on* the horizontal and vertical lines through $(X_i, Y_i)$; these are tied in exactly one coordinate with $(X_i, Y_i)$. How should we take these into account? It seems reasonable to count them at half their number, on the grounds that they lie exactly poised between concordance and discordance. Thus the total number of intersections concordant with $(X_i, Y_i)$, plus half the number tied in one coordinate, all divided by $n^2-1$, the total number of intersections excluding $(X_i, Y_i)$, is

$$\frac{1}{n^2-1}\left\{2R_{X_i}R_{Y_i}-(n+1)(R_{X_i}+R_{Y_i})+n(n+1)\right\}.\tag{12.2}$$

Next, average this over the $n$ $(X_i, Y_i)$'s to obtain

$$\frac{1}{n(n^2-1)}\left\{2\sum R_{X_i}R_{Y_i} - n(n+1)^2 + n^2(n+1)\right\}$$
$$= \frac{1}{n(n^2-1)}\left\{2\sum R_{X_i}R_{Y_i} - n(n+1)\right\}. \tag{12.3}$$

This is a fairly reasonable estimator of $\iota_c$; we might call it $i_c$. Now take 6 times $i_c$ and subtract 3, to obtain, as an estimator of $\rho_S$

$$r_S = \frac{1}{n(n^2-1)}\left\{12\sum R_{X_i}R_{Y_i} - 6n(n+1) - 3n(n^2-1)\right\}$$
$$= \frac{1}{n(n^2-1)}\left\{12\sum R_{X_i}R_{Y_i} - 3n(n+1)^2\right\}. \tag{12.4}$$

This is the famous Spearman sample rank correlation coefficient. Two more usual forms of it are

$$r_S = \frac{\frac{1}{n}\sum\left(R_{X_i} - \frac{n+1}{2}\right)\left(R_{Y_i} - \frac{n+1}{2}\right)}{(n^2-1)/12} \tag{12.5}$$

or the ordinary correlation coefficient *computed in terms of the ranks;* and

$$r_S = 1 - 6\,\frac{\sum (R_{X_i} - R_{Y_i})^2}{n(n^2-1)}, \tag{12.6}$$

a form often convenient for computation. The equivalence of the above three forms is easily verified by elementary algebraic manipulation.

The above motivation for $r_S$ contains a number of arbitrary elements, for example the manner of counting intersections tied in one coordinate with $(X_i, Y_i)$. Let us now consider a less arbitrarily motivated estimator for $\rho_S$; this turns out to be not $r_S$, but rather a linear combination of $r_S$ and $t$. However, for large $n$ it becomes indistinguishable from $r_S$.

This more natural estimator is approached just as we approached the estimator for $\tau$. We begin with the definition $\rho_S = 2\omega_c - 1$, and ask about estimation of $\omega_c$.

### 13. A MORE NATURAL ESTIMATOR OF $\rho_S$ AND FURTHER REMARKS ABOUT $r_S$

A natural estimator of $\omega_c$, $w_c$, may be obtained by looking at all $\binom{n}{3}$ unordered sample triplets, $((X_i, Y_i), (X_j, Y_j), (X_k, Y_k))$, with no tied subscripts, and scoring 1 or 0 according as there are $\leqq 1$ inversions of rank order or $\geqq 2$ such inversions. Then compute $w_c = \text{total score}/\binom{n}{3}$. In order to bring this into more explicit form, adopt the following device. (I assume $n \geqq 3$, so that this approach has meaning.)

Count all concordances between the $n(n-1)(n-2)$ pairs

$$(X_i, Y_i),\ (X_j, Y_k) \qquad i \neq j,\ i \neq k,\ j \neq k. \tag{13.1}$$

We know, from Table 825, that each triple scoring 1 from the above paragraph contributes 4 to this count, and each triple scoring 0 from the above paragraph contributes 2 to this count. If the count suggested here yields $c$, say, then

$$c = 4\binom{n}{3}w_c + 2\binom{n}{3}(1 - w_c) \tag{13.2}$$

or

$$c\Big/\binom{n}{3} = 2w_c + 2; \qquad w_c = \frac{c}{2\binom{n}{3}} - 1. \tag{13.3}$$

Now ask how much $(X_i, Y_i)$ contributes to the count $c$. If we did *not* have the restriction $j \neq k$ it is clear (just as before) that the number of $(X_j, Y_k)$ points below and to the left of $(X_i, Y_i)$ would be

$$(R_{X_i} - 1)(R_{Y_i} - 1) \tag{13.4}$$

where $R_{X_i}$ and $R_{Y_i}$ are the ranks of $X_i$ among the $X$'s and of $Y_i$ among the $Y$'s. Similarly the number of $(X_j, Y_k)$ points above and to the right of $(X_i, Y_i)$ would be

$$(n - R_{X_i})(n - R_{Y_i}). \tag{13.5}$$

Hence the contribution of $(X_i, Y_i)$ to $c$ is

$$\begin{aligned}&(R_{X_i} - 1)(R_{Y_i} - 1) + (n - R_{X_i})(n - R_{Y_i})\\ &\qquad - \text{no. } (X_j, Y_j)\text{'s concordant with } (X_i, Y_i),\end{aligned} \tag{13.6}$$

where the last term subtracts off the correct integer to take account of the restriction $j \neq k$. We obtain then

$$\begin{aligned}&2R_{X_i}R_{Y_i} - (n + 1)(R_{X_i} + R_{Y_i}) + n^2 + 1\\ &\qquad - \text{no. } (X_j, Y_j)\text{'s concordant with } (X_i, Y_i).\end{aligned} \tag{13.7}$$

If we sum over $i$, we obtain

$$\begin{aligned} c &= 2\sum R_{X_i}R_{Y_i} - n(n + 1)^2 + n^3 + n - n(n - 1)P_c\\ &= 2\sum R_{X_i}R_{Y_i} - 2n^2 - n(n - 1)P_c\end{aligned} \tag{13.8}$$

since $\sum R_{X_i} = \sum R_{Y_i} = n(n+1)/2$, and since each concordant pair is counted twice in the last subtractive term.

Hence

$$\begin{aligned} w_c &= \frac{1}{\binom{n}{3}}\sum R_{X_i}R_{Y_i} - \frac{n^2}{\binom{n}{3}} - 1 - \frac{3}{n - 2}P_c\\ &= \frac{1}{\binom{n}{3}}\left\{\sum R_{X_i}R_{Y_i} - n^2 - \frac{1}{6}n(n - 1)(n - 2)\right\} - \frac{3}{n - 2}P_c \qquad (13.9)\\ &= \frac{6}{n(n - 1)(n - 2)}\sum R_{X_i}R_{Y_i} - \frac{(n + 1)(n + 2)}{(n - 1)(n - 2)} - \frac{3}{n - 2}P_c.\end{aligned}$$

Next, to estimate $\rho_S$, consider $2w_c-1$, and—for the sake of symmetry—replace $P_c$ by $(t+1)/2$. We find

$$\begin{aligned} 2w_c - 1 &= \frac{12}{n(n-1)(n-2)} \sum R_{X_i} R_{Y_i} - \frac{2(n+1)(n+2)}{(n-1)(n-2)} \\ &\quad - 1 - \frac{3}{n-2} t - \frac{3}{n-2} \qquad (13.10) \\ &= \frac{1}{n(n-1)(n-2)} \left\{ 12 \sum R_{X_i} R_{Y_i} - 3n(n+1)^2 \right\} - \frac{3}{n-2} t \\ &= \frac{n+1}{n-2} r_S - \frac{3}{n-2} t. \end{aligned}$$

Note that, for large $n$, this is virtually the same as $r_S$.

Note also that, since the probability is $\omega_c$ that a random triple of sample points shows 0 or 1 inversions in rank order, $EW_c = \omega_c$, and consequently the expected value of the above estimator is $\rho_S$. Hence

$$\rho_S = \frac{n+1}{n-2} Er_S - \frac{3}{n-2} \tau \qquad (13.11)$$

or

$$Er_S = \frac{n-2}{n+1} \rho_S + \frac{3}{n+1} \tau. \qquad (13.12)$$

This shows that $r_S$ is in general a biased estimator of $\rho_S$, but that the bias rapidly goes to zero as $n$ grows.

The $\rho_S$ estimator $2w_c - 1 = [(n+1)/(n-2)]r_S - [3/(n-2)]t$ has two advantages when compared to $r_S$: it is more naturally motivated and it is unbiased. On the other hand, $r_s$ is more easily computed from a sample and is much more commonly used. Konijn [53] has called $2w_c - 1$ the unbiased grade correlation. It, together with $r_S$, has been studied carefully by Hoeffding [44], where general expressions for the variances of both quantities are given.

If $X$ and $Y$ are *independent*, Var $r_S = 1/(n-1)$ and Var $(2w_c - 1) = (n^2-3)/[n(n-1)(n-2)]$.

If $X$ and $Y$ are jointly normally distributed, the variance of $r_S$ may be approximated by the first terms of a series expansion; see Kendall [51, second ed., p. 130]. The distribution of $r_S$ in the bivariate normal case has been investigated by Fieller, Hartley, and Pearson [30a], using empirical sampling methods. In the same article the inverse hyperbolic tangent function is considered as a normalizing and variance-stabilizing transformation.

For a test of independence, $r_S$ may be used as a test statistic. Its exact distribution under the hypothesis of independence has been tabulated for $n$ through 8 by M. G. Kendall et al [52] and for $n=9$, 10 by S. T. David et al [16]. (See also Olds [65] and Thornton [84]. The distributions are tabulated in Kendall's monograph [51].) Hotelling and Pabst [46] have shown that, when independence holds,

$$\sqrt{n} r_S \qquad (13.13)$$

is asymptotically unit-normal. The more general asymptotic distribution of $r_S$ has been discussed by Hoeffding [44].

When independence holds, M. G. Kendall has found that a reasonable approximation to the distribution of $r_S$ may be obtained by treating

$$r_S \sqrt{\frac{n-2}{1-r_S^2}}$$

as distributed according to Student's $t$ distribution with $n-2$ degrees of freedom. A table of one and five per cent two-sided critical values, for $n=5(1)40$, based on this approximation, is given by Litchfield and Wilcoxon [62].

The asymptotic efficiency (Pitman sense) of the test of independence based on $r_S$, in the normal case, is $9/\pi^2 \cong .91$ relative to that of the test based on the correlation coefficient. This is the same value as that for the test based on $t$.

## 14. RELATIVE MERITS OF $\tau$ AND $\rho_S$, $t$ AND $r_S$

As between $r_S$ and $t$ qua estimators, they are not really in general competition, for they are estimators of *different* population quantities. It may be noted that $\tau$ is simpler to interpret than $\rho_S$. Some authors consider it important that $r_S$ is easier to compute than $t$.

As between $r_S$ and $t$ qua test statistics for a test of independence, it may be pointed out that they presumably have power against different sorts of alternatives to the null hypothesis of independence. A rather extreme example is the population in which the mass is spread smoothly along two curves as in Fig. 830, with $a=\frac{1}{2}$. Here $\tau=0$ and $\rho_S=\frac{1}{4}$. If populations akin to that pictured were the important alternatives to independence, $r_S$ would undoubtedly be a better test statistic than $t$.

Konijn [53] presents an interesting and novel approach to the power of tests for independence based on $q$, $t$, $r_S$, and $w_c$. In this approach the family of alternatives is taken to be those bivariate distributions derived by linear transformations from bivariate distributions with independent marginals. Konijn, in paper to appear in *Sankhya*, has considered another interesting restricted family of bivariate distributions. Another approach to power has been suggested in a recent paper by Barton and David [5].

In the bivariate normal case, the estimators of $\rho$ based on $t$ and $r_S$ have a correlation approaching unity as $n \to \infty$.

It is clear that many other ordinal measures of association might be proposed. For example, one might base such a measure on the probability that all pairs out of three observations are concordant. Or one might base a measure on $\Pi_{cc}$. And so on.

On the grounds of simplicity of interpretation, reasonable sensitivity to form of distribution, and relative simplicity of sampling theory, I prefer the use of $\tau$ and $t$ to that of $\rho_S$ and $r_S$.

## 15. GENERAL DISTRIBUTIONS, AND ANOTHER APPROACH TO $t$ IN THE NO-TIE CASE

If the assumption that $X$ and $Y$ have continuous marginal distributions be dropped, then many of the manipulations presented or suggested in the preceding sections become false or incompletely defined. For then ties in one coordinate or the other may occur with positive probability. However, $\tau$ and $\rho_S$ may still be defined.

All we need do, in the case of $\tau$, is to redefine $\Pi_c$ and $\Pi_d$ as the probabilities of the same chance events as before, but now *conditionally* on the absence of ties. Thus, we need only set, for example,

$$\Pi_c = \Pr\{X_1 - X_2)(Y_1 - Y_2) > 0 \mid X_1 \neq X_2 \text{ and } Y_1 \neq Y_2\}. \qquad (15.1)$$

This quantity may be considered as well defined in general; the only exceptions are the degenerate cases in which the entire distribution is concentrated on a horizontal or vertical line. The definitions hold perfectly well in the continuous case earlier discussed, for then the probability of the condition is unity.

What is the natural sample estimator for $\tau$ in the general case? Let us look first at that for $\Pi_c$. As before, we start by observing the relative number of sample pairs for which concordance obtains. But now we must divide by the relative number of sample pairs satisfying the no-ties condition. This suggests the quantity

$$P_c = \frac{\text{no. concordant sample pairs}}{\text{no. no-tie sample pairs}}\cdot \qquad (15.2)$$

To say that the pair of observations $(X_i, Y_i)$, $(X_j, Y_j)$ $(i \neq j)$ is a no-tie pair is to say that $X_i \neq X_j$ and $Y_i \neq Y_j$. The analogous definition for $P_d$ is clear, and so $t$ may be defined for the general case as $P_c - P_d$, just as before.

The above variant of $t$ for tied observations, or some slight modification of it, has been suggested occasionally in the literature, perhaps first by Deuchler (see section 17). A recent discussion of it is presented by Adler [1]. For approaches to $t$ when ties are present, using other denominators than that of (15.2), see Kendall [51, Chapter 3 and the references there given].

A particularly interesting kind of noncontinuous distribution is that in which $X$ and $Y$ can each take only a finite number of values. Without loss of generality (because of ordinal invariance) we may suppose that $X$ takes values $1, 2, \cdots, \alpha$ and $Y$ takes values $1, 2, \cdots, \beta$. The joint distribution of $X$ and $Y$ may then be described by the $\alpha\beta$ probabilities $\rho_{ab}$ $(a = 1, \cdots \alpha, b = 1, \cdots, \beta)$, the probabilities that $X = a$ and $Y = b$. This is just the cross classification situation of [35] with order in both classifications.

The probability of a tie in one or both coordinates when taking two observations is

$$\Pi_t = \sum_a \rho_{a.}^2 + \sum_b \rho_{.b}^2 - \sum_a \sum_b \rho_{ab}^2. \qquad (15.3)$$

Here $\rho_{a.} = \sum_b \rho_{ab}$ and $\rho_{.b} = \sum_a \rho_{ab}$, the marginal probabilities that $X = a$ and that $Y = b$ respectively. It is easily shown that

$$\begin{aligned} \Pi_c &= \frac{2}{1 - \Pi_t} \sum_a \sum_b \rho_{ab} \left\{ \sum_{a'>a} \sum_{b'>b} \rho_{a'b'} \right\} \\ \Pi_d &= \frac{2}{1 - \Pi_t} \sum_a \sum_b \rho_{ab} \left\{ \sum_{a'>a} \sum_{b'<b} \rho_{a'b'} \right\}. \end{aligned} \qquad (15.4)$$

See [35] for a detailed discussion of the approach in this case, but note that there are a few minor differences of notation between [35] and the present paper. In particular $\Pi_s$ of [35] is here denoted by $\Pi_c(1-\Pi_t)$, and $\Pi_d$ of [35] is here $\Pi_d(1-\Pi_t)$. Further, "$a$" refers here to columns and "$b$" to rows; in [35] the opposite usage holds. In [35], $\Pi_c-\Pi_d$ of this paper is called $\gamma$.

From this point of view another motivation for $t$ in the *continuous* case may be given: $t$ is just the value of $\tau$ for the sample pattern of points, each taken with probability $1/n$. For consider the sample of $n$ $(X_i, Y_i)$'s as defining a discrete distribution in which $X$ can take values $X_1, \cdots, X_n$, in which $Y$ can take values $Y_1, \cdots, Y_n$, and for which $\rho_{ab}=1/n$ when $(a, b)=(X_i, Y_i)$ and zero otherwise. This discrete distribution may be represented as an $n\times n$ table of probabilities in which all entries are zero except that exactly one entry in each row and column has $\rho_{ab}=1/n$.

We readily compute that, in this case

$$\begin{aligned} \Pi_t &= 1/n \\ 1/(1-\Pi_t) &= n/(n-1) \\ \Pi_c &= \frac{1}{\binom{n}{2}} \text{ no. concordant unordered pairs } ((X_i, Y_i), (X_j, Y_j)) \\ \Pi_d &= \frac{1}{\binom{n}{2}} \text{ no. discordant unordered pairs } ((X_i, Y_i), (X_j, Y_j)) \end{aligned} \tag{15.5}$$

so that $\tau$ for the uniform discrete distribution on the points of the sample $\{(X_i, Y_i)\}$ is precisely $t$ for the points considered qua sample.

## 16. $\rho_S$ IN THE GENERAL CASE, AND ANOTHER APPROACH TO $r_S$ IN THE NO-TIE CASE

Let us now turn to analogous manipulations for $\rho_S$ and $r_S$. For any bivariate distribution, we may consider the following disjoint chance events relating to three independent observations:

$$\begin{aligned} &C\text{: at least one of the three is clearly concordant with the other two} \\ &D\text{: at least one of the three is clearly discordant with the other two.} \end{aligned} \tag{16.1}$$

To say that $(X_i, Y_i)$ and $(X_j, Y_j)$ are clearly concordant (discordant) is to say that $(X_i-X_j)(Y_i-Y_j)>0(<0)$. Emphasis is on the strong inequality.

If the marginals are continuous, then $\omega_c=\Pr\{C\}$ and $\omega_d=\Pr\{D\}$. But if ties can occur, it seems natural to generalize the definitions of $\omega_c$ and $\omega_d$, consistently with (15.2), thus,

$$\begin{aligned} \omega_c &= \Pr\{C \mid C \text{ or } D\} = \Pr\{C\}/[\Pr\{C\}+\Pr\{D\}], \\ \omega_d &= \Pr\{D \mid C \text{ or } D\} = \Pr\{D\}/[\Pr\{C\}+\Pr\{D\}], \end{aligned} \tag{16.2}$$

to take account of the fact that patterns may occur that fall into neither of the events $C$ or $D$. For example, the pattern of Fig. 847 is of this kind. On the other hand, some patterns with ties may fall into $C$ or $D$ (but not both). For example, that of Fig. 847a falls into $C$.

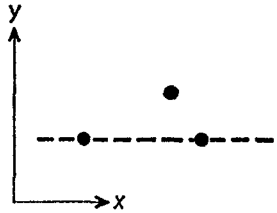



FIG. 847. Pattern falling into neither $C$ or $D$.

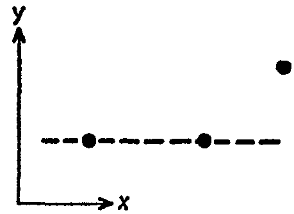



FIG. 847a. Pattern falling into $C$.

Dots are observations. Lower two observations tied in $y$ coordinates.

Adopting (16.2), it is natural to define $\rho_S$ as $\omega_c - \omega_d$ generally. In the no-ties case this definition coincides with our earlier one. Reasonable estimators of $\omega_c$, $\omega_d$, and $\rho_S$ from a random sample of $n(\geqq 3)$ would be, following section 15, to look at all $\binom{n}{3}$ unordered triples from the sample and to set

$$w_c = \frac{\text{no. of the } \binom{n}{3} \text{ triples in } C}{\text{no. of the } \binom{n}{3} \text{ triples in } C \text{ or } D}$$

$$w_d = \frac{\text{no. of the } \binom{n}{3} \text{ triples in } D}{\text{no. of } \binom{n}{3} \text{ triples in } C \text{ or } D}. \tag{16.3}$$

estimator of $\rho_S$

$$= \frac{\text{no. of } \binom{n}{3} \text{ triples in } C - \text{no. of } \binom{n}{3} \text{ triples in } D}{\text{no. of } \binom{n}{3} \text{ triples in } C \text{ or } D}.$$

Note that, if ties can occur, it is possible that two observations will be tied in *both* coordinates. In such a case, if the third observation of a triple is tied in neither coordinate with the two, the triple must be in either $C$ or $D$.

Next, let us ask what the generalized $\rho_S$ will be for a cross classification. Notation will be the same as that of section 15, and the computation will be carried out in two steps for simplicity.

First, we recall a classical formula of elementary probability. If $A_1$, $A_2$, $A_3$ are three chance events, then,

$$\begin{aligned}\Pr\{A_1 \text{ or } A_2 \text{ or } A_3\} &= \Pr\{A_1\} + \Pr\{A_2\} + \Pr\{A_3\} - \Pr\{A_1 \text{ and } A_2\} \\ &\quad - \Pr\{A_1 \text{ and } A_3\} - \Pr\{A_2 \text{ and } A_3\} + \Pr\{A_1 \text{ and } A_2 \text{ and } A_3\}.\end{aligned} \tag{16.4}$$

Next, let the event $A_1$ be defined as follows:

$A_1$: of three observations, $(X_i, Y_i)$, $i=1, 2, 3$, observation 1 is concordant with observations 2 and 3.

$A_2$ and $A_3$ are similarly defined, with observations 2 and 3 respectively playing the role of observation 1 in the definition of $A_1$.

It is then clear that $C=A_1$ or $A_2$ or $A_3$, so that (using symmetry, and the fact that "$A_1$ and $A_2$" means that all three observations are concordant)

$$\begin{aligned}\Pr\{C\} &= 3\Pr\{A_1\} - 3\Pr\{\text{all three conc.}\} + \Pr\{\text{all three conc.}\}\\ &= 3\Pi_{cc} - 2\Pi_{ccc}\end{aligned} \tag{16.5}$$

where $\Pi_{cc}$ has its earlier meaning, $\Pr\{A_i\}$, and $\Pi_{ccc}$ is the probability that, of three independent observations, all are concordant. (It is convenient here to maintain *un*conditional senses for $\Pi_{cc}$ and $\Pi_{ccc}$.)

Next, let us introduce the notation

$$\begin{aligned}\mathrm{I}_{ab} &= \sum_{a'>a}\sum_{b'>b}\rho_{a'b'}\\ \mathrm{II}_{ab} &= \sum_{a'<b}\sum_{b'>b}\rho_{a'b'}\\ \mathrm{III}_{ab} &= \sum_{a'<a}\sum_{b'<b}\rho_{a'b'}\\ \mathrm{IV}_{ab} &= \sum_{a'>b}\sum_{b'<b}\rho_{a'b'}.\end{aligned} \tag{16.6}$$

The mnemonic for this notation is the conventional numbering of the quadranst around $(X, Y)=(a, b)$.

It is clear that

$$\Pi_{cc} = \sum_a\sum_b \rho_{ab}(\mathrm{I}_{ab} + \mathrm{III}_{ab})^2 \tag{16.7}$$

by summing the quantities $\Pr\{A_1 \text{ and } (X_1, Y_1)=(a, b)\}$. Similarly, by summing the six equal quantities of form $\Pr\{$all three concordant, with $(X_1, Y_1)$ in lower left position, $(X_2, Y_2)$ in middle, and $(X_3, Y_3)$ in upper right$\}$, each of which in turn is taken as a sum of probabilities with the middle observation fixed, we obtain

$$\Pi_{ccc} = 6\sum_a\sum_b \rho_{ab}\mathrm{I}_{ab}\mathrm{III}_{ab}. \tag{16.8}$$

Substituting (16.7) and (16.8) into (16.5) we have

$$\Pr(C) = 3\sum_a\sum_b \rho_{ab}(\mathrm{I}_{ab} - \mathrm{III}_{ab})^2, \tag{16.9}$$

and, similarly,

$$\Pr(D) = 3\sum_a\sum_b \rho_{ab}(\mathrm{II}_{ab} - \mathrm{IV}_{ab})^2.$$

Thus, for a cross classification, we may define a measure of association, analogous to $\rho_S$, as

$$\frac{\sum\sum \rho_{ab}[(\mathrm{I}_{ab} - \mathrm{III}_{ab})^2 - (\mathrm{II}_{ab} - \mathrm{IV}_{ab})^2]}{\sum\sum \rho_{ab}[(\mathrm{I}_{ab} - \mathrm{III}_{ab})^2 + (\mathrm{II}_{ab} - \mathrm{IV}_{ab})^2]}. \tag{16.10}$$

This quantity might be useful in the analysis of cross classifications, and it is offered as another measure of association for such analyses, along with the measures of [35]. It is a direct analog of $\rho_S$ in the no-ties case for the wholly discrete case, and it has a clear operational interpretation, though not as direct as that of the corresponding analog to $\tau$. The computation of (16.10) is somewhat harder than that of the analog to $\tau$. (16.10) has the following properties:

1. (16.10) is defined unless some $\rho_{a.}$ or $\rho_{.b}=1$.
2. (16.10) lies between $-1$ and 1 inclusive.
3. (16.10) is 1 if and only if Pr $\{D\}=0$. This means that there is no pair of positive cell probabilities, $\rho_{ab}$ and $\rho_{a'b'}$, such that $(a, b)$ is discordant with $(a', b')$.
4. (16.10) is $-1$ if and only if Pr $\{C\}=0$. This means that there is no pair of positive cell probabilities, $\rho_{ab}$ and $\rho_{a'b'}$, such that $(a, b)$ is concordant with $(a', b')$.
5. (16.10) need *not* be zero in the case of independence. However, if at least one set of marginals is uniform (i.e., all $\rho_{a.}=1/\alpha$ or all $\rho_{.b}=1/\beta$) then (16.10) is zero under independence, but not conversely in general.
6. In the $2\times2$ case, (16.10) is

$$\frac{\rho_{11}\rho_{22}(\rho_{11}+\rho_{22})-\rho_{12}\rho_{21}(\rho_{12}+\rho_{21})}{\rho_{11}\rho_{22}(\rho_{11}+\rho_{22})+\rho_{12}\rho_{21}(\rho_{12}+\rho_{21})} \tag{16.11}$$

Property 5 is unhappy. It reflects the fact that, if the coordinates of two observations in a class $C$ pattern are cross-switched ($X_1\leftrightarrow Y_2$, $X_2\leftrightarrow Y_1$), the resulting pattern may be in neither class $C$ nor $D$. To surmount this difficulty, we might first change the cross classification so as to make one set of marginals equal, in the manner discussed by section 5.4 of [35]. Or, more basically, we might redefine $C$ as the set of all patterns with concordance between one of the observations and the other two *and no ties* at all. Although this complicates a bit the formulation of the analog to (16.10), I shall describe the reformulation shortly.

First, however, let us show, analogously to the last part of section 15, that $r_S$ in the no-ties case may be motivated as the value of (16.10) for the sample, each point taken with probability $1/n$. Consider, then, the sample of $(X_i, Y_i)$'s, with no marginal ties, as defining a discrete distribution in which $X$ can take values $X_1, \cdots, X_n$, $Y$ can take values $Y_1, \cdots, Y_n$, and $\rho_{ab}=0$ unless $(a, b)=(X_i, Y_i)$, when $\rho_{ab}=1/n$.

For this case we note that, for $(a, b)$ such that $\rho_{ab}=1/n$,

$$\begin{aligned} \mathrm{I}_{ab}+\mathrm{IV}_{ab} &= (n-R_a)/n \\ \mathrm{I}_{ab}+\mathrm{II}_{ab} &= (n-R_b)/n \\ \mathrm{II}_{ab}+\mathrm{III}_{ab} &= (R_a-1)/n \\ \mathrm{III}_{ab}+\mathrm{IV}_{ab} &= (R_b-1)/n \end{aligned} \tag{16.12}$$

where $R_a$ and $R_b$ are the ranks of the coordinates of that $(X_i, Y_i)$ that equals $(a, b)$. To see why (16.12) holds, one need, for example, only note that

$n(I_{ab}+IV_{ab})$ is the number of observations with $x$ coordinate greater than that of the $(X_i, Y_i)$ equal to $(a, b)$.

From (16.12) we find that (when $\rho_{ab}=1/n$)

$$\mathrm{I}_{ab} - \mathrm{III}_{ab} = \frac{n - R_b}{n} - \frac{R_a - 1}{n} = (n - R_a - R_b + 1)/n$$

and that

$$\mathrm{II}_{ab} - \mathrm{IV}_{ab} = \frac{n - R_b}{n} - \frac{n - R_a}{n} = (R_a - R_b)/n.$$

Substituting in (16.10), and carrying out routine algebra, we find that in this case the denominator of (16.10) is

$$\frac{1}{3}\,\frac{n^2 - 1}{n^2}$$

and that the numerator of (16.10) is

$$\frac{4}{n^3}\sum R_{X_iY_i} - \frac{(n+1)^2}{n^2}$$

whence (16.10), for the uniform distribution over the points $\{(X_i, Y_i)\}$, is just $r_S$ for that sample.

Let us now reformulate the definitions of $C$ and $D$, thus constructing a reasonable second analog to $\rho_S$ in the general case. Define, for three independent observations, the disjoint chance events

$$\begin{aligned} &C^*\text{: at least one of the three is concordant with the other two,}\\ &\qquad\text{and there are } \textit{no} \text{ ties}\\ &D^*\text{: at least one of the three is discordant with the other two,}\\ &\qquad\text{and there are } \textit{no} \text{ ties.} \end{aligned} \tag{16.13}$$

Based on (16.13), the natural generalization of $\rho_S$ would be, not

$$[\Pr\{C\} - \Pr\{D\}]/[\Pr\{C\} + \Pr\{D\}]$$

as in the earlier part of this section, but

$$[\Pr\{C^*\} - \Pr\{D^*\}]/[\Pr\{C^*\} + \Pr\{D^*\}].$$

Both possible generalizations, of course, coincide with $\rho_S$ in the no-ties case. There is no difficulty, in principle, about stating a reasonable estimator for the new generalization: one simply takes the formulas of (16.3) and replaces $C$ by $C^*$, and $D$ by $D^*$.

Next, what will this second generalization of $\rho_S$ be for a cross classification? If we define, in our present spirit, $\Pi_{cc}{}^*$ as the probability that, of three independent observations, observations 2 and 3 are concordant with observation 1 *and* there are no ties among the three, then, in the same way that (16.5) was derived,

$$\Pr\{C^*\} = 3\Pi_{cc}{}^* - 2\Pi_{ccc}.$$

Next, introduce notation analogous to that of (16.6), as follows:

$$\begin{aligned} \mathrm{I}_{ab}^{*2} &= \sum_{\substack{a',a''>a \\ a'\neq a''}}\sum \sum_{\substack{b',b''>b \\ b'\neq b''}}\sum \rho_{a'b'}\rho_{a''b''} \\ \mathrm{II}_{ab}^{*2} &= \sum_{\substack{a',a''<a \\ a'\neq a''}}\sum \sum_{\substack{b',b''>b \\ b'\neq b''}}\sum \rho_{a'b'}\rho_{a''b''} \end{aligned} \tag{16.14}$$

with similar definitions for $\mathrm{III}_{ab}^{*2}$ and $\mathrm{IV}_{ab}^{*2}$. Their interpretations are simple: $\mathrm{I}_{ab}^{*2}$, for example, is the probability that two independent observations from the cross classification are both in the first quadrant with respect to the $(a, b)$ cell *and* are untied in either coordinate.

It is then clear that

$$\Pi_{cc}^* = \sum_a \sum_b \rho_{ab}(\mathrm{I}_{ab}^{*2} + 2\mathrm{I}_{ab}\mathrm{III}_{ab} + \mathrm{III}_{ab}^{*2}) \tag{16.15}$$

and, by direct computation, that the measure of association for cross classifications, following this second route of generalization, is

$$\frac{\sum\sum \rho_{ab}([\mathrm{I}_{ab}^{*2}-2\mathrm{I}_{ab}\mathrm{III}_{ab}+\mathrm{III}_{ab}^{*2})-(\mathrm{II}_{ab}^{*2}-2\mathrm{II}_{ab}\mathrm{IV}_{ab}+\mathrm{IV}_{ab}^{*2})]}{\sum\sum \rho_{ab}[(\mathrm{I}_{ab}^{*2}-2\mathrm{I}_{ab}\mathrm{III}_{ab}+\mathrm{III}_{ab}^{*2})+(\mathrm{II}_{ab}^{*2}-2\mathrm{II}_{ab}\mathrm{IV}_{ab}+\mathrm{IV}_{ab}^{*2})]}\cdot \tag{16.16}$$

Note that (16.16) would be the same as (16.10) if the starred quantities were unstarred. The quantity (16.16) has the following properties:

1. A necessary and sufficient condition that (16.16) be meaningful is the following pair of statements: (a) there are more than two positive $\rho_{a.}$'s and more than two positive $\rho_{.b}$'s; (b) the entire mass of the cross classification is not concentrated on a single row and column.
2. (16.16) lies between $-1$ and 1 inclusive.
3. (16.16) is 1 if and only if $\Pr\{D^*\}=0$. This means that every triple of positive cell probabilities, $\rho_{a_1b_1}$, $\rho_{a_2b_2}$, and $\rho_{a_3b_3}$, without ties between the subscript coordinates, has one of the subscript pairs concordant with the other two.
4. (16.16) is $-1$ if and only if $\Pr\{C^*\}=0$. This means the same as above, but with "discordant" replacing "concordant."
5. (16.16) is zero in the case of independence, but not in general conversely. That (16.16) is zero under independence follows from general considerations, or it may be demonstrated in detail by setting $\rho_{ab}=\rho_{a.}\rho_{.b}$.

For other approaches to $\rho_S$ in the cross classification case, see "Student" [80], Hoeffding [41], and Kendall [51].

Finally, we may ask what (16.16) becomes when we begin with a no-ties $n$-fold sample $\{(X_i, Y_i)\}$ and compute (16.16) for the corresponding $n\times n$ cross classification with $\rho_{ab}=0$ unless $(a, b)=(X_i, Y_i)$, when $\rho_{ab}=1/n$. We suppose that $n\geqq 3$. Notice here that, for an $(a, b)$ with $\rho_{ab}=1/n$ (and we need consider no others),

$$\mathrm{I}_{ab}^{*2} = \mathrm{I}_{ab}^2 - \mathrm{I}_{ab}/n,$$

for two observations in the first quadrant with respect to $(a, b)$ can tie marginally if and only if they fall in the same cell. Similar relationships hold for the other three quadrants. Hence for our particular case, (16.16) is

$$\frac{\sum\sum \rho_{ab}[(\mathrm{I}_{ab}-\mathrm{III}_{ab})^2-(\mathrm{II}_{ab}-\mathrm{IV}_{ab})^2]-\frac{1}{n}\sum\sum \rho_{ab}[\mathrm{I}_{ab}+\mathrm{III}_{ab}-\mathrm{II}_{ab}-\mathrm{IV}_{ab}]}{\sum\sum \rho_{ab}[(\mathrm{I}_{ab}-\mathrm{III}_{ab})^2+(\mathrm{II}_{ab}-\mathrm{IV}_{ab})^2]-\frac{1}{n}\sum\sum \rho_{ab}[\mathrm{I}_{ab}+\mathrm{III}_{ab}+\mathrm{II}_{ab}+\mathrm{IV}_{ab}]}. \tag{16.17}$$

Now the first summations in numerator and denominator may be expressed in terms of ranks, just as we did for (16.10). The second summations are even simpler, since

$$\sum\sum \rho_{ab}(\mathrm{I}_{ab}+\mathrm{III}_{ab}) = \frac{2}{n^2}\binom{n}{2}P_c = \frac{n-1}{n}P_c$$

$$\sum\sum \rho_{ab}(\mathrm{II}_{ab}+\mathrm{IV}_{ab}) = \frac{2}{n^2}\binom{n}{2}P_d = \frac{n-1}{n}P_d$$

where $P_c$ and $P_d$ refer to the $\{(X_i, Y_i)\}$ sample. Substituting, we find that (16.17) is

$$\frac{\frac{4}{n^3}\sum R_{X_i}R_{Y_i} - \frac{(n+1)^2}{n^2} - \frac{n-1}{n^2}t}{\frac{1}{3}\frac{n^2-1}{n^2} - \frac{n-1}{n^2}} = \frac{\frac{n(n^2-1)}{3n^3}r_s - \frac{n-1}{n^2}t}{\frac{1}{3n^2}(n-1)(n-2)} \tag{16.18}$$

$$= \frac{n+1}{n-2}r_s - \frac{3}{n-2}t. \qquad \text{(compare (13.10))}$$

In short, (16.16) for a no-ties sample, considered as a special cross classification, is just (13.10), our more natural estimator of $\rho_S$, for the sample itself.

Since this second generalization of $\rho_S$ is more reasonable than the first in its behavior under independence, we have an additional small argument in favor of (13.10) rather than $r_S$ as an estimator for $\rho_S$, the population quantity to which both converge stochastically.

## 17. HISTORICAL COMMENTS

All the ordinal measures of association discussed above had their beginnings, to the best of my knowledge, in the last years of the nineteenth century, and in the early years of the present century. Forms of these ordinal measures entered the statistical literature only a few years after Francis Galton, Frank Edgeworth,[3] and Karl Pearson had fashioned the correlation coefficient as a tool of statistical analysis. Of course, formal probabilistic discussion of quantities closely related to the correlation coefficient had existed for some time; for discussions of this early work see H. Walker [88, Chap. 5], and K. Pearson [68] and [69]. Pearson, in his biography of Galton, says [67, Vol. II, p. 392,

[3] Edgeworth's contributions to the study and use of the correlation coefficient have not been as widely recognized as their merits may deserve. A discussion and summary of Edgeworth's work on correlation is given by Bowley [11, Chapter 9]. Two of Edgeworth's major articles on the subject are [26] and [27]. Bowley gives an extensive bibliography. K. Pearson [68] was critical of Edgeworth's work on correlation.

and Vol. IIIA, p. 3] that Galton had first tried correlation of ranks or grades but had forsaken that approach in favor of what later became standard bivariate normal correlation theory. This shift is quite understandable when we notice that Galton thought primarily in terms of linear regression. Nonetheless, Galton's own measure of correlation was itself a function of marginal interquartile ranges and the slopes of the lines of conditional medians. See [67, Vol. IIIA, pp. 50–57], for details.

Most of the first papers on ordinal measures of association were wholly in terms of the *sample values;* and indeed this emphasis still continues. It is only in relatively recent years that much attention has been given to the *population meanings* of the measures. In fact, one motivation for this paper is to stress the importance of population meanings. In contrast to this, the population meaning of the correlation coefficient has been emphasized from the time of Galton. Another feature of much work on sample measures of association has been its emphasis on the use of such sample statistics for estimating the correlation coefficient, under the assumption of normality. This is closely related to the lack of discussion of population meaning; for we cannot estimate an undefined quantity.

A quadrant measure of association was first proposed, so far as I am aware, by Sheppard in 1899 [77] who *did* consider its population meaning. However, Fechner, in 1897 had proposed a similar quantity but in a more complex context, that of double time series. See [29, pp. 386–98]; also discussion by O. Anderson [4, pp. 249–50]. Since then, quadrant measures have been discussed from time to time, for example by Thorndike in the psychological literature [81], [82, p. 155], [83], and by Cochran [11a]. An extensive discussion of sampling theory for $q$, the sample analog of the quadrant measure, was given by Blomqvist [10].

It is undoubtedly possible to devise a variety of measures of association related to $\rho$; for example, one might modify $\rho$ so as to stress the "outer" parts of the quadrants along lines suggested, in the hypothesis testing context, by Olmstead and Tukey in their corner test [66].

The essential idea behind the measure of association $\tau$ was first suggested, I believe, by Fechner in 1897 [29, particularly pp. 372–5], although Fechner was mainly concerned, not with association in a bivariate population, but with association between two time-series. (For details on Fechner's work and its relation to $\tau$, see Risser-Traynard [72, p. 109 (Ed. 1), p. 6 (Ed. 2, Liv. II)], and Salvemini [73].) Fechner's suggestion for the double time-series case appears to have become known in France; see March [63] and Lenoir [55, p. 69], for example. Next, $\binom{n}{2}P_c$, a simple linear function of $t$, was proposed by G. F. Lipps in 1905 [60], and $\binom{n}{2}t$ was discussed by him in 1906 [61]. Lipps obtained the first two moments of these quantities under the hypothesis of independence, and suggested their use as a test of independence.

Meanwhile, in France, Binet and his colleagues had proposed, [8] and [9], measuring association by a function of the ranks, essentially the same function as that later called Spearman's foot-rule. These French psychologists appear shortly to have stopped using such measures, see [78, p. 86], although continued interest is suggested by Binet's footnote to [76, p. 492]. A few years later, in

1904, the English psychologist, Spearman, became interested in sample measures of association based on ranks, [78] and [79]. Spearman suggested two measures. The first was $r_S$, motivated by him as the sample correlation coefficient between the ranks. The second was an analogous quantity, but based on $\sum |R_{X_i} - R_{Y_i}|$. This is the so-called Spearman foot-rule; I have not discussed it because it does not seem to estimate a population quantity having a simple probabilistic interpretation.

I return now to $\tau$ and $t$. The German educational psychologist, G. Deuchler, considered $t$ at length in a series of papers beginning in 1909: [20], [21], [22], [23], and [24]. Deuchler began with Lipps' suggestions and carried the discussion much further. His major paper appears to be that of 1914 [21]. In it, Deuchler considered the exact distribution of $t$ under the hypothesis of independence and obtained essentially the same recursion formula as that later worked out by Kendall. By means of this, the exact distribution of $t$ under independence was developed numerically for $n=2, 3, 4, 5$. Deuchler recognized the now familiar relation between $t$ and numbers of inversions, and he developed the generating function for the distribution of number of inversions under the hypothesis of independence. The question of ties was dealt with at length, in a way related both to that later used by Kendall and to that discussed in section 15. The heavily tied case was considered here and also in [24]; for more details see the comments in [36]. A method for simplifying the computation of $t$ was worked out. Lipps' expressions for mean and variance under independence were rederived by the generating function method. Then Deuchler turned to non-independence, and gave, as the mean and variance of $t$, $\tau$ and

$$2(1 - \tau^2)[2n + 5 + (2n - 1)\tau]/[3n(n - 1)(\tau + 3)],$$

respectively.

The above variance expression is different from our (11.3) and it appears to be in error. While no derivation for it was given in [21], the reader was referred for one to a monograph by Deuchler in which the material of his journal articles was expounded in greater detail. This monograph has never been published, but, through the courtesy of Professor Deuchler's widow, I have been permitted to examine it and to retain a microfilm copy of it. In the section on non-independence, Deuchler restricted himself to what he calls "regular dependence," a very special kind of dependence, and, under this restriction, $\Pi_{cc}$ is related to $\Pi_c$ in such a way that Deuchler's variance expression and (11.3) are the same.

The meaning of "regular dependence" seems to be the following: Let $\tilde{Y}_1, \tilde{Y}_2, \cdots$ be the $Y_i$'s corresponding to the *ordered* $X_i$'s in an $n$-fold random sample. $\tilde{Y}_1$ is the $Y$ observation associated with Min $X_i$, etc. Now let $p_1, p_2, \cdots$, be a sequence of nonnegative numbers such that Pr $\{\tilde{Y}_j$ is lowest $Y$ observation$\} = p_j / \sum_{i=1}^{n} p_i$. Assume $p_1 > 0$. Conditionally on $\tilde{Y}_j$ being the lowest $Y$, assume that the probability that $\tilde{Y}_{j'}$ be lowest among the remaining $n-1$ $Y$'s is $p_{j'}/\sum_1^{n-1} p_i$ (if $j'<j$) or $p_{j'-1}/\sum_1^{n-1} p_i$ (if $j'>j$). And so on seriatim. In general, Pr $\{\tilde{Y}_{j_1} < \tilde{Y}_{j_2} < \cdots < \tilde{Y}_{j_{k-1}} < \tilde{Y}_{j_k} <$ other $\tilde{Y}$'s$| \tilde{Y}_{j_1} < \tilde{Y}_{j_2} < \cdots < \tilde{Y}_{j_{k-1}} <$ other $\tilde{Y}$'s$\}$ is $p_s/\sum_1^{n-k+1} p_i$, where $s = j_k$ minus the number of inequalities, $j_1 < j_k, j_2 < j_k, \cdots, j_{k-1} < j_k$, that hold for the particular sequence of $j$'s of interest. The heuristic idea is to work conditionally at the $k$th step on the remaining $(X_i, Y_i)$'s and to suppose that $p_1, p_2, \cdots, p_{n-k+1}$ give the relative probabilities that the $\tilde{Y}$'s of the *remaining* observations be least. This is the "regularity."

One can compute the $p$'s in term of $\tau$ from the above structure. For example if $p_1=1$, then $p_2=(1-\tau)/(1+\tau)$, $p_3=(1-\tau)/(1+\tau)^2$, $p_4=(1-\tau)(3+\tau)/[3(1+\tau)^3]$; in general Deuchler gives an expression equivalent to

$$p_i = \frac{\prod_{j=0}^{i-2} [(j+1) + (j-1)\tau]}{(i-1)!(1+\tau)^{i-1}} .$$

Now, for this "regular dependence" structure, Deuchler's variance is correct and coincides with (11.3). However the structure itself may be criticized, not only for its great restrictiveness, but also because, if one adds natural symmetrical assumptions for greatest (instead of least), then $\tau$ *must* be zero. The idea of regular dependence, nevertheless, seems interesting, and one reason for describing it here is that some modification of it might be a useful restriction on the bivariate distributions of interest. The search for such useful restrictions has recently been exemplified by Konijn [53]. Another article in the same direction is that of Barton and David [5].

Deuchler also considered in [20] the approximation of the distribution of $t$ in general. A critical discussion of the two measures proposed by Spearman was presented, and finally a detailed discussion of $2\times 2$ tables appeared. The $2\times 2$ table question was also extensively discussed in [22]. In [23] Deuchler applied various measures of association, including $t$, to the study of arithmetical skills in children. A biographical note about Deuchler is given in [54].

Although the suggestions of Lipps and Deuchler were discussed in the German psychometric literature of the time (see, for example, Betz [7], Wirth [91], and Valentiner [85] and [86]), these suggestions do not seem to have aroused much interest in other countries or in other fields. For example, I have found no references to them in the standard German statistical treatises, such as those of E. Czuber. However, in Switzerland, especially in medical and psychiatric work, Lipps' original work does seem to have had some influence. I have run across a constellation of articles by Swiss authors, including Bersot [6], Haemig [39], de Montet [18] and [19], Piguet [70] and Michaud [64], in which use of Lipps' suggestion was made.

In 1924, Esscher [28] independently suggested $\tau$ and $t$ as measures of association, and gave a clear statement of the population meaning of $\tau$. Esscher restricted his considerations to the bivariate normal case, but this is inessential for his basic statements. In the bivariate normal case, Esscher obtained the variance of $\sin(\pi/2)t$ and an inequality for this variance.

Lindeberg, a little later, [56] and [57], considered $\tau$ and $t$ without the normality restriction, gave a clear statement of the meaning of $\tau$, and presented —so far as I know for the first time—the variance of $t$ in general.

In 1928 or before, S. D. Holmes proposed $t$ as an approximation to the ordinary correlation coefficient. Holmes' approach has been described by P. Sandiford [75, Appendix B], and recently by H. D. Griffin [37]. The approach is graphical in nature and represents a neat way of computing $t$ that seems to be generally unknown.

The quantity $\Pi_c$ was independently proposed by de Finetti in 1937 [17] as a measure of association. De Finetti motivated and interpreted $\Pi_c$ in essentially the same way as does section 4 of this paper.

In 1938, Kendall [50] independently proposed $t$, and began a series of papers dealing extensively with ordinal measures of correlation. A thorough discussion is given in Kendall's monograph [51]. Kendall, and other English workers in this area, have written extensively on one topic not treated here at all, the distribution of $t$ and $r_S$ when sampling *without* replacement from a finite population.

Hoeffding, in a fundamental paper on asymptotic distribution theory [44], states in passing the population interpretations of both $\tau$ and $\rho_S$. Kendall [51, 2nd ed., chap. 9] briefly discusses the population interpretations of $\tau$ and $\rho_S$, apparently basing his discussion on the doctoral thesis of R. M. Sundrum.

Other measures of association have been suggested from time to time. For example, Lipmann [59] proposed various measures based on the order statistics of $|R_{X_i}-R_{Y_i}|$; but see critical comments by Wirth [92]. Von Schelling [87] proposed a curious kind of measure that does not seem to have aroused interest elsewhere. About forty years ago Gini [34] proposed a symmetrized version of Spearman's foot-rule that has been frequently discussed in the Italian literature since then; for example see Amato [3] and Salvemini [73] and [74]. A crude form of this quantity had been discussed by Tönnies in 1909; see [36] for details. Related discussions have been given by Watkins [89] and Julin [47]. There is an extensive Italian literature, centered about the work of Gini, on the whole subject of measures of association, and there have been a number of interesting contributions by Fréchet; for more detail on these and other contributions, and for bibliography, I refer to [36].

Measures of association based upon the Shannon-Wiener information concept have been proposed in recent years. For a discussion of these I refer to a paper by E. H. Linfoot [58]. Fieller, Hartley, and Pearson [30a] discuss the sample ordinal measure of association obtained by computing the correlation coefficient, not of the observations, but of the corresponding mean normal order statistics. This is an application of a proposal made earlier by Fisher and Yates.

Measures of partial and multiple association that are closely related to the bivariate ordinal measures of association discussed here have been proposed from time to time. Discussions of this topic will be found in Kendall's monograph [51] and in an article by Goodman and me [35].

All of the measures of association discussed in detail in this paper have the property that they may be zero even when $X$ and $Y$ are stochastically dependent. This has led some authors to seek measures of association that are zero if and only if $X$ and $Y$ are independent. In particular, the works of Fréchet, Gini, Steffensen, Cramér, and Pollaczek-Geiringer cited in [36] bear on this point. Hoeffding's [40, 41, and 42] should be mentioned here, as well as a later paper [45]. See also Féron [30].

Daniels [13] suggested a formal synthesis for certain ordinal sample measures of association by considering them as sample correlation coefficients from an $n^2$-fold sample constructed from the original $n$-fold sample, $(X_i, Y_i), i=1, \cdots, n$, via functions $a$ and $b$: $(a(X_i, X_j), b(X_i, X_j)); i, j=1, \cdots, n$. It is required that $a(X_i, X_i)=b(X_i, X_i)=0$ and that $a(X_i, X_j)=-a(X_j, X_i)$ and $b(X_i, X_j) = -b(X_i, X_j)$. Konijn [53, pp. 306–7], has given, in concise terms, a similar synthesis. A related synthesis, in terms of general measures of dispersion, has been suggested by Amato [2] and [3].

Kendall's monograph [51], contains an extensive discussion of recent developments regarding ordinal measures of association, and the reader may be referred to this book for a discussion of recent work not mentioned here.

There has been an enormous amount of statistical work with the goal of making precise and useful our intuitive notions of stochastic association. I have tried to survey with reasonable completeness that portion of this work dealing with ordinally invariant measures of association, but the literature is so extensive and scattered that this survey is almost surely not completely comprehensive.

## REFERENCES

[1] Adler, Leta McKinney, "A modification of Kendall's tau for the case of arbitrary ties in both rankings," *Journal of the American Statistical Association*, 52 (1957), 33–5.

[2] Amato, Vittorio, "Sulla determinazione degli indici di cograduazione mediante gli indici di variabilità," pp. 26–42 in *Studi Sulle Relazioni Statistiche, Studi e Monografie della Società Italiana di Economia Demografia e Statistica*, Number 6 (1952–3).

[3] Amato, Vittorio, "Un criterio per la determinazione di indici di cograduazione tra serie statistiche con ripetizioni," *Statistica* (Italy), 14 (1954), 54–66.

[4] Anderson, Oskar, *Probleme der Statistischen Methodenlehre in den Sozialwissenschaften*, Second edition, Physica-Verlag, Würzburg, 1954. No. 6 of *Einzelschriftender Deutschen Statistischen Gesellschaft.*

[5] Barton, D. E., and David, F. N., " 'Spearman's rho' and the matching problem," *British Journal of Statistical Psychology*, 9 (1956), 69–73.

[6] Bersot, Hri., "Variabilité et corrélations organiques. Nouvelle étude du réflexe plantaire. Chap. IV, Détermination du sens de la variation. Méthodes des $+-$et des $= \neq$," *Schweizer Archiv für Neurologie und Psychiatrie (Archives Suisses de Neurologie et de Psychiatrie)*, 5 (1919), 305–24. Chap. V, 6 (1920), 37–79, is also of some interest.

[7] Betz, W., *Über Korrelation*, J. A. Barth, Leipzig, 1927. Zweite, überarbeitete Auflage. First edition was in *Beihefte zur Zeitschrift für Angewandte Psychologie*, 3 (1911), 88 pp.

[8] Binet, A., and Vaschide, N., "Corrélation des épreuves physiques," *L'Année Psychologique*, 4 (1897–8), 142–72.

[9] Binet, A., and Henri, V., *La Fatigue Intellectuelle*, Schleicher Frères, Paris, 1898.

[10] Blomqvist, Nils, "On a measure of dependence between two random variables," *Annals of Mathematical Statistics*, 21 (1950), 593–600.

[11] Bowley, A. L., *F. Y. Edgeworth's Contributions to Mathematical Statistics*, Royal Statistical Society, London, 1928.

[11a] Cochran, W. G., "The efficiencies of the binomial series tests of significance of a mean and of a correlation coefficient," *Journal of the Royal Statistical Society*, 100 (1937), 69–73.

[12] Cramér, Harald, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[13] Daniels, H. E., "The relation between measures of correlation in the universe of sample permutations," *Biometrika*, 33 (1944), 129–35.

[14] Daniels, H. E., "Rank correlation and population models," *Journal of the Royal Statistical Society*, Series B, 12 (1950), 171–81.

[15] Daniels, H. E., and Kendall, M. G., "The significance of rank correlations where parental correlation exists," *Biometrika*, 34 (1947), 197–208.

[16] David, S. T., Kendall, M. G., and Stuart, A., "Some questions of distribution in the theory of rank correlation," *Biometrika*, 38 (1951), 131–40.

[17] De Finetti, Bruno, "A proposito di 'correlazione'," *Supplemento Statistico ai Nuovi Problemi di Politica Storia ed Economia*, 3 (1–2), (1937).

[18] De Montet, Ch., "La détermination quantitative de la variabilité physiologique," *Correspondenz-Blatt für Schweizer Aerzte*, 46 (52) (1916), 1729–41 and 1761–71.

[19] De Montet, Ch., "Le calcul des variations et la notion de corrélation en biologie exposés à l'aide de recherches sur la sensibilité à la pression dans les lésions de certains nerfs périphériques," *Zur Diagnose und Behandlung der Spätfolgen von Kriegsverletzungen, Mitteilungen aus der Armeesanitätsanstalt für Internierte in Luzern (Schweiz)*, II Lieferung (1919), 107–48. Ed. by Veraguth, Otto; Brun, Hans; and Hössly, Hans. Pub. Rascher et Cie., Zürich.

[20] Deuchler, G., "Beiträge zur Erforschung der Reaktionsformen," *Psychologische Studien*, 4 (1909), 353–430.

[21] Deuchler, Gustav, "Über die Methoden der Korrelationsrechnung in der Pädagogik und Psychologie," *Zeitschrift für Pädagogische Psychologie und Experimentelle Pädagogik*, 15 (1914), 114–31, 145–59, and 229–42.

[22] Deuchler, Gustav, "Über die Bestimmung einseitiger Abhängigkeit in pädagogisch-psychologischen Tatbeständen mit alternativer Variabilität," *Zeitschrift für Pädagogische Psychologie und Experimentelle Pädagogik*, 16 (1915), 550–66.

[23] Deuchler, Gustav, "Beiträge zur Psychologie der Rechenübung und Rechenfertigkeit," *Zeitschrift für Pädagogische Psychologie und Experimentelle Pädagogik*, 17 (1916), 86–95 and 420–49.

[24] Deuchler, Gustav, "Über die Bestimmung von Rangkorrelationen aus Zeugnisnoten," *Zeitschrift für Angewandte Psychologie*, 12 (1917), 395–439.

[25] Durbin, J., and Stuart, A., "Inversions and rank correlations," *Journal of the Royal Statistical Society, Series B*, 13 (1951), 303–9.

[26] Edgeworth, F. Y., "Correlated averages," *Philosophical Magazine*, (5) 34 (1892), 190–204.

[27] Edgworth, F. Y., "Exercises in the calculation of errors," *Philosophical Magazine*, (5) 36 (1893), 98–111.

[28] Esscher, Frederik, "On a method of determining correlation from the ranks of variates," *Skandinavisk Aktuarietidskrift*, 7 (1924), 201–19.

[29] Fechner, Gustav Theodor, *Kollektivmasslehre*, Wilhelm Engelmann, Leipzig, 1897. (Published posthumously; completed and edited by G. F. Lipps.)

[30] Féron, Robert, "Information, régression, corrélation," *Publications de l'Institut de Statistique de l'Université de Paris*, 5 (1956), 113–215.

[30a] Fieller, E. C., Hartley, H. O., and Pearson, E. S., "Tests for rank correlation coefficients. I," *Biometrika*, 44 (1957), 470–81.

[31] Fréchet, Maurice, "Sur le coefficient, dit de corrélation," *Comptes-Rendus de L'Académie des Sciences de Paris*, 197 (1933), 1268–9.

[32] Fréchet, Maurice, "Sur deux relations simples entre le 'coefficient' de corrélation et le 'rapport' de corrélation," *Comptes-rendus du Deuxième Congrès des Mathématiciens des Pays Slaves*, in *Časopis Pro Pěstování Matematiky a Fysiky Část Vědecka*, 64 (1934–5), 209–10.

[33] Fréchet, Maurice [and others], "Sur l'usage du coefficient de corrélation. Communication, commentaires et rapport," *Bulletin de l'Institut International de Statistique*, 28(2) (1935), 24–53.

[34] Gini, Corrado, "Di una misura delle relazioni tra le graduatorie di due caratteri." Appendix to: Mancini, A., *Le Elezioni Generali Politiche del 1913 nel Comune di Roma*, Ludovico Cecchini, Rome, 1914.

[35] Goodman, Leo A., and Kruskal, William H., "Measures of association for cross classifications," *Journal of the American Statistical Association*, 49 (1954), 732–64.

[36] Goodman, Leo A., and Kruskal, William H., "Measures of association for cross classifications II: Further discussion and references," to be published in *Journal of the American Statistical Association.*

[37] Griffin, Harold D., "Graphic computation of tau as a coefficient of disarray," *Journal of the American Statistical Association,* 53 (1958), 441–7.

[38] Guldberg, Alf, "On correlation," *Norsk Matematisk Forenings Skrifter,* I(5) (1921), 11–23.

[39] Haemig, E., "Systematische Abhängigkeitsbestimmungen," *Mitteillungen der Vereinigung Schweizerischer Verisicherungsmathematiker (Bulletin de l'Association des Actuaires Suisses),* 12 (1917), 299–333.

[40] Höffding, Wassilij, "Massstabinvariante Korrelationstheorie," *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin,* 5(3) (1940), 181–233.

[41] Höffding, Wassilij, "Massstabinvariante Korrelationsmasse für diskontinuierliche Verteilungen," *Archiv für Mathematische Wirtschafts- und Sozialforschung,* 7 (1941), 49–70.

[42] Höffding, Wassilij, "Stochastische Abhängigkeit und funktionaler Zusammenhang," *Skandinavisk Aktuarietidskrift,* 25 (1942), 200–27.

[43] Hoeffding, Wassily, "On the distribution of the rank correlation coefficient $\tau$ when the variates are not independent," *Biometrika,* 34 (1947), 183–96.

[44] Hoeffding, Wassily, "A class of statistics with asymptotically normal distributions," *Annals of Mathematical Statistics,* 19 (1948), 293–325.

[45] Hoeffding, Wassily, "A non-parametric test of independence," *Annals of Mathematical Statistics,* 19 (1948), 546–57.

[46] Hotelling, Harold, and Pabst, Margaret Richards, "Rank correlations and tests of significance involving no assumptions of normality," *Annals of Mathematical Statistics,* 7 (1936), 29–43.

[47] Julin, A., "De quelques méthodes de détermination de la covariation," *Revue de l'Institut International de Statistique,* 3 (1935), 251–77.

[48] Kaarsemaker, L., and van Wijngaarden, A., "Tables for use in rank correlation," *Statistica* (Netherlands), 7 (1953), 41–54.

[49] Kapteyn, J. C., "Definition of the correlation coefficient," *Monthly Notices of the Royal Astronomical Society,* 72 (1912), 518–25.

[50] Kendall, M. G., "A new measure of rank correlation," *Biometrika,* 30 (1938), 81–93.

[51] Kendall, Maurice G., *Rank Correlation Methods,* Charles Griffin, London, and Hafner, New York, 1948 (first edition) and 1955 (second edition).

[52] Kendall, M. G., Kendall, S. F. H., and Babington Smith, B., "The distribution of Spearman's coefficient of rank correlation in a universe in which all rankings occur an equal number of times," *Biometrika,* 30 (1938), 251–73.

[53] Konijn, H. S., "On the power of certain tests for independence in bivariate populations," *Annals of Mathematical Statistics,* 27 (1956), 300–23. Corrections in same journal, 29 (1958) 935–6. For related work see: Konijn, H. S., "A class of two-dimensional random variables and distribution functions," *Sankhyā* 18 (1957), 167–72.

[54] Kruskal, William H., "Historical note on the Wilcoxon unpaired two-sample test," *Journal of the American Statistical Association,* 52 (1957), 356–60.

[55] Lenoir, Marcel, *Études sur la Formation et le Mouvement des Prix,* M. Giard and E. Brière, Paris, 1913.

[56] Lindeberg, J. W., "Über die Korrelation," *VI Skandinaviska Mathematikerkongress,* Copenhagen, 1925, 437–46.

[57] Lindeberg, J. W., "Some remarks on the mean error of the percentage of correlation," *Nordic Statistical Journal,* 1 (1929), 137–41.

[58] Linfoot, E. H., "An informational measure of correlation," *Information and Control,* 1 (1957), 85–9.

[59] Lipmann, Otto, *Abzählende Methoden und ihre Verwendung in der Psychologischen Statistik,* J. A. Barth, Leipzig, 1921.

[60] Lipps, Gottl. Friedr., "Die Bestimmung der Abhängigkeit zwischen den Merkmalen eines Gegenstandes," *Berichte über die Vorhandlungen der Königlich Sachsischen*

*Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse*, 57 (1905), 1–32.

[61] Lipps, G. F., *Die Psychischen Massmethoden*, F. Vieweg und Sohn, Braunschweig, 1906. Volume 10 of the monograph series, *Die Wissenschaft.*

[62] Litchfield, John T., Jr., and Wilcoxon, Frank, "The rank correlation method," *Analytical Chemistry*, 27 (1955), 299–300.

[63] March, Lucien, "Comparaison numérique de courbes statistiques," *Journal de la Société de Statistique de Paris*, 46 (1905), 255–77 and 306–11.

[64] Michaud, L., "La valeur du calcul des probabilités pour la médicine interne," *Schweizerische Medizinische Wochenschrift*, 1 (1920), 743–51.

[65] Olds, E. G., "Distributions of sums of squares of rank differences for small numbers of individuals," *Annals of Mathematical Statistics*, 9 (1938), 133–48.

[66] Olmstead, P. S., and Tukey, John W., "A corner test for association," *Annals of Mathematical Statistics*, 18 (1947), 495–513.

[67] Pearson, Karl, *The Life, Letters, and Labours of Francis Galton*, Cambridge, Vol. I, 1914; Vol. II, 1924; Vol. IIIA, 1930; Vol. IIIB, 1930.

[68] Pearson, Karl, "Notes on the history of correlation," *Biometrika*, 13 (1920–1), 25–45.

[69] Pearson, Karl, "The contribution of Giovanni Plana to the normal bivariate frequency surface," *Biometrika*, 20A (1928), 295–8.

[70] Piguet, Charles-Abel, *Essai sur la détermination exacte des rapports existant entre certains facteurs morbides dans le diabète sucré*, Thesis, Université de Lausanne, Clinique Médicale, Leysin, 1917.

[71] Quenouille, M. A., *Associated Measurements*, Butterworths Scientific Publications, London, 1952.

[72] Risser, R., and Traynard, C.-E., *Les Principes de la Statistique Mathématique*, Gauthier-Villars, Paris, 1933. Fascicule IV of Tome I, *Les Principes de la Théorie des Probabilités*, of: Borel, Émile (ed.), *Traité du Calcul des Probabilités et de ses Applications*. Second Edition: Livre I, 1957; Livre II, 1958.

[73] Salvemini, Tommaso, "Sui vari indici di cograduazione," *Statistica* (Italy), 11(2) (1951), 133–54.

[74] Salvemini, T., "Indici di connessione e indici di concordanza," *Genus*, 9 (1950–2), 283–92.

[75] Sandiford, Peter, *Educational Psychology, An Objective Study*, Longmans, Green and Co., London, New York, and Toronto, 1928.

[76] Sée, A., "Une formule mathématique applicable aux recherches sur la psychologie," *Bulletin de la Société Libre pour l'Étude Psychologique de l'Enfant*, 4 (1904), 492–8.

[77] Sheppard, W. F., "On the application of the theory of error to cases of normal distribution and normal correlation," *Philosophical Transactions of the Royal Society of London* (A), 192 (1899), 101–67. Errata on p. 531.

[78] Spearman, C., "The proof and measurement of association between two things," *American Journal of Psychology*, 15 (1904), 72–101.

[79] Spearman, C., " 'Footrule' for measuring correlation," *British Journal of Psychology*, 2 (1906), 89–108.

[80] "Student" [W. S. Gosset], "An experimental determination of the probable error of Dr Spearman's correlation coefficients," *Biometrika*, 13 (1920–1), 263–82. Reprinted in: Pearson, E. S., and Wishart, John (ed.), *"Student's" Collected Papers*, Biometrika Office, University College, London, 1942; pp. 70–89.

[81] Thorndike, Edward L., "Empirical studies in the theory of measurement," *Archives of Psychology*, 3 (1907), 45 pp.

[82] Thorndike, Edward L., *An Introduction to the Theory of Mental and Social Measurements*, Teachers College, Columbia University, New York, Second Edition, 1919.

[83] Thorndike, Edward L., "On correlations between measurements which are not normally distributed," *Journal of Educational Psychology*, 28 (1937), 367–70.

[84] Thornton, G. R., "The significance of rank difference coefficients of correlation," *Psychometrika*, 8 (1943), 211–22.

[85] Valentiner, S., "Die Verwendung des Masses des Verdachtes auf systematische

Fehler in der experimentellen Pädagogik," *Sitzungsberichte der Gesellschaft zur Beförderung der Gesammten Naturwissenschaften zu Marburg*, (1920), 43–6.

[86] Valentiner, S., "Über das Mass der Abhängigkeit von Begabungen (Korrelationslehre)," *Zeitschrift für Angewandte Psychologie*, 18 (1921), 19–32.

[87] Von Schelling, H., "Korrelationsmessung auf Grund der Anordnung der Beobachtungen," *Zeitschrift für Angewandte Mathematik und Mechanik*, 12 (1932), 377–80.

[88] Walker, Helen M., *Studies in the History of Statistical Method*, Williams and Wilkins, Baltimore, 1929.

[89] Watkins, G. P., "An ordinal index of correlation," *Journal of the American Statistical Association*, 28 (1933), 139–51.

[90] Whitefield, J. W., "Intra-class rank correlation," *Biometrika*, 36 (1949), 463–7.

[91] Wirth, Wilhelm, *Spezielle Psychophysische Massmethoden*, Urban and Schwarzenberg, Berlin and Vienna, 1920. Part of Abderhalden, Emil (ed.), *Handbuch der Biologischen Arbeitsmethoden;* Lieferung 4, Abt. 6, Methoden der Experimentellen Psychologie, A; Heft 1.

[92] Wirth, W., "Die psychotechnische Brauchbarkeit des Spearman'schen Korrelationskoeffizienten zumal für Augenmassprüfungen," *Industrielle Psychotechnik*, 2 (1925), 22–31.

[93] Wolfowitz, J., "Non-parametric statistical inference," pp. 93–113 of *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley and Los Angeles, 1949.