MEASURES OF ASSOCIATION FOR CROSS CLASSIFICATIONS III: APPROXIMATE SAMPLING THEORY*

LEO A. GOODMAN AND WILLIAM H. KRUSKAL
University of Chicago

The population measures of association for cross classifications, discussed in the authors' prior publications, have sample analogues that are approximately normally distributed for large samples. (Some qualifications and restrictions are necessary.) These large sample normal distributions with their associated standard errors, are derived for various measures of association and various methods of sampling. It is explained how the large sample normality may be used to test hypotheses about the measures and about differences between them, and to construct corresponding confidence intervals. Numerical results are given about the adequacy of the large sample normal approximations. In order to facilitate extension of the large sample results to other measures of association, and to other modes of sampling, than those treated here, the basic manipulative tools of large sample theory are explained and illustrated.

CONTENTS

APPENDIX

* This research was supported in part by the Army Research Office, the Office of Naval Research, and the Air Force Office of Scientific Research by Contract No. Nonr-2121(23); and in part by Research Grant NSF-G21058 from the Division of Mathematical, Physical, and Engineering Sciences of the National Science Foundation.

Part of Mr. Goodman's work on this paper was done at the Statistical Laboratory of the University of Cambridge under a Fulbright Award and a Social Science Research Council Fellowship. Part of Mr. Kruskal's work on this paper was done at the Department of Statistics, University of California, Berkeley.

We thank the following persons for helpful comments: H. E. Daniels, B. L. Fox, R. Kozelka, J. Mincer, I. R. Savage, R. Sneyers, and A. Stuart. We also wish to acknowledge computational work on this paper by M. DeGroot, W. Goldfarb, and G. Pinkham.

1. INTRODUCTION AND SUMMARY

PROBLEMS connected with measuring the degree of association between two or more cross classifications or polytomies were considered by us in [8]. A number of possible measures or indexes of association were discussed for situations in which the meaning of the term "degree of association" is not completely clear-cut within some precisely stated model. The central theme of [8] was that measures of association should have operationally meaningful interpretations that are relevant in the contexts of empirical investigations in which the measures are used.[1]

A supplementary discussion [9] presented further measures of association, together with historical and bibliographical material.

The discussion in references [8] and [9] supposed almost throughout that the parent population is known, so that no sampling problems arise. In the present paper, we develop, by asymptotic (i.e., large sample) methods, approximate sampling theory for the measures considered in [8]; without having such a theory in practical form, the measures of [8] are of limited use. We use this sampling theory in connection with testing hypotheses about the measures and establishing confidence intervals for the measures. We also include here some material about the adequacy of the asymptotic approximations; we expect to present further material in a later publication.

The notation of [8] will be used freely, but we shall generally try to recapitulate for the reader's convenience. We restrict ourselves to sample sizes fixed in advance, and most, but not all, of this paper deals with the case of two polytomies or cross classifications, A with α classes, and B with β classes. The population or true frequency for the cell with A classification A_a and B classification B_b is denoted by ρ_{ab}; we set $\rho_{a\cdot} = \sum_b \rho_{ab}$ and $\rho_{\cdot b} = \sum_a \rho_{ab}$.

In developing an asymptotic sampling theory for the measures proposed in [8], a number of considerations arise.

i. Sampling methods. There are several possible sampling methods. For example, one may choose a random sample, in either the sense of "with replacement" (infinite population) or "without replacement," from a population of individuals that is cross classified into the $\alpha\beta$ cells obtained by crossing classifications A and B. In this case, the sampling method leads to a multinomial (with replacement) or to a generalized hypergeometric (without replacement) distribution.

---

[1] The measures of association considered in our papers may be appropriate in situations where little or no structural information is available about the true relative frequencies in the cells of the cross classification. If, contrariwise, a structural parametric model is assumed, it will often be the case that one or more of its parameters will have obvious meanings as measures of association within the terms of the assumed model.

We do not think it meaningful to speak of *the* most appropriate measure of association for most of the unstructured situations we have in mind. What is central is that the association measures used should have meaningful interpretations; it is quite likely that several measures, each with its own interpretation, might all be useful in a given situation.

Alternatively, one might sample independently *within* each A_a class or row of the $A \times B$ cross classification, obtaining α independent multinomial or generalized hypergeometric distributions. Then the relative sample sizes for the α classes of classification A would be of importance. Similarly, one might sample independently within each B_b class or column. There are still other possibilities, but we shall not deal with them here.

For the sake of simplicity, we assume infinite populations throughout; that is, sampling with replacement in the technical sense. Much of the work refers to the case of a multinomial sample over all the $\alpha\beta$ cells, but we do consider some cases of independent sampling in the rows or columns.

ii. Auxiliary knowledge. Auxiliary knowledge may vary considerably. Thus one may know marginal totals, the $\rho_{a\cdot}$'s, and/or the $\rho_{\cdot b}$'s. (For treatments of this, see [5], [25], and [6].) Or one may know the values of the b subscripts maximizing ρ_{ab} without actually knowing the numerical values of the ρ's. We generally assume no auxiliary knowledge, except for uniqueness assumptions to be described, but in Section 4 we do consider some cases in which auxiliary knowledge is utilized.

iii. Choice of estimators. The question of what estimator to use for a given measure is resolved here by using the obvious sample analogue of the population measure. In all cases but one, (4.2.1), this is the maximum likelihood estimator.[2] For testing and confidence intervals we again have used intuitively straightforward procedures based on the point estimators and their distributions.

iv. Uniqueness of maxima. Many of the asymptotic results depend on assumptions like this: there is just one value of b maximizing $\rho_{\cdot b}$; we shall generically say that such assumptions are those of uniqueness of maxima. Even when there is a unique value of b maximizing $\rho_{\cdot b}$, there may be other values of b for which $\rho_{\cdot b}$ is very near the maximum. In such cases, particularly large sample sizes may be needed before the asymptotic distributions we discuss become good approximations to the actual distributions.

v. Asymptotic approximations and their possible modifications. For each estimator and sampling method, our procedure is to find a function of the sample, an approximate standard error (ASE), such that the difference between the estimator and the true value of the measure being estimated, divided by the ASE, is for large samples approximately unit-normal (normal with zero mean and unit standard deviation);

$$(\text{estimator} - \text{true value})/(\text{ASE}) \approx N(0, 1)$$

for large samples.[3] Given such an approximation (corresponding to convergence in distribution), one may modify the estimator, the ASE, or both in many ways

[2] This paper studies multi-parameter situations, and, in fact, we generally consider $\alpha\beta-1$ independent parameters, one for the probability of each cell of the cross classification, modified by the restriction that these probabilities sum to one. We adopt the usual convention that, if $(\hat{\theta}_1, \cdots, \hat{\theta}_k)$ is the maximum likelihood estimator of $(\theta_1, \cdots, \theta_k)$, then by "the maximum likelihood estimator of $f(\theta_1, \cdots, \theta_k)$" we mean $f(\hat{\theta}_1, \cdots, \hat{\theta}_k)$. This convention is justified by the invariance of maximum likelihood estimation under reparameterization.

[3] Other approximations than normal ones are in principle possible.

without destroying mathematical convergence to unit-normality, but with possible improvement in the normal approximation for finite samples. It is also possible to consider a transformation of the true value and the estimator. Such modifications or transformations have been widely used to improve asymptotic approximations. (For example, corrections for continuity may be viewed from this standpoint.) We shall present a few possible transformations with their corresponding ASE's, and we expect to present further material in a later publication.

2. NOTATION AND PRELIMINARIES

A sample of n individuals is drawn in some specified manner. The A and B classifications of each member of the sample are observed. Let N_{ab} be the number of sample individuals that fall in the (A_a, B_b) cell; that is, N_{ab} is the number of individuals having A classification A_a and B classification B_b. Thus

$$\sum_{a=1}^{\alpha} \sum_{b=1}^{\beta} N_{ab} = n.$$

In the case of nonrestricted sampling, multinomial over the entire $\alpha \times \beta$ cross classification table, the marginals will be denoted in the conventional manner,

$$N_{a\cdot} = \sum_b N_{ab}, \qquad N_{\cdot b} = \sum_a N_{ab}. \tag{2.1}$$

In most other sampling methods, at least one set of these marginals will not be random but fixed in advance. In general, we shall use capital Latin letters for random variables and lower case Latin letters for corresponding fixed numbers, e.g., $n_{a\cdot} = \sum_b N_{ab}$ for fixed row marginals. The Latin letters to be used will, whenever feasible, be related to the Greek letters used for corresponding population quantities.

Thus R_{ab} (corresponding to ρ_{ab}) will be used for the proportion, N_{ab}/n, of observations in the (A_a, B_b) cell, $R_{a\cdot}$ for $N_{a\cdot}/n = \sum_b R_{ab}$ (when the row marginals are random), and so on; $\sum\sum_{ab} R_{ab} = 1$. It is convenient to work with the R_{ab}'s for present purposes, but the more important formulas will also be given in terms of the N_{ab}'s for convenience in applications. Special mention must be made of the notation for maxima. We denote by N_{am} the maximum over b of N_{ab}, with analogous notation for other maxima as follows:

$$\begin{aligned} N_{am} &= \operatorname*{Max}_b N_{ab} \\ N_{mb} &= \operatorname*{Max}_a N_{ab} \\ N_{m\cdot} &= \operatorname*{Max}_a N_{a\cdot} \quad (\text{or } n_{m\cdot} = \operatorname*{Max}_a n_{a\cdot}) \\ N_{\cdot m} &= \operatorname*{Max}_b N_{\cdot b} \quad (\text{or } n_{\cdot m} = \operatorname*{Max}_b n_{\cdot b}). \end{aligned} \tag{2.2}$$

The notation for R_{am}, $R_{\cdot m}$, etc. will be similar. It should be noted that a symbol like "$N_{m\cdot}$" is a single unit, meaning "the largest $N_{a\cdot}$".

From now until further notice (in Section 4), we assume multinomial sampling over all the $\alpha\beta$ cells. We shall throughout be considering asymptotic behavior as $n \rightarrow \infty$, so that in principle "n" should be attached to all symbols for random variables to indicate that we have in mind a sequence of samples: $N_{ab}^{(n)}$. For the sake of simplicity, however, we shall omit the "n."

One fact is basic: the $\alpha\beta$ random variables $\sqrt{n}(R_{ab}-\rho_{ab})$ jointly converge in distribution[4] to the multivariate normal distribution with all means zero, and with variances and covariances given by

$$\begin{aligned} &\operatorname{Var}[\sqrt{n}(R_{ab} - \rho_{ab})] = \rho_{ab}(1 - \rho_{ab}) \\ &\operatorname{Cov}[\sqrt{n}(R_{ab} - \rho_{ab}), \sqrt{n}(R_{a'b'} - \rho_{a'b'})] = - \rho_{ab}\rho_{a'b'} \quad (a \neq a' \text{ or } b \neq b'). \end{aligned} \tag{2.3}$$

(These means, variances, and covariances also are correct for any finite n.) A reference to this basic fact is p. 419 of [2]. It follows from the above that each R_{ab} converges in probability[4] to ρ_{ab}.

We now make, until further notice, the following assumptions about the population:

$$\begin{aligned} &\text{For each } a,\ \rho_{am} = \rho_{ab} \text{ for a unique value of } b. \\ &\text{For each } b,\ \rho_{mb} = \rho_{ab} \text{ for a unique value of } a. \\ &\rho_{\cdot m} = \rho_{\cdot b} \text{ for a unique value of } b. \\ &\rho_{m\cdot} = \rho_{a\cdot} \text{ for a unique value of } a. \end{aligned} \tag{2.4}$$

These are the assumptions mentioned in the preceding section. Without them no useful asymptotic theory for the distribution of estimators of the λ coefficients seems possible at the present time. (We need not always make all four assumptions; for example, in the case of λ_b the assumptions relating to ρ_{am} and $\rho_{\cdot m}$ will suffice. None of these uniqueness assumptions are needed in the case of γ.) An assumption that will be made throughout is that the population or true value of the measure of association in question is well defined.

Under these assumptions, and by convergence in probability, the probability that $R_{am}=R_{ab}$ approaches unity for that value of b such that $\rho_{am}=\rho_{ab}$. Similarly, the probability that $R_{\cdot m}=R_{\cdot b}$ approaches unity for that value of b such that $\rho_{\cdot m}=\rho_{\cdot b}$. Hence, for our asymptotic purposes, we may act as if R_{am} is taken on at that value of b such that $\rho_{am}=\rho_{ab}$ (see Section A4). Similarly, we may act as if $R_{\cdot m}$ is taken on at that value of b such that $\rho_{\cdot m}=\rho_{\cdot b}$. The same statements hold of course for R_{mb} and $R_{\cdot m}$.[5]

---

[4] This *convergence in distribution* means that the probability that the $\sqrt{n}$ $(R_{ab}-\rho_{ab})$ together satisfy any fixed (measurable) set of conditions has the limit, as $n \rightarrow \infty$, given by the probability that random variables X_{ab} satisfy the same set of conditions, where the X_{ab} are governed by the indicated multivariate normal distribution. To say that R_{ab} *converges in probability* to ρ_{ab} means that the probability that $\rho_{ab}-\epsilon_1 \leq R_{ab} \leq \rho_{ab}+\epsilon_2$ approaches unity as $n \rightarrow \infty$, for any positive ϵ_1 and ϵ_2.

[5] Exact distribution theory for the observed maximum frequency in a sample from a multinomial distribution seems intractable. Related material is discussed in [10], [16], and [17].

It would be impracticable to attempt to give here asymptotic distributions relevant to estimators of *each* measure discussed in [8] under *many* sampling methods and *many* assumptions about auxiliary knowledge. Rather, we shall present distributions, for cases that seem important to us, in a way that we hope will enable others to work out similar asymptotic distributions as may be required. Of the various sampling methods, probably the one most frequently found is multinomial sampling over the entire $A \times B$ double polytomy; hence we begin with, and devote most space to, that method.

Although this paper is organized around traditional ideas of hypothesis testing and estimation, the asymptotic distributions presented may also be useful in connection with other approaches to statistical inference, for example, the likelihood-ratio and the neo-Bayesian approaches.

3. MULTINOMIAL SAMPLING OVER THE WHOLE DOUBLE POLYTOMY

3.1. The Index λ_b

An index of association called λ_b, suggested in [8] as appropriate in some situations where asymmetry obtains and order is immaterial, is

$$\lambda_b = \frac{\sum_a \rho_{am} - \rho_{\cdot m}}{1 - \rho_{\cdot m}}. \tag{3.1.1}$$

This measure is the relative decrease in probability of erroneous guessing of B_b (when presented with random individuals) as between A_a unknown and A_a known.

We now discuss the maximum likelihood estimator of λ_b,

$$L_b = \frac{\sum_a R_{am} - R_{\cdot m}}{1 - R_{\cdot m}} = \frac{\sum_a N_{am} - N_{\cdot m}}{n - N_{\cdot m}}. \tag{3.1.2}$$

L_b is defined except when $R_{\cdot m} = 1$. We assume $\rho_{\cdot m} \neq 1$, and by the argument of Section A4 we may neglect, for asymptotic purposes, the possibility that $R_{\cdot m} = 1$. What to do should $R_{\cdot m} = 1$ in a finite sample will be discussed later in this Section.

It is shown in Section A5 that $\sqrt{n}(L_b - \lambda_b)$ is asymptotically normal with mean zero and variance

$$(1 - \sum \rho_{am})(\sum \rho_{am} + \rho_{\cdot m} - 2 \sum{}^r \rho_{am})/(1 - \rho_{\cdot m})^3, \tag{3.1.3}$$

where $\sum^r \rho_{am}$ denotes the sum of the ρ_{am}'s over those values of a such that ρ_{am} is taken on in that column in which $\rho_{\cdot m}$ is taken on. Since we have assumed that no ties exist among the contenders for $\text{Max}_b\, \rho_{ab}$ and $\text{Max}_b\, \rho_{\cdot b}$, the definition of $\sum^r \rho_{am}$ is unambiguous. The variance (3.1.3) is zero if and only if λ_b is zero or one.

To clarify the meaning of $\sum^r \rho_{am}$ we now give a simple example. Suppose the ρ_{ab} table is the 3×4 table

a \ b	1	2	3	4	
1	.14*	.05	.04	.04	.27
2	.04	.18*	.06	.04	.32
3	.04	.05	.24*	.08	.41
	.22	.28	.34	.26	1.00

where the marginal totals appear beyond the double lines. The ρ_{am}'s are indicated by asterisks in their cells; $\rho_{\cdot m}$ is .34, and $\sum^r \rho_{am}$ is .24.

It will also be convenient later to use the notation $\sum^c \rho_{mb}$ to mean the sum of the ρ_{mb}'s over those values of b such that ρ_{mb} is taken on in that row in which $\rho_{m\cdot}$ is taken on. In the above example the ρ_{mb}'s are .14, .18, .24, .08 respectively from left to right, $\rho_{m\cdot}$ is .41, and $\sum^c \rho_{mb} = .24 + .08 = .32$.

We shall also use the notations $\sum^r R_{am}$, $\sum^c R_{mb}$ and $\sum^r N_{am}$, $\sum^c N_{mb}$. They are defined just as above, but for the R_{ab}'s and N_{ab}'s respectively. When we work in terms of the R's or N's, ties may of course exist (although we neglect them for purposes of asymptotic theory) and a procedure for handling them will be suggested in the next Section.

It follows from (3.1.3), in a manner described in Section A5, that the following quantity is asymptotically unit-normal (normal with zero mean and unit variance):

$$\sqrt{n}(L_b - \lambda_b)\sqrt{\frac{(1 - R_{\cdot m})^3}{(1 - \sum R_{am})(\sum R_{am} + R_{\cdot m} - 2\sum^r R_{am})}}, \qquad (3.1.4)$$

under the following assumptions, some of which repeat earlier statements:

i) Multinomial sampling over the entire double polytomy;
ii) ρ_{am}'s and $\rho_{\cdot m}$ unique;
iii) $\rho_{\cdot m} \neq 1$ (i.e., λ_b is well defined); and
iv) $\lambda_b \neq 0$ or 1.

For computational convenience, we give another form of (3.1.4),

$$(L_b - \lambda_b)\sqrt{\frac{(n - N_{\cdot m})^3}{(n - \sum N_{am})(\sum N_{am} + N_{\cdot m} - 2\sum^r N_{am})}}. \qquad (3.1.4a)$$

We note that, to the present level of asymptotic approximation, if $\lambda_b = 0$, then $L_b = 0$. (Indeed, if $\lambda_b = 0$, the probability that $L_b = 0$ has the limit 1.) If $\lambda_b = 1$, then $L_b = 1$ without any asymptotic approximation, providing L_b is well defined.

3.2. Use of Asymptotic Unit-normality

For n large, $L_b - \lambda_b$ divided by an ASE is approximately unit-normal. The ASE, which we denote by $g(N_{ab}$'s) to emphasize its dependence on the sample, is the reciprocal of the square root factor in (3.1.4a). Then the probability that $(L_b-\lambda_b)/g(N_{ab}\text{'s})$ lies in an interval (c, d) is approximately $\Phi(d)-\Phi(c)$, where Φ is the unit-normal cumulative distribution function, ubiquitously tabled. For example, the probability that $(L_b-\lambda_b)/g(N_{ab}\text{'s})$ lies between -1.96 and 1.96 is approximately .95. Thus we may readily set up approximate confidence intervals for λ_b. Suppose that we seek an approximate confidence interval, symmetric about L_b, on the $1-\alpha$ level of confidence. Let $K_{\alpha/2}$ be the upper $100(\alpha/2)\%$ point for the unit-normal distribution (e.g., $K_{\alpha/2}=1.96$ for $\alpha=.05$). Then

$$\Pr\{-K_{\alpha/2} \leq (L_b - \lambda_b)/g(N_{ab}\text{'s}) \leq K_{\alpha/2}\} \cong 1 - \alpha, \tag{3.2.1}$$

or, equivalently,

$$\Pr\{L_b - K_{\alpha/2}g(N_{ab}\text{'s}) \leq \lambda_b \leq L_b + K_{\alpha/2}g(N_{ab}\text{'s})\} \cong 1 - \alpha, \tag{3.2.2}$$

so that $L_b \pm K_{\alpha/2}g(N_{ab}\text{'s})$ gives a confidence interval approximately on the $1-\alpha$ level of confidence. If the interval happens to go beyond 0 or 1, such inadmissible values would be excluded.

Similarly, we may test the null hypothesis that $\lambda_b=\lambda_b^{(0)}$ (when $\lambda_b^{(0)}\neq 0$),[6] on approximately the α level of significance, by rejecting the null hypothesis just when $\lambda_b^{(0)}$ lies outside of the interval $L_b \pm K_{\alpha/2}g(N_{ab}\text{'s})$.

To test the special hypothesis $\lambda_b=1$, one accepts when $L_b=1$ and otherwise rejects, for any level of significance. (See the final paragraph of Section 3.1.) Because of our level of asymptotic approximation,[6] one accepts the special hypothesis $\lambda_b=0$ when $L_b=0$ and otherwise rejects, for any significance level. Thus, approximate confidence intervals for λ_b, as described above, should exclude the points $\lambda_b=0$ and $\lambda_b=1$ *unless* $L_b=0$ or 1. In the later cases, the confidence interval consists of 0 or 1 (respectively) alone.

One-sided confidence intervals and tests may be readily obtained in the same manner.

It is also possible to obtain confidence intervals for the *difference* between the values of λ_b in two tables. Suppose that we have two independent multinomial samples, one from each of the tables: $\{N_{ab}^{(1)}\}$, $\{N_{ab}^{(2)}\}$. Let $L_b^{(i)}$ and $\lambda_b^{(i)}$ $(i=1, 2)$ be the estimated and true values of λ_b. If the assumptions after (3.14) are satisfied for both tables, then

$$\frac{(L_b^{(1)} - L_b^{(2)}) - (\lambda_b^{(1)} - \lambda_b^{(2)})}{\sqrt{[g(N_{ab}^{(1)}\text{'s})]^2 + [g(N_{ab}^{(2)}\text{'s})]^2}} \tag{3.2.3}$$

[6] The situation here is much like that of asymptotic distributions for the squared sample correlation coefficient or the squared sample multiple correlation coefficient [2, p. 415]. When the population parameter is zero, the asymptotic distribution appropriate for other cases degenerates and puts all its mass on zero. By changing the power of n as a scaling factor, one can often, in such cases, obtain a nondegenerate limit distribution.

In addition to this parallelism between asymptotic distribution theory for L_b and the squared sample correlation coefficient, there is an interpretive relationship between the two population quantities. If ρ is the population correlation coefficient, then $1-\rho^2$ is the ratio of "unexplained variability" in one variate, when the other is known, to "unexplained variability" when the other is not known. ("Unexplained variability" here refers to expected squared deviation around the best linear predictor and around the best constant predictor, respectively [18, p. 817].) Similarly, $1-\lambda_b$ is the ratio of "unexplained variability" in predicting the B classification, but here measured in terms of error probabilities for prediction from the A classification and from nothing, respectively [8, p. 741].

There are, of course, differences between ρ^2 and λ_b. For example, ρ itself—unsquared—may be positive or negative and its sign gives information about the sense of the association. For λ_b it is meaningless to speak of the sign or sense of association, since λ_b is invariant under permutations of rows (columns) among themselves. Another difference is that ρ is symmetric between the two variates while λ_b is not.

is asymptotically unit-normal. Thus

$$L_b^{(1)} - L_b^{(2)} \pm K_{\alpha/2}\sqrt{[g(N_{ab}^{(1)}\text{'s})]^2 + [g(N_{ab}^{(2)}\text{'s})]^2}$$

gives a confidence interval for $\lambda_b^{(1)} - \lambda_b^{(2)}$ approximately on the $1-\alpha$ level of confidence.

Similarly, we may test the null hypothesis $\lambda_b^{(1)} - \lambda_b^{(2)} = \Delta$, on approximately the α level of significance, by rejecting the null hypothesis just when the above confidence interval for $\lambda_b^{(1)} - \lambda_b^{(2)}$ fails to cover Δ. In particular, we may test the null hypothesis that the difference between $\lambda_b^{(1)}$ and $\lambda_b^{(1)}$ is zero (i.e., that $\lambda_b^{(1)} = \lambda_b^{(2)}$), on approximately the α level of significance, by rejecting this hypothesis when the above confidence interval fails to include zero. This test of the null hypothesis that $\lambda_b^{(1)} = \lambda_b^{(2)}$ can be generalized in order to obtain a test of the null hypothesis that the values of λ_b in k tables are all equal, i.e., that $\lambda_b^{(1)} = \lambda_b^{(2)} = \cdots = \lambda_b^{(k)}$, where $\lambda_b^{(i)}$ is the true value of λ_b for the ith table. First we note that, if the assumptions stated in Section 3.1 are satisfied for each of the k tables, and if the null hypothesis is in fact true, then the statistic

$$\sum_{i=1}^{k} (L_b^{(i)} - \bar{L})^2 / [g(N_{ab}^{(i)}\text{'s})]^2 \tag{3.2.4}$$

will have approximately ($n^{(i)} \to \infty$) the chi-square distribution[7] with $k-1$ degrees of freedom, where $L_b^{(i)}$ ($i = 1, 2, \cdots, k$) denotes the estimated value of λ_b in the ith table, $n^{(i)}$ is the sample size in the ith table, $g(N_{ab}^{(i)}\text{'s})$ denotes the estimate of the asymptotic standard deviation of $L_b^{(i)}$ (the $g(N_{ab}^{(i)}\text{'s})$ are maximum likelihood estimates), and where

$$\bar{L} = \left\{ \sum_{i=1}^{k} \left\{ L_b^{(i)} [g(N_{ab}^{(i)}\text{'s})]^{-2} \right\} \right\} \Big/ \left\{ \sum_{i=1}^{k} [g(N_{ab}^{(i)}\text{'s})]^{-2} \right\}.$$

We may test the null hypothesis that $\lambda_b^{(1)} = \lambda_b^{(2)} = \cdots = \lambda_b^{(k)}$, on approximately the α level of significance, by rejecting this hypothesis just when the statistic (3.2.4) is larger than the upper 100α per cent point of the chi-square distribution with $k-1$ degrees of freedom.

We digress to comment briefly on the possibility of using, not L_b, but some monotone transform of it for purposes of hypothesis testing, interval estimation, etc. One considers such transformations in the hope of bettering the asymptotic normal approximation, of simplifying the asymptotic variance, or of making the asymptotic variance more nearly constant. We record here expressions like (3.1.4a) for three transformations; the following three quantities are asymptotically unit-normal:

---

[7] The derivation of this asymptotic chi-square distribution (under the null hypothesis) may be broken into two parts. First, the asymptotic joint distribution of the k quantities like (3.1.4) is shown (via Section A2) to be the same as the corresponding joint distribution, but with true asymptotic variances replacing the sample estimators in (3.1.4). Second, in the asymptotic distribution of these modified quantities, a standard (weighted) average is considered, along with the corresponding standard, weighted sum of squares of residuals. Finally, (3.2.4) and its distribution are obtained by another application of Section A2 in a generalized form.

$$[\log(1 - L_b) - \log(1 - \lambda_b)]\sqrt{\frac{(n - N_{\cdot m})(n - \sum N_{am})}{\sum N_{am} + N_{\cdot m} - 2\sum^r N_{am}}}, \quad (3.2.5)$$

where the logarithms are to base e,

$$[\sqrt{1 - L_b} - \sqrt{1 - \lambda_b}]2(n - N_{\cdot m})/\sqrt{\sum N_{am} + N_{\cdot m} - 2\sum^r N_{am}}, \quad (3.2.6)$$

$$[\sqrt{L_b} - \sqrt{\lambda_b}][L_b/(1 - L_b)]^{1/2}2(n - N_{\cdot m})/\sqrt{\sum N_{am} + N_{\cdot m} - 2\sum^r N_{am}}. \quad (3.2.7)$$

The stated asymptotic normality follows directly from that of (3.1.4) by an application of Section A3. The methods of using (3.2.5)–(3.2.7) for inference about λ_b are straight-forward analogies to the methods for using (3.1.4).

All the further approximations in this paper may be used in the same ways as those described above. In every case the estimator of the measure of association minus its true value, all divided by a function of the N_{ab}'s, is approximately unit-normal for large n. Hence the above description of statistical procedures will not be repeated each time.

We now give a numerical example of the use of the suggested approximation based on (3.1.4). We drew a random sample of 50 from the population given by the table in Section 3.1.

Random sample of 50 from above population

Numbers are observed N_{ab}'s, and marginal $N_{a\cdot}$'s and $N_{\cdot b}$'s

	b = 1	2	3	4	
a = 1	8	5	3	3	19
2	0	8	1	0	9
3	0	4	14	4	22
	8	17	18	7	50

$$\sum N_{am} = 8 + 8 + 14 = 30, \quad N_{\cdot m} = 18, \quad \sum^r N_{am} = 14$$

$$L_b = \frac{30 - 18}{50 - 18} = \frac{12}{32} = .3750$$

$$g(N_{ab}\text{'s})^{-1} = \sqrt{\frac{(50 - 18)^3}{(50 - 30)(30 + 18 - 2 \times 14)}} = 9.0510$$

$$1.96/9.0510 = .2166$$

95% approximate confidence interval: $.1584 \leq \lambda_b \leq .5916$.

The confidence interval obtained is rather wide, but on the other hand the sample is not very large relative to the number of cells and their probabilities. The population value of λ_b, 1/3, is covered by the confidence interval for the chosen sample.

When samples are of moderate size, like the above, one often obtains ties among the maximum $N_{\cdot b}$'s and among the maximum N_{ab}'s for some a or a's. Although these ties disappear asymptotically by our assumption of no ties among the true maximum $\rho_{\cdot b}$'s or ρ_{ab}'s, ties must nonetheless be dealt with in real samples. Ties will affect our suggested procedure only via $\sum^r N_{am}$; they will not affect L_b itself, but only $g(N_{ab}$'s). For example, consider the following sample drawn from the same population as above. Here $\sum^r N_{am}$ might be 10 (if column 2's $N_{\cdot 2}=16$ is taken as $N_{\cdot m}$) or 14 (if column 3's $N_{\cdot 3}=16$ is taken as $N_{\cdot m}$). The sample and the two alternative computations are shown below.

Second random sample of 50 from above population

		b: 1	2	3	4	
a	1	9	2	1	1	13
	2	0	10	1	0	11
	3	2	4	14	6	26
		11	16	16	7	50

$$L_b = \frac{33-16}{50-16} = .5000$$

$$g(N_{ab}\text{'s})^{-1} = \text{either } \sqrt{\frac{34^3}{17\times(33+16-2\times 10)}} \text{ or } \sqrt{\frac{34^3}{17\times(33+16-2\times 14)}}$$

$$= \text{either } 8.9288 \text{ or } 10.4926$$

95% approximate confidence interval:

$$\text{either } .2805 \leq \lambda_b \leq .7195$$

$$\text{or } .3132 \leq \lambda_b \leq .6868.$$

When ties occur they may be resolved by the flip of a fair coin; this is the method used in the random sampling discussed in Section 3.8. Other methods are possible and perhaps better. For example, one might average the two or more possible values of $g(N_{ab}$'s), or one might take the largest possible value of $g(N_{ab}$'s). This topic requires further investigation.

Another problem that may arise in real samples, even when the assumptions for our asymptotic statements are true, is that $N_{\cdot m}$ may be n (i.e., $R_{\cdot m}=1$). This means that all the observations fall in one column. Unlike ties, this should happen very rarely in the usual sort of application we envisage. There seems to be no reasonable way of estimating λ_b in this case, and in fact L_b is not defined. Thus we suggest that, when $N_{\cdot m}=n$, the confidence interval be taken as the trivial one of all possible values of λ_b: $0\leq\lambda_b\leq 1$, and that any null hypothesis be accepted. To give a confidence interval from 0 to 1 inclusive is, of course, just a way of saying that nothing has been learned from the sample about λ_b.

3.3. The Index λ_a

If the roles of columns and rows be interchanged, we have the index λ_a and its estimator L_a. Everything is exactly the same as for λ_b and L_b, except for a systematic interchange of notation.

3.4. The Index λ

In [8], the symmetrical version of λ_a and λ_b was

$$\lambda = \frac{\sum_a \rho_{am} + \sum_b \rho_{mb} - \rho_{\cdot m} - \rho_{m\cdot}}{2 - (\rho_{\cdot m} + \rho_{m\cdot})} \cdot \tag{3.4.1}$$

The maximum likelihood estimator of λ is

$$\begin{aligned} L &= \frac{\sum_a R_{am} + \sum_b R_{mb} - R_{\cdot m} - R_{m\cdot}}{2 - (R_{\cdot m} + R_{m\cdot})} \\ &= \frac{\sum_a N_{am} + \sum_b N_{mb} - N_{\cdot m} - N_{m\cdot}}{2n - N_{\cdot m} - N_{m\cdot}} . \end{aligned} \tag{3.4.2}$$

We assume that λ is determinate, that is, that $\rho_{\cdot m}+\rho_{m\cdot}\neq 2$, or in other words that the entire population does not lie in one cell of the $A \times B$ double polytomy.

We show in Section A6 that (provided λ is determinate and $\neq 0$ or 1) the following quantity is asymptotically unit normal:

$$(L - \lambda) \frac{\sqrt{n}(2 - U_\cdot)^2}{\sqrt{\begin{aligned}&(2 - U_\cdot)(2 - U_\Sigma)(U_\cdot + U_\Sigma + 4 - 2U_*) \\ &\quad - 2(2 - U_\cdot)^2\left(1 - \textstyle\sum^* R_{am}\right) - 2(2 - U_\Sigma)^2(1 - R_{**})\end{aligned}}}, \tag{3.4.3}$$

where

$$\begin{aligned} U_\cdot &= R_{\cdot m} + R_{m\cdot}, \\ U_\Sigma &= \sum R_{am} + \sum R_{mb}, \\ U_* &= \textstyle\sum^r R_{am} + \sum^c R_{mb} + R_{*m} + R_{m*}, \end{aligned} \tag{3.4.4}$$

with the asterisked notation defined as follows:

$\sum^* R_{am}$ = sum of those R_{am}'s that also appear as R_{mb}'s, i.e., $\sum\sum R_{ab}$ over all (a, b) such that $R_{ab}=R_{am}=R_{mb}$,

R_{**} = that R_{ab} that appears both in the row for which $R_{a\cdot}$ is maximum and in the column for which $R_{\cdot b}$ is maximum,

R_{*m} = that R_{am} in the same row as that for which $R_{a\cdot}$ is maximum,

R_{m*} = that R_{mb} in the same column as that for which $R_{\cdot b}$ is maximum.

This notation is easier to use than to write down formally. An example of its use follows for the first sample of 50 described in Section 3.2.

First random sample in Section 3.2.

$$\sum N_{am} = 30, \qquad \sum N_{mb} = 34, \qquad N_{\cdot m} = 18, \qquad N_{m\cdot} = 22$$

$$L = \frac{30 + 34 - 18 - 22}{100 - 18 - 22} = .4000$$

$$U_{\cdot} = (18 + 22)/50 = .8000 \qquad U_{\Sigma} = (30 + 34)/50 = 1.2800$$

$$\sum{}^{r} R_{am} = 14/50 = .2800 \qquad \sum{}^{c} R_{mb} = (14 + 4)/50 = .3600$$

$$R_{*m} = 14/50 = .2800 \qquad R_{m*} = 14/50 = .2800$$

$$U_{*} = (14 + 18 + 14 + 14)/50 = 1.200$$

$$\sum{}^{*} R_{am} = (8 + 8 + 14)/50 = .600 \qquad R_{**} = 14/50 = .2800$$

$$g(N_{ab}\text{'s})^{-1} = \frac{\sqrt{50}(2 - .80)^2}{\sqrt{\begin{array}{l}(2 - .80)(2 - 1.28)(.80 + 1.28 + 4 - 2.4) \\ \qquad - 2(2 - .80)^2(1 - .60) - 2(2 - 1.28)^2(1 - .28)\end{array}}}$$

$$\cong 8.9964$$

95% approximate confidence interval for λ: $.1821 \leq \lambda \leq .6179$. Population value of $\lambda = .3600$.

As before, ties may occur that make the various quantities entering into the right fractional factor of (3.4.3) ambiguous, although such ambiguities will disappear asymptotically under our assumptions of unique maxima.

Again, as with L_b, when $\lambda = 0$ or 1, our asymptotic expressions degenerate. If $\lambda = 0$, the square root factor of (3.4.3) becomes 0; to our level of approximation, if $\lambda = 0$, L is 0. If $\lambda = 1$, the population is wholly concentrated in cells no two of which lie in the same row or column; hence the sample will be similarly concentrated, and L will always be 1 without any asymptotic approximation, providing that it is well defined. Thus, as before, if $L \neq 0$ or 1, and if a confidence interval, computed in the described manner, includes 0 or 1 or points beyond 0 or 1, such values should be removed from it. If $L = 0$ or 1, the confidence interval is just the single number 0 or 1 respectively.

If all the observations lie in a single cell, then L is indeterminate. This should be very infrequent for applications of the kind we have in mind. We suggest that, when this occurs, the confidence interval should be the entire interval [0, 1], and that any null hypothesis be accepted.

3.5. The Index γ

In [8], an index of association called γ was suggested as appropriate in some situations where both classifications have intrinsic and relevant order. The definition of γ was

$$\gamma = \frac{\Pi_s - \Pi_d}{1 - \Pi_t} = \frac{2\Pi_s + \Pi_t - 1}{1 - \Pi_t}, \tag{3.5.1}$$

where

$$\Pi_s = 2\sum_a \sum_b \rho_{ab}\Big\{\sum_{a'>b}\ \sum_{b'>b} \rho_{a'b'}\Big\}$$

$$\Pi_d = 2\sum_a \sum_b \rho_{ab}\Big\{\sum_{a'>a}\ \sum_{b'<b} \rho_{a'b'}\Big\} \tag{3.5.2}$$

$$\Pi_t = 1 - \Pi_s - \Pi_d = \sum_a \sum_b \rho_{ab}\big\{\rho_{a\cdot} + \rho_{\cdot b} - \rho_{ab}\big\}$$

$$= \sum_a \rho_{a\cdot}^2 + \sum_b \rho_{\cdot b}^2 - \sum_a \sum_b \rho_{ab}^2.$$

Here Π_s ("s" for "same") is the probability that two randomly chosen individuals will have the same order in both classifications (concordance), Π_d ("d" for "different") is the probability that they will have different orders (discordance) and Π_t ("t" for "tie") is the probability that one or both classifications will be the same so that order is not clearly defined. We propose the estimation of γ by its maximum likelihood estimator,

$$G = \frac{P_s - P_d}{1 - P_t} = \frac{2P_s - 1 + P_t}{1 - P_t}, \tag{3.5.3}$$

where the P's are the sample analogs of the Π's, as follows:

$$P_s = 2\sum_a \sum_b R_{ab}\Big\{\sum_{a'>a}\ \sum_{b'>b} R_{a'b'}\Big\}$$

$$= \frac{2}{n^2}\sum_a \sum_b N_{ab}\Big\{\sum_{a'>a}\ \sum_{b'>b} N_{a'b'}\Big\},$$

$$P_d = 2\sum_a \sum_b R_{ab}\Big\{\sum_{a'>a}\ \sum_{b'<b} R_{a'b'}\Big\}$$

$$= \frac{2}{n^2}\sum_a \sum_b N_{ab}\Big\{\sum_{a'>a}\ \sum_{b<'b} N_{a'b'}\Big\}, \tag{3.5.4}$$

$$P_t = 1 - P_s - P_d = \sum_a \sum_b R_{ab}\big\{R_{a\cdot} + R_{\cdot b} - R_{ab}\big\}$$

$$= \frac{1}{n^2}\sum_a \sum_b N_{ab}\big\{N_{a\cdot} + N_{\cdot b} - N_{ab}\big\}$$

$$= \sum_a R_{a\cdot}^2 + \sum_b R_{\cdot b}^2 - \sum_a \sum_b R_{ab}^2$$

$$= \frac{1}{n^2}\Big[\sum_a N_{a\cdot}^2 + \sum_b N_{\cdot b}^2 - \sum_a \sum_b N_{ab}^2\Big].$$

As before, we suppose multinomial sampling over the entire tableau, and we assume $\Pi_t \neq 1$.

It is shown in Section A7 that $\sqrt{n}(G-\gamma)$ is asymptotically normal with zero mean and variance[8]

[8] This variance for the 2×2 case, in which γ is the same as Yule's Q, was obtained by Yule in 1900 [27, p. 285]. For a recent discussion of the 2×2 case, see [22] and [7].

$$\frac{16}{(1-\Pi_t)^4}\left\{\Pi_s^2\Pi_{dd} - 2\Pi_s\Pi_d\Pi_{sd} + \Pi_d^2\Pi_{ss}\right\}, \tag{3.5.5}$$

where

$$\begin{aligned}
\Pi_{ss} &= \sum_a \sum_b \rho_{ab}\Big\{\sum_{a'>a}\sum_{b'>b}\rho_{a'b'} + \sum_{a'<a}\sum_{b'<b}\rho_{a'b'}\Big\}^2 \\
\Pi_{sd} &= \sum_a \sum_b \rho_{ab}\Big\{\sum_{a'>a}\sum_{b'>b}\rho_{a'b'} + \sum_{a'<a}\sum_{b'<b}\rho_{a'b'}\Big\} \\
&\qquad\qquad \cdot\Big\{\sum_{a'>a}\sum_{b'<b}\rho_{a'b'} + \sum_{a'<a}\sum_{b'>b}\rho_{a'b'}\Big\}, \\
\Pi_{sd} &= \sum_a \sum_b \rho_{ab}\Big\{\sum_{a'>a}\sum_{b'<b}\rho_{a'b'} + \sum_{a'<a}\sum_{b'>b}\rho_{a'b'}\Big\}^2.
\end{aligned} \tag{3.5.6}$$

These doubly subscripted Π's are readily interpreted as follows. Suppose we draw three individuals at random and independently from the population in question. Then

- Π_{ss} is the probability that the second and third individuals both have "positive" sign relationships (i.e., are both concordant) with the first.
- Π_{dd} is the probability that the second and third individuals both have "negative" sign relationships (i.e., are both discordant) with the first.
- Π_{sd} is the probability that the second individual has "positive" sign relationships with the first (i.e., the first and second are concordant) but that the third has "negative" sign relationships with the first (i.e., the first and third are discordant).[9]

Hence, if we define P_{ss}, P_{sd}, and P_{dd} just as the corresponding Π's but with R_{ab} replacing ρ_{ab}, and if we assume that (3.5.5) is $\neq 0$, it follows that

$$\sqrt{n}(G-\gamma) \Big/ \sqrt{\frac{16}{(1-P_t)^4}\left\{P_s^2P_{dd} - 2P_sP_dP_{sd} + P_d^2P_{ss}\right\}} \tag{3.5.7}$$

is asymptotically unit-normal. (Note that we do not require any uniqueness of maxima assumptions here. On the other hand, the assumption that $(3.5.5) \neq 0$ does not seem to have any very simple interpretation. Comments on the meaning of this assumption are given in Section A7.)

If $G=1$, then the denominator of (3.5.7) (i.e., the estimator of $\sqrt{(3.5.5)}$) will be equal to zero. For this particular situation, a possible statistical procedure, if n is large, is to give as the confidence interval the degenerate interval $\gamma=1$. If $\gamma<1$, the probability that $G=1$ vanishes as $n\rightarrow\infty$, while if $\gamma=1$, G will always be 1, providing that it is well defined. On the other hand, if γ is near 1, G may frequently equal 1 unless n is very large, so that the magnitude of n for our asymptotic theory to work depends critically on γ when γ is near 1. Similar comments can also be made when $G=-1$. In the particular situation where G is undefined, which will happen very rarely in the sort of application we envisage, the point of view presented at the end of Section 3.2 (for the case where L_b is undefined) can be applied.

[9] Note that Π_{sd} has been defined unsymmetrically with respect to the second and third individuals; this is the reason for the 2 in the middle term of the last factor of (3.5.5).

If in computing P_{ss}, etc., we use N_{ab} instead of R_{ab}, we emerge with $n^3\ P_{ss}$, etc. Hence, denoting by P_{ss}, P_s, etc., the quantities corresponding to P_{ss}, P_s, etc., but computed in terms of N_{ab}'s, we find that

$$P_{ss} = \mathsf{P}_{ss}/n^3 \qquad P_s = \mathsf{P}_s/n^2, \text{ etc.};$$

that

$$G = \frac{\mathsf{P}_s - \mathsf{P}_d}{n^2 - \mathsf{P}_t} = \frac{\mathsf{P}_s - \mathsf{P}_d}{\mathsf{P}_s + \mathsf{P}_d};$$

and that (3.5.7) is the same as

$$(G - \gamma)\,\frac{(n^2 - \mathsf{P}_t)^2}{4\sqrt{\mathsf{P}_s^2\mathsf{P}_{dd} - 2\mathsf{P}_s\mathsf{P}_d\mathsf{P}_{sd} + \mathsf{P}_d^2\mathsf{P}_{ss}}}\,. \tag{3.5.8}$$

This is perhaps the most convenient form for computation. As an example, let us treat the first random sample used in Section 3.2, but now with each population polytomy thought of as ordered. A convenient way to organize the computations is the following.

First set down the N_{ab} table. Then compute what we might call the S (for "same") table. This $\alpha\times\beta$ table contains in its (a, b) cell the sum of all $N_{a'b'}$ such that $a'>a$ and $b'>b$, plus the sum of all $N_{a'b'}$ such that $a'<a$ and $b'<b$. Then compute the D (for 'different') table. This $\alpha\times\beta$ table contains in its (a, b) cell the sum of all $N_{a'b'}$ such that $a'>a$ and $b'<b$ plus the sum of all $N_{a'b'}$ such that $a'<a$ and $b'>b$. Note that, if we were only computing G itself, we could use simpler S and D tables, entailing but one pair of inequalities.

Thus we have for our example

N_{ab} Table

8	5	3	3
0	8	1	0
0	4	14	4

S Table

31	19	4	0
22	26	17	16
0	8	21	25

D Table

0	0	12	27
11	6	7	18
20	7	3	0

where, e.g., the 31 in the upper left corner of the S table is found by adding $8+1+0+4+14+4=31$, and the 7 in the second row and third column of the D table is found by adding $0+4+3=7$.

P_s is found by multiplying each entry of the N_{ab} table by the corresponding entry of the S table and adding the products.

P_d is found by multiplying each entry of the N_{ab} table by the corresponding entry of the D table and adding the products.

P_{ss} is found by multiplying each entry of the N_{ab} table by the square of the corresponding entry of the S table and adding the products.

P_{dd} is found by multiplying each entry of the N_{ab} table by the square of the corresponding entry of the D table and adding the products.

P_{sd} is found by taking the sum of the triple products of corresponding terms from the N_{ab}, S, and D tables.

With the possible exception of P_{sd}, all these numbers may be found very rapidly with the aid of a table of squares and a desk computer. For our example we have

$$\mathsf{P}_s = 1{,}006, \qquad \mathsf{P}_d = 242$$

$$\mathsf{P}_{ss} = 24{,}168, \qquad \mathsf{P}_{sd} = 2{,}617, \qquad \mathsf{P}_{dd} = 3{,}278.$$

A simple numerical check starts by computation of a T (for 'tie') table. Here in the (a, b) cell one puts $N_{a\cdot}+N_{\cdot b}-N_{ab}$. In our example we have

T Table

19	31	34	23
17	18	26	16
30	35	26	25

(The T table itself may be checked by observing that the sum of its entries must be $n(\alpha+\beta-1)$. In the above case the sum should be 300, as it is.) To use this check compute the following quantities:

P_t = sum of products of N_{ab}'s by corresponding entries in T table,
P_{st} = sum of triple products of N_{ab}'s by corresponding entries in S and T tables,
P_{dt} = sum of triple products of N_{ab}'s by corresponding entries in D and T tables,
P_{tt} = sum of products of N_{ab}'s by *squares* of corresponding entries in T table.

In our case we have

$$\mathsf{P}_t = 1{,}252, \qquad \mathsf{P}_{st} = 23{,}515, \qquad \mathsf{P}_{dt} = 6{,}205, \qquad \mathsf{P}_{tt} = 32{,}880.$$

The following relationships then hold and serve as a partial check:

$$\mathsf{P}_s + \mathsf{P}_d + \mathsf{P}_t = n^2,$$

$$\mathsf{P}_{ss} + 2\mathsf{P}_{sd} + \mathsf{P}_{dd} + 2\mathsf{P}_{st} + 2\mathsf{P}_{dt} + \mathsf{P}_{tt} = n^3.$$

Some other relations that may be used for more detailed checking are

$$n\mathsf{P}_s = \mathsf{P}_{ss} + \mathsf{P}_{sd} + \mathsf{P}_{st},$$

$$n\mathsf{P}_d = \mathsf{P}_{sd} + \mathsf{P}_{dd} + \mathsf{P}_{dt},$$

$$n\mathsf{P}_t = \mathsf{P}_{st} + \mathsf{P}_{dt} + \mathsf{P}_{tt}.$$

These all hold for the computations of our example.

In systematizing the above computations it may be convenient to write separately the cross-product tables of the S and D tables, the S and T tables, and the D and T tables. In our example, these are

$S\times D$

0	0	48	0
242	156	119	288
0	56	63	0

$S\times T$

589	589	136	0
374	468	442	256
0	280	546	625

$D\times T$

0	0	408	621
187	108	182	288
600	245	78	0

From the above numbers we compute that

$$G = \frac{1{,}006 - 242}{1{,}006 + 242} = .6122$$

(to four places), and that, via (3.5.8),

$$(.6122 - \gamma)\frac{1{,}557{,}504}{4\sqrt{3{,}458{,}600{,}992}} = (.6122 - \gamma)6.6209$$

may be considered as an observation from an approximately unit-normal population. Hence we may readily establish approximate confidence limits for γ, say on the 95% level of confidence. These limits, for our sample, say that γ lies between the numbers

$$.6122 \pm \frac{1.96}{6.621},$$

or that γ (to 3 places) lies between .316 and .908. In particular, this means that G differs from zero with statistical significance on the 5% level of significance. (The true value of γ for the cross classification from which the sample was drawn is .4889 to four places.)

The computations described and exemplified in the preceding pages are rather tedious because of the many arithmetical operations. For a rapid significance test or a crude confidence interval, the computations may be much curtailed (at the expense of power).

This curtailment is possible because of the existence of a simple upper bound for the asymptotic variance of $\sqrt{n}\,(G-\gamma)$. We show in Section A7 that the asymptotic variance of $\sqrt{n}\,(G-\gamma)$, whose exact value is given by (3.5.5), is always less than or equal to

$$2(1 - \gamma^2)/(1 - \Pi_t). \tag{3.5.9}$$

This upper bound is closely related to a bound obtained by Daniels and Kendall for a somewhat different problem (see [3], [4]), and to a bound presented by A. Stuart [23] for a measure of association that is similar to our γ [8, pp. 750–1]. The method of obtaining (3.5.9), which is developed in Section A7, is somewhat different from, and perhaps simpler than, the methods used by earlier writers.

As an example of this upper bound, consider the 3×4 population of Section 3.2, from which was drawn the sample we have just discussed. For that population, (3.5.5) turns out to be 1.259 and (3.5.9) to be 2.920, so that the upper bound for asymptotic variance is about 2.3 times the actual asymptotic variance. For most uses, however, the relevant ratio is the square root of 2.3, about 1.5.

It follows from (3.5.9) that *conservative* asymptotic tests and confidence intervals may be obtained by considering as a unit-normal quantity

$$\sqrt{n}(G - \gamma)\sqrt{\frac{1 - P_t}{2(1 - G^2)}}, \tag{3.5.10}$$

or, in another notation,

$$(G-\gamma)\sqrt{\frac{n^2-\mathsf{P}_t}{2n(1-G^2)}}\,. \tag{3.5.11}$$

By a conservative test we mean a test whose probability of falsely rejecting the null hypothesis is known only to be $\leq$ the nominal level of significance. By a conservative confidence interval we mean a confidence interval whose probability of covering the true value being estimated is known only to be $\geq$ the nominal confidence level. This traditional notion of conservatism is, of course, asymmetrical. If we are conservative about significance level, we may lose power.

As an example, let us return to the numerical work a few paragraphs back that led to the 95% confidence interval (.316, .908) from a particular sample. For that sample, using the present cruder approximation,

$$(.6122-\gamma)\sqrt{\frac{2{,}500-1{,}252}{100(1-.6122^2)}}=(.6122-\gamma)\times 4.468$$

may be considered as an observation from (approximately) a normal distribution with zero mean and variance less than or equal to unity. This results in the following conservative asymptotic confidence interval at the 95% level of confidence: (.174, 1.051). We would of course change the right-hand end point to obtain (.174, 1.000). (Note that, by a familiar argument, 1.000 itself is excluded from the interval.)

Because of the simple form of (3.5.9), one may consider a variant procedure, much like that of the familiar quadratic confidence procedure for binomial proportions. The probability is (asymptotically) $\geq 1-\alpha$ that

$$-K_{\alpha/2}\leq\sqrt{n}(G-\gamma)\sqrt{\frac{(1-P_t)}{2(1-\gamma^2)}}\leq K_{\alpha/2}, \tag{3.5.12}$$

where the quantity in the middle is like (3.5.10) except that G^2 in (3.5.10) is replaced by γ^2 here. The statement (3.5.12) is equivalent to

$$(G-\gamma)^2\left[\frac{n^2-\mathsf{P}_t}{2n(1-\gamma^2)}\right]\leq K^2_{\alpha/2}. \tag{3.5.13}$$

Simplifying, we obtain the quadratic inequality

$$\gamma^2[n^2-\mathsf{P}_t+2nK^2_{\alpha/2}]-\gamma[2G(n^2-\mathsf{P}_t)]+G^2(n^2-\mathsf{P}_t)-2nK^2_{\alpha/2}\leq 0. \tag{3.5.14}$$

It is readily shown that, for large n, the probability is nearly one that the values of γ satisfying (3.5.14) form an interval. It is a conservative approximate confidence interval for γ.

In our numerical example, (3.5.14) becomes

$$1{,}632\gamma^2-1{,}528\gamma+83.6\leq 0,$$

and the two real roots of the above quadratic expression are (.058, .878).

The numerical results from our particular sample may be recapitulated as follows:

95% asymptotic confidence interval based on estimate of variance via ((3.5.8): (.316, .908)

95% conservative asymptotic confidence interval based on estimated variance bound via (3.5.11): (.174, 1.000)

95% conservative asymptotic confidence interval based on estimated variance bound via (3.5.14): (.058, .878)

As would be expected, the confidence interval obtained via (3.5.8) is appreciably narrower than the other two. This reflects the fact that the upper bound for asymptotic variance (3.5.9) may be considerably larger than the true asymptotic variance. The two conservative methods give intervals of about the same length but in different positions. If we compare the uncurtailed interval based on (3.5.11), that is (.174, 1.051), with the interval based on (3.5.14), we see that the former is longer than the latter. This might have been expected since G in the denominator of (3.5.11) is subject to sampling variability, while in (3.5.14) the true value γ appears instead of G.

On the whole, we recommend the use of (3.5.8) because of its more precise results. It is true that it requires a tedious (although not difficult) computation, but in most serious studies this amount of computation would be a negligible cost unless it had to be repeated many times. It may well be that better bounds than (3.5.9) will be found that permit simplified computations without appreciable loss of precision. Daniels [4] has shown that the upper bound first given by Daniels and Kendall [3], which is related to (3.5.9) though appropriate for a somewhat different problem, is in general a poor one, although it is sometimes attainable. (See also [14].) Nevertheless, when n is large, even this bound can be good enough for some practical purposes [23], which suggests that the bound (3.5.9) presented here can also be good enough for such practical purposes.

We mentioned earlier in this section that the upper bound (3.5.9) for the variance of G was closely related to a bound presented by Stuart [23] for an estimate of a measure of association that he has suggested. It was also noted in [8, pp. 750–1] that Stuart's measure of association was closely related to γ. (The denominator $1-\Pi_t$ in (3.5.1) does not appear in Stuart's measure, and in its place we find a quantity that depends on the minimum number m of rows and columns in the cross classification table [8].)[10] It therefore seemed worthwhile to include here some numerical comparison of the two measures and the

[10] Stuart's denominator is introduced in order that his measure of association, τ_c, may attain, or nearly attain, the absolute value 1 when the entire cross classification population lies in a longest diagonal. The absolute value 1 is attained, following Stuart, just when the following three conditions are met: (1) population size is a multiple of m, (2) the population lies entirely in cells along a longest diagonal of the cross classification table, and (3) the frequencies in these diagonal cells are equal. This characteristic of τ_c is rather different from the corresponding characteristic of γ, which has absolute value 1 when (but not only when) the population is concentrated on *any* diagonal of the cross classification; in particular, the cells of the diagonal need not have equal or nearly equal frequencies. This difference between τ_c and γ may be relevant in deciding which measures to use in applications.

Stuart defines τ_c for a finite population and considers sampling without replacement, while we consider sampling with replacement. Sampling theory for either τ_c or γ could, of course, be considered for sampling both with and without replacement.

bounds for their corresponding variances. For this purpose, the data presented by Stuart [23, pp. 8–10] has been re-examined. Stuart's data (Table 1 in [23]) refers to 4×4 cross classifications between left and right eye vision for a group of employees in Royal Ordnance factories; he gives two cross classifications, one for men and one for women. The following table summarizes information about t_c, G, and estimated bounds on standard error for Stuart's data:

	t_c	G	Estimated Bound on Standard Error	
			for t_c	for G
Men	+0.629	+0.776	0.029	0.022
Women	+0.633	+0.798	0.019	0.014
			For differences	
Difference	0.004	0.022	0.035	0.027

We note that the estimates G are larger than the corresponding estimates t_c and the estimated bounds on standard errors are smaller. These differences no doubt reflect in part the fact that t_c and G estimate somewhat different population measures of association.

In closing this section, we mention that, in testing the null hypothesis $\gamma = 0$, some modification and simplification of the asymptotic variance formula (3.5.8) is possible, since $\Pi_s = \Pi_d$ when the null hypothesis is, in fact, true. In this situation, the denominator of (3.5.8) might be replaced by the following, which is asymptotically equivalent to that denominator under the null hypothesis:

$$2(\mathsf{P}_s + \mathsf{P}_d)\sqrt{\mathsf{P}_{dd} - 2\mathsf{P}_{sd} + \mathsf{P}_{ss}}. \tag{3.5.15}$$

The statistic (3.5.15) is simpler to compute than the denominator of (3.5.8), and its use as a replacement for the denominator of (3.5.8) will not greatly affect, when n is large, the level of significance of the correspondingly modified test, although it will affect, to an unknown extent, the power of this test. Further study of the use of (3.5.15), when the null hypothesis is $\gamma = 0$, would be worthwhile. (A related discussion in the binomial context is given in [20].)

If we wished to test the stricter null hypothesis of independence between the two polytomies (independence implies that $\gamma = 0$, but $\gamma = 0$ does not imply independence), further modification and simplification is possible, in the sense that (3.5.15) could be replaced by a statistic which would be a function of only the marginal N's and which would be asymptotically equivalent to (3.5.15) under the null hypothesis.

3.6. Measures of Reliability

In [8], a number of measures of reliability or agreement were suggested as possibly useful when A_1 is the same class as B_1, A_2 as B_2, and so on, but where assignment to class is by two different methods. In these situations $\alpha = \beta$.

For some of the proposed measures of reliability, namely

$$\sum_{a=1}^{\alpha} \rho_{aa} \quad \text{and} \quad \sum_{|a-b|\leq I} \rho_{ab},$$

where I is a small integer, there is relatively little difficulty in working with the sampling distributions of the sample analogs

$$\sum_{a=1}^{\alpha} R_{aa} \quad \text{and} \quad \sum_{|a-b|\leq I} R_{ab}. \tag{3.6.1}$$

Each of these will, under over-all multinomial sampling, have a binomial distribution with binomial probability equal to the population value of the measure and sample size equal to n. Thus familiar procedures for estimating and testing binomial probabilities may be used.

A slightly more complex measure, of possible interest in the unordered case, was also suggested in [8]. It is

$$\lambda_r = \frac{\sum \rho_{aa} - \frac{1}{2}(\rho_{M\cdot} + \rho_{\cdot M})}{1 - \frac{1}{2}(\rho_{M\cdot} + \rho_{\cdot M})}, \tag{3.6.2}$$

where

$$\rho_{M\cdot} + \rho_{\cdot M} = \operatorname*{Max}_a\ (\rho_{a\cdot} + \rho_{\cdot a}).$$

The sample analogue of λ_r is

$$L_r = \frac{\sum R_{aa} - \frac{1}{2}(R_{M\cdot} + R_{\cdot M})}{1 - \frac{1}{2}(R_{M\cdot} + R_{\cdot M})}, \tag{3.6.3}$$

where

$$R_{M\cdot} + R_{\cdot M} = \operatorname*{Max}_a\ (R_{a\cdot} + R_{\cdot a}).$$

Assume now (i) that λ_r is well-defined, (ii) that there is a unique modal class (i.e., that $\rho_{a\cdot}+\rho_{\cdot a}=\rho_{M\cdot}+\rho_{\cdot M}$ for only one a), and (iii) that $\lambda_r \neq \pm 1$. Then the methods of the preceding sections and the Appendix may be applied to show that

$$\begin{aligned}
&(L_r - \lambda_r)(1 - \tfrac{1}{2}S_r)^2 \left[\frac{1 - D_r}{n} \left\{ D_r + \tfrac{1}{4}S_r(1 - D_r - S_r) - R_{MM}(\tfrac{3}{2} + \tfrac{1}{2}D_r - S_r) \right\} \right]^{-1/2} \\
&= (L_r - \lambda_r)(1 - \tfrac{1}{2}S_r)^2 \left[\frac{1 - D_r}{n} \left\{ (1 - \tfrac{1}{2}S_r)(\tfrac{1}{2}S_r + D_r - 2R_{MM}) \right.\right. \\
&\qquad\qquad \left.\left. - \tfrac{1}{4}(1 - D_r)(S_r - 2R_{MM}) \right\} \right]^{-1/2}
\end{aligned} \tag{3.6.4}$$

is asymptotically unit-normal. The new notation is defined as follows:

$$\begin{aligned}
D_r &= \sum R_{aa}, \qquad\qquad S_r = R_{M\cdot} + R_{\cdot M}, \\
R_{MM} &= \text{that } R_{aa} \text{ such that } R_{a\cdot} + R_{\cdot a} = R_{M\cdot} + R_{\cdot M}.
\end{aligned} \tag{3.6.5}$$

For large n, R_{MM} is uniquely defined with high probability. If a tie occurs we suggest a random choice.

3.7. Partial and Multiple Association

As final examples of asymptotic sampling theory in the case of fully multinomial sampling, we consider one of the coefficients of partial association suggested in Section 11 of [8] for a three-way classification, together with the coefficient of multiple association suggested in Section 12 of [8].

Suppose that there are three polytomies: $A_1, \cdots, A_\alpha$; $B_1, \cdots, B_\beta$; and $C_1, \cdots, C_\gamma$. If an individual is chosen at random from the fixed triply polytomous population of interest, the probability that he will simultaneously fall in the categories A_a, B_b, and C_c is ρ_{abc}. A measure of partial association between the A and B polytomies, averaged over the C polytomy, is

$$\lambda_b'(A,B\,|\,C) = \frac{\sum_a \sum_c \rho_{amc} - \sum_c \rho_{\cdot mc}}{1 - \sum_c \rho_{\cdot mc}}, \tag{3.7.1}$$

where dots mean summation over the dotted subscript, where $\rho_{amc} = \text{Max}\ \rho_{abc}$, and where $\rho_{\cdot mc} = \text{Max}\ \rho_{\cdot bc}$. This measure is the relative decrease in probability of error of guessing the B category of an individual if we know both his A and C categories as against knowing only his C category. Thus it refers to optimal prediction with the C category always known (hence "partial" association). It is asymmetric in that only prediction of B categories are considered, and it is unchanged by independent permutations of the classes within each polytomy (hence appropriate in some situations where there is no natural ordering of these classes).

For convenience we may simply write λ_b' instead of λ_b' $(A, B\,|\,C)$ in this Section, but in applications the arguments should probably be retained, for the six possible asymetrical λ''s obtained from an $\alpha\times\beta\times\gamma$ cross classification will in general all have both different numerical values and different interpretations.

As before, we assume that all relevant maxima are unique, i.e., that

For each a and c, $\rho_{amc} = \rho_{abc}$ for a unique value of b.

For each c, $\rho_{\cdot mc} = \rho_{\cdot bc}$ for a unique value of b.

We also assume that λ_b' is well defined, i.e., that $\sum_c \rho_{\cdot mc} \neq 1$.

Suppose that a sample of n is drawn from our population of interest and that each member of the sample is assigned to one of the $\alpha\beta\gamma$ cells (A_a, B_b, C_c) according to observation (without error) of its categories in the three polytomies. Suppose further that sampling is with replacement, or, alternatively, that the population of interest is infinite or very large. Let N_{abc} be the number of individuals among the n in the sample that fall into the (A_a, B_b, C_c) cell. Then $\sum_a \sum_b \sum_c N_{abc} = n$ and the N_{abc}'s have jointly a multinomial distribution with cell probabilities ρ_{abc}. Denote by R_{abc} the quantity N_{abc}/n.

Just as in Section 2, the $\alpha\beta\gamma$ random variables $\sqrt{n}$ $(R_{abc} - \rho_{abc})$ are jointly asymptotically normal, and we may assume, for asymptotic purposes, that R_{amc} and $R_{\cdot mc}$ are taken on at the values of b corresponding to ρ_{amc} and $\rho_{\cdot mc}$.

In this context, we now discuss the approximate distribution of the maximum likelihood estimator of λ_b',

$$L_b' = \frac{\sum_a \sum_c R_{amc} - \sum_c R_{\cdot mc}}{1 - \sum_c R_{\cdot mc}} \cdot \tag{3.7.2}$$

Of course, even if λ_b' is defined, L_b' may not be, but, as in Sections 3.1 and 3.2 we may neglect this possibility for asymptotic purposes. As in prior sections, one may show that $\sqrt{n}\ (L_b' - \lambda_b')$ is asymptotically normal with zero mean and variance

$$(1 - \sum_c \sum_a \rho_{amc})(\sum_c \sum_a \rho_{amc} + \sum_c \rho_{\cdot mc} - 2 \sum_c \sum_a{}^r \rho_{amc})/(1 - \sum_c \rho_{\cdot mc})^3, \tag{3.7.3}$$

where

$$\sum_a{}^r \rho_{amc}$$

denotes the sum of the ρ_{amc}'s (for fixed c) over these values of a such that ρ_{amc} is taken on for that value of b for which $\rho_{\cdot mc}$ is taken on. Note the strong similarity between (3.1.3) and (3.7.3); the latter is just like the former except that all terms (but unity) have an additional summation over c. We note that (3.7.3) is zero if and only if λ_b' is zero or one.

Hence, the following quantity is asymptotically unit-normal:

$$\sqrt{n}[L_b'(A, B \mid C) - \lambda_b'(A, B \mid C)] \cdot \sqrt{\frac{(1 - \sum_c R_{\cdot mc})^3}{(1 - \sum_c \sum_a R_{amc})(\sum_c \sum_a R_{amc} + \sum_c R_{\cdot mc} - 2 \sum_c \sum_a{}^r R_{amc})}} \tag{3.7.4}$$

provided that our earlier assumptions hold and that $\lambda_b' \neq 0$ or 1. If $\lambda_b' = 0$, then, to the present level of asymptotic approximation, $L_b' = 0$. If $\lambda_b' = 1$, then $L_b' = 1$ without any asymptotic approximation, providing L_b' is defined.

We note that the following quantities, are also asymptotically unit-normal:

$$\sqrt{n}[\log(1 - L_b') - \log(1 - \lambda_b')]\sqrt{(1 - \sum_c R_{\cdot mc})(1 - \sum_c \sum_a R_{amc})/D}, \tag{3.7.5}$$

$$\sqrt{n}[\sqrt{1 - L_b'} - \sqrt{1 - \lambda_b'}]2(1 - \sum_c R_{\cdot mc})/\sqrt{D}, \tag{3.7.6}$$

$$\sqrt{n}[\sqrt{L_b'} - \sqrt{\lambda_b'}][L_b'/(1 - L_b')]^{1/2} 2(1 - \sum_c R_{\cdot mc})/\sqrt{D}, \tag{3.7.7}$$

where

$$D = \sum_c \sum_a R_{amc} + \sum_c R_{\cdot mc} - 2 \sum_c \sum_a{}^r R_{amc},$$

and logarithms are to base e.

We conclude this Section with some comments about the following measure of multiple association between B and (A, C) together [8, Sec. 12]:

$$\lambda_b''(B; A, C) = \frac{\sum_a \sum_c \rho_{amc} - \rho_{\cdot m \cdot}}{1 - \rho_{\cdot m \cdot}},$$

where $\rho_{\cdot m \cdot} = \text{Max } \rho_{\cdot b \cdot}$. This quantity is exactly λ_b itself, computed from the $(\alpha\gamma) \times \beta$ cross classification in which the $\alpha\gamma$ cells of the A, C cross classification are thought of as making a single classification. (Note: On p. 762 of [8], the rearranged tableau is shown for A against (B, C) rather than for B against (A, C).) The interpretation of λ_b'' $(B; A, C)$ is the usual one for λ_b, but now comparing errors in predicting B, knowing A *and* C against knowing nothing.

It is interesting to note on analogy (brought to our attention by J. Mincer) with classical correlation analysis. In the present context, the marginal λ_b, relative to prediction of B from C, is

$$\lambda_b(B; C) = \frac{\sum_c \rho_{\cdot mc} - \rho_{\cdot m \cdot}}{1 - \rho_{\cdot m \cdot}},$$

so we see that

$$1 - \lambda_b''(B; A, C) = [1 - \lambda_b(B; C)][1 - \lambda_b'(B, A \mid C)].$$

This is completely analogous to the classical relationship between multiple and partial correlation coefficients,

$$1 - R_{b \cdot ac}^2 = [1 - \rho_{bc}^2][1 - \rho_{ba \cdot c}^2];$$

see, for example, [2], p. 307. It is interesting to note that, although $R_{b \cdot ac}$ may be expressed as a function of ρ_{ba}, ρ_{bc}, and ρ_{ac}, the analogous relationship does not hold for λ_b' $(B, A \mid C)$.[11]

Finally, we note that the asymptotic distribution of the sample analog of λ_b'' needs no fresh discussion here. For λ_b'' is really λ_b for the cross classification between (say) B and (A, C); hence the material of Section 3.1 applies directly.

3.8. Sampling Experiments

This section presents the results of some sampling experiments that bear on the adequacy of the asymptotic approximations given in the prior sections.

Table 3.8.1 describes the sampling experiments that have been done; in each case, reference is made to the figure that summarizes the results of the experiment graphically. These figures are drawn on normal probability paper, the straight lines represent the standard normal distribution, and each dot has its abscissa equal to the value of a computed statistic and its ordinate equal to the proportion of computed statistics less than or equal to the abscissa. Thus the dots give the "corners" of the observed sample cumulative distribution functions. Deviations of the dots from the straight line arise from two sources: (1) inadequacy of the asymptotic approximation, and (2) sampling fluctuations.

[11] The simple relationships between the ρ's are inherent in the geometry of classical correlation theory, and are not tied to the assumption of normality. The last sentence of Section 12 in [8] is corrected by this remark.

TABLE 3.8.1. SAMPLING EXPERIMENTS

Sample Measure of Association	Statistic Computed from Each Sample	Sample Size	Number of Samples	Population	Figure
L_b	(3.1.4)	200	50	3×4 cross classification of Sec. 3.2	3.8.1
L_b	(3.1.4)	100	50	3×4 cross classification of Sec. 3.2	3.8.2
L_b	(3.1.4)	100	100 (iv)	3×4 cross classification of Sec. 3.2	3.8.3
L_a (i)	(3.3.4)	200	50	3×4 cross classification of Sec. 3.2	3.8.4
L_a	(3.3.4)	100	50	3×4 cross classification of Sec. 3.2	3.8.5
L_a	(3.3.4)	100	100 (iv)	3×4 cross classification of Sec. 3.2	3.8.6
L_b	(3.1.4)	200	50	2×3 cross classification (ii)	3.8.7
L_a (i)	(3.3.4)	200	50	2×3 cross classification (ii)	3.8.8
G	$\dfrac{\sqrt{n}\,(G-\gamma)}{\sqrt{(3.5.5)}}$ (iii)	50	100	3×4 cross classification of Sec. 3.2	3.8.9
G	$\dfrac{\sqrt{n}\,(G-\gamma)}{\sqrt{(3.5.5)}}$	200	50	3×4 cross classification of Sec. 3.2	3.8.10
G	$\dfrac{\sqrt{n}\,(G-\gamma)}{\sqrt{(3.5.5)}}$	200	50	2×3 cross classification (ii)	3.8.11

Notes to the table

(i) L_a for the 3×4 cross classification is, of course, L_b for the transposed (4×3) cross classification. In fact, the same samples were used.

(ii) The 2×3 cross classification used here was

		b: 1	2	3	
a	1	.04	.23	.18	.45
	2	.18	.05	.32	.55
		.22	.28	.50	1.00

For this population, $\lambda_b=.10$, $\lambda_a=.40$, and $\lambda\approx.03$.

(iii) This statistic is the deviation of G from its population value, γ, divided by the *population* asymptotic std. dev. of G. Thus the results here are not directly applicable, in general, since (3.5.5) would hardly ever be known in practice. The corresponding practical statistic, (3.5.7), requires considerably more computation; sampling results for it are presented in Table 3.8.2.

(iv) The 100 samples include the 50 samples of the line just above.

TABLE 3.8.2. THE EMPIRICAL AVERAGE NUMBER OF TIMES PER ONE HUNDRED SAMPLES THAT $\sqrt{n}\ (G-\gamma)$ DIVIDED BY THE SQUARE ROOT OF (3.5.5) EXCEEDED THE VALUES FROM THE UNIT-NORMAL DISTRIBUTION AT THE TWO-SIDED .05, .10, AND .25 LEVELS, WHEN THE SAMPLE SIZE $n=50$

Values of γ	$-.01$ to $+.01$	.20 to .29	.30 to .39	.40 to .49	.50 to .59	.60 to .69	.70 to .79	.80 to .89	.90 to .92	.99 to 1.00
Number of samples	300	800	1900	700	1000	2000	1400	700	300	900
Significance levels										
.25	22	29	26	25	30	30	31	30	26	11
.10	10	13	11	9	12	12	12	9	8	6
.05	6	7	5	4	5	5	5	4	3	5

THE EMPIRICAL AVERAGE NUMBER OF TIMES PER ONE HUNDRED SAMPLES THE STATISTIC (3.5.7) EXCEEDED THE VALUES FROM THE UNIT-NORMAL DISTRIBUTION AT THE TWO-SIDED .05, .10, AND .25 LEVELS, WHEN THE SAMPLE SIZE $n=50$

Values of γ	$-.01$ to $+.01$	.20 to .29	.30 to .39	.40 to .49	.50 to .59	.60 to .69	.70 to .79	.80 to .89	.90 to .92	.99 to 1.00
Number of samples	300	800	1900	700	1000	2000	1400	700	300	900
Significance levels										
.25	23	30	29	25	34	33	37	36	32	67
.10	12	17	15	13	18	21	23	24	21	65
.05	6	10	10	7	12	16	18	19	17	65

Note that for the cross classifications considered, with 6 or 12 cells, sample sizes of 50 and 100 would not ordinarily be considered large. The sample is spread over all cells, while, at least for L_b and L_a, only some cells determine the sample measure of association.

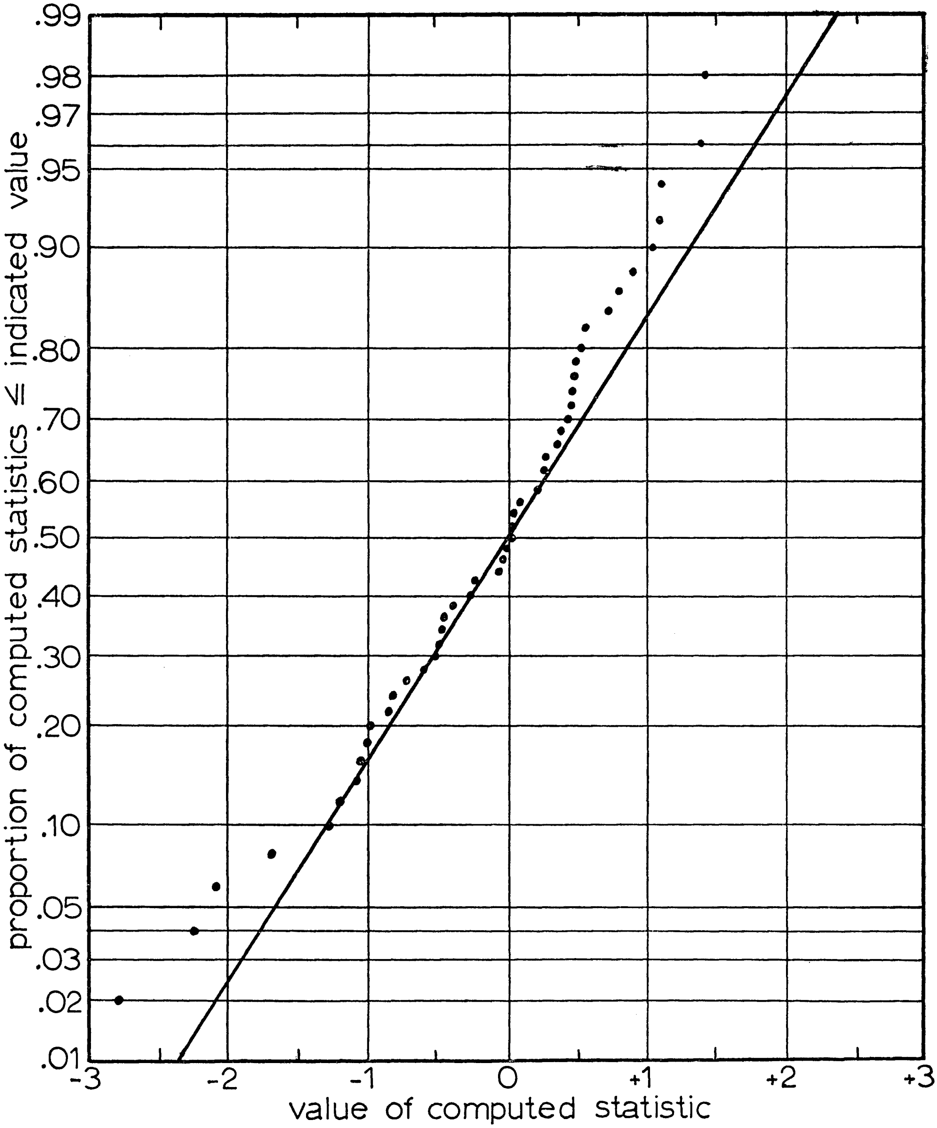
In the light of these considerations, we find the sampling results encouraging. For L_a and L_b, the dots on the figures lie generally near the standard normal straight line. The tail probabilities are particularly important for most statistical applications, and we note that some of the deviations in the tails appear large. In examining the figures, however, keep in mind that the scale near the tails is magnified because of the use of normal probability paper.

We next summarize in Table 3.8.2 some sampling results about G that were obtained by Miss Irene Rosenthal (Institute of Child Welfare, University of California, Berkeley) and that we reproduce here with her kind permission.[12]

Miss Rosenthal's work related to two statistics $\sqrt{n}\ (G-\gamma)/\sqrt{(3.5.5)}$ and (3.5.7). She considered a large number of 5×5 cross classifications, categorized by their values of γ. In Table 3.8.2, the populations are not separated, but results are grouped by ranges of γ values. Table 3.8.2 compares the two-sided tail probabilities from asymptotic unit normality with the observed relative fre-

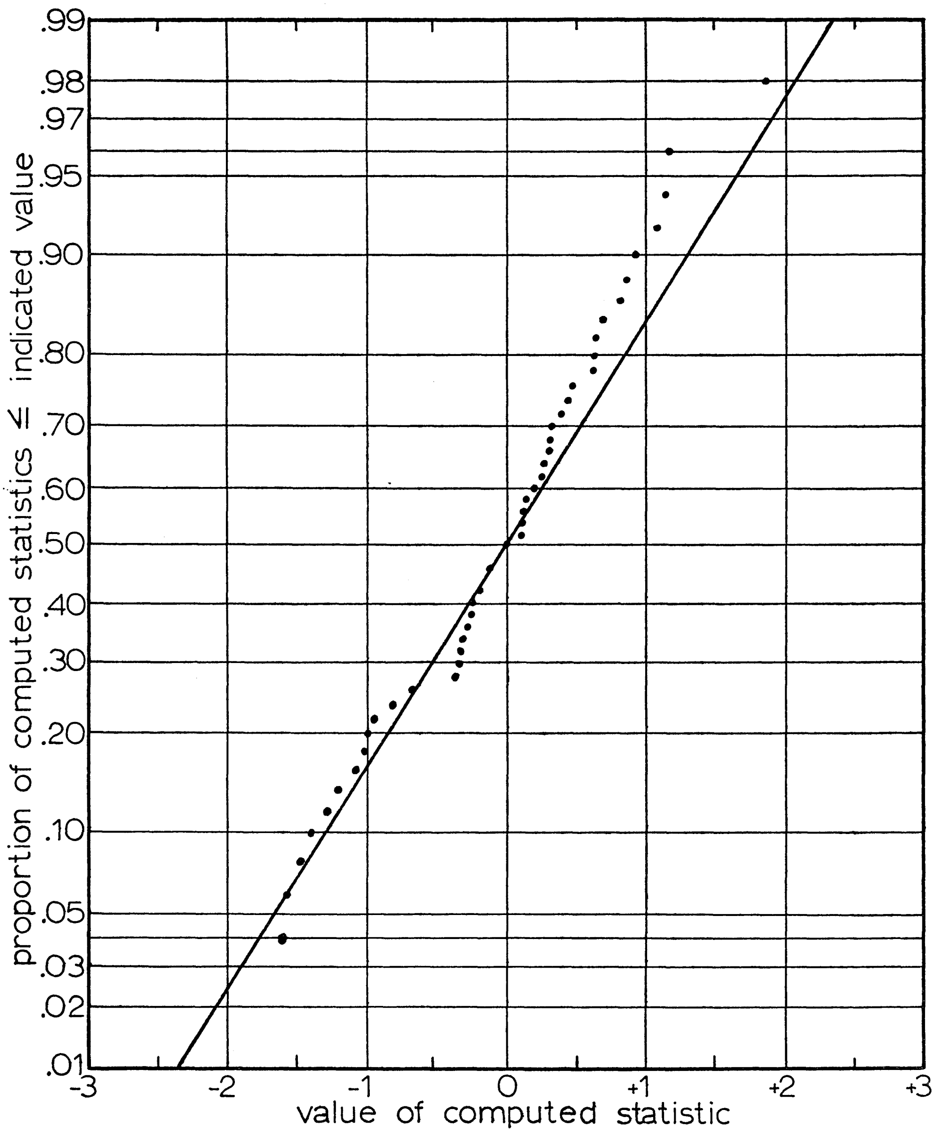
[12] Miss Rosenthal has in preparation a manuscript giving more detailed information about her sampling experiments.

FIG. 3.8.1. Results of Sampling Experiment for (3.1.4). 50 Samples, Each of 200 Cross Classified Observations.



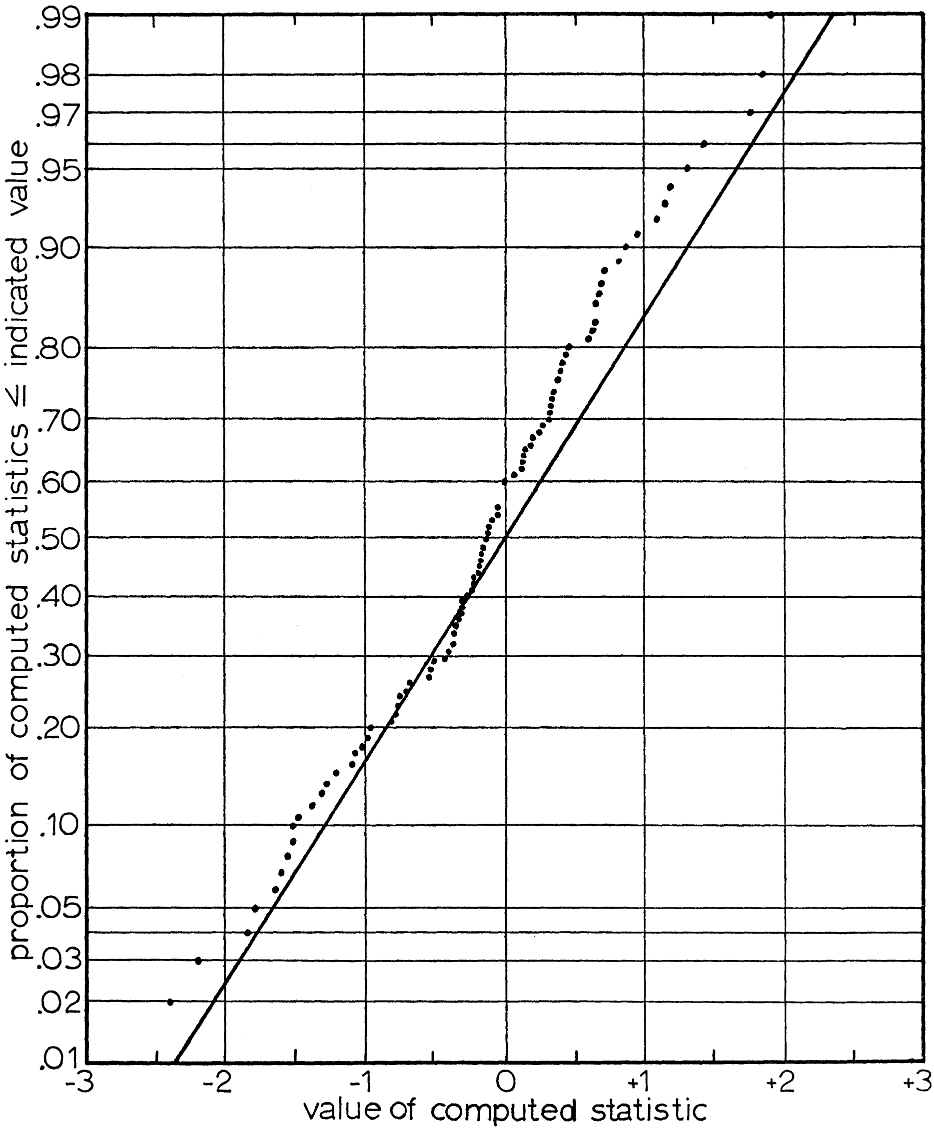
Straight line is unit-normal cumulative.
Unplotted maximum observed value is 1.64.

Fig. 3.8.2. Results of Sampling Experiment for (3.1.4). 50 Samples, Each of 100 Cross Classified Observations.



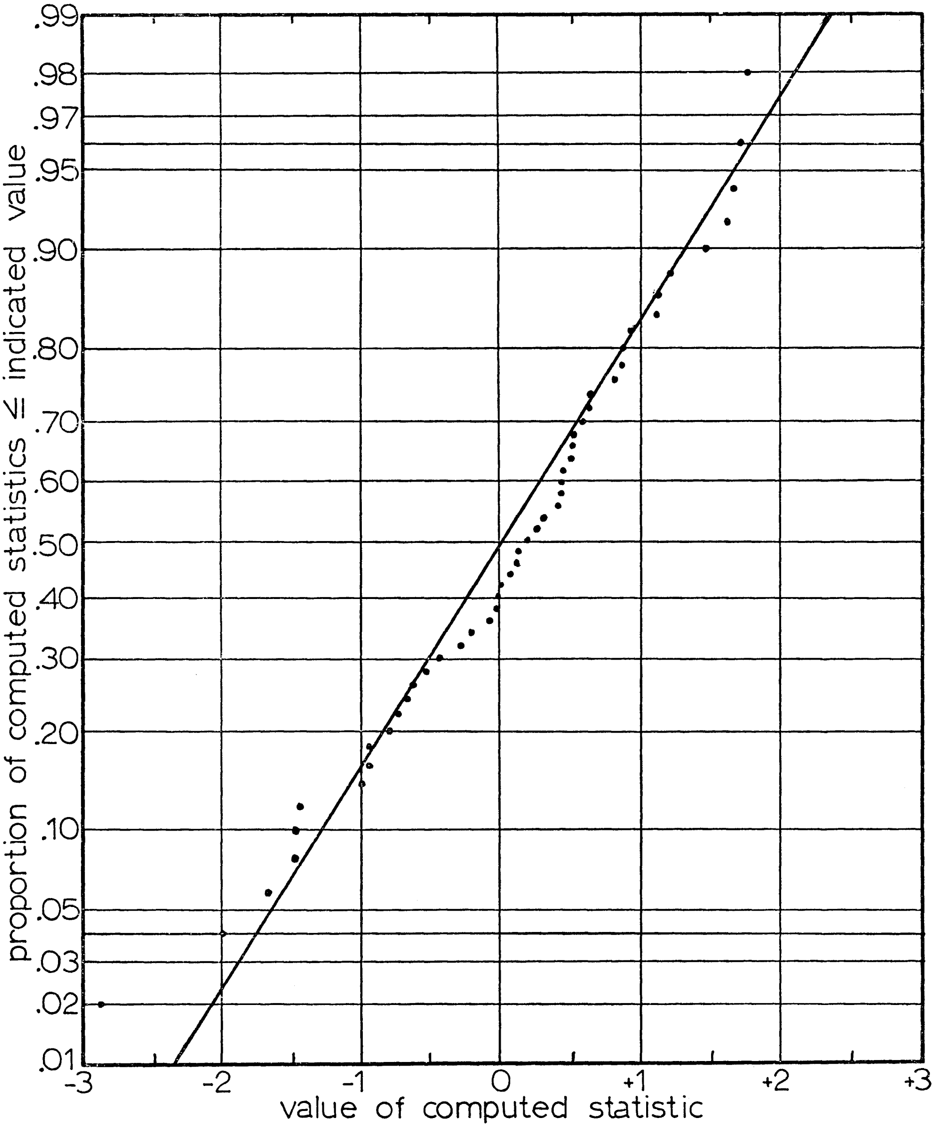
Straight line is unit-normal cumulative.
Unplotted maximum observed value is 2.05; minimum −4.75.

FIG. 3.8.3. Results of Sampling Experiment for (3.1.4). 100 Samples, Each of 100 Cross Classified Observations.



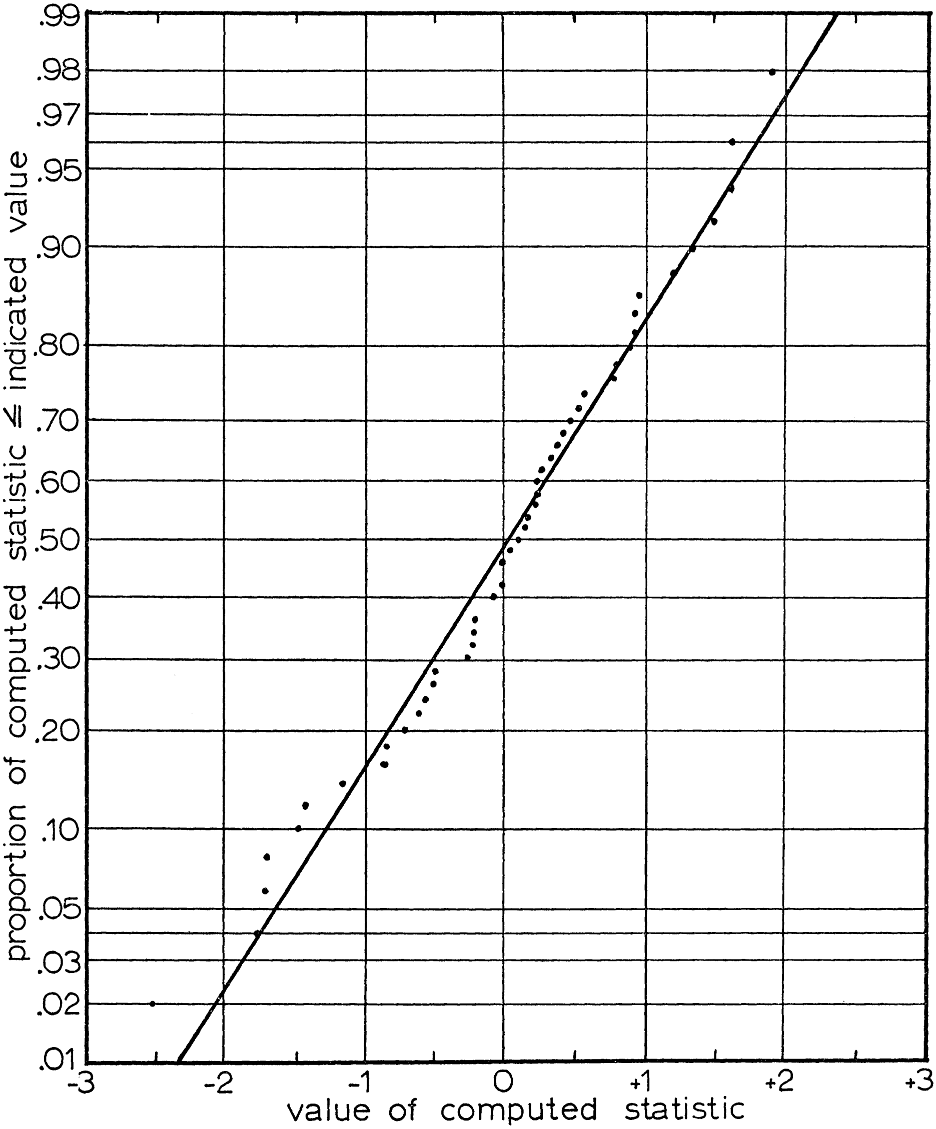
Straight line is unit-normal cumulative.
Unplotted maximum observed value is 2.05; minimum −4.75.

FIG. 3.8.4. Results of Sampling Experiment for (3.3.4). 50 Samples, Each of 200 Cross Classified Observations.



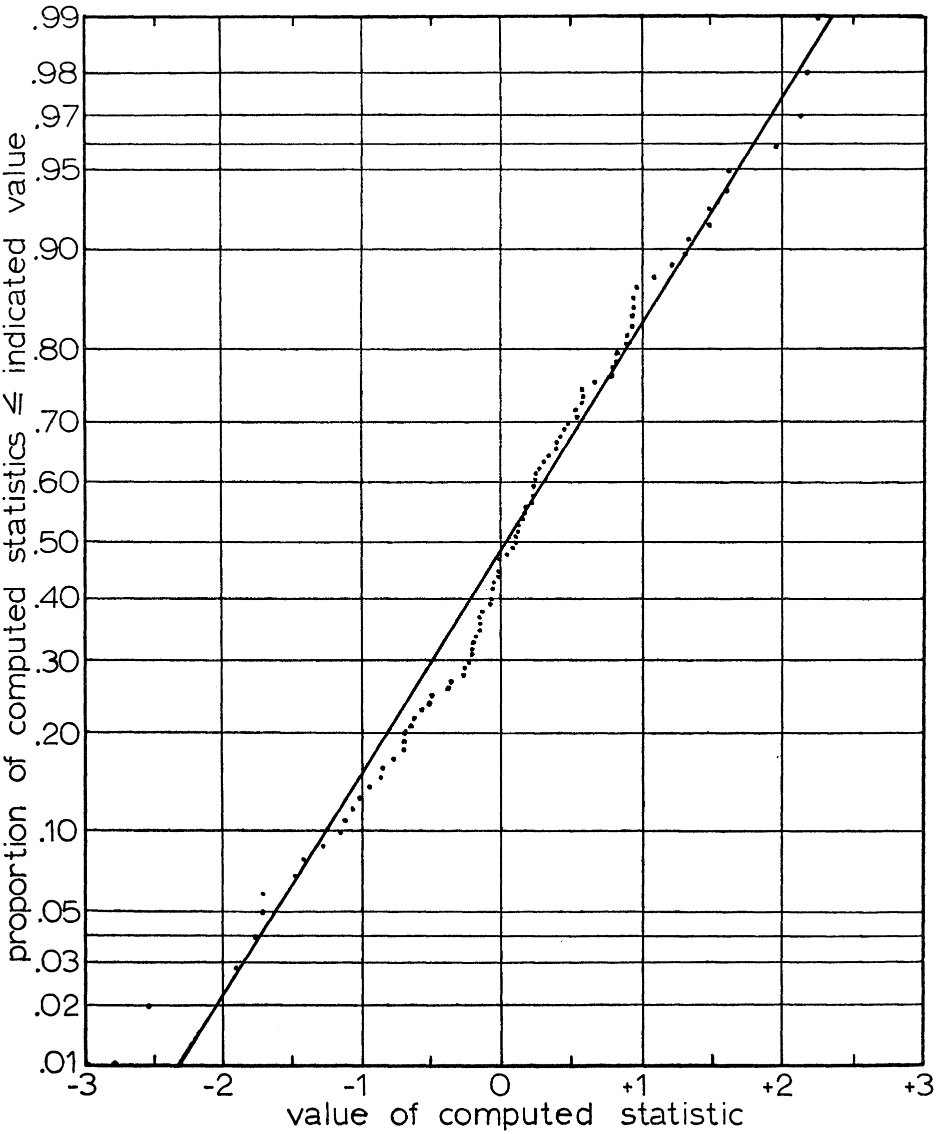
Straight line is unit-normal cumulative.
Unplotted maximum observed value is 2.33.

Fig. 3.8.5. Results of Sampling Experiment for (3.3.4). 50 Samples, Each of 100 Cross Classified Observations.



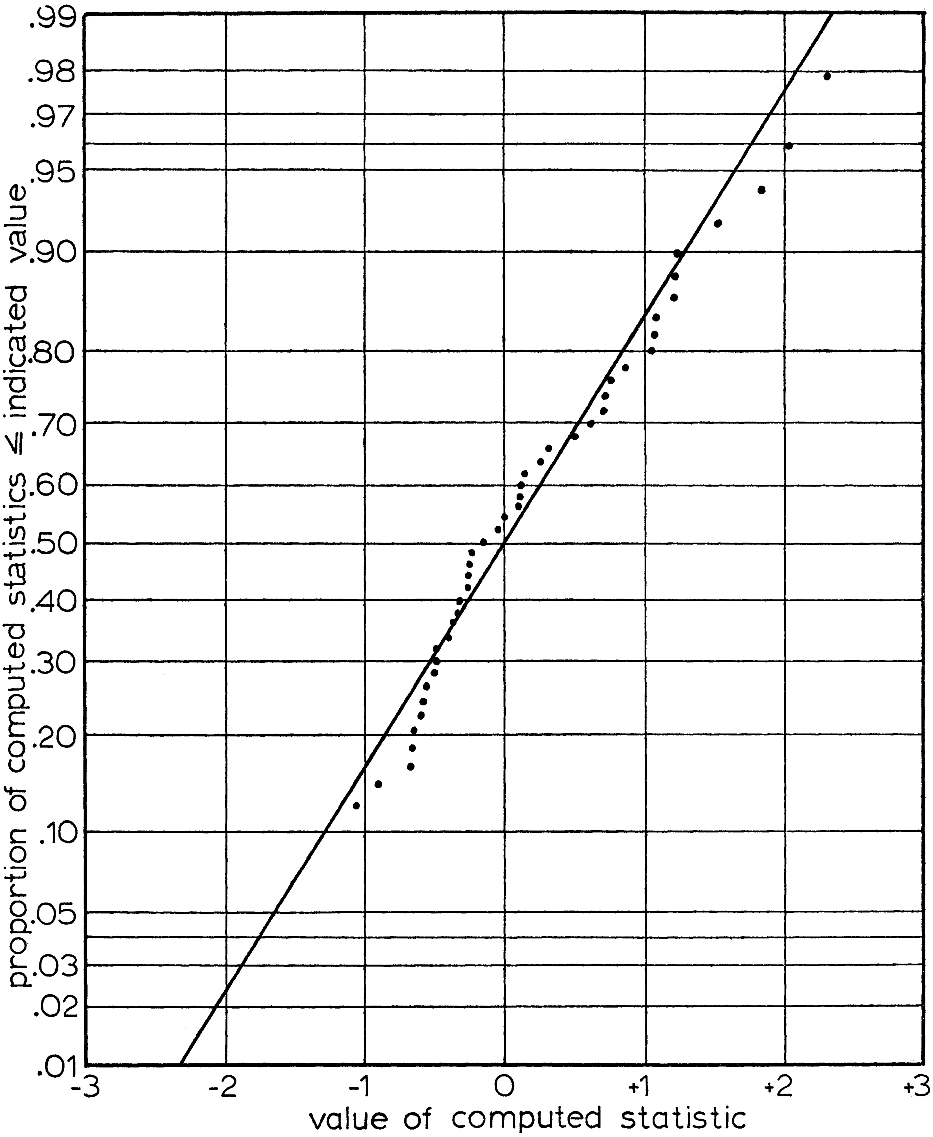
Straight line is unit-normal cumulative.
Unplotted maximum observed value is 2.19.

FIG. 3.8.6. Results of Sampling Experiment for (3.3.4). 100 Samples, Each of 100 Cross Classified Observations.



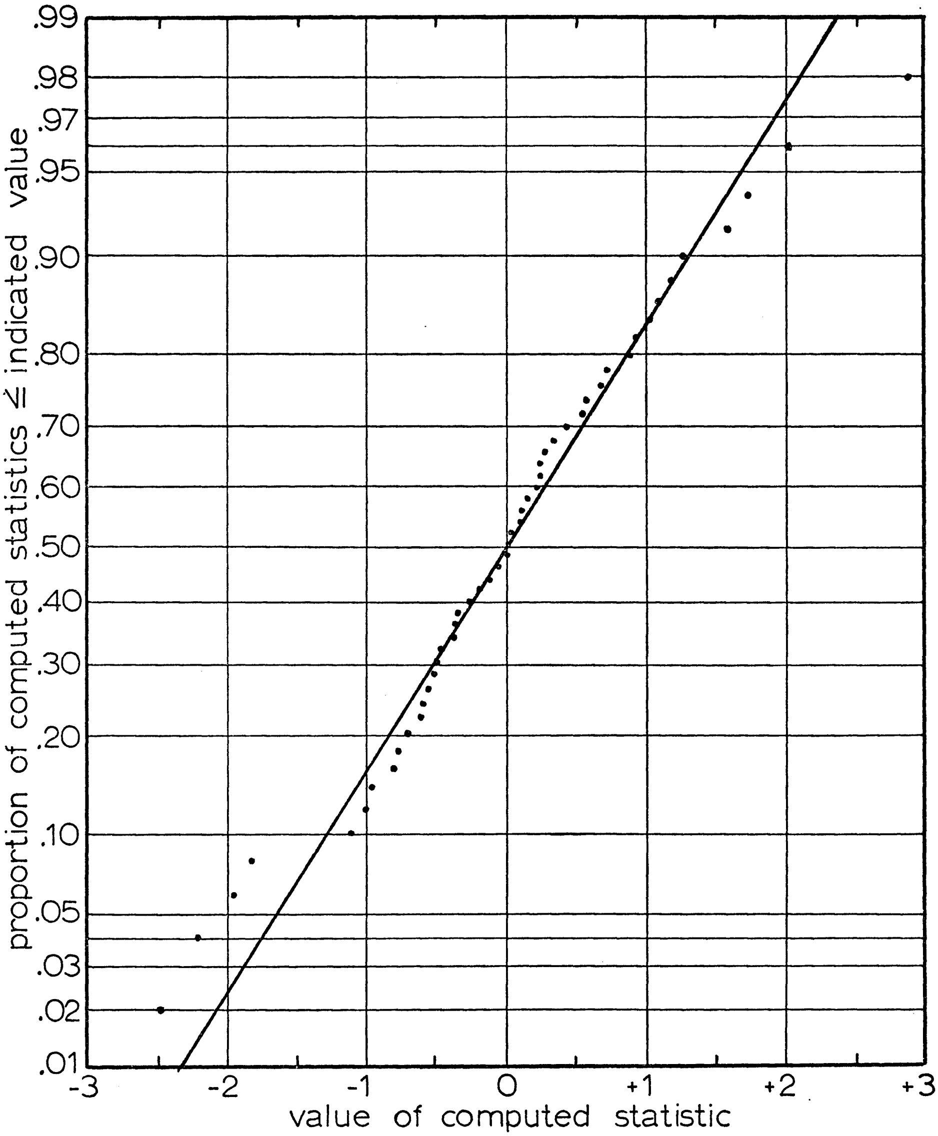
Straight line is unit-normal cumulative.
Unplotted maximum observed value is **2.61.**

FIG. 3.8.7. Results of Sampling Experiment for (3.1.4). 50 Samples, Each of 200 Cross Classified Observations.



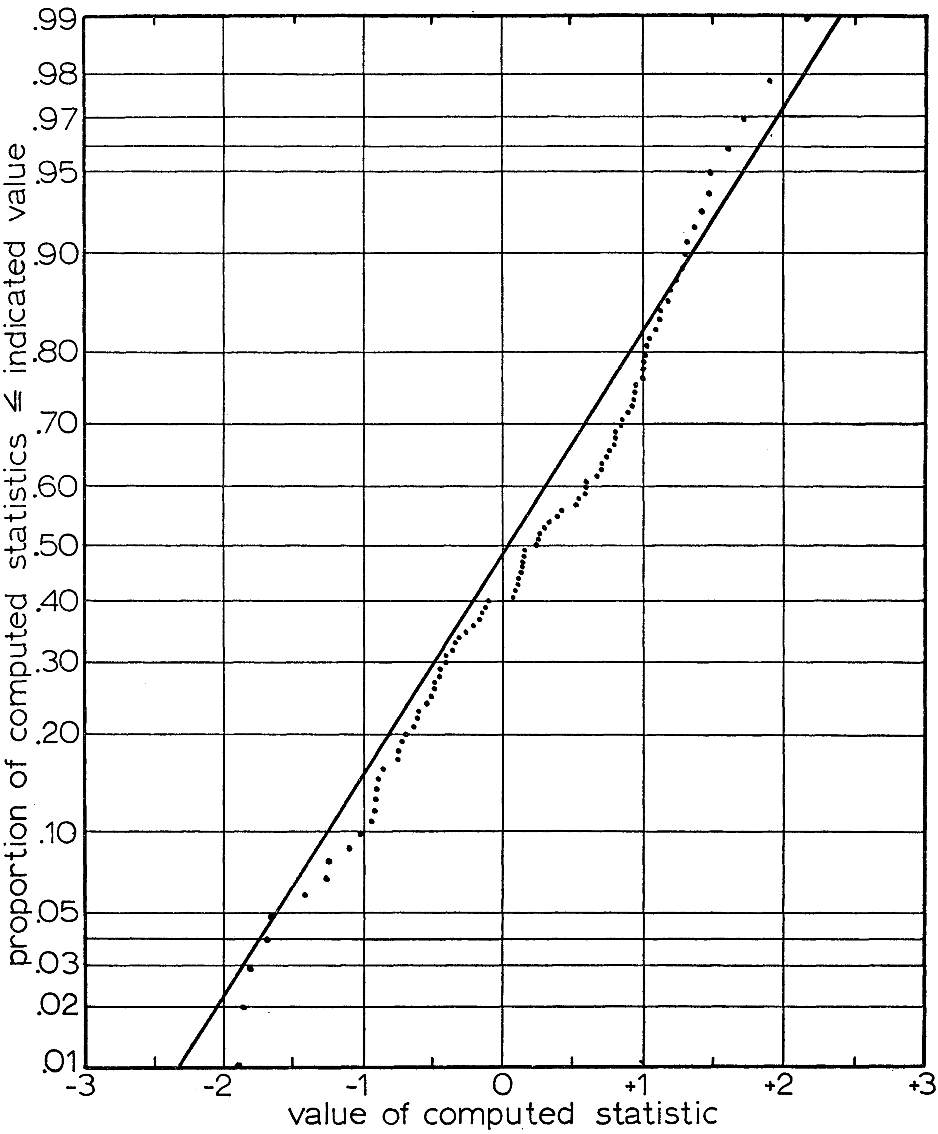
Straight line is unit-normal cumulative.
Unplotted maximum observed value is 2.88; five minimum values are $-\infty$.

FIG. 3.8.8. Results of Sampling Experiment for (3.3.4). 50 Samples, Each of 200 Cross Classified Observations.



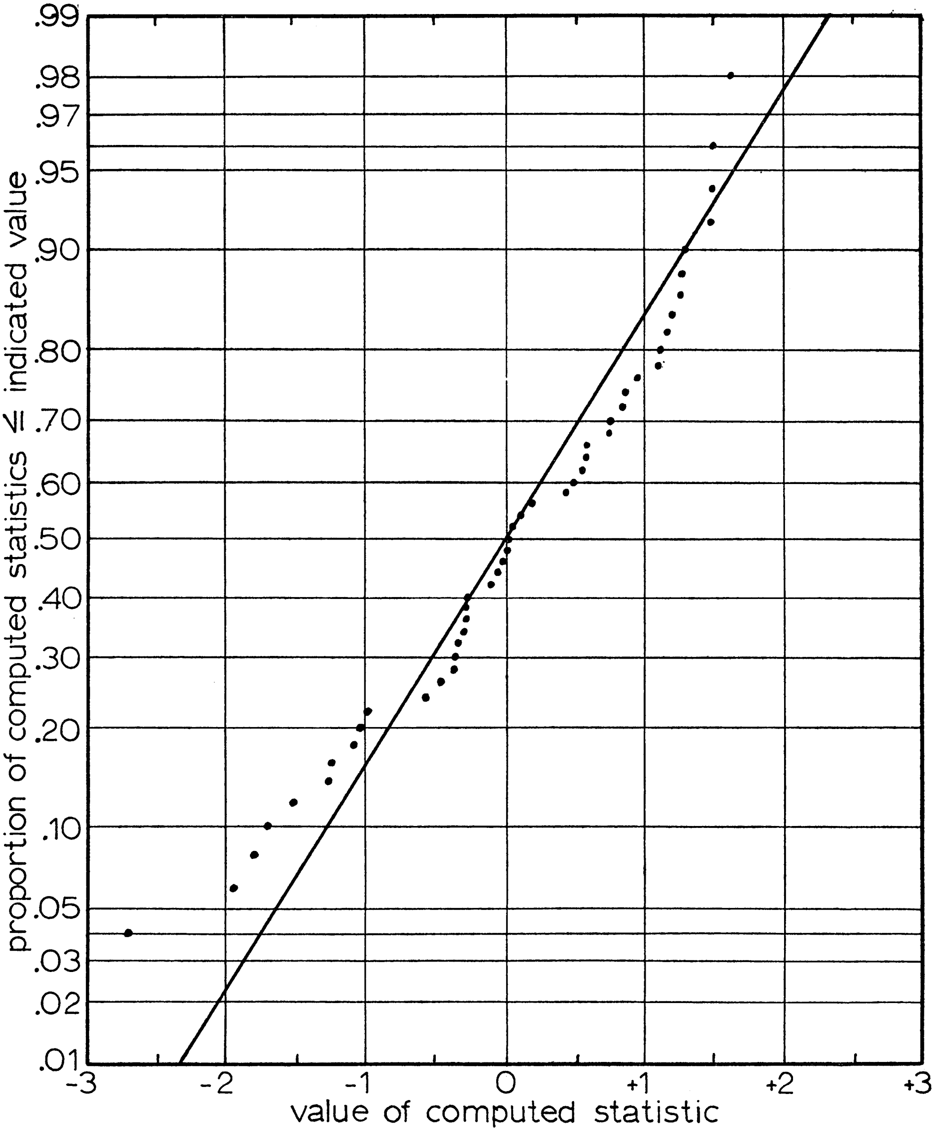
Straight line is unit-normal cumulative.
Unplotted maximum observed value is 3.01.

FIG. 3.8.9. Results of Sampling Experiment for $\sqrt{n}(G-\gamma)/\sqrt{(3.5.5)}$. 100 Samples, Each of 50 Cross Classified Observations.



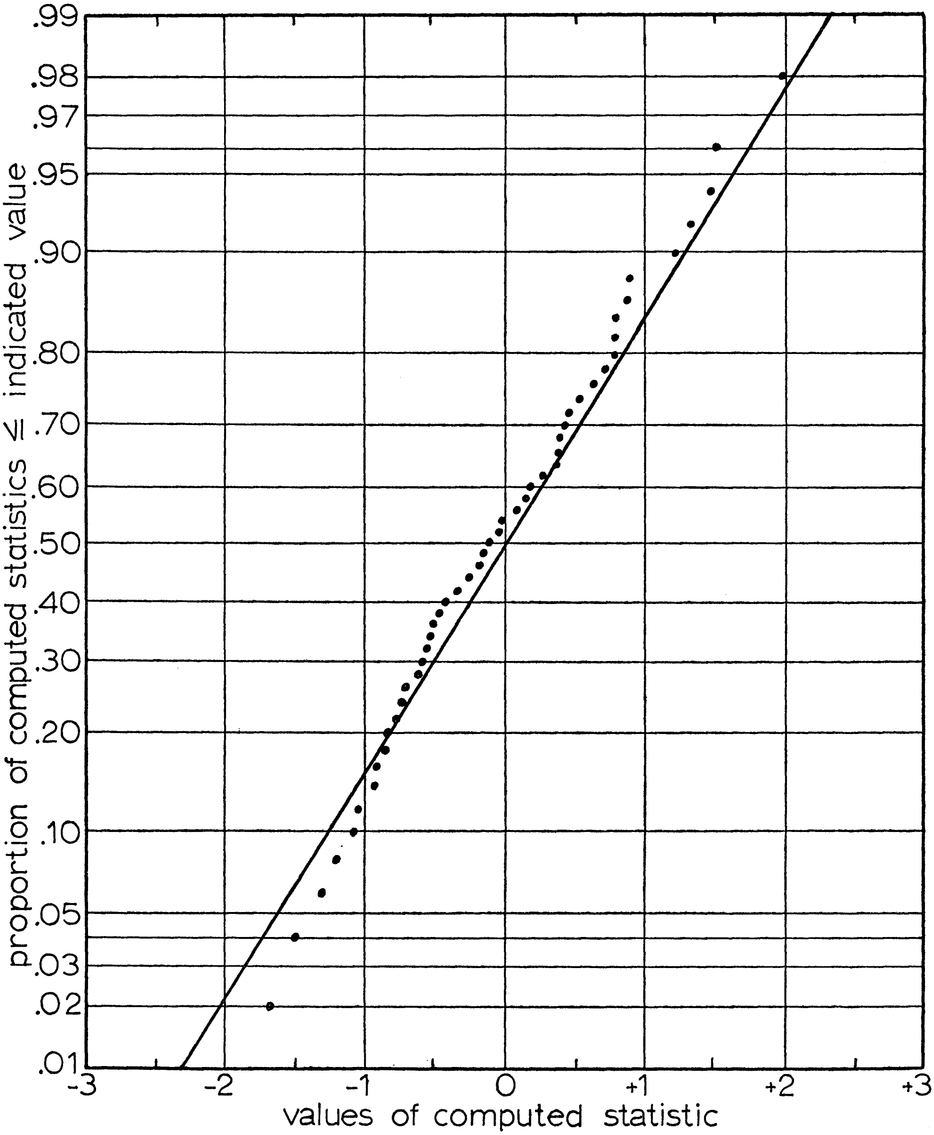
Straight line is unit-normal cumulative.
Unplotted maximum observed value is 2.23.

FIG. 3.8.10. Results of Sampling Experiment for $\sqrt{n}(G-\gamma)/\sqrt{(3.5.5)}$. 50 Samples, Each of 200 Cross Classified Observations.



Straight line is unit-normal cumulative.
Unplotted maximum observed value is 1.75; minimum −3.14.

FIG. 3.8.11. Results of Sampling Experiment for $\sqrt{n}(G-\gamma)/\sqrt{(3.5.5)}$. 50 Samples, Each of 200 Cross Classified Observations.



Straight line is unit-normal cumulative.
Unplotted maximum observed value is 2.83.

quencies from her sampling. The sample size, n, was 50 throughout, and 100 random samples were drawn from each of the 5×5 populations. Thus, for example, we see from Table 3.8.2 that there were eight 5×5 populations with γ values in the range [.20, .29].

The evidence of Table 3.8.2 suggests that the distribution of $\sqrt{n}\ (G-\gamma)/\sqrt{(3.5.5)}$ is reasonably well approximated by the unit normal distribution, for populations of the kind considered by Miss Rosenthal.

The distribution of (3.5.7), on the other hand, seems to be reasonably well approximated by the unit normal distribution, for populations of Miss Rosenthal's kind, only for values of $|\gamma|$ less than .5, and, even for such values of γ, there is a persistent tendency to have more frequent observations in the tails than the unit normal approximation predicts. This behavior is consistent with the presence in (3.5.7) of a random denominator. Of course, 50 can hardly be considered a large sample size, especially for 5×5 cross classifications! The distribution of (3.5.7) is more relevant to practice than that of $\sqrt{n}\ (G-\gamma)/\sqrt{(3.5.5)}$, and we plan to discuss (3.5.7) further in a subsequent publication.

4. MULTINOMIAL SAMPLING WITHIN EACH ROW (COLUMN) OF THE DOUBLE POLYTOMY

4.1. Preliminaries

Instead of sampling with replacement over all the cells of an $A\times B$ double polytomy, thus obtaining a sample point governed by a multinomial distribution, it may be necessary, desirable, or efficient to preassign and fix the sample size within each row (or, alternatively, column) and sample independently within each row (column) with replacement. Thus one will obtain a sample point governed by a product of α (or β) independent multinomial distributions.[13] For definiteness, suppose that the separate multinomials are within each row.

Since the sample sizes within rows are fixed, the N_{ab}'s are now subject to the α additional restrictions $\sum_b N_{ab}=n_{a\cdot}$, in addition to the over-all restriction $\sum_a\sum_b N_{ab}=n$. Within the ath row, the quantities $\sqrt{n_{a\cdot}}\ [(N_{ab}/n_{a\cdot})-(\rho_{ab}/\rho_{a\cdot})]$ have zero means, variances $(\rho_{ab}/\rho_{a\cdot})-(\rho_{ab}/\rho_{a\cdot})^2$, and covariances $-(\rho_{ab}/\rho_{a\cdot})(\rho_{ab'}/\rho_{a\cdot})$ for $b\neq b'$. As $n_{a\cdot}\to\infty$, the distribution of the quantities in question approaches that multivariate normal distribution with zero means and with the same variance-covariance structure. The N_{ab}'s in different rows are independent, both for finite samples and asymptotically. In considering asymptotic distributions for all rows together, we shall assume that, as $n\to\infty$, the ratios $n_{a\cdot}/n$ approach definite limits unequal to zero or one.

For this sampling procedure, and without ancillary knowledge, it is impossible to estimate the ρ_{ab}'s themselves, since the distribution of the sample point depends only upon the row-wise *conditional* probabilities $\rho_{ab}/\rho_{a\cdot}$. Hence, if we want to estimate or make tests on such measures as λ_b, λ, γ, etc. ,which themselves do not depend on the ρ_{ab}'s via the conditional probabilities $\rho_{ab}/\rho_{a\cdot}$, we must either assume the $\rho_{a\cdot}$'s known or perform a separate experiment to ob-

[13] One can also obtain such a product of distributions by starting with a multinomial distribution over the whole $\alpha\times\beta$ tableau, and then considering the conditional distribution *given* $N_{1\cdot}=n_{1\cdot}, \cdots, N_{\alpha\cdot}=n_{\alpha\cdot}$. This device and its extensions, are often used for the presentation of tests and the computation of so-called P-values in tests of independence with $\alpha\times\beta$ cross classifications.

tain estimates for them. If, however, we wish to estimate or make tests on a measure such as λ_b^* [8, Sec. 5.4], which is a function only of the ratios $\rho_{ab}/\rho_{a\cdot}$, then we may proceed without the assumption of knowledge of marginals and without performing a separate experiment. In the next three sub-sections we shall present examples of the above procedures.

4.2. The Index λ_b, With Marginal Row Probabilities Known

Perhaps the simplest case of the foregoing general discussion is that in which we are concerned with λ_b and in which

i) there is separate multinomial sampling in the rows,
ii) the row marginals, $\rho_{a\cdot}$, are known (we assume that they are positive),
iii) the maximum column marginal, $\rho_{\cdot m}$, is known, and
iv) sampling rates in the several rows are such that $n_{a\cdot}=n\rho_{a\cdot}$; that is, the sample sizes in the rows are proportional to the known row marginals.

Such a case might arise if the marginal probabilities are known, say, from census data, and if the sampling rates in the rows may be determined at our convenience.[14] We shall drop assumption (iii) shortly, but the development is simpler if we first suppose $\rho_{\cdot m}$ known.

A natural estimator for $\rho_{ab}/\rho_{a\cdot}$ is $N_{ab}/n_{a\cdot}=N_{ab}/(n\rho_{a\cdot})$. Hence N_{ab}/n is a natural estimator for ρ_{ab}. The corresponding estimator for λ_b is

$$L_b = \frac{\sum (N_{am}/n) - \rho_{\cdot m}}{1 - \rho_{\cdot m}}\,. \tag{4.2.1}$$

Strictly speaking, we should attach an added symbol to "L_b" in order to emphasize the fact that it is a different quantity from L_b of Section 3.1. We shall, however, refrain from doing this because of the already bothersome prolixity of notation.

The distribution of this new L_b depends only on the sum of the N_{am}'s. As before, we may, for asymptotic purposes, suppose that N_{am} is taken on in the same cell of the ath row in which ρ_{am} is taken. Hence, the quantities $\sqrt{n_{a\cdot}}\,[(N_{am}/n_{a\cdot})-(\rho_{am}/\rho_{a\cdot})]=\sqrt{n}\;[(N_{am}/n)-\rho_{am}]/\sqrt{\rho_{a\cdot}}$ are asymptotically independent normal deviates with zero means and variances $(\rho_{am}/\rho_{a\cdot})[1-(\rho_{am}/\rho_{a\cdot})]$. (The expression for variance and the independence hold also for finite samples.) From this, we see that the quantities $\sqrt{n}\,[(N_{am}/n)-\rho_{am}]$ are asymptotically independent with zero means and variances $(\rho_{am}/\rho_{a\cdot})(\rho_{a\cdot}-\rho_{am})$. Hence, $\sqrt{n}\{\sum(N_{am}/n)-\sum\rho_{am}\}$ is asymptotically normal with zero mean and variance $\sum_a(\rho_{am}/\rho_{a\cdot})(\rho_{a\cdot}-\rho_{am})$. Since

$$\sqrt{n}(L_b-\lambda_b) = \sqrt{n}\left\{\sum (N_{am}/n) - \sum \rho_{am}\right\}/(1-\rho_{\cdot m}),$$

we have that $\sqrt{n}(L_b-\lambda_b)$ is asymptotically normal with zero mean and variance

$$\frac{1}{(1-\rho_{\cdot m})^2}\sum_a \frac{\rho_{am}}{\rho_{a\cdot}}(\rho_{a\cdot}-\rho_{am}). \tag{4.2.2}$$

This is zero if and only if $\lambda_b=1$.

[14] Of course $n\rho_{a\cdot}$ will not in general be an integer, but for large n the difference will be unimportant, and asymptotically it makes no difference at all. We should also point out that setting $n_{a\cdot}=n\rho_{a\cdot}$ here is purely for convenience; we have not examined the question of whether or not this choice is good from the point of view of power.

Thus, if $\lambda_b \neq 1$, by the same argument as before, the following quantity is asymptotically unit-normal:

$$\sqrt{n}(L_b - \lambda_b) \frac{1 - \rho_{\cdot m}}{\sqrt{\sum \{\rho_{a\cdot}[N_{am}/n_{a\cdot}][1 - (N_{am}/n_{a\cdot})]\}}}, \qquad (4.2.3a)$$

which may also be written as

$$n^{3/2}(L_b - \lambda_b) \frac{1 - \rho_{\cdot m}}{\sqrt{\sum \{N_{am}(n\rho_{a\cdot} - N_{am})/\rho_{a\cdot}\}}} \cdot \qquad (4.2.3b)$$

As an example of the method of using the normal approximation, suppose that $\alpha = \beta = 2$, $n = 30$, $n_{1\cdot} = 18$, $n_{2\cdot} = 12$, and that $\rho_{\cdot m} = 0.5$. (Note that $\rho_{\cdot m}$ need not be unique in this context.) Suppose further that the sample values turn out to be

$$N_{11} = 7 \qquad N_{12} = 11$$
$$N_{21} = 0 \qquad N_{22} = 12.$$

Then L_b is

$$\frac{1}{1 - .50}\left[\frac{(.6)(11)}{18} + \frac{(.4)(12)}{12} - .50\right] = .533.$$

The approximate 95% symmetrical confidence statement is that the following quantity lies between -1.96 and $+1.96$:

$$\frac{\sqrt{30}(.533 - \lambda_b)}{\sqrt{\left[\frac{(.6)(11)}{18}\frac{7}{18} + \frac{(.4)(12)}{12}\frac{0}{12}\right] \Big/ (1 - .50)^2}} = \frac{.533 - \lambda_b}{.138},$$

that is to say

$$.263 \leq \lambda_b \leq .803.$$

In order to obtain some idea of the adequacy of the normal approximation, the actual distribution of (4.2.3) was obtained in the following case:

$\alpha = \beta = 2$ $\qquad n_{1\cdot} = 18$ $\qquad n_{2\cdot} = 12$ $\qquad n = 30$

$\rho_{11} = .42$	$\rho_{12} = .18$	$\rho_{1\cdot} = .6$
$\rho_{21} = .08$	$\rho_{22} = .32$	$\rho_{2\cdot} = .4$
$\rho_{\cdot 1} = .50$	$\rho_{\cdot 2} = .50$	

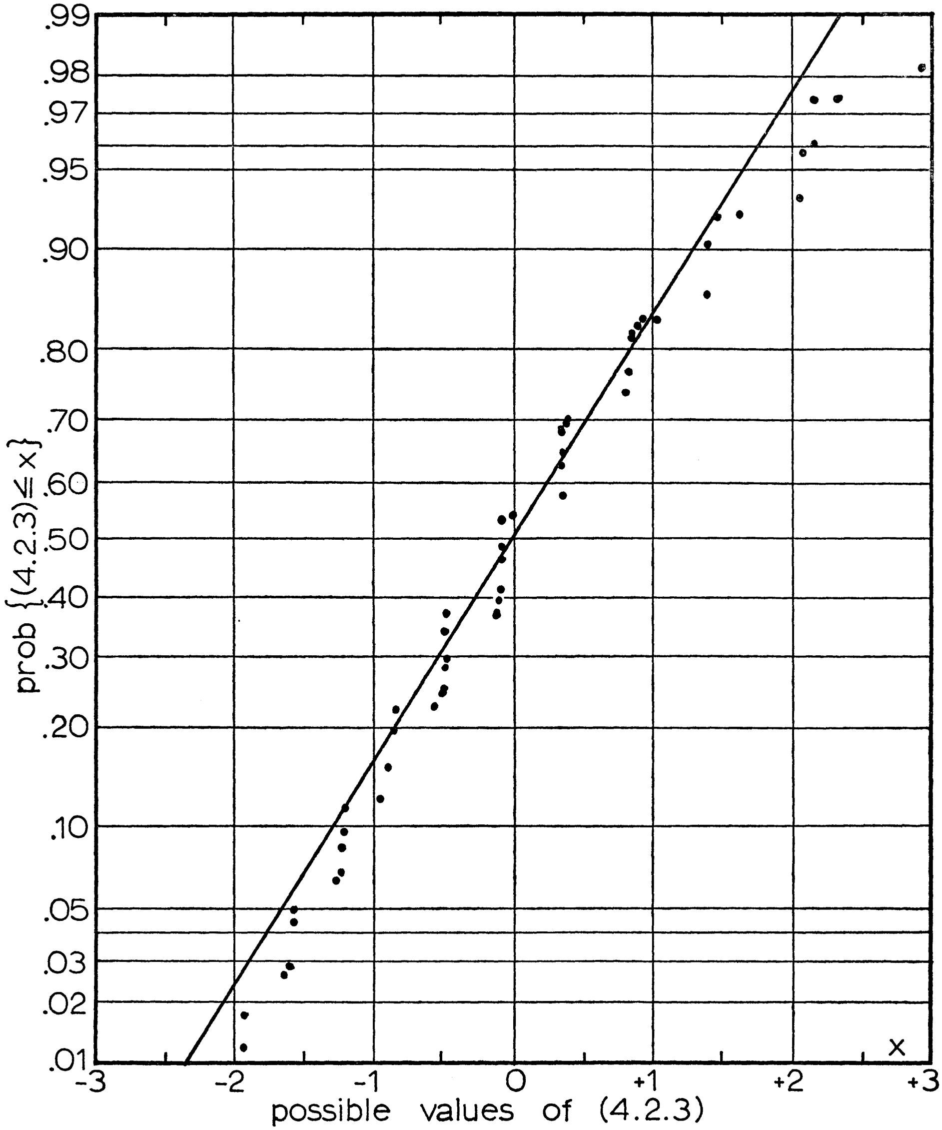
$\lambda_b = .48$

and the results[15] are shown in Figure 4.2.1. The dots in Figure 4.2.1 appear at the "corners" of the discrete cumulative distribution graph of (4.2.3); for example, the probability that $(4.2.3) \leq 1.3717$ (to four places) is .86164 (to five places).

The approximation seems quite satisfactory for most statistical purposes in

[15] Because of the small numbers, it was feasible to find the exact distribution rather than resort to random sampling.

FIG. 4.2.1. Cumulative Distribution of (4.2.3).



Straight line is unit-normal cumulative.

Three points omitted at left (lowest is at -2.63) and seven points omitted at right (highest is at 7.10).

this instance, and the sample size is by no means huge. Of course, this is only one case, and a rather special one at that.

It seems worthwhile to discuss briefly the results when assumption (iii) is dropped. In this case, $\rho_{\cdot m}$ is naturally estimated by the maximum likelihood estimator

$$\operatorname*{Max}_{b}\,(N_{\cdot b}/n) = N_{\cdot m}/n = R_{\cdot m}, \tag{4.2.4}$$

and λ_b by

$$L_b = \frac{\sum R_{am} - R_{\cdot m}}{1 - R_{\cdot m}}, \tag{4.2.5}$$

which is the *same* as (3.1.2). However, the probability structure is different because of independence from row-to-row.

Assuming that λ_b is well-defined, and that $\rho_{\cdot m}$ is unique, we may compute, much as in Section 3.1, that $\sqrt{n}(L_b-\lambda_b)$ is now asymptotically normal with zero mean and variance

$$\begin{aligned}(1 - \rho_{\cdot m})^2 \sum \rho_{am}[1 - (\rho_{am}/\rho_{a\cdot})] + (1 - \sum \rho_{am})^2 \sum \rho_{a(\cdot m)}[1 - (\rho_{a(\cdot m)}/\rho_{a\cdot})] \\ - 2(1 - \rho_{\cdot m})(1 - \sum \rho_{am})\left\{ \sum{}^{r} \rho_{am} - \sum (\rho_{am}\rho_{a(\cdot m)}/\rho_{a\cdot})\right\},\end{aligned} \tag{4.2.6}$$

all divided by $(1-\rho_{\cdot m})^4$, where $\rho_{a(\cdot m)}$ denotes that ρ_{ab} in the ath row of the column for which $\rho_{\cdot b}$ is maximum. The asymptotic variance just stated is zero (assuming that λ_b is well-defined) if and only if $\lambda_b=0$ or 1. Consequently, $\sqrt{n}(L_b-\lambda_b)$ divided by the square root of the sample analogue of the variance stated above, under the assumption that λ_b is well-defined and $\neq 0$ or 1, is asymptotically unit normal. Approximate tests and confidence intervals may thus be found just as before.

Similar, but more complex, results may be obtained for the case in which only (i) and (ii) stated at the beginning of this section hold; that is, for the case in which there is separate multinomial sampling in the rows and the row marginals are known, but nothing is known of the column marginals and the row-wise sampling rates are *not* proportional to the row marginals.

4.3. The Index λ_b^*

An index of association, related to λ_b, is

$$\lambda_b^* = \frac{\sum \overset{*}{\rho}_{am} - \overset{*}{\rho}_{\cdot m}}{1 - \rho^*_{\cdot m}}, \tag{4.3.1}$$

where

$$\overset{*}{\rho}_{ab} = \rho_{ab}/(\alpha\rho_{a\cdot}), \quad \overset{*}{\rho}_{am} = \operatorname*{Max}_{b}\, \overset{*}{\rho}_{ab}, \quad \text{and} \quad \overset{*}{\rho}_{\cdot m} = \operatorname*{Max}_{b} \sum_{a} \overset{*}{\rho}_{ab}.$$

The use of this index was motivated in [8] by its independence of the row marginals $\rho_{a\cdot}$.

If we have separate multinomial sampling in the rows, a natural estimator of ρ_{ab}^* is $N_{ab}/(\alpha n_{a\cdot})$, and a natural estimator of λ_b^* is[16]

$$L_b^+ = \frac{\sum R_{am}^* - R_{\cdot m}^*}{1 - R_{\cdot m}^*}. \tag{4.3.2}$$

where

$$R_{ab}^* = R_{ab}/(\alpha R_{a\cdot}) = N_{ab}/(\alpha n_{a\cdot}), \quad R_{am}^* = \operatorname*{Max}_b R_{ab}^*, \quad \text{and} \quad R_{\cdot m}^* = \operatorname*{Max}_b \sum_a R_{ab}^*.$$

In order to approximate asymptotically the behavior of L_b^+ in this situation, we suppose that the $n_{a\cdot}$'s, or row-wise sampling rates, increase together so that $n_{a\cdot}/n \to \sigma_a$. With this hypothesis, we have from Section 4.1 that the quantities

$$\begin{aligned}\sqrt{n}(R_{ab}^* - \rho_{ab}^*) &= \sqrt{n}[(N_{ab}/n_{a\cdot}) - (\rho_{ab}/\rho_{a\cdot})]/\alpha \\ &\approx \sqrt{n_{a\cdot}}[(N_{ab}/n_{a\cdot}) - (\rho_{ab}/\rho_{a\cdot})]/(\alpha\sqrt{\sigma_a})\end{aligned}$$

are independent between rows, have zero means, and are asymptotically normal with variances

$$(\rho_{ab}/\rho_{a\cdot})[1 - (\rho_{ab}/\rho_{a\cdot})]/(\alpha^2\sigma_a) = \rho_{ab}^*(1 - \alpha\rho_{ab}^*)/(\alpha\sigma_a).$$

As before, assume that ρ_{am}^* and $\rho_{\cdot m}^*$ are unique; whence, by a straightforward manipulation, $\sqrt{n}(L_b^+ - \lambda_b^*)$ is asymptotically normal with zero mean and variance

$$\begin{aligned}(1 - \rho_{\cdot m}^*)^2 \sum [\rho_{am}^*(1 - \alpha\rho_{am}^*)/(\alpha\sigma_a)] + (1 - \rho_{am}^*)^2 \sum [\rho_{a(\cdot m)}^*(1 - \alpha\rho_{a(\cdot m)}^*)/(\alpha\sigma_a)] \\ - 2(1 - \rho_{\cdot m}^*)(1 - \rho_{am}^*)\left\{\sum{}^r \rho_{am}^*/(\alpha\sigma_a) - \sum \rho_{am}^*\rho_{a(\cdot m)}^*/\sigma_a\right\},\end{aligned}$$

all divided by $(1 - \rho_{\cdot m}^*)^4$. Here $\rho_{a(\cdot m)}^*$ is that ρ_{ab}^* in the ath row of the column for which $\sum_a \rho_{ab}^*$ is maximum; and $\sum^r$ means a sum over those values of a for which $\rho_{am}^* = \rho_{a(\cdot m)}^*$. The asymptotic variance just stated is zero (assuming that λ_b^* is well defined) if and only if λ_b^* is 0 or 1.

Consequently, $\sqrt{n}(L_b^+ - \lambda_b^*)$, divided by the square root of the sample analogue of the variance stated above, under the assumption that λ_b^* is well defined and $\neq 0$ or 1, is asymptotically unit-normal. Approximate tests and confidence intervals may be found as before. If the sampling rate is the same in each row, i.e., if $\sigma_a = 1/\alpha$, then the asymptotic variance simplifies somewhat.

4.4. The Index τ_b, with Marginal Row Probabilities Known

In Section 9 of [8], we mentioned a measure of association based, not on optimal prediction, but on proportional prediction in a manner there explained. This measure is

$$\tau_b = \frac{\sum_a \sum_b \rho_{ab}^2/\rho_{a\cdot} - \sum_b \rho_{\cdot b}^2}{1 - \sum_b \rho_{\cdot b}^2}. \tag{4.4.1}$$

[16] The reason for using a plus sign as superscript, instead of an asterisk, is to emphasize that we are here dealing with separate multinomial sampling in the rows of the cross classification.

When conditions (i), (ii), and (iv), stated in Section 4.2, hold, a natural estimator for τ_b is

$$t_b = \frac{\sum_a \sum_b R_{ab}^2/r_{a\cdot} - \sum_b R_{\cdot b}^2}{1 - \sum_b R_{\cdot b}^2}, \tag{4.4.2}$$

where $R_{ab}=N_{ab}/n$, $r_{a\cdot}=n_{a\cdot}/n$, $R_{\cdot b}=N_{\cdot b}/n$. (The symbol t_b should not be confused with Stuart's t_c, referred to in Section 3.5; these measures of association are quite different.) It can be seen that $\sqrt{n}(t_b-\tau_b)$ is asymptotically normally distributed with zero mean and variance

$$4\left\{ \sum_a \left[\frac{\rho_{ab}}{\rho_{a\cdot}}(1 - \sum_b \rho_{\cdot b}^2) - \rho_{\cdot b}\left(1 - \sum_{ab} \frac{\rho_{ab}^2}{\rho_{a\cdot}}\right)\right]^2 \rho_{ab} - \left[\sum_{ab} \frac{\rho_{ab}^2}{\rho_{a\cdot}} - \sum_b \rho_{\cdot b}^2 \right]^2 \right\}, \tag{4.4.3}$$

all divided by $(1-\sum_b \rho_{\cdot b}^2)^4$. Consequently, assuming that (4.4.1) is well defined and that (4.4.3) is different from zero, $\sqrt{n}(t_b-\tau_b)$ divided by the square root of the sample analogue of the variance (4.4.3) stated above is asymptotically unit-normal.

Similar, but more complex, results may be obtained in the situation where there is separate multinomial sampling in the rows and the row marginals are known, but the row-wise sampling rates are not necessarily proportional to the row marginals; i.e., the case where conditions (i) and (ii) of Section 4.2 hold true.

5. FURTHER REMARKS

The methods exemplified in this paper may be applied to other measures of association (or, more generally, to other sample measures), to other sampling procedures, to circumstances in which other kinds of outside knowledge exist, etc. One of our purposes has been to present these methods in a manner permitting their use by a wide class of research workers. (In some cases, different asymptotic methods may also be useful; for example, see Hoeffding [12].) In particular, one can obtain asymptotic approximations for the distributions of the traditional measures of association [8, Sec. 4] under other assumptions than that of independence. See [1], [27], [22], [21], [26], [15], [7]. Such approximations for the traditional measures have not been widely used. Perhaps one reason has been the almost obsessive interest in testing the null hypothesis of independence. Another possible reason is that the variance formulas obtained have been unwieldy and "too troublesome" to apply regularly in practice. See, for example, the editorial footnote on page 385 of [15].

We have not touched upon the question of power for the tests and confidence interval methods here discussed. The study of approximate power would be an appropriate next step.

6. REFERENCES

[1] Bulmer, M. G., "Confidence intervals for distance in the analysis of variance," *Biometrika*, 45 (1958), 360–9.

[2] Cramér, Harald, *Mathematical Methods of Statistics*. Princeton, New Jersey: Princeton University Press, 1946.

[3] Daniels, H. E., and Kendall, M. G., "The significance of rank correlations where parental correlation exists," *Biometrika*, 34 (1947), 198–208.

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

[5] El-Bradry, M. A., and Stephan, F. F., "On adjusting sample tabulations to census counts," *Journal of the American Statistical Association*, 50 (1955), 738–62.

[6] Friedlander, D., "A technique for estimating a contingency table, given the marginal totals and some supplementary data," *Journal of the Royal Statistical Society, Series A*, 124 (1961), 412–20.

[7] Goodman, Leo A., "On methods for comparing contingency tables," *Journal of the Royal Statistical Society, Series A*, 126 (1963), 94–108.

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

[9] Goodman, Leo A., and Kruskal, William H., "Measures of association for cross classifications. II: Further discussion and references," *Journal of the American Statistical Association*, 54 (1959), 123–63.

[10] Greenwood, Robert E., and Glasgow, Mark O., "Distribution of maximum and minimum frequencies in a sample drawn from a multinomial distribution," *Annals of Mathematical Statistics*, 21 (1950), 416–24.

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

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

[13] Hoeffding, Wassily, and Robbins, Herbert, "The central limit theorem for dependent random variables," *Duke Mathematical Journal*, 15 (1948), 773–80.

[14] Hoeffding, Wassily, "An upper bound for the variance of Kendall's 'tau' and of related statistics," pp. 258–64 in *Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling*, edited by Ingram Olkin and others, Stanford University Press, Stanford, 1960.

[15] Kondo, Tsutomu, "On the standard error of the mean square contingency," *Biometrika*, 21 (1929), 377–428.

[16] Kozelka, Robert M., "On some special order statistics from the multinomial distribution," Unpublished Ph.D. dissertation, Harvard University, 1952.

[17] Kozelka, Robert M., "Approximate upper percentage points for extreme values in multinomial sampling," *Annals of Mathematical Statistics*, 27 (1956), 507–12.

[18] Kruskal, William H., "Ordinal measures of association," *Journal of the American Statistical Association*, 53 (1958), 814–61.

[19] Mann, H. B., and Wald, A., "On stochastic limit and order relationships," *Annals of Mathematical Statistics*, 14 (1943), 217–26. (See also Wald, Abraham, *Selected Papers in Statistics and Probability*, McGraw-Hill Book Company, New York, 1955.)

[20] Noether, Gottfried E., "Two confidence intervals for the ratio of two probabilities and some measures of effectiveness," *Journal of the American Statistical Association*, 52 (1957), 36–45.

[21] Pearson, Karl, "On the probable error of a coefficient of mean square contingency," *Biometrika*, 10 (1915), 570–3.

[22] Sneyers, R., "Remarques à propos d'une généralisation de l'indice de similitude de M. Bouët," *Archiv für Meteorologie, Geophysik und Bioklimatologie, Series A: Meteorologie und Geophysik*, Band 11, 1. Heft, 1959, 126–35.

[23] Stuart, A., "The estimation and comparison of strengths of association in contingency tables," *Biometrika*, 40 (1953), 105–10.

[24] Sundrum, R. M., "Moments of the rank correlation coefficient τ in the general case," *Biometrika*, 40 (1953), 409–20.

[25] Weichselberger, K., "Über die Parameterschätzungen bei Kontingenztafeln, deren Randsummen vorgegeben sind," *Metrika*, 2 (1959), 100–30 and 198–229.

[26] Young, Andrew W., and Pearson, Karl, "On the probable error of a coefficient of contingency without approximation," *Biometrika*, 11 (1915), 215–30.

[27] Yule, G. Udny, "On the association of attributes in statistics with illustrations from the material from the childhood society, &c.," *Philosophical Transactions of the Royal Society of London, Series A*, 194 (1900), 257–319.

APPENDIX

A1. *Introduction.* The purpose of this appendix is to state some useful methods for deriving asymptotic distributions, to exemplify their use by outlining some of the derivations whose end-products are given in the preceding text, and to present other auxiliary material that seems inappropriate for inclusion in the text. An explicit statement of these methods may also be convenient for readers who wish to work out asymptotic distributions for other measures of association or for other sampling procedures than the ones discussed in this article.

A2. *A Basic Convergence Theorem.*

Theorem. If $\{X_n\}$ and $\{Y_n\}$ are two sequences of random variables, and if X is a random variable and y a constant, such that X_n converges in distribution to the distribution of X and Y_n converges in probability to y, then

X_n+Y_n converges in distribution to the distribution of $X+y$,
X_nY_n converges in distribution to the distribution of Xy,
X_n/Y_n converges in distribution to the distribution of X/y, (provided, in the last case, that $y\neq 0$).

(Convergence in distribution of X_n to X means that $\lim_{n\to\infty} \Pr\{X_n\leq x\} = \Pr\{X\leq x\}$ for every x at which $\Pr\{X\leq x\}$ is continuous. In the work of this paper the qualification may be neglected, for the limiting distributions are all normal, and hence continuous except for degenerate singular cases. For convergence in probability, see footnote 4.)

This result is stated, and an outline of its proof given, by Cramér [2, pp. 254–5]; it is a special case of earlier results by Mann and Wald (see [19], Theorem 5 and discussion on p. 224). The essential point is that if Z_n is a sequence of vector-valued random variables converging in distribution to the distribution of Z, and if f is a continuous vector-valued function (continuity may be weakened), then $f(Z_n)$ converges in distribution to the distribution of $f(Z)$.

This result has been applied in a number of places in the present text; e.g., in all places where a consistent estimator of the true asymptotic standard deviation of a statistic has been used, instead of the true asymptotic standard deviation itself, as the divisor of the difference between the statistic and its corresponding population value, without affecting the asymptotic distribution obtained.

A3. *The Delta Method Theorem.* We state this theorem for the case of two sequences of random variables, but its extension to more than two sequences is immediate. The proof of a more special form may be found in [2], pp. 366–7, and the proof of a somewhat more general form is in [13], pp. 777–8.

Theorem (Delta Method). Assume that W_n and V_n are two sequences of random variables ($n=1, 2, \cdots$) and that w and v are constants such that the pairs

$$\{\sqrt{n}(W_n - w), \sqrt{n}(V_n - v)\}$$

converge in distribution (bivariate sense) to the bivariate normal distribution with zero means, variances σ_{ww} and σ_{vv}, and covariance σ_{wv}.

Let $f(s, t)$ be a function with continuous first partial derivatives at (w, v).

Then $\sqrt{n}[f(W_n, V_n)-f(w, v)]$ is asymptotically normal, and in fact has the same asymptotic distribution as

$$\sqrt{n}\{a_w(W_n - w) + a_v(V_n - v)\},$$

where a_w, a_v are the partial derivatives of f with respect to its first and second arguments respectively, evaluated at (w, v). Thus the asymptotic distribution is the normal distribution with zero mean and variance

$$a_w^2\sigma_{ww} + 2a_wa_v\sigma_{wv} + a_v^2\sigma_{vv}.$$

Some of the results in this paper (e.g., the asymptotic variance (3.5.5)) were obtained with the aid of the delta method theorem. Some other results could be obtained either by applying this general theorem, or by more direct considerations (e.g., Section A5), since it is sometimes possible to arrange matters so that the function $f(s, t)$ is of the form a_ws+a_vt to begin with or so that a quantity simply related to $f(s, t)$ is of this linear form. We shall be frequently dealing with linear combinations of the $\sqrt{n}\ R_{ab}$'s, and we observe that the covariance structure of the $\sqrt{n}\ R_{ab}$'s is independent of n and is the same as the asymptotic covariance structure. Hence the variance of a linear combination of $\sqrt{n}\ R_{ab}$'s is exactly equal to its asymptotic variance.

A4. *Sample Maxima.* The fact that we may proceed, for asymptotic purposes, under the assumption that the R_{am}'s, $R_{\cdot m}$'s, etc., are taken on at the same columns and rows respectively as are the ρ_{am}'s, $\rho_{\cdot m}$'s, etc., is a consequence of the following.

Lemma. Let P_n ($n=1, 2, \cdots$) be a sequence of probability distributions, and let V be a chance event such that $\lim_{n\to\infty} P_n(V)$ exists; call it $P(V)$. Let W be a chance event such that $\lim_{n\to\infty} P_n(W)=1$. Then $\lim_{n\to\infty} P_n(V\cap W) =P(V)$. (The symbol "$\cap$" (read "and") means set-theoretic intersection.)

Proof: We note that

$$P_n(V) = P_n(V \cap W) + P_n(V \cap CW),$$

where CW is the complement of W. But $P_n(V\cap CW)\leq P_n(CW)$, which has the limit zero. This lemma is also a special case of the theorem of Section A2, obtained by replacing the probability of an event by the probability that its characteristic function is unity.

To apply this lemma, let P_n be the joint distribution of the random variables $\sqrt{n}(R_{ab}-\rho_{ab})$. Let W be the event defined by the requirement that the various sample maxima are taken on at the same columns and rows as the corresponding true maxima. V may be any chance event. For example, V might be the event that some function of the $\sqrt{n}(R_{ab}-\rho_{ab})$'s is $\leq$ a given constant. The hypotheses of the lemma are satisfied.

We are interested in the limit of $P_n(V)$. By applying the lemma, we see that this limit is equal to the limit of $P_n(V\cap W)$. This in turn is equal to the limit of $P_n(V'\cap W)$, where V' is like V except that R_{am} is replaced by that R_{ab} such that $\rho_{ab}=\rho_{am}$ (assuming uniqueness) and so on, since the event $V\cap W$ is exactly the same as the event $V'\cap W$. Finally, reapplying the lemma, the limit of $P_n(V'\cap W)$ is equal to the limit of $P_n(V')$, which is often easy to compute and which then gives us the desired limit of $P_n(V)$.

The following sections will illustrate applications of this lemma and the prior general theorems.

A5. *Asymptotic Behavior of* L_b. To examine the asymptotic behavior of L_b we write $\sqrt{n}(L_b-\lambda_b)$ as follows:

$$\begin{aligned}\sqrt{n}(L_b - \lambda_b) &= \sqrt{n}\left(\frac{\sum R_{am} - R_{\cdot m}}{1 - R_{\cdot m}} - \frac{\sum \rho_{am} - \rho_{\cdot m}}{1 - \rho_{\cdot m}}\right)\\ &= \sqrt{n}\,\frac{(\sum R_{am} - R_{\cdot m})(1 - \rho_{\cdot m}) - (\sum \rho_{am} - \rho_{\cdot m})(1 - R_{\cdot m})}{(1 - R_{\cdot m})(1 - \rho_{\cdot m})} \qquad \text{(A5.1)}\\ &= \sqrt{n}\,\frac{[\sum (R_{am} - \rho_{am})](1 - \rho_{\cdot m}) - (R_{\cdot m} - \rho_{\cdot m})(1 - \sum \rho_{am})}{(1 - \rho_{\cdot m})^2}\cdot\frac{1 - \rho_{\cdot m}}{1 - R_{\cdot m}}\,.\end{aligned}$$

The right-most factor, $(1-\rho_{\cdot m})/(1-R_{\cdot m})$, converges in probability to unity. Hence, if we can find the asymptotic distribution of what remains after omitting the right-most factor, we know, by Section A2, that it is the same as the asymptotic distribution *with* the right-most factor.

Since the ρ_{am}'s and $\rho_{\cdot m}$ are constants, our search then is effectively for the asymptotic distribution of $\sqrt{n}$ times a linear combination of the R_{am}'s and $R_{\cdot m}$; namely, that linear combination appearing in the numerator of the first fraction above. Let us call this numerator Δ. We already know, by the lemma presented in Section A4, that, for asymptotic purposes, R_{am} may be considered equal to R_{ab} for that value of b satisfying $\rho_{am}=\rho_{ab}$ and that similarly $R_{\cdot m}$ may be considered equal to $R_{\cdot b}$ for that value of b satisfying $\rho_{\cdot m}=\rho_{\cdot b}$.

Hence, by Section A3 and the facts just noted, it follows that $\sqrt{n}\Delta$ is asymptotically normal with mean zero and variance

$$\begin{aligned}&(1 - \rho_{\cdot m})^2(\textstyle\sum \rho_{am})(1 - \sum \rho_{am})\\ &\quad + (1 - \textstyle\sum \rho_{am})^2\rho_{\cdot m}(1 - \rho_{\cdot m}) \qquad \text{(A5.2)}\\ &\quad - 2(1 - \rho_{\cdot m})(1 - \textstyle\sum \rho_{am})[\sum^r \rho_{am} - \rho_{\cdot m} \sum \rho_{am}].\end{aligned}$$

For the reader's convenience, we insert here the argument leading to the above expression; similar arguments can be presented to obtain derivations of the asymptotic variances, presented in the text, for some of the other statistics discussed.

The first line of (A5.2) is the variance of the first term of $\sqrt{n}\Delta$; viz., $\sqrt{n}(1-\rho_{\cdot m})\sum(R_{am}-\rho_{am})$. The constant term, $(1-\rho_{\cdot m})$, is squared, and, since $\sum R_{am}$ and $1-\sum R_{am}$ may be considered as the proportions of successes and failures in n independent Bernouilli trials with constant probabilities $\sum\rho_{am}$ and $1-\sum\rho_{am}$, the variance of $\sqrt{n}\sum(R_{am}-\rho_{am})$ or of $\sqrt{n}\sum R_{am}$ is simply $(\sum\rho_{am})(1-\sum\rho_{am})$.

The second line of (A5.2) is the variance of the second term of $\sqrt{n}\Delta$; viz., $\sqrt{n}(1-\sum\rho_{am})(R_{\cdot m}-\rho_{\cdot m})$. It may be written down in the same manner as the first line.

The third line of (A5.2) is minus twice the covariance of the two terms of $\sqrt{n}\Delta$. The constants afford no difficulty, but the quantity in square brackets, the covariance of $\sqrt{n}\sum R_{am}$ and $\sqrt{n}R_{\cdot m}$, may be troublesome to check. Note that, in general, if $R_1, R_2, \cdots, R_k$ are multinomial proportions corresponding to the probabilities $\rho_1, \rho_2, \cdots, \rho_k$, then

$$\begin{aligned}\operatorname{Cov}(R_1+R_2, R_1+R_3) &= \operatorname{Var} R_1 + \operatorname{Cov}(R_1, R_2) + \operatorname{Cov}(R_1, R_3) + \operatorname{Cov}(R_2, R_3)\\ &= n^{-1}\{\rho_1(1-\rho_1) - \rho_1\rho_2 - \rho_1\rho_3 - \rho_2\rho_3\}\\ &= n^{-1}\{\rho_1 - (\rho_1+\rho_2)(\rho_1+\rho_3)\}.\end{aligned}$$

To apply this, take R_1 as the sum of those R_{am}'s that also appear as summands of $R_{\cdot m}$, R_2 as $\sum R_{am}-R_1$, and R_3 as $R_{\cdot m}-R_1$. We can therefore write down the third line above immediately.

Simplifying the above expression for the variance of the asymptotic distribution of $\sqrt{n}\Delta$, we obtain

$$(1-\rho_{\cdot m})(1-\sum\rho_{am})(\sum\rho_{am}+\rho_{\cdot m}-2\sum{}^{r}\rho_{am}). \tag{A5.3}$$

Note that the last factor is just the sum of the ρ_{am}'s not summands in $\rho_{\cdot m}$, plus those summands in $\rho_{\cdot m}$ that are not ρ_{am}'s. (A5.3) is zero if and only if $\lambda_b=0$.

Hence, as $n\to\infty$, the distribution of $\sqrt{n}(L_b-\lambda_b)$ will approach the normal distribution with mean zero and variance (3.1.3). This variance is zero if and only if $\lambda_b=0$ or 1. It is indeterminate if $\rho_{\cdot m}=1$, but in this case λ_b itself is indeterminate. Finally, if $\lambda_b\neq 0$ or 1, it follows from Section A2 that (3.1.4) is asymptotically unit normal.

We note that (3.1.3) could also be obtained by direct application of the delta method to the first line of (A5.1). We have that

$$\frac{\partial}{\partial R_{\cdot m}}\left(\frac{\sum R_{am}-R_{\cdot m}}{1-R_{\cdot m}}\right) = -\frac{1-\sum R_{am}}{(1-R_{\cdot m})^2},$$

and

$$\frac{\partial}{\partial\sum R_{am}}\left(\frac{\sum R_{am}-R_{\cdot m}}{1-R_{\cdot m}}\right) = \frac{1}{1-R_{\cdot m}}.$$

Evaluating these derivatives for the population values, we obtain $-(1-\sum\rho_{am})/(1-\rho_{\cdot m})^2$ and $1/(1-\rho_{\cdot m})$, respectively. If we then computed the quadratic form required by the delta method, and used the appropriate variances and covariances given a few paragraphs back, we would emerge with (3.1.3). These two paths to the same goal serve as mutual checks.

A6. *Asymptotic Behavior of L.* Just as in Section A5, we write $\sqrt{n}(L-\lambda)$ as one fraction and change the denominator to a function of the ρ's while at the same time multiplying by a quantity approaching unity in probability. Neglecting this last quantity, as we may by Section A2, we deal with

$$\begin{aligned}\frac{\sqrt{n}}{(2-\rho_{\cdot m}-\rho_{m\cdot})^2}\{&(2-\rho_{\cdot m}-\rho_{m\cdot})(\sum R_{am}+\sum R_{mb})\\ &+(\sum \rho_{am}+\sum \rho_{mb}-2)(R_{\cdot m}+R_{m\cdot})+2(\rho_{\cdot m}+\rho_{m\cdot}-\sum \rho_{am}-\sum \rho_{mb})\}.\end{aligned} \tag{A6.1}$$

This quantity is asymptotically normal with zero mean and with a variance that is equal to the following divided by $(2-\rho_{\cdot m}-\rho_{m\cdot})^4$:

$$\begin{aligned}&(2-\rho_{\cdot m}-\rho_{m\cdot})^2[\sum \rho_{am}(1-\sum \rho_{am})+\sum \rho_{mb}(1-\sum \rho_{mb})\\ &\qquad\qquad+2\sum{}^{*}\rho_{am}-2(\sum \rho_{am})(\sum \rho_{mb})]\\ &+(2-\sum \rho_{am}-\sum \rho_{mb})^2[\rho_{\cdot m}(1-\rho_{\cdot m})+\rho_{m\cdot}(1-\rho_{m\cdot})+2\rho_{**}-2\rho_{m\cdot}\rho_{\cdot m}]\\ &-2(2-\rho_{\cdot m}-\rho_{m\cdot})(2-\sum \rho_{am}-\sum \rho_{mb})[\sum{}^{r}\rho_{am}-\rho_{\cdot m}\sum \rho_{am}+\sum{}^{c}\rho_{mb}\\ &\qquad-\rho_{m\cdot}\sum \rho_{mb}+\rho_{*m}-\rho_{m\cdot}\sum \rho_{am}+\rho_{m*}-\rho_{\cdot m}\sum \rho_{mb}],\end{aligned} \tag{A6.2}$$

where the notation involving asterisks is defined in Section 3.4 in terms of sample R's. Simplifying (A6.2) we obtain the quantity (A6.3),[17]

$$\begin{aligned}&(2-\rho_{\cdot m}-\rho_{m\cdot})^2[(\sum \rho_{am}+\sum \rho_{mb})(1-\sum \rho_{am}-\sum \rho_{mb})+2\sum{}^{*}2\rho_{am}]\\ &+(2-\sum \rho_{am}-\sum \rho_{mb})^2[(\rho_{\cdot m}+\rho_{m\cdot})(1-\rho_{\cdot m}-\rho_{m\cdot})+2\rho_{**}]\\ &-2(2-\rho_{\cdot m}-\rho_{m\cdot})(2-\sum \rho_{am}-\sum \rho_{mb})[\sum{}^{r}\rho_{am}+\sum{}^{c}\rho_{mb}+\rho_{*m}\\ &\qquad+\rho_{m*}-(\rho_{\cdot m}+\rho_{m\cdot})(\sum \rho_{am}+\sum \rho_{mb})].\end{aligned} \tag{A6.3}$$

If, for simplicity, we let

$$\begin{aligned}\Upsilon_{\cdot}&=\rho_{\cdot m}+\rho_{m\cdot},\\ \Upsilon_{\Sigma}&=\sum \rho_{am}+\sum \rho_{mb},\\ \Upsilon_{*}&=\sum{}^{r}\rho_{am}+\sum{}^{c}\rho_{mb}+\rho_{*m}+\rho_{m*},\end{aligned}$$

then (A6.3) is equal to both the following quantities:

$$\begin{aligned}&(2-\Upsilon_{\cdot})^2(\Upsilon_{\Sigma}+2\sum{}^{*}\rho_{am})+(2-\Upsilon_{\Sigma})^2(\Upsilon_{\cdot}+2\rho_{**})-4(\Upsilon_{\Sigma}-\Upsilon_{\cdot})^2\\ &\qquad-2(2-\Upsilon_{\cdot})(2-\Upsilon_{\Sigma})\Upsilon_{*}\end{aligned} \tag{A6.4a}$$

$$\begin{aligned}&(2-\Upsilon_{\cdot})(2-\Upsilon_{\Sigma})(\Upsilon_{\cdot}+\Upsilon_{\Sigma}+4-2\Upsilon_{*})-2(2-\Upsilon_{\cdot})^2(1-\sum{}^{*}\rho_{am})\\ &\qquad-2(2-\Upsilon_{\Sigma})^2(1-\rho_{**}).\end{aligned} \tag{A6.4b}$$

[17] Actually we could write (A6.3) directly by the use of general formulas such as that of Section A5. For example it is easy to check that

$$\mathrm{Var}[(R_1+R_2)+(R_1+R_3)]=n^{-1}[(2\rho_1+\rho_2+\rho_3)(1-2\rho_1-\rho_2-\rho_3)].$$

The variance of the asymptotic normal distribution for $\sqrt{n}(L-\lambda)$ is either of the above divided by $(2-\Upsilon.)^4$.

It follows from Section A2 that, provided (A6.4) is not zero, (3.4.3) is asymptotically unit normal.

Finally, we show that (A6.4) is zero if and only if $\lambda=0$ or 1. We are concerned with the variance of

$$(2 - \Upsilon.)U_\Sigma + (\Upsilon_\Sigma - 2)U., \tag{A6.5}$$

the random quantity in (A6.1). To say that this has zero variance is to say that it is constant; we consider the various possible cases.

(A): $2-\Upsilon.=0$. This says that $\rho_{\cdot m}=\rho_{m\cdot}=1$, or that there is only one non-zero ρ_{ab}, and it equals unity. This degenerate case, in which λ is not even defined, we have precluded by assumption.

(B): $2-\Upsilon_\Sigma=0$. This says that $\sum\rho_{am}=\sum\rho_{mb}=1$, or that each row and column has at most one non-zero ρ_{ab}. Then $\lambda=1$, $L=1$ always, and $U_\Sigma=2$ always. Thus, in this special case, $\lambda=1$ and the variance of (A6.5) is zero.

(C): $2-\Upsilon.\neq0$, $2-\Upsilon_\Sigma\neq0$. If neither coefficient of (A6.5) is zero, at least one row or column has two or more cells with non-zero ρ_{ab}'s. Without loss of generality, suppose that ρ_{11} and ρ_{12} are both non-zero. Then there is positive probability for each of the following $n+1$ sample points: $R_{11}=k/n$, $R_{12}=(n-k)/n$ $(k=0, 1, \cdots, n)$. If $k\leq n/2$, $U_\Sigma=U.=1+[(n-k)/n]$; while if $k\geq n/2$, $U_\Sigma=U.=1+[k/n]$. Hence, for such sample points,

$$\text{(A6.5)} = (\Upsilon_\Sigma - \Upsilon.)[1 + \text{Max}\,(k, n - k)/n], \qquad (k = 0, 1, \cdots, n).$$

If (A6.5) is constant and $n\geq2$, it follows that $\Upsilon_\Sigma=\Upsilon.$, or $\lambda=0$.

Conversely, it is easy to see that if $\lambda=0$, all the ρ_{am}'s must appear in the same column, and all the ρ_{mb}'s in the same row. Hence $\Upsilon.=\Upsilon_\Sigma$, $\Upsilon_*=\Upsilon.+2\rho_{**}$, and $\Sigma^*\rho_{am}=\rho_{**}=\rho_{*m}=\rho_{m*}$. Hence, substituting in (A6.4), we obtain zero variance.

A7. *Asymptotic Behavior of G.* As before, we simplify

$$\sqrt{n}(G - \gamma) = \sqrt{n}\left[\frac{P_s - P_d}{1 - P_t} - \frac{\Pi_s - \Pi_d}{1 - \Pi_t}\right]$$

by writing it as a single fraction, and then replacing the denominator by the quantity to which it converges in probability. This leaves us with

$$\frac{2}{(1 - \Pi_t)^2}\sqrt{n}[P_s\Pi_d - P_d\Pi_s] \tag{A7.1}$$

as the quantity whose asymptotic distribution is desired. We assume that $\Pi_t<1$. By use of the delta method (Section A3), we see that $\sqrt{n}[P_s\Pi_d-P_d\Pi_s]$ is asymptotically normal with zero mean and with a variance computed in the following manner. The random variable P_s, considered as a function of the R_{ab}'s, may be partially differentiated with respect to R_{ab} as follows:

$$\frac{\partial P_s}{\partial R_{\mathbf{ab}}} = 2 \sum_a \sum_b \sum_{a'>a} \sum_{b'>b} \frac{\partial}{\partial R_{\mathbf{ab}}} R_{ab} R_{a'b'}.$$

Unless $(a, b) = (\mathbf{a}, \mathbf{b})$ or $(a', b') = (\mathbf{a}, \mathbf{b})$, the summand is zero. (Both equalities cannot hold simultaneously.) Hence, we obtain

$$\frac{\partial P_s}{\partial R_{\mathbf{ab}}} = 2 \sum_{a'>\mathbf{a}} \sum_{b'>\mathbf{b}} R_{a'b'} + 2 \sum_{a'<\mathbf{a}} \sum_{b'<\mathbf{b}} R_{a'b'}. \tag{A7.2a}$$

Similarly,

$$\frac{\partial P_d}{\partial R_{\mathbf{ab}}} = 2 \sum_{a'>\mathbf{a}} \sum_{b'<\mathbf{b}} R_{a'b'} + 2 \sum_{a'<\mathbf{a}} \sum_{b'>\mathbf{b}} R_{a'b'}. \tag{A7.2b}$$

Evaluating these at $R_{a'b'} = \rho_{a'b'}$, and changing $(\mathbf{a}, \mathbf{b})$ to (a, b) we obtain the differential coefficients

$$\begin{aligned} 2\mathcal{R}_{ab}^{(s)} &= \left[\frac{\partial P_s}{\partial R_{ab}}\right]_{R_{a'b'}=\rho_{a'b'}} = 2 \sum_{a'>a} \sum_{b'>b} \rho_{a'b'} + 2 \sum_{a'<a} \sum_{b'<b} \rho_{a'b'}, \\ 2\mathcal{R}_{ab}^{(d)} &= \left[\frac{\partial P_d}{\partial R_{ab}}\right]_{R_{a'b'}=\rho_{a'b'}} = 2 \sum_{a'>a} \sum_{b'<b} \rho_{a'b'} + 2 \sum_{a'<a} \sum_{b'>b} \rho_{a'b'}. \end{aligned} \tag{A7.3}$$

Essentially these quantities appear in (3.5.6).

Hence the derivative of $P_s\Pi_d - P_d\Pi_s$ with respect to R_{ab}, evaluated for the true $\rho_{a'b'}$'s, is

$$2[\Pi_d \mathcal{R}_{ab}^{(s)} - \Pi_s \mathcal{R}_{ab}^{(d)}],$$

and the asymptotic variance of $\sqrt{n}\,[P_s\Pi_d - P_d\Pi_s]$ is the same as the asymptotic variance of

$$\sum_a \sum_b 2[\Pi_d \mathcal{R}_{ab}^{(s)} - \Pi_s \mathcal{R}_{ab}^{(d)}]\sqrt{n}(R_{ab} - \rho_{ab}). \tag{A7.4}$$

But this quantity is

$$4 \sum_a \sum_b \sum_{a'} \sum_{b'} [\Pi_d \mathcal{R}_{ab}^{(s)} - \Pi_s \mathcal{R}_{ab}^{(d)}][\Pi_d \mathcal{R}_{a'b'}^{(s)} - \Pi_s \mathcal{R}_{a'b'}^{(d)}][\delta_{ab,a'b'}\rho_{ab} - \rho_{ab}\rho_{a'b'}], \tag{A7.5}$$

where $\delta_{ab,a'b'} = 0$ unless $(a, b) = (a', b')$, in which case it is unity. Now consider the eight terms of the above sum, obtained by multiplying out. The four terms that involve $\delta_{ab,a'b'}$ are

$$4 \sum_a \sum_b \Pi_d^2 \mathcal{R}_{ab}^{(s)^2} \rho_{ab} = 4\Pi_d^2 \Pi_{ss},$$

$$-4 \sum_a \sum_b \Pi_s \Pi_d \mathcal{R}_{ab}^{(s)} \mathcal{R}_{ab}^{(d)} \rho_{ab} = -4\,\Pi_s\Pi_d\Pi_{sd} \text{ (two of these), and}$$

$$4 \sum_a \sum_b \Pi_s^2 \mathcal{R}_{ab}^{(d)^2} \rho_{ab} = 4\Pi_s^2 \Pi_{dd};$$

whereas the other terms sum to zero, for they are equal to

$$4\left\{\sum_a \sum_b [\Pi_d \mathcal{R}_{ab}^{(s)} - \Pi_s \mathcal{R}_{ab}^{(d)}]\rho_{ab}\right\}^2$$

and

$$\Pi_d \sum_a \sum_b \rho_{ab} \mathcal{R}_{ab}^{(s)} = \Pi_d \Pi_s = \Pi_s \sum_a \sum_b \rho_{ab} \mathcal{R}_{ab}^{(d)}.$$

Hence the asymptotic variance of $\sqrt{n}[P_s \Pi_d - P_d \Pi_s]$ is

$$4[\Pi_d^2 \Pi_{ss} - 2\Pi_s \Pi_d \Pi_{sd} + \Pi_s^2 \Pi_{dd}].$$

Finally, we must multiply this by the square of the constant coefficient in (A7.1) to obtain (3.5.5), the asymptotic variance of $\sqrt{n}(G-\gamma)$.

It is interesting to note that (3.5.5) reduces, in the continuous case, to

$$16\{\Pi_s^2 \Pi_d - \Pi_s^2 \Pi_{sd} - 2\Pi_s \Pi_d \Pi_{sd} + \Pi_d^2 \Pi_s - \Pi_d^2 \Pi_{sd}\} = 16\{\Pi_{ss} - \Pi_s^2\}, \quad \text{(A7.6)}$$

since, in this case, $\Pi_t = 0$, $\Pi_s = \Pi_{ss} + \Pi_{sd}$, $\Pi_d = \Pi_{dd} + \Pi_{sd}$, $\Pi_s + \Pi_d = 1$. Formula (A7.6) is, in fact, the asymptotic variance for $\sqrt{n}(t-\tau)$, where t is Kendall's t (see [11], [24], [18]). This is natural, since G, in the continuous case, is [18] essentially Kendall's t.

We now turn to the bound on the asymptotic variance stated in Section 3.5. This quantity is also a bound on the actual (small-sample) variance of $\sqrt{n}$ $(G-\gamma)$, at least for even-sized samples [4]. We shall prove here that (3.5.5) is less than or equal to the bound (3.5.9), $2(1-\gamma^2)/(1-\Pi_t)$; i.e., that

$$\{\Pi_s^2 \Pi_{dd} - 2\Pi_s \Pi_d \Pi_{sd} + \Pi_d^2 \Pi_{ss}\} \leq (1 - \Pi_t)^3 (1 - \gamma^2)/8. \quad \text{(A7.7)}$$

Proof: Let

$$V_{ij} = \begin{cases} -\Pi_d & \text{if observations } i \text{ and } j \text{ are fully concordant.} \\ \Pi_s & \text{if observations } i \text{ and } j \text{ are fully discordant.} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$E\{V_{ij}\} = -\Pi_d \Pi_s + \Pi_s \Pi_d = 0,$$

$$\text{Var}\{V_{ij}\} = \Pi_d^2 \Pi_s + \Pi_s^2 \Pi_d = \Pi_s \Pi_d (\Pi_d + \Pi_s) = \Pi_s \Pi_d (1 - \Pi_t),$$

$$\text{Cov}\{V_{ij}, V_{ik}\} = \Pi_d^2 \Pi_{ss} - 2\Pi_s \Pi_d \Pi_{sd} + \Pi_s^2 \Pi_{dd},$$

for $j \neq k$. Since

$$\begin{aligned} &\text{Var}\{2(V_{12} + V_{34}) - (V_{13} + V_{14} + V_{23} + V_{24})\} \\ &\quad = 12 \, \text{Var}\{V_{12}\} - 24 \, \text{Cov}\{V_{12}, V_{13}\} \geq 0, \end{aligned}$$

we have that

$$\Pi_s\Pi_d(1 - \Pi_t) \geq 2\{\Pi_d^2\Pi_{ss} - 2\Pi_s\Pi_d\Pi_{sd} + \Pi_s^2\Pi_{dd}\}.$$

We also note that

$$1 - \gamma^2 = 1 - [(\Pi_s - \Pi_d)/(\Pi_s + \Pi_d)]^2 = 4\Pi_s\Pi_d/(1 - \Pi_t)^2.$$

Thus,

$$\Pi_d^2\Pi_{ss} - 2\Pi_s\Pi_d\Pi_{sd} + \Pi_d^2\Pi_{dd} \leq (1 - \gamma^2)(1 - \Pi_t)^3/8.$$

We conclude with some remarks on the meaning of the assumption that (3.5.5), the asymptotic variance, is not zero. It is clear from (A7.5) and the following lines that (3.5.5) = 0 if and only if

$$\Pi_d \mathcal{R}_{ab}^{(s)} - \Pi_s \mathcal{R}_{ab}^{(d)} = 0$$

for all (a, b) such that $\rho_{ab} \neq 0$. Multiplying through by $\rho_{ab}\mathcal{R}_{ab}^{(s)}$ or $\rho_{ab}\mathcal{R}_{ab}^{(d)}$ and summing, we find that, if (3.5.5) = 0, then

$$\Pi_d\Pi_{ss} = \Pi_s\Pi_{sd}, \qquad \Pi_s\Pi_{dd} = \Pi_d\Pi_{sd}.$$

The converse is immediate. We may write these two statements as

$$\Pi_{ss}/\Pi_s = \Pi_{sd}/\Pi_d, \quad \text{and} \quad \Pi_{dd}/\Pi_d = \Pi_{sd}/\Pi_s,$$

providing that Π_s and Π_d are different from zero. From these we see that (3.5.5) is zero if and only if, taking three individuals, 1, 2, 3, from the population at random,

$$\Pr\left\{\begin{array}{c}\text{1 and 3 have "positive"}\\ \text{sign relation}\end{array}\middle|\begin{array}{c}\text{1 and 2 have "positive"}\\ \text{sign relation}\end{array}\right\}$$
$$= \Pr\left\{\begin{array}{c}\text{1 and 3 have "positive"}\\ \text{sign relation}\end{array}\middle|\begin{array}{c}\text{1 and 2 have "negative"}\\ \text{sign relation}\end{array}\right\}$$

and

$$\Pr\left\{\begin{array}{c}\text{1 and 3 have "negative"}\\ \text{sign relation}\end{array}\middle|\begin{array}{c}\text{1 and 2 have "negative"}\\ \text{sign relation}\end{array}\right\}$$
$$= \Pr\left\{\begin{array}{c}\text{1 and 3 have "negative"}\\ \text{sign relation}\end{array}\middle|\begin{array}{c}\text{1 and 2 have "positive"}\\ \text{sign relation}\end{array}\right\}.$$

We suggest that this is an unlikely state of affairs in most applications. For example, if the four corner cells of the cross classification have positive probabilities ($\rho_{11}, \rho_{1\beta}, \rho_{\alpha 1}, \rho_{\alpha\beta} > 0$), then (3.5.5) must be positive.