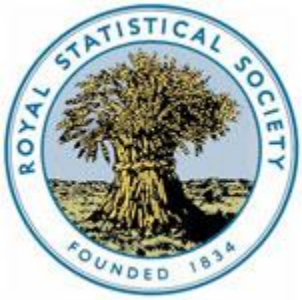

A New Index of Fit Based on Mixture Methods for the Analysis of Contingency Tables
Author(s): Tamas Rudas, Clifford C. Clogg and Bruce G. Lindsay
Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 56, No. 4 (1994), pp. 623-639
Published by: Wiley for the Royal Statistical Society
Stable URL: http://www.jstor.org/stable/2346187
Accessed: 23-03-2017 09:44 UTC

J. R. Statist. Soc. B (1994)
56, *No.* 4, *pp.* 623–639

A New Index of Fit Based on Mixture Methods for the Analysis of Contingency Tables

By TAMÁS RUDAS†, CLIFFORD C. CLOGG and BRUCE G. LINDSAY

Eötvös University, Budapest, Hungary *Pennsylvania State University, University Park, USA*

[Received July 1991. Final revision June 1993]

SUMMARY

A framework based on mixture methods is proposed for evaluating goodness of fit in the analysis of contingency tables. For a given model H applied to a contingency table $\mathbf{P}$, we consider the two-point mixture $\mathbf{P} = (1-\pi)\mathbf{\Pi}_1 + \pi\mathbf{\Pi}_2$, with π the mixing proportion $(0 \leqslant \pi \leqslant 1)$ and $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ the tables of probabilities for each latent class or component. In the *unstructured* approach recommended here, the mixture model applies H to $\mathbf{\Pi}_1$ but does not impose any restrictions on $\mathbf{\Pi}_2$. A contingency table $\mathbf{P}$ can generally be represented as such a two-point mixture for an interval of π-values. We define our index of lack of fit, π^*, to be the smallest such π, i.e. π^* is the fraction of the population that cannot be described by model H. This approach can be contrasted with the *structured* approach that applies model H to *both* $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ and leads to conventional latent class models when H is the hypothesis of independence. The case where H is the hypothesis of row–column independence and $\mathbf{P}$ is a two-way contingency table is covered in detail, but the procedure is quite general.

Keywords: CONTINGENCY TABLE; EM ALGORITHM; GOODNESS OF FIT; LATENT CLASSES; MISCLASSIFICATION

1. INTRODUCTION

The problem of evaluating goodness of fit in the analysis of contingency tables is considered from a new viewpoint by using mixture methods and simple modifications of latent class analysis. The techniques proposed can be used as a supplement to the conventional analysis based on χ^2-statistics or quantities derived from them. The approach focuses attention on the substantive importance of the discrepancy between the model and the data, allows comparisons across samples or studies and permits an evaluation of the model that downplays the role of sample size. The approach is attractive in the common situation where the sample size is very large, but it can also be applied to the situation where n is not large.

Throughout we suppose that a relatively simple model H is considered and that the goal is to evaluate how well this particular model describes the data. H might correspond to some standard null model, like independence or quasi-independence, or to some non-null model having simple structure. Examples of the latter include log-linear models excluding at least some higher order interactions if multidimensional tables are considered (Agresti, 1990), or association models for two-way tables (Goodman, 1984, 1985) or for multiway tables (Becker and Clogg, 1989),

†*Address for correspondence*: Statistics Group, Institute of Sociology, Eötvös University, Pollack M. ter 10, H-1088 Budapest, Hungary.
E-mail: H163RUD@ELLA.HU

 0035–9246/94/56623

depending on the context. We give a new index called π^*, the mixing weight from a special *unstructured* two-point mixture, that represents the fraction of the population intrinsically outside model H. A method for calculating the maximum likelihood estimator of π^* is given along with interval estimators derived from the relevant profile likelihood. Although we restrict attention to the analysis of contingency tables here, in Section 9 we note that this index of fit derived from mixture concepts has a wide range of potential applications in other settings.

The method is closely related to standard procedures for latent class analysis (Lazarsfeld and Henry, 1968; Goodman, 1974; Clogg and Goodman, 1984; Clogg, 1988; Haberman, 1979). The relationship to the conventional latent class model for two-way contingency tables in Clogg (1981) or Goodman (1987) will be brought out in the examples; our proposed method is contrasted with the conventional latent class model (or the *structured* approach) in Section 8.

For a contingency table with N cells, denote the cell probabilities as $\mathbf{P} \equiv \{P_h, h=1, \ldots, N\}$. Multinomial sampling is assumed with expected frequencies $\mathbf{F} \equiv \{F_h, h=1, \ldots, N\}$, with $F_h = nP_h$ for the hth cell. Observed frequencies are denoted as $\mathbf{f}$, or f_h for the cell frequency, with $\Sigma_h f_h = n$, the sample size. The conventional approach for assessing goodness of fit is covered in several standard sources (see, for example, Bishop *et al.* (1975) and Agresti (1990)). First, the maximum likelihood estimate of $\mathbf{P}$ (or $\mathbf{F}$) is obtained. Second, the agreement between the model and the data is assessed with conventional goodness-of-fit statistics, such as the Pearson statistic, the likelihood ratio statistic or other members of the power divergence family (Read and Cressie, 1988). Third, in cases where the fit statistics provide evidence for lack of fit, residual analysis might be used. Residual analysis usually entails examination of cell-by-cell components of χ^2-statistics or quantities closely related to them. Other commonly used procedures for model evaluation include collapsing categories and refitting the model, comparing measures of association for H and competing models, and using indexes based on prediction criteria or information criteria such as the Akaike information criterion (see Atkinson (1981)). Many of the available procedures are reviewed in an applied context by Fowlkes *et al.* (1988). To our knowledge, indexes of 'fit' based on prediction or information criteria extract a penalty for sample size but do not remove the effect of sample size as such, and it is difficult to interpret the magnitudes of those indexes in substantive ways.

Conventional methods for evaluating contingency table models rely on χ^2-statistics or quantities derived from them. When the sample size n is not large, the usual asymptotic theory justifying χ^2-approximations might not be appropriate; see, for example, Rudas (1986) and Read and Cressie (1988). Alternatives to the usual goodness-of-fit statistics are important to consider in these situations.

Assessing goodness of fit in cases where the sample size is very large presents a different problem. In this case the asymptotic justification for χ^2-statistics is secure, but different fit statistics (Neyman, Pearson, likelihood ratio, etc.) can give different impressions about the magnitude of the discrepancy between the model and the data when the model is not true. The model will usually be 'rejected' in favour of some more complicated model if the sample is sufficiently large. Goodness-of-fit statistics are usually not informative when the sample size is very large. We begin with two examples that illustrate the need for alternatives.

TABLE 1
Cross-classification of eye colour and hair colour†

Eye colour	*Hair colour*			
	Black	*Brunette*	*Red*	*Blonde*
Brown	68	119	26	7
Blue	20	84	17	94
Hazel	15	54	14	10
Green	5	29	14	16

†$n = 592$; source, Snee (1974) and Diaconis and Efron (1985).

2. TWO EXAMPLES

Diaconis and Efron (1985) provided a volume test interpretation of the Pearson statistic (for the independence hypothesis applied to a two-way table) that pertains to model evaluation in large or very large samples. Table 1, considered earlier by Snee (1974), is a 4×4 table cross-classifying eye colour and hair colour. The sample size $n = 592$ which is certainly large, although perhaps not very large. The Pearson statistic for the independence model is $X^2 = 138.290$ on 9 degrees of freedom, and the likelihood ratio statistic is $L^2 = 146.444$; the index of dissimilarity is $D = 0.184$. The model would be rejected on the basis of these quantities. (The statistic $D = (\Sigma_h |f_h - \hat{F}_h|)/2n$ is often used to remove the sample size effect, especially for data in the social sciences. A D-value close to 0 indicates a good fit, but the upper bound of D is less than 1 and varies with the model being considered.)

Diaconis and Efron (1985) found that, among all 4×4 tables with $n = 592$ (no margins fixed), approximately 10% have X^2-values less than 138.29 when sampling uniformly from the relevant universe. They concluded that the given 4×4 table does not lie particularly close to independence, confirming the inference drawn from the conventional goodness-of-fit statistics.

Table 2 was originally published in Cramer (1946). This 5×4 table cross-classifies number of children by (grouped) annual income levels. The sample size $n = 25263$, which is very large. The fit statistics are $X^2 = 568.566$ and $L^2 = 569.420$ on 12 degrees of freedom, and $D = 0.056$. Both χ^2-statistics obviously lead to rejection using conventional criteria. However, the D-value is relatively small indicating that

TABLE 2
Cross-classification of number of children by annual income†

No. of children	*Annual income*			
	0–1	*1–2*	*2–3*	*3+*
0	2161	3577	2184	1636
1	2755	5081	2222	1052
2	936	1753	640	306
3	225	419	96	38
4+	39	98	31	14

†$n = 25263$; source, Cramer (1946) and Diaconis and Efron (1985).

only about 6% of the observations would have to be reallocated to match the predictions of the model. Diaconis and Efron (1985) found that, among all 5×4 tables with $n = 25263$ (no margins fixed), the proportion of those with $X^2 < 568.576$ is 2.1×10^{-7}. They concluded that the observed table is close to independence, which is quite at odds with the conclusion drawn from the χ^2-values.

The qualitative inferences obtained from our method lead to conclusions that are consistent with those reached by Diaconis and Efron even though we proceed from an entirely different starting point. Our approach gives a simple index of fit—the mixing weight π^* defined in Section 4 as the fraction of the population outside the model—that allows comparisons across samples or across models. Note that the Diaconis–Efron method would require modification if some other statistic besides the Pearson statistic were used, if some hypothesis other than independence were considered or if the table analysed pertained to more than two variables and some hypothesis besides independence or conditional independence were considered. Our approach can be applied to arbitrary contingency tables (not just two-way tables), and it can be applied to examine virtually any model (not just the independence model).

3. FINITE MIXTURE APPROACH AND LATENT CLASS ANALYSIS: GENERAL CONSIDERATIONS

The family of models that we propose for evaluation purposes is

$$P_h = (1-\pi)\Pi_{1h} + \pi\Pi_{2h}; \qquad \boldsymbol{\Pi}_1 \in H,\ \boldsymbol{\Pi}_2 \text{ unspecified, for } h = 1, \ldots, N, \quad (1)$$

with π denoting the mixing weight. π is the fraction of the population outside model H and $1-\pi$ is therefore the fraction of the population described by H. For each π, let H_π denote the model described by equation (1) with π fixed, with H_0 therefore equal to H and H_1 the completely unrestricted model.

This model is a two-point mixture. The mixing distribution can be represented by a dichotomous (Bernoulli) variable, say X, with $P(X=1) = 1-\pi$ and $P(X=2) = \pi$. Let Y_h be the indicator variable for cell h, i.e. $Y_h = 1$ if cell h is observed and $Y_h = 0$ otherwise. The quantity Π_{th} refers to the conditional probability of cell h given that the observation is drawn from latent class t, $t = 1, 2$ and $h = 1, \ldots, N$, i.e. $\Pi_{th} = P(Y_h = 1 \mid X = t)$. The mixing proportion π can take on any value in the interval $0 \leqslant \pi \leqslant 1$, but we suppose that π is relatively small. When $\pi = 0$ the original model is obtained, so model (1) is a generalization of the model originally specified for $\mathbf{P}$.

If the model H had been assumed for both latent classes (i.e. $\boldsymbol{\Pi}_1 \in H$, $\boldsymbol{\Pi}_2 \in H$), then model (1) is a two-class latent structure model. If H is the model of independence between (among) factors, the usual latent class model is obtained (Lazarsfeld and Henry, 1968; Goodman, 1974). However, H need not be the model of independence; other models can be used. The lack of restrictions on $\boldsymbol{\Pi}_2$ ensures that the model for $\mathbf{P}$ is a two-point mixture that generalizes both H and the usual two-class latent structure in these other cases.

4. 'SATURATED' TWO-POINT MIXTURE WITH IDENTIFIED MIXING WEIGHT π^*

Let us consider for the moment how researchers ordinarily proceed in assessing the evidence for or against a particular model in practice. If the model (H) is

rejected, then other models that are more comprehensive are considered, e.g. a set of models $H_1, \ldots, H_v$, each of which contains H as a special case. Of course, if the sample is sufficiently large, this procedure will often lead to a saturated model, say H_S, with the property that $\hat{F}_h = f_h$ for all h. For the class of hierarchical log-linear models, for example, H might be the model fitting all two-factor marginals (i.e. including all two-factor interactions), and $H_1, \ldots, H_v$ might be models including higher order interactions in succession. If the sample size is sufficiently large, in many or most cases the researcher will conclude that only model H_S is congruent with the data. How can we summarize the magnitude of interactions left out in the model originally specified for the data? This question is related to but different from the analysis of residuals; residuals under model H are normally assessed assuming that the model holds true at least approximately.

We now consider an approach based on the two-point mixture of model (1) that can be tied to the objectives implicit in standard practice, i.e. we wish to measure the degree of congruency of model H with the data to summarize the importance of factors, like higher order interactions, that are left out of H.

First note that as π varies the class of models H_π varies from H_0, the restricted model, to H_1, the completely unrestricted model. Moreover, the class of models grows monotonically, because $H_\pi \subset H_{\pi'}$ for $\pi < \pi'$. This follows by an elementary argument. Rewrite $\mathbf{P} = (1-\pi)\mathbf{\Pi}_1 + \pi\mathbf{\Pi}_2$ as $\mathbf{P} = (1-\pi')\mathbf{\Pi}_1 + (\pi'-\pi)\mathbf{\Pi}_1 + \pi\mathbf{\Pi}_2 = (1-\pi')\mathbf{\Pi}_1 + \pi'\mathbf{\Pi}_2'$, say. Thus, if the data are not fitted well by H, we can expect that as π increases the model H_π will become an adequate fit for π sufficiently large. This simple idea is exploited next to construct an index of fit based on the mixture model construction.

For any distribution $\mathbf{P}$, we associate with it a value π^* that represents the smallest π such that $\mathbf{P} \in H_\pi$, i.e. we define the functional $\pi^*(\mathbf{P})$ as

$$\pi^*(\mathbf{P}) = \inf\{\pi\colon \mathbf{P} = (1-\pi)\mathbf{\Pi}_1 + \pi\mathbf{\Pi}_2;\ \mathbf{\Pi}_1 \in H,\ \mathbf{\Pi}_2 \text{ arbitrary}\}. \tag{2}$$

In words, π^* is the smallest π such that $\mathbf{P}$ can be fitted exactly by model H_π. Note that the nested nature of the H_π-models implies that $\mathbf{P} \in H_{\pi'}$ for $\pi' \in (\pi^*, 1]$. Let $\mathbf{f}$ denote an observed frequency distribution, and let $\hat{\mathbf{P}} = \mathbf{f}/n$ be the unrestricted maximum likelihood estimator of $\mathbf{P}$. Now define $\hat{\pi}^*$ by

$$\hat{\pi}^* = \pi^*(\hat{\mathbf{P}}). \tag{3}$$

Note that $\hat{\pi}^*$ is uniquely defined by virtue of definition (2) using the property that $\hat{\mathbf{P}}$ is unique. It follows that the maximum likelihood estimator of $\pi^*(\mathbf{P})$ is $\hat{\pi}^*$ as defined by equation (3).

The proposed index of fit described above can be interpreted easily and naturally in terms of the mixture representation. Essentially, the lack of fit of H is summarized in $\mathbf{\Pi}_2$, and the relative *size* of the parameter values (or interactions) omitted from H is summarized by the mixing weight π^*. Stated another way, $1-\pi^*$ is the proportion of the population *intrinsically described* by H and π^* is the proportion *intrinsically not described* by H. The quantity $\hat{\pi}^*$ defined above is the relevant maximum likelihood estimator.

Several analogues to our approach can be found. Goodman (1975) considered a scaling model derived from Guttman's scaling model (analogous to our model H) and defined the part of the population where this model does not hold as the intrinsically unscalable class. Clogg and Sawyer (1981) presented this model as a

two-point mixture; the generalization to a wide class of scaling models was given in Dayton and Macready (1980). In these references as well as in most references on latent class analysis, the usual interpretation is that *misclassification* or measurement error accounts for the failure of H. As in these references, the quantity π^* could be viewed as a measure of the proportion of the population measured with error (or misclassified). See also Espeland and Hui (1987) for a similar perspective. The model and the diagnostic parameter π can also be recast in terms of the resistance concept of Ylvisaker (1977).

An important property of the proposed index of the lack of fit is that if we have two models H and K, with $H \subset K$, then $\hat{\pi}_K^* \leqslant \hat{\pi}_H^*$. This follows because $H \subset K$ implies $H_\pi \subset K_\pi$, so $\mathbf{P} \in H_{\pi_H^*}$ implies $\mathbf{P} \in K_{\pi_H^*}$, i.e. the lack-of-fit index π^* decreases as we consider progressively more comprehensive models.

Our method of finding π^* defined in equation (3) is as follows. For each fixed π we maximize the likelihood over the model H_π (by the EM algorithm; see Section 6). The nestedness of the models H_π ensures that the maximized likelihood is monotonically increasing as π increases. Equivalently, the log-likelihood ratio statistic $L^2(\pi)$, for testing H_π against H_1, decreases in π. The estimator $\hat{\pi}^*$ is then the smallest π for which $L^2(\pi) = 0$.

To summarize, we propose π^* as a simple index of lack of fit which describes the magnitude of 'left-out interactions' or 'left-out terms'. It is the minimum proportion of the population for which the two-point mixture is saturated, the fraction of the population intrinsically outside the model H.

5. REANALYSIS OF EXAMPLES USING THE UNSTRUCTURED TWO-POINT MIXTURE APPROACH

We consider first a modification of the method just described where π is fixed at particular values. In this case, the EM algorithm is simple to apply; see Section 6. The model to be evaluated is row–column independence (H), as before. H was assumed for the first latent class (i.e. for $\mathbf{\Pi}_1$), and the second latent class was unrestricted or unspecified. In Table 3 the χ^2-statistics for various fixed values of π along with $\hat{\pi}^*$ are given for the data in Table 1. Note that when $\pi = 0$ the original model is obtained. A monotonic reduction in L^2-values is evident for increasing values of π, and $\hat{\pi}^* = 0.298$ indicates that the original table is far from independence in that nearly 30% of the population would have to be regarded as outside the model (or as misclassified). For any value of $\pi \geqslant \hat{\pi}^*$, the two-point mixture will also be saturated yielding fit statistics of 0. A lower (approximate) 95% bound for π^* is $\hat{\pi}_L^* = 0.236$, as noted in Table 3. (The construction of the lower confidence limit $\hat{\pi}_L^*$ is taken up in Section 6.)

Table 4 gives the analogous results for the data in Table 2. A value of π of about 0.09 leads to an acceptable fit; the approximate 95% lower bound for π^* is 0.091, and $\hat{\pi}^* = 0.104$. Given the potential for misclassification in either or both factors in Table 2, a misclassification rate of the order of 10% is reasonable, but the main conclusion is that the data in Table 2 are closer to the hypothesis of row–column independence than are the data in Table 1. Our analysis gives the same qualitative impression about the suitability of row–column independence for each set of data as the Diaconis–Efron approach. Other models could be evaluated in the same way, and whether the $\hat{\pi}^*$-value of 0.104 is sufficiently small to denote acceptance of

TABLE 3
Fit statistics for the unstructured mixture model applied to the data in Table 1†

π	X^2	L^2
0.00	138.29	146.44
0.10	47.35	48.67
0.15	23.74	24.36
0.20	8.55	8.75
0.236 ($=\hat{\pi}_L^*$)	2.57	2.66
0.25	1.38	1.44
0.26	0.83	0.87
0.27	0.42	0.43
0.28	0.16	0.16
0.29	0.02	0.02
0.298 ($=\hat{\pi}^*$)	0.00	0.00
$\pi \geqslant 0.298$	0.00	0.00

†See the text for the definitions of $\hat{\pi}^*$ and $\hat{\pi}_L^*$.

row–column independence depends on judgmental factors, including the objectives of the analysis.

It is interesting to apply this approach to the 2×2 table where explicit formulae can be obtained. For the table of frequencies, $\{f_{11}=a,\ f_{12}=b,\ f_{21}=c,\ f_{22}=d\}$, consider the case where $ad-bc>0$ and $a>d$. It can be verified that $\hat{\pi}^*=(ad-bc)/an$, where $n=a+b+c+d$, and the frequencies corresponding to Π_1 are $\{a,\ b,\ c,\ bc/a\}$, and the frequencies corresponding to Π_2 are $\{0,\ 0,\ 0,\ d-bc/a\}$. (These results change in obvious ways for other cases; if $a=d$, the expression for $\hat{\pi}^*$ is unchanged but two solutions for the Π-tables in the mixture equation are possible.) Consider the 2×2 table, $\{f_{11}=60, f_{12}=20, f_{21}=20, f_{22}=60\}$, $X^2=40.00$. The value of $\hat{\pi}^*$ is $\frac{1}{3}$, which is of course unchanged if the frequencies are multiplied by a constant such as 0.1 or 100. This case is illuminating because it illustrates the indeterminacy (or lack of identification) of the other parameters in the H_π-model which arises in this case because $a=d$. The minimum π is obtained

TABLE 4
Fit statistics for the unstructured mixture model applied to the data in Table 2†

π	X^2	L^2
0.00	568.57	569.42
0.07	22.37	22.64
0.08	10.38	10.49
0.09	3.07	3.09
0.091 ($=\hat{\pi}_L^*$)	2.58	2.59
0.10	0.11	0.10
0.104 ($=\hat{\pi}^*$)	0.00	0.00
$\pi \geqslant 0.104$	0.00	0.00

†See the text for the definitions of $\hat{\pi}^*$ and $\hat{\pi}_L^*$.

with frequencies $\{60, 20, 20, 6\frac{2}{3}\}$ corresponding to Π_1 and frequencies $\{0, 0, 0, 53\frac{1}{3}\}$ corresponding to Π_2 ($\hat{\pi}^* = 53\frac{1}{3}/160$), or with frequencies $\{6\frac{2}{3}, 20, 20, 60\}$ corresponding to Π_1 and frequencies $\{53\frac{1}{3}, 0, 0, 0\}$ corresponding to Π_2. If the frequencies had been $\{6, 2, 2, 6\}$, we would obtain $X^2 = 4.0$, giving a dramatically different impression about significance. But the value of $\hat{\pi}^*$ would be the same, conveying the fact that about 30% of the population (estimate) is outside the model. The cell residuals take on absolute value 20 for the first case and absolute value 2 for the second case. (For the 2×2 table, of course, there is just one non-redundant residual because there is 1 degree of freedom.)

6. INTERVAL ESTIMATION

Suppose that we adopt the point of view that the model H is to be rejected if it fails to explain the response of some (relatively large) fraction of the population, say π_0^*. Then we wish to test the null hypothesis $\pi^* \leqslant \pi_0^*$ against $\pi^* > \pi_0^*$.

We here consider the asymptotic properties of the likelihood ratio statistic $L^2(\pi_0^*)$ (and therefore of equivalent statistics) for testing this hypothesis. First note that, under the alternative, the likelihood will be maximized with fitted probabilities equal to the observed proportions. If $\pi_0^* = 0$, we are therefore exactly in the setting of the standard likelihood ratio test of model H against the unrestricted multinomial alternative, and so under regularity the likelihood ratio statistic $L^2(0)$ has an asymptotic χ^2-distribution with degrees of freedom equal to the number of cells N minus the number of non-redundant parameters in H minus 1. Denote this quantity as DF.

If we wish to test the null hypothesis $\pi^* \leqslant \pi_0^*$ against the alternative $\pi^* > \pi_0^*$ for π_0^* strictly positive, however, the asymptotic distribution theory is quite different. We claim that, under the condition $\pi^* = \pi_0^*$, $L^2(\pi_0^*)$ has, asymptotically, a mixed χ^2-distribution, with probability 0.5 equal to 0 and with probability 0.5 equal to $\chi^2_{(1)}$. Under this claim, a test with $\alpha = 0.05$ would reject the null hypothesis whenever the test statistic exceeds $\chi^2_{(1)}(0.90) = 2.70$. Inverting this test procedure to obtain a lower 95% confidence limit $\hat{\pi}_L^*$ amounts to finding the value of π^* for which $L^2(\pi^*)$ attains the value 2.70 (see Tables 3 and 4).

Before proving our claim we note that the results in the previous two paragraphs show that the limiting distributions in question are discontinuous in π_0^* at $\pi_0^* = 0$. Therefore in practice we must be cautious about using $\hat{\pi}_L^*$ when π_0^* is close to 0 and the sample size is small. In such cases, we recommend one of two methods:

(a) use, instead of the critical value 2.70, the critical value of the χ^2-statistic with DF degrees of freedom (the reference distribution for the hypothesis that $\pi_0^* = 0$), which is a conservative procedure, or
(b) simulate the null distribution.

The distributional claim for $\pi_0^* > 0$ must be argued geometrically. The set of all cell probabilities $\{\mathbf{P}\}$ forms an $(N-1)$-dimensional simplex. The probability model H_π, for $\pi > 0$, is a closed set, if H is closed, with an $(N-1)$-dimensional interior. On the boundary of the null hypothesis (i.e. when $\pi^*(\mathbf{P}) = \pi_0^*$), the distribution $\mathbf{P}$ is on the boundary $\mathscr{B}$ of $H_{\pi_0^*}$. Under regularity, the boundary $\mathscr{B}$ is a locally smooth surface of dimension $N-2$, and so has a tangent hyperplane within the simplex. From here we may follow Self and Liang (1987) (see case 5). Asymptot-

ically, using the limiting normal distribution of the observed cell proportions $\mathbf{f}/n$ implies that with probability 0.5 this observed vector is inside $H_{\pi_0^*}$ and so $L^2(\pi_0^*)=0$, and with probability 0.5 it is outside, giving the $\chi^2_{(1)}$-distribution, under the null hypothesis.

The discontinuity in the limiting distribution arises because the dimension of $H=H_0$ does not match that of H_π for $\pi > 0$.

7. MAXIMUM LIKELIHOOD ALGORITHMS

The EM algorithm of Dempster *et al.* (1977) can be used to calculate maximum likelihood estimates for the mixture models presented here; also see Hartley and Hocking (1971) and Little and Rubin (1987). Other algorithms could be used, but because all parameters are not identified in general (e.g. values for the unspecified multinomial Π_2) the EM algorithm is attractive because matrix inversion and information matrices are not required. We show how this approach can be applied with trivial modification of routines that would be used to find maximum likelihood estimates of any model H for the observed table.

For illustration suppose that a two-way contingency table is considered with I rows and J columns. Suppose further that H is the hypothesis of row–column independence. The complete data array is thus an $I\times J$ contingency table whereas the incomplete data array is $I\times J\times 2$. As earlier, the conditional distribution in the first layer is Π_1 (where H is posited) whereas in the second it is Π_2 (unspecified). The two layers have proportionate distribution $1-\pi$ and π, which for the moment is considered fixed. Let Q_{ijk}, $i=1, \ldots, I$, $j=1, \ldots, J$, $k=1, 2$, denote the cell probabilities in the incomplete array. Only the marginal Q_{ij+} can be observed, i.e. $Q_{ij+}=f_{ij}/n=p_{ij}$, the observed proportion in cell (i, j).

Let $Q^{(0)}_{ijk}$ denote initial estimates such that $Q^{(0)}_{++1}=1-\pi$ and $Q^{(0)}_{++2}=\pi$. Initial values can be obtained in a variety of ways, but a simple strategy is as follows. For the first latent class, set

$$Q^{(0)}_{ij1} = (1-\pi)\hat{P}^{(H)}_{ij}, \tag{4}$$

where $\hat{P}^{(H)}_{ij}$ denotes the maximum likelihood estimate of P_{ij} under H. (For H equivalent to row–column independence, $\hat{P}^{(H)}_{ij}=f_{i+}f_{+j}/n^2$.) For the second latent class, set

$$Q^{(0)}_{ij2} = \pi(IJ)^{-1}. \tag{5}$$

The conditions $Q^{(0)}_{++1}=1-\pi$ and $Q^{(0)}_{++2}=\pi$ are satisfied with this choice of starting values.

At cycle s, the E- (expectation) step of the algorithm is defined by

$$g^{(s)}_{ijk} = p_{ij}Q^{(s)}_{ijk}/Q^{(s)}_{ij+}, \tag{6}$$

for all i, j and k. The M- (maximization) step is

$$Q^{(s+1)}_{ij1} = (1-\pi)Q^{(s)}_{ij|1}, \tag{7}$$

where $Q^{(s)}_{ij|1}$ denotes the maximum likelihood estimate of Π_1 in cell (i, j) under hypothesis H, i.e.

$$Q^{(s)}_{ij|1} = Q^{(s)}_{i+|1}Q^{(s)}_{+j|1}, \tag{8}$$

where $Q^{(s)}_{i+|1} = g^{(s)}_{i+1}/(1-\pi)$ and $Q^{(s)}_{+j|1} = g^{(s)}_{+j1}/(1-\pi)$. This update pertains to the first latent class. For the second latent class (or layer), the update is

$$Q^{(s+1)}_{ij2} = \pi(g^{(s)}_{ij2}/g^{(s)}_{++2}). \tag{9}$$

Note how the second set of updates is unspecified by a model or hypothesis; no constraints are applied in contrast with the case of conventional latent class models. Repeated cycling between the E- and M-steps defines the algorithm, with updates defined in the usual way.

The generalization to an arbitrary model H and an arbitrary contingency table is straightforward. As before, let h denote a cell in the table. Q_{ijk} above is replaced by Q_{hk}, for $h=1, \ldots, N$ and $k=1, 2$, with $Q_{h+} = f_h/n = p_h$ observed. Pick initial values using the generalization of equation (4), i.e. $Q^{(0)}_{h1} = (1-\pi)\hat{P}^H_h$, where $\hat{P}^H_h$ denotes the maximum likelihood estimate of P_h under model H. Formula (5) is replaced by $Q^{(0)}_{h2} = \pi/N$. The E-step of equation (6) is replaced by $g^{(s)}_{hk} = p_h Q^{(s)}_{hk}/Q^{(s)}_{h+}$ for all h and k. The M-step of equation (7) for the first latent class is replaced by $Q^{(s+1)}_{h1} = (1-\pi)Q^{(s)}_{h|1}$ where $Q^{(s)}_{h|1}$ is the maximum likelihood estimate of Π_{1h} under H. The algorithm used to find maximum likelihood estimates under H for the observed table can be used directly to find these values. The M-step for the second (unrestricted) latent class, following equation (9), is $Q^{(s+1)}_{h2} = \pi(g^{(s)}_{h2}/g^{(s)}_{+2})$. For π specified in advance, this algorithm could be appended easily to the routine for finding maximum likelihood estimates under H.

The algorithm can be modified to find $\hat{\pi}^*$. We used a procedure that estimates the profile likelihood for increasing values of π; see the results in Tables 3 and 4. The solution is obtained as the smallest possible value of π for which the likelihood ratio statistic $L^2(\pi) = 0$. Pick a suitably small value for the trial value, e.g. $\pi^{*(0)} = 0.05$. Then apply the algorithm defined by equations (4)–(9) with the value π replaced by $\pi^{*(0)}$. If the model is already saturated ($L^2 = 0$), then begin with a smaller value of the mixing weight. Otherwise, proceed in increments, with, say, $\pi^{*(1)} = 0.10$. Reapply the algorithm for fixed π defined by equations (4)–(9), and reiterate to find the smallest possible value with $L^2 = 0$.

We could alternatively use the method of 'binary search' or 'line splitting' exploiting the fact that $\hat{\pi}^* \in [0, 1]$ and $L^2(\pi) = 0$ for $\pi > \hat{\pi}^*$. Pick $\pi^{*(0)} = 0.5$, say, and if $L^2 = 0$ pick $\pi^{*(1)} = 0.25$; if $L^2(\pi^{*(1)}) > 0$ pick $\pi^{*(2)} = (\pi^{*(0)} + \pi^{*(1)})/2 = 0.375$ (otherwise pick $\pi^{*(2)} = (0 + \pi^{*(1)})/2 = 0.125$); continue halving successive trial values until convergence to the desired accuracy occurs.

The computational labour could also be reduced by letting the value of π^* at loop t be

$$\pi^{*(t)} = \pi^{*(t-1)} + a_t$$

where $a_t \geqslant 0$. After loop t^* (with $t^* > 3$, say), pick a_{t^*+1} by fitting, say, a quadratic function to the $\{a_1, a_2, \ldots, a_{t^*}\}$ using the values $\{L^2_{(0)} - L^2_{(1)}, L^2_{(1)} - L^2_{(2)}, \ldots, L^2_{(t^*-1)} - L^2_{(t^*)}\}$. The procedure continues until the smallest value for which $L^2 = 0$ is obtained. For each loop $t \geqslant 1$, the starting values for Q_{ij1} (see equation (4)) can be taken as $(1-\pi^{*(t)})\hat{\Pi}_1$, where $\hat{\Pi}_1$ is the estimate of Π_1 obtained at loop $t-1$.

The computational burden depends on the true value of $\hat{\pi}^*$, the starting value for this parameter and the complexity of the M-step for the first latent class (the

calculations required to find maximum likelihood estimates under hypothesis H). It is likely that improvements in the algorithm used to find $\hat{\pi}^*$ can be made.

8. ANALYSIS USING CONVENTIONAL LATENT CLASS MODELS

Conventional latent class analysis (Goodman, 1974) is closely related to the approach suggested above, but it is based on a *structured* representation of the lack of fit of H. In latent class analysis we assume that H applies to *both* Π_1 and Π_2 and that H is the hypothesis of independence. The assumption of within latent class independence ('local independence') is not necessary. For example, Formann (1993) considered the case where H is the fixed distance model incorporating a special kind of row–column interaction in each latent class, and Agresti (1991) considered the case where H is the linear-by-linear interaction model. The conventional approach based on local independence is considered next along with some of its limitations.

Applying the conventional approach to the two examples of Section 2 leads to the latent class models for two-way contingency tables given in Clogg (1981); also see Goodman (1987) and van der Heijden *et al.* (1992). Let cell $h=(i, j)$, for $i=1, \ldots, I$ and $j=1, \ldots, J$, where I and J denote the number of rows and the number of columns respectively. Independence within latent class t ($t=1, 2$) means that $\Pi_{th} = \Pi_{tij} = \pi_{A(i)|X(t)}\pi_{B(j)|X(t)}$ where, for example, $\pi_{A(i)|X(t)} = P(A=i|X=t)$, and A and B refer to the row and column variables respectively. (These conditional probabilities should not be confused with the marginal probabilities of the mixing variable, $1-\pi$ and π.) Table 5 gives maximum likelihood estimates of the parameters for this model applied to the data in Table 1, and Table 6 gives the corresponding entries for the model applied to the data in Table 2.

The two-class latent structure applied to a two-way table is not identifiable unless two restrictions are imposed, and some of the ways that identification can be

TABLE 5

Parameter values for the two-class latent structure model applied to the data in Table 1†

Parameter	*Latent class t*		*Marginal proportion*
	$t=1$	$t=2$	
$\pi_{A(i)\|X(t)}$			
$i=1$	0.550	*0.000*	0.372
$i=2$	0.165	0.776	0.363
$i=3$	0.197	0.074	0.157
$i=4$	0.088	0.150	0.108
$\pi_{B(j)\|X(t)}$			
$j=1$	0.270	*0.000*	0.182
$j=2$	0.561	0.321	0.483
$j=3$	0.139	0.081	0.120
$j=4$	0.031	0.598	0.215
$\{(1-\pi), \pi\}$	0.676	0.324	

†$X^2=14.899$, $L^2=14.174$, $D=0.047$, 4 degrees of freedom. Entries in italics denote restrictions used to identify $\hat{\pi}$ and the other parameters of the model.

TABLE 6
Parameter values for the two-class latent structure model applied to the data in Table 2†

Parameter	*Latent class t*		*Marginal proportion*
	t=1	*t=2*	
$\pi_{A(i)\|X(t)}$			
$i=1$	0.225	0.284	0.242
$i=2$	0.373	0.580	0.433
$i=3$	0.233	0.135	0.205
$i=4$	0.169	*0.000*	0.121
$\pi_{B(j)\|X(t)}$			
$j=1$	0.530	*0.004*	0.378
$j=2$	0.358	0.642	0.440
$j=3$	0.098	0.258	0.144
$j=4$	0.011	0.081	0.031
$j=5$	*0.004*	0.014	0.007
$\{(1-\pi), \pi\}$	0.712	0.288	

†$X^2=19.021$, $L^2=18.537$, $D=0.008$, 6 degrees of freedom. Entries in italics denote restrictions used to identify $\hat{\pi}$ and the other parameters in the model; compare the restrictions with those used in Table 5.

achieved are illustrated in these tables. In Table 5, the restrictions used to identify parameter values were

$$\pi_{A(1)|X(2)} = \pi_{B(1)|X(2)} = 0.0,$$

i.e. two of the conditional probabilities pertaining to the second latent class were set at boundary values. In Table 6, we used the restrictions

$$\pi_{A(4)|X(2)} = 0.0, \qquad \hat{\pi}_{B(5)|X(1)} = \hat{\pi}_{B(1)|X(2)},$$

for a total of two restrictions, one a fixed (boundary) restriction and the other an equality constraint involving conditional probabilities from both latent classes. The magnitude of $\hat{\pi}$ is affected by the restrictions imposed (Goodman, 1987); however, the fitted values and hence the magnitudes of fit statistics are not affected. In this case, the estimate of $\hat{\pi}$ is virtually the same for each table (0.32 *versus* 0.29), so this approach does not give a clear signal about the different levels of fit of model H in the two tables. In contrast, the unstructured approach based on the saturated two-point mixture shows clearly that H is more satisfactory for the second set of data. The main complication is that the value of $\hat{\pi}$ in the conventional latent class model depends on the (possibly arbitrary) restrictions used.

For many other contingency tables, the two-class latent structure will have identifiable parameters, and this *structured* approach could be applied in these other situations without the ambiguity created by underidentification. For example, the two-class latent structure has identifiable parameters when more than two variables are considered. (The model parameters are identified even for a $2\times2\times2$ table, although the model has 0 degrees of freedom or is 'saturated' in this case; see Lazarsfeld and Henry (1968) and Goodman (1974).)

Although the two-class latent structure does not fit either set of data when judged by conventional criteria ($\chi^2_{4,0.95}=9.5$ and $\chi^2_{6,0.95}=12.6$), the magnitudes of the fit statistics are reduced dramatically in each case, so much so that a two-point mixture of this type is *almost* congruent with the data in each instance. Whereas conventional latent class analysis represents an obvious way to develop final models for the data, the approach considered earlier provides a more direct method of evaluating the goodness of fit.

9. CONCLUSION AND SOME EXTENSIONS

Although the approach put forth deals with goodness of fit in the analysis of contingency tables, it can be applied in many other settings. To illustrate, consider a possible analysis involving the bivariate normal distribution. For an arbitrary bivariate normal distribution, suppose that the hypothesis H to be examined is independence. Without loss of generality, suppose that (X, Y) follows the bivariate normal distribution with mean (0, 0) and $\text{var}(X)=\sigma_1^2$, $\text{var}(Y)=\sigma_2^2$, $\text{cov}(X, Y)=\rho\sigma_1\sigma_2$. Let $f_{XY}(x, y)$ denote the density function. If H is the hypothesis of independence (i.e. $\rho=0$), then the representation corresponding to model (1) is

$$f_{XY}(x,y) = (1-\pi^*)\, g_X(x)\, g_Y(y) + \pi^* \{f_{\text{LOF}}(x,y)\} \tag{10}$$

where $f_{\text{LOF}}(x, y)$ denotes the unspecified joint density for lack of fit of the independence model and $g_X(x)$ and $g_Y(y)$ denote arbitrary univariate normal densities. It is shown in Appendix A that

$$\pi^* = 1 - \{(1-|\rho|)/(1+|\rho|)\}^{1/2} \tag{11}$$

if $|\rho|<1$. The fraction π^* outside the model of independence is thus a monotonically increasing function of the correlation ρ and hence measures the distance from independence. For $|\rho|$-values in the set {0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}, the corresponding π^*-values are {0.00, 0.095, 0.184, 0.266, 0.345, 0.423, 0.500, 0.580, 0.667, 0.771}; for π^*-values in the set {0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}, the corresponding $|\rho|$-values are {0.00, 0.105, 0.220, 0.342, 0.471, 0.600, 0.724, 0.835, 0.923, 0.980}, i.e. a correlation of 0.5 corresponds to the fraction 0.423 outside independence, whereas a π^*-value of 0.5 corresponds to $|\rho|=0.6$. Note that π^* can be estimated consistently by substituting the maximum likelihood estimator of ρ in the above expression. The index π^* can be interpreted as in the body of the paper (i.e. as the fraction of the population outside the independence relationship), which is perhaps a simpler interpretation of the degree of non-independence in the bivariate normal distribution than is the magnitude of $|\rho|$. It is also interesting to note that $\log(1-\pi^*)=\frac{1}{2}\log(1-|\rho|)/(1+|\rho|)$, which is essentially the same as Fisher's transform of the correlation coefficient. In this case, the asymptotic distribution for $\hat{\pi}^*$ is known by virtue of the relationship to Fisher's Z, with the sample estimator $\hat{\rho}$ replacing the parameter ρ in the above expression. These surprising results for the bivariate normal distribution indicate that the mixture-based index for assessing lack of fit can be easily and usefully extended beyond the contingency table setting. For a two-way contingency table that represents a discretization of continuous variables, this analysis also provides an

interpretation of π^* in terms of the correlation in an underlying bivariate normal model (see Goodman (1985)).

We conclude with some issues that need to be examined further.

(a) Although $\hat{\pi}^*$ is consistent, it will tend to be biased upwards in finite samples if π^* is relatively small. To see this, note that, even if H holds true so that $\pi_0^* = 0$, $\hat{\pi}^*$ will be greater than 0 for finite n with probability approaching 1. The bias will be inconsequential for very large samples, but a detailed investigation would be appropriate.

(b) A measure of precision for $\hat{\pi}^*$ ought also to be considered. Clearly, calculations of the standard error derived in the usual way (e.g. from assessing the curvature of the log-likelihood at the maximum) are not relevant since the profile log-likelihood is flat for $\pi > \hat{\pi}^*$. We note that the lower confidence limit $\hat{\pi}_L^*$ provides inferential information that is independent of the bias in $\hat{\pi}^*$ and, via $\hat{\pi}^* - \hat{\pi}_L^*$, gives an indication of the magnitude of the error in $\hat{\pi}^*$.

(c) The possible effect of sampling zeros in $\mathbf{f}$ ought to be investigated further. If $f_h = 0$ for cell h, for example, then $\hat{\Pi}_{1h} = \hat{\Pi}_{2h} = 0$ and the value of $\hat{\pi}^*$ might be affected by this (in small sample situations). The effect of sampling zeros is not obvious, however. Consider the 2×2 table $\{a = n, b = c = d = 0\}$, where n is small. This table satisfies row–column independence, with two zero marginals and three sampling zeros. Here $\hat{\pi}^* = 0$ and the estimate of Π_1 is equal to the observed distribution; there is no inflation in the estimator of π^* due to sampling zeros in this case. The effect of sampling zeros will depend on the structure of the data as well as the suitability of the model H for the data. In cases where H is row–column independence, a sampling zero will make the estimate of the row or the column total in Π_1 equal to 0. And, in general, this will tend to increase $\hat{\pi}^*$ by an amount that is directly related to the smaller of the observed row marginal proportion and the observed column marginal proportion pertaining to the cell with a sampling zero. If both of these marginal probabilities are relatively large, then the sampling zero is an indication of a possibly extreme departure from row–column independence, and we would expect the value of $\hat{\pi}^*$ to be relatively large as a consequence. Other approaches that ought to be investigated include replacing sampling zeros by small positive flattening constants or redefining model H by regarding the sampling zeros as structural zeros.

(d) The approach put forth here provides implicitly a new definition of *residual analysis* that ought to be explored further. Note that cell residuals under model H are summarized in Π_2. The non-zero elements of $\hat{\Pi}_2$ provide information about sources of lack of fit that are related to but different from the information provided in the ordinary residuals obtained by fitting H directly to $\mathbf{f}$.

(e) Finally, the approach ought to be extended to a wider class of models, such as the class of generalized linear models.

We believe that our mixture approach and the index π^* derived from it deserve to be considered as a supplement to or as a replacement for some of the other popular published indices of fit.

ACKNOWLEDGEMENTS

Rudas's research was supported in part by grant OTKA 2513 from the Hungarian National Science Foundation and was initiated while Rudas was Visiting Professor of Statistics at Pennsylvania State University. Clogg's research was supported in part by grant SES-9011973 from the National Science Foundation, Lindsay's research was supported in part by grant DMS-910685-A01 from the National Science Foundation, and Clogg and Lindsay were supported in part by the Population Research Institute, Pennsylvania State University, with core support from the National Institute of Child Health and Human Development (grant 1-HD28263-01). The authors are indebted to M. Becker, W. Eddy, J. Landwehr, M. Meyer, C. Mallows, the Associate Editor and a referee for helpful comments.

APPENDIX A: INDEX π^* AS MEASURE OF CORRELATION IN BIVARIATE NORMAL DISTRIBUTION

The conditions are those provided at the beginning of Section 9. Without loss of generality, assume that g_X corresponds to an $N(0, \sigma_1^2/\lambda_1)$ density and that g_Y corresponds to an $N(0, \sigma_2^2/\lambda_2)$ density, with $\lambda_1 = \lambda + \epsilon$, $\lambda_2 = \lambda - \epsilon$, $\lambda_1 > 0$ and $\lambda_2 > 0$. (g_X and g_Y do not refer to the corresponding marginal densities derived from f_{XY}.) Because $f_{\mathrm{LOF}}(x, y) \geqslant 0$ and $\pi^* \geqslant 0$, equation (10) is equivalent to

$$f_{XY}(x, y) \geqslant c\, g_X(x)\, g_Y(y) \tag{12}$$

for all x and y. We seek the largest constant c with $0 \leqslant c \leqslant 1$ (where $c = 1 - \pi^*$) such that inequality (12) holds. Hence we define

$$c = \inf_{x,y}\{f_{XY}(x, y)/g_X(x)\, g_Y(y)\}. \tag{13}$$

Now consider the scale change $x_1 = x/\sigma_1$, $x_2 = y/\sigma_2$, with Jacobian $\sigma_1\sigma_2$, applied to both sides of inequality (12). After simplification we obtain

$$(1-\rho^2)^{-1/2} \exp(-Q_J/2) \geqslant c(\lambda^2 - \epsilon^2)^{1/2} \exp(-Q_I/2), \tag{14}$$

making use of the fact that $\lambda_1\lambda_2 = \lambda^2 - \epsilon^2$, where $Q_J = (x_1^2 + x_2^2 - 2\rho x_1 x_2)/(1-\rho^2)$ and $Q_I = \lambda_1 x_1^2 + \lambda_2 x_2^2$. Using definition (13) we can therefore write

$$c = (\lambda^2 - \epsilon^2)^{-1/2}(1-\rho^2)^{-1/2} \inf_{x_1, x_2}\{\exp(-Q/2)\} \tag{15}$$

where $Q = Q_J - Q_I = (x_1^2 + x_2^2 - 2\rho x_1 x_2)/(1-\rho^2) - \lambda_1 x_1^2 - \lambda_2 x_2^2$, a quadratic form in x_1 and x_2. Q can be written as $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ with $\mathbf{A} = \{a_{ij}\}$ where $a_{11} = (1-\rho)^{-1} - (\lambda + \epsilon)$, $a_{22} = (1-\rho^2)^{-1} - (\lambda - \epsilon)$ and $a_{12} = a_{21} = -\rho/(1-\rho)^2$.

If $\mathbf{A}$ has any positive eigenvalues, then by choosing $\mathbf{x}$ proportional to an eigenvector associated with a positive eigenvalue we can make the term $\exp(-\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}/2)$ as small as we like, so $\mathbf{A}$ must be negative semidefinite ($-\mathbf{A}/2$ must be positive semidefinite) if c is to be greater than 0. If $\mathbf{A}$ is negative semidefinite, then $-\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}/2$ takes on its minimal value, 0, at $\mathbf{x} = \mathbf{0}$, so that

$$c = (\lambda^2 - \epsilon^2)^{-1/2}(1-\rho^2)^{-1/2}. \tag{16}$$

Thus we need to choose λ and ϵ (or λ_1 and λ_2) to maximize $(\lambda^2 - \epsilon^2)^{-1/2}$ subject to the condition that $\mathbf{A}$ is negative semidefinite.

Given (λ, ϵ), the eigenvalues of $\mathbf{A}$ are

$$(1-\rho^2)^{-1} - \lambda \pm \{\epsilon^2 + \rho^2/(1-\rho^2)^2\}^{1/2}.$$

Setting the larger eigenvalue to 0, we have

$$\lambda = \{\epsilon^2 + \rho^2/(1-\rho^2)^2\}^{1/2} + (1-\rho^2)^{-1}. \quad (17)$$

We wish to minimize $\lambda^2 - \epsilon^2$, which for the above λ is

$$\lambda^2 - \epsilon^2 = (1-\rho^2)^{-2} + \rho^2(1-\rho^2)^{-2} + 2(1-\rho^2)^{-1}\{\epsilon^2 + \rho^2(1-\rho^2)^{-2}\}^{1/2}.$$

This is minimized at $\epsilon = 0$, and substituting in equation (17) with $\epsilon = 0$ we obtain $\lambda = |\rho|/(1-\rho^2) + 1/(1-\rho^2) = 1/(1-|\rho|)$. This gives the result in equation (11) on substitution in equation (16).

REFERENCES

Agresti, A. (1990) *Categorical Data Analysis*. New York: Wiley.

——— (1991) Modelling patterns of agreement and disagreement. *Statist. Meth. Med. Res.*, **1**, 201–218.

Atkinson, A. C. (1981) Likelihood ratios, posterior odds and information criteria. *J. Econometr.*, **16**, 15–20.

Becker, M. P. and Clogg, C. C. (1989) Analysis of sets of two-way contingency tables using association models. *J. Am. Statist. Ass.*, **84**, 142–151.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975) *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: Massachusetts Institute of Technology Press.

Clogg, C. C. (1981) Latent structure models of mobility. *Am. J. Sociol.*, **86**, 836–868.

——— (1988) Latent class models for measuring. In *Latent Trait and Latent Class Models* (eds R. Langeheine and J. Rost), pp. 173–205. New York: Plenum.

Clogg, C. C. and Goodman, L. A. (1984) Latent structure analysis of a set of multidimensional contingency tables. *J. Am. Statist. Ass.*, **79**, 762–771.

Clogg, C. C. and Sawyer, D. O. (1981) A comparison of alternative models for analyzing the scalability of response patterns. In *Sociological Methodology 1981* (ed. S. Leinhardt), pp. 240–280. San Francisco: Jossey-Bass.

Cramer, H. (1946) *Mathematical Methods in Statistics*. Princeton: Princeton University Press.

Dayton, C. M. and Macready, G. B. (1980) A scaling model with response errors and intrinsically unscalable respondents. *Psychometrika*, **45**, 343–356.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc.* B, **39**, 1–22.

Diaconis, P. and Efron, B. (1985) Testing for independence in a two-way contingency table: new interpretations of the chi-square statistic. *Ann. Statist.*, **13**, 845–874.

Espeland, M. A. and Hui, S. L. (1987) A general approach to analyzing epidemiologic data that contain misclassification errors. *Biometrics*, **43**, 1001–1012.

Formann, A. (1993) Fixed-distance latent class models for the analysis of sets of two-way contingency tables. *Biometrics*, **49**, 511–521.

Fowlkes, E. B., Freeny, A. E. and Landwehr, J. M. (1988) Evaluating logistic models for large contingency tables. *J. Am. Statist. Ass.*, **83**, 611–622.

Goodman, L. A. (1974) Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, **61**, 215–231.

——— (1975) A new model for scaling response patterns: an application of the quasi-independence concept. *J. Am. Statist. Ass.*, **70**, 755–768.

——— (1984) *The Analysis of Cross-classified Data having Ordered Categories*. Cambridge: Harvard University Press.

——— (1985) The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models and asymmetry models for contingency tables with or without missing entries. *Ann. Statist.*, **13**, 10–69.

——— (1987) New methods for analyzing the intrinsic character of qualitative variables using cross-classified data. *Am. J. Sociol.*, **93**, 529–583.

Haberman, S. J. (1979) *Analysis of Qualitative Data*, vol. II, *New Developments*. New York: Academic Press.

Hartley, H. O. and Hocking, R. R. (1971) The analysis of incomplete data. *Biometrics*, **27**, 783–808.

van der Heijden, P. G. M., Mooijaart, A. and de Leeuw, J. (1992) Constrained latent budget analysis. In *Sociological Methodology 1992* (ed. P. V. Marsden), pp. 279–320. Oxford: Basil Blackwell.

Lazarsfeld, P. F. and Henry, N. W. (1968) *Latent Structure Analysis*. Boston: Houghton Mifflin.

Little, R. J. A. and Rubin, D. B. (1987) *Statistical Analysis with Missing Data*. New York: Wiley.

Read, T. R. C. and Cressie, N. A. C. (1988) *Goodness-of-fit Statistics for Discrete Multivariate Analysis*. New York: Springer.

Rudas, T. (1986) A Monte Carlo comparison of the small sample behaviour of the Pearson, the likelihood ratio and the Cressie–Read statistics. *J. Statist. Computn Simuln*, **24**, 107–120.

Self, S. G. and Liang, K.-Y. (1987) Asymptotic properties of maximum likelihood estimators and likelihood-ratio tests under nonstandard conditions. *J. Am. Statist. Ass.*, **82**, 605–610.

Snee, R. (1974) Graphical display of two-way contingency tables. *Am. Statistn*, **38**, 9–12.

Ylvisaker, D. (1977) Test resistance. *J. Am. Statist. Ass.*, **72**, 551–556.