

Data Information in Contingency Tables: A Fallacy of Hierarchical Loglinear Models

Philip E. Cheng, Jiun W. Liou, Michelle Liou and John A. D. Aston
Academia Sinica

Abstract: Information identities derived from entropy and relative entropy can be useful in statistical inference. For discrete data analyses, a recent study by the authors showed that the fundamental likelihood structure with categorical variables can be expressed in different yet equivalent information decompositions in terms of relative entropy. This clarifies an essential difference between the classical analysis of variance and the analysis of discrete data, revealing a fallacy in the analysis of hierarchical loglinear models. The discussion here is focused on the likelihood information of a three-way contingency table, without loss of generality. A classical three-way categorical data example is examined to illustrate the findings.

Key words: Entropy, likelihood ratio test, loglinear models, mutual information.

1. Introduction

The analysis of contingency tables with multi-way classifications originates from the historical development of statistical inference with 2×2 tables. In the initial extension to the case of $2 \times 2 \times K$ tables, Bartlett (1935) discussed testing for three-way interaction and derived an estimate of the common odds ratio suggested by R. A. Fisher. Norton (1945) and Simpson (1951) supplied interpretations of varied interactions which led to the well-known Simpson's paradox (Blyth, 1972). Roy and Kastenbaum (1956) showed that Bartlett's procedure is an implicit maximum likelihood estimation (MLE), conditioned upon the fixed margins of each 2×2 table. The celebrated analysis of variance (ANOVA, Fisher, 1925) inspired discussions of partitioning chi-squares within the contingency tables, notably by Lancaster (1951), Mood (1950), and Claringbold (1961), among others. In related research in biostatistics, Cochran (1954), Woolf (1955) and Mantel and Haenszel (1959) developed chi-square tests for no association between two variables across levels of the third variable. These early studies led to further analyses of three-way tables, which include estimating the common odds ratio, testing zero interaction and testing no association across strata, for

examples, Kullback (1959), Gart (1962), Darroch (1962), Lewis (1962), Plackett (1962), Birch (1963, 1964), and Goodman (1964).

The classical method of partitioning the chi-squares for the three-way ($I \times J \times K$) contingency table does not provide a convenient test of the null hypothesis that the three-way interaction is zero (Lancaster, 1951). The null distribution of Lancaster's test statistic need not be asymptotically chi-square distributed (Plackett, 1962). A similar remark can also be applied to the methods of partitioning chi-squares by Kullback (1959) and Claringbold (1961). Since partitions of chi-squares are closely related to the likelihood ratio tests (Wilks, 1935), maximum likelihood estimation of association and interaction significantly influenced the early studies of multi-way contingency tables (Roy and Kastenbaum, 1956). In particular, it led to the development of a likelihood ratio test for zero interaction in a multi-way contingency table (Darroch, 1962). This is given a close examination below, and also in Section 4.

Suppose that individuals of a sample are classified according to three categorical variables $\{X\}$, $\{Y\}$, $\{Z\}$ with classified levels: $i = 1, \dots, I$, $j = 1, \dots, J$, $k = 1, \dots, K$, respectively. Denote the joint probability density by $p_{ijk} = P(X = i, Y = j, Z = k)$, which satisfies that $\sum_{ijk} p_{ijk} = p_{\dots} = 1$. Define

$$p_{ij\cdot} = \sum_k p_{ijk} = P(X = i, Y = j), \quad p_{i\cdot\cdot} = \sum_{j,k} p_{ijk} = P(X = i), \quad (1.1)$$

as the marginal probabilities, and define $p_{i\cdot k}$, $p_{\cdot jk}$, $p_{\cdot j\cdot}$ and $p_{\cdot\cdot k}$ analogously.

Also, let n_{ijk} denote the number of individuals classified to the cell ($X = i, Y = j, Z = k$). A classical test of primary interest concerns the notion of no (or, zero) three-way interaction in the full $I \times J \times K$ table (Bartlett, 1935). This is defined by the set of $(I-1)(J-1)(K-1)$ conditions on the joint probabilities:

$$\frac{p_{ijk} p_{IJK}}{p_{iJk} p_{Ijk}} = \frac{p_{ijK} p_{IJK}}{p_{iJK} p_{IjK}} \quad (i = 1, \dots, I-1, j = 1, \dots, J-1, k = 1, \dots, K-1). \quad (1.2)$$

In addition to providing a likelihood ratio test for (1.2) (via solving the maximum likelihood estimates by the Lagrange multiplier method), Roy and Kastenbaum (1956, pp. 750-753) recommended an equivalence test for the independence between one variable and the remaining two. Specifically, they argued that the combination of (1.2) with the hypotheses:

$$p_{i\cdot k} = p_{i\cdot} p_{\cdot\cdot k} \quad \text{and} \quad p_{\cdot jk} = p_{\cdot j\cdot} p_{\cdot\cdot k} \quad (i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K),$$

is equivalent to the hypothesis: for all (i, j, k) ,

$$p_{ijk} = p_{ij\cdot} p_{\cdot\cdot k} \quad (1.3)$$

It was taken for granted that equation (1.3) is directly analogous to a property of interactions in the analysis of variance (Darroch, 1962). Unexpectedly, the equivalence of (1.3) is invalid, a fact that has not been clarified in the statistics literature.

The goal of this study is to illustrate the fallacy of (1.3) using likelihood information identities. For three-way tables, the data information structure will be analyzed in Section 2, using a basic orthogonal decomposition in terms of the *mutual information*, also known as the *relative entropy* to engineers (Shannon, 1948), and, the *Kullback-Leibler divergence* (1951) to statisticians. It is shown that equation (1.3) is in conflict with the information identities illustrated in Section 2. The validity of the information identities is also examined against a data example, a $2 \times 2 \times 2$ contingency table (Snedecor, 1958). The calculations given in Sections 3 and 4 will illustrate the errors incurred when using (1.3) and (4.1), which led to invalid analyses in the classical hierarchical loglinear models.

Analysis of variance was designed to measure deviations from the means of continuous variables, which are expressed by partitioning sums of squares (Fisher, 1925). In contrast, analysis of categorical data measures deviations from uniform association, or independence, and varied interactions between the categorical variables, which are solely defined by the likelihood factorization. A basic likelihood information approach to categorical data analysis, coined linear information models, will provide a remedy, and this will be discussed in a forthcoming study.

2. Discrete Data Likelihood Identities

Let (X, Y) be a discrete random vector, where X and Y are categorical variables. Let $f(x_i, y_j)$ ($i = 1, \dots, I$ and $j = 1, 2, \dots, J$), be the joint probability density function (j.p.d.f.), and $f(x_i)$ and $g(y_j)$ are the marginal p.d.f. of X and Y , respectively. The Shannon entropy (1949) of the random variables is defined as

$$H(X) = - \sum_{x_i} f(x_i) \log f(x_i), \quad H(Y) = - \sum_{y_j} g(y_j) \log g(y_j),$$

and

$$H(X, Y) = - \sum_{(x_i, y_j)} f(x_i, y_j) \log f(x_i, y_j). \quad (2.1)$$

This is widely known as the foundation of information theory in engineering, physics, probability, and statistics (see for example, Gray, 1990). A well-known derivative of the Shannon entropy is the relative entropy, also known as the Kullback-Leibler (KL) divergence. By definition, (2.1) is equivalent to the basic equation:

$$H(X) + H(Y) = H(X, Y) + I(X; Y), \quad (2.2)$$

where $I(X; Y)$ defines the so-called *mutual information*:

$$I(X; Y) \equiv \sum_{i,j} f(x_i, y_j) \log[f(x_i, y_j) / f(x_i)g(y_j)]. \quad (2.3)$$

Essentially, identity (2.3) defines the minimum KL-divergence, from the joint p.d.f. of X and Y to the product space of marginal p.d.f., where X and Y are independent (cf. Cheng, Liou, Aston, and Arthur, 2006). The observed cell frequencies n_{ij} , the total counts of the vectors ($X = i, Y = j$), characterize the mutual information by the test statistic:

$$\hat{I}(X, Y) = \sum_{i=1}^I \sum_{j=1}^J n_{ij} \log\left(\frac{n_{ij}}{n_{..} \hat{f}(x_i, y_j)}\right) = \sum_{i=1}^I \sum_{j=1}^J n_{ij} \log\left(\frac{n_{ij} n_{..}}{n_{i.} n_{.j}}\right), \quad (2.4)$$

where $\hat{f}(x_i, y_j)$ is the MLE under the null hypothesis of independence. Twice the mutual information (2.4) is asymptotically chi-square distributed with $(I - 1)(J - 1)$ d.o.f., which is the same test for no interaction between X and Y . It is noteworthy that the likelihood ratio test statistic (2.4), being the sample analog of $I(X; Y)$, is an average of empirical log-likelihood; and therefore, the sample version of (2.2) does not include a constant term as the loglinear model does with standard ANOVA. This is the basic difference between the linear equation (2.2) and the saturated loglinear model in two variables.

Equations (2.2) and (2.3) admit obvious extensions to the general case with multiple variables. In particular, for the joint distribution of three variables, the analog of (2.2) is

$$H(X) + H(Y) + H(Z) = H(X, Y, Z) + I(X; Y; Z), \quad (2.5)$$

It is plain that equations (2.2) and (2.5) represent a fundamental linear information model. For $I \times J \times K$ contingency tables, it is analogous to (2.4) that twice the sample mutual information, the MLE $2\hat{I}(X; Y; Z)$, is asymptotically chi-square distributed with d.o.f. equal to $IJK - (I + J + K) + 2$, under the null hypothesis of mutual independence. In the sequel, the “hat” notation for the MLE, the sample mutual information, will be omitted for short.

Taking any two variables, say, (Y, Z) as a single component, a decomposition of (2.5) yields

$$I(X; Y; Z) = I(X; (Y, Z)) + I(Y; Z). \quad (2.6)$$

This initial step of information decomposition examines whether one variable, say, X , is independent of the pair (Y, Z) ; and the first term of (2.6) is tested against the chi-square distribution with $(I - 1)(JK - 1)$ d.o.f. Next, decompose this term to yield

$$I(X; (Y, Z)) = I(X; Z) + I(X; Y | Z) = I(X; Y) + I(X; Z | Y). \quad (2.7)$$

By (2.6) and (2.7), the mutual information between X , Y and Z is the sum of three orthogonal components. For example, the first sum of (2.7) yields the equation:

$$I(X;Y;Z) = I(X;Z) + I(Y;Z) + I(X;Y | Z). \quad (2.8)$$

The last term of (2.8) measures the conditional independence between X and Y across the variable Z , which is tested against a chi-square distribution with $(I-1)(J-1)K$ d.o.f. Clearly, the right-hand side of (2.8) can be expressed by three equivalent decompositions, using each variable as the conditioning variable. It is notable that the unique three-way interaction in a three-way table is included within a term of conditional mutual information, say, $I(X;Y | Z)$. That is,

$$I(X;Y | Z) = Int(X;Y;Z) + I(X;Y \parallel Z), \quad (2.9)$$

where $Int(X;Y;Z)$ measures the non-uniform association between X and Y , across Z , which is the unique three-way interaction with $(I-1)(J-1)(K-1)$ d.o.f.; and $I(X;Y \parallel Z)$ defines the uniform association between X and Y , that remains with respect to Z , with $(I-1)(J-1)$ d.o.f. The latter is commonly called the *partial association* between X and Y , given Z (for example, Birch, 1964; Goodman, 1969). There are obviously three different forms of (2.9), one for each conditioning variable.

It follows from equations (2.5) to (2.9) that a saturated linear model in the log-likelihood of three variables must be one of three information-equivalent models. Each model consists of three main (one-way) effects, two two-way (first-order) interaction effects, and a three-way (second-order) effect of conditional mutual information. By (2.9), a three-way effect consists of the unique interaction, and a conditional uniform association. Extension of this decomposition to multi-way contingency tables can be similarly formulated. An important message is conveyed: the present characterization of a saturated linear information model, based on identities (2.5) to (2.9), has not been discussed for use in the statistics literature, nor in the hierarchical loglinear models. Computational effects of this drawback will be illustrated using a classical data example in Section 3.

3. Discrete Data Information: An Example

Table 1 below is a three-way contingency table that had been analyzed by Mood (1950), and discussed by Snedecor (Table 1, 1958), which also included an ANOVA method by Lancaster (1951). It is a typical case of a three-way table and is used to illustrate the main findings of this study. The original notations of the three variables, $\{A, B, C\}$ will be used, instead of $\{X, Y, Z\}$ as used in Sections 1 and 2.

Table 1: A Three-Way Contingency Table

	C_1		C_2	
	B_1	B_2	B_1	B_2
A_1	79	62	73	168
A_2	177	121	81	75

A digest of Snedecor's discussion about Table 1 is useful for the present study: "In non-specific terms it may be said that the second-order interactions of Bartlett and Lancaster answer the same type of question, and that Mood tests combinations of the Lancaster interactions. The discrepancies observed have two sources: (1) the models used by the three investigators are different, and (2) different approximations to the chi-square distribution are employed."

Mood (1950) initiated the study of Table 1 and considered two cases: "completely random sampling" and "fixed totals in one classification, say, variable C ". In both cases, the measures of independence (no association) are estimated from the marginal variable totals. Thus, for the $2 \times 2 \times 2$ Table 1, it is the same in either case that (i) there are 4 degrees of freedom (d.o.f.) for testing whether all three variables $\{A, B, C\}$ are mutually independent; and (ii) there are 3 d.o.f. for testing whether a variable C is independent of the other two variables $\{A, B\}$. Accordingly, based on maximum likelihood estimation, Mood calculated the chi-squares of these two tests to be (i) $\chi_{4,M}^2 = 111.10$, and (ii) $\chi_{3,M}^2 = 86.72$, where the second subscript M refers to the author, Mood. Both chi-squares correspond to extremely small probabilities such that the independence hypotheses are rejected.

Lancaster (1951) analyzed an ANOVA approach to decomposing the chi-square of the mutual independence. For the present data with three variables, the decomposed chi-squares were calculated and listed in Table 2 of Snedecor (1958, p. 561). They are $\chi_{AB}^2 = 24.10$, $\chi_{BC}^2 = 31.80$, $\chi_{AC}^2 = 68.30$, and $\chi_{ABC}^2 = 7.80$, where each chi-square has 1 d.o.f., giving the total chi-square for mutual independence to be (i) $\chi_{4,L}^2 = 132.00$, and the measure of association between variables $\{A, B\}$ and C to be (ii) $\chi_{3,L}^2 = \chi_{AC}^2 + \chi_{BC}^2 + \chi_{ABC}^2 = 107.90$. These two chi-square values differ greatly from those calculated by Mood as given above, although insignificantly in probability. Here, the notation χ_{AB}^2 defines the chi-square of the two-way independence between the variables A and B ; and, an approximating likelihood-ratio test statistic to χ_{AB}^2 is twice the mutual information $I(A, B)$ defined in (2.3). Likewise, $\chi_{ABC}^2 \simeq \chi_{(I-1)(J-1)(K-1)}^2$, approximated by $2 \text{Int}(A; B; C)$, denotes the chi-square of the three-way interaction discussed by Bartlett (1935).

By using the classical method of "expected" frequency chi-squares, Snedecor gave his answers to Table 1 as (i) $\chi_{4,S}^2 = 132.00$, (ii) $\chi_{3,S}^2 = 93.73$, and the

three-way interaction $\chi_{ABC}^2 = 19.57$. Besides the difference in the two previous total chi-squares of (i) that were given by Mood and Lancaster, the three chi-square values of (ii) are all different. As stated above, Snedecor only argued that the discrepancies between the two sets of calculations given by Mood and Lancaster were attributable to using different loglinear models, but he did not provide explanations about the conditions of the hypotheses and models in use. In fact, it can be illustrated that these are direct calculations of chi-squares out of the original data, before identifying or using a fitted model.

The sizes of data cells of Table 1 were sufficiently large to warrant close approximations between the decomposed chi-squares and the corresponding likelihood ratio test statistics, namely, twice the KL divergence defined in Section 2. In accordance with equations (2.6) to (2.8), a complete list of the conditional and unconditional mutual information (KL divergence) is provided below. The two-way effects are

$$I(A; B) = 12.11, \quad I(B; C) = 16.01, \quad \text{and} \quad I(A; C) = 34.77. \quad (3.1)$$

The three-way conditional mutual information terms are

$$I(A; B \mid C) = 9.52, \quad I(B; C \mid A) = 13.41, \quad \text{and} \quad I(A; C \mid B) = 32.18. \quad (3.2)$$

The second-order (three-way) interaction, computed by the well-known method of *iterative proportional fitting* (IPF, Deming and Stephan, 1940), and the total mutual information are respectively

$$Int(A; B; C) \simeq (0.5) \chi_{ABC}^2 = 3.41, \quad \text{and} \quad I(A; B; C) = 60.30. \quad (3.3)$$

The corresponding chi-squares are *twice* the divergence values of (3.1) to (3.3). Since the early computations were carried out about a half century ago, it is not surprising that the current computations (3.1) to (3.3) could yield *different yet correct* answers. Specifically, the mutual independence is: (i) $\chi_{4, KL}^2 = 2I(A; B; C) = 120.60$, which sits between the two values, 111.10 and 132.00, calculated by Mood (1950) and Lancaster (1951), respectively. And, the measure of no association between variable C and $\{A, B\}$ is: (ii) $\chi_{3, KL}^2 = 2I(C; (A, B)) = 96.37$. The latter is obtained from the identity (2.6), which yields $I(C; (A, B)) = I(A; B; C) - I(A; B)$ according to (3.1) and (3.2). For the same Table 1, none of the KL-divergences (i) and (ii) were close to the corresponding values calculated by the three authors as stated above. It is worth noting that the values of three two-way effects in (3.1) are quite close to those given by Lancaster (1951). However, by (3.3), the correct three-way interaction effect is found to be $\chi_{ABC}^2 = 6.82$, with p value close to 0.009, which differs from Lancaster's $\chi_{ABC}^2 = 7.80$, with $p \simeq 0.005$.

To confirm that the values calculated from (3.1) to (3.3) are correct and reliable, one can simply check that the computations of all the three equally symmetric forms of (2.6) to (2.8) are equal. In this case, it is checked that the three differences between the corresponding terms of (3.1) and (3.2) are equal, that is,

$$I(A; B) - I(A; B | C) = I(B; C) - I(B; C | A) = I(A; C) - I(A; C | B) = 2.59.$$

4. A Fallacy of Hierarchical Loglinear Models

It is well known that hierarchical analyses of the loglinear models in three variables could accommodate eight distinct parameters in a saturated model. Taking away the constant mean, it defines the decomposition of chi-squares to be attributed to three main effects, three two-way effects, and the unique three-way interaction. Although the chi-square decomposition in the multi-way contingency tables had been studied long before statistical inferences were discussed for the loglinear models, it is surprising that the inconclusive and varied calculations of the partitioned chi-squares, as exemplified in Section 3, were not rigorously analyzed. As a consequence, equation (1.3) given by Roy and Kastenbaum (1956) led to the following development of a likelihood ratio test.

Darroch (1962, p. 260) derived a likelihood-ratio test for the three-way interaction effect based on the identity:

$$Z_{ABC} = Z_I - Z_{AB} - Z_{BC} - Z_{CA}, \quad (4.1)$$

where $Z_{ABC} = 2 \text{Int}(A; B; C)$ is the three-way interaction, $Z_I = 2I(A; B; C)$ is the joint mutual information of (2.6), and $Z_{AB} = 2I(A; B)$, Z_{BC} and Z_{CA} are the obvious two-way effects. As a matter of fact, (4.1) was obtained from (1.3) and (2.5), based on the argument that the partition bears an obvious analogy to the ANOVA of the term $\chi_{3,L}^2$ (cf. Section 3) discussed by Lancaster (1951). Thus, on heuristic grounds, a test for no three-way interaction can be based on $Z_{ABC} \simeq \chi_{(I-1)(J-1)(K-1)}^2$ (Darroch, 1962, p. 259). Equivalently, (4.1) asserts that testing zero three-way interaction is equivalent to testing that the three two-way effects altogether is equal to the effect of mutual independence among the three variables. This is simply not true, because (1.3) is incorrect, a priori, in view of the information identities (2.7) to (2.9).

For numerical evidence of the invalidity of (4.1), hence (1.3), a data example is observed from Table 1 of Section 3. The computation using equations (3.1) to (3.3) evaluates the left-hand side of (4.1) to be $Z_{ABC} = 6.82$. And, the right-hand side of (4.1) is found to be $2(60.30 - 12.11 - 16.01 - 34.77) = -5.18 < 0$. The two sides of (4.1) are unequal, and the illogical negative value is obtained, because

(4.1) violates the identity (2.8), which disallows the inclusion of all the two-way effects within the likelihood information decomposition.

There is however a useful interpretation: the right-hand side of (4.1) can be expressed as an entropy identity,

$$\begin{aligned} & H(X) + H(Y) + H(Z) - H(X, Y) - H(Y, Z) - H(Z, X) + H(X, Y, Z) \\ = & I(X; Y) + I(Y; Z) + I(Z; X) - I(X; Y; Z). \end{aligned} \quad (4.2)$$

The left-hand side of equation (4.2) is seemingly a Venn diagram in terms of entropy, but its value need not be nonnegative as illustrated above. Nevertheless, an ideal Venn diagram can be formulated in terms of relative entropy. It is actually an equivalent identity to equation (4.2):

$$\begin{aligned} & I(X) + I(Y) + I(Z) - I(X; Y) - I(Y; Z) - I(Z; X) + I(X; Y; Z) \\ = & H(X, Y) + H(Y, Z) + H(Z, X) - H(X, Y, Z) \geq 0, \end{aligned} \quad (4.3)$$

where $I(X) = H(X)$ is commonly defined for notational convenience.

Since 1962, the current finding with formulae (1.3) and (4.1) has been overlooked in the statistics literature, as is not found in the studies of Lewis (1962, p.100), Birch (1963, p.223), Goodman (1969, 1978), Bishop, Fienberg, and Holland (1975), Hagenaars (1993), Christensen (1997), and Agresti (2002). In the elementary analysis of a 2×2 table, it is also remarkable that partition of chi-squares (associated with three degrees of freedom) could only yield equations in distribution by approximation (Kendall and Stuart, 1978, (33.117); Rao, 1973, (6d.2.6)), while exact equations are given in terms of relative entropy. The task of examining the deviation from the mean in the classical ANOVA is intrinsically different from testing association between categorical variables. It is clarified in this study that information identities (2.5) to (2.9) provide a key to the analysis of loglinear models. This introduces the *linear information models* which will be presented in a forthcoming study.

Acknowledgments

The authors are grateful for the discussion of Section 2 with the Brain Imaging Group, Institute of Statistical Science, Academia Sinica, on March 18, 2005; and also, for the follow-up discussions with Dr. M. T. Chao.

References

- Agresti, A. (2002). *Categorical Data Analysis, 2nd ed.* Wiley.

- Bartlett, M. S. (1935). Contingency table interactions. *J. Roy. Statist. Soc. B* **2**, 248-252.
- Birch, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc. B* **25**, 220-233.
- Birch, M. W. (1964). The detection of partial association. *J. Roy. Statist. Soc. B* **26**, 313-324.
- Bishop, Y. M. M., Fienberg, S. E. and Holland. P. W. (1975). *Discrete Multivariate Analysis*. MIT Press.
- Blyth, C. R. (1972). On Simpson's paradox and the sure thing principle. *J. Amer. Statist. Assoc.* **67**, 364-366.
- Cheng, P. E., Liou, M., Aston, J. A. D. and Tsai, Arthur C. (2006). Information identities and testing hypotheses: Power analysis for contingency tables. *Statistica Sinica*, tentatively accepted.
- Christensen, R. (1997). *Loglinear Models and Logistic Regression*, Springer-Verlag.
- Claringbold, P. J. (1961). The use of orthogonal polynomials in the partition of chi-square. *Aust. J. Statist.* **3**, 48-63.
- Cochran, W. G. (1954). Some methods for strengthening the common chi-square tests. *Biometrics* **24**, 315-327.
- Darroch, J. N. (1962). Interactions in multifactor contingency tables. *J. Roy. Statist. Soc. B* **24**, 251-263.
- Deming, W. E. and Stephan F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* **11**, 427-444.
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd.
- Gart, J. J. (1962). On the combination of relative risks. *Biometrics* **18**, 601-610.
- Goodman, L. A. (1964). Simple methods for analyzing three-factor interaction in contingency tables. *J. Amer. Statist. Assoc.* **59**, 319-352.
- Goodman, L. A. (1969). On partitioning χ^2 and detecting partial association in three-way contingency tables. *J. Roy. Statist. Soc. B* **31**, 486-498.
- Goodman, L. A. (1978). *Analyzing Qualitative/Categorical Data*. Abt Associates Inc.
- Gray, R. M. (1990). *Entropy and information theory*. Springer-Verlag.
- Hagenaars, J. A. (1993). *Loglinear Models with Latent Variables*. Sage.
- Kendall, M. G. and Stuart, A. (1979). *The Advanced Theory of Statistics. 4th Edition*, Vol. 2, Charles Griffin.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.* **22**, 79-86.

- Kullback, S. (1959). *Information Theory and Statistics*. Wiley.
- Lancaster, H. O. (1951). Complex contingency tables treated by the partition of chi-square. *J. Roy. Statist. Soc. B* **13**, 242-249.
- Lewis, B. N. (1962). On the analysis of interaction in multi-dimensional contingency tables. *J. Roy. Statist. Soc. A* **125**, 88-117.
- Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst.* **22**, 719-748.
- Mood, A. M. (1950). *Introduction to the Theory of Statistics*. McGraw-Hill.
- Norton, H. W. (1945). Calculation of chi-square for complex contingency tables. *J. Amer. Statist. Assoc.* **40**, 251-258.
- Plackett, R. L. (1962). A note on interactions in contingency tables. *J. Roy. Statist. Soc. B* **24**, 162-166.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. 2nd edition, Wiley.
- Roy, S. N. and Kastenbaum, M. A. (1956). On the hypothesis of no "interaction" in a multi-way contingency table. *Ann. Math. Statist.* **27**, 749-757.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Sys. Tech. Journal* **27**, 379-423; 623-656.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *J. Roy. Statist. Soc. B* **13**, 238-241.
- Snedecor, G. W. (1958). Chi-squares of Bartlett, Mood, and Lancaster in a 2^3 contingency table. *Biometrics* **14**, 560-562.
- Wilks, S. S. (1935). The likelihood test of independence in contingency tables. *Ann. Math. Statist.* **6**, 190-196.
- Woolf, B. (1955). On estimating the relation between blood group and disease. *Ann. Human Genetics* **19**, 251-253.

Received March 25, 2005; accepted December 6, 2005.

Philip E. Cheng
Institute of Statistical Science
Academia Sinica
Taipei, 11529, Taiwan
pcheng@stat.sinica.edu.tw

Jiun W. Liou
Institute of Statistical Science
Academia Sinica
Taipei, 11529, Taiwan
needgen@stat.sinica.edu.tw

Michelle Liou
Institute of Statistical Science
Academia Sinica
Taipei, 11529, Taiwan
mliou@stat.sinica.edu.tw

John A. D. Aston
Institute of Statistical Science
Academia Sinica
Taipei, 11529, Taiwan
jaston@stat.sinica.edu.tw