Geometry of Mutual Information in Three-Way Contingency Tables

Philip E. Cheng, and **Michelle Liou**, Academia Sinica, **John A.D. Aston**†, University of Warwick

1 The $2 \times 2 \times K$ Contingency Table - Setup and Notation

- $(X,\ Y,\ Z)$ denote the three-way categorical vector,
- $(X_k,\ Y_k)$ denote pairs of dichotomous variables, where Z is the K-level $(k = 1,\ \dots,\ K)$ stratum variable.
- observed data are frequency counts n_{ijk} of subjects having condition i, ($i = 1$ (case), 2 (control)), and exposure j ($j = 1$ (exposed), 2 (non-exposed)), which fall in stratum k, $k = 1,\ \dots,\ K$.
- $U = \{U_k = (n_{11k},\ n_{12k};\ n_{21k},\ n_{22k}),\ k = 1,\ \dots,\ K,\}$ denote the observed K strata of 2×2 tables.

A dot notation will be used for summation over a subscript, say, $n_{\cdots} = n$ denotes the total sample size, $n_{1\cdot k}$ is the number of cases in stratum k, and $n_{\cdot 2k}$ is the total number of non-exposed subjects in stratum k, and so on.

	Z=0		Z=1	
	Y=0	Y=1	Y=0	Y=1
X=0	n_{000}	n_{010}	n_{001}	n_{011}
X=1	n_{100}	n_{110}	n_{101}	n_{111}

Table 1: An example of a $2 \times 2 \times 2$ contingency table

2 Testing Hypotheses

Let the odds ratios of the 2×2 tables be defined by $\psi_k = p_{11k}p_{22k} / p_{12k}p_{21k}$, $k = 1,\ \dots,\ K$, where $p_{ijk} = P(X = i, Y = j, Z = k)$, $i, j = 1$ or 2, are the cell proportions.

- Conditional Independence

$$H_0:\ \psi_k = 1,\ \text{for } k \in \{1,\ \dots,\ K\}. \tag{1}$$

- Common Odds Ratio (COR)

$$H_1:\ \psi_k = \psi,\quad \text{for } k \in \{1,\ \dots,\ K\}, \tag{2}$$

for a positive constant ψ.

- Uniform Association

Given a COR ψ,

$$H_2: \psi = 1, \tag{3}$$

As can be seen $H_2 = (H_0|H_1)$.

3 Classical Tests

- H_0 - the Pearson chi-square test

$$\chi^2_{PE} = \sum_{k=1}^{K}\sum_{i,j=1}^{2} \frac{(n_{ijk} - n_{i\cdot k}n_{\cdot jk} / n_{\cdot\cdot k})^2}{n_{i\cdot k}n_{\cdot jk} / n_{\cdot\cdot k}}. \tag{4}$$

It approximates the chi-square distribution with K d.f., denoted χ^2_K.

- H_1 - Breslow-Day test

$$\chi^2_{BD} = \sum_k \frac{e_k^2}{var(n_{11k}|\psi_{MH})}. \tag{5}$$

where the adjusted cell estimates e_k and the denominator variance can easily be found (e.g., Agresti 2002, p. 232), with

$$\psi_{MH} = \frac{\sum_{k=1}^{K}(n_{11k}n_{22k} / n_{\cdot\cdot k})}{\sum_{k=1}^{K}(n_{12k}n_{21k} / n_{\cdot\cdot k})}. \tag{6}$$

The B-D test approximates the chi-square distribution with $K - 1$ d.f.

- H_2 - Cochran-Mantel-Haenszel test, often wrongly believed to test H_0

$$\chi^2_{CMH} = \frac{\left(\sum_{k=1}^{K} n_{11k} - \sum_{k=1}^{K} n_{1\cdot k}n_{\cdot 1k} / n_{\cdot\cdot k}\right)^2}{\sum_{k=1}^{K}\{n_{1\cdot k}n_{2\cdot k}n_{\cdot 1k}n_{\cdot 2k} / n^2_{\cdot\cdot k}(n_{\cdot\cdot k} - 1)\}}. \tag{7}$$

The CMH test approximates the chi-square distribution with 1 d.f.

4 Information Identity

- $(X,\ Y,\ Z)$ be the variables of a three-way $I \times J \times K$ contingency table.
- $f(i, j, k) = P(X = i,\ Y = j,\ Z = k)$, $f(i)$, $g(j)$, $h(k)$; $i = 1,\ \dots,\ I$, $j = 1,\ \dots,\ J$, $k = 1,\ \dots,\ K$, denote the joint and marginal probability density functions (p.d.f.).

$$H(X) + H(Y) + H(Z) = I(X,\ Y,\ Z) + H(X,\ Y,\ Z), \tag{8}$$

where

- $H(X,\ Y,\ Z) = -\sum_{(i,j,k)} f(i, j, k) \cdot \log f(i, j, k)$ is the joint entropy, and marginal entropies
- $I(X,\ Y,\ Z) = \sum_{(i,j,k)} f(i, j, k) \cdot \log\{f(i, j, k)/f(i)g(j)h(k)\}$ denotes the mutual information between the three variables.

Furthermore, $I(X,\ Y,\ Z)$ admits three equivalent expressions

$$\begin{aligned}\log\left\{\frac{f(i,j,k)}{f(i)g(j)h(k)}\right\} &= \log\left\{\frac{f(i,k)}{f(i)h(k)}\right\} + \log\left\{\frac{f(i,j,k)}{f(i,k)g(j)}\right\} \\ &= \log\left\{\frac{f(i,k)}{f(i)h(k)}\right\} + \log\left\{\frac{f(j,k)}{g(j)h(k)}\right\} \\ &\quad + \log\left\{\frac{f(i,j,k)/h(k)}{f(i|k)f(j|k)}\right\},\end{aligned} \tag{9}$$

where convenient notations $f(i, j)$ and $f(i|j)$ are used to denote j.p.d.f. and conditional p.d.f., respectively.

By taking expectations of the sampling versions of both sides of the above, an orthogonal decomposition of the mutual information using Z as the (common) conditioning variable (CV) is expressed as

$$I(X,\ Y,\ Z) = I(X,\ Z) + I(Y,\ Z) + I(X,\ Y \mid Z). \tag{10}$$

$$I(X,\ Y \mid Z) = Int(X,\ Y,\ Z) + I(X,\ Y \parallel Z). \tag{11}$$

The first summand $Int(X,\ Y,\ Z)$ on the r.h.s. of (11) defines the three-way interaction between X and Y, across Z

5 Likelihood Ratio Tests

Let the conditional MLE under H_0 be denoted by $W_k = (n^*_{11k},\ n^*_{12k};\ n^*_{21k},\ n^*_{22k})$, $k = 1,\ \dots,\ K$, where $n^*_{ijk} = n_{i\cdot k}n_{\cdot jk}/n_{\cdot\cdot k}$ are the conditional MLEs of the cell proportions given the margins, which are the sufficient statistics, of each 2×2 table.

The first term on the r.h.s. of (11) characterizes the conditional MLE under H_1 by $V = \{V_k = (\hat{n}_{11k},\ \hat{n}_{12k};\ \hat{n}_{21k},\ \hat{n}_{22k}),\ k = 1,\ \dots,\ K\}$, which can be computed by the IPF (Deming and Stephan, 1940) scheme.

- H_0:

$$D_0 = 2D(U : W) = 2\sum_{k=1}^{K}\sum_{i=1}^{2}\sum_{j=1}^{2} n_{ijk}\log(n_{ijk} / n^*_{ijk}) \cong \chi^2_K(H_0). \tag{12}$$

- H_1:

$$D_1 = 2D(U : V) = 2\sum_{k=1}^{K}\sum_{j}\sum_{i} n_{ijk}\log(n_{ijk} / \hat{n}_{ijk}) \cong \chi^2_{K-1}(H_1). \tag{13}$$

- H_2:

$$D_2 = 2D(V : W) = 2\sum_{k=1}^{K}\sum_{j}\sum_{i} \hat{n}_{ijk}\log(\hat{n}_{ijk} / n^*_{ijk}) \cong \chi^2_1(H_0 \mid H_1), \tag{14}$$

6 Power Analysis for LR Tests

Theorem 1. *Let U be a $2 \times 2 \times K$ table. Let $W' \in H'$ be another $2 \times 2 \times K$ table, having the same table totals as those of U, sample odds ratios $(\psi_1,\ \dots,\ \psi_K)$, and consecutive three-way sample interactions $1 \neq \gamma_i = \psi_i / \psi_{i+1} > 0$, $i = 1,\ \dots,\ K-1$. Then, there is a unique $2 \times 2 \times K$ table V', $V' \in H'_1$, having the same table margins as those of U, such that the following holds*

$$D(U : W') = D(U : V') + D(V' : W'). \tag{15}$$

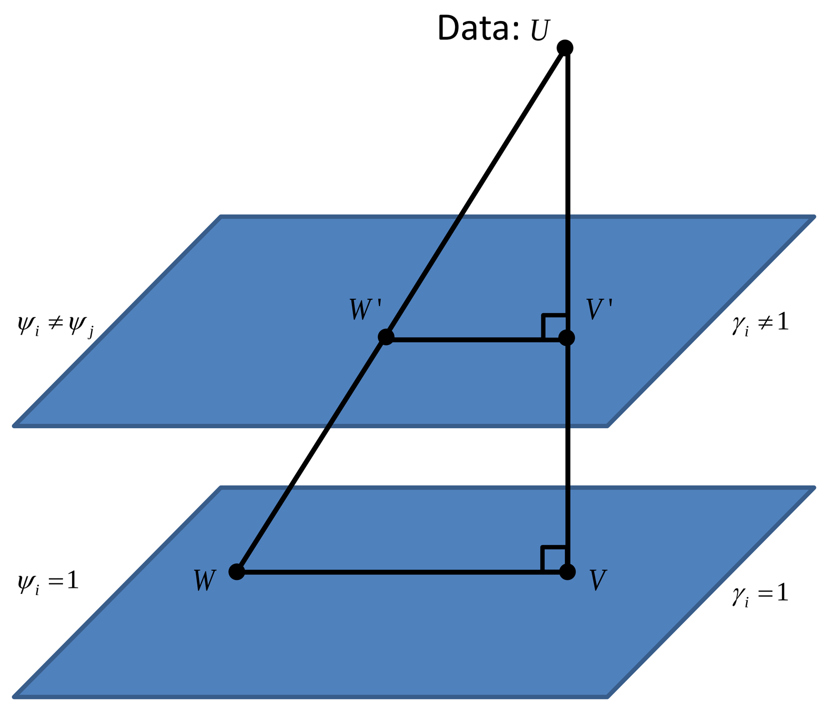

Figure 1: Null Hypotheses: $D(U : W) = 0 = D(U : V) + D(V : W)$, $\gamma_i = 1$;Alternative Hypotheses: $D(U : W') = 0 = D(U : V') + D(V' : W')$, $\gamma_i \neq 1$.

Corollary 2. *For $K = 2$, the statistic $D(U : V')$ tests for a specific value of the interaction parameter $\gamma(\neq 1)$, and provides an interval estimation for the parameter γ of the observed data U.*

7 An Example

	Poland		U.S.	
Allele freq.\Genotype	C	G	C	G
Case	62	419	48	447
Control	92	371	51	445

Table 2: Data

Data of two 2×2 tables are genotypes and allele frequencies for certain polymorphisms in the Polish and U.S. samples. (Ardlie, et al. 2002, Table 2).

The authors' analysis:

- Sample odds ratios 0.597 and 0.937 for the two tables
- COR estimate $\psi_{MH} = 0.719$, with a 95% confidence interval (0.60, 0.87).
- CMH test yields $\chi^2_{CMH} = 5.88$ with $p = 0.015$ (or $\chi^2_{MH} = 5.56$ with $p = 0.018$)

Authors' conclusion: **"the two odds ratios are different"**.

- $D_0 = 8.55$ with $p = 0.014$, $K = 2$ d.f.
- $D_1 = 2.646$ with $p = 0.104$, and the conditional MLE $\hat{\psi} = 0.718$; further, $\psi_{MH} = 0.719$ and $\chi^2_{BD} = 2.653$, $p = 0.103$.
- $D_2 = 5.905$ with $p = 0.015$, which is significant at level $\alpha_2 \approx \alpha/2 = 0.025$.

Conclusion: **There is evidence that the odds ratios differ from one, but no evidence that they differ from each other.**

References

A Agresti. (2002) Categorical Data Analysis, New Jersey: Wiley

KC Ardlie, KL Lunetta and M Seielstad (2002 *Am. J. Hum. Genet.*, 71, 304-311.

PE Cheng, M Liou and JAD Aston. (2010) Likelihood Ratio Tests in Three-Way Tables, *JASA*, in press.

WE Deming and FF Stephan. (1940) *Ann. Math. Statist.*, 11, 427-444.