Journal of the Korean Statistical Society (1998), **27**: 2, pp 205–220

Tests For and Against a Positive Dependence Restriction in Two-Way Ordered Contingency Tables

Myongsik Oh[1]

ABSTRACT

Dependence concepts for ordered two-way contingency tables have been of considerable interest. We consider a dependence concept which is less restrictive than likelihood ratio dependence and more restrictive than regression dependence. Maximum likelihood estimation of cell probability under this dependence restriction is studied. The likelihood ratio statistics for and against this dependence are proposed and their large sample distributions are derived. A real data is analyzed to illustrate the estimation and testing procedures.

Keywords: Chi-bar-square distribution; least favorable distribution; level probability; likelihood ratio test; uniform conditional stochastic ordering

1. INTRODUCTION

Dependence concepts for ordered two-way contingency tables have received considerable interest since Lehmann (1966) first introduced three types of dependence concepts namely, quadrant, regression, and likelihood ratio dependence. Douglas *et al.* (1990) reinterpreted dependence concepts in terms of various classes of odds ratios, which include local odds ratio, local-global odds ratio and global odds ratio. Requiring a collection of each of three classes of odds ratios to be bigger than or equal to 1 corresponds to positive likelihood ratio dependence, positive regression dependence and positive quadrant dependence, respectively.

Statistical inference concerning these dependence concepts in two-way ordered contingency tables have been studied widely. For positive likelihood ratio dependence, which is also called total positivity of order 2 or trend, several tests have been proposed. Cohen and Sackrowitz (1991), Hirotsu (1982), Lee (1990) and Patefield (1982) are among others. Nguyen and Sampson (1987) studied testing problem for positive quadrant dependence in ordinal contingency tables.

[1]Department of Statistics, Pusan University of Foreign Studies, Pusan, 608-738, Korea.

Unfortunately, no explicit forms of maximum likelihood estimates of cell probabilities under each of three types of positive dependence restrictions have been found yet except the case either r or c are 2. Readers may refer Oh (1995) for detailed estimation procedures with various sampling schemes for the latter case. Dykstra and Lemke (1988) studied iterative algorithm for finding maximum likelihood estimate under likelihood ratio ordering. Feltz and Dykstra (1985) studied iterative method for finding maximum likelihood estimate under regression dependence for several survival functions. This algorithm can be used for finding maximum likelihood estimate of cell frequencies under positive regression dependence restriction. We are not aware that any iterative algorithm for finding maximum likelihood estimate under positive quadrant dependence restriction has been already developed.

Consider a type of dependence concept which is less restrictive than likelihood ratio dependence and more restrictive than regression dependence. As it will be explained later the benefit of using this dependence concept is mainly due to that a closed form of maximum likelihood estimate under the restriction is available. To illustrate this dependence concept we let $p_{ij} > 0$, $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, c$ and $\sum_{ij} p_{ij} = 1$. Consider the following restriction. For $i = 1, \ldots, r-1$, $j = 1, \ldots, c-1$,

$$\frac{p_{ij} \cdot \sum_{\ell=i+1}^{r} p_{\ell,j+1}}{p_{i,j+1} \cdot \sum_{\ell=i+1}^{r} p_{\ell j}} \geq 1. \tag{1.1}$$

The cross product ratio in (1.1), which we will call continuation odds ratio, is based on continuation-ratio logit models. It is not difficult to show that (1.1) is equivalent to

$$\frac{\sum_{\ell=i}^{r} p_{\ell j}}{\sum_{\ell=i}^{r} p_{\ell,j+1}} \text{ is nondecreasing in } i \text{ for each } j = 1, \ldots, c-1. \tag{1.2}$$

Then restriction (1.2) can be interpreted as follows. The ratio of two survival functions of the conditional distributions of columns j and $j+1$ is nondecreasing in i. This is the so-called uniform stochastic ordering. We say that the conditional distribution of column j is uniformly stochastically greater than the conditional distribution of column $j+1$. In this sense, we call restriction (1.2) uniform conditional stochastic ordering. Shaked (1977) refers to this restriction as conditional hazard rate decreasing.

In this paper we are going to discuss statistical inference procedures under (1.1). Section 2 deals with the maximum likelihood estimation under restriction

(1.1) and the strong consistency of the estimators. In Section 3, we consider testing problems. The likelihood ratio statistic for testing independence against restrictions (1.1), where the independence hypothesis is reintepreted as

$$H_0 : \frac{p_{ij} \cdot \sum_{\ell=i+1}^{r} p_{\ell,j+1}}{p_{i,j+1} \cdot \sum_{\ell=i+1}^{r} p_{\ell j}} = 1, \text{ for } i = 1, \dots, r-1, j = 1, \dots, c-1. \quad (1.3)$$

We also consider likelihood ratio test of restriction (1.1) against all alternatives. The main results from Section 3 are the asymptotic null distributions of two likelihood ratio statistics and their upper bounds of the right taill probabilities, which we call least favorable configurations. In section 4, a real data is analyzed in order to illustrate the estimation and test procedures.

2. MAXIMUM LIKELIHOOD ESTIMATION

Suppose $p_{ij} > 0, i = 1, 2, \dots, r, j = 1, 2, \dots, c$ and $\sum_{ij} p_{ij} = 1$. Let n be the total count, i.e., sample size and $\hat{p}_{ij}$ be the relative frequency of an event having probability p_{ij}. Then the likelihood function is proportional to $\prod_{i=1}^{r} \prod_{j=1}^{c} {p_{ij}}^{n\hat{p}_{ij}}$. It is convenient to use a one-to-one transformation of the parameter space by setting

$$\theta_{ij} = \frac{p_{ij}}{\sum_{\ell=i}^{r} p_{\ell j}}, \phi_j = \sum_{\ell=1}^{r} p_{\ell j}, \text{ for } i = 1, \dots, r-1, j = 1, \dots, c.$$

Then $p_{1j} = \theta_{1j}\phi_j$, $p_{ij} = \phi_j \theta_{ij} \prod_{\ell=1}^{i-1}(1 - \theta_{\ell j}), i = 2, \dots, r-1$, $p_{rj} = \phi_j \prod_{\ell=1}^{r-1}(1 - \theta_{\ell j})$ for $j = 1, 2, \dots, c$. The basic restriction becomes

$$0 < \theta_{ij} < 1, 0 < \phi_j < 1, i = 1, \dots, r-1, j = 1, \dots, c, \text{ and } \sum_{j=1}^{c} \phi_j = 1.$$

It is not difficult to show that (1.3) is equivalent to $H_0 : \ \theta_{i1} = \theta_{i2} = \cdots = \theta_{ic}$ for $i = 1, 2, \dots, r-1$ and (1.1) is equivalent to

$$H_1 : \ \theta_{i1} \geq \theta_{i2} \geq \cdots \geq \theta_{ic} \text{ for } i = 1, 2, \dots, r-1. \quad (2.1)$$

Let $\succeq$ be the partial order on $\{1, 2, \dots, c\}$ defined by $i \succeq j$ if $1 \leq i < j \leq c$. Then (2.1) is equivalent to that $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ic})$ is isotonic with respect to $\succeq$ for $i = 1, 2, \dots, r-1$.

The likelihood function, in terms of θ_{ij} and ϕ_j, is proportional to

$$\prod_{i=1}^{r-1}\left[\prod_{j=1}^{c}\theta_{ij}{}^{n\hat{p}_{ij}}(1-\theta_{ij})^{n\sum_{\ell=i+1}^{r}\hat{p}_{\ell j}}\right]\cdot\prod_{j=1}^{c}\phi_j{}^{n\sum_{i=1}^{r}\hat{p}_{ij}}. \tag{2.2}$$

We note that (2.2) is a product of two components; one is a function of only θ_{ij}'s and the other is a function of ϕ_j's. The inside the brackets can be viewed as a products of c independent binomial likelihood functions. The second part is a multinomial likelihood function.

The maximum likelihood estimates of θ_{ij}'s and ϕ_j's under H_0 are given by θ_{ij}° and ϕ_j°, where

$$\begin{aligned}\theta_{i1}^{\circ} &= \theta_{i2}^{\circ}=\cdots=\theta_{ic}^{\circ}=\frac{\sum_{j=1}^{c}\hat{p}_{ij}}{\sum_{\ell=i}^{r}\sum_{j=1}^{c}\hat{p}_{\ell j}}, i=1,\ldots,r-1,\\ \phi_j^{\circ} &= \sum_{i=1}^{r}\hat{p}_{ij}, j=1,2,\ldots,c.\end{aligned}$$

We now consider maximum likelihood estimation under (2.1). Since (2.1) does not involve ϕ_j's, we can maximize (2.2) by maximizing the two parts separately. The second part is just a multinomial problem since no restriction except basic restriction is imposed. The maximum likelihood estimate, ϕ_j^*, of ϕ_j is equal to ϕ_j°.

Next we consider estimation of θ_{ij}'s under (2.1). Since (2.1) does not relate θ_{ij}'s for different values of i, the maximum of the first part in (2.2) can be achieved by maximizing $r-1$ binomial likelihood functions independently. Let $\hat{\boldsymbol{\theta}}_i$ and $\hat{\mathbf{p}}_i$ be sample statistics for $\boldsymbol{\theta}_i$ and $\mathbf{p}_i=(\sum_{\ell=i}^{r}p_{\ell 1},\sum_{\ell=i}^{r}p_{\ell 2},\cdots,\sum_{\ell=i}^{r}p_{\ell c})$, respectively. As in Example 1.5.1 of Robertson, Wright and Dykstra (1988), the maximum likelihood estimate of $\boldsymbol{\theta}_i$ under (2.1) is the isotonic regression of $\hat{\boldsymbol{\theta}}_i$ onto the closed convex cone $\mathcal{I}=\{\mathbf{x}\in\mathbf{R}^c : x_1\geq x_2\geq\cdots\geq x_c\}$ with the weights $\hat{\mathbf{p}}_i$, which we will denote by $E_{\hat{\mathbf{p}}_i}\left(\hat{\boldsymbol{\theta}}_i|\mathcal{I}\right)$, where jth component is given by

$$E_{\hat{\mathbf{p}}_i}\left(\hat{\boldsymbol{\theta}}_i|\mathcal{I}\right)_j=\max_{1\leq\alpha\leq j}\min_{j\leq\beta\leq k}\frac{\sum_{\ell=\alpha}^{\beta}\hat{p}_{i\ell}}{\sum_{\ell=\alpha}^{\beta}\sum_{k=i}^{r}\hat{p}_{k\ell}}.$$

This leads to the following theorem.

Theorem 2.1. *The maximum likelihood estimate of* $\mathbf{p}$ *under restriction (1.1) is given by* $\mathbf{p}^*$, *where* $\mathbf{p}^*$ *is obtained by evaluating* $\mathbf{p}$ *at* $\boldsymbol{\theta}_i=\boldsymbol{\theta}_i^*$ *and* $\phi_j=\phi_j^*$ *for* $i=1,2,\ldots,r-1, j=1,2,\ldots,c$ *with*

$$\boldsymbol{\theta}_i^*=E_{\hat{\mathbf{p}}_i}\left(\hat{\boldsymbol{\theta}}_i|\mathcal{I}\right).$$

For computation of $E_{\hat{\mathbf{p}}_i}\left(\hat{\boldsymbol{\theta}}_i|\mathcal{I}\right)$, the pool-adjacent-violators algorithm (PAVA) can be used. See page 8-11 of Robertson *et al.* (1988) for the description of the PAVA and an example.

Next we show that the maximum likelihood estimate under H_1 is strongly consistent. Corollary to Theorem 1.4.4 of Robertson *et al.* (1988) shows that the projection operator $E_{\mathbf{w}}(\mathbf{x}|\mathcal{I})$ is continuous with respect to both of $\mathbf{w}$ and $\mathbf{x}$. Now it follows from the strong law of large numbers that the restricted maximum likelihood estimate, $\mathbf{p}^*$, converges to $\mathbf{p}$ with probability one if H_1 is true.

Theorem 2.2. *The maximum likelihood estimator,* $\mathbf{p}^*$*, of* $\mathbf{p}$ *under* H_1 *is a strongly consistent estimator, i.e.,* $Pr[\lim_{n\to\infty}\mathbf{p}^* = \mathbf{p}] = 1$.

3. LIKELIHOOD RATIO TESTS

3.1. Test of Independence Against Dependence

Consider first the problem of testing the null hypothesis H_0 against the alternative hypothesis H_1. The likelihood ratio statistic is

$$\Lambda_{01} = \frac{\sup_{\mathbf{p}\in H_0} L(\mathbf{p})}{\sup_{\mathbf{p}\in H_1} L(\mathbf{p})} = \frac{\prod_{i=1}^{r-1}\left[\prod_{j=1}^{c}(\theta_{ij}^{\circ})^{n\hat{p}_{ij}}(1-\theta_{ij}^{\circ})^{n\sum_{\ell=i+1}^{r}\hat{p}_{\ell j}}\right]}{\prod_{i=1}^{r-1}\left[\prod_{j=1}^{c}(\theta_{ij}^{*})^{n\hat{p}_{ij}}(1-\theta_{ij}^{*})^{n\sum_{\ell=i+1}^{r}\hat{p}_{\ell j}}\right]}.$$

The test rejects H_0 for large value of $T_{01} = -2\ln\Lambda_{01}$, i.e.,

$$\begin{aligned} T_{01} &= 2n\sum_{i=1}^{r-1}\left[\sum_{j=1}^{c}\hat{p}_{ij}(\ln\theta_{ij}^{*} - \ln\theta_{ij}^{\circ}) + \right. \\ &\qquad \left. \sum_{j=1}^{c}\Big(\sum_{\ell=i+1}^{r}\hat{p}_{\ell j}\Big)(\ln(1-\theta_{ij}^{*}) - \ln(1-\theta_{ij}^{\circ}))\right]. \end{aligned} \tag{3.1}$$

Next we find the asymptotic distribution of T_{01} under H_0. Expanding $\ln\theta_{ij}^{*}$ and $\ln\theta_{ij}^{\circ}$ about $\hat{\theta}_{ij}$ and $\ln(1-\theta_{ij}^{*})$ and $\ln(1-\theta_{ij}^{\circ})$ about $1-\hat{\theta}_{ij}$ and using properties of isotonic regression (see Theorem 1.3.2 of Robertson *et al.* 1988), (3.1) is rewritten as

$$n\sum_{i=1}^{r-1}\left[\sum_{j=1}^{c}(\theta_{ij}^{\circ}-\hat{\theta}_{ij})^2\left(\frac{\hat{p}_{ij}}{\beta_{ij}^2} + \frac{\sum_{\ell=i+1}^{r}\hat{p}_{ij}}{\delta_{ij}^2}\right) - \sum_{j=1}^{c}(\theta_{ij}^{*}-\hat{\theta}_{ij})^2\left(\frac{\hat{p}_{ij}}{\alpha_{ij}^2} + \frac{\sum_{\ell=i+1}^{r}\hat{p}_{ij}}{\gamma_{ij}^2}\right)\right]$$

where α_{ij} (β_{ij}) is between $\hat{\theta}_{ij}$ and θ^*_{ij} ($\hat{\theta}_{ij}$ and θ°_{ij}) and γ^*_{ij} (δ_{ij}) is between $1-\hat{\theta}_{ij}$ and $1-\theta^*_{ij}$ ($1-\hat{\theta}_{ij}$ and $1-\theta^\circ_{ij}$). Note that α_{ij} and β_{ij} converge to θ_{ij} and γ_{ij} and δ_{ij} converge to $1-\theta_{ij}$ with probability one under H_0. This fact implies that, for sufficiently large n,

$$\frac{\hat{p}_{ij}}{\beta^2_{ij}} + \frac{\sum_{\ell=i+1}^r \hat{p}_{ij}}{\delta^2_{ij}} \approx \frac{\hat{p}_{ij}}{\alpha^2_{ij}} + \frac{\sum_{\ell=i+1}^r \hat{p}_{ij}}{\gamma^2_{ij}} \approx \frac{(\sum_{\ell=i}^r \hat{p}_{\ell j})^3}{\hat{p}_{ij}\sum_{\ell=i+1}^r \hat{p}_{\ell j}}.$$

We note that if H_0 is true then

$$\frac{(\sum_{\ell=i}^r p_{\ell j})^2}{p_{ij}\sum_{\ell=i+1}^r p_{\ell j}} = \frac{(\sum_{\ell=i}^r p_{\ell\cdot})^2}{p_{i\cdot}\sum_{\ell=i+1}^r p_{\ell\cdot}} \text{ for } j=1,2,\ldots,c,$$

where $p_{i\cdot} = \sum_{j=1}^c p_{ij}$. For sufficiently large n,

$$\frac{(\sum_{\ell=i}^r \hat{p}_{\ell j})^3}{\hat{p}_{ij}\sum_{\ell=i+1}^r \hat{p}_{\ell j}} \approx \frac{(\sum_{\ell=i}^r p_{\ell\cdot})^2}{p_{i\cdot}\sum_{\ell=i+1}^r p_{\ell\cdot}} \sum_{\ell=i}^r \hat{p}_{\ell j} \text{ for } j=1,2,\ldots,c.$$

Now we have

$$T_{01} \approx n\sum_{i=1}^{r-1}\sum_{j=1}^{c}\left[(\theta^\circ_{ij}-\hat{\theta}_{ij})^2 - (\theta^*_{ij}-\hat{\theta}_{ij})^2\right] \frac{(\sum_{\ell=i}^r p_{\ell\cdot})^2}{p_{i\cdot}\sum_{\ell=i+1}^r p_{\ell\cdot}} \sum_{\ell=i}^r \hat{p}_{\ell j}. \tag{3.2}$$

It follows from Theorem 1.3.2 of Robertson *et al.* (1988) that

$$\sum_{j=1}^c \hat{\theta}_{ij}\theta^\circ_{ij}\sum_{\ell=i}^r \hat{p}_{\ell j} = \sum_{j=1}^c \theta^*_{ij}\theta^\circ_{ij}\sum_{\ell=i}^r \hat{p}_{\ell j} \text{ and } \sum_{j=1}^c \theta^*_{ij}\theta^\circ_{ij}\sum_{\ell=i}^r \hat{p}_{\ell j} = \sum_{j=1}^c {\theta^*_{ij}}^2\sum_{\ell=i}^r \hat{p}_{\ell j}.$$

Hence (3.2) is equal to

$$n\sum_{i=1}^{r-1}\frac{(\sum_{\ell=i}^r p_{\ell\cdot})^2}{p_{i\cdot}\sum_{\ell=i+1}^r p_{\ell\cdot}}\sum_{j=1}^c(\theta^\circ_{ij}-\theta^*_{ij})^2\sum_{\ell=i}^r\hat{p}_{\ell j} =$$

$$\sum_{i=1}^{r-1}\frac{(\sum_{\ell=i}^r p_{\ell\cdot})^2}{p_{i\cdot}\sum_{\ell=i+1}^r p_{\ell\cdot}}\sum_{j=1}^c\left\{E_{\hat{\mathbf{p}}_i}(\sqrt{n}(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i)|\mathcal{I})_j - E_{\hat{\mathbf{p}}_i}(\sqrt{n}(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i)|\mathcal{C})_j\right\}^2(\sum_{\ell=i}^r\hat{p}_{\ell j})$$

where $\mathcal{C} = \{\mathbf{x}\in\mathbf{R}^c : x_1 = x_2 = \cdots = x_c\}$.

By the straightforward (but tedious) use of the delta method and using Theorem 2.2 we can show that $\sqrt{n}\left((\hat{\boldsymbol{\theta}}_1,\hat{\boldsymbol{\theta}}_2,\ldots\hat{\boldsymbol{\theta}}_{r-1}) - (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_{r-1})\right)$ converges in distribution to a normal random vector with mean $\mathbf{0}$ and variance $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is a block diagonal matrix with block elements $\boldsymbol{\Sigma}_i, i=1,\ldots,r-1$ defined as

$$\boldsymbol{\Sigma}_i = diag\left\{\frac{p_{ij}(\sum_{\ell=i+1}^r p_{\ell j})}{(\sum_{\ell=i}^r p_{\ell j})^3}, j=1,2,\ldots,c\right\}.$$

Now T_{01} converges in distribution to

$$\sum_{i=1}^{r-1}\sum_{j=1}^{c}\{E_{\mathbf{p}_i}(\mathbf{X}_i|\mathcal{I})_j - E_{\mathbf{p}_i}(\mathbf{X}_i|\mathcal{C})_j\}^2 \frac{(\sum_{\ell=i}^{r} p_{\ell\cdot})^2(\sum_{\ell=i}^{r} p_{\ell j})}{p_{i\cdot}(\sum_{\ell=i+1}^{r} p_{\ell\cdot})},$$

where $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ic})$ and X_{ij} is an independent normal variable with the common mean 0 and the variance $((\sum_{\ell=i}^{r} p_{\ell\cdot})^2 \sum_{\ell=i}^{r} p_{\ell j})/p_{i\cdot}(\sum_{\ell=i+1}^{r} p_{\ell\cdot})$. For each i let

$$U_i = \sum_{j=1}^{c}\{E_{\mathbf{p}_i}(\mathbf{X}_i|\mathcal{I})_j - E_{\mathbf{p}_i}(\mathbf{X}_i|\mathcal{C})_j\}^2 \frac{(\sum_{\ell=i}^{r} p_{\ell\cdot})^2(\sum_{\ell=i}^{r} p_{\ell j})}{p_{i\cdot}(\sum_{\ell=i+1}^{r} p_{\ell\cdot})}.$$

Then by Theorem 2.3.1 of Robertson *et al.* (1988), U_i has a chi-bar-square distribution, i.e., for all t,

$$Pr[U_i \geq t] = \sum_{\ell=1}^{c} P(\ell, c; \mathbf{p}_i) Pr[\chi^2_{\ell-1} \geq t], \tag{3.3}$$

where χ^2_ν denotes a chi-square random variable with ν degrees of freedom and $P(\ell, c; \mathbf{w})$ is the level probability that $E_{\mathbf{w}}(\mathbf{Y}|\mathcal{I})$ takes on ℓ distinct values, where $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_c)$ consists of independent random variables and Y_i is $N(0, 1/w_i)$ We note that U_i's are independent. We also note that if H_0 is true then $\mathbf{p}_i = \sum_{\ell=i}^{r} p_{\ell\cdot}(p_{\cdot 1}, p_{\cdot 2}, \ldots, p_{\cdot c})$. Since level probabilities do not change their values if weights are multiplied by a positive constant, the weights in $P(\ell, c; \mathbf{p}_i)$ can be replaced by $(p_{\cdot 1}, p_{\cdot 2}, \ldots, p_{\cdot c})$. Then the distributions of U_i's are identical. Hence the aymptotic null distribution of T_{01} is a convolution of $r-1$ independent, identical chi-bar-square distributions. The next theorem summarizes the above argument.

Theorem 3.1. *Under H_0, for all t,*

$$\lim_{n\to\infty} Pr[T_{01} \geq t] = \sum_{\ell=r-1}^{(r-1)c} C_\ell Pr[\chi^2_{\ell-r+1} \geq t],$$

where C_ℓ is called mixing coefficient defined as

$$C_\ell = \sum_{\ell_1+\ell_2+\ldots+\ell_{r-1}=\ell}\ \prod_{k=1}^{r-1} P(\ell_k, c; (p_{\cdot 1}, p_{\cdot 2}, \ldots, p_{\cdot c})).$$

In the problem of testing the equality of several distribution functions against uniform stochastic ordering, Dykstra, Kochar and Robertson (1991) show that the asymptotic null distribution of the test statistic does not depend upon the common distribution function. It is, however, of interest to observe that the asymptotic null distribution of T_{01} depends upon the unknown parameter $\mathbf{p}$.

Since the distribution of T_{01} depends upon unknown quantities we need to figure out how to find critical value. One method is to find the least favorable distribution, which is stochastically largest within the class of asymptotic null distribution. The test based on the least favorable distribution is, however, very conservative.

We now find the least favorable distribution. Note that U_i's are independent, identically distributed. Moreover they are all nonnegative. Hence the least favorable distribution can be obtained by finding each of least favorable configurations which maximize $Pr[U_i \geq t]$'s, respectively. It follows from Theorem 1 of Robertson and Wright (1982) that, for all t,

$$Pr[U_i \geq t] \leq \sum_{\ell=1}^{c} \binom{c-1}{\ell-1} 2^{-c+1} Pr[\chi^2_{\ell-1} \geq t].$$

Now we have the following theorem.

Theorem 3.2. *Under H_0, for all t,*

$$\lim_{n\to\infty} Pr[T_{01} \geq t] \leq \sum_{\ell=r-1}^{(r-1)c} C_\ell Pr[\chi^2_{\ell-r+1} \geq t],$$

where

$$C_\ell = \sum_{\ell_1+\ell_2+\cdots+\ell_{r-1}=\ell} \binom{c-1}{\ell_1-1}\binom{c-1}{\ell_2-1}\cdots\binom{c-1}{\ell_{r-1}-1} 2^{-(r-1)(c-1)}.$$

Oh (1994) studied methods for approximating the null hypothesis distributions of several test statistics by using estimate of the unknown quantity on which the null distribution depends. We recommend that the quantity in (3.3) be approximated by

$$\sum_{\ell=1}^{c} P(\ell, c; (\hat{p}_{\cdot 1}, \hat{p}_{\cdot 2}, \ldots, \hat{p}_{\cdot c})) Pr[\chi^2_{\ell-1} \geq t]. \tag{3.4}$$

Using these methods we may show that (3.4) generally provides a very good approximation to the asymptotic null distribution of T_{01}.

3.2. Test of Dependence Against All Alternatives

Next we consider the problem of testing the null hypothesis H_1 against all alternatives, i.e., H_2, which imposes no restrictions other than the basic restrictions. The likelihood ratio statistic is

$$\Lambda_{12} = \frac{\prod_{i=1}^{r-1}\left[\prod_{j=1}^{c}(\theta_{ij}^*)^{n\hat{p}_{ij}}(1-\theta_{ij}^*)^{n\sum_{\ell=i+1}^{r}\hat{p}_{ij}}\right]}{\prod_{i=1}^{r-1}\left[\prod_{j=1}^{c}(\hat{\theta}_{ij})^{n\hat{p}_{ij}}(1-\hat{\theta}_{ij})^{n\sum_{\ell=i+1}^{r}\hat{p}_{ij}}\right]}.$$

The test rejects H_1 for large value of $T_{12} = -2\ln\Lambda_{12}$, i.e.,

$$T_{12} = 2n\sum_{i=1}^{r-1}\left[\sum_{j=1}^{c}\hat{p}_{ij}(\ln\hat{\theta}_{ij} - \ln\theta_{ij}^*) + \sum_{j=1}^{c}(\sum_{\ell=i+1}^{r}\hat{p}_{\ell j})(\ln(1-\hat{\theta}_{ij}) - \ln(1-\theta_{ij}^*))\right].$$

Expanding $\ln\theta_{ij}^*$ and $\ln(1-\theta_{ij}^*)$ about $\hat{\theta}_{ij}$ and $1-\hat{\theta}_{ij}$, respectively, and writing in terms of projection operators we have

$$T_{12} = \sum_{i=1}^{r-1}\sum_{j=1}^{c}\left[\sqrt{n}(\hat{\boldsymbol{\theta}}_i)_j - \sqrt{n}E_{\hat{\mathbf{p}}_i}(\hat{\boldsymbol{\theta}}_i|\mathcal{I})_j\right]^2\left\{\frac{\hat{p}_{ij}}{\sigma_{ij}^2} + \frac{\sum_{\ell=i+1}^{r}\hat{p}_{\ell j}}{\rho_{ij}^2}\right\},$$

where σ_{ij} (ρ_{ij}) is between $\hat{\theta}_{ij}$ and θ_{ij}^* ($1-\hat{\theta}_{ij}$ and $1-\theta_{ij}^*$).

Consider the computation of $E_{\hat{\mathbf{p}}_i}(\hat{\boldsymbol{\theta}}_i|\mathcal{I})$. Suppose $\theta_{ij} > \theta_{i,j+1}$. For sufficiently large n we have $\hat{\theta}_{ij} > \hat{\theta}_{i,j+1}$ with probability one. This implies that no amalgamation between $\hat{\theta}_{ij}$ and $\hat{\theta}_{i,j+1}$ occurs when computing $E_{\hat{\mathbf{p}}_i}(\hat{\boldsymbol{\theta}}_i|\mathcal{I})$. Define a binary relation $\succeq_{\boldsymbol{\theta}_i}$ on $\{1,2,\dots,c\}$ by

$$a \succeq_{\boldsymbol{\theta}_i} b \text{ only when } \theta_{ia} = \theta_{ib} \text{ and } a \succeq b.$$

It is not difficult to show that $\succeq_{\boldsymbol{\theta}_i}$ is a partial order on $\{1,2,\dots,c\}$. Define $\mathcal{I}_{\boldsymbol{\theta}_i}$ by $\{\mathbf{x}\in\mathbf{R}^c : x_a \geq x_b \text{ if } a \succeq_{\boldsymbol{\theta}_i} b\}$. Using the strong law of large numbers and following the lines of Lemmas A and B of Section 5.2 in Robertson *et al.* (1988) we have

$$E_{\hat{\mathbf{p}}_i}(\hat{\boldsymbol{\theta}}_i|\mathcal{I}) = E_{\hat{\mathbf{p}}_i}(\hat{\boldsymbol{\theta}}_i|\mathcal{I}_{\boldsymbol{\theta}_i})$$

provided n is sufficiently large. Then

$$T_{12} \approx \sum_{i=1}^{r-1}\sum_{j=1}^{c}\left[E_{\hat{\mathbf{p}}_i}(\sqrt{n}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)|\mathcal{I}_{\boldsymbol{\theta}_i})_j - \sqrt{n}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)_j\right]^2\left\{\frac{\hat{p}_{ij}}{\sigma_{ij}^2} + \frac{\sum_{\ell=i+1}^{r}\hat{p}_{\ell j}}{\rho_{ij}^2}\right\}$$

and hence T_{12} converges in distribution to

$$\sum_{i=1}^{r-1}\sum_{j=1}^{c}\left[E_{\mathbf{p}_i}(\mathbf{X}_i|\mathcal{I}_{\boldsymbol{\theta}_i})_j - X_{ij}\right]^2\left\{\frac{(\sum_{\ell=i}^{r} p_{\ell j})^3}{p_{ij}\sum_{\ell=i+1}^{r} p_{\ell j}}\right\}.$$

Let $V_i = \sum_{j=1}^{c}\left[E_{\mathbf{p}_i}(\mathbf{X}_i|\mathcal{I}_{\boldsymbol{\theta}_i})_j - X_{ij}\right]^2\left\{(\sum_{\ell=i}^{r} p_{\ell j})^3/(p_{ij}\sum_{\ell=i+1}^{r} p_{\ell j})\right\}$. Then by Theorem 2.3.1 of Robertson *et al.* (1988), V_i has a chi-bar-square distribution. Specifically, under H_1, for all t,

$$Pr[V_i \geq t] = \sum_{\ell=1}^{c} P(\ell, c; \mathbf{p}_i, \succeq_{\boldsymbol{\theta}_i})Pr[\chi^2_{c-\ell} \geq t], \tag{3.5}$$

where $P(\ell, c; \mathbf{p}_i, \succeq_{\boldsymbol{\theta}_i})$ is level probability defined as earlier except different partial order. Since V_i's are independent, the asymptotic null distribution of T_{12} is a convolution of $(r-1)$ independent chi-bar-square distributions. The distribution obviously depends upon unknown parameter $\mathbf{p}_i$'s. This leads to the following theorem.

Theorem 3.3. *Under H_1, for all t,*

$$\lim_{n\to\infty} Pr[T_{12} \geq t] = \sum_{\ell=r-1}^{(r-1)c} C_\ell Pr[\chi^2_{\ell-r+1} \geq t],$$

where

$$C_\ell = \sum_{\ell_1+\ell_2+\ldots+\ell_{r-1}=\ell}\ \prod_{k=1}^{r-1} P(\ell_k, c; \mathbf{p}_k, \succeq_{\boldsymbol{\theta}_k}).$$

Next we find the least favorable distribution of T_{12}. Note that $\mathcal{I} \subset \mathcal{I}_{\boldsymbol{\theta}_i}$. The (3.5) is maximized when $\mathcal{I} = \mathcal{I}_{\boldsymbol{\theta}_i}$, which requires that

$$\theta_{i1} = \theta_{i2} = \cdots = \theta_{ic}. \tag{3.6}$$

Note that $\mathbf{p}_i = (\phi_1\prod_{\ell=1}^{i-1}(1-\theta_{\ell 1}), \phi_2\prod_{\ell=1}^{i-1}(1-\theta_{\ell 2}), \ldots, \phi_c\prod_{\ell=1}^{i-1}(1-\theta_{\ell c}))$. If (3.6) holds, then $\mathbf{p}_i = K\cdot(\phi_1, \phi_2, \ldots, \phi_c)$, where K is the common value of $\prod_{\ell=1}^{i-1}(1-\theta_{\ell j})$. Since level probabilities do not change their values if weights are multiplied by a positive constant, if (3.6) holds then (3.5) is equal to

$$\sum_{\ell=1}^{c} P(\ell, c; (\phi_1, \phi_2, \ldots, \phi_c), \succeq)Pr[\chi^2_{c-\ell} \geq t]$$

and hence the distributions of V_i's become identical. It follows from Theorem 1 of Robertson and Wright (1982) that, under the condition (3.6),

$$Pr[V_i \geq t] \leq \frac{1}{2}\left\{Pr[\chi^2_{c-2} \geq t] + Pr[\chi^2_{c-1} \geq t]\right\}.$$

Then the least favorable distribution of T_{12} becomes a convolution of $(r-1)$ independent, identical chi-bar-square distributions. We summarize the above argument in the following theorem.

Theorem 3.4. *Under H_1, for all t,*

$$\lim_{n\to\infty} Pr[T_{12} \geq t] \leq \sum_{\ell=0}^{r-1} \binom{r-1}{\ell} 2^{-r+1} Pr[\chi^2_{(c-1)(r-1)-\ell} \geq t]. \tag{3.7}$$

Critical value can be obtained from (3.7) for a very conservative test. Rather than using this conservative test one can use an approximate test which is implemented by approximating the null distribution using the similar method studied by Oh(1994). Such test procedure may be found in Oh(1998).

4. EXAMPLE

To illustrate the estimation and testing procedures discussed in earlier sections, we examine a data set, taken from Srole *et al.* (1978 p 289), which describes the relationship between mental impairment and parents socioeconomic status for a sample of residents of Manhatten. We would like to test whether there is a tendency for mental health to be better at higher level of parents' socioeconomic status.

Table 4.1 : Mental Health Status and Parents' Socioeconomic Status

	Parents' Socioeconomic Status					
Mental Health	A(high)	B	C	D	E	F(low)
well	64	57	57	72	36	21
mild	94	94	105	141	97	71
moderate	58	54	65	77	54	54
impaired	46	40	60	94	78	71

Table 4.2 shows the estimated continuation odds ratios. For comparisons sake, we list the local odds ratios. The odds ratios which are less than 1 are

Table 4.2 : Estimated Continuation Odds Ratios and Local Odds Ratios

i	Odds ratio	j: 1	2	3	4	5
1	Continuation	1.0661	1.2234	1.0739	1.4679	1.4672
	Local	1.1228	1.1170	1.0631	1.3759	1.2548
2	Continuation	0.9038+	1.1905	1.0187	1.1221	1.2937
	Local	0.9310+	1.0776	0.8822+	1.0194	1.3662
3	Continuation	0.9340+	1.2462	1.3225	1.1832	0.9103+
	Local	0.9340+	1.2462	1.3225	1.1832	0.9103+

Table 4.3 : Computational Details for $\hat{\theta}_{ij}$, θ^*_{ij}, θ°_{ij}, and ϕ_i

i		j: 1	2	3	4	5	6	θ°_{ij}
1	$\hat{\theta}_{ij}$	0.2443	0.2327	0.1986	0.1875	0.1358	0.0968	0.1849
	θ^*_{ij}	0.2443	0.2327	0.1986	0.1875	0.1358	0.0968	
2	$\hat{\theta}_{ij}$	0.4747†	0.5000†	0.4565	0.4519	0.4236	0.3622	0.4449
	θ^*_{ij}	0.4870‡	0.4870‡	0.4565	0.4519	0.4236	0.3622	
3	$\hat{\theta}_{ij}$	0.5577†	0.5745†	0.5200	0.4503	0.4091†	0.4320†	0.4820
	θ^*_{ij}	0.5657‡	0.5657‡	0.5200	0.4503	0.4202‡	0.4202‡	
ϕ_j		0.1578	0.1476	0.1729	0.2313	0.1596	0.1307	

marked by +. The fact that 12 out of 15 odds ratios are bigger than 1 indicates that there is a tendency being positively dependent in the sense of dependence concept discussed earlier.

The computational details are found in Table 4.3. The order-violated parameters are marked by † and the corresponding maximum likelihood estimates are marked by ‡. The restricted maximum likelihood estimate of θ_{21} and θ_{22}, for instance, are given by

$$\theta^*_{21} = \theta^*_{22} = \frac{p_{21} + p_{22}}{\sum_{\ell=2}^{4}(p_{\ell 1} + p_{\ell 2})}.$$

The computed value of likelihood ratio statistic $T_{01} = -2\ln\Lambda_{01}$ is 46.97682.

To find p-value of the test, we first need to compute the level probabilities, $P(\ell, 6; \mathbf{p}_i), \ell = 1, 2, \ldots, 6$. We note that $P(\ell, 6; \mathbf{p}_i)$ is unknown. As mentioned

Table 4.4 : Mixing Coefficients Used in a Convolution of Chi-bar-square Distributions

ℓ	3	4	5	6	7	8	9	10
approx.	.00390	.02738	.08771	.17007	.22364	.21167	.14926	.08000
equal	.00463	.03171	.09855	.18388	.23126	.20772	.13781	.06886
least	.00003	.00046	.00320	.01389	.04166	.09164	.15274	.19638
ℓ	11	12	13	14	15	16	17	18
approx.	.03292	.01042	.00252	.00046	.00006	.00001	.00000	.00000
equal	.02616	.00756	.00017	.00027	.00003	.00000	.00000	.00000
least	.19638	.15274	.09164	.04166	.01389	.00320	.00046	.00003

earlier we can approximate it by plugging in $\hat{\mathbf{p}}_i$ and it provides generally a good approximation. It is, however, still unable to compute the level probabilities since no explicit formula for computing level probability, $P(\ell, k; \mathbf{w})$ with $k \geq 6$ and arbitrary weights $\mathbf{w}$, is available. Robertson and Wright (1983) studied approximation of level probabilities for the simple order. They suggest that the equal weights approximation can be used if the ratio of the maximum weight to the minimum weight is less than 3.4. We may also use the equal weights level probabilities. For the equal weights case see Robertson *et al.* (1988). Based on these level probabilities we compute mixing coefficients, C_ℓ's, defined in Theorems 3.1 and 3.2. Table 4.4 lists the mixing coefficients.

The p-values of this test, which are 2.1×10^{-8}, 1.8×10^{-8}, and 3.9×10^{-7} for each method of computing level probabilities, are close to zero. Clearly there is a strong evidence supporting the tendency for mental health to be better at higher levels of parents' socioeconomic status.

5. CONCLUDING REMARKS

In this paper we have discussed statistical inference concerning a new type of dependence concept which is based on continuation odds ratios. There are, however, three other types of continuation odds ratios namely,

$$\frac{p_{ij} \cdot \sum_{\ell=j+1}^{c} p_{i+1,\ell}}{p_{i+1,j} \cdot \sum_{\ell=j+1}^{c} p_{i\ell}}, \quad \frac{\sum_{\ell=1}^{i} p_{\ell,j} \cdot p_{i+1,j+1}}{\sum_{\ell=1}^{i} p_{\ell,j+1} \cdot p_{i+1,j}}, \text{ and } \frac{\sum_{\ell=1}^{j} p_{i\ell} \cdot p_{i+1,j+1}}{\sum_{\ell=1}^{j} p_{i+1,\ell} \cdot p_{i,j+1}}.$$

Requiring each type of odds ratios to be greater than or equal to 1 corresponds to each specific type of dependence concepts, respectively. These four types of dependence concepts have virtually the same structure and properties. But there is no heirarchical relationship among these four dependence concepts. By symmetry argument we can easily implement the estimation and test procedures for each type of dependence concept.

In the analysis of categorical data, multinomial or Poisson models are generally assumed. In this paper, we assume a single multinomial model, i.e., $p_{ij} > 0$ and $\sum_{ij} p_{ij} = 1$. One might, however, be interested in other sampling models. Assume that $p_{ij} > 0$ and $\sum_i p_{ij} = 1$ for $j = 1, 2, \ldots, c$. This is the so-called column product multinomial model. Inference under this model is the problem of uniform stochastic ordering among several multinomial populations. Statistical inference for this problem may be found in Dykstra *et al.* (1991) and Park, Lee and Robertson (1998). Next assume that $p_{ij} > 0$ and $\sum_j p_{ij} = 1$ for $i = 1, 2, \ldots, r$. This is row product multinomial model. Unfortunately, we have been unable to find an explicit form of maximum likelihood estimate under restriction (1.1). The Poisson model assumes that each observation in a cell is distributed as Poisson distribution with mean $\lambda_{ij} > 0$. If we assume the number of observations - sample sizes are the same for all cells, we can use the estimation procedure for a single multinomial model with minor modification. But we were unable to find restricted maximum likelihood estimate if the assumption of equal sample sizes is relaxed.

References

Cohen, A., and H. B. Sachrowitz (1991). "Tests for independence in contingency tables with ordered categories," *Journal of Multivariate Analysis*, **36**, 56-67.

Douglas, R., S. E. Fienberg, M. T. Lee, A. R. Sampson, and L. R. Whitaker (1990). Positive dependence concepts for ordinal contingency tables, In *Topics in Statistical Dependence* (H. W. Block, A. R. Sampson, T. H. Savits eds.). IMS Lecture/Monograph series, Hayward, CA, 189-202.

Dykstra, R. L., and J. H. Lemke (1988). "Duality of *I* projections and Maximum likelihood estimation for log-linear models under cone constraints," *Journal of the American Statistical Association*, **83**, 546-554.

Dykstra, R. L., S. C. Kochar, and T. Robertson (1991). "Statistical inference for uniform stochastic ordering in several populations," *The Annals of Statistics*, **19**, 870-888.

Feltz, P. D., and R. L. Dykstra (1985). "Maximum likelihood estimation of the survial functions of N stochastically ordered random variables," *Journal of the American Statistical Association*, **80**, 1012-1019.

Hirotsu, C. (1982). "Use of cumulative efficient scores for testing ordered alternatives in discrete models," *Biometrika*, **69**, 567-577.

Lee, M. -L. T. (1990). Tests of independence against likelihood ratio dependence in ordered contingency tables. In *Topics in Statistical Dependence* (H. W. Block, A. R. Sampson, T. H. Savits. eds.), IMS Lecture Notes/Monograph series, Hayward, CA, 351-357.

Lehmann. E. L. (1966). "Some concepts of dependence," *Annals of Mathematical Statistics*, **37**, 1137-1153.

Nguyen, T., and A. R. Sampson (1987). "Testing for positive quadrant dependence in ordinal contingency tables," *Naval Research Logistics Quarterly* **34**, 859-877.

Oh, M. (1994). Statistical tests concerning a set of multinomial parameters under order restrictions: Approximations to null hypotheses distributions, Ph. D. Thesis, The University of Iowa, Iowa City, Iowa.

Oh, M. (1995). "On maximum likelihood estimation of cell probabilities in $2 \times k$ contingency tables under negative dependence restrictions with various sampling scheme," *Communications in Statistics-Theory and Methods*, **24**(8), 2127-2143.

Oh, M. (1998). Approximate tests for uniform stochastic ordering in several multinomial populations, unpublished manuscript.

Park, C. G., C. I. C. Lee, and T. Robertson (1998). "Goodness-of-fit test for uniform stochastic ordering among several distributions," *The Canadian Journal of Statistics*, **26**, 69-81.

Patefield, W. M. (1982). "Exact tests for trends in ordered contingency tables," *Applied Statistics*, **31**, 32-43.

Robertson, T., and F. T. Wright (1982). "Bounds on mixtures of distribution arising in order restricted inference," *The Annals of Statistics*, **10**, 302-306.

Robertson, T., and F. T. Wright (1983). "On approximation of the level probabilities and associated distributions in order restricted inference," *Biometrika*, **70**, 597-606.

Robertson, T., F. T. Wright, and R. L. Dykstra (1988). *Order Restricted Statistical Inference*, Wiley, Chichester.

Shaked, M. (1977). "A family of concepts of dependence for bivariate distributions," *Journal of the American Statistical Association*, **72**, 642-650.

Srole, L., T. S. Langner, S. T. Michael, P. Kirkpatrick, M. K. Opler, and T. A. C. Rennie (1978). *Mental Health in the Metropolis: The Midtown Manhattan Study* rev. ed., New York University Press, New York.