Statistica Sinica **18**(2008), 1165-1183

SOME ASSOCIATION MEASURES AND THEIR COLLAPSIBILITY

Xianchao Xie[1], Zongming Ma[2] and Zhi Geng[1]

[1]*Peking University and* [2]*Stanford University*

Abstract: For two variables X and Y with arbitrary distributions, we consider three general association measures, the mixed derivative of interaction, the partial derivative of the conditional distribution function and the partial derivative of the conditional expectation. The sign of an association measure between X and Y may sometimes be reversed after marginalization over a third variable W. In this paper, we first compare the stringency of these measures for evaluating a positive association. Then we present the condition for avoiding the effect reversal after marginalization over W. Further we show that a modification of the condition can be used for collapsibility of the association measures over W.

Key words and phrases: Association measure, collapsibility, Yule-Simpson paradox.

1. Introduction

For arbitrary distributions of continuous, discrete, or even mixed type variables, three general measures of association between two variables X and Y are discussed in this paper. The first measure is the mixed derivative of interaction, $\partial^2 \log f(x, y)/\partial x \partial y$ where f denotes a probability density function (Holland and Wang (1987) and Whittaker (1990)). The second one is the partial derivative of the conditional distribution function, $\partial F(y|x)/\partial x$ where F denotes a cumulative distribution function, proposed by Cox and Wermuth (2003) and called the distribution dependence. The third one is defined as the partial derivative of the conditional expectation, $\partial E(Y|x)/\partial x$, called expectation dependence below, when the expectation exists. When variables X and Y have a joint normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we have

$$\frac{\partial E(Y|x)}{\partial x} = \rho_{XY}\frac{\sigma_Y}{\sigma_X}, \qquad \frac{\partial^2 \log f(x,y)}{\partial x \partial y} = \frac{\rho_{XY}}{(1-\rho_{XY}^2)\sigma_X\sigma_Y},$$

$$\frac{\partial F(y|x)}{\partial x} = -\frac{\rho_{XY}}{\sqrt{2\pi}(1-\rho_{XY}^2)\sigma_X\sigma_Y}\exp\left\{-\frac{[y-E(Y|x)]^2}{2\sigma_{Y|X}^2}\right\},$$

where μ_X denotes the mean of X, σ_X^2 the variance of X, ρ_{XY} the correlation coefficient, and $E(Y|x) = \mu_Y + (\rho_{XY}\sigma_Y/\sigma_X)(x - \mu_X)$.

If X and/or Y is discrete, then the partial differentiation for these measures is replaced by differencing between adjacent levels. For example, if X is binary and Y has three values 0, 1 and 2, then the expectation dependence becomes $\sum_{j=1}^{2} j[P(Y = j|X = 1) - P(Y = j|X = 0)]$, the distribution dependence becomes $P(Y \leq j|X = 1) - P(Y \leq j|X = 0)$ for $j = 0$ and 1, and the mixed derivative of interaction becomes the log odds ratios $\log P(X = 1, Y = j + 1)P(X = 0, Y = j)/[P(X = 0, Y = j + 1)P(X = 1, Y = j)]$ for $j = 0$ and 1.

It can be shown that the following statements are equivalent:

1. there is no mixed derivative of interaction, that is, $\partial^2 \log f(x, y)/\partial x \partial y = 0$ for all x and y;
2. there is no distribution dependence, that is, $\partial F(y|x)/\partial x = 0$ for all x and y; and
3. X and Y are independent, denoted as $X \perp\!\!\!\perp Y$.

See Whittaker (1990, Proposition 2.3.1) for the equivalence of the first and third statements. The equivalence of the second and the third is immediate since $\partial F(y|x)/\partial x = 0$ for all x and y if and only if $F(y|x) = F(y|x')$ for all x and x'. The above equivalences also hold conditionally on a third variable W. However, for expectation dependence, $\partial E(Y|x)/\partial x = 0$ for all x is only necessary but not sufficient for the independence $X \perp\!\!\!\perp Y$.

An association between X and Y may be a consequence of the fact that X associates with W which in turn associates with Y. On the other hand, omitting W may sometimes reverse the sign of an association measure between X and Y. This effect reversal of an association measure is also called the Yule-Simpson Paradox (Yule (1903) and Simpson (1951)). Cox and Wermuth (2003) presented a general condition (Y and W are conditionally independent given X, denoted as $Y \perp\!\!\!\perp W|X$; or $X \perp\!\!\!\perp W$) for avoiding the effect reversal of distribution dependence after marginalization over W.

In this paper, we first compare the stringency of these measures to evaluate a positive association between X and Y. Note that the mixed derivative of interaction can be seen as a local-local measure, the distribution dependence as a local-global measure, and the expectation dependence as a local-expected measure. Here 'local' and 'global' are in the sense similar to those of odds ratios for ordinal variables in Agresti (1984). We show for these three association measures that a positive association is the most stringent when the mixed derivative of interaction is positive for all x and y, and that a positive association is the weakest when the expectation dependence is positive for all x. Then, we propose that the two independencies given by Cox and Wermuth (2003) are also applicable separately to other association measures for avoiding the effect reversal after marginalization over W. In many studies, we may wish to study whether or not an observed or unobserved W influences the association between X and Y, and we may also wish to discretize a continuous W without changing the original association measure. We say that an association measure between X and Y is simply collapsible over W if the measure conditional on W remains unchanged after marginalization over W. Further we say that the measure is uniformly collapsible over W if it remains unchanged conditional on any interval or subset of W. Uniform collapsibility of distribution dependence was discussed in Ma, Xie and Geng (2006). This paper discusses uniform collapsibility of the mixed derivative of interaction and the expectation dependence, and it shows that modifications of Cox and Wermuth's general condition is a necessary and sufficient condition for uniform collapsibility of these two measures. Throughout our discussion, we assume that the joint distribution of Y, X and W is such that differentiation and integration are interchangeable.

Section 2 compares the stringency of these association measures for evaluating a positive association, and shows the conditions for avoiding the effect reversal after marginalization over W. Section 3 defines collapsibility of association measures, and presents the necessary and sufficient conditions for uniform collapsibility of association measures over a discrete or continuous W. In Section 4, we apply the collapsibility of association measures to linear models to illustrate collapsibility of parameters in the models. We discuss multivariate cases in Section 5. Finally a discussion is given in Section 6. All proofs of theorems are given in the Appendix.

2. Stringency of Association Measures and Conditions for Avoiding Effect Reversal

If $\partial F(y|x)/\partial x \leq 0$ for all y and x, with strict inequality in a region of positive probability, then we have $P(Y > y|X = x) \geq P(Y > y|X = x')$ for $x > x'$ and all y, and we say that the distribution dependence of Y on X is stochastically increasing with X. If $\partial E(Y|x)/\partial x \geq 0$ for all y and x, with strict inequality in a region of positive probability, then we have $E(Y|x) \geq E(Y|x')$ for $x > x'$ and all y, and we say that the expectation dependence of Y on X is average increasing with X. It can be seen that the sign of these association measures indicates the direction of the association between X and Y. The following theorem compares a positive mixed derivative of interaction, a stochastic increasing dependence, an average increasing dependence, and a positive correlation for their stringency.

Theorem 1. *The association measures have the following implications*

$$\frac{\partial^2 \log f(x,y)}{\partial x \partial y} \geq 0,\ \forall x, y \ \Rightarrow\ \frac{\partial F(y|x)}{\partial x} \leq 0,\ \forall x, y$$
$$\Rightarrow\ \frac{\partial E(Y|x)}{\partial x} \geq 0,\ \forall x \ \Rightarrow\ \rho_{XY} \geq 0.$$

Further, for any two of the four inequalities, strict inequality holds with positive probability for the right inequality if strict inequality holds with positive probability for the left one.

The first implication of a positive mixed derivative of interaction to a stochastic increasing dependence has been shown by Mari and Kotz (2001). If both X and Y are binary or both are normal variables, all converses of the implications in Theorem 1 are also true; in general, any converse of the implications is not true. From Theorem 1, it can be seen that a positive mixed derivative of interaction is the most stringent, a stochastic increasing dependence the second, an average increasing dependence the third, and a positive correlation is the weakest positive association measure.

The conditional measures are defined by the conditional distribution of X and Y given a third variable W. For example, the mixed derivative of interaction conditional on W is defined as $\partial^2 \log f(x,y|w)/\partial x \partial y$, where $f(x,y|w)$ is the conditional density of Y and X given $W = w$. Even if an association measure has the same sign conditionally on any value of W, the sign may be reversed by marginalizing over W, called the effect reversal of this association measure. For example, one can have $\partial F(y|x,w)/\partial x \leq 0$ for all x, y and w, but $\partial F(y|x)/\partial x > 0$ for some x or y. Cox and Wermuth (2003) proposed the general condition ($X \perp\!\!\!\perp W$ or $Y \perp\!\!\!\perp W|X$) for avoiding the effect reversal of distribution dependence. We show below that the independencies ($X \perp\!\!\!\perp W$ and $Y \perp\!\!\!\perp W|X$) are also applicable separately to some other association measures.

Theorem 2. *If $X \perp\!\!\!\perp W$, then the effect reversals of the distribution dependence and the expectation dependence is avoided. If $Y \perp\!\!\!\perp W|X$, then the effect reversals of the mixed derivative of interaction, the distribution dependence, and the expectation dependence are avoided.*

In balanced data where W has the same distribution conditional on X, effect reversal cannot arise for these measures except for the mixed derivative of interaction for any distribution. A counterexample for the mixed derivative of interaction can be shown easily by considering log odds ratios conditional on W when Y has three levels, and X and W are binary.

3. Collapsibility of Association Measures

In this section we discuss conditions for collapsibility of association measures. For a discrete or continuous W, we define homogeneity and collapsibility of an association measure as follows. We say that an association measure between X and Y is homogeneous over W if the conditional association measure between X and Y, given $W = w$, equals that given $W = w'$ for all $w \neq w'$. For example, the expectation dependence is homogeneous over W if $\partial E(Y|x, w)/\partial x = \partial E(Y|x, w')/\partial x$ for all x and $w \neq w'$. The simple collapsibility of an association measure between X and Y means that the conditional association measure between X and Y, given $W = w$, equals the marginal association measure between X and Y for all w. For example, the simple collapsibility of the expectation dependence means that $\partial E(Y|x, w)/\partial x = \partial E(Y|x)/\partial x$ for all x and w.

Definition 1. An association measure between X and Y is uniformly collapsible over W if the measure conditional on $W \in \mathcal{I}$ for any $\mathcal{I}$ equals the measure obtained after marginalization over W, where $\mathcal{I}$ is a subset of levels for a nominal background variable W, a subset of consecutive levels $(i, i+1, \ldots, i+j)$ for an ordinal discrete background variable W, or an interval for a continuous background variable W.

For example, the conditional expectation dependence is uniformly collapsible over W if $\partial E(Y|x, W \in \mathcal{I})/\partial x = \partial E(Y|x)/\partial x$ for all x and any $\mathcal{I}$. Note that uniform collapsibility is defined for the general case where W may be discrete or continuous, and it coincides with strong and consecutive collapsibility when W is nominal and ordinal respectively (Geng (1992) and Geng and Asano (1993)). When W is binary, uniform collapsibility and simple collapsibility coincide.

From the definitions, it can be seen that uniform collapsibility implies simple collapsibility, which in turn implies homogeneity. Uniform collapsibility can be used to group levels of a discrete W or to discretize a continuous W. If the domain of W can be partitioned into K regions $\mathcal{I}_1, \ldots, \mathcal{I}_K$, and the association measure is uniformly collapsible separately for each region $\mathcal{I}_k$, then W can be recategorized into a crude variable with K levels, such that the association measure in each region is the same as the original association measure.

Theorem 3. *Expectation dependence is uniformly collapsible over W if and only if*
(a) $E(Y|x, w) = E(Y|x, w')$ *for all x and $w \neq w'$, or*
(b) $X \perp\!\!\!\perp W$ *and the expectation dependence is homogeneous over W.*

If Y is a binary response, X is a binary treatment and W is discrete, expectation dependence specializes to the risk difference $P(Y = 1|X = 1, W =$

$w) - P(Y = 1|X = 0, W = w)$. The condition in Theorem 3 is similar to that for simple and strong collapsibility of relative risks presented by Wermuth (1987, Propositions 1 and 4) and Geng (1992, Thm. 2).

Theorem 4. *The mixed derivative of interaction is uniformly collapsible over W if and only if* (a) $Y \perp\!\!\!\perp W|X$ *or* (b) $X \perp\!\!\!\perp W|Y$.

When Y and X are binary and W is discrete, the mixed derivative of interaction specializes to the log odds ratio. The condition in Theorem 4 has been shown to be necessary and sufficient for simple collapsibility of equal odds-ratios in $2\times 2\times 2$ contingency tables by Whittemore (1978), and for strong collapsibility of odds ratios in $2 \times 2 \times K$ contingency tables by Ducharme and Lepage (1986, Them. 1).

4. Applications to Regression Models

In this section, we apply the conditions for collapsibility of association measures presented in the previous section to linear and logistic regression models for collapsibility of parameters.

Let Y be a continuous dependent variable in a linear regression model, or a binary response with values 0 and 1 in a logistic regression model. X may be a continuous or discrete independent variable. In the following subsections, we consider different regression models separately when W is continuous or discrete.

4.1. Linear regression models

For a discrete W with I levels, assume the linear regression model of Y on X, conditional on W,

$$E(Y|X = x, W = i) = \alpha(i) + \beta(i)x,$$

for $i = 1, \ldots, I$, and suppose $P(W = i) > 0$ for any i. Especially, when $\beta(i) = \beta(j) = \beta$ for all $i \neq j$, we call this a parallel linear regression model. When X, Y and W have a homogenous conditional Gaussian (HCG) distribution for which the continuous variables have a joint normal distribution with a common covariance matrix, conditionally on the discrete variables, the expectation of Y conditional on X and W has a parallel regression model. For a discrete W, we only consider the parallel regression model below.

For a continuous W, assume the linear regression model of Y on X, conditional on W,

$$E(Y|X = x, W = w) = \alpha + \beta x + \gamma w,$$

for all x and w, and suppose $f(w) > 0$ for any w.

We say that the regression coefficient β is uniformly collapsible over W if the partially marginal regression model

$$E(Y|X=x, W\in\mathcal{I}) = \alpha(\mathcal{I}) + \beta(\mathcal{I})x,$$

always obtains and $\beta(\mathcal{I}) = \beta$ for any $\mathcal{I}$, where $\mathcal{I}$ is a subset of W's levels for a discrete W, a set of consecutive levels $(i, i+1, \ldots, i+j)$ for an ordinal discrete W, or an interval in W's domain for a continuous W. For the linear regression model, the uniform collapsibility of β is equivalent to that the expectation dependence $\partial E(Y|x,w)/\partial x$ is uniformly collapsible over W. In particular, when $\mathcal{I}$ is the whole domain of W, we denote the marginal regression as

$$E(Y|x) = \widetilde{\alpha} + \widetilde{\beta}x.$$

We say that the regression coefficient β is simply collapsible over W if the marginal regression model holds and $\tilde{\beta} = \beta$, which is equivalent to that the expectation dependence is simply collapsible over W. From Theorem 3, we have the following corollaries.

Corollary 1. *The regression coefficient β is uniformly collapsible over W if and only if*
(a) *for a discrete W, $\alpha(i) = \alpha(j)$ for all $i \neq j$; for a continuous W, $\gamma = 0$; or*
(b) $W \perp\!\!\!\perp X$.

Corollary 2. *Suppose that X, Y and W have a joint normal distribution for a continuous W, or that X, Y and W have a HCG distribution for a discrete W. Then the regression coefficient β is uniformly collapsible over W if and only if* (a) $Y \perp\!\!\!\perp W|X$ *or* (b) $W \perp\!\!\!\perp X$.

Corollaries 1 and 2 extend the results on collapsibility of linear models over a discrete W (Wermuth (1989) and Geng and Asano (1993)) to the cases where W may be a continuous or an ordinal discrete variable.

4.2. Logistic regression models

For a discrete W with I levels, assume that Y, X and W have a strictly positive joint density function, and that a logistic regression model of Y on X, conditional on $W = i$, is

$$\log\frac{P(Y=1|X=x,W=i)}{P(Y=0|X=x,W=i)} = \alpha(i) + \beta(i)x$$

for $i = 1, \ldots, I$. For a continuous W, assume that Y, X and W have a strictly positive joint density function, and that a logistic regression model of Y on X, conditional on $W = w$, is

$$\log \frac{P(Y=1|X=x, W=w)}{P(Y=0|X=x, W=w)} = \alpha + \beta x + \gamma w.$$

We say that the logistic regression coefficient β is uniformly collapsible over W if the partially marginal logistic regression model

$$\log \frac{P(Y=1|X=x, W \in \mathcal{I})}{P(Y=0|X=x, W \in \mathcal{I})} = \alpha(\mathcal{I}) + \beta(\mathcal{I})x$$

always obtains and $\beta(\mathcal{I}) = \beta$ for any $\mathcal{I}$, where $\mathcal{I}$ is a subset of W's levels for a discrete W, a set of consecutive levels $(i, i+1, \ldots, i+j)$ for an ordinal discrete W, or an interval of W's domain for a continuous W. In particular, when $\mathcal{I}$ is the full domain of W, we denote the marginal logistic regression as

$$\log \frac{P(Y=1|X=x)}{P(Y=0|X=x)} = \widetilde{\alpha} + \widetilde{\beta}x.$$

We say that β is simply collapsible over W if the marginal regression model holds and $\tilde{\beta} = \beta(i)$ for a discrete W, or $\tilde{\beta} = \beta$ for a continuous W.

Notice that the partial differentiation with respect to the binary Y specializes to differencing between $Y = 1$ and $Y = 0$. In logistic regression models, the mixed derivative of marginal or conditional interaction becomes the logistic regression coefficient. Thus uniform collapsibility of the logistic regression coefficient β is a particular case of unform collapsibility of the mixed derivative of interaction. From Theorem 4, we have the following corollary.

Corollary 3. *The logistic regression coefficient β is uniformly collapsible over W if and only if* (a) $Y \perp\!\!\!\perp W|X$ *or* (b) $X \perp\!\!\!\perp W|Y$.

Corollary 3 extends strong and consecutive collapsibility of logistic models over a discrete and an ordinal discrete W (Guo and Geng (1995)) and Guo, Geng and Shi (2003)) to uniform collapsibility where W may also be a continuous variable.

5. Generalization to Multivariate Cases

In the previous sections, we considered the cases that Y, X and W are univariate. As Cox and Wermuth (2003) mentioned, multivariate responses Y can often be treated one component at a time, and multivariate X are most simply studied one contrast at a time while holding other contrasts fixed. We next consider multivariate $W = (W_1, \ldots, W_p)$, say.

Definition 2. An association measure between X and Y is simply collapsible over $W = (W_1, \ldots, W_p)$ if the conditional association measure between X and

Y given $(w_1, \ldots, w_p)$ is the marginal association measure between X and Y for all $(w_1, \ldots, w_p)$. Further an association measure between X and Y is uniformly collapsible over W if the measure, conditionally on any $(W_1 \in \mathcal{I}_1, \ldots, W_p \in \mathcal{I}_p)$, is the marginal association measure between X and Y, where $\mathcal{I}_k$ is a subset of levels for a nominal W_k, a subset of consecutive levels $(i, i+1, \ldots, i+j)$ for an ordinal discrete W_k, or an interval for a continuous W_k.

From Theorems 3 and 4, and from Theorem 1 of Ma, Xie and Geng (2006), treating W as a single variable, we can immediately obtain sufficient conditions for uniform collapsibility of the mixed derivative of interaction, the expectation dependence and the distribution dependence. However, these conditions are no longer necessary for uniform collapsibility since crossed pooling of a multivariate variable's levels (e.g., $(w_1, w_2) \in \{(1,2),(2,1)\}$) is not required for uniform collapsibility.

The sufficient conditions can be weakened for simple collapsibility by partitioning W into two disjoint sets U and V (i.e., $W = (U, V)$) and applying the conditions for uniform collapsibility to U and V successively. For example, first applying condition (a) of Theorem 4 to U conditional on V, and then applying condition (b) of Theorem 4 to V, we obtain that the mixed derivative of interaction is simply collapsible over W. Notice that it is not sufficient for uniform collapsibility over W, since the application of condition (a) of Theorem 4 is conditional on V. In this way we obtain the following result.

Corollary 4. *For simple collapsibility, we have:*

1. *The mixed derivative of interaction is simply collapsible over* W *if either*
   (a) $Y \perp\!\!\!\perp U | (X, V)$ *and* $X \perp\!\!\!\perp V | Y$*, or*
   (b) $Y \perp\!\!\!\perp U | X$ *and* $X \perp\!\!\!\perp V | (Y, U)$*;*
2. *The expectation dependence is simply collapsible over* W *if either*
   (a) $Y \perp\!\!\!\perp U | (X, V)$*,* $X \perp\!\!\!\perp V$ *and the expectation dependence is homogeneous over* V*, or*
   (b) $Y \perp\!\!\!\perp U | X$*,* $X \perp\!\!\!\perp V | U$ *and the expectation dependence is homogeneous over* V *conditional on* U*;*
3. *The distribution dependence is simply collapsible over* W *if either*
   (a) $Y \perp\!\!\!\perp U | (X, V)$*,* $X \perp\!\!\!\perp V$ *and the distribution dependence is homogeneous over* V*, or*
   (b) $Y \perp\!\!\!\perp U | X$*,* $X \perp\!\!\!\perp V | U$ *and the distribution dependence is homogeneous over* V *conditional on* U*.*

Note that in the case of a joint Gaussian distribution, homogeneity of expectation dependence is always satisfied, and thus the independence conditions alone are sufficient for simple collapsibility of regression coefficients. These conditions could also be derived by repeated application of the partial inversion operator to the covariance matrix, a new calculus for real-valued square matrices introduced by Wermuth, Wiedenbeck and Cox (2006) to study the matrix representations of various multivariate statistical models.

If sets U and V are independent, conditionally on X (i.e. $U \perp\!\!\!\perp V | X$), then sufficient condition (a) for simple collapsibility in Corollary 4 is also sufficient for uniform collapsibility. Especially, conditional independence $U \perp\!\!\!\perp V | X$ holds in balanced data when V has the same distribution among groups defined by X and U.

Theorem 5. *Suppose that $U \perp\!\!\!\perp V | X$. Then we have:*

1. *The mixed derivative of interaction is uniformly collapsible over W if $Y \perp\!\!\!\perp U | (X, V)$ and $X \perp\!\!\!\perp V | Y$;*
2. *The expectation dependence is uniformly collapsible over W if $Y \perp\!\!\!\perp U | (X, V)$, $X \perp\!\!\!\perp V$, and the expectation dependence is homogeneous over V.*
3. *The distribution dependence is uniformly collapsible over W if $Y \perp\!\!\!\perp U | (X, V)$, $X \perp\!\!\!\perp V$, and the distribution dependence is homogeneous over V.*

For the mixed derivative of interaction, we can immediately obtain from the symmetry of X and Y that another sufficient condition of uniform collapsibility over W is that $U \perp\!\!\!\perp V | Y$, $X \perp\!\!\!\perp V | (Y, U)$, and $Y \perp\!\!\!\perp U | X$.

6. Discussion

The conditions for collapsibility and for avoiding effect reversal can be used for data analysis, causal inference, observational and experimental designs; the conditions for uniform collapsibility can also be used to discretize a continuous W, or to pool levels of a discrete W without changing the original association. The marginal independence $X \perp\!\!\!\perp W$ may often be achieved by proportional allocation of individuals to treatments (or to quasi-treatments), while any independence involving a response variable Y can in general not be achieved at the planning stage of a study. Thus the conditions discussed in the paper are of quite different importance for applications.

Acknowledgements

We would like to thank an associate editor and the referees for their valuable comments and suggestions that improved the presentation and structure of this paper. This research was supported by NSFC (10771007, 10431010), NBRP 2003CB715900 and 863 Project of China (2007AA01Z437).

Appendix: Proofs of Theorems

A.1. Proof of Theorem 1

First we prove that $\partial^2 \log f(x,y)/\partial x \partial y \geq 0 \Rightarrow \partial F(y|x)/\partial x \geq 0$. Since

$$\partial^2 \log f(x,y)/\partial x \partial y = \partial^2 [\log f(y|x) + \log f(x)]/\partial x \partial y = \partial^2 \log f(y|x)/\partial x \partial y \geq 0,$$

we obtain that $[\partial f(y|x)/\partial x]/f(y|x) \geq [\partial f(y'|x)/\partial x]/f(y'|x)$ for $y \geq y'$ and x. Thus we have for any x and any monotonically increasing function $g(\cdot)$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\frac{\partial f(y|x)/\partial x}{f(y|x)} - \frac{\partial f(y'|x)/\partial x}{f(y'|x)}\right][g(y)-g(y')]f(y|x)f(y'|x)dydy' \geq 0.$$

Under the regularity condition that integration and derivation are interchangeable, the above inequality implies that $\partial E[g(Y)|x]/\partial x \geq 0$ for any x. Letting $g(\cdot) = -I_{(-\infty,y)}(\cdot)$, we obtain $\partial F(y|x)/\partial x \leq 0$.

Next we show that $\partial F(y|x)/\partial x \leq 0 \Rightarrow \partial E(Y|x)/\partial x \geq 0$. By Fubini's Theorem, we have

$$E(Y|x) = \int_0^{\infty}[1-F(y|x)]dy - \int_{-\infty}^{0} F(y|x)dy\ .$$

Since $\partial F[y|x]/\partial x \geq 0$ implies that $F(y|x) \leq F(y|x')$ for $x > x'$, we get $E(Y|x) \geq E(Y|x')$, which in turn implies $\partial E(Y|x)/\partial x \geq 0$.

Finally we show that $\partial E(Y|x)/\partial x \geq 0 \Rightarrow \rho_{XY} \geq 0$. It is sufficient to show that $E(XY) - E(X)E(Y)$ is non-negative. By simple manipulation, we have

$$\begin{aligned}
E(XY) - E(X)E(Y) &= E[E(Y|X)X] - E(X)E[E(Y|X)] \\
&= \int_{-\infty}^{\infty} E(Y|x)x dF(x) - \int_{-\infty}^{\infty} x' dF(x') \int_{-\infty}^{\infty} E(Y|x)dF(x) \\
&= \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[E(Y|x) - E(Y|x')](x-x')dF(x)dF(x') \\
&= \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{x'}^{x}\frac{\partial E(Y|t)}{\partial t}dt\right](x-x')dF(x)dF(x'),
\end{aligned}$$

which is non-negative.

In the above proof, it can also be shown that strict inequality holds with positive probability for a right inequality if strict inequality holds with positive probability for a left one.

A.2. Proof of Theorem 2

For the distribution dependence, these two conditions have been shown by Cox and Wermuth (2003).

For the expectation dependence, suppose that $\partial E(Y|x,w)/\partial x \geq 0$ for all x and w. By some simple manipulation, we have

$$\frac{\partial E(Y|x)}{\partial x} = \int \left[\frac{\partial E(Y|x,w)}{\partial x} f(w|x) + E(Y|x,w) \frac{\partial f(w|x)}{\partial x} \right] dw \ .$$

If $X \perp\!\!\!\perp W$, then the second term in the integral equals zero, and thus we have

$$\frac{\partial E(Y|x)}{\partial x} = \int \left[\frac{\partial E(Y|x,w)}{\partial x} f(w|x) \right] dw \geq 0.$$

If $Y \perp\!\!\!\perp W|X$, then $\partial E(Y|x)/\partial x = \partial E(Y|x,w)/\partial x \geq 0$.

For the mixed derivation of interaction, if $Y \perp\!\!\!\perp W|X$ then

$$\partial^2 \log f(x,y)/\partial x \partial y = \partial^2 \log f(y|x)/\partial x \partial y = \partial^2 \log f(x,y|w)/\partial x \partial y.$$

Thus it cannot be reversed after marginalization over W.

A.3. Proof of Theorem 3

First we give some lemmas which can be proved by simple manipulations.

Lemma 1. *Suppose that $\mathcal{I}_1$ and $\mathcal{I}_2$ are mutually disjoint sets, then we have*

1. $E(Y|x, W \in \mathcal{I}_1) = E(Y|x, W \in \mathcal{I}_2)$ *if and only if* $E(Y|x, W \in \mathcal{I}_1 \cup \mathcal{I}_2) = E(Y|x, W \in \mathcal{I}_1)$*;*
2. *For the three equalities* $P(W \in \mathcal{I}_i|x) = P(W \in \mathcal{I}_i)$ *for* $i = 1,\ 2$ *and* $P(W \in \mathcal{I}_1 \cup \mathcal{I}_2|x) = P(W \in \mathcal{I}_1 \cup \mathcal{I}_2)$*, any two of them imply the third.*

For any positive integer n, we partition $(-n, n]$ into $2n \times 2^n$ intervals $\mathcal{I}_i^{(n)} = (a_i, b_i]$, where $a_1 = -n$, $b_{2n \times 2^n} = n$, $a_{i+1} = b_i$ and $b_i - a_i = 2^{-n}$, $i = 1, \cdots, 2n^n$. Then define $\Gamma^n = \left\{ \mathcal{I}_i^{(n)} : i = 1, \ldots, 2n \times 2^n \right\}$ and $\Gamma = \bigcup_{n=1}^{\infty} \Gamma^n$.

Lemma 2. *For any positive integer n, we have the following implications:*

$$E(Y|x, W \in \mathcal{I}) = E(Y|x),\ \forall \mathcal{I} \in \Gamma_{n+1} \Rightarrow E(Y|x, W \in \mathcal{I}) = E(Y|x),\ \forall \mathcal{I} \in \Gamma_n;$$
$$P(W \in \mathcal{I}|x) = P(W \in \mathcal{I}),\ \forall \mathcal{I} \in \Gamma_{n+1} \Rightarrow P(W \in \mathcal{I}|x) = P(W \in \mathcal{I}),\ \forall \mathcal{I} \in \Gamma_n.$$

Now recalling that uniform collapsibility and simple collapsibility are equivalent when the background variable W is binary, we have the following lemma.

Lemma 3. *When W is binary, the derivative measure of expectation dependence is uniformly collapsible over W if and only if*

1. *$[E(Y|x,w) - E(Y|x)]\partial P(W = w|x)/\partial x = 0$ for all x and w, and*

2. *the derivative measure of expectation dependence is homogeneous over W.*

Proof. We first rewrite

$$\frac{\partial E(Y|x)}{\partial x} = \frac{\partial}{\partial x}\sum_{w=0}^{1} E(Y|x,w)P(W=w|x)$$
$$= \sum_w \left\{ \frac{\partial E(Y|x,w)}{\partial x} P(W=w|x) + E(Y|x,w)\frac{\partial P(W=w|x)}{\partial x} \right\}$$
$$= \sum_w \frac{\partial E(Y|x,w)}{\partial x} P(W=w|x)$$
$$+ \{E(Y|x,W=1) - E(Y|x,W=0)\} \frac{\partial P(W=1|x)}{\partial x}.$$

For necessity, we have homogeneity from collapsibility, and thus we obtain

$$\frac{\partial E(Y|x)}{\partial x} = \frac{\partial E(Y|x)}{\partial x} + \{E(Y|x,W=1) - E(Y|x,W=0)\} \frac{\partial P(W=1|x)}{\partial x}.$$

This implies $[E(Y|x,W=1) - E(Y|x,W=0)]\,\partial P(W=1|x)/\partial x = 0$, which is equivalent to Condition 1. Condition 2 can be obtained directly from the definition of uniform collapsibility.

For sufficiency, in the case of a binary W, uniform collapsibility is reduced to simple collapsibility, and thus we need only show $\partial E(Y|x)/\partial x = \partial E(Y|x,w)/\partial x$ for all x and w. This can be obtained from the first formula in the proof.

Proof of Theorem 3. For sufficiency, note that for any subset $\mathcal{I}$ given in Definition 1, we have

$$\frac{\partial E(Y|x, W \in \mathcal{I})}{\partial x} = \frac{\partial}{\partial x}\left[\frac{\int_{w\in\mathcal{I}} E(Y|x,w)P(w|x)\mathrm{d}w}{P(W\in\mathcal{I}|x)}\right]. \tag{1}$$

If $\partial E(Y|x,w)/\partial w = 0$ for all x and w, we have $E(Y|x) = E(Y|x,w)$ for all x and w, and thus

$$\frac{\partial}{\partial x}\left[\frac{\int_{w\in\mathcal{I}} E(Y|x,w)P(w|x)\mathrm{d}w}{P(W\in\mathcal{I}|x)}\right] = \frac{\partial E(Y|x)}{\partial x}.$$

If $X \perp\!\!\!\perp W$, we can rewrite (1) as

$$\frac{\partial E(Y|x, W\in\mathcal{I})}{\partial x} = \frac{1}{P(W\in\mathcal{I})}\frac{\partial}{\partial x}\int_{w\in\mathcal{I}} E(Y|x,w)P(w)\mathrm{d}w.$$

Since the derivative measure is homogeneous, it becomes

$$\frac{1}{P(W \in \mathcal{I})} \int_{w \in \mathcal{I}} \frac{\partial E(Y|x,w)}{\partial x} P(w) \mathrm{d}w = \frac{\partial E(Y|x,w)}{\partial x}$$

for any w. Thus sufficiency is proved.

For the necessity part, we discuss separately the cases that W is binary, nominal (including ordinal) with more than two levels, and continuous.

(i) *The case of a binary* W

According to Lemma 3, we need only to show that the two conditions of Lemma 3 imply Condition (a) or (b). Suppose that (a) does not hold, that is, there exists x_* such that $E(Y|x_*, W = 1) - E(Y|x_*) \neq 0$. According to the definition of uniform collapsibility of the derivative measure of expectation dependence, we have that $\partial[E(Y|x, W = 1) - E(Y|x)]/\partial x = 0$ for all x. Thus we obtain $E(Y|x, W = 1) - E(Y|x) \neq 0$ for all x. From Condition 1 of Lemma 3, we have $\partial P(W = 1|x)/\partial x = 0$ for all x, which means $X \perp\!\!\!\perp W$. The homogeneity of the measure is directly shown by Definition 1. We have thus shown that at least one of (a) or (b) holds.

(ii) *The case that* W *is nominal or ordinal with more than two levels*

We use mathematical induction to show that Condition (a) or (b) holds. The homogeneity of the derivative measure of expectation dependence follows directly from Definition 1. Assume that $W \in \{1, \ldots, K\}$ with $K \geq 3$. In the following proof of the necessity, we only use subsets of W's consecutive levels $(i, i+1, \ldots, i+j)$ for a nominal W in an arbitrary level ordering.

First, we consider the case of $K = 3$. In case (i), we have shown that Condition (a) or (b) of Theorem 3 holds for $K = 2$. Applying it repeatedly to the case of $K = 3$, we have that for a binary background variable with two levels, one combined level $\{i, i+1\}$ and the other single level $\{1, 2, 3\} \setminus \{i, i+1\}$ for $i = 1$ or 2,

$$E(Y|x, W \in \{i, i+1\}) = E(Y|x) \tag{2}$$

or

$$P(W \in \{i, i+1\}|x) = P(W \in \{i, i+1\}), \tag{3}$$

and that for a binary background with two single levels i and $i+1$:

$$E(Y|x, W = i) = E(Y|x, W = i+1) = E(Y|x, W \in \{i, i+1\}) \tag{4}$$

or

$$P(W = i|x, W \in \{i, i+1\}) = P(W = i|W \in \{i, i+1\}). \tag{5}$$

We show below that the above equations implies Condition (a) or (b). For simplicity, let $(j)_i$ denote that equation (j) holds for i. For example, $(2)_1$ means that (2) holds for $i = 1$. Enumerate all possible equations as follows

$$\begin{array}{cccc} (2)_1 & (2)_2 & (4)_1 & (4)_2 \\ (3)_1 & (3)_2 & (5)_1 & (5)_2, \end{array}$$

where the first row corresponds to Condition (a), and the second (b), for each binary background. Thus, at least one equation holds for each column. From this, we obtain that at least two equations in $\{(2)_1, (2)_2, (4)_1, (4)_2\}$ hold, or at least two in $\{(3)_1, (3)_2, (5)_1, (5)_2\}$ hold. From the first result in Lemma 1, we can show that any two in $\{(2)_1, (2)_2, (4)_1, (4)_2\}$ imply that Condition (a) holds. From the second result in Lemma 1, it can be shown that each pair of $\{(3)_1, (3)_2\}$ and $\{(3)_i, (5)_j\}$ for $i,\ j = 1,\ 2$, implies that condition (b) holds. Below, we show that $\{(5)_1, (5)_2\}$ also implies $X \perp\!\!\!\perp W$. If the pair $\{(5)_1, (5)_2\}$ holds, then for $i = 1$ and 2,

$$P(W = i|x)/P(W \in \{1, 2\}|x) = P(W = i)/P(W \in \{1, 2\})$$

and, for $i = 2$ and 3,

$$P(W = i|x)/P(W \in \{2, 3\}|x) = P(W = i)/P(W \in \{2, 3\}).$$

Dividing them side-by-side, we obtain

$$P(W \in \{1, 2\}|x)/P(W \in \{2, 3\}|x) = P(W \in \{1, 2\})/P(W \in \{2, 3\}),$$

which implies

$$P(W \in \{1, 2\}|x)/P(W \in \{1, 2\}) = P(W \in \{2, 3\}|x)/P(W \in \{2, 3\}).$$

From the above equations, we have that for $i = 1$, 2 and 3,

$$P(W = i) = P(W = i|x)P(W \in \{1, 2\})/P(W \in \{1, 2\}|x).$$

Summing over i, we get $P(W \in \{1, 2\})/P(W \in \{1, 2\}|x) = 1$. Thus we have proved that $P(W = i) = P(W = i|x)$ for $i = 1$, 2 and 3, i.e., Condition (b) holds.

Next we suppose that the conclusion is true for $K = n$. As in the above proof going from $K = 2$ to $K = 3$, we merge consecutive levels, j and $j+1$, of W to obtain a background W_j with n levels: $1, \ldots, j-1, \{j, j+1\}, j+2, \ldots, n+1$. By supposition, Condition (a) or (b) holds for W_j. Since $K = n+1 \geq 4$, one of (a) and (b) holds for at least two different backgrounds, say $W_{j'}$ and $W_{j''}$. Below we show that (a) or (b) holds for the case of $K = n + 1$.

If (a) holds for $W_{j'}$ and $W_{j''}$, we have $E(Y|x, W_{j'} = \{j', j'+1\}) = E(Y|x, W_{j'} = i) = E(Y|x)$ for $i \neq j'$ or $j'+1$, and $E(Y|x, W_{j''} = \{j'', j''+1\}) = E(Y|x, W_{j''} = i) = E(Y|x)$ for $i \neq j''$ or $j'' + 1$. By the first part of Lemma 1, we have $E(Y|x, W = i) = E(Y|x)$, $i = 1, \ldots, n+1$, that is, Condition (a) holds.

If Condition (b) holds for $W_{j'}$ and $W_{j''}$, we have $P(W_{j'} = \{j', j'+1\}|x) = P(W_j = \{j', j'+1\})$ and $P(W_{j'} = i|x) = P(W_{j'} = i)$ for $i \neq j'$ or $j'+1$ and $P(W_{j''} = \{j'', j''+1\}|x) = P(W_{j''} = \{j'', j''+1\})$ and $P(W_{j''} = i|x) = P(W_{j''} = i)$ for $i \neq j''$, $j''+1$. By the second part of Lemma 1, we have that $P(W = i|x) = P(W = i)$ for $i = 1, \ldots, n+1$, that is, (b) holds.

Thus the necessity for $K = n+1$ holds, and we have proved necessity.

(iii) *The case of a continuous W*

Define a sequence of ordinal random variables

$$Z_n = -nI_{W \in (-\infty, -n]} + nI_{W \in (n, \infty]} + \sum_{i=1}^{2n \times 2^n} b_i I_{W \in \mathcal{I}_i^{(n)}}, \; n = 1, 2, \cdots .$$

Then Z_n converges pointwise to W as $n \to \infty$.

Uniform collapsibility of the expectation dependence over W implies uniform collapsibility over Z_n. According to case (ii), we obtain that for each Z_n, condition (a) or (b) holds.

Condition (a) holding for Z_n implies that for any $z_n \in Ran(Z_n)$,

$$E(Y|x, Z_n = z_n) = E(Y|x). \tag{6}$$

By Lemma 2, (6) also holds for all Z_k with $k \leq n$. Condition (b) holding for Z_n implies that for any $z_n \in Ran(Z_n)$,

$$P(Z_n \leq z_n|x) = P(Z_n \leq z_n). \tag{7}$$

Again by Lemma 2, (7) also holds for all Z_k with $k \leq n$. Thus we have that (6) holds for all $n < \infty$ or that (7) holds for all $n < \infty$. We consider these situations separately.

Suppose (6) holds for all $n < \infty$. By the Continuous Mapping Theorem, we have

$$E(Y|x, W) = \lim_{n \to \infty} E(Y|x, Z_n) = \lim_{n \to \infty} E(Y|x) = E(Y|x).$$

Thus Condition (a) holds for W.

Next suppose (7) holds for all $n < \infty$. Again, by applying the Continuous Mapping Theorem twice, we get

$$F(W|x) = \lim_{n \to \infty} F(Z_n|x) = \lim_{n \to \infty} F(Z_n) = F(W).$$

This means that (b) holds for W.

A.4. Proof of Theorem 4

For sufficiency, note that for any subset $\mathcal{I}$ specified in Definition 1, we have

$$\begin{aligned}\frac{\partial^2 \log f(x,y|W\in\mathcal{I})}{\partial x\partial y} &= \frac{\partial^2 \log[f(x|y,W\in\mathcal{I})f(y|W\in\mathcal{I})]}{\partial x\partial y}\\ &= \frac{\partial^2 \log[f(y|x,W\in\mathcal{I})f(x|W\in\mathcal{I})]}{\partial x\partial y}.\end{aligned}$$

If $Y \perp\!\!\!\perp W|X$, we get

$$\frac{\partial^2 \log[f(y|x,W\in\mathcal{I})f(x|W\in\mathcal{I})]}{\partial x\partial y} = \frac{\partial^2 \log[f(y|x)f(x|W\in\mathcal{I})]}{\partial x\partial y} = \frac{\partial^2 \log f(x,y)}{\partial x\partial y}.$$

If $X \perp\!\!\!\perp W|Y$, we get

$$\frac{\partial^2 \log[f(x|y,W\in\mathcal{I})f(y|W\in\mathcal{I})]}{\partial x\partial y} = \frac{\partial^2 \log[f(x|y)f(y|W\in\mathcal{I})]}{\partial x\partial y} = \frac{\partial^2 \log f(x,y)}{\partial x\partial y}.$$

Thus sufficiency is proved.

For necessity, as in the proof of Theorem 3, we divide our proof into three parts. When W is binary we have, for $i=0$ and 1,

$$\frac{\partial^2 \log f(x,y|W=i)}{\partial x\partial y} = \frac{\partial^2 \log[f(W=i|x,y)f(x,y)]}{\partial x\partial y} = \frac{\partial^2 \log f(x,y)}{\partial x\partial y}.$$

Thus we have $\partial^2 \log f(W=i|x,y)/\partial x\partial y = 0$ for $i=0$ and 1, which can be rewritten as

$$f(W=i|x,y)\frac{\partial^2 f(W=i|x,y)}{\partial x\partial y} = \frac{\partial f(W=i|x,y)}{\partial x}\frac{\partial f(W=i|x,y)}{\partial y}. \tag{8}$$

From (8), we get

$$\frac{\partial^2 f(W=1|x,y)}{\partial x\partial y} = \frac{\partial f(W=1|x,y)}{\partial x}\frac{\partial f(W=1|x,y)}{\partial y} = 0. \tag{9}$$

Actually, (9) implies $Y \perp\!\!\!\perp W|X$ or $X \perp\!\!\!\perp W|Y$. If $X \not\perp\!\!\!\perp W|Y$, then there exist x_* and y_* such that $\partial f(W=1|x_*,y_*)/\partial x \neq 0$. Since $\partial^2 f(W=1|x,y)/\partial x\partial y = 0$, we have $\partial f(W=1|x_*,y)/\partial x \neq 0$ for all y, and furthermore $\partial f(W=1|x,y)/\partial x \neq 0$ for all x and y. Thus, according to (9), we have $\partial f(W=1|x,y)/\partial y = 0$ for all x and y, which means $Y \perp\!\!\!\perp W|X$.

Similar to the proof of Theorem 3, we can prove the necessity for the cases that W is nominal or ordinal with more than two levels, or W is continuous.

A.5. Proof of Theorem 5

For mixed derivative of interaction, we have from $U \perp\!\!\!\perp V|X$ and $Y \perp\!\!\!\perp U|(X, V)$ that

$$\begin{aligned}
&\frac{\partial^2 \log f(x, y|U \in \mathcal{I}_1, V \in \mathcal{I}_2)}{\partial x \partial y} = \frac{\partial^2 \log f(x, y, U \in \mathcal{I}_1, V \in \mathcal{I}_2)}{\partial x \partial y} \\
&= \frac{\partial^2 \log \int_{u \in \mathcal{I}_1, v \in \mathcal{I}_2} f(y|x, u, v) f(x, u, v) dudv}{\partial x \partial y} \\
&= \frac{\partial^2 \log \int_{u \in \mathcal{I}_1, v \in \mathcal{I}_2} f(y|x, v) f(x, v) f(u|x) dudv}{\partial x \partial y} \\
&= \frac{\partial^2 \log\{\int_{v \in \mathcal{I}_2} f(y|x, v) f(x, v) dv \int_{u \in \mathcal{I}_1} f(u|x) du\}}{\partial x \partial y} \\
&= \frac{\partial^2 \log \int_{v \in \mathcal{I}_2} f(y|x, v) f(x, v) dv}{\partial x \partial y} \\
&= \frac{\partial^2 \log f(x, y|V \in \mathcal{I}_2)}{\partial x \partial y}.
\end{aligned}$$

Further, from $X \perp\!\!\!\perp V|Y$, we have

$$\frac{\partial^2 \log f(x, y|V \in \mathcal{I}_2)}{\partial x \partial y} = \frac{\partial^2 [\log f(x|y, V \in \mathcal{I}_2) + \log f(y|V \in \mathcal{I}_2)]}{\partial x \partial y} = \frac{\partial^2 \log f(x, y)}{\partial x \partial y}.$$

We proved that the mixed derivative of interaction is uniformly collapsible over U and V.

Next, for the expectation dependence we have, from $U \perp\!\!\!\perp V|X$ and $Y \perp\!\!\!\perp U|(X, V)$,

$$\begin{aligned}
&\frac{\partial E(Y|x, U \in \mathcal{I}_1, V \in \mathcal{I}_2)}{\partial x} = \frac{\partial}{\partial x}\left[\frac{\int_{u \in \mathcal{I}_1, v \in \mathcal{I}_2} E(Y|x, u, v) f(u, v|x) dudv}{P(U \in \mathcal{I}_1, V \in \mathcal{I}_2|x)}\right] \\
&= \frac{\partial}{\partial x}\left[\frac{\int_{u \in \mathcal{I}_1, v \in \mathcal{I}_2} E(Y|x, v) f(v|x) f(u|x) dudv}{P(U \in \mathcal{I}_1|x) P(V \in \mathcal{I}_2|x)}\right] \\
&= \frac{\partial}{\partial x}\left[\frac{\int_{v \in \mathcal{I}_2} E(Y|x, v) f(v|x) dv}{P(V \in \mathcal{I}_2|x)}\right] = \frac{\partial E(Y|x, V \in \mathcal{I}_2)}{\partial x}.
\end{aligned}$$

Since $X \perp\!\!\!\perp V$ and the expectation dependence is homogeneous over V, we have from Theorem 3 that $\partial E(Y|x, V \in \mathcal{I}_2)/\partial x = \partial E(Y|x)/\partial x$. Thus, under the conditions stated in 2, we have proved that the expectation dependence is uniformly collapsible over U and V.

Replacing $E(\cdot)$ with $F(\cdot)$ in the above proof, we can immediately prove sufficiency of conditions in 3 for uniform collapsibility of the distribution dependence over U and V.

References

Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. Wiley, New York.

Cox, D. R. and Wermuth, N. (2003). A general condition for avoiding effect reversal after marginalization. *J. Roy. Statist. Soc. Ser. B* **65**, 937-941.

Ducharme, G. R. and Lepage, Y. (1986). Testing collapsibility in contingency tables. *J. Roy. Statist. Soc. Ser. B* **48**, 197-205.

Geng, Z. (1992). Collapsibility of relative risks in contingency tables with a response variable. *J. Roy. Statist. Soc. Ser. B* **54**, 585-593.

Geng, Z. and Asano, C. (1993). Strong collapsibility of association measures in linear models. *J. Roy. Statist. Soc. Ser. B* **55**, 741-747.

Guo, J. H. and Geng, Z. (1995). Collapsibility of logistic regression coefficients. *J. Roy. Statist. Soc. Ser. B* **57**, 263-267.

Guo, J. H., Geng, Z. and Shi, N. Z. (2003). Consecutive collapsibility of logistic regression coefficients over an ordinal background. *J. Statist. Plann. Inference* **115**, 59-67.

Holland, P. W. and Wang, Y. J. (1987). Dependence functions for continuous bivariate densities. *Comm. Statist. Theory Methods* **16**, 863-876.

Ma, Z. M., Xie, X. C. and Geng, Z. (2006). Collapsibility of distribution dependence. *J. Roy. Statist. Soc. Ser. B* **68**, 127-133.

Mari, D. D. and Kotz, S. (2001). *Correlation and Dependence.* Imperial College Press, London.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *J. Roy. Statist. Soc. Ser. B* **13**, 238-241.

Wermuth, N. (1987). Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable. *J. Roy. Statist. Soc. Ser. B* **49**, 353-364.

Wermuth, N. (1989). Moderating effects of subgroups in linear models. *Biometrika* **76**, 81-92.

Wermuth, N., Wiedenbeck, M. and Cox, D. R. (2006). Partial inversion for linear systems and partial closure of independence graphs. *BIT Numerical Mathematics* **46**, 883-901.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics.* Wiley, New York.

Whittemore, A. S. (1978). Collapsibility of multidimensional contingency tables. *J. Roy. Statist. Soc. Ser. B* **40**, 328-340.

Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika* **2**, 121-134.

School of Mathematical Sciences, LMAM, Peking University, Beijing 100871, China.

E-mail: xie1981@gmail.com

Department of Statistics, Stanford University, Stanford, CA 94305, USA.

E-mail: zongming@stanford.edu

School of Mathematical Sciences, LMAM, Peking University, Beijing 100871, China.

E-mail: zgeng@math.pku.edu.cn

(Received October 2006; accepted March 2007)