2-2006

Uniform Collapsibility of Distribution Dependence Over a Nominal, Ordinal or Continuous Background

Zongming Ma
University of Pennsylvania

Xianchao Xie

Zhi Geng

Recommended Citation

Ma, Z., Xie, X., & Geng, Z. (2006). Uniform Collapsibility of Distribution Dependence Over a Nominal, Ordinal or Continuous Background. *Journal of the Royal Statistical Society: Series B, 68* (1), 127-133. http://dx.doi.org/10.1111/j.1467-9868.2005.00536.x

Uniform Collapsibility of Distribution Dependence Over a Nominal, Ordinal or Continuous Background

Abstract
Cox and Wermuth proposed that the partial derivative of the conditional distribution function of a random variable *Y* given another *X* is used for measuring association between two variables with arbitrary distributions. This paper shows the condition for collapsibility of the association measure.

Keywords
collapsibility, distribution dependence, Yule-Simpson paradox

Disciplines
Statistical Methodology

Uniform collapsibility of distribution dependence over a nominal, ordinal or continuous background

Zongming Ma, Xianchao Xie and Zhi Geng
LMAM, Peking University, Beijing 100871, China
zgeng@math.pku.edu.cn

Summary. Cox and Wermuth (2003) proposed that the partial derivative of the conditional distribution function of a random variable Y given another X can be used for measuring association between the two variables with arbitrary distributions. This paper shows the condition for collapsibility of the association measure.

Keywords: Collapsibility; Distribution dependence; Yule-Simpson paradox

1. Introduction

An association measure may be reversed by omitting a background variable, which is called Yule-Simpson Paradox (Yule, 1903; Simpson, 1951). For simplicity, we consider three random variables Y, X and W, such as a response of interest, a treatment and a background variable. Cox and Wermuth (2003) proposed the partial derivative of the conditional distribution function $F(y|x)$ of Y given $X = x$ with respect to a continuous X

$$\frac{\partial F(y|x)}{\partial x}$$

as a general measure of dependence of Y on X with an arbitrary density, called the distribution dependence below. If X is discrete, the partial differentiation is replaced by differencing between adjacent levels, that is, $F(y|x+1) - F(y|x)$. Further, if Y is a binary response and X is a binary treatment, the difference becomes the risk difference $P(Y = 1|X = 1) - P(Y = 1|X = 0)$. If the partial derivation satisfies

$\partial F(y|x)/\partial x \leq 0$ for all y and x with strict inequality in a region of positive probability, then $P(Y > y|X = x)$ is increasing in x for all y, and we say that the distribution dependence of Y on X is stochastically increasing with X. Similarly the distribution dependence of Y on X conditional on W can be defined as $\partial F(y|x,w)/\partial x$, where $F(y|x,w)$ is the conditional distribution function of Y given $X = x$ and $W = w$. When W is not observed or omitted, the distribution dependence of Y on X may be reversed, that is, $\partial F(y|x,w)/\partial x \leq 0$ (or ≥ 0) for all y, x and w, but $\partial F(y|x)/\partial x > 0$ (or < 0) for some y or x. Cox and Wermuth (2003) showed that either $Y \perp\!\!\!\perp W|X$ or $X \perp\!\!\!\perp W$ is a sufficient condition for avoiding distribution dependence reversal after marginalization over W.

In many investigations, such as epidemiological studies, we wish to study whether a background variable W may influence the dependence of Y on X, or whether W is a confounder. In a clinical study, we may also wish to discretize a continuous background variable without changing the original dependence of response on treatment. In this paper, we discuss collapsibility of distribution dependence. We say that a distribution dependence is collapsible if the dependence conditional on W remains unchanged after marginalization over W. We show a condition for collapsibility of distribution dependence, which is a revision of Cox and Wermuth's (2003) condition for avoiding dependence reversal.

Section 2 defines collapsibility of distribution dependence and shows a condition for the collapsibility over a discrete or continuous background variable W. The proof of our main result is given in Appendix.

2. Collapsibility of distribution dependence

For both cases of discrete and continuous background variables, we define homogeneity and collapsibility of distribution dependence as follows. We say that the conditional distribution dependence $\partial F(y|x,w)/\partial x$ is homogeneous over the background variable W if $\partial F(y|x,w)/\partial x = \partial F(y|x,w')/\partial x$ for all y, x and $w \neq w'$. The simple collapsibility of distribution dependence means that the conditional distribution dependence $\partial F(y|x,w)/\partial x$ equals the marginal distribution dependence

$\partial F(y|x)/\partial x$ for all y, x and w. When the distribution dependence is not simply collapsible over W, we call W a moderating variable.

Definition 1. The conditional distribution dependence $\partial F(y|x,w)/\partial x$ is uniformly collapsible over W if $\partial F(y|x, W \in \mathcal{I})/\partial x = \partial F(y|x)/\partial x$ for all y, x and $\mathcal{I}$, 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.

When the distribution dependence is not uniformly collapsible over W, we call W an occasional moderating variable. Assume that the distribution of Y, X and W satisfies the regular condition, that is, differentiation and integration are exchangeable.

Theorem 1. The distribution dependence is uniformly collapsible over W if and only if

(a) $Y \perp\!\!\!\perp W|X$ or

(b) $X \perp\!\!\!\perp W$ and the distribution dependence is homogeneous over W.

In a special case where Y is a binary response, X is a binary treatment and W is a discrete background variable, the distribution dependence is replaced by the risk difference. Theorem 1 gives a necessary and sufficient condition for uniform collapsibility of risk difference over the discrete background variable W. The condition coincides with the condition for collapsibility of relative risks for Y with respect to X presented by Wermuth (1987, propositions 1 and 4) and Geng (1992, theorem 2).

From definitions, it can be seen that uniform collapsibility implies simple collapsibility, and simple collapsibility in turn implies no reversal. Cox and Wermuth (2003) showed that either $Y \perp\!\!\!\perp W|X$ or $X \perp\!\!\!\perp W$ is a sufficient condition for avoiding reversal of distribution dependence. Theorem 1 shows that $Y \perp\!\!\!\perp W|X$ is also a sufficient condition for both uniform collapsibility and simple collapsibility, but $X \perp\!\!\!\perp W$ is not unless the dependence is homogeneous. Theorem 1 can also be used to group levels of a discrete background variable or to discretize a continuous background variable. If the domain of W can be partitioned into K regions $\mathcal{I}_1, \ldots, \mathcal{I}_K$, and the condition in theorem 1 is satisfied separately for each region $\mathcal{I}_k$, then the background variable

W can be recategorized into a crude background variable with K levels such that the dependence in each region keeps the same as the original distribution dependence.

Cox and Wermuth (2003) discussed the general case with multivariate Y, X and W. Their argumentation also applies to our results on simple collapsibility and confounding.

Acknowledgements

This research was supported by NSFC, NBRP 2003CB715900 and the Presidential funds of Peking University.

Appendix: Proof of theorem 1

First we give the notations and lemmas to be used in proofs of theorems 1. For an arbitrary natural number n, we partition the interval $(-n, n]$ uniformly into $2n \times 2^n$ small 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 $b_i - a_i = 1/2^n$ for $i = 1, \ldots, 2n \times 2^n$. Let Γ^n denote the class $\{\mathcal{I}_i^{(n)} : i = 1, \ldots, 2n \times 2^n\}$ and Γ denote the union $\bigcup_{n=1}^{\infty} \Gamma^n$.

Lemma 1. Suppose that $\mathcal{I}_1$ and $\mathcal{I}_2$ are two disjoint subsets of W. Then we have:

(1) If $F(y|x, W \in \mathcal{I}_1) = F(y|x, W \in \mathcal{I}_2)$, then $F(y|x, W \in \mathcal{I}_1 \cup \mathcal{I}_2) = F(y|x, W \in \mathcal{I}_1)$;

(2) If $F(y|x, W \in \mathcal{I}_1) = F(y|x, W \in \mathcal{I}_1 \cup \mathcal{I}_2)$, then $F(y|x, W \in \mathcal{I}_2) = F(y|x, W \in \mathcal{I}_1 \cup \mathcal{I}_2)$;

(3) If $P(W \in \mathcal{I}_i|x) = P(W \in \mathcal{I}_i)$ for $i = 1$ and 2, then $P(W \in \mathcal{I}_1 \cup \mathcal{I}_2|x) = P(W \in \mathcal{I}_1 \cup \mathcal{I}_2)$;

(4) If $P(W \in \mathcal{I}_1|x) = P(W \in \mathcal{I}_1)$ and $P(W \in \mathcal{I}_1 \cup \mathcal{I}_2|x) = P(W \in \mathcal{I}_1 \cup \mathcal{I}_2)$, then $P(W \in \mathcal{I}_2|x) = P(W \in \mathcal{I}_2)$.

Proof. The proof is obvious and thus omitted. □

Lemma 2. For an arbitrary natural number n, we have:

(1) If $F(y|x, W \in \mathcal{I}_\alpha^{(n+1)}) = F(y|x)$ for all α, then $F(y|x, W \in \mathcal{I}_\beta^{(n)}) = F(y|x)$ for all β;

(2) If $P(W \in \mathcal{I}_{\alpha}^{(n+1)}|x) = P(W \in \mathcal{I}_{\alpha}^{(n+1)})$ for all α, then $P(W \in \mathcal{I}_{\beta}^{(n)}|x) = P(W \in \mathcal{I}_{\beta}^{(n)})$ for all β.

Proof. Let $\mathcal{I}_{\beta}^{n} = \mathcal{I}_{\alpha_1}^{(n+1)} \cup \mathcal{I}_{\alpha_2}^{(n+1)}$. For the result 1, we have

$$\begin{aligned} & F(y|x, w \in \mathcal{I}_{\beta}^{(n)}) = \frac{P(Y \leq y, w \in \mathcal{I}_{\beta}^{(n)}|x)}{P(w \in \mathcal{I}_{\beta}^{(n)}|x)} \\ = & \frac{P(Y \leq y, w \in \mathcal{I}_{\alpha_1}^{(n+1)}|x) + P(Y \leq y, w \in \mathcal{I}_{\alpha_2}^{(n+1)}|x)}{P(w \in \mathcal{I}_{\beta}^{(n)}|x)} \\ = & \frac{F(y|x, w \in \mathcal{I}_{\alpha_1}^{(n+1)})P(w \in \mathcal{I}_{\alpha_1}^{(n+1)}|x) + F(y|x, w \in \mathcal{I}_{\alpha_2}^{(n+1)})P(w \in \mathcal{I}_{\alpha_2}^{(n+1)}|x)}{P(w \in \mathcal{I}_{\beta}^{(n)}|x)} \\ = & \frac{F(y|x)[P(w \in \mathcal{I}_{\alpha_1}^{(n+1)}|x) + P(w \in \mathcal{I}_{\alpha_2}^{(n+1)}|x)]}{P(w \in \mathcal{I}_{\beta}^{(n)}|x)} = F(y|x). \end{aligned}$$

For the result 2, we have

$$\begin{aligned} P(W \in \mathcal{I}_{\beta}^{(n)}|x) &= P(W \in \mathcal{I}_{\alpha_1}^{(n+1)}|x) + P(W \in \mathcal{I}_{\alpha_2}^{(n+1)}|x) \\ &= P(W \in \mathcal{I}_{\alpha_1}^{(n+1)}) + P(W \in \mathcal{I}_{\alpha_2}^{(n+1)}) \\ &= P(W \in \mathcal{I}_{\beta}^{(n)}). \end{aligned}$$

Lemma 3. Suppose that W is binary. Then the distribution dependence is uniformly collapsible over W if and only if

(1) $[F(y|x, w) - F(y|x)]\partial P(W = w|x)/\partial x = 0$ for all y, x and w, and

(2) the distribution dependence is homogeneous over W.

Proof. For the necessity, we have

$$\begin{aligned} & \frac{\partial F(y|x)}{\partial x} = \frac{\partial}{\partial x}[P(Y \leq y, W = 0|x) + P(Y \leq y, W = 1|x)] \\ = & \frac{\partial}{\partial x}[F(y|x, W = 0)P(W = 0|x) + F(y|x, W = 1)P(W = 1|x)] \\ = & \frac{\partial}{\partial x}F(y|x, W = 0)P(W = 0|x) + \frac{\partial}{\partial x}F(y|x, W = 1)P(W = 1|x) \\ & + F(y|x, W = 0)\frac{\partial}{\partial x}P(W = 0|x) + F(y|x, W = 1)\frac{\partial}{\partial x}P(W = 1|x) \\ = & \frac{\partial F(y|x)}{\partial x} + F(y|x, W = 0)\frac{\partial}{\partial x}P(W = 0|x) + F(y|x, W = 1)\frac{\partial}{\partial x}P(W = 1|x). \end{aligned}$$

Thus we have $[F(y|x,W=0)-F(y|x,W=1)]\partial P(W=w|x)/\partial x=0$, that is, the condition (1). Condition (2) is immediately from the definition of uniform collapsibility.

For the sufficiency, we only need to show that $\partial F(y|x)/\partial x=\partial F(y|x,w)/\partial x$ for all y, x and w since W is binary. We have

$$\frac{\partial F(y|x)}{\partial x}=\sum_{w=0}^{1}\frac{\partial}{\partial x}[F(y|x,w)P(w|x)]$$
$$=\sum_{w=0}^{1}\frac{\partial}{\partial x}F(y|x,w)P(w|x)+\sum_{w=0}^{1}F(y|x,w)\frac{\partial}{\partial x}P(w|x),$$

where $\sum_{w=0}^{1}F(y|x,w)\partial P(w|x)/\partial x=\sum_{w=0}^{1}F(y|x)\partial P(w|x)/\partial x=0$ according to condition (1). Thus according to the homogeneity of distribution dependence, we have

$$\frac{\partial}{\partial x}F(y|x)=\sum_{w=0}^{1}\frac{\partial}{\partial x}F(y|x,w)P(w|x)$$
$$=\frac{\partial}{\partial x}F(y|x,w)\sum_{w=0}^{1}P(w|x)=\frac{\partial}{\partial x}F(y|x,w).\quad\square$$

Hereunder we simply call $\mathcal{I}$ a normal set of W if $\mathcal{I}$ is a set of consecutive levels for a discrete W or $\mathcal{I}$ is an interval for a continuous W.

Proof of Theorem 1. For sufficiency, we consider the conditions (a) and (b) separately. For any subset $\mathcal{I}$ defined in definition 1, we have

$$\frac{\partial F(y|x,w\in\mathcal{I})}{\partial x}=\frac{\partial}{\partial x}\left[\frac{P(Y\le y,w\in\mathcal{I},|x)}{P(w\in\mathcal{I}|x)}\right].\qquad(1)$$

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

$$\frac{\partial}{\partial x}\left[\frac{F(y|x)P(w\in\mathcal{I},|x)}{P(w\in\mathcal{I}|x)}\right]=\frac{\partial F(y|x)}{\partial x}.$$

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

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

Since $\partial F(y|x,w)/\partial x$ is homogeneous, it becomes

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

for any w. Thus we showed the sufficiency.

For the necessity, we discuss separately the cases that W is binary, nominal (including ordinal) with more than 2 levels and continuous. For the case that W is binary, according to lemma 3, we need only to show that two conditions of lemma 3 implies (a) or (b). Suppose that (a) does not hold, that is, there exists y_0 and x_0 such that $F(y_0|x_0, W=1) - F(y_0|x_0) \neq 0$. According to the definition of uniform collapsibility, we have that for all x

$$\frac{\partial}{\partial x}[F(y_0|x, W=1) - F(y_0|x)] = 0,$$

that is, $F(y_0|x, W=1) - F(y_0|x)$ does not involve x. Thus we obtain that for all x

$$F(y_0|x, W=1) - F(y_0|x) = F(y_0|x_0, W=1) - F(y_0|x_0) \neq 0.$$

Condition (1) in lemma 3 implies that for all x

$$\frac{\partial}{\partial x} P(W=1|x) = 0,$$

which implies $P(W=1|x) = P(W=1)$, i.e. $X \perp\!\!\!\perp W$. The homogeneity of distribution dependence is directly from the definition of uniform collapsibility. We showed that at least one of the conditions (a) and (b) holds.

For the case that W is nominal or ordinal with more than 2 levels, we use the method of mathematical induction to show that $Y \perp\!\!\!\perp W|X$ or $X \perp\!\!\!\perp W$ holds. The homogeneity of distribution dependence is directly from the definition of uniform collapsibility. Assume that $W \in \{1, \ldots, K\}$ $(K \geq 3)$. In the following proof of the necessity, we only use subsets of W's consecutive levels $(i, i+1, \ldots, i+j)$ (i.e., normal sets) for a nominal W in an arbitrary level ordering.

First we consider the case of $K = 3$. According to the above result for a binary W, we have that for one combined level $\{i, i+1\}$ and the other single level $\{1,2,3\} \setminus \{i, i+1\}$, where $i = 1$ or 2:

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

or

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

and that for two single levels i and $i+1$

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

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

$$\begin{array}{cccc} (2.1) & (2.2) & (4.1) & (4.2) \\ (3.1) & (3.2) & (5.1) & (5.2), \end{array}$$

and we find that at least one equation holds for each column. Thus 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. According to lemma 1, we can easily find that any two in $\{(2.1), (2.2), (4.1), (4.2)\}$ implies $Y \perp\!\!\!\perp W|X$, and that each pair of $\{(3.1), (3.2)\}$ and $\{(3.i), (5.j)\}$ for $i, j = 1, 2$ implies $X \perp\!\!\!\perp W$. Below we show that $\{(5.1), (5.2)\}$ also implies $X \perp\!\!\!\perp W$. If the pair $\{(5.1), (5.2)\}$ holds, we have that 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 gets $P(W \in \{1, 2\}|x)/P(W \in \{2, 3\}|x) = P(W \in \{1, 2\})/P(W \in \{2, 3\})$, 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 two equations, we have $P(W = i) = P(W = i|x)P(W \in \{1, 2\})/P(W \in \{1, 2\}|x)$ for $i = 1$, 2 and 3. Summing over W gets $P(W \in \{1, 2\})/P(W \in \{1, 2\}|x) = 1$. So we get that $P(W = i) = P(W = i|x)$ for $i = 1$, 2 and 3, that is, $X \perp\!\!\!\perp W$.

Next we assume that the conclusion is true for $K = n$. For $K = n+1$, we merge the two consecutive levels of W, j and $j+1$, into one level $\{j, j+1\}$ for $1 \le j \le n$, and define a new variable W_j with n levels: $1, \dots, j-1, \{j, j+1\}, j+2, \dots, n+1$. According to definition 1, distribution dependence is also uniformly collapsible over W_j. Thus, for each j, we have $Y \perp\!\!\!\perp W_j | X$ or $X \perp\!\!\!\perp W_j$. Since $K = n+1 \ge 4$, at least one of $Y \perp\!\!\!\perp W_j | X$ and $X \perp\!\!\!\perp W_j$ holds for two different j, say j' and j''.

If $Y \perp\!\!\!\perp W_j | X$ for $j = j'$ and j'', we have that $F(y|x, W \in \{j', j'+1\}) = F(y|x, W_{j'} = \{j', j'+1\}) = F(y|x, W = i) = F(y|x, W_j = i) = F(y|x)$ for $i \ne j'$ or $j'+1$, and that $F(y|x, W \in \{j'', j''+1\}) = F(y|x, W_{j''} = \{j'', j''+1\}) = F(y|x, W = i) = F(y|x, W_{j''} = i) = F(y|x)$ for $i \ne j''$ or $j''+1$. According to lemma 1, we have $F(y|x, W = i) = F(y|x), i = 1, \dots, n+1$, that is, $Y \perp\!\!\!\perp W | X$.

If $X \perp\!\!\!\perp W_j$ for $j = j'$ and j'', we have that $P(W \in \{j', j'+1\}|x) = P(W_{j'} = \{j', j'+1\}|x) = P(W_j = \{j', j'+1\}) = P(W \in \{j', j'+1\})$ and $P(W = i|x) = P(W_j = i|x) = P(W_{j'} = i) = P(W = i)$ for $i \ne j'$ or $j'+1$, and that $P(W \in \{j'', j''+1\}|x) = P(W_{j''} = \{j'', j''+1\}|x) = P(W_{j''} = \{j'', j''+1\}) = P(W \in \{j'', j''+1\})$ and $P(W = i|x) = P(W_{j''} = i|x) = P(W_{j''} = i) = P(W = i)$ for $i \ne j''$, $j''+1$. Also according to lemma 1, we have that $P(W = i|x) = P(W = i)$ for $i = 1, \dots, n+1$, that is, $X \perp\!\!\!\perp W$.

Thus we showed the necessity for $K = n+1$, and thus we proved the necessity for a nominal or an ordinal W.

Finally, for the case that W is continuous, we can define a discrete random variable Z_n with $2n \times 2^n + 2$ levels for an arbitrary natural number n. $Z_n = 0$ denotes $W \in (-\infty, n]$, $Z_n = i$ denotes $W \in \mathcal{I}_i^{(n)}$ for $i = 1, \dots, 2n \times 2^n$, and $Z_n = 2n \times 2^n + 1$ denotes $W \in (n, \infty)$. Uniformly collapsible of distribution dependence over W implies uniformly collapsible over Z_n. According to the above results for a discrete W, we obtain that $Y \perp\!\!\!\perp Z_n | X$ or $X \perp\!\!\!\perp Z_n$ holds.

$Y \perp\!\!\!\perp Z_n | X$ implies that for $i = 1, \dots, 2n \times 2^n$

$$F(y|x, W \in \mathcal{I}_i^{(n)}) = F(y|x, Z_n = i) = F(y|x). \tag{6}$$

According to lemma 2, equation (6) also holds for all Z_k where $k \le n$. $X \perp\!\!\!\perp Z_n$ implies that for $i = 1, 2, \cdots, 2n \times 2^n$,

$$P(W \in \mathcal{I}_i^{(n)}|x) = P(Z_n = i|x) = P(W \in \mathcal{I}_i^{(n)}). \tag{7}$$

According to lemma 2, equation (7) also holds for all Z_k where $k \leq n$. Thus we have that equation (6) holds for all $n < \infty$ or that equation (7) holds for all $n < \infty$. Now we consider these two situations separately.

First consider the case that equation (6) holds for all $n < \infty$, that is, $F(y|x, W \in \mathcal{I}) = F(y|x)$ for all x, y and $\mathcal{I} \in \Gamma$. Let $f(\cdot)$ denote the density of $F(\cdot)$. Then we also have $f(y|x, W \in \mathcal{I}) = f(y|x)$. For any real value w_0, there exists a natural number n_0 such that $|w_0| \leq n_0$, and there also exists a sequence of intervals $\{\mathcal{I}_n\}_{n=n_0}^{\infty}$ such that for $n \geq n_0$

$$\mathcal{I}_n \in \Gamma^n, \quad \mathcal{I}_{n+1} \in \Gamma^{n+1}, \quad \mathcal{I}_{n+1} \subseteq \mathcal{I}_n, \quad \bigcap_{n=n_0}^{\infty} \overline{\mathcal{I}_n} = \{w_0\},$$

where $\overline{\mathcal{I}_n}$ denotes the closure of $\mathcal{I}$. From the mean value theorem, we have

$$f(y|x) = f(y|x, W \in \mathcal{I}_n) = \frac{\int_{\mathcal{I}_n} f(y, w|x)\mathrm{d}w}{\int_{\mathcal{I}_n} f(w|x)\mathrm{d}w} = \frac{f(y, w_{n_1}|x)\frac{1}{2^n}}{f(w_{n_2}|x)\frac{1}{2^n}} = \frac{f(y, w_{n_1}|x)}{f(w_{n_2}|x)},$$

which, as $n \to \infty$, tends to

$$\frac{f(y, w_0|x)}{f(w_0|x)} = f(y|x, w_0).$$

Thus we showed that $f(y|x) = f(y|x, w)$ for any x, y and w, which means $Y \perp\!\!\!\perp X|W$.

Next consider the case that equation (7) holds for all $n < \infty$, that is, $P(W \in \mathcal{I}|x) = P(W \in \mathcal{I})$ for all x and $\mathcal{I} \in \Gamma$. For any real value w_0, there exists a sequence $\{\mathcal{I}_n\}_{n=1}^{\infty} \subset \Gamma$ such that

$$(-\infty, w_0) \subset \bigcup_{n=1}^{\infty} \mathcal{I}_n \subset (-\infty, w_0], \quad \mathcal{I}_n \bigcap \mathcal{I}_m = \emptyset$$

for all $n \neq m$. Thus, we have that for all x

$$P(W \leq w_0|x) = \sum_{n=1}^{\infty} P(W \in \mathcal{I}_n|x) = \sum_{n=1}^{\infty} P(W \in \mathcal{I}_n) = P(W \leq w_0).$$

Thus we showed that $X \perp\!\!\!\perp W$. The homogeneity of distribution dependence are directly from the definition of uniform collapsibility. Thus we proved the necessity for a continuous W. $\square$

References

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

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

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

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

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