Ann Inst Stat Math (2017) 69:1155–1176
DOI 10.1007/s10463-016-0580-y

Collapsibility of some association measures and survival models

P. Vellaisamy[1]

Received: 7 October 2015 / Revised: 9 July 2016 / Published online: 18 August 2016

Abstract Collapsibility deals with the conditions under which a conditional (on a covariate W) measure of association between two random variables Y and X equals the marginal measure of association. In this paper, we discuss the average collapsibility of certain well-known measures of association, and also with respect to a new measure of association. The concept of average collapsibility is more general than collapsibility, and requires that the conditional average of an association measure equals the corresponding marginal measure. Sufficient conditions for the average collapsibility of the association measures are obtained. Some interesting counterexamples are constructed and applications to linear, Poisson, logistic and negative binomial regression models are discussed. An extension to the case of multivariate covariate W is also analyzed. Finally, we discuss the collapsibility conditions of some dependence measures for survival models and illustrate them for the case of linear transformation models.

Keywords Average collapsibility · Conditional distributions · Linear and non-linear regression models · Measures of association · Linear transformation model

1 Introduction

The study of association between two random variables arises in several applications. Several measures, nonparametric in nature, have been proposed in the literature. Often, the random variables of interest, say Y and X, may be associated because of their

✉ P. Vellaisamy
pv@math.iitb.ac.in

[1] Department of Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

association with another variable W, called a covariate or a background variable. In this case, we need to investigate the conditional association measure between Y and X given W, and compare it with the marginal association measure between Y and X. It is in general possible that the conditional association measure may be positive, while the marginal association measure may be negative. Such an effect reversal is called the Yule–Simpson paradox attributed to Yule (1903) and Simpson (1951). When Yule–Simpson paradox or the effect reversal does not occur, and a conditional measure of association equals the marginal measure, we say that the measure is collapsible over the covariate W. Collapsibility is an important issue associated with data analysis, analysis of contingency tables, causal inference, regression analysis, epidemiological studies and the design of experiments; see, for example, (Cox and Wermuth 2003; Ma et al. 2006; Xie et al. 2008) for applications and discussions.

There have been several notions of collapsibility, namely simple, strong and uniform collapsibility. These issues have been addressed in several different contexts such as the analysis of contingency tables, regression models and association measures; see, for example, (Bishop 1971; Cox 2003; Cox and Wermuth 2003; Geng 1992; Ma et al. 2006; Vellaisamy and Vijay 2007, 2008; Wermuth 1987, 1989; Whittemore 1978; Xie et al. 2008). Cox and Wermuth (2003) studied the concept of distribution dependence and discussed the conditions under which no effect reversal occurs.

We next define the (simple) collapsibility and the uniform collapsibility with respect to the conditional expectation dependence function $\partial E(Y|x, w)/\partial x$ [(see Xie et al. (2008)] and they follow similarly for other measures.

Definition 1 The conditional expectation dependence function $\partial E(Y|x, w)/\partial x$ is said to be homogeneous over W if

$$\frac{\partial E(Y|x, w)}{\partial x} = \frac{\partial E(Y|x, w')}{\partial x},$$

for all x and $w' \neq w$.

Definition 2 The conditional expectation dependence function $\partial E(Y|x, w)/\partial x$ is said to be simple collapsible over W if

$$\frac{\partial E(Y|x, w)}{\partial x} = \frac{\partial E(Y|x)}{\partial x}, \text{ for all } x \text{ and } w,$$

and uniformly collapsible over W if

$$\frac{\partial E(Y|x, W \in A)}{\partial x} = \frac{\partial E(Y|x)}{\partial x}$$

for all x and A, where A is a subset of levels when W is nominal, a subset of consecutive levels $(i, i+1, \ldots, i+j)$ when W is ordinal, and is an interval when W is a continuous random variable.

Observe that the collapsibility of a measure implies the homogeneity and the uniform collapsibility implies the collapsibility. In other words, both the collapsibility and the uniform collapsibility require the condition of homogeneity.

Xie et al. (2008) discussed the simple collapsibility and the uniform collapsibility of the following association measures:

$$
\begin{array}{lll}
\text{(i)} & \dfrac{\partial}{\partial x} E(Y \mid x) & \text{(expectation dependence)} \\
\text{(ii)} & \dfrac{\partial^2}{\partial x \partial y} \log f(x, y) & \text{(mixed derivative of interaction)} \\
\text{(iii)} & \dfrac{\partial}{\partial x} F(y \mid x) & \text{(distribution dependence).}
\end{array}
$$

They discussed also the stringency of the above measures for positive association, studied the conditions for no effect reversal (after marginalization over W) and obtained the necessary and sufficient conditions for uniform collapsibility of mixed derivative of interaction, among other results. Recently, Vellaisamy (2012) investigated the average collapsibility of distribution dependence and the quantile regression coefficients. It is shown that average collapsibility is a general concept and coincides with collapsibility under the condition of homogeneity. In addition, the average collapsibility does not require more stringent conditions than the simple collapsibility which requires an additional condition of homogeneity and thus restricts the class of joint distributions.

In the context of contingency tables, the relative risk (rr) is an important measure of association which is widely used in economics, epidemiology, sociology and probabilistic expert systems [see Geng (1992)]. Consider a 2×2 table corresponding to binary variables Y and X with p_{ij} as its cell probabilities. The rr is defined as $r = p_{1|1}/p_{1|2}$, where for example $p_{1|1} = P(Y = 1|X = 1)$. Similarly, when we have a $2 \times 2 \times 2$ table with an additional binary covariate W, let $r(1)$ and $r(2)$ denote the rr corresponding to the levels $W = 1$ and $W = 2$, respectively. The three-dimensional table is collapsible over W with respect to rr if (a) $r(1) = r(2) = r^*$ (consistency over W) and (b) $r_m = r^*$, where r_m denotes the rr corresponding to the marginal 2×2 table, obtained by summing over the levels of W. Note that consistency is a rather stronger condition which may not hold in many situations.

To motivate average collapsibility, consider for instance the following $2 \times 2 \times 2$ table:

		W	
X	Y	1	2
1	1	4	1
	2	1	4
2	1	2	2
	2	3	3

From the above table, we have $r(1) = 2$ and $r(2) = 0.5$; hence, the consistency as well as the collapsibility of rr over W does not hold. In addition, from the marginal table corresponding to Y and X, we have $r_m = 1.25$. However, $E(r(W)) = \frac{1}{2}r(1) + \frac{1}{2}r(2) =$

1.25, where we have used the empirical distribution $P(W = 1) = P(W = 2) = \frac{1}{2}$ of W to compute the expectation. Thus, we see that the average collapsibility $E(r(W)) = r_m = 1$ holds.

In this paper, the average collapsibility of the association measures expectation dependence and mixed derivative of interaction, studied by Xie et al. (2008), are investigated in detail. These measures have connections to linear regression and logistic regression models. In addition, a new measure of association, namely

$$\text{(iv)} \quad \frac{\partial}{\partial x} \log E(Y \mid x) \quad (\log - \text{expectation dependence})$$

is introduced and its average collapsibility conditions are investigated. This new measure has a direct application to Poisson and negative binomial regression models. Some of the results are then extended to the case of multivariate covariate W.

Finally, we consider the collapsibility results for certain survival models. Recently, Di Serio et al. (2009) discussed the Simpson's paradox for the survival probability and hazard rate functions for the survival models, and observed some interesting results. But, the collapsibility issues have not been addressed in the literature so far. We derive sufficient conditions for the average collapsibility of survival probability and hazard rate dependence functions and also apply them to the case of linear transformation models. Note that the linear transformation models include Cox's (1972) proportional hazard rate models and Pettitt's (1984) proportional odds model as special cases.

2 Average collapsibility of association measures

Let (Y, X, W) be a random vector, where our interest is mainly on the association between Y and X, and W is treated as a covariate. We assume for simplicity that X and W are continuous, unless stated otherwise. Note that Y has a monotone (increasing) regression function of X if $E(Y|X = x)$ is increasing in x or equivalently the expectation dependence function (EDF) $\partial E(Y \mid X = x)/\partial x \geq 0$. We first discuss the average collapsibility results for the EDF and introduce the following definition.

Definition 3 The conditional expectation dependence function $\partial E(Y|x, w)/\partial x$ is average collapsible over W if

$$E_{W|x}\left(\frac{\partial}{\partial x} E(Y|x, W)\right) = \frac{\partial}{\partial x} E(Y|x), \quad \text{for all } x. \tag{1}$$

The following result gives sufficient conditions for the average collapsibility of EDF. In the sequel, $Y \perp\!\!\!\perp X$ and $Y \perp\!\!\!\perp X|W$, respectively, denote the independence of X and Y, and the conditional independence of Y and X given W. We assume henceforth all the partial derivatives exist and are continuous so that the differentiation and integration can be interchanged. In addition, for simplicity, we will henceforth denote $E(Y|X = x, W = w)$ by $E(Y|x, w)$ and similar notations for conditional distributions/densities.

Theorem 1 *The conditional EDF measure $\partial E(Y|x, w)/\partial x$ is average collapsible over W if either*

(i) *$E(Y|x, w)$ is independent of w, or*
(ii) *$X \perp\!\!\!\perp W$*
holds.

Remark 1 (i) The condition that $E(Y|x, w)$ is independent of w implies the homogeneity of EDF and in this case both uniform collapsibility [Part (a) of Theorem 3.4 of Xie et al. (2008)] and average collapsibility hold. However, when the EDF is not homogeneous over w, average collapsibility may still hold if (and only if) $X \perp\!\!\!\perp W$, while the simple collapsibility and the uniform collapsibility do not hold as they require the homogeneity condition.

(ii) Observe also that the condition $E(Y|x, w)$ is independent of w is a weaker condition than $Y \perp\!\!\!\perp W|X$, usually required for other notions of collapsibility. For example, when $W > 0$, and $(Y|x, w) \sim U(x - w, x + w)$, we have $E(Y|x, w) = x$ for all w. But, the conditional density of $(Y|x, w)$ is

$$f(y|x, w) = \frac{1}{2w}, \quad x - w < y < x + w, \tag{2}$$

showing that Y and W are not conditionally independent given X.

First, we discuss an application of Theorem 1.

Example 1 Let X_1, X_2 and X_3 be iid $N(0, 1)$ variables. Define now

$$W = X_1; \quad X = X_1 X_2; \quad Y = X_1 X_2 + X_1 X_3,$$

so that $X = W X_2$ and $Y = X + W X_3$. Then it can be seen that

$$(X|w) \sim N(0, w^2); \quad (Y|x, w) \sim N(x, w^2).$$

Hence, $E(Y|x, w) = x$ which is independent of w. By Theorem 1, the average collapsibility of *EDF* $\partial E(Y|x, w)/\partial x$ holds. Note here that Y and W are not conditionally independent given X.

We next show that condition (i) or (ii) is only sufficient, but not necessary. Hereafter, $\phi(z)$ and $\Phi(z)$ denote, respectively, the density and the distribution function of $Z \sim N(0, 1)$.

Example 2 Suppose $(Y|x, w)$ follows uniform $U(0, (x^2 + (w - x)^2))$ so that

$$F(y|x, w) = y(x^2 + (w - x)^2)^{-1}, \quad 0 < y < (x^2 + (w - x)^2) \tag{3}$$

and $E(Y|x, w) = \frac{1}{2}(x^2 + (w - x)^2)$. Assume also $(W|X = x) \sim N(x, 1)$ so that

$$\frac{\partial}{\partial x} f(w|x) = -\phi'(w - x) = (w - x)\phi(w - x). \tag{4}$$

Then

$$\begin{aligned}
&\int E(Y|x,w)\frac{\partial}{\partial x}f(w|x)dw\\
&=\frac{1}{2}\int_{-\infty}^{\infty}(x^2+(w-x)^2)(w-x)\phi(w-x)dw\\
&=\frac{1}{2}\left[x^2\int_{-\infty}^{\infty}(w-x)\phi(w-x)dw+\int_{-\infty}^{\infty}(w-x)^3\phi(w-x)dw\right]\\
&=\frac{1}{2}\left[x^2\int_{-\infty}^{\infty}t\phi(t)dt+\int_{-\infty}^{\infty}t^3\phi(t)dt\right]\\
&=0,\quad \text{for all } x. \qquad (5)
\end{aligned}$$

Thus, using (34) given in the Appendix, the average collapsibility of EDF over W holds, but neither condition (i) nor condition (ii) is satisfied.

As another example, one can take $X \sim N(0,1)$, $(W|x) \sim N(x,1)$ and $(Y|x,w) \sim N(x^2+(w-x)^2, 1)$. Then again (34) is satisfied and hence the average collapsibility of EDF holds.

An implication of Theorem 1 to linear regression models follows.

Linear regression Consider the following conditional and marginal linear regression models, respectively (see Xie et al. (2008)):

$$E(Y|X=x,W=w)=\begin{cases}\alpha(w)+\beta(w)x, & \text{if } W \text{ is discrete}\\ \alpha+\beta x+\gamma w, & \text{if } W \text{ is continuous}\end{cases}$$

and

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

Then

$$\frac{\partial}{\partial x}E(Y \mid X=x,W=w)=\begin{cases}\beta(w), & \text{if } W \text{ is discrete}\\ \beta, & \text{if } W \text{ is continuous}\end{cases}$$

and

$$\frac{\partial}{\partial x}E(Y \mid x)=\tilde{\beta}.$$

We say that the regression coefficient $\beta(w)$ (or β) is simple collapsible if $\beta(w) = \tilde{\beta}$ for all w (or $\beta = \tilde{\beta}$). In addition, it is said to be average collapsible if

$$E_{W|x}(\beta(W))=\tilde{\beta}\ \ (\text{or } E_{W|x}(\beta)=\tilde{\beta}),\quad \text{for all } x. \qquad (6)$$

Thus, the average collapsibility of EDF reduces to the average collapsibility of regression coefficients, in the case of linear regression models.

Remark 2 Vellaisamy and Vijay (2007) defined the average collapsibility of regression coefficients $\beta(w)$ as $E_W(\beta(W)) = \tilde{\beta}$ and discussed the conditions under which it holds. However, the definition of average collapsibility given in (6) is more natural as it involves the joint distribution of W and X. Note also that $E_{W|x}(\beta(W)) = \tilde{\beta}$ for all x implies $E_W(\beta(W)) = \tilde{\beta}$, but not necessarily conversely.

Next, we look at the average collapsibility of mixed derivative of interaction (MDI). Since

$$\frac{\partial^2}{\partial x \partial y} \log f(x, y) = \frac{\partial^2}{\partial x \partial y} \log f(y|x), \quad \text{for all } x \text{ and } y, \tag{7}$$

it follows from Proposition 3.2.1 of Whittaker (1990) that

$$\frac{\partial^2}{\partial x \partial y} \log f(y|x) = 0 \quad \text{for all } x \text{ and } y \iff Y \perp\!\!\!\perp X.$$

In view of (7), the MDI henceforth stands for $\partial^2 \log f(y|x)/\partial x \partial y$, which motivates the following definition of average collapsibility.

Definition 4 The conditional MDI measure $\partial^2 \log f(y|x, w)/\partial x \partial y$ is said to be average collapsible over W if

$$E_{W|x}\left(\frac{\partial^2}{\partial x \partial y} \log f(y|x, W)\right) = \frac{\partial^2}{\partial x \partial y} \log f(y|x), \quad \text{for all } (y, x).$$

It is assumed that $\log f(y|x)$ has continuous partial derivatives so that

$$\frac{\partial^2}{\partial x \partial y} \log f(y|x) = \frac{\partial^2}{\partial y \partial x} \log f(y|x) \quad \text{for all } (y, x).$$

The following result provides a set of sufficient conditions for the average collapsibility of MDI.

Theorem 2 *The MDI measure is average collapsible over W if either*

(i) $Y \perp\!\!\!\perp W|X$, *or*
(ii) $X \perp\!\!\!\perp W|Y$
holds.

Xie et al. (2008) showed that condition (i) or (ii) in Theorem 2 is necessary and sufficient for uniform collapsibility. The following counterexample shows that they are only sufficient, but not necessary, for the average collapsibility.

Example 3 Let $X > 0$ and $(W|X = x) \sim N(x, 1)$. Assume that

$$f(y|x, w) = xy^{x-1}(x^2 + (w - x)^2), \;\; 0 < y < (x^2 + (w - x)^2)^{-1/x}, \tag{8}$$

which can easily be seen to be a valid density.

Then

$$\frac{\partial^2}{\partial x \partial y} \log f(y|x, w) = \frac{1}{y} = E_{W|x} \left(\frac{\partial^2}{\partial x \partial y} \log f(y|x, W) \right). \tag{9}$$

Since $(W|X = x) \sim N(x, 1)$, it follows that the marginal density of $(Y|X = x)$ is

$$\begin{aligned} f(y|x) &= \int_{-\infty}^{\infty} f(y|x, w) f(w|x) dw \\ &= xy^{x-1} \left[\int_{-\infty}^{\infty} x^2 \phi(w - x) dw + \int_{-\infty}^{\infty} (w - x)^2 \phi(w - x) dw \right] \\ &= xy^{x-1}(x^2 + 1), \end{aligned}$$

which is also a valid density on $0 < y < (x^2 + 1)^{-1/x}$.

In addition, it follows from (9)

$$\frac{\partial^2}{\partial x \partial y} \log f(y|x) = \frac{1}{y} = E_{W|x} \left(\frac{\partial^2}{\partial x \partial y} \log f(y|x, W) \right).$$

Thus, average collapsibility of MDI holds, although the condition (i) is not satisfied. Note, however, the condition (ii) is satisfied.

It was rather challenging to construct the counterexample 3, as it requires, in addition to the other conditions, the interchange of log and integration, which holds very rarely. Observe also that in Example 3,

$$\frac{\partial}{\partial y} \log f(y|x, w) = \frac{\partial}{\partial y} \log f(y|x), \quad \text{for all } (y, x),$$

which leads to the average collapsibility. This observation leads to the following result which generalizes Theorem 2 whose proof is immediate.

Theorem 3 *The MDI measure is average collapsible over W if either*

$$\begin{aligned} &\text{(i)} \ \frac{\partial}{\partial y} \log f(y|x, w) = \frac{\partial}{\partial y} \log f(y|x), \quad \textit{for all } (y, x), \quad \textit{or} \\ &\text{(ii)} \ \frac{\partial}{\partial x} \log f(y|x, w) = \frac{\partial}{\partial x} \log f(y|x), \quad \textit{for all } (y, x) \end{aligned}$$

holds.

Remark 3 (i) Observe that the sufficient conditions given in Theorem 3 are weaker than the ones given in Theorem 2.

(ii) It will be of practical interest to develop certain statistical tests to verify the conditions (i) and (ii). Indeed, there is not much work devoted to the testing aspects of collapsibility in general, due to the complexities involved.

Some additional examples for Theorem 3. Let $f(y|x, w)$ be as in Example 3 and consider, for $\lambda > 0$, the tempered normal density

$$t_\lambda(w|x) = c_\lambda(x)e^{-\lambda w}\phi(w - x), \quad \text{for } x > 0, w \in \mathbb{R},$$

where

$$c_\lambda(x) = \left(\int_{-\infty}^{\infty} e^{-\lambda w}\phi(w - x)dw\right)^{-1} = e^{(x^2-(x-\lambda)^2)/2},$$

that is, $t_\lambda(w|x) = \phi(w - x + \lambda)$. Then the corresponding marginal density of $(Y|X = x)$ is

$$\begin{aligned} f_\lambda(y|x) &= xy^{x-1}\left[\int_{\infty}^{\infty} x^2\phi(w - x + \lambda)dw + \int_{\infty}^{\infty}(w - x)^2\phi(w - x + \lambda)dw\right] \\ &= xy^{x-1}(x^2 + \lambda^2 + 1), \end{aligned}$$

which is also a valid density on $0 < y < (x^2 + \lambda^2 + 1)^{-1/x}$. Thus, the average collapsibility of MDI holds for the family $\{t_\lambda(w|x)\}$, $\lambda > 0$, also.

Next, we discuss the connection of Theorem 2 to logistic regression models.

Logistic regression Let Y be a binary variable and consider the following conditional and marginal logistic regression models (Vellaisamy and Vijay 2007; Xie et al. 2008) considered in the literature:

$$\log\left(\frac{f(1|x, w)}{f(0|x, w)}\right) = \begin{cases} \alpha(w) + \beta(w)x, & \text{if } W \text{ is discrete} \\ \alpha + \beta x + \gamma w, & \text{if } W \text{ is continuous} \end{cases}$$

and

$$\log\left(\frac{f(1|x)}{f(0|x)}\right) = \tilde{\alpha} + \tilde{\beta}x.$$

We say the logistic regression coefficient is simple collapsible if

$$\tilde{\beta} = \begin{cases} \beta(w) \text{ for all } w, & \text{if } W \text{ is discrete} \\ \beta, & \text{if } W \text{ is continuous.} \end{cases}$$

In addition, we say $\beta(w)$ or β is said to be average collapsible if $E_{W|x}(\beta(W)) = \tilde{\beta}$, when W is discrete and $E_{W|x}(\beta) = \tilde{\beta}$, when W is continuous.

Since Y is binary, the partial derivative is replaced by the difference between the adjacent levels of Y [see Cox (2003)] so that

$$
\begin{aligned}
\frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}\log f(y|x,w)\right) &= \frac{\partial}{\partial x}(\log f(1|x,w)-\log f(0|x,w))\\
&= \frac{\partial}{\partial x}\log\left(\frac{f(1|x,w)}{f(0|x,w)}\right)\\
&= \begin{cases} \frac{\partial}{\partial x}(\alpha(w)+\beta(w)x)=\beta(w), & \text{if } W \text{ is discrete}\\ \frac{\partial}{\partial x}(\alpha+\beta x+\gamma w)=\beta, & \text{if } W \text{ is continuous,}\end{cases}
\end{aligned}
$$

the logistic regression coefficients corresponding to both the cases of W. From Theorem 2, we now conclude that $\beta(w)$ or β is average collapsible if $(i) Y \perp\!\!\!\perp W|X$ or $(ii) X \perp\!\!\!\perp W|Y$ holds.

Finally, we discuss a new measure called log-expectation dependence (LED) between $Y > 0$ and X, defined by $\partial \log E(Y|x,w)/\partial x$, where it is assumed that $0 < E(Y|x,w) < \infty$, for all (x,w). First note that for all x,

$$
\begin{aligned}
\frac{\partial}{\partial x}\log E(Y|x)=0 &\Longleftrightarrow \frac{\partial}{\partial x}E(Y|x)=0\\
&\Longleftrightarrow \int y\frac{\partial}{\partial x}\,(dF(y|x))=0\\
&\Longleftrightarrow dF(y|x)=dF(y|x^*)\ \text{ for } \ \text{all } y,x \ \text{ and } x^*\\
&\Longleftrightarrow Y \perp\!\!\!\perp X.
\end{aligned}
$$

In addition, by Theorem 1 of Xie et al. (2008),

$$
\frac{\partial}{\partial x}\log E(Y|x)\geq 0 \Longrightarrow \frac{\partial}{\partial x}E(Y|x)\geq 0 \Longrightarrow \rho(Y,X)\geq 0,
$$

where $\rho(Y,X)$ is the correlation coefficient between Y and X.

Assume now $E(Y|x)\neq 0$ for all x. Since,

$$
\frac{\partial}{\partial x}\log E(Y|x)=\frac{\frac{\partial}{\partial x}E(Y|x)}{E(Y|x)},
$$

we have LED is nonnegative if and only if $E(Y|x)$ is positive (negative) and ED is nonnegative (nonpositive) for all x. In other words, LED is nonnegative if and only if $E(Y|x)$ is positive (negative) and nondecreasing (nonincreasing).

Next, we discuss the collapsibility issues for the LED measure and hence the following definition.

Definition 5 The conditional LED measure $\partial\log E(Y|x,w)/\partial x$ is simple collapsible over W if

$$
\frac{\partial}{\partial x}\log E(Y|x,w)=\frac{\partial}{\partial x}\log E(Y|x),\quad \text{for all } x \text{ and } w \tag{10}
$$

and average collapsible over W if

$$E_{W|x}\left(\frac{\partial}{\partial x}\log E(Y|x,W)\right)=\frac{\partial}{\partial x}\log E(Y|x),\quad \text{for all } x. \tag{11}$$

Theorem 4 *The LED measure is simple collapsible and hence average collapsible over W if $E(Y|x,w)$ is homogeneous over w.*

Note that the condition of homogeneity of $E(Y|x,w)$ is a weaker condition than $Y \perp\!\!\!\perp W|X$, a condition usually required for other measures. In addition, it seems difficult to obtain other weaker sufficient conditions than the ones given in Theorem 4.

The LED measure has relevance to Poisson and negative binomial (NB) regression models, seen as follows.

Poisson regression. Consider the Poisson regression model defined by

$$(Y|X=x,W=w)\sim Poi(\lambda(x,w)),$$

where the mean

$$E(Y|x,w)=\lambda(x,w)=\begin{cases} e^{\alpha(w)+\beta(w)x}, & \text{if } W \text{ is discrete}\\ e^{\alpha+\beta x+\gamma w}, & \text{if } W \text{ is continuous.}\end{cases}$$

Then

$$\frac{\partial}{\partial x}(\log E(Y|x,w))=\begin{cases}\beta(w), & \text{if } W \text{ is discrete}\\ \beta, & \text{if } W \text{ is continuous.}\end{cases}$$

Let $(Y|x)\sim Poi(e^{\tilde{\alpha}+\tilde{\beta}x})$, the marginal Poisson regression model. Then

$$\log E(Y|x)=\tilde{\alpha}+\tilde{\beta}x;\quad \frac{\partial}{\partial x}\log E(Y|x)=\tilde{\beta}.$$

Hence, by Theorem 4, the average collapsibility of Poisson regression coefficient $\beta(w)$ (or β) holds, that is,

$$E_{W|x}(\beta(W))=\tilde{\beta}(or\, E_{W|x}(\beta)=\tilde{\beta})$$

when $\lambda(x,w)$ does not depend on w. The latter condition holds when for example $\gamma=0$ which does not in general mean that $Y \perp\!\!\!\perp W|X$.

The following interesting example shows that average collapsibility may hold, even when $E(Y|x,w)$ depends on w.

Example 4 Let $X>0$ and $(Y|X=x,W=w)\sim P(\lambda(x)w)$, where $\lambda(x)=exp(\alpha+\beta x)$. Then $E(Y|x,w)=\lambda(x)w$ and

$$\frac{\partial}{\partial x}\log E(Y|x,w)=\beta=E_{W|x}\left(\frac{\partial}{\partial x}\log E(Y|x,W)\right). \tag{12}$$

Let now $(W|X=x) \sim G(x,x)$, the gamma distribution with mean unity. Then it is known that

$$(Y|x) \sim NB\left(x, \frac{x}{x+\lambda(x)}\right),$$

the negative binomial (NB) distribution with

$$P(Y=y|x) = \frac{\Gamma(y+x)}{y!\,\Gamma(x)}\left(\frac{x}{x+\lambda(x)}\right)^{x}\left(\frac{\lambda(x)}{x+\lambda(x)}\right)^{y}, \quad y=0,1,\cdots.$$

When $x=r$, an integer, Y denotes the number of failures before rth success in a sequence of independent Bernoulli trials. Hence,

$$E(Y|x) = \lambda(x); \quad \frac{\partial}{\partial x}\log E(Y|x) = \beta. \tag{13}$$

Thus, from (12) and (13), the average collapsibility holds. Note here the covariates W and X are not independent.

Negative binomial regression Suppose in Example 4 we assume in addition that the unobservable W is independent of X and $W \sim G(\theta,\theta)$. Then again

$$(Y|X=x) \sim NB\left(\theta, \frac{\theta}{\theta+\lambda(x)}\right); \quad E(Y|x) = \lambda(x). \tag{14}$$

The model (14) is the usual NB regression model. Thus, the average collapsibility of the LED function corresponds to that of the NB regression coefficient β. It is interesting to note that when the unobserved covariate W follows the gamma distribution with mean unity, the average collapsibility of the NB regression coefficient holds, even when W and X are not independent (Example 4).
Note, however, in the negative binomial regression,

$$Var(Y|x) = \lambda(x)\left(1+\frac{\lambda(x)}{\theta}\right) > \lambda(x) = E(Y|x), \tag{15}$$

unlike the Poisson regression case. Thus, whenever the data exhibit over dispersion (variance exceeds mean), the negative binomial regression model is more appropriate than the Poisson regression model for the analysis of data.

3 The multivariate case

In this section, we consider an extension to the multivariate case. The case of multivariate response Y may be considered by treating one component at a time (Cox and Wermuth 2003; Xie et al. 2008) and similarly the covariate X may also be considered one component at a time, while keeping other components fixed. Therefore, we consider here only the case of multivariate random vector $W=(W_1,\ldots,W_p)$.

A conditional measure of association, say, $\frac{\partial}{\partial x}(E(Y|x, w)$ is simple collapsible over W if

$$\frac{\partial}{\partial x}(E(Y|x, w)) = \frac{\partial}{\partial x}(E(Y|x)), \quad \text{for all } x \text{ and } w = (w_1, \ldots, w_p).$$

and average collapsible over W if

$$E_{W|x}\left(\frac{\partial}{\partial x}(E(Y|x, W))\right) = \frac{\partial}{\partial x}(E(Y|x)), \quad \text{for all } x.$$

The definition of average collapsibility of other measures of association remains the same, except that W is now a p-variate random vector.

Let $W = (W_1, W_2)$, where W_1 has q components and W_2 has $(p-q)$ components. We now have the following result for the EDF and MDI and the corresponding results for LED follow easily when $E(Y|x, w)$ is homogeneous over w.

Theorem 5 *The following results hold:*

(a) *The EDF is average collapsible over W if $(i) Y \perp\!\!\!\perp W_1|(X, W_2)$ and $(ii) X \perp\!\!\!\perp W_2$ hold.*
(b) *The MDI is average collapsible over W if $(i) Y \perp\!\!\!\perp W_1|(X, W_2)$ and $(ii) X \perp\!\!\!\perp W_2|Y$ hold.*

Remark 4 (i) By symmetry, the average collapsibility of MDI holds when X and Y are interchanged in conditions (i) and (ii) of Part (b) of Theorem 5. In addition, in view of Theorem 2, the condition (ii) of part (b) can be replaced by $Y \perp\!\!\!\perp W_2|X$.

(ii) Xie et al. (2008) established the uniform collapsibility of DDF and EDF under an additional condition of homogeneity of these measures. Thus, average collapsibility holds under less restrictive conditions and hence is applicable to a larger class of conditional distributions that may arise in practical applications.

Remark 5 Wang et al. (2009) discussed the concept of uniform non-confounding for a multivariate covariate and obtained sufficient conditions for it to hold. They discussed also an algorithm which extends Greenland et al. (1999) approach to the multiple subset of potential confounders. Starting from the sufficient set, say Z, which contains multiple potential confounders, their algorithm repeatedly deletes non-confounders from Z as much as possible. In view of the above-mentioned points, it would be of interest to consider a more general case and derive an iterative procedure for checking the average collapsibility over a multivariate W, similar to the works of Wang et al. (2009).

4 Collapsibility results for survival models

In this section, we discuss the collapsibility of some dependence measures that arise in the context of survival analysis. Let T denote the lifetime and X and Y be the covariates of interest that are associated with T. Suppose our interest is to study the effect of the covariate X on T, which arises in several situations. As mentioned in Di Serio et al. (2009), increasing the value of X may have positive effect on T, for each subgroup corresponding to another covariate Y, but may have a negative effect where there is no conditioning on Y. This phenomenon, known as Simpson's paradox (Simpson 1951) has been recently investigated by Di Serio et al. (2009) in the context of survival analysis, with respect to linear transformation models. Let $P(T > t + s|T > t, x, y)$ and $h(t|x, y) = \lim_{h\to 0} P(t < T < t + h|T > t, x, y)$ be the conditional (on Y) survival probability and the hazard rate functions. Similarly, let $P(T > t+s|T > t, x)$ and $h(t|x)$ be the corresponding marginal characteristics based on covariate X alone.

Di Serio et al. (2009) discussed the conditions under which $P(T > t + s|T > t, x, y)$ is decreasing in x for all y, but $P(T > t + s|T > t, x)$ is increasing in x. They discussed also the conditions under which $h(t|x, y)$ is increasing in x and for all y, while $h(t|x)$ is decreasing in x. That is, they discussed the occurrence of Simpson's paradox when the lifetime T and the covariates X and Y follow the linear transformation (LT) model defined by

$$K(T) = -\beta_{tx}X - \beta_{ty}Y + W, \tag{16}$$

where $W \perp\!\!\!\perp (X, Y)$, with special emphasis on Cox's regression model. Note that the LT model is more general and includes Cox's (1972) proportional hazards model, and Pettit's (1983) proportional odds model, as special cases. It is of interest to know the conditions under which the survival probability and hazard rate functions based on a single covariate X are reasonable especially in the context of another covariate Y. This study is in a sense complementary to the investigations by Di Serio et al. (2009) who discussed the conditions under which Simpson's paradox occurs. In situations where the covariate Y is not observed, average collapsibility characterizes the conditions under which the inference based on X alone will on the average be the same as the one based on both X and Y.

We first formulate the appropriate definitions.

Definition 6 We call $\partial P(T > t + s|T > t, x)/\partial x$ and $\partial h(t|x)/\partial x$, respectively, the survival dependence function and the hazard rate dependence function. Similarly, $\partial P(T > t + s|T > t, x, y)/\partial x$ and $\partial h(t|x, y)/\partial x$ are called the corresponding conditional (on Y) dependence functions.

Definition 7 The conditional survival dependence function $\frac{\partial}{\partial x}P(T > t + s|T > t, x, y)$ is average collapsible over covariate (Y) if

$$\frac{\partial}{\partial x}P(T > t + s|T > t, x) = E_{\{Y|T>t,x\}}\left(\frac{\partial}{\partial x}P(T > t + s|T > t, x, Y)\right), \quad \text{for all } (t, s, x) \tag{17}$$

and the hazard rate dependence function $\partial h(t|x, y)/\partial x$ is average collapsible over Y if

$$\frac{\partial}{\partial x}h(t|x, y) = E_{\{Y|T>t,x\}}\left(\frac{\partial}{\partial x}h(t|x, Y)\right), \quad \text{for all } (t, x). \tag{18}$$

First note that, assuming Y is also continuous for simplicity,

$$P(T > t+s|T > t, x) = \int P(T > t+s|T > t, x, y) f(y|T > t, x) dy.$$

Partial differentiation with respect to x gives us

$$\begin{aligned} \frac{\partial}{\partial x} P(T > t+s|T > t, x) &= \int \Big(\frac{\partial}{\partial x} P(T > t+s|T > t, x, y) \Big) f(y|T > t, x) dy \\ &\quad + \int P(T > t+s|T > t, x, y)) \frac{\partial}{\partial x} f(y|T > t, x) dy \\ &= E_{\{Y|T>t,x\}} \Big(\frac{\partial}{\partial x} P(T > t+s|T > t, x, Y) \Big) \\ &\quad + \int P(T > t+s|T > t, x, y)) \frac{\partial}{\partial x} f(y|T > t, x) dy. \end{aligned} \tag{19}$$

Thus, average collapsibility of survival dependence function holds if

$$\int P(T > t+s|T > t, x, y)) \frac{\partial}{\partial x} f(y|T > t, x) dy \equiv 0, \tag{20}$$

for all (t, s, x).

Remark 6 Suppose now the average collapsibility holds so that from (19) and (20), we get

$$\frac{\partial}{\partial x} P(T > t+s|T > t, x) = \int \left\{ \frac{\partial}{\partial x} P(T > t+s|T > t, x, y) \right\} f(y|T > t, x) dy$$
$$\text{for all } (t, s, x).$$

Then

$$\frac{\partial}{\partial x} P(T > t+s|T > t, x, y) \leq 0 \quad \text{for all } y \text{ and } (t, s, x)$$

implies

$$\frac{\partial}{\partial x} P(T > t+s|T < t, x) \leq 0 \;\text{ for all } (t, s, x),$$

showing that the characteristic reversal or Simpson's paradox does not occur.

Consider next the hazard rate function. Note that

$$
\begin{aligned}
h(t|x) &= \lim_{s\to 0} \int P(t < T < t+s|T > t, x, y) f(y|T > t, x) dy \\
&= \int \left(\lim_{s\to 0} P(t < T < t+s|T > t, x, y) \right) f(y|T > t, x) dy \\
&= \int h(t|x, y) f(y|T > t, x) dy. \qquad (21)
\end{aligned}
$$

Therefore, as in the case of conditional survival probability,

$$
\begin{aligned}
\frac{\partial}{\partial x} h(t|x) = E_{\{Y|T>t,x\}} \Big(\frac{\partial}{\partial x} h(t|x, Y) \Big) \\
+ \int h(t|x, y) \frac{\partial}{\partial x} f(y|T > t, x) dy.
\end{aligned}
$$

Thus, the average collapsibility (over Y) of hazard rate function holds if

$$
\int h(t|x, y) \frac{\partial}{\partial x} f(y|T > t, x) dy \equiv 0 \quad \text{for all } (t, x). \qquad (22)
$$

The following result provides general sufficient conditions for the average collapsibility.

Theorem 6 *The conditional survival dependence function $\partial P(t > t+s|T > t, x, y)/\partial x$ is average collapsible over Y if*

$$
\begin{aligned}
&(i) \ \frac{P(T > t+s|x, y)}{P(T > t|x, y)} \ \textit{is homogeneous over } y, \textit{ or} \\
&(ii) \ Y \perp\!\!\!\perp X|T.
\end{aligned}
$$

Observe also that the condition $(T \perp\!\!\!\perp Y|X)$ implies (i), but not conversely.

The next result for the hazard rate dependence function can be proved in a similar manner and hence the proof is omitted.

Theorem 7 *The conditional hazard rate dependence function $\partial h(t|x, y)/\partial x$ is average collapsible over Y if*

(i) $h(t|x, y)$ is homogeneous over y or (ii) $Y \perp\!\!\!\perp X|T$ holds.

Application to linear transformation models

Let the failure time T and the covariates X and Y follow the well-known linear transformation (LT) model [see Di Serio et al. (2009)] defined by

$$
K(T) = \beta_{tx} X - \beta_{ty} Y + W, \qquad (23)
$$

where $W \perp\!\!\!\perp (X, Y)$. Di Serio et al. (2009) discussed the conditions under which Simpson's paradox occurs for the conditional survival probability and the conditional hazard rate, where (T, Y, X) follow (23). They showed the model is more general and reduces to Cox's (1972) proportional hazard rate models when $W \sim F_W(t) = 1 - \exp(-e^t)$ (Gumbel-type) and Pettitt's (1984) proportional odds model when $W \sim F_W(t) = e^t/(1 + e^t)$ (logistic), where $t \in \mathbb{R}$.

First, we discuss an extension of Cochran (1938) result to the LT model. Observe first that LT model can also be viewed as a regression model with

$$E(K(T)|X, Y) = \mu_W - \beta_{tx}X - \beta_{ty}Y, \tag{24}$$

where $\mu_W = E(W)$ is the mean of W. It is of interest first to know the conditions under which the marginal model $E(K(T)|X)$ is also an LT model. Since,

$$\begin{aligned} E(K(T)|X) &= E_{Y|X}(\mu_W - \beta_{tx}X - \beta_{ty}Y) \\ &= \mu_W - \beta_{tx}X - \beta_{ty}E(Y|X), \end{aligned} \tag{25}$$

we see that $E(K(T)|X)$ is linear if and only if $E(Y|X)$ is also linear. Suppose

$$\tilde{K}(T) = \tilde{\alpha}_{tx} - \tilde{\beta}_{tx}X + \tilde{W},$$

where $\tilde{W} \perp\!\!\!\perp X$, is the marginal LT model. Then

$$E(\tilde{K}(T)|X) = \mu_{\tilde{W}} + \tilde{\alpha}_{tx} - \tilde{\beta}_{tx}X \tag{26}$$

is the marginal LT regression model. Assume now

$$E(Y|X) = \alpha_{yx} + \beta_{yx}X, \tag{27}$$

the linear regression of Y on X. Substitution of (27) into (25) leads to

$$E(\tilde{K}(T)|X) = (\mu_W - \alpha_{yx}\beta_{t_y}) - (\beta_{tx} + \beta_{ty}\beta_{yx})X. \tag{28}$$

A comparison of the coefficients in (26) and (28) yields

$$\tilde{\alpha}_{tx} = \mu_W - \mu_{\tilde{W}} - \beta_{ty}\alpha_{yx} \tag{29}$$

and

$$\tilde{\beta}_{tx} = \beta_{tx} + \beta_{ty}\beta_{yx}, \tag{30}$$

which give us the relationships between the coefficients of the conditional and the marginal LT models. Observe that the result (30) is similar to Cochran (1938) result for linear regression coefficients.

Definition 8 The regression coefficient β_{tx} is said to be simple collapsible over Y if $\tilde{\beta}_{tx} = \beta_{tx}$.

Note from (30) the simple collapsibility of β_{tx} holds if $\beta_{yx} = 0$. Since $\beta_{yx} = Cov(Y, X)/V(X)$, it is clear that $\beta_{yx} = 0$ holds if $Cov(Y, X) = 0$, that is, when X and Y are uncorrelated.
Finally, an application of Theorem 6 follows.

Example 5 Let (T, X, Y) follow LT model defined by

$$K(T) = -\beta_{tx}X - \beta_{ty}Y + W, \tag{31}$$

where $W \perp\!\!\!\perp (X, Y)$. In this case, we have,

$$\begin{aligned} P(T > t + s|T > t, x, y) &= \frac{P(T > t + s|x, y)}{P(T > t|x, y)} \\ &= \frac{\overline{F}_W(K(t + s) + \beta_{tx}x + \beta_{ty}y)}{\overline{F}_W(K(t) + \beta_{tx}x + \beta_{ty}y)}, \end{aligned} \tag{32}$$

which follows from equation (7) of Di Serio et al. (2009).

Suppose now $W \sim E(\lambda)$, the exponential distribution with mean λ^{-1}. Then

$$\begin{aligned} \frac{P(T > t + s|x, y)}{P(T > t|x, y)} &= \frac{e^{-\lambda(K(t+s) + \beta_{tx}x + \beta_{ty}y)}}{e^{-\lambda(K(t) + \beta_{tx}x + \beta_{ty}y)}} \\ &= e^{-\lambda(K(t+s) - K(t))}, \end{aligned}$$

which does not depend on y. Hence, by Part (i) of Theorem 1, the average collapsibility of conditional survival dependence function $\partial P(T > t + s|T > t, x, y)/\partial x$ holds.

Note also in the above example,

$$P(T > t|x, y) = e^{-\lambda(K(t) + \beta_{tx}x + \beta_{ty}y)},$$

showing that T and Y are not conditionally independent.

In addition, for the model (31) and $f_W(w) = \lambda e^{-\lambda w}$, for $w > 0$, we have

$$\begin{aligned} h(t|x, y) &= \frac{f_W(K(t) + \beta_{tx}x + \beta_{ty}y)}{\overline{F}_W(K(t) + \beta_{tx}x + \beta_{ty}y)} K'(t) \\ &= \lambda K'(t), \end{aligned}$$

which is homogeneous over y. Using Part (i) of Theorem 7, the average collapsibility of conditional hazard rate dependence $\partial h(t|x, y)/\partial x$ also holds.

Acknowledgements A part of this work was done while the author was visiting the Department of Statistics and Probability, Michigan State University, during Summer 2015. The author is grateful to both the referees for several helpful comments and encouragements. This research is partially supported by a DST research Grant No. SR/MS:706/10.

APPENDIX: Proofs

Proof of Theorem 1 Note that

$$\begin{aligned}\frac{\partial}{\partial x}E(Y|x) &= \frac{\partial}{\partial x}\int E(Y|x,w)f(w|x)dw \\ &= \int\left[\frac{\partial}{\partial x}E(Y|x,w)\right]f(w|x)dw + \int E(Y|x,w)\frac{\partial}{\partial x}f(w|x)dw \\ &= E_{W|x}\left(\frac{\partial}{\partial x}E(Y|x,W)\right) + \int E(Y|x,w)\frac{\partial}{\partial x}f(w|x)dw. \qquad (33)\end{aligned}$$

Hence, average collapsibility holds if and only if

$$\int E(Y|x,w)\frac{\partial}{\partial x}f(w|x)dw = 0, \quad \text{for all } x. \qquad (34)$$

Assume now condition (i) holds so that

$$E(Y|x,w) = h(x), \quad \text{for all } x \text{ and } w, (say).$$

Then

$$\begin{aligned}\int E(Y|x,w)\frac{\partial}{\partial x}f(w|x)dw &= h(x)\int \frac{\partial}{\partial x}f(w|x)dw \\ &= h(x)\frac{\partial}{\partial x}\int f(w|x)dw \\ &= 0, \quad \text{for all } x.\end{aligned}$$

The average collapsibility follows from (34).
Assume next condition (ii) holds. Then obviously,

$$\int E(Y|x,w)\frac{\partial}{\partial x}f(w|x)dw = 0, \quad \text{for all } x,$$

and so average collapsibility holds again. □

Proof of Theorem 2 Since

$$\frac{\partial^2}{\partial x\partial y}\log f(y|x) = \frac{\partial}{\partial x}\left(\frac{\frac{\partial}{\partial y}f(y|x)}{f(y|x)}\right), \qquad (35)$$

average collapsibility of MDI holds if and only if

$$E_{W|x}\left(\frac{\partial}{\partial x}\left(\frac{\frac{\partial}{\partial y}f(y|x,W)}{f(y|x,W)}\right)\right) = \frac{\partial}{\partial x}\left(\frac{\frac{\partial}{\partial y}f(y|x)}{f(y|x)}\right) \quad \text{for all } (y,x). \qquad (36)$$

Note that condition (i) implies

$$f(y|x,w)=f(y|x)\Longrightarrow\frac{\partial}{\partial y}f(y|x,w)=\frac{\partial}{\partial y}f(y|x),\quad\text{for all }(y,x,w).$$

Thus, Eq. (36) holds.

Observe also that

$$\frac{\partial^2}{\partial x\partial y}\log f(x,y)=\frac{\partial}{\partial y}\left(\frac{\frac{\partial}{\partial x}f(x|y)}{f(x|y)}\right),$$

which is the same as equation (35) with x and y interchanged. Thus, the condition (ii) also implies the average collapsibility of *MDI*. □

Proof of Theorem 4 Let

$$E(Y|x,w)=h_1(x)\quad\text{for all }x\text{ and }w. \tag{37}$$

Then

$$E(Y|x)=E_{W|x}(E(Y|x,W))=E_{W|x}(h_1(x))=h_1(x). \tag{38}$$

Thus, from (37) and (38),

$$E(Y|x)=E(Y|x,w),\text{ for all }x\text{ and }w,$$

and hence simple collapsibility holds.

In addition, since

$$\frac{\partial}{\partial x}\log E(Y|x,w)=\frac{\partial}{\partial x}\log E(Y|x),$$

the average collapsibility also holds. □

Proof of Theorem 5 (a): Observe that

$$\begin{aligned}
&E_{W|x}\left(\frac{\partial}{\partial x}E(Y|x,W)\right)\\
&\quad=\int_{w_2}\int_{w_1}\left(\frac{\partial}{\partial x}E(Y|x,w)\right)dF(w_1,w_2|x)\\
&\quad=\int_{w_2}\int_{w_1}\left(\frac{\partial}{\partial x}E(Y|x,w_1,w_2)\right)dF(w_1|x,w_2)\,dF(w_2|x)\\
&\quad=\int_{w_2}\left(\int_{w_1}\frac{\partial}{\partial x}E(Y|x,w_2)\,dF(w_1|x,w_2)\right)dF(w_2|x)(\because Y \perp\!\!\!\perp W_1|(X,W_2))\\
&\quad=\int_{w_2}\left(\frac{\partial}{\partial x}E(Y|x,w_2)\right)dF(w_2|x)
\end{aligned}$$

$$= E_{W_2|x}\left(\frac{\partial}{\partial x}E(Y|x, W_2)\right)$$
$$= \frac{\partial}{\partial x}E(Y|x), \quad \text{for all } x,$$

using $X \perp\!\!\!\perp W_2$ and condition (ii) of Theorem 1.

(b): Note that

$$\begin{aligned}
&E_{W|x}\left(\frac{\partial^2}{\partial x \partial y}\log f(y|x, W)\right)\\
&= \int_{w_2}\left(\int_{w_1}\frac{\partial^2}{\partial x \partial y}\log f(y|x, w_1, w_2)\, dF(w_1|x, w_2)\right) dF(w_2|x)\\
&= \int_{w_2}\left(\int_{w_1}\frac{\partial^2}{\partial x \partial y}\log f(y|x, w_2)\, dF(w_1|x, w_2)\right) dF(w_2|x)(\because Y \perp\!\!\!\perp W_1|(X, W_2))\\
&= \int_{w_2}\left(\frac{\partial^2}{\partial x \partial y}\log f(y|x, w_2)\right) dF(w_2|x)\\
&= E_{W_2|x}\left(\frac{\partial^2}{\partial x \partial y}\log f(y|x, W_2)\right)\\
&= \frac{\partial^2}{\partial x \partial y}\log f(y|x), \quad \text{for all } x \text{ and } y,
\end{aligned}$$

which follows using $X \perp\!\!\!\perp W_2|Y$ and condition (ii) of Theorem 2. □

Proof of Theorem 6 Integrating the left-hand side of (20) by parts, we get

$$\begin{aligned}
&\int P(T > t+s|T > t, x, y)\frac{\partial}{\partial x}f(y|T > t, x)dy\\
&= \left[P(T > t+s|T > t, x, y)\frac{\partial}{\partial x}F(y|T > t, x)\right]_{-\infty}^{\infty}\\
&\quad - \int \frac{\partial}{\partial y}P(T > t+s|T > t, x, y)\frac{\partial}{\partial x}F(y|T > t, x)dy\\
&= -\int \frac{\partial}{\partial y}P(T > t+s|T > t, x, y)\frac{\partial}{\partial x}F(y|T > t, x)dy.
\end{aligned}$$

Hence, (20) holds if either

$$\frac{\partial}{\partial y}P(T > t+s|T > t, x, y) = \frac{\partial}{\partial y}\left(\frac{P(T > t+s|x, y)}{P(T > t|x, y)}\right) \equiv 0, \quad \text{for all } (t, s, x, y), \tag{39}$$

or

$$\frac{\partial}{\partial x}\Big(F(y|T > t, x)\Big) \equiv 0, \quad \text{for all } (t, x, y). \tag{40}$$

Note that Eqs. (39) and (40), respectively, hold when conditions (i) and (ii) are satisfied. The result now follows. □

References

Bishop, Y. M. M. (1971). Effects of collapsing multidimensional contingency tables. *Biometrics*, *27*, 545–562.

Cochran, W. G. (1938). The omission or addition of an independent variable in multiple linear regression. *Journal of the Royal Statistical Society, Supplementary*, *5*, 171–176.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series, B*, *34*, 187–220.

Cox, D. R. (2003). Conditional and marginal association for binary random variables. *Biometrika*, *90*, 982–984.

Cox, D. R., Wermuth, N. (2003). A general condition for avoiding effect reversal after marginalization. *Journal of the Royal Statistical Society, Series B*, *65*, 937–941.

Di Serio, C., Rinott, Y., Scarsini, M. (2009). Simpson's paradox in survival models. *Scandinavian Journal of Statistics*, *36*, 463–480.

Geng, Z. (1992). Collapsibility of relative risk in contingency tables with a response variable. *Journal of the Royal Statistical Society, Series B*, *54*, 585–593.

Greenland, S., Pearl, J., Robbins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, *10*, 37–48.

Ma, Z., Xie, X., Geng, Z. (2006). Collapsibility of distribution dependence. *Journal of the Royal Statistical Society, Series B*, *68*, 127–133.

Mantel, N., Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*, 719–748.

Pettitt, A. N. (1984). Proportional odds models for survival data and estimates using ranks. *Applied Statistics, 33,* 169–175.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B*, *13*, 238–241.

Vellaisamy, P. (2012). Average collapsibility of distribution dependence and quantile regression coefficients. *Scandinavian Journal of Statistics*, *39*, 153–165.

Vellaisamy, P., Vijay, V. (2007). Some collapsibility results for n-dimensional contingency tables. *Annals of the Institute of Statistical Mathematics*, *59*, 557–576.

Vellaisamy, P., Vijay, V. (2008). Collapsibility of regression coefficients and its extensions. *Journal of Statistical Planning and Inference*, *138*, 982–994.

Wang, X., Geng, Z., Chen, H., Xie, X. (2009). Detecting multiple confounders. *Journal of Statistical Planning and Inference*, *139*, 1073–1081.

Wermuth, N. (1987). Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable. *Journal of the Royal Statistical Society, Series B*, *49*, 353–364.

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

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

Whittemore, A. S. (1978). Collapsibility of multidimensional contingency tables. *Journal of the Royal Statistical Society, Series B*, *40*, 328–340.

Xie, X., Ma, Z., Geng, Z. (2008). Some association measures and their collapsibility. *Statistica Sinica*, *18*, 1165–1183.

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