Published in final edited form as:

Stat Methods Med Res. 2018 November ; 27(11): 3224–3235. doi:10.1177/0962280217695345.

Conditional Independence Test by Generalized Kendall's tau with Generalized Odds Ratio

Shuang Ji¹, Jing Ning², Jing Qin³, and Dean Follmann³

¹Morgan Stanley, New York, NY, USA

²Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

³Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Rockville, MD, USA

Abstract

Determining conditional dependence is a challenging but important task in both model building and in applications such as genetic association studies and graphical models. Research on this topic has focused on kernel-based methods or has used categorical conditioning variables because of the challenge of the curse of dimensionality. To overcome this challenge, we propose a class of tests for conditional independence without any restriction on the distribution of the conditioning variables. The proposed test statistic can be treated as a generalized weighted Kendall's tau, in which the generalized odds ratio is utilized as a weight function to account for the distance between different values of the conditioning variables. The test procedure has desirable asymptotic properties and is easy to implement. We evaluate the finite sample performance of the proposed test through simulation studies and illustrate it using two real data examples.

Keywords

Conditional independence; Generalized odds ratio; Generalized Kendall's tau; U-statistics

1. Introduction

1.1. Motivation

Statistical conditional independence is a fundamental premise in various applications or is a necessary assumption to ensure the correctness of inference in statistical models. It is much more challenging to test for conditional independence than for unconditional independence, as pointed out by Bergsma (2004)¹. Let Y , S and X denote three sets of random variables. The conditional independence between Y and S , given X , denoted by $Y \perp S | X$, means that given X , further information about Y does not provide any additional information about S . One commonly used method to measure conditional dependence is Pearson's partial

correlation coefficient $\rho_{Y \cdot S \cdot X}$, defined as the Pearson correlation between the errors from the regressions $Y = g(X) + \epsilon$ and $S = h(X) + \epsilon_2$. When X is a scalar variable,

$$\rho_{Y \cdot S \cdot X} = \frac{\rho(Y, S) - \rho(Y, X)\rho(S, X)}{\sqrt{\{1 - \rho^2(Y, X)\}\{1 - \rho^2(S, X)\}}},$$

where $\rho(W_1, W_2)$ denotes the Pearson correlation between two random variables, W_1 and W_2 . We can similarly define Kendall's partial tau and Spearman's partial rank correlation by replacing $\rho(\cdot, \cdot)$ with appropriate coefficients. However, test statistics based on such measures are not always consistent estimators of zero under the null hypothesis $Y \perp S | X^2$, and therefore may lead to inflated type I error rates. Approaches that compare the conditional distribution of Y , given both X and S , with that of Y , given only X , have also received a fair amount of attention. Su and White (2007)³ and Su and White (2008)⁴ considered the differences in the corresponding density functions and characteristic functions. Without requiring estimation on the distribution of Y , given $\{X, S\}$, Song (2009)⁵ developed a test on the basis of a Rosenblatt transformation, in which the conditional density functions of Y and S , given some function of X , need to be estimated. Huang (2010)⁶ defined the "maximal nonlinear conditional correlation" and constructed a test statistic by a weighted average of this coefficient, given each chosen value of X . Tests of this class often involve smoothing techniques and are not easy to implement in general^{7,8}.

Kendall's tau, the probability of the difference between the concordant pairs and the discordant pairs, is one of the most popular methods for testing the unconditional independence between two variables. Here, concordant and discordant pairs describe the relationship between pairs of observations. Specifically, for two paired observations (Y_i, S_i) and (Y_j, S_j) , the two pairs are concordant if $(Y_i - Y_j)(S_i - S_j) > 0$, and are discordant if $(Y_i - Y_j)(S_i - S_j) < 0$. A number of studies have been conducted to extend Kendall's tau for the conditional independence test. When the conditioning variables are categorical, Korn (1984)⁹ and Taylor (1987)¹⁰ generalized Kendall's tau to a weighted sum of Kendall's tau over all values of X for testing conditional independence. However, for continuous X , there is no ideal solution because of the challenge of high dimensionality. Goodman (1959)¹¹ and Quade (1974)¹² considered a partial version of Kendall's tau by assessing the concordance and discordance of (Y_i, S_i) and (Y_j, S_j) , where X_i and X_j are close to each other. Similarly, Gripenberg (1992)¹³ proposed a partial rank correlation by comparing consecutive pairs of (Y_i, S_i) and (Y_{i+1}, S_{i+1}) , which are ordered by the value of X . In addition to lacking clear interpretations of the estimators, these methods do not fully use all possible pairwise comparisons and may not be efficient in terms of statistical power. Zhu and others (2012)¹⁴ developed a weighted U-statistic that incorporates the distance between any X_i and X_j by utilizing a nonparametric weight function, for which determining the bandwidth is often computationally intensive.

In view of the importance of the test for conditional independence, it is worth pursuing an easy-to-implement test procedure that has an advantage over the existing tests based on Kendall's tau in terms of generality. We have developed a new test for conditional independence based on a generalized Kendall's tau using a generalized odds ratio proposed

by Liang and Qin (2000)¹⁵. The proposed test imposes no restriction on the distribution of the conditioning variable X . When X is continuous, our test conducts comparisons between all possible pairs, regardless of the distance between the values of X . In addition, our test does not require estimating the conditional distribution of Y , given $\{X, S\}$. We establish the null distribution of the generalized Kendall's tau by U-statistic theories and provide a closed form for its limiting variance.

1.2. Examples

This section contains six examples to illustrate our targeted problem.

Example 1: Validation of surrogate endpoints—Surrogate endpoints are often used in the evaluation of experimental treatments or other interventions. The surrogate endpoints replace or supplement the true endpoints when the surrogate endpoints can be measured earlier, more conveniently or more frequently¹⁶. Prentice (1989)¹⁷ proposed a formal definition of surrogate endpoints and outlined a set of criteria. A key component for a valid surrogate endpoint is that the true endpoint must be conditionally independent of the treatment assignment, given the surrogate endpoint. Such conditional independence ensures that all factors that influence the true endpoint act solely through the surrogate. Hence, determining the best procedure by which to ascertain conditional independence is an essential task in the evaluation of surrogate endpoints.

Example 2: Markov transition model in longitudinal data analysis—Markov transition models are often used in analyzing longitudinal data¹⁸, in which the conditional independence is the fundamental assumption to account for the withinsubject correlation. A q -order Markov transition model assumes that for a series of observations $\{Y_j\}_{j=1}^m$ over time points $\{t_j\}_{j=1}^m$, the conditional distribution of Y_j given the history up to Y_{j-1} , depends on only the last q observations up to Y_{j-1} . For a first-order transition model ($q = 1$), this means that $Y_j \perp \{Y_1, \dots, Y_{j-2}\} | Y_{j-1}$. Hence, testing for conditional independence is an important component of model checking when using Markov transition models. Our proposed method can be directly applied for testing a 1-order Markov assumption, and can be extended for a q -order ($q > 1$) Markov assumption by combining the test statistics for each component¹⁹.

Example 3: Conditional independence screening for biomarker discovery—Genomic research has generated a large number of candidate biomarkers that have potential use in cancer diagnosis and prognosis. When established risk factors exist, researchers seek to determine whether the new biomarkers can improve risk prediction beyond what is possible with the established risk factors. For example, when evaluating the risk of progression from liver cirrhosis to hepatocellular carcinoma, the question of interest is to identify whether blood-based assays of serum biomarkers (e.g., leptin, adiponectin) are conditionally independent of the risk of developing hepatocellular carcinoma given the established risk factors, including the baseline Child-Turcotte-Pugh score and the baseline liver biopsy Ishak score²⁰. We let Y denote the binary outcome of disease progression from liver cirrhosis to hepatocellular carcinoma, X denote the set of the established risk factors

via the standard logistic regression models²¹, and S denote the candidate biomarkers. The scientific question is then whether the candidate biomarkers S , which are often expensive to measure, can improve risk prediction, given the information on X .

Example 4: Association test in a genetic study—Association analysis in many genetic studies has emerged as a powerful tool for identifying genes associated with the disease of interest. Genetic studies generally measure covariates (X), such as environmental factors, which may affect the relationship between the genes (S) and traits (Y). Ignoring such potentially important covariates in the association analysis may result in misleading associations between the genes and traits or affect the test power¹⁴. The test for conditional independence can identify genes with pathogenic associations such that they are still associated with the traits after adjusting the covariates.

Example 5: Validation of instrumental variables—Instrumental variables provide a way to correct for nonrandom treatment assignment and are commonly used in economics and in observational medical studies to draw causal conclusions about a treatment²². An instrumental variable's only impact on the outcome is through the treatment. Formally, an instrument S for outcome Y from treatment X satisfies two conditions: 1) $Y \perp S|X$ and 2) $X \not\perp S$. The latter condition is easy to examine while the former condition can be examined using our proposed methods.

2. Main Results

2.1. Test procedure

Without loss of generality, suppose X is a scalar variable. If the conditioning variable X is discrete, then Kendall's tau can be applied for each value of X , and a weighted sum of Kendall's tau over the possible values of X can be used to test for conditional independence¹⁰. As X is continuous, it would be difficult to find two subjects who have the same value of X . Although some work has been done to compare concordant and discordant pairs when the values of X are close, determining whether the pairs are close based on the values of X is challenging. Also, when only close pairs are compared and utilized in the test procedures, we may lose statistical efficiency. To enable pairwise comparison regardless of the distances between the values of X , we employ the generalized odds ratio proposed by Liang and Qin (2000)¹⁵. For a pair of subjects with measurements (y_i, x_i, s_i) and (y_j, x_j, s_j) , the generalized odds ratio is defined as

$$R(y_i, x_i, y_j, x_j; \beta) = \frac{f_Y(y_j | x_i; \beta) f_Y(y_i | x_j; \beta)}{f_Y(y_i | x_i; \beta) f_Y(y_j | x_j; \beta)},$$

where β is the parameter involved in the conditional distribution of Y given X . For brevity, hereinafter we denote $f_Y(y|x; \beta)$ by $f_Y(y|x)$ and $R(y_i, x_i, y_j, x_j; \beta)$ by $R_{ij}(y, x; \beta)$. Our goal is to evaluate whether further information about S provides any additional information about Y given X and the established conditional distribution of Y given X .

Under the assumption of conditional independence $Y \perp\!\!\!\perp S|X$, for any function $\eta(Y_i, X_i, S_i, Y_j, X_j, S_j)$, we have

$$\begin{aligned} & E\left\{\eta(Y_i, X_i, S_i, Y_j, X_j, S_j) - \eta(Y_j, X_j, S_j, Y_i, X_i, S_i) R_{ij}(\mathbf{y}, \mathbf{x}; \beta) \middle| x_i, s_i, x_j, s_j\right\} \\ &= \iint \left\{\eta(y_i, x_i, s_i, y_j, x_j, s_j) - \eta(y_j, x_j, s_j, y_i, x_i, s_i) R_{ij}(\mathbf{y}, \mathbf{x}; \beta)\right\} f_Y(y_i | x_i) f_Y(y_j | x_j) dy_i dy_j \\ &= \iint \left\{\eta(y_i, x_i, s_i, y_j, x_j, s_j) f_Y(y_i | x_i) f_Y(y_j | x_j) \right. \\ &\quad \left. - \eta(y_j, x_j, s_j, y_i, x_i, s_i) f_Y(y_j | x_j) f_Y(y_i | x_i)\right\} dy_i dy_j = 0. \end{aligned}$$

Then by the double expectation theorem, we can show that

$$E\left\{\eta(Y_i, X_i, S_i, Y_j, X_j, S_j) - \eta(Y_j, X_j, S_j, Y_i, X_i, S_i) R_{ij}(\mathbf{y}, \mathbf{x}; \beta)\right\} = 0. \quad (1)$$

Note that the function $\eta(Y_i, X_i, S_i, Y_j, X_j, S_j)$ is the function $\eta(Y_j, X_j, S_j, Y_i, X_i, S_i)$ with the roles of Y_i and Y_j switched. Equation (1) implies that the generalized odds ratio can adjust for the distance between x_i and x_j such that the adjusted difference by switching the roles of Y_i and Y_j has a mean of zero under the conditional independence assumption. Accordingly, we can construct the following class of test statistics for testing $H_0: Y \perp\!\!\!\perp S|X$,

$$T^Y(\beta) = \frac{2}{n(n-1)} \sum_{i < j} \left\{ \eta(y_i, x_i, s_i, y_j, x_j, s_j) - \eta(y_j, x_j, s_j, y_i, x_i, s_i) R_{ij}(\mathbf{y}, \mathbf{x}; \beta) \right\}. \quad (2)$$

Motivated by Kendall's tau, a special case in this class is attained by letting

$$\eta(y_i, x_i, s_i, y_j, x_j, s_j) = I\left\{(y_i - y_j)(s_i - s_j) > 0\right\},$$

and the resulting test statistic is

$$T^Y(\beta) = \frac{2}{n(n-1)} \sum_{i < j} \xi(y_i, x_i, s_i, y_j, x_j, s_j; \beta), \quad (3)$$

where $\xi(y_i, x_i, s_i, y_j, x_j, s_j; \beta) =$

$$I\left\{(y_i - y_j)(s_i - s_j) > 0\right\} - I\left\{(y_i - y_j)(s_i - s_j) < 0\right\} R_{ij}(\mathbf{y}, \mathbf{x}; \beta).$$

This test statistic calculates the difference between concordant pairs and weighted discordant pairs, in which the generalized odds ratio is used as the weight function to account for the effect of the conditioning variable X . Note that $R_{ij}(\mathbf{y}, \mathbf{x}; \beta) = 1$ when $x_i = x_j$, implying that there is no need to adjust the discordant indicator for pairs that share the same value of X . In a very special case with constant values of X , $T^Y(\beta)$ reduces to the conventional Kendall's tau. Therefore, we term $T^Y(\beta)$ as a generalized Kendall's tau statistic. The test statistic

involves an unknown quantity, β . However, under $H_0: Y \perp S|X$, β can be readily estimated by $\hat{\beta}$, the maximizer of the conditional log-likelihood:

$$l_Y(\beta) = \sum_{i=1}^n \log f_Y(y_i|x_i; \beta).$$

We have established the asymptotic property of $T^Y(\hat{\beta})$ under H_0 after accounting for the additional variation due to the estimated value of β . From equation (1) and the fact that $\hat{\beta}$ is the maximum-likelihood estimator for the true value β_0 , it is easy to see that, under mild regularity conditions for the consistency of $\hat{\beta}$, we have $T^Y(\hat{\beta}) \xrightarrow{P} 0$. The large sample properties of the test statistic are summarized in Theorem 1, with proofs given in the Appendix.

Theorem 1.—Under the null hypothesis and regularity conditions A(1)-A(4) in the Appendix,

$$\sqrt{n}T^Y(\hat{\beta}) \xrightarrow{D} N(0, \sigma^Y),$$

$$\text{where } \sigma^Y = E \left[2\zeta(Y, X, S) - A_1 \cdot J_1^{-1} \frac{\partial f_Y(y|x; \beta_0)}{\partial \beta} \right]^2, J_1 = E \left[-\frac{\partial^2 l_Y(\beta_0)}{\partial \beta \partial \beta^T} \right] / n,$$

$$\zeta(y_i, x_i, s_i) = E \left\{ \xi(y_i, x_i, s_i, Y_j, X_j, S_j; \beta_0) \middle| y_i, x_i, s_i \right\},$$

$$\text{and } A_1 = E \left[I \left\{ (Y_i - Y_j)(S_i - S_j) < 0 \right\} \frac{\partial R_{ij}(Y, X; \beta_0)}{\partial \beta} \right].$$

2. 2. Alternative tests by symmetry

It is worth noting that the test statistic in equation (3) is symmetric with respect to Y and S . By switching the roles of Y and S , we can develop an alternative test. Under the null hypothesis $Y \perp S|X$, we have

$$\begin{aligned} & E \left\{ \eta(Y_i, X_i, S_i, Y_j, X_j, S_j) - \eta(Y_j, X_j, S_j, Y_i, X_i, S_i) R_{ij}(s, x; \alpha) \right\} \\ &= E \left\{ E \left\{ \eta(Y_i, X_i, S_i, Y_j, X_j, S_j) - \eta(Y_j, X_j, S_j, Y_i, X_i, S_i) R_{ij}(s, x; \alpha) \middle| x_i, y_i, x_j, y_j \right\} \right\} \\ &= E \left[\iint \left\{ \eta(y_i, x_i, s_i, y_j, x_j, s_j) - \eta(y_i, x_i, s_j, y_j, x_j, s_i) R_{ij}(s, x; \alpha) \right\} \right. \\ &\quad \left. \times f_S(s_i|x_i) f_S(s_j|x_j) ds_i ds_j \right] \\ &= E \left[\iint \left\{ \eta(y_i, x_i, s_i, y_j, x_j, s_j) f_S(s_i|x_i) f_S(s_j|x_j) \right. \right. \\ &\quad \left. \left. - \eta(y_i, x_i, s_j, y_j, x_j, s_i) f_S(s_j|x_j) f_S(s_i|x_i) \right\} ds_i ds_j \right] = 0 \end{aligned}$$

where

$$R_{ij}(s, \mathbf{x}; \alpha) = \frac{f_S(s_j | x_i; \alpha) f_S(s_i | x_j; \alpha)}{f_S(s_i | x_i; \alpha) f_S(s_j | x_j; \alpha)},$$

and α is the parameter involved in the conditional distribution of S given X .

The corresponding test statistic can be defined as

$$T^S(\alpha) = \frac{2}{n(n-1)} \sum_{i < j} \xi^*(y_i, x_i, s_i, y_j, x_j, s_j; \alpha), \quad (4)$$

where $\xi^*(y_i, x_i, s_i, y_j, x_j, s_j; \alpha) = I\{(y_i - y_j)(s_i - s_j) > 0\} - I\{(y_i - y_j)(s_i - s_j) < 0\} R_{ij}(s, \mathbf{x}; \alpha)$.

By utilizing both (3) and (4), we can obtain a combined test statistic

$$T^C(\beta, \alpha) = \frac{2}{n(n-1)} \sum_{i < j} \xi^C(y_i, x_i, s_i, y_j, x_j, s_j; \beta, \alpha), \quad (5)$$

where $\xi^C(y_i, x_i, s_i, y_j, x_j, s_j; \beta, \alpha) =$

$$2I\{(y_i - y_j)(s_i - s_j) > 0\} - I\{(y_i - y_j)(s_i - s_j) < 0\} R_{ij}(y, s, \mathbf{x}; \beta, \alpha)$$

with

$$R_{ij}(y, s, \mathbf{x}; \beta, \alpha) = R_{ij}(y, \mathbf{x}; \beta) + R_{ij}(s, \mathbf{x}; \alpha).$$

Under the null hypothesis $H_0: Y \perp S | X$, the unknown parameter α can be similarly estimated by maximizing the conditional log-likelihood

$$l_S(\alpha) = \sum_{i=1}^n \log f_S(s_i | x_i; \alpha).$$

After plugging in the estimators $\hat{\beta}$ and $\hat{\alpha}$, the asymptotic distributions of test statistics $T^S(\hat{\alpha})$ and $T^C(\hat{\beta}, \hat{\alpha})$ can be established in a manner similar to that of $T^Y(\hat{\beta})$ under H_0 .

Corollary 1.—Under the null hypothesis and regularity conditions A(1)-A(8) in the Appendix, we have

$$\sqrt{n}T^S(\hat{\alpha}) \xrightarrow{D} N(0, \sigma^S) \text{ and } \sqrt{n}T^C(\hat{\beta}, \hat{\alpha}) \xrightarrow{D} N(0, \sigma^C),$$

where

$$\begin{aligned}
\sigma^S &= E \left\{ \zeta(Y, X, S) - A_2 \cdot J_2^{-1} \frac{\partial f_S(s|X; \alpha_0)}{\partial \alpha} \right\}^2, \\
\sigma^C &= E \left\{ 2\zeta(Y, X, S) - A_1 \cdot J_1^{-1} \frac{\partial f_Y(y|X; \beta_0)}{\partial \beta} - A_2 \cdot J_2^{-1} \frac{\partial f_S(s|X; \alpha_0)}{\partial \alpha} \right\}^2, \\
J_2 &= E \left\{ -\frac{\partial^2 \sum_{i=1}^n \log f_S(s_i|x_i; \alpha_0)}{\partial \alpha \partial \alpha^T} \right\} / n \\
A_2 &= E \left\{ I \left\{ (Y_i - Y_j)(S_i - S_j) < 0 \right\} \frac{\partial R_{ij}(S, X; \alpha_0)}{\partial \alpha} \right\}.
\end{aligned}$$

While the proposed tests based on (3), (4) and (5) are valid, an interesting question is which one leads to a more efficient test for conditional independence. The test based on (5) appears to be more powerful than the others; however, it requires the correct specification of both conditional distributions $f_Y(y|x; \beta)$ and $f_S(s|x; \alpha)$, and thus may lose some robustness as a result. Next, we will use empirical studies to compare the performance of three tests.

3. Simulation Studies

We used Monte Carlo simulations to evaluate and compare the performance (type I error rate and power) of the proposed tests and Taylor's test¹⁰ under various scenarios. We used small ($n=50$) to moderate ($n=400$) sample sizes. We used 5000 and 1000 replications of the tests to calculate the type I error rate and power, respectively.

We generated independent triplets of (Y_i, S_i, X_i) , $i = 1, \dots, n$. For the distribution of X_i we considered a continuous case, where X_i followed $Unif(0,1)$, and an ordinal case, where X_i followed $Discrete Unif(0, 1; M)$ with $\Pr(X_i = \frac{m}{M}) = \frac{1}{M+1}$, $m = 0, \dots, M$. We generated the variable S_i from the model $S_i = 2 - X_i + \delta_i$, where δ_i followed a standard normal distribution $N(0, 1)$, independently of X_i . For the variable Y_i we also considered two cases: (a) Y_i was continuous, generated from $Y_i = 1 + X_i + kS_i + \epsilon_i$, where ϵ_i followed $N(0, 1)$, independently of X_i and δ_i ; and (b) Y_i was binary, generated from $\text{logit}(Y_i) = 0.5 + 0.5X_i + kS_i$. In both cases, k was a constant that ranged from 0 to 1, indicating the strength of the conditional dependence between Y and S given X .

Tables 1 and 2 list the rates of rejecting the null hypothesis at a significance level of 0.05 based on simulations with continuous Y . In Table 1, the conditioning variable X was generated from a standard uniform distribution. To apply Taylor's test¹⁰, which was only applicable to the cases with discrete X , we dichotomized the variable X using cutoff values of 0.7 or 0.8. We fitted a linear regression model to derive the generalized odds ratio in the proposed tests. As seen in Table 1, when both Y and X were continuous, each of the three proposed tests based on T^Y , T^S and T^C performed satisfactorily and comparably. The type I error rates were well maintained in the null scenario ($k = 0$), especially for larger sample sizes. When $k = 0.6$, powers greater than 95% were achieved at $n = 200$. Unlike the proposed tests, Taylor's test could not preserve the specified significance level. For instance, when the cutoff value to dichotomize X was 0.8 and the sample size was 400, the rate of rejection

achieved by Taylor's test was 0.142, which is much larger than the nominal value of 0.05. We conducted additional simulations by increasing the sample size. When the sample size increased to 1000, the rate of rejection achieved by Taylor's test increased to 0.280, while those of the three proposed tests were close to 0.05. As expected, the power of the proposed tests increased with the increasing sample size and the increasing strength of the conditional dependence between Y and S given X (i.e., increasing value of k). Somewhat surprisingly, the three proposed tests produced comparable powers, suggesting that the combined test T^C may not gain additional efficiency compared with the tests T^Y and T^S when both Y and X are continuous variables.

For the scenarios with discrete values of variable X from $Unif(0, 1; M)$, we conducted several sets of simulations by choosing different values of M over $[2, 10]$. Similar findings were observed across the scenarios with different values of M . In Table 2, we present simulation results when $M = 5$. Under this scenario, the type I error rates for both the proposed tests and Taylor's test were reasonably close to the nominal value of 0.05. For the power comparison, the proposed tests exhibited comparable power and had slightly superior power compared to that of Taylor's test. In addition, the simulation results suggest that our methods are most advantageous over Taylor's test when $M = 10$, i.e., when the distribution of X is closest to the uniform distribution. This is expected, as Taylor's test is only applicable for categorical X , while our test statistics work regardless of the distribution of X .

In Tables 3 and 4, we present the rate of rejecting the null hypothesis at a significance level of 0.05 for binary values of Y , with $X \sim Unif(0, 1)$ and with $X \sim Discrete\ Unif(0, 1; M)$ respectively. To apply Taylor's test when $X \sim Unif(0, 1)$, we again dichotomized the variable X using cutoff values of 0.7 or 0.8. We fitted a logistic regression model to derive the generalized odds ratio in the proposed tests. We observed results from the proposed tests for binary values of Y that were similar to those for continuous values of Y : the proposed tests preserved the nominal error rate under the null hypothesis and had similar performance levels under the alternative hypothesis regardless of the distribution of X . It is interesting that the rejection rates achieved by Taylor's test were close to 0.05 even when the underlying conditional covariate X was first generated from the uniform distribution and then was dichotomized in the test procedure. However, when the sample size increased to 1000, the rejection rates achieved by Taylor's test remained slightly inflated (around 0.06), and did not converge to the nominal value, which was different than the results achieved by the proposed tests. For the power comparison, the proposed tests exhibited superior power compared to Taylor's test in the scenarios with continuous values of X . For example, in Table 3, the proposed tests achieved 15 to 22% greater power than Taylor's test for the scenario with $k = 0.3$ and $n = 400$. When both Y and X were discrete, the validity of Taylor's test, which was originally proposed for this scenario, was not in doubt (see Table 4). Our methods, which were designed for a more general purpose, showed comparable performance in this scenario. Note that the proposed test had slightly larger type I error rates when the sample size was 50. A possible explanation is that the performance of the proposed tests partially relies on the maximum likelihood estimators of the regression model and may need a reasonably large sample size to have satisfactory performance.

4. Data Application

We applied the proposed tests to data from two studies. The first application concerns the clearance rate of the cardiac drug digoxin in patients treated for heart failure. The digoxin clearance rates of 35 consecutive patients under treatment for heart failure were reported by Halkin and others (1975)²³ and further analyzed by Bergsma (2013)²⁴. The hypothesis of interest is whether digoxin clearance is independent of urine flow after conditioning for creatinine clearance. Considering the skewness of the marginal distributions, we first used log transformation on the three variables of interest (digoxin clearance, urine flow, and creatinine clearance), and fit a regression model of the log digoxin clearance on the log creatinine clearance to derive the generalized odds ratio. We then computed the generalized Kendall's tau test statistic from a pairwise comparison of the log digoxin clearance and the log urine flow, adjusted by the generalized odds ratio. The normalized test statistic was 5.81, with a p-value < 0.001 , which suggested that the digoxin clearance was not independent of the urine flow even after controlling for the creatinine clearance. As an alternative approach, we regressed the log urine flow on the log creatinine clearance and then derived the other two test statistics. We obtained the same conclusion from the alternative approach, that the digoxin clearance and urine flow were statistically dependent even when given the information on the creatinine clearance.

The purpose of the second application is to test the conditional independence for model building in longitudinal data analysis. We applied our method to the AIDS Clinical Trials Group 175 study²⁵, a randomized clinical trial that compared monotherapy with combination therapy in the treatment of adults with HIV who had CD4 T cell counts between 200 and 500 per cubic millimeter of blood. Among the 2139 patients enrolled in this trial, 1093 were treated by monotherapy (either zidovudine or didanosine) and 1046 were treated by combination therapy (zidovudine and didanosine or zidovudine and zalcitabine). One important outcome was the CD4 count, which was measured at baseline, at 20 weeks and at the last visit at 96 weeks. Complete records of each patient's CD4 counts at baseline and at 20 weeks were available, while 37% of the records at 96 weeks were missing. To construct an appropriate Markov transition model to evaluate the treatment effects on the CD4 count, we were interested in testing the conditional independence between a patient's CD4 count at 96 weeks and that at baseline, given his/her CD4 count at 20 weeks.

We applied log transformation to the CD4 counts to ensure the normality of the outcomes. The logarithms of the CD4 counts at baseline, at 20 weeks and at 96 weeks are denoted by Y , X and S , respectively. Our test for $Y \perp S | X$ was conducted in two phases. To fully utilize the observed information, we first regressed Y on X and obtained the generalized odds ratio. At that stage, we did not include S ; therefore, the skewness and missingness of S had no impact on our analysis. We then computed the generalized Kendall's tau test statistic from a pairwise comparison of Y and S . This phase was non-parametric, and only data from patients who had complete records of CD4 counts at all three time points were included.

Our analysis resulted in a normalized test statistic of -0.96 , with a 0.34 p-value. To explore the potential influence on the variability of the results, we conducted a sensitivity analysis.

Figure 1 displays the diagnostics of the linear regression of Y on X , which suggest adequate model fitting, except that the tails of the residuals appear to be somewhat heavy. After removing the candidate(s) for outliers, observation(s) with an absolute value of the residual greater than 1.5 (1 subject) or 1.0 (7 subjects), we obtained test statistics of -0.34 (variance = 0.069) and -0.49 (variance = 0.147), respectively. Both cases resulted in a p -value of 0.19, indicating failure to reject the null hypothesis $Y \perp S|X$, which was consistent with the findings we observed before removing the outliers. Therefore, we concluded that, given the CD4 count at 20 weeks, the CD4 count at baseline did not provide any additional information toward the CD4 count at 96 weeks. Accordingly, it appeared that the use of a first-order transition model of the CD4 counts was adequate for the longitudinal data.

5. Discussion

Conditional independence is often required in statistical analysis, but is infrequently tested before being assumed. One possible reason is that we lack a simple unifying approach for conducting the test. We have built a new framework for testing conditional independence that is based on a generalized Kendall's tau that incorporates the generalized odds ratio. The implementation of the proposed test procedure is straightforward and is greatly facilitated by the closed-form asymptotic nature of our test statistic under the null hypothesis.

To estimate the generalized odds ratio, which involves estimating the conditional density of Y given X , we adopted a parametric approach. Therefore, it is important to conduct proper diagnostics to validate the adequacy of the parametric model we assumed. Considering the main goal of our study is to evaluate whether further information about S provides any additional information about Y given X and the established conditional distribution of Y given X , we chose the parametric model to estimate the conditional density of Y given X . For example, in biomarker discovery studies (Example 3), the density function of the disease status given the established biomarkers, obtained by using logistic regression, has been well studied in the existing literature. Our test is to evaluate whether the new biomarkers can improve risk prediction beyond what is possible with the established risk factors. Alternatively, we can consider using nonparametric regression approaches, such as kernel smoothing and smoothing splines, to estimate the conditional densities. This will substantially impact the large sample property of the test statistic and merits future research.

This work presents a general framework for conditional independence testing that can be extended to numerous settings. Thus, there remain many interesting extensions and questions to study in future work, including how to accommodate complications such as truncation and censoring to enable testing for conditional independence in survival data.

Acknowledgments

The work was supported in part by the U.S. National Institute of Health (NIH) grants CA193878 and CA016672.

APPENDIX

A.1. Regularity conditions

Without loss of generality, we let β be a scalar for simplicity of notation. We assume the following regularity conditions:

A(1) For each $y \in \mathcal{Y}$, $f_Y(y|x; \beta)$ is three times differentiable with respect to β , the third derivative is continuous in β , and $\int f_Y(y|x; \beta)dx$ can be differentiated three times under the integral sign.

A(2) The parameter β is in a compact set whose interior contains the true value β_0 .

A(3) There exists a positive number C and a function $M(y)$ ($E_{\beta_0}[M(y)] < \infty$) such that

$$\left| \partial^3 f_Y(y|x; \beta) / \partial^3 \beta \right| \leq M(y)$$

for all $y \in \mathcal{Y}$ and $\beta \in (\beta_0 - C, \beta_0 + C)$.

A(4) $E\{\xi(Y_1, X_1, S_1, Y_2, X_2, S_2; \beta)\}^2 < \infty$

A(5) For each $s \in \mathcal{S}$, $f_S(s|x; \alpha)$ is three times differentiable with respect to α , the third derivative is continuous in α , and $\int f_S(s|x; \alpha)dx$ can be differentiated three times under the integral sign.

A(6) The parameter α is in a compact set whose interior contains the true value α_0 .

A(7) There exists a positive number C and a function $M(s)$ ($E_{\alpha_0}[M(s)] < \infty$) such that

$$\left| \partial^3 f_S(s|x; \alpha) / \partial^3 \alpha \right| \leq M(s)$$

for all $s \in \mathcal{S}$ and $\alpha \in (\alpha_0 - C, \alpha_0 + C)$.

A(8) $E\{\xi^*(Y_1, X_1, S_1, Y_2, X_2, S_2; \alpha)\}^2 < \infty$

A.2. Proof of Theorem 1

Note that the estimator $\hat{\beta}$ satisfies $\partial l_Y(\hat{\beta}) / \partial \beta = 0$. Then by Taylor's expansion and regularity conditions A(1)-A(2), we have

$$\sqrt{n}(\hat{\beta} - \beta) = J_1^{-1} \sum_{i=1}^n \frac{\partial \log f_Y(y_i|x_i; \beta)}{\partial \beta} + o_p(1). \quad (6)$$

Given the true parameter value β_0 , the test statistic

$$T^Y(\beta_0) = \binom{n}{2}^{-1} \sum_{i < j} \xi(y_i, x_i, s_i, y_j, x_j, s_j; \beta_0)$$

is a second-order U-statistic. Using Powell's projection theorem and regularity condition A(8), we can show that

$$T^Y(\beta_0) = \frac{2}{n} \sum_{i=1}^n \zeta(y_i, x_i, s_i) + o_p(n^{-1/2}). \quad (7)$$

To prove the asymptotic normality of $T^Y(\hat{\beta})$, we study the asymptotic normality of the difference between $T^Y(\hat{\beta})$ and $T^Y(\beta_0)$. Therefore, we have

$$T^Y(\hat{\beta}) - T^Y(\beta_0) \quad (8)$$

$$\begin{aligned} &= -\frac{2}{n(n-1)} \sum_{i < j} I\{(y_i - y_j)(s_i - s_j) < 0\} \{R_{ij}(\mathbf{y}, \mathbf{x}; \hat{\beta}) - R_{ij}(\mathbf{y}, \mathbf{x}; \beta_0)\} \\ &= -\frac{2}{n(n-1)} \sum_{i < j} I\{(y_i - y_j)(s_i - s_j) < 0\} \cdot \frac{\partial R_{ij}(\mathbf{y}, \mathbf{x}; \beta_0)}{\partial \beta} \cdot (\hat{\beta} - \beta_0) + o_p(\hat{\beta} - \beta_0) \\ &= -A_1 \cdot J_1^{-1} \cdot \frac{1}{n} \sum_{i=1}^n \frac{\partial f_Y(y_i | x_i; \beta_0)}{\partial \beta} + o_p(n^{-1/2}). \end{aligned} \quad (9)$$

Combining (8) and (7), we derive the asymptotic independent and identically distributed representation of test statistic $T^Y(\hat{\beta})$,

$$T^Y(\hat{\beta}) = \frac{1}{n} \sum_{i=1}^n \left\{ 2\zeta(y_i, x_i, s_i) - A_1 \cdot J_1^{-1} \cdot \frac{\partial f_Y(y_i | x_i; \beta_0)}{\partial \beta} \right\} + o_p(n^{-1/2}). \quad (10)$$

This, coupled with the consistency of $T^Y(\hat{\beta})$, indicates that

$$\sqrt{n}T^Y(\hat{\beta}) \xrightarrow{D} N(0, \sigma^Y)$$

with $\sigma^Y = E\left\{2\zeta(Y, X, S) - A_1 \cdot J_1^{-1} \cdot \partial f_Y(y|X; \beta_0)/\partial \beta\right\}^2$.

The asymptotic properties of test statistics $T^S(\hat{\alpha})$ and $T^C(\hat{\beta}, \hat{\alpha})$ can be shown by the techniques used in the proof of Theorem 1; we omit the details here.

References

- [1]. Bergsma WP. Testing conditional independence for continuous random variables. EURANDOM-report 2004; : 2004–049.
- [2]. Korn EL. The ranges of limiting values of some partial correlations under conditional independence. *The American Statistician* 1984; 38: 61–62.
- [3]. Su L and White H. A consistent characteristic functionbased test for conditional independence. *Journal of Econometrics* 2007; 141: 807–834.
- [4]. Su L and White H. A nonparametric hellinger metric test for conditional independence. *Econometric Theory* 2008; 24: 829–864.
- [5]. Song K Testing conditional independence via rosenblatt transforms. *The Annals of Statistics* 2009; 37: 4011–4045.
- [6]. Huang TM. Testing conditional independence using maximal nonlinear conditional correlation. *The Annals of Statistics* 2010; 38: 2047–2091.
- [7]. Fukumizu K, Gretton A, Sun X et al. Kernel measures of conditional dependence. In *NIPS*, volume 20 pp. 489–496.
- [8]. Zhang K, Peters J, Janzing D et al. Kernel-based conditional independence test and application in causal discovery. *arXiv preprint arXiv:12023775* 2012.
- [9]. Korn EL. kendall's tau with a blocking variable. *Biometrics* 1984; 40: 209–214. [PubMed: 6733229]
- [10]. Taylor JMG. Kendall's and Spearman's correlation coefficient in the presence of a blocking variable. *Biometrics* 1987; 43: 409–415. [PubMed: 3607205]
- [11]. Goodman LA. Partial tests for partial taus. *Biometrika* 1959; 46: 425–432.
- [12]. Quade D Nonparametric partial correlation In Blalock HM (ed.) *Measurements in the Social Sciences: Theories and Strategies*. Chicago: Aldine, 1974 pp. 369–398.
- [13]. Gripenberg G Confidence intervals for partial rank correlations. *Journal of the American Statistical Association* 1992; 87: 546–551.
- [14]. Zhu W, Jiang Y and Zhang H. Nonparametric covariate-adjusted association tests based on the generalized kendall's tau. *Journal of the American Statistical Association* 2012; 107: 1–11. [PubMed: 22745516]
- [15]. Liang KY and Qin J. Regression analysis under non-standard situations: a pairwise pseudolikelihood approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2000; 62(4): 773–786.
- [16]. Fleming TR, Prentice RL, Pepe MS et al. Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and aids research. *Statistics in medicine* 1994; 13(9): 955–968. [PubMed: 8047747]
- [17]. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in medicine* 1989; 8(4): 431–440. [PubMed: 2727467]
- [18]. Zeger SL and Liang KY. An overview of methods for the analysis of longitudinal data. *Statistics in Medicine* 1992; 11(14–15): 1825–1839. [PubMed: 1480876]
- [19]. Zhang H, Liu CT and Wang X. An association test for multiple traits based on the generalized kendalls tau. *Journal of the American Statistical Association* 2010; 105(490): 473–481. [PubMed: 20711441]
- [20]. Lok AS, Seeff LB, Morgan AM, di Bisceglie T Rand et al. Incidence of hepatocellular carcinoma and associated risk factors in hepatitis c-related advanced liver disease. *Gastroenterology* 2009; 136: 138–148. [PubMed: 18848939]
- [21]. Xiao J, Zhao Y, Varghese RS et al. Evaluation of metabolite biomarkers for hepatocellular carcinoma through stratified analysis by gender, race, and alcoholic cirrhosis. *Cancer Epidemiology Biomarkers & Prevention* 2014; 23(1): 64–72.
- [22]. Greenland S An introduction to instrumental variables for epidemiologists. *International journal of epidemiology* 2000; 29(4): 722–729. [PubMed: 10922351]
- [23]. Halkin H, Sheiner LB, Peck CC et al. Determinants of the renal clearance of digoxin. *Clin Pharmacol Theor* 1975; 17: 385–394.

- [24]. Bergsma W Nonparametric testing of conditional independence by means of the partial copula. arXiv:11014607 2013.
- [25]. Hammer SM, Katzenstein DA, Hughes MD et al. A trial comparing nucleoside monotherapy with combination therapy in hiv-infected adults with cd4 cell counts from 200 to 500 per cubic millimeter. aids clinical trials group study 175 study team. N Engl J Med 1996; 335: 1081–90. [PubMed: 8813038]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.
Diagnostics of the model fitting $Y \sim X$

Table 1.

Rejection rates for testing $Y \perp S|X$ with continuous Y and $X \sim \text{Unif}(0, 1)$.

$k = 0$					
	T^Y	T^S	T^C	$T_{taylor}^{0.7}$	$T_{taylor}^{0.8}$
n=50	0.069	0.071	0.072	0.055	0.059
n=100	0.056	0.060	0.061	0.061	0.072
n=200	0.066	0.054	0.054	0.078	0.102
n=400	0.053	0.053	0.053	0.097	0.142
$k = 0.15$					
	T^Y	T^S	T^C	$T_{taylor}^{0.7}$	$T_{taylor}^{0.8}$
n=50	0.201	0.205	0.212	0.106	0.087
n=100	0.290	0.306	0.295	0.182	0.150
n=200	0.500	0.495	0.496	0.302	0.241
n=400	0.793	0.797	0.804	0.548	0.437
$k = 0.3$					
	T^Y	T^S	T^C	$T_{taylor}^{0.7}$	$T_{taylor}^{0.8}$
n=50	0.516	0.525	0.542	0.338	0.305
n=100	0.785	0.792	0.794	0.614	0.574
n=200	0.970	0.974	0.971	0.903	0.883
n=400	1	1	1	0.907	0.995

T^Y : test based on $R(y, x; \beta)$; T^S : test based on $R(s, x; a)$; T^C : test based on $R(y, s, x; \beta, a)$; T_{taylor} : test by Taylor (1987)

Table 2.Rejection rates for testing $Y \perp S|X$ with continuous Y and $X \sim \text{Discrete Unif}(0,1; M=5)$.

$k = 0$				
	T^Y	T^S	T^C	T_{taylor}
n=50	0.066	0.066	0.067	0.048
n=100	0.059	0.061	0.061	0.049
n=200	0.056	0.056	0.054	0.052
n=400	0.053	0.054	0.051	0.051
$k = 0.15$				
	T^Y	T^S	T^C	T_{taylor}
n=50	0.203	0.202	0.208	0.125
n=100	0.334	0.336	0.335	0.261
n=200	0.547	0.536	0.541	0.490
n=400	0.809	0.808	0.814	0.788
$k = 0.3$				
	T^Y	T^S	T^C	T_{taylor}
n=50	0.512	0.552	0.539	0.372
n=100	0.801	0.798	0.802	0.727
n=200	0.973	0.971	0.974	0.966
n=400	0.999	1	1	1

T^Y : test based on $R(y, x; \beta)$; T^S : test based on $R(s, x; a)$; T^C : test based on $R(y, s, x; \beta, a)$; T_{taylor} : test by Taylor (1987)

Table 3.Rejection rates for testing $Y \perp S|X$ with binary Y and $X \sim \text{Unif}(0,1)$.

$k = 0$					
	T^Y	T^S	T^C	$T_{taylor}^{0.7}$	$T_{taylor}^{0.8}$
n=50	0.075	0.080	0.081	0.054	0.057
n=100	0.051	0.054	0.051	0.041	0.043
n=200	0.058	0.061	0.061	0.054	0.054
n=400	0.051	0.053	0.055	0.055	0.054
$k = 0.3$					
	T^Y	T^S	T^C	$T_{taylor}^{0.7}$	$T_{taylor}^{0.8}$
n=50	0.184	0.198	0.191	0.131	0.142
n=100	0.223	0.234	0.228	0.179	0.178
n=200	0.408	0.412	0.412	0.327	0.310
n=400	0.676	0.674	0.677	0.586	0.556
$k = 0.6$					
	T^Y	T^S	T^C	$T_{taylor}^{0.7}$	$T_{taylor}^{0.8}$
n=50	0.360	0.371	0.368	0.284	0.290
n=100	0.567	0.575	0.583	0.526	0.523
n=200	0.842	0.849	0.851	0.819	0.796
n=400	0.993	0.991	0.995	0.988	0.986

T^Y : test based on $R(y, x; \beta)$; T^S : test based on $R(s, x; a)$; T^C : test based on $R(y, s, x; \beta, a)$.

Table 4.Rejection rates for testing $Y \perp S|X$ with binary Y and $X \sim \text{Discrete Unif}(0,1)$.

$k = 0.15$				
	T^Y	T^S	T^C	T_{taylor}
n=50	0.074	0.079	0.079	0.050
n=100	0.048	0.052	0.049	0.039
n=200	0.059	0.060	0.062	0.053
n=400	0.050	0.056	0.052	0.048
$k = 0.3$				
	T^Y	T^S	T^C	T_{taylor}
n=50	0.182	0.205	0.199	0.123
n=100	0.215	0.236	0.226	0.196
n=200	0.404	0.420	0.409	0.392
n=400	0.671	0.674	0.678	0.659
$k = 0.6$				
	T^Y	T^S	T^C	T_{taylor}
n=50	0.361	0.387	0.378	0.275
n=100	0.552	0.590	0.586	0.499
n=200	0.845	0.860	0.864	0.834
n=400	0.992	0.989	0.993	0.987

T^Y : test based on $R(y, x; \beta)$; T^S : test c based on $R(s, x; a)$; T^C : test based on $R(y, s, x; \beta, a)$; T_{taylor} : test by Taylor (1987)