TESTING CONDITIONAL INDEPENDENCE FOR CONTINUOUS RANDOM VARIABLES

Wicher P. Bergsma[†]

3-12-2004

Abstract: A common statistical problem is the testing of independence of two (response) variables conditionally on a third (control) variable. In the first part of this paper, we extend Hoeffding's concept of estimability of degree r to testability of degree r, and show that independence is testable of degree two, while conditional independence is not testable of any degree if the control variable is continuous. Hence, in a well-defined sense, conditional independence is much harder to test than independence. In the second part of the paper, a new method is introduced for the nonparametric testing of conditional independence of continuous responses given an arbitrary, not necessarily continuous, control variable. The method allows the automatic conversion of any test of independence to a test of conditional independence. Hence, robust tests and tests with power against broad ranges of alternatives can be used, which are favorable properties not shared by the most commonly used test, namely the one based on the partial correlation coefficient. The method is based on a new concept, the partial copula, which is an average of the conditional copulas. The feasibility of the approach is demonstrated by an example with medical data.

Keywords: tests of conditional independence, estimability and testability of degree r, partial copula, partial correlation
AMS Subject Classification: Primary 62H15; secondary 62E99

[†]EURANDOM, PO Box 513, 5600 MB Eindhoven

Contents

1 Introduction

For a given triple of random variables (X, Y, Z) we consider the problem of the nonparametric testing of the hypothesis of conditional independence of Y and Z controlling for X based on n independent and identically distributed (iid) data points $(X_1, Y_1, Z_1), \ldots, (X_n, Y_n, Z_n)$. Following Dawid (1979), this hypothesis is denoted as $Y \perp\!\!\!\perp Z|X$.

For the testing of unconditional independence between two random variables a wide array of tests is available, with the best known ones based on the Pearson correlation, Spearman's rank correlation, or Kendall's tau. Tests of unconditional independence which have asymptotic power against all alternatives were proposed by Hoeffding (1948b) and Schweizer and Wolff (1981).

In contrast, for the (nonparametric) testing of conditional independence there appears to be only one commonly used method, namely the test based on the partial correlation coefficient. With $g(x) = E(Y|X = x)$ and $h(x) = E(Z|X = x)$, it is defined as

$$\rho(Y, Z|X) = \frac{\rho(g(X), h(X)) - \rho(X, g(X))\rho(X, h(X))}{\sqrt{(1 - \rho(X, g(X))^2)(1 - \rho(X, h(X))^2)}} \tag{1}$$

In fact, $\rho(Y, Z|X)$ equals the correlation between the errors in the regressions $y = g(x)+\varepsilon_1$ and $z = h(x)+\varepsilon_2$. Evaluation of the test requires the estimation of the regression curves g and h. An alternative method, which does not appear to be used often, was proposed by Goodman (1959) and is based on a partial version of Kendall's tau, using the number of local concordant and discordant pairs of observations. This test was further discussed in Goodman and Grunfeld (1961) and Gripenberg (1992). Some well-known other tests are based on the linear partial correlation coefficient, which is (1) with g and h replaced by the identity function, or on Kendall's partial tau (Kendall, 1942). However, it is well-known that these coefficients are not necessarily zero under conditional independence unless certain restrictive conditions are met, severely limiting their applicability (Korn, 1984).

Note that the above remark regarding the limited number of tests of conditional independence applies only if the control variable X is continuous. If X is categorical, with sufficiently many observations per category, it is not difficult to devise a test of conditional independence: for each category, a test of independence can be done, and these tests can be combined in various ways. If all three variables are categorical, log-linear techniques can be used (cf. Lauritzen, 1996).

Summarizing, a flexible array of tests of conditional independence with a continuous control variable appears to be unavailable in the literature, and this paper aims to fill that gap. First, in Section 2, the theoretical difficulties with the testing of conditional independence are investigated and it is shown that with a continuous control variable, the problem is in a well-defined sense harder than the testing of unconditional independence. Then in Section 3 a new testing methodology is presented, using which any test of unconditional independence can be used to construct a test of conditional independence. Evaluation of the test requires the estimation of certain conditional marginal quantiles of Y given X and of Z given X. Although this is about the same amount of work as the estimation of the regression curves in (1), it leads to a much more flexible class of test statistics. In Section 4, the feasibility of the approach is demonstrated by an example with medical data.

2 Degree of testability of (conditional) independence

In this section we first introduce the concept of testability of degree r, which is an extension of Hoeffding's concept of estimability of degree r, and then apply it to the hypotheses of independence and conditional independence. It is shown that independence is testable of degree 2, while conditional independence is not testable of degree r for any r. Finally, it is shown that if the (conditional) marginal distributions are known, both hypotheses are testable of degree 1.

2.1 Testability of degree r

Hoeffding (1948a) defined the concept of estimability of degree r, which we restate as follows:

Definition 1 *For a set of probability measures $\mathcal{P}$, a parameter $\theta : \mathcal{P} \to \mathbf{R}$ is called* estimable of degree r *if r is the smallest number for which there is a function $h : \mathbf{R}^r \to \mathbf{R}$ such that, for all $P \in \mathcal{P}$,*

$$\theta(P) = Eh(X_1, \ldots, X_r)$$

if the X_i are iid according to P.

Related to Hoeffding's concept, we introduce the concept of testability of degree r:

Definition 2 *Let $\mathcal{P}$ be a set of probability measures and $(\mathcal{P}_1, \mathcal{P}_2)$ a partition of $\mathcal{P}$. Then $\mathcal{P}_1$ is called* testable of degree r against the alternative $\mathcal{P}_2$ *if r is the smallest number for which there is an estimable parameter θ of degree r such that*

T1: *For all $P \in \mathcal{P}_1$, $\theta(P) = 0$.*

T2: *There exists a $P \in \mathcal{P}_2$ such that $\theta(P) \neq 0$.*

In the two subsections below we shall see how the testability concept applies to the independence and conditional independence hypotheses, with and without assumptions on the marginal distributions.

2.2 Testability of (conditional) independence without assumptions

Let Ω_1, Ω_2 and Ω_3 be given sets, each containing at least two elements. Suppose X has a distribution on Ω_1 and, for all $x \in \Omega_1$, the pair of random variables (Y, Z) has a joint distribution on $\Omega_2 \times \Omega_3$ given $X = x$. Note that it is implied that (Y, Z) has a marginal distribution on $\Omega_2 \times \Omega_3$.

The random variables Y and Z are *independent*, denoted $Y \perp\!\!\!\perp Z$, if

$$\Pr(Y \in A, Z \in B) = \Pr(Y \in A) \Pr(Z \in B)$$

for all measurable $A \subseteq \Omega_2$ and $B \subseteq \Omega_3$. If independence does not hold, this is denoted as $Y \not\perp\!\!\!\perp Z$.

For ease of exposition we shall, in the sequel, sometimes say "the hypothesis $Y \perp\!\!\!\perp Z$" to refer to the set of probability measures on $\Omega_2 \times \Omega_3$ for which independence holds.

Concerning independence, we have the following theorem:

Theorem 1 *The independence hypothesis $Y \perp\!\!\!\perp Z$ is testable of degree 2 against the alternative $Y \not\perp\!\!\!\perp Z$.*

The next example illustrates the theorem, and is part of its proof, by giving an estimable parameter of degree 2 which is zero under independence but nonzero for certain alternatives.

Example 1 *Let (A_1, A_2) be a partition of Ω_2 and let (B_1, B_2) be a partition of Ω_3. Now consider the function*

$$\begin{aligned} h[(y_1, z_1), (y_2, z_2)] \;=\; & I(y_1 \in A_1, z_1 \in B_1) I(y_2 \in A_2, z_2 \in B_2) \\ & - I(y_1 \in A_1, z_1 \in B_2) I(y_2 \in A_2, z_2 \in B_1) \end{aligned}$$

where I is the indicator function, equalling 1 if its argument is true and 0 otherwise. Then if $Y \perp\!\!\!\perp Z$ and if (Y_1, Z_1) and (Y_2, Z_2) are iid and distributed as (Y, Z), it is straightforward to verify that

$$Eh[(Y_1, Z_1), (Y_2, Z_2)] = 0$$

Furthermore, if (Y', Z') are such that

$$\Pr(Y' \in A_1, Z' \in B_1) = \Pr(Y' \in A_2, Z' \in B_2) = \frac{1}{2}$$

then for iid (Y_i', Z_i') distributed as (Y', Z'),

$$Eh[(Y_1', Z_1'), (Y_2', Z_2')] = \frac{1}{4} \neq 0$$

Hence, the estimable parameter θ of degree 2 based on the function h satisfies the conditions T1 and T2.

Proof of Theorem 1: Let $\mathcal{P}$ be the set of probability measures on $\Omega_2 \times \Omega_3$. Further, let $\mathcal{P}_1 \subseteq \mathcal{P}$ be the set of probability measures satisfying the independence hypothesis and let $\mathcal{P}_2$ be its complement in $\mathcal{P}$. Suppose θ is an estimable parameter of degree 1 which satisfies condition T1 and which is based on a certain function h. Let $P \in \mathcal{P}$ be degenerate satisfying $P(a, b) = 1$ for certain $a \in \Omega_2$ and $b \in \Omega_3$. Then it immediately follows that $P \in \mathcal{P}_1$. Therefore, by condition T1, $\theta(P) = 0$, implying $h(a, b) = 0$. Since a and b were arbitrary, h is zero on its domain and hence $\theta(P) = 0$ for all $P \in \mathcal{P}$. Therefore condition T2 does not hold, and so, since θ was chosen arbitrarily, the independence hypothesis is not testable of degree 1. That it is testable of degree 2 follows from Example 1. □

The random variables Y and Z are *conditionally independent* given X, denoted $Y \perp\!\!\!\perp Z|X$, if

$$\Pr(Y \in A, Z \in B|X = x) = \Pr(Y \in A|X = x) \Pr(Z \in B|X = x)$$

for all $x \in \Omega_1$ and measurable $A \subseteq \Omega_2$ and $B \subseteq \Omega_3$. If conditional independence does not hold, this is denoted as $Y \not\perp\!\!\!\perp Z|X$.

The random variable X is called *continuous* if $P(X = x) = 0$ for all x in the domain of X. Theorem 2 shows that, with a continuous control variable, testing conditional independence is, in a well-defined sense, much harder than testing independence.

Theorem 2 *Under the restriction that X is continuous, there is no r such that the conditional independence hypothesis $Y \perp\!\!\!\perp Z|X$ is testable of degree r against the alternative $Y \not\perp\!\!\!\perp Z|X$.*

Proof of Theorem 2: Let $\mathcal{P}$ be the set of probability measures on $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$ with continuous marginal distribution on Ω_1, i.e.,

$$P(\{x\} \times \Omega_2 \times \Omega_3) = 0$$

for all $P \in \mathcal{P}$ and $x \in \Omega_1$. It is assumed that for all $x \in \Omega_1$ the conditional measures on $\{x\} \times \Omega_2 \times \Omega_3$ are also defined. Let $\mathcal{P}_1$ be the set of probability measures in $\mathcal{P}$ satisfying conditional independence and let $\mathcal{P}_2$ be its complement in $\mathcal{P}$.

Suppose, for certain r, θ is an estimable parameter of degree r which satisfies condition T1 and which is based on a certain function h. For arbitrary $(x_1, y_1, z_1), \ldots, (x_r, y_r, z_r)$ in Ω with $x_i \neq x_j$ if $i \neq j$ let $P \in \mathcal{P}$ be the degenerate distribution satisfying

$$P[(x_1, y_1, z_1), \ldots, (x_r, y_r, z_r)] = 1$$

Then it immediately follows that $P \in \mathcal{P}_1$. Therefore, by condition T1, $\theta(P) = 0$, implying

$$h[(x_1, y_1, z_1), \ldots, (x_r, y_r, z_r)] = 0$$

Since the (x_i, y_i, z_i) were arbitrary, h is zero on its domain except possibly on a set with measure zero (where $x_i = x_j$ for some $i \neq j$). Hence $\theta(P) = 0$ for all $P \in \mathcal{P}$, so condition T2 does not hold for θ. Therefore, since θ was chosen arbitrarily, $\mathcal{P}_1$ is not testable of degree r against $\mathcal{P}_2$. Since r was arbitrary, the theorem is proven. □

An intuition for Theorem 2 is as follows. If X is continuous, then an iid sample $\{(X_i, Y_i, Z_i)\}$ has, with probability one, at most one observed (Y, Z) pair for any value of X. Theorem 1 indicates that at least two pairs would be needed in order to have any 'information' on the conditional dependence for the corresponding value of X.

Summarizing, Theorems 1 and 2 indicate a fundamental difference between the testing of independence and of conditional independence with a continuous control variable.

2.3 Testability of (conditional) independence with assumptions

In this section we show that given assumptions about the (conditional) marginal distributions, (conditional) independence hypotheses are testable of degree 1. This result is especially useful for the conditional independence hypothesis with a continuous control variable, since, by Theorem 2, it is not testable of any degree. However, it also indicates a fundamental difficulty: assumptions *must* be made in order to be able to obtain a test. For the testing of independence this is, of course, not necessary.

A distribution is called degenerate if all its probability mass is concentrated on one point. The next theorem shows the effect of incorporating assumptions about the marginal distributions of Y and Z on the testability of $Y \perp\!\!\!\perp Z$.

Theorem 3 *Suppose Y and Z have given non-degenerate marginal distributions. Then the hypothesis $Y \perp\!\!\!\perp Z$ is testable of degree 1 against the alternative $Y \not\perp\!\!\!\perp Z$.*

To prove the theorem, it suffices to give an example of an estimable parameter of degree 1 satisfying the conditions T1 and T2, as is done next.

Example 2 *If Y and Z have given non-degenerate marginal distributions, there is an $A \subset \Omega_2$ such that*

$$0 < \Pr(Y \in A) < 1$$

and a $B \subset \Omega_3$ such that

$$0 < \Pr(Z \in B) < 1$$

Using the shorthand $p = \Pr(Y \in A)$ and $q = \Pr(Z \in B)$, consider the function

$$h(y, z) = I(y \in A, z \in B) - pq$$

Then

$$\theta = Eh(Y, Z) = \Pr(Y \in A, Z \in B) - pq$$

is zero if $Y \perp\!\!\!\perp Z$ and nonzero if, for example,

$$\Pr(Y \in A, Z \in B) = \min(p, q)$$

Hence, the estimable parameter θ of degree 1 based on the function h satisfies the conditions T1 and T2.

Theorem 2 suggests the hypothesis $Y \perp\!\!\!\perp Z|X$ is 'untestable.' However, as indicated by Theorem 4, appropriate assumptions about the conditional distributions of Y and Z given X render the hypothesis testable.

Theorem 4 *Suppose for all $x \in \Omega_1$, both the marginal distributions of Y and Z given $X = x$ are known and non-degenerate. Then the hypothesis $Y \perp\!\!\!\perp Z|X$ is testable of degree 1 against the alternative $Y \not\perp\!\!\!\perp Z|X$.*

The following example illustrates and proves the theorem for arbitrary Ω_1, Ω_2 and Ω_3.

Example 3 *Suppose, for all x, Y and Z have given non-degenerate marginal distributions given $X = x$. Then for any x there is an $A(x) \subset \Omega_2$ such that*

$$0 < \Pr[Y \in A(x)|X = x] < 1$$

and a $B(x) \subset \Omega_3$ such that

$$0 < \Pr[Z \in B(x)|X = x] < 1$$

Using the shorthand $p(x) = \Pr[Y \in A(x)|X = x]$ and $q(x) = \Pr[Z \in B(x)|X = x]$, consider the function

$$h(x, y, z) = I[y \in A(x), z \in B(x)] - p(x)q(x)$$

Then, for all x,

$$Eh(x, Y, Z) = \Pr[Y \in A(x), Z \in B(x)] - p(x)q(x)$$

is zero if $Y \perp\!\!\!\perp Z|X$ and nonzero if, for example,

$$\Pr(Y \in A(x), Z \in B(x)|X = x) = \min(p(x), q(x)) \tag{2}$$

for all x. Hence,

$$\theta = Eh(X, Y, Z) = \int Eh(x, Y, Z)dP_1(x)$$

is zero if $Y \perp\!\!\!\perp Z|X$ and nonzero if (2) holds. Therefore, the estimable parameter θ of degree 1 based on the function h satisfies the conditions T1 and T2.

3 The partial copula

In this section an automatic conversion procedure is presented by which any test of unconditional independence can be used to obtain a test of conditional independence. For this purpose, the partial copula is introduced, which has the salient feature that independence holds for the partial copula if conditional independence holds for the responses given the control. A kernel estimation method for the partial copula is introduced, which involves estimating conditional marginal distributions of the responses given the control.

3.1 Definition and basic properties

Suppose Y and Z are real-valued random variables with conditional distribution functions

$$F_{2|1}(y|x) = \Pr(Y \leq y|X = x)$$
$$F_{3|1}(z|x) = \Pr(Z \leq z|X = x)$$

A basic property of

$$U = F_{2|1}(Y|X)$$

and

$$V = F_{3|1}(Z|X)$$

is given in the following lemma.

Lemma 1 *Suppose, for all x, $F_{2|1}(y|x)$ is continuous in y and $F_{3|1}(z|x)$ is continuous in z. Then U and V have uniform marginal distributions.*

Proof of Lemma 1: By continuity of $F_{2|1}(y|x)$ in y, and with F_1 the marginal distribution function of X,

$$\begin{aligned} \Pr(U \leq u) &= \Pr(F_{2|1}(Y|X) \leq u) \\ &= \int \Pr(F_{2|1}(Y|x) \leq u) dF_1(x) \\ &= \int u dF_1(x) \\ &= u \end{aligned}$$

i.e., the marginal distribution of U is uniform. The uniformity of the distribution of V is shown analogously. □

The importance of the introduction of U and V lies in the following theorem.

Theorem 5 *Suppose, for all x, $F_{2|1}(y|x)$ is continuous in y and $F_{3|1}(z|x)$ is continuous in z. Then $Y \perp\!\!\!\perp Z|X$ implies $U \perp\!\!\!\perp V$.*

Proof of Theorem 5: By Lemma 1, U and V are uniformly distributed. Hence if $Y \perp\!\!\!\perp Z|X$ the joint distribution function of U and V simplifies as follows:

$$\begin{aligned}
\Pr(U \leq u, V \leq v) &= \Pr(F_{2|1}(Y|X) \leq u, F_{3|1}(Z|X) \leq v) \\
&= \int \Pr(F_{2|1}(Y|x) \leq u, F_{3|1}(Z|X) \leq v) dF_1(x) \\
&= \int \Pr(F_{2|1}(Y|x) \leq u) \Pr(F_{3|1}(Z|X) \leq v) dF_1(x) \\
&= \int uv dF_1(x) \\
&= uv \\
&= \Pr(U \leq u) \Pr(V \leq v)
\end{aligned}$$

This completes the proof. □

For continuous random variables Y and Z with marginal distribution functions F and G, the *copula* of their joint distribution is defined as the joint distribution of $F(Y)$ and $G(Z)$. The copula is said to contain the grade (or rank) association between Y and Z (for an overview, see Nelsen, 1998). For example, Kendall's tau and Spearman's rho are functions of the copula. The following definition gives an extension of the copula concept.

Definition 3 *The joint distribution of U and V is called the* partial copula *of the distribution of Y and Z given X.*

Note that the conditional copula, denoted $C_{23|1}$, is given as

$$C_{23|1}(u, v|x) = \Pr(F_{2|1}(Y|X = x) \leq u, F_{3|1}(Z|X = x) \leq v)$$

and that the partial copula, say G_{23}, is the average conditional copula, given by the formula

$$G_{23}(u, v) = EC_{23|1}(u, v|X) = \int C_{23|1}(u, v|x) dF_1(x) \tag{3}$$

Theorem 5 implies that a test of (marginal) independence of U and V is a test of conditional independence of Y and Z given X. It should be noted

that since $U \perp\!\!\!\perp V$ does not imply $Y \perp\!\!\!\perp Z|X$, a test of the hypothesis $U \perp\!\!\!\perp V$ cannot have power against all alternatives of the hypothesis $Y \perp\!\!\!\perp Z|X$. In particular, this is so for alternatives with interaction, that is, where the association between Y and Z depends on the value of X. We should expect most power against alternatives with a constant conditional copula, i.e., alternatives for which the joint distribution of $(F_{2|1}(Y|x), F_{3|1}(Z|x))$ does not depend on x. A test of independence of U and V can be done by any standard procedure.

An example of the derivation of U and V in a parametric setting is given next.

Example 4 *Suppose the distribution of Y given $X = x$ is exponential with scale parameter $\lambda(x)$, and the distribution of Z given $X = x$ is exponential with scale parameter $\mu(x)$, i.e.,*

$$F_{2|1}(y|x) = 1 - e^{-\lambda(x)y}$$
$$F_{3|1}(z|x) = 1 - e^{-\mu(x)z}$$

Then by Theorem 5,

$$U = 1 - e^{-\lambda(X)Y}$$

and

$$V = 1 - e^{-\mu(X)Z}$$

are independent if $Y \perp\!\!\!\perp Z|X$.

The next subsection discusses the nonparametric estimation of the partial copula.

3.2 Kernel estimation

In practice, the conditional marginal distribution functions $F_{2|1}$ and $F_{3|1}$ are often unknown. Simple kernel estimators are:

$$\hat{F}_{2|1}(y|x) = \frac{\sum_{i=1}^{n} K_2[(x - X_i)/h_2]J(Y_i, y)}{\sum_{i=1}^{n} K_2[(x - X_i)/h_2]} \tag{4}$$

and

$$\hat{F}_{3|1}(y|x) = \frac{\sum_{i=1}^{n} K_3[(x - X_i)/h_3]J(Y_i, y)}{\sum_{i=1}^{n} K_3[(x - X_i)/h_3]} \tag{5}$$

where $h > 0$ is the bandwidth, usually dependent on n, and K the kernel function, which can be a density symmetric around zero and

$$J(x,y) = \begin{cases} 0 & x < y \\ \frac{1}{2} & x = y \\ 1 & x > y \end{cases}$$

A suitable choice for K is often the standard normal distribution.

Consider the new observations (U_i, V_i), given as

$$U_i = \hat{F}_{2|1}(Y_i|X_i) \quad (6)$$

$$V_i = \hat{F}_{3|1}(Z_i|X_i) \quad (7)$$

Now a test of independence of U and V based on the (U_i, V_i) is a test of conditional independence of Y and Z given X based on the (X_i, Y_i, Z_i). An example is given in the next section. A heuristic argument that the (U_i, V_i) may be treated as iid observations for sufficiently large n is as follows. Standard results can be used to show that, if $h_2 n$ and $h_3 n$ go to zero at a sufficiently fast rate, and under suitable (light) regularity conditions, both

$$\left(\sqrt{n}(\hat{F}_{2|1}(y|x) - F_{2|1}(y|x)), \sqrt{n}(\hat{F}_{2|1}(y'|x') - F_{2|1}(y'|x')\right)$$

and

$$\left(\sqrt{n}(\hat{F}_{3|1}(z|x) - F_{3|1}(z|x)), \sqrt{n}(\hat{F}_{3|1}(z'|x') - F_{3|1}(z'|x')\right)$$

have an asymptotic bivariate normal distribution with correlation equal to zero for all $(x, y, z) \neq (x', y', z')$.

4 Example

Table 1 shows data on 35 consecutive patients under treatment for heart failure with the drug digoxin. The data are from Halkin, Sheiner, Peck, and Melmon (1975). Of medical interest is the hypothesis that digoxin clearance is independent of urine flow controlling for creatinine clearance, i.e., $Y \perp\!\!\!\perp Z|X$. Edwards (2000) based his analyses on the partial correlation coefficient (1) assuming linear regression functions g and h. Then (1) reduces to

$$\rho_{YZ|X} = \frac{\rho_{YZ} - \rho_{XY}\rho_{XZ}}{\sqrt{(1-\rho_{XY}^2)(1-\rho_{XZ}^2)}}$$

X	Y	Z	X	Y	Z
19.5	17.5	0.74	66.8	37.5	0.50
24.7	34.8	0.43	72.4	50.1	0.97
26.5	11.4	0.11	80.9	50.2	1.02
31.1	29.3	1.48	82.0	50.0	0.95
31.3	13.9	0.97	82.7	31.8	0.76
31.8	31.6	1.12	87.9	55.4	1.06
34.1	20.7	1.77	101.5	110.6	1.38
36.6	34.1	0.70	105.0	114.4	1.85
42.4	25.0	0.93	110.5	69.3	2.25
42.8	47.4	2.50	114.2	84.8	1.76
44.2	31.8	0.89	117.8	63.9	1.60
49.7	36.1	0.52	122.6	76.1	0.88
51.3	22.7	0.33	127.9	112.8	1.70
55.0	30.7	0.80	135.6	82.2	0.98
55.9	42.5	1.02	136.0	46.8	0.94
61.2	42.4	0.56	153.5	137.7	1.76
63.1	61.1	0.93	201.1	76.1	0.87
63.7	38.2	0.44			

Table 1: Digoxin clearance data. Clearances are given in ml/min/1.73m^2, urine flow in ml/min. Source: Halkin et al. (1975).
Note: X= Creatinine clearance, Y= digoxin clearance, Z= urine flow.

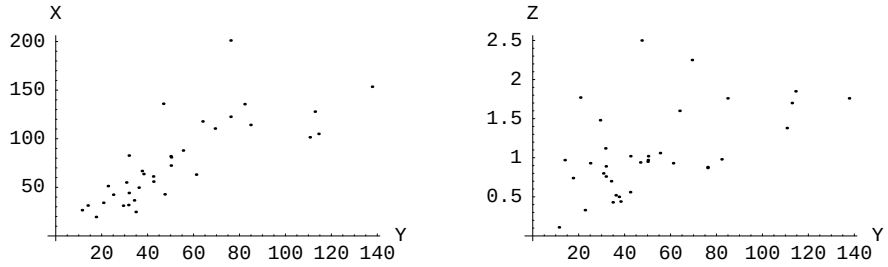

Figure 1: Scatterplots of the (Y_i, X_i) and the (Z_i, X_i)

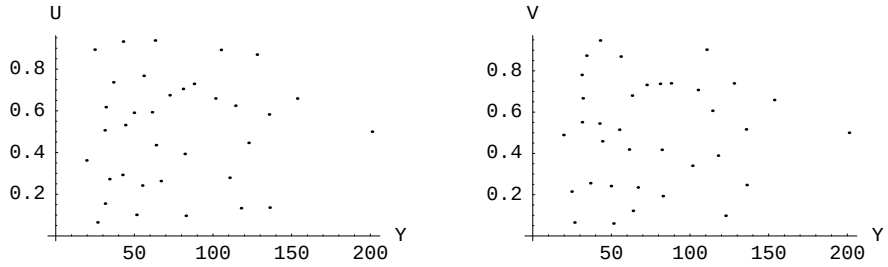

Figure 2: Scatterplots of the (X_i, U_i) and (X_i, V_i)

A visual inspection of the marginal scatter plots, shown in Figure 1, indicates that the linearity assumption appears to be reasonable. However, if the linear model is wrong, a biased estimate of the partial correlation may result, resulting in a biased test for conditional independence. One alternative approach is to perform a nonparametric regression and correlate the errors. Potential disadvantages of this procedure are its sensitivity to outliers, and its sensitivity to only a limited number of alternatives.

Testing conditional independence using the estimated partial copula potentially overcomes these disadvantages. We estimated the partial copula using formulas (4) and (5) with a standard normal kernel and bandwidth 10. This means that 95% of the weight is formed by approximately 7 of the 35 observations. (A justification of this bandwidth is given later.) The new observations (U_i, V_i) are given by Formulas (6) and (7). In Figure 2 scatter plots are given of the pairs (X_i, U_i) and (X_i, V_i), respectively. A visual inspection of both pictures seems to confirm that the effect of X has been removed, that is, independence seems to hold.

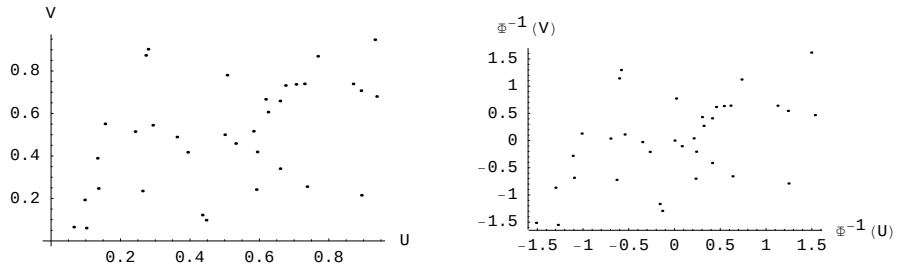

Figure 3: Scatterplots of the (U_i, V_i) and the $(\Phi^{-1}(U_i), \Phi^{-1}(V_i))$

A scatterplot of the estimated partial copula is given in Figure 3 on the left hand side. Since uniform marginals are arbitrary, we have transformed the marginals to standard normals in the picture on the right hand side. In both pictures there appears to be some dependence present in the data. We shall test independence by testing the significance of four statistics, which are described next. First, we estimate the correlation for the partial copula as

$$r_{UV} = \frac{12}{n} \sum_{i=1}^{n} (U_i - 1/2)(V_i - 1/2)$$

The 12 appears because $1/12$ is the variance of both U and V, which are uniformly distributed. Note that r_{UV} also estimates the Spearman rank correlation, since U and V have uniform marginals. The "normal correlation," that is, the correlation based on transforming the marginals to a normal distribution, is estimated as

$$r^*_{UV} = \frac{1}{n} \sum_{i=1}^{n} (\Phi^{-1}(U_i) - 1/2)(\Phi^{-1}(V_i) - 1/2)$$

where Φ is the distribution function of the standard normal distribution. The estimated values and p-values are given in Table 2. The p-values were calculated using a bootstrap approximation of the permutation test.

An important potential advantage of using r and r^* compared to a nonparametric estimate of the partial correlation (1) is the robustness of r and r^*, that is, their insensitivity to outliers.

Let G_{23} be the distribution function of the partial copula, i.e., the joint distribution function of U and V as given by (3), and let G_2 and G_3 be the corresponding marginal distribution functions. A variant of Hoeffding's

coefficient (Hoeffding, 1948b) measuring the dependence between U and V is

$$H_{UV} = \int \left[G_{23}(y,z) - G_2(y)G_3(z)\right]^2 dydz$$

To obtain Hoeffding's coefficient, $dydz$ has to be replaced by $dG_{23}(y,z)$. Both Hoeffding's coefficient and H_{UV} share the property that they are non-negative and equal to zero if and only if U and V are independent. A convenient way to estimate H_{UV} is by using a formula developed by (Bergsma, 2004), given as

$$H_{UV} = E\left(|U_1' - U_2'| - |U_1' - U_3'|\right)\left(|V_1' - V_2'| - |V_2' - V_4'|\right)$$

where $(U_1', V_1'), \ldots, (U_4', V_4')$ are iid and distributed as (U, V). Hence, the unbiased U-statistic estimator of H_{UV} is given as

$$\hat{H}_{UV} = \begin{pmatrix} n \\ 4 \end{pmatrix}^{-1} \sum \left(|U_i - U_j| - |U_i - U_k|\right)\left(|V_i - V_j| - |V_j - V_l|\right)$$

where the sum is taken over all i, j, k, l unequal. Like the correlation, H_{UV} is also evaluated for both the partial copula with uniform and with normal marginals, the latter denoted as H_{UV}^*. In Table 2, estimated values of H and H^* are given, together with p-values for the hypothesis that they are zero. The p-values were calculated by a bootstrap approximation of the permutation test.

The reason we have chosen H_{UV} rather than Hoeffding's coefficient is to avoid unnecessary discretization. Hoeffding's test is based on the ranks of the observations (like Kendall's tau and Spearman's rho). One of the reasons to use a rank test is to deal with outliers and to control the marginals by making them uniform. In the present case, the (theoretical) distribution of U and V is already uniform, thereby making a further ranking of the (U_i, V_i) unnecessary, and this would only cause unnecessary discretization of the data. For the same reason, we have not considered Kendall's tau and Spearman's rho.

An important potential advantage of using H and H^* to test for independence between U and V is that they yield asymptotic power against all alternatives with dependence. This is something that cannot be achieved with the partial correlation coefficient (1).

A justification of the choice of bandwidth has not been given. As the bandwidth approaches zero, the (U_i, V_i) converge to $(\frac{1}{2}, \frac{1}{2})$. Thus, all of the estimators above converge to zero. The bandwidth should be chosen large

Coefficient	Estimated value	p-value
ρ_{UV}	.372	.0062
ρ^*_{UV}	.314	.0031
H_{UV}	.656	.0032
H^*_{UV}	.651	.0077

Table 2: Estimated values of some coefficients and their bootstrap p-values. Based on the data of Table 1

enough that at least one point in the neighborhood has sufficient weight. For crossvalidation purposes, the estimated parameters are plotted as a function of the bandwidth h in Figure 4. For all coefficients, the method appears to break down when the bandwidth $h < 6$, and the value $h = 10$ appears to be reasonable. Note that a crossvalidation should always be performed.

5 Summary

In Section 2, the concept of testability of degree r, which is related to Hoeffding's concept of estimability of degree r, was defined. It was shown that independence is testable of degree 2 while, for a continuous control variable, conditional independence is not testable of any degree. Then it was shown that, if the conditional marginal distributions of responses given control are known, the conditional independence hypothesis is testable of degree 1. Hence, the testing of conditional independence is more difficult than the testing of unconditional independence, in the sense that the former requires assessment of the conditional marginal distributions.

In Section 3, the results of Section 2 were used to derive a practical testing procedure, which makes it possible to convert an arbitrary test of independence to a test of conditional independence. This was done by introducing the partial copula which is a function of the original trivariate distribution. Like the copula, the partial copula is a bivariate distribution with uniform marginals. Additionally, it has the property of satisfying independence if conditional independence holds for the original distribution. Hence, any test of independence applied to the partial copula is a test of conditional independence applied to the original distribution. Estimation of the partial copula requires the conditional marginal distributions of responses given control which are usually unknown in practice, and a kernel estimator was proposed. Thus, a wide range of tests is obtained whose evaluation is

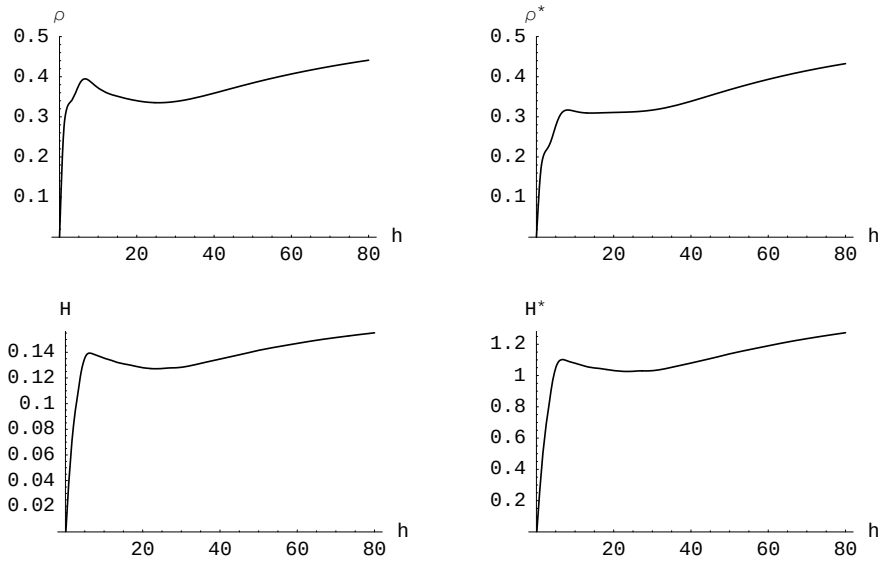

Figure 4: Estimated values of several coefficients in the estimated partial copula as a function of the bandwidth h for crossvalidation

no more difficult than nonparametric estimation of the partial correlation coefficient.

The method was illustrated by an example in Section 4. Two tests related to the rank correlation were described. These directly compete with the partial correlation, but have the advantage of robustness. Two other tests related to a test of Hoeffding (1948b) were also described, and these have the advantage of asymptotic power against a broad range of alternatives.

References

Bergsma, W. P. (2004). Absolute correlation. *Unpublished manuscript.*

Dawid, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc. Ser. B*, *41*, 1-31.

Edwards, D. (2000). *Introduction to graphical modelling.* New York: Springer verlag.

Goodman, L. A. (1959). Partial tests for partial tau. *Biometrika*, *46*, 425-432.

Goodman, L. A., & Grunfeld, Y. (1961). Some nonparametric tests for comovements between time series. *Journal of the American Statistical Association*, *56*, 11-26.

Gripenberg, G. (1992). Partial rank correlations. *Journal of the American Statistical Association*, *87*, 546-551.

Halkin, H., Sheiner, L. B., Peck, C. C., & Melmon, K. L. (1975). Determinants of the renal clearance of digoxin. *Clin. Pharmocol. Theor.*, *17*, 385-394.

Hoeffding, W. (1948a). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, *19*, 293-325.

Hoeffding, W. (1948b). A non-parametric test of independence. *Annals of Mathematical Statistics*, *19*, 546-557.

Kendall, M. G. (1942). Partial rank correlation. *Biometrika*, *32*, 277-283.

Korn, E. L. (1984). The ranges of limiting values of some partial correlations under conditional independence. *The American Statistician*, *38*, 61-62.

Lauritzen, S. L. (1996). *Graphical models.* Oxford: Clarendon Press.

Nelsen, R. B. (1998). *An introduction to copulas.* New York: Springer.

Schweizer, B., & Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *Annals of Statistics*, *9*, 879-885.