Nonparametric testing of conditional independence

Wicher Bergsma

November 14, 2012

Abstract

Conditional independence of Y and Z given X holds if and only if the following two conditions hold:

- CI1: The expected conditional covariance between arbitrary functions of Y and Z given X is zero (where the expectation is taken with respect to X).
- CI2: The conditional covariance between arbitrary functions of Y and Z given X does not depend on X.

Based on this decomposition, we propose a simple method to test conditional independence for the case that the variables are Euclidean and the distribution of Y given X is continuous. The latter implies that conditional probability integral transforms can be used to convert Y to a variable independent of X.

For example, if $Y = (Y_1, Y_2)' \in \mathbb{R}^2$, then

$$Y^* = (Y_1^*, Y_2^*)' = [F_{Y_1|X}(Y|X), F_{Y_2|XY_1}(Y_2|X, Y_1)]'$$

is independent of X.

This transformation greatly simplifies conditional independence testing. In particular, condition CI1 above is equivalent to marginal independence between Y^* and Z, and CI2 is equivalent to absence of three variable (Lancaster) interaction among X, Y^*, and Z. The first condition can hence be tested using any test of marginal independence. For Lancaster interaction, no tests are available as far as we are aware, and so we propose two new ones. In both cases, exact p-values can be computed using the permutation test.

In practice the conditional probability integral transforms will need to be estimated, which can be done using a simple Nadaraya-Watson estimator. We show that using appropriate estimators this estimation has no effect on asymptotic distributions of the relevant test statistics.

The advantages of our method compared to recently published alternative methods are outlined.