The Annals of Applied Statistics
2009, Vol. 3, No. 4, 1279–1281
DOI: 10.1214/00-AOAS312C
Main article DOI: 10.1214/09-AOAS312

DISCUSSION OF: BROWNIAN DISTANCE COVARIANCE

BY LESLIE COPE

Johns Hopkins University

I read *Distance Covariance*, by Drs. Szekely and Rizzo, with great interest. This is an elegant contribution to statistical theory; the three-way equivalence between a weighted expectation of the difference between Brownian covariance and two very different formulations of $\mathcal{V}^2$ is very attractive, and together with the examples make a strong case for distance covariance.

But like many statisticians, I spend much of my working life analyzing genomic data sets and so am interested in how distance covariance and correlation might be used in high dimensional data with relatively small sample sizes. In these applications it is often more important to characterize the relationships between genes than to formally test for independence. And the Pearson correlation coefficient, complimented by a well-developed and widely-used theory of linear models and matrix methods, is highly applicable on such data sets. The restriction to linear relationships between variables is arguably even an advantage; while Pearson's correlation may not capture all dependencies, we know a great deal about the interpretation of results from its application.

It is, of course, not possible to settle the question here, but some preliminary thoughts follow on the potential utility of distance covariance, and particularly the scaled distance correlation, in this setting.

Using the author's notation, if (X, Y) is a pair of random variables (vectors) and $(\mathbf{X}, \mathbf{Y})$ a sample drawn from the joint distribution, the dependence statistics A_{kl} and B_{kl} are centered, interpoint distance matrices for $\mathbf{X}$ and $\mathbf{Y}$ respectively, and $\mathcal{V}^2(\mathbf{X}, \mathbf{Y})$ is the mean product moment of the entries in these two matrices. Thus, the empirical distance covariance is a cross-variable covariance of within-variable interpoint distances, and the distance correlation is the same, appropriately scaled. In practice, this is similar to the *correlation of correlations* used by Lee et al. (2003) and Parmigiani et al. (2004) to quantify the reproducibility of results obtained on different microarray platforms or from independent gene expression studies, but is more general, since it can be applied even to two, scalar-valued random variables, and because of its potential to capture nonlinear as well as linear dependencies.

This representation of the distance correlation offers some intuition into the characteristics of the statistic. It is reflected in Theorem 4(iii), stating that $R(\mathbf{X}, \mathbf{Y}) = 1$ only if $\mathbf{Y}$ can be obtained from $\mathbf{X}$ by orthogonal transformation, since these rigid transformations preserve interpoint distances up to a scaling factor. It explains the ability, demonstrated in the first few examples, to capture non-monotone relationships between two variables; samples with similar wavelength