The Annals of Applied Statistics
2009, Vol. 3, No. 4, 1303–1308
DOI: 10.1214/09-AOAS312REJ
Main article DOI: 10.1214/09-AOAS312

REJOINDER: BROWNIAN DISTANCE COVARIANCE

BY GÁBOR J. SZÉKELY AND MARIA L. RIZZO

Bowling Green State University, Hungarian Academy of Sciences and Bowling Green State University

First of all we want to thank the editor, Michael Newton, for leading the review and discussion of our work.

We also want to thank all discussants for their interesting comments. Some of them are in fact short research papers that expand the scope of Brownian Distance Covariance. Many of the comments emphasized the existence of some competing notions like maximal correlation; others requested further clarifications or suggested several extensions. Most of the comments were theoretical in nature. We do hope that once our new correlation is applied in practice we shall receive comments from the broader community of applied statisticians. Let us now continue with replies to the discussions collectively by grouping the topics.

1. Unbiased distance covariance.

In the discussion Cope observes that the distance dependence statistics are biased, and that this bias may be substantial and increasing with dimension. As he points out, in genomic studies, high dimension and small sample sizes are common.

In this section we present an unbiased estimator of the population distance covariance, define a corrected distance correlation statistic C_n, and propose a simple decision rule for the high dimension, small sample size situation.

The expected value of $\mathcal{V}_n^2$ is $E[\mathcal{V}_n^2(\mathbf{X},\mathbf{Y})] = \frac{n-1}{n^2}[(n-2)\mathcal{V}^2(X,Y) + \mu_1\mu_2]$, where $\mu_1 = E|X-X'|$ and $\mu_2 = E|Y-Y'|$. An unbiased estimator of $\mathcal{V}^2(X,Y)$ can be defined as follows.

DEFINITION 1.

$$U_n(\mathbf{X},\mathbf{Y}) = \frac{n^2}{(n-1)(n-2)}\left[\mathcal{V}_n^2(\mathbf{X},\mathbf{Y}) - \frac{T_2}{n-1}\right], \qquad n \geq 3,$$

where T_2 is the statistic defined in Theorem 1.

We proposed to normalize the V-statistic $n\mathcal{V}_n^2$ by dividing by T_2. Under independence, it follows from Corollary 2(i) that

$$\frac{nU_n}{T_2} = \frac{n^2}{(n-1)(n-2)}\left[\frac{n\mathcal{V}_n^2}{T_2} - \frac{n}{n-1}\right] \xrightarrow{\mathscr{D}} \sum_{k=0}^{\infty}\lambda_k(Z_k^2-1) \quad \text{as } n\to\infty,$$

which is the limiting distribution of the corresponding U-statistic.