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.

A modified distance correlation statistic C_n can be defined by substituting in the original definition of $\mathcal{R}_n^2$ the unbiased estimators U_n. It can be shown that $U_n(\mathbf{X}, \mathbf{X}) \geq 0$ for $n \geq 3$, so that $U_n(\mathbf{X})U_n(\mathbf{Y}) > 0$ whenever $\mathcal{V}_n^2(\mathbf{X})\mathcal{V}_n^2(\mathbf{Y}) > 0$, $n \geq 3$.

DEFINITION 2.   The corrected distance correlation for sample sizes $n \geq 3$ is

$$C_n(\mathbf{X}, \mathbf{Y}) = \begin{cases} \dfrac{U_n(\mathbf{X}, \mathbf{Y})}{\sqrt{U_n(\mathbf{X})U_n(\mathbf{Y})}}, & U_n(\mathbf{X})U_n(\mathbf{Y}) > 0; \\ 0, & \text{otherwise.} \end{cases}$$

If $n = 1$ or $n = 2$ define $C_n = 1$.

If X and Y are independent, $(p + q)/n$ is large and n is moderately large, one can compare nC_n with percentiles of a Normal$(0, \sigma^2 = 2)$ distribution, under very general conditions on the distributions of X and Y.

2. Other measures of dependence, old and new.
Bickel and Xu mentioned canonical correlation ρ, rank correlation r and Rényi correlation R. Of these, only R is the one which vanishes if and only if X and Y are independent. A big advantage of dCor vs R is that dCor is much easier to compute. In the discussion there is a method to approximate Rényi's R, but frankly we do not think that the simplicity of computing or even approximating R is comparable to the simplicity of computing Pearson correlation. Part of the reason is that there is no explicit formula for computing R in general. On the other hand, we have an explicit formula to compute dCov, and practitioners or applied statisticians should find it easy to use.

For the first named author it was heartwarming to see several references to Rényi because Rényi was his first advisor and mentor. In his 1959 paper, Rényi [5] characterized R with seven "natural postulates." His last postulate is that the dependence measure equals the absolute value of Pearson correlation for bivariate normal distributions. This axiom does not hold in our case, although dCor is a deterministic function of Pearson correlation. It would be nice to extend Rényi's theorem and prove a joint characterization of R and dCor.

Bickel and Xu remind us that "if $R = 1$ then then there exist nontrivial functions f and g such that $P(f(X) = g(Y)) = 1 \ldots$." However, the following example suggests that this is not necessarily a desirable property. Consider random variables $X = \sin kU$ and $Y = \sin mU$, where U is uniformly distributed on $(0, 2\pi)$, and k, m are distinct positive integers. Their Pearson correlation is 0, yet for Chebyshev polynomials $\{T_k\}$, we have

$$T_k(\cos 2mU) = T_m(\cos 2kU)$$
$$= T_k(1 - 2Y^2) = T_m(1 - 2X^2).$$

Thus, $R = 1$ even though in many cases X and Y are heuristically quite unrelated: neither f nor g is invertible, Y is not a function of X and vice versa; exceptions are when m is an odd multiple of k. Our simulations suggest that $0 < \mathrm{dCor} < 1/3$ for the examples above, and reaches its maximum when $m = 3k$.

Because X and Y are not independent, it is not surprising that the CLT does not hold for

$$S_n = \sin U + \sin 2U + \cdots + \sin nU.$$

Nevertheless, it can be surprising that S_n tends to C/2 in distribution, as $n \to \infty$, where C is a standard Cauchy random variable. (It is not a misprint that we did not divide by $\sqrt{n}$; here we do not need any kind of normalization.) For the proof of this result and generalizations to other "trigonometric coins," other orthogonal series, and finite Fourier series, see "Trigonometric Coins" [8]. The general infinite Fourier series case is an open problem. One of the advantages of dCov is that in terms of $\mathrm{dCov} = 0$ type conditions we can prove general CLTs for strongly stationary series (Székely and Bakirov [7]).

Further dependence measures can be found in the discussion of Gretton, Fukuzimu and Sriperumbudur. We recognize the theoretical importance of RKHS-based dependence measures, but they do not look as simple as our distance covariance, and they do not seem to be formal extensions of Brownian distance covariance because our weight function (2.4) is not integrable.

3. Generalizations to metric spaces.

One can easily extend the definition of Brownian distance covariance via formula (2.8) to all metric spaces; all we need is to replace the Euclidean distances between observations with their metric distances. Thus in principle we can measure the dependence between two samples where the sample elements come from two arbitrary metric spaces. In order to prove counterparts of our theorems, we need further restrictions. One of the possible approaches is to try to represent the abstract samples in finite dimensional Euclidean spaces such that the distances a_{kl}, b_{kl} become interpoint distances in these Euclidean spaces. Necessary and sufficient conditions are established in the multidimensional scaling literature (see, e.g., Mardia, Kent and Bibby [3], Chapter 14). When such a representation is possible, many theorems in this paper can be extended to measuring and testing independence of random vectors that take values in abstract metric spaces. For example, the metric space extension is applicable for testing independence of categorical data. They are not in Euclidean spaces, but their association can be used as a distance.

A very important area of applications is how to measure the dependence of stochastic processes. In this respect, infinite dimensional extensions of our paper are crucial, so we commend the discussion of Kosorok. Because of his work we now have an extension of our theorems to certain Hilbert spaces.

4. Invariance.   Our test statistic is scale invariant and also rotation invariant. Cramér–von Mises type test statistics, mentioned, for example, in Rémillard's discussion Section 2, are not rotation invariant. This is a major problem if one wants to extend the measure to metric spaces. Let us emphasize that our test *procedure* is invariant with respect to marginal distributions, even though the test statistic is not. On the other hand, it is true that we can easily make our dependence measure even more invariant (invariant with respect to the marginals and with respect to monotone transformations) if we apply the transformations suggested in Section 1 of Rémillard. The negative side of this is that we might lose power, especially if the sample size is small.

Rémillard asked if certain dependence measures can be written in our form. The general answer is no, because the well-known measures such as Kendall's tau and many other rank based measures do not characterize independence, or the statistics are not rotation invariant (e.g., Cramér–von Mises), or like maximal correlation they do not have an explicit computing formula, or may not be defined for arbitrary dimension (e.g., Feuerverger's measure [2]).

Invariance with respect to monotone transformations in one dimension suggests rank type tests such as Feuerverger [2], but they have the disadvantage of being one-dimensional. We can also eliminate all kinds of moment conditions by transforming X and Y to bounded random variables first and then compute their distance covariance, but then there is an arbitrariness in choosing these bounded functions. In one dimension the rank is a natural choice. Section 2 of Rémillard's discussion proposes a natural rank based transformation for the multivariate case.

5. Applications.   Genovese asks about the generality or required conditions for the test of nonlinearity, Example 6. The application of dCov to testing for nonlinearity requires only the linear model $Y = X\beta + \varepsilon$ can be estimated, and that observations $(\mathbf{X}, \hat{\varepsilon})$ are i.i.d. The existence of first moments is implicit in the linear model specification. Distance covariance is defined in arbitrary dimension, so the procedure can be applied to models with a multivariate response. This expands the scope of the test, because models can often be specified with a multivariate response and i.i.d. errors.

The extension of distance covariance methods to non-i.i.d. samples would be very important for applications; see, e.g., Rémillard's discussion Section 3 on the application to time series: *Serial Brownian Distance Correlation*. We agree with Rémillard that "there are still many interesting avenues" to explore in this context.

6. Simplicity/complexity.   Our formula (2.8) to compute dCov is not only simple, it has an obvious formal similarity to Pearson product moment covariance, except that we need to average n^2 products. Genovese comments that the $O(n^2)$ computational complexity of $\mathcal{R}_n$ or $\mathcal{V}_n$ can be burdensome for very large n. However, the simplicity of the computing formula (2.8) in terms of products $A_{kl}B_{kl}$

provides economies of reusable computations. The distances need only be computed once in the permutation test implementation, as the permutation of sample indices of **Y** corresponds to permutations of indices of B_{kl}, for example.

If we compare the complexity of our statistic (2.8) to the complexity of other measures of dependence (including, e.g., RKHS-based methods suggested by the discussants Gretton et al., or our own measure proposed in Bakirov, Rizzo and Székely [1]), then the superiority of Brownian distance covariance is clear. On top of that, one can compute dCov even if the X sample and the Y sample are in completely different metric spaces, because it is not necessary to add or multiply the sample elements; we need only operations on their real valued distances. This is a significant advantage if we want to measure the dependence of apples and oranges, even infinite dimensional ones.

7. Distance covariance vs product-moment covariance and how to teach them. After noticing that Pearson and distance covariance are two different special cases of a general notion of covariance with respect to stochastic processes, we have not explored the boundaries of this generalization. We focused on the two most natural and simplest cases: Brownian covariance and Pearson covariance. Feuerverger raises some interesting questions in this direction at the end of his discussion. Rémillard also raises some questions on the role of stochastic processes U, V. Genovese's discussion sheds some light on these questions. Although we have not yet explored the frontiers of these extensions, these questions and the research of Genovese on this topic are indeed interesting.

For more than a century Pearson correlation has dominated the world of measuring dependence. Even though we know that for nonnormal distributions, product-moment correlation does not characterize independence (does not really measure what we want) for reasons of simplicity, perhaps, it is the first and sometimes the only measure of dependence that students may see. Here Genovese raises a good pedagogical question: should distance correlation be introduced in our teaching at an introductory level? Indeed, we agree that the idea of distance correlation is understandable even at the undergraduate level (without proofs), and one could then continue with product-moment correlation for normal distributions obtained with exponent $\alpha = 2$.

8. Final comments. Our test of independence is implemented in R as part of the "energy" package [4, 6]. The explanation of this cover name is that Newton's potential energy is a function of the Euclidean distances between objects in a gravitational space. In *energy statistics* the "objects" are the elements of the statistical sample, and the statistics are functions of the Euclidean distances between the sample elements. These statistics, the statistical potential energies, govern the cosmos of our paper.

REFERENCES

[1] BAKIROV, N. K., RIZZO, M. L. and SZÉKELY, G. J. (2006). A multivariate nonparametric test of independence. *J. Multivariate Anal.* **93** 1742–1756. MR2298886

[2] FEUERVERGER, A. (1993). A consistent test for bivariate dependence. *Int. Statist. Rev.* **61** 419–433.

[3] MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*. Academic Press, London. MR0560319

[4] R DEVELOPMENT CORE TEAM (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available at http://www.R-project.org.

[5] RÉNYI, A. (1959). On measures of dependence. *Acta Math. Acad. Sci. Hungary* **10** 441–451. MR0115203

[6] RIZZO, M. L. and SZÉKELY, G. J. (2008). energy: E-statistics (energy statistics). R package version 1.1-0. Available at http://cran.case.edu/web/packages/energy/.

[7] SZÉKELY, G. J. and BAKIROV, N. K. (2008). Brownian covariance and CLT for stationary sequences. Technical Report No. 08-01. Dept. Mathematics & Statistics, Bowling Green State University.

[8] SZÉKELY, G. J., MÓRI, T. F., PHADKE, V. and ZIRBEL, C. (2008). Trigonometric coins. Unpublished manuscript.

DEPARTMENT OF MATHEMATICS AND STATISTICS
BOWLING GREEN STATE UNIVERSITY
BOWLING GREEN, OHIO 43403
USA
AND
RÉNYI INSTITUTE OF MATHEMATICS
HUNGARIAN ACADEMY OF SCIENCES
HUNGARY
E-MAIL: gszekely@nsf.gov

DEPARTMENT OF MATHEMATICS AND STATISTICS
BOWLING GREEN STATE UNIVERSITY
BOWLING GREEN, OHIO 43403
USA
E-MAIL: mrizzo@bgsu.edu