Submitted to the Annals of Probability

ERRATA TO "DISTANCE COVARIANCE IN METRIC SPACES"

By Russell Lyons*

We correct several statements and proofs in our paper, *Ann. Probab.* **41**, no. 5 (2013), 3284–3305.

We are grateful to Martin Emil Jakobsen for asking us about several gaps in our proofs. Most of the mistakes and gaps below were noted first by Jakobsen, whom we thank also for his help in checking these errata. His 2017 master's thesis Jakobsen (2017) contains more details of some of our proofs, as well as some extensions.

(i) In general, the metric spaces $(\mathscr{X}, d)$ should be assumed separable: separability ensures that $(x, y) \mapsto d(x, y)$ is measurable with respect to the product Borel σ-field.

(ii) The display after (2.1) should have $d_\mu(x, x')^2$ in place of $d_\mu(x, x')$.

(iii) The proof of Proposition 2.6 is not correct; we do not know whether the statement is correct with only the assumption of finite first moments. The proof does give a.s. convergence to $\operatorname{dcov}(\theta)$ of the U-statistics for the kernel h given in that proof.

(Jakobsen, 2017, Theorem 5.5) implies that if the marginals of θ have finite 5/3-moments, then the convergence of V-statistics as stated in Proposition 2.6 is correct.

On the other hand, if $\mathscr{X}$ and $\mathscr{Y}$ have negative type, then the statement of Proposition 2.6 is easily proved as follows: The strong law of large numbers in Hilbert space tells us that if μ and ν denote the marginals of θ, and μ_n, ν_n the corresponding marginals of θ_n, then $\beta_{\phi,\psi}(\theta_n - \mu_n \times \nu_n) \to \beta_{\phi,\psi}(\theta - \mu \times \nu)$ a.s. Thus, the result follows from Proposition 3.7.

(iv) The conclusion of Theorem 2.7 is not quite correct, but should be changed to $n \operatorname{dcov}(\theta_n) \Rightarrow \sum_i \lambda_i \big(Z_i^2 - 1\big) + D(\mu)D(\nu)$, omitting the claim

*Research partially supported by NSF grant DMS-1612363

MSC 2010 subject classifications: Primary 62H20, 62G20, 51K99; secondary 62H15, 30L05

Keywords and phrases: Negative type, hypothesis testing, independence, distance correlation, Brownian covariance

that $\sum_i \lambda_i = D(\mu)D(\nu)$. The difficulty with the proof is that the operator in question need not be trace class. Also, the proof of the first paragraph is not quite right; it should read as follows:

The triangle inequality gives that

$$|f(z_1, z_2, z_3, z_4)| \le 2 \min\big\{d(z_2, z_3), d(z_1, z_4)\big\}\,,$$

whence

$$f(z_1, z_2, z_3, z_4)^2 \le 4\, d(z_1, z_4) \cdot d(z_2, z_3)\,,$$

which shows that $f(X^1, X^2, X^3, X^4)$ has finite second moment. Therefore, when $\theta = \mu \times \nu$, $h\big((X^1, Y^1), \ldots, (X^6, Y^6)\big)$ has finite second moment.

One can also show that n times the U-statistics for h converge in distribution to $\sum_i \lambda_i\big(Z_i^2 - 1\big)$. See (Jakobsen, 2017, Theorem 5.10) for details.

We now present an example where the operator of Theorem 2.7 is *not* trace class. Given a probability measure μ on a metric space $\mathscr{X}$ with finite first moment, write T_μ for the Hilbert–Schmidt operator that sends $F \in L^2(\mu)$ to the function

$$x \mapsto \int_{\mathscr{X}} d_\mu(x, x') F(x')\, d\mu(x')\,.$$

Since $\theta = \mu \times \nu$, the operator of Theorem 2.7 is the tensor product $T_\mu \otimes T_\nu$. The eigenvalues of the tensor product are the products of the eigenvalues, whence the tensor product is trace class iff each of the terms is. We will take $\mathscr{X} = \mathscr{Y}$ to be constructed out of Paley graphs. Recall that for a prime q that is congruent to 1 modulo 4, the **Paley graph** G_q is a Cayley graph of the group $\mathbb{Z}_q$ with respect to the generators $S := \{k^2\,;\ k \ne 0\}$; that is, G_q has vertex set $\{0, 1, \ldots, q-1\}$ and j, k are joined by an edge when $j - k$ is a nonzero square modulo q. Write $q = 4t + 1$. The Paley graph G_q is **strongly regular** with parameters $(q, 2t, t-1, t)$, meaning that the number of vertices is q, the degree is $2t$, each pair of adjacent vertices has $t - 1$ common neighbors, and each pair of nonadjacent vertices has t common neighbors. In particular, all graph distances are 0, 1, or 2. The adjacency matrix has eigenvalues $2t$ (with multiplicity 1) and $(-1 \pm \sqrt{q})/2$ (with multiplicity $2t$ each). For these facts, see (Brouwer and Haemers, 2012, Proposition 9.1.1 and Theorem 9.1.3). It follows easily from this that the eigenvalues of the distance matrix of G_q are $6t$ and $(-5 \pm \sqrt{q})/2$, with the corresponding multiplicities. Let μ_q denote the uniform measure on the vertices of G_q. The preceding facts imply that the trace norm of the operator T_{μ_q} is asymptotic to $\sqrt{q}/2$ as $q \to \infty$. Choose a sequence of primes $q_n \equiv 1 \pmod 4$ so that $\sum_n 1/\sqrt{q_n} = 1/c < \infty$. Now let $\mathscr{X}$ be the graph formed from the disjoint union of all G_{q_n}, with an edge added between each pair of vertices that belong to different G_{q_n}. Let μ be the probability measure that puts mass $c/q_n^{3/2}$ on each vertex of G_{q_n}. The eigenvectors of T_{μ_q} are constant or are orthogonal to the constants. Therefore, a short calculation shows that for any eigenvector $v \perp \mathbf{1}$ of T_{μ_q} with eigenvalue λ, the function

$$f_n(x) := \begin{cases} v(x) & \text{if } x \in G_{q_n}, \\ 0 & \text{otherwise} \end{cases}$$

is an eigenvector of T_μ with eigenvalue $\lambda \cdot c/\sqrt{q_n}$. Therefore, T_μ is not trace class, whence neither is the operator $T_\mu \otimes T_\mu$ of Theorem 2.7.

(v) The change in (iv) leads to changing (2.5) of Corollary 2.8 to

$$\frac{n \operatorname{dcov}(\theta_n)}{D(\mu_n)D(\nu_n)} \Rightarrow \frac{\sum_i \lambda_i \big(Z_i^2 - 1\big)}{D(\mu)D(\nu)} + 1\,.$$

Similarly, the change in (iii) leads us to require stronger hypotheses in order to apply Proposition 2.6 if $\operatorname{dcov}(\theta) \neq 0$.

(vi) On the other hand, if $\mathscr{X}$ and $\mathscr{Y}$ have negative type, then the original conclusions of Theorem 2.7 and Corollary 2.8 all hold. It suffices to show then that the operator in question is trace class. To see that it is, we actually give another proof of the final part of Theorem 2.7.

We suppose that $\theta = \mu \times \nu$. Let the embeddings ϕ and ψ witness the negative type of $\mathscr{X}$ and $\mathscr{Y}$. For vectors $\xi, \alpha_1, \alpha_2 \in H \otimes H$, write

$$M_{\alpha_1,\alpha_2}(\xi) := \langle \alpha_1, \xi \rangle \langle \xi, \alpha_2 \rangle\,.$$

We claim that $n \operatorname{dcov}(\theta_n) \Rightarrow 4\|\zeta\|^2$, where ζ is a centered Gaussian $H \otimes H$-valued random variable with covariance

$$\big(\alpha_1, \alpha_2\big) \mapsto \int M_{\alpha_1,\alpha_2}\big([\phi(x) - \beta_\phi(\mu)] \otimes [\psi(y) - \beta_\psi(\nu)]\big)\, d\theta(x,y)$$

and

$$\mathbf{E}\big[4\|\zeta\|^2\big] = D(\mu)D(\nu)\,.$$

Indeed, write $\widehat{\phi} := \phi - \beta_\phi(\mu)$ and $\widehat{\psi} := \psi - \beta_\psi(\nu)$. Then

$$\beta_{\phi,\psi}(\theta_n - \mu_n \times \nu_n) = \beta_{\widehat{\phi},\widehat{\psi}}(\theta_n) - \beta_{\widehat{\phi},\widehat{\psi}}(\mu_n \times \nu_n)\,.$$

Since $\mathbf{E}\big[\beta_{\widehat{\phi},\widehat{\psi}}(\theta_n)\big] = \beta_{\phi,\psi}(\theta) - \beta_{\phi,\psi}(\mu \times \nu) = 0$, the central limit theorem yields $\sqrt{n}\,\beta_{\widehat{\phi},\widehat{\psi}}(\theta_n) \Rightarrow \zeta$. Also, by the central limit theorem

$$\sqrt{n}\,\beta_{\widehat{\phi},\widehat{\psi}}(\mu_n \times \nu_n) = \big(n^{1/4}\beta_{\widehat{\phi}}(\mu_n)\big) \otimes \big(n^{1/4}\beta_{\widehat{\psi}}(\nu_n)\big) \to 0$$

in probability. Thus the claim follows from Proposition 3.7.

Since the covariance operator of a Hilbert-space valued Gaussian random variable is trace class, it has eigenvalues $\lambda_i/2 \ge 0$ that allow one to write $4\|\zeta\|^2$ as $\sum_i \lambda_i Z_i^2$, where Z_i are independent standard normal random variables and $\sum_i \lambda_i = D(\mu)D(\nu)$. We claim that λ_i are the same as the non-0 eigenvalues in Theorem 2.7. It suffices to prove that the representation of a distribution as $\sum_j c_j(Z_j^2 - 1)$ is unique if Z_j are independent standard normal random variables, $c_j \ne 0$, and $\sum_j c_j^2 < \infty$. Indeed, the square of the characteristic function of Z_j^2 is $t \mapsto 1/(1-2it)$. Therefore, the square of the characteristic function of $\sum_j c_j(Z_j^2 - 1)$ is $t \mapsto \prod_j e^{-2ic_j t}/(1-2ic_j t)$, which has a unique meromorphic extension $f(z) := \prod_j e^{-2ic_j z}/(1-2ic_j z)$. Since the poles of f are at $(2ic_j)^{-1}$, this proves the desired uniqueness.

(vii) The example given in Remark 3.3 is incorrect. A correct example is the following. Let e_k and f_k all be orthonormal for $k \ge 1$. Write $v_0 := e_1$ and $v_k := -e_k + e_{k+1}/2$ for $k \ge 1$, and likewise write $w_0 := f_1$ and $w_k := -f_k + f_{k+1}/2$ for $k \ge 1$. Then the collection $\{v_k, w_k\,;\ k \ge 0\}$ has no obtuse angles and is affinely independent (which is the same as the property that the barycenter determines every finitely supported probability measure on the set), yet $(1/3)\big(v_0 + \sum_{k\ge1} v_k/2^{k-1}\big) = \mathbf{0} = (1/3)\big(w_0 + \sum_{k\ge1} w_k/2^{k-1}\big)$.

(viii) We assumed that our Hilbert spaces were real just after Proposition 3.5, but we should have assumed this just before Remark 3.4. Note that distances in the realification of a complex Hilbert space are the same as in the original space.

(ix) The proof of Lemma 3.8 has a gap, since we did not show that ν_k has a finite first moment. Here is a reformulation of the lemma and its proof to remedy this. It incorporates Lemma 3.9 as well.

LEMMA 3.8. *Let $\mathscr{X}$, $\mathscr{Y}$ have strong negative type. Then there are embeddings ϕ and ψ that witness strong negative type such that $\beta_{\phi\otimes\psi}$ is injective on $M^{1,1}(\mathscr{X} \times \mathscr{Y})$.*

PROOF. Let $\phi' : \mathscr{X} \to H$ and $\psi' : \mathscr{Y} \to H$ witness strong negative type. Define $\phi(x) := \big(\phi'(x) - \phi'(o), 1\big)$ and $\psi(x) := \big(\psi'(x), 1\big)$. These embed into $H \oplus \mathbb{R}$. Then β_ϕ and β_ψ are injective on $M^1(\mathscr{X})$ and $M^1(\mathscr{Y})$, respectively, and $\phi(o) = (0, 1)$.

Let $\theta \in M^{1,1}(\mathscr{X} \times \mathscr{Y})$ satisfy $\beta_{\phi\otimes\psi}(\theta) = 0$. For $k \in H$, define the bounded linear map $T_k : H \otimes H \to H$ by linearity, continuity, and

$$T_k(u \otimes v) := \langle u, k\rangle v\,.$$

More precisely, one uses the above definition on $e_i \otimes e_j$ for an orthonormal basis $\{e_i\}$ of H and then extends. Also, define

$$\nu_k(B) := \int \langle \phi(x), k \rangle \mathbf{1}_B(y)\, d\theta(x,y) \qquad (B \subseteq \mathscr{Y} \text{ Borel}).$$

We claim that $\nu_k = 0$ for all $k \in H$.

Let F be the closed linear span of the image of ϕ. For $k \perp F$, it is immediate that $\nu_k = 0$. Consider $k = \phi(z)$ for some z. Then $2\langle \phi(x), \phi(z) \rangle = d(x,z) - d(o,x) - d(o,z) + 2$; the triangle inequality shows that this lies in $[2 - 2\,d(o,z), 2]$. Thus, if $f \geq 0$,

$$\int f\, d|\nu_k| \leq \big(1 + d(o,z)\big) \int f(y)\, d|\theta|(x,y).$$

Now use $f(y) := d(o,y)$ to see that ν_k has finite first moment. Since

$$\begin{aligned}\beta_\psi(\nu_k) &= \int \langle \phi(x), k \rangle \psi(y)\, d\theta(x,y) = \int T_k\big(\phi(x) \otimes \psi(y)\big)\, d\theta(x,y) \\ &= T_k\big(\beta_{\phi \otimes \psi}(\theta)\big) = 0,\end{aligned}$$

this implies that $\nu_k = 0$ by injectivity of β_ψ. Since $k \mapsto \nu_k$ is linear, we also have that $\nu_k = 0$ for all k in the linear span of the image of ϕ. It follows that the same holds for all $k \in F$, and therefore for all $k \in H$, because

$$\begin{aligned}|\nu_k|(B) &\leq \int \|\phi(x)\| \cdot \|k\| \mathbf{1}_B(y)\, d|\theta|(x,y) \\ &\leq \|k\| \left(\int [d(o,x) + 1]\, d\mu(x) \right)^{1/2} \nu(B)^{1/2},\end{aligned}$$

where the marginals of $|\theta|$ are μ and ν.

Since $\nu_k = 0$ for each $k \in H$, we obtain that for every Borel $B \subseteq \mathscr{Y}$,

$$\int \phi(x) \mathbf{1}_B(y)\, d\theta(x,y) = 0.$$

Defining

$$\mu_B(A) := \theta(A \times B) \qquad (A \subseteq \mathscr{X} \text{ Borel}),$$

we have $\beta_\phi(\mu_B) = \int \phi(x) \mathbf{1}_B(y)\, d\theta(x,y) = 0$, whence $\mu_B = 0$ by injectivity of β_ϕ. In other words $\theta(A \times B) = 0$ for every pair of Borel sets A and B. Since such product sets generate the product σ-field on $\mathscr{X} \times \mathscr{Y}$, it follows that $\theta = 0$. $\square$

(x) The justification of the claim at the end of the penultimate paragraph in the proof of Theorem 3.16, which says, "we have that for ρ-a.e. w, for all $v \in \mathbb{R}^K$ and all $s \in \mathbb{R}$, $\beta(\mu)\big((v, w), s\big) = 0$", is the following. Let $M := \int \|u\|_2 d|\mu|(u)$ and $\epsilon > 0$. Let $v, v' \in \mathbb{R}^K$ with $\|v - v'\|_2 < \epsilon^2/M$. Let $w \in \mathbb{R}^\infty$ and $i = 1, 2$. Since $(v, w)(u) = \langle u_{\le K}, v\rangle + w(u_{>K})$ and $\mu_i\{u\,;\ \|u_{\le K}\|_2 > M/\epsilon\} < \epsilon$, we have

$$\begin{aligned}\mu_i\big\{u\,;\ (v', w)(u) \le s - \epsilon\big\} - \epsilon < \mu_i\big\{u\,;\ (v, w)(u) \le s\big\}\\ < \mu_i\big\{u\,;\ (v', w)(u) \le s + \epsilon\big\} + \epsilon\,.\end{aligned}$$

Fix w such that $\beta(\mu)\big((v', w), s\big) = 0$ for λ^K-a.e. $v' \in \mathbb{R}^K$ and for λ-a.e. $s \in \mathbb{R}$. Then for every $v \in \mathbb{R}^K$, we may choose $v' \in \mathbb{R}^K$ with $\|v - v'\|_2 < \epsilon^2/M$ and $\beta(\mu)\big((v', w), t\big) = 0$ for $t = s \pm \epsilon$. It follows that $\big|\beta(\mu)\big((v, w), s\big)\big| < 2\epsilon$. Since this holds for all $\epsilon > 0$ and λ-a.e. $s \in \mathbb{R}$, it follows that $\beta(\mu)\big((v, w), s\big) = 0$ for λ-a.e. $s \in \mathbb{R}$. Finally, since $\beta(\mu)\big((v, w), s\big)$ is continuous from the right, we obtain the same equality for all $s \in \mathbb{R}$.

References.

Brouwer, A. E. and Haemers, W. H. (2012). *Spectra of Graphs*. *Universitext*. Springer, New York. MR2882891

Jakobsen, M. E. (2017). Distance Covariance in Metric Spaces: Non-Parametric Independence Testing in Metric Spaces Master's thesis, University of Copenhagen https://arxiv.org/abs/1706.03490.

Department of Mathematics,
831 E. 3rd St.,
Indiana University,
Bloomington, IN 47405-7106
E-mail: rdlyons@indiana.edu
URL: http://pages.iu.edu/~rdlyons