Appendix of Local Distance Correlation

Cencheng Shen[*1], Joshua T. Vogelstein[†2], and Carey E. Priebe[‡3]

[1]Department of Statistics, Temple University
[2]Department of Biomedical Engineering and Institute of Computational Medicine, Johns Hopkins University
[3]Department of Applied Mathematics and Statistics, Johns Hopkins University

December 7, 2015

[*]cshen@temple.edu
[†]jovo@jhu.edu
[‡]cep@jhu.edu

1 Proofs

Theorem 1. *Local distance correlation is consistent for testing independence against all alternatives, i.e., the testing power $\beta \to 1$ as $n \to \infty$.*

Proof. The power of local distance correlation satisfies

$$\beta = \max_{k,l}\{\beta(dCorr_{kl}), k, l \in [2, \ldots, n]\} \geq \beta(dCorr_n). \tag{1}$$

When testing independence against any alternative, as $n \to \infty$ we have $\beta(dCorr_n) \to 1$ by Szekely et al. (2007). It follows that $\beta \to 1$, and local distance correlation is also consistent.

Note that local modified distance correlation $\{mdCorr_{kl}\}$ is also consistent, by the consistency of global modified distance correlation in Szekely and Rizzo (2013). $\square$

Next we prove a number of useful propositions for local distance correlation. In particular, proposition 1-4 indicate that one should fix neighborhood ratio $(\frac{k}{n}, \frac{l}{n})$ rather than neighborhood size (k, l) when determining the optimal neighborhood.

Proposition 1. *1. For any k, l we have*

$$dVar_k(\mathcal{X}) = O(p),$$
$$dVar_l(\mathcal{Y}) = O(q),$$
$$dCov_{kl}(\mathcal{X}, \mathcal{Y}) = O(\min\{p, q\}),$$
$$|dCorr_{kl}(\mathcal{X}, \mathcal{Y})| \leq 1,$$

where $p = \frac{k}{n}$ and $q = \frac{l}{n}$.

2. For any fixed (p, q) with $k = pn, l = qn$, we have as $n \to \infty$,

$$dCov_{kl}(\mathcal{X}, \mathcal{Y}) \to E(dCov_{kl}(\mathcal{X}, \mathcal{Y})) = E(A_{ij}^H B_{ij}^H I(A_{ij} < \epsilon_{A_j}) I(B_{ij} < \epsilon_{B_j})),$$

$$dVar_k(\mathcal{X}) \to E(dVar_k(\mathcal{X})) = E(A_{ij}^H A_{ij}^H I(A_{ij} < \epsilon_{A_j})),$$

$$dVar_l(\mathcal{Y}) \to E(dVar_l(\mathcal{Y})) = E(B_{ij}^H B_{ij}^H I(B_{ij} < \epsilon_{B_j})),$$

where ϵ_{A_j} and ϵ_{B_j} are two random variables satisfying $Prob(A_{ij} < \epsilon_{A_j} | X_j) = p$ and $Prob(B_{ij} < \epsilon_{B_j} | Y_j) = q$.

3. For any k, l, we have

$$dVar_k(\mathcal{X}) = 0 \Leftrightarrow X \text{ is degenerate},$$

$$dVar_l(\mathcal{Y}) = 0 \Leftrightarrow Y \text{ is degenerate}.$$

When neither X nor Y is degenerate, we further have

$$dVar_k(\mathcal{X}) \to 0 \Leftrightarrow p \to 0,$$

$$dVar_l(\mathcal{Y}) \to 0 \Leftrightarrow q \to 0,$$

$$dCov_{kl}(\mathcal{X}, \mathcal{Y}) \to 0 \Leftrightarrow \min\{p, q\} \to 0.$$

4. Suppose X and Y are non-degenerate, and neither p nor q converges to 0. Then we have

$$dCorr_{kl}(\mathcal{X}, \mathcal{Y}) \to \frac{E(dCov_{kl}(\mathcal{X}, \mathcal{Y}))}{\sqrt{E(dVar_k(\mathcal{X})) E(dVar_l(\mathcal{Y}))}}$$

$$= \frac{E(A_{ij}^H B_{ij}^H I(A_{ij} < \epsilon_{A_j}) I(B_{ij} < \epsilon_{B_j}))}{\sqrt{E(A_{ij}^H A_{ij}^H I(A_{ij} < \epsilon_{A_j})) E(B_{ij}^H B_{ij}^H I(B_{ij} < \epsilon_{B_j}))}}.$$

5. Suppose X and Y are non-degenerate. Then under the null hypothesis that X is independent of Y, we have

$$E(dCov_{kl}(\mathcal{X}, \mathcal{Y})) \geq 0,$$

3

with equality holds if and only if $\max\{p,q\} = 1$.

It follows that $dCov_{kl}$ and $dCorr_{kl}$ are asymptotically no less than 0 for any k, l.

6. In the permutation test, the power of local distance covariance is the same as the power of local distance correlation, because $\beta(dCorr_{kl}) = \beta(dCov_{kl})$ for all k, l. The p-value of local distance covariance is also the same as that of local distance correlation.

Proof. 1.

For local distance covariance, we have

$$
\begin{aligned}
|dCov_{kl}(\mathcal{X}, \mathcal{Y})| &= \frac{1}{n^2} | \sum_{i,j=1}^{n} A_{ij}^H B_{ij}^H I(r(A_{ij}) < k) I(r(B_{ij}) < l) | \\
&\leq \frac{1}{n^2} \max_{i,j} |A_{ij}^H B_{ij}^H| \cdot \min\{kn, ln\} \\
&= O(\min\{p, q\}),
\end{aligned}
$$

where the last equality follows from the finite second moments assumption such that $A_{ij}^H B_{ij}^H$ is bounded. Similarly we can derive that $dVar_k(\mathcal{X}) = O(p)$ and $dVar_l(\mathcal{Y}) = O(q)$.

Since $|dCov_{kl}| \leq \sqrt{dVar_k \cdot dVar_l}$ by the Cauchy-Schwartz inequality, we immediately have $|dCorr_{kl}| \leq 1$ for any k, l.

2.

For each i, j, $A_{ij} = \|X_i - X_j\|_2$ such that A_{ij} is only correlated with $A_{is}, A_{si}, A_{js}, A_{sj}$ for $s = 1, \ldots, n$. Thus there are at most $4n$ out of n^2 distance entries in A that are correlated with each A_{ij}.

Since $A^H = HAH$, we have

$$
A_{ij}^H = A_{ij} - \bar{A}_{i\cdot} - \bar{A}_{\cdot j} + \bar{A}_{\cdot\cdot},
$$

where $\bar{A}_{i\cdot} = \sum_{s=1}^{n} \frac{A_{is}}{n}$, $\bar{A}_{\cdot j} = \sum_{s=1}^{n} \frac{A_{sj}}{n}$, and $\bar{A}_{\cdot\cdot} = \sum_{s,t=1}^{n} \frac{A_{st}}{n^2}$. Thus for each A_{ij}^H, it is correlated with all of A_{st}^H at least through the centering terms.

4

But when $i, j \notin \{s, t\}$, we have

$$
\begin{aligned}
Cov(A_{ij}, A_{st}^H) &= Cov(A_{ij}, A_{st} - \bar{A}_{s\cdot} - \bar{A}_{\cdot t} + \bar{A}_{\cdot\cdot}) \\
&= Cov(A_{ij}, -\frac{A_{si} + A_{sj}}{n} - \frac{A_{it} + A_{jt}}{n} + 2\sum_{s=1}^{n} \frac{A_{is} + A_{sj}}{n^2}) \\
&= O(\frac{Var(A_{ij})}{n}) \\
&= O(\frac{1}{n}),
\end{aligned}
$$

where the last equality follows from the finite second moments of X such that $Var(A_{ij})$ is bounded above. Similarly we have

$$
\begin{aligned}
Cov(A_{ij}I_{ij}, A_{st}^H I_{st}) &= Cov(A_{ij}I_{ij}, A_{st}I_{st} - \bar{A}_{s\cdot}I_{st} - \bar{A}_{\cdot t}I_{st} + \bar{A}_{\cdot\cdot}I_{st}) \\
&= Cov(A_{ij}I_{ij}, A_{st}I_{st}) + Cov(A_{ij}I_{ij}, -\frac{A_{si} + A_{sj}}{n}I_{st} \\
&\quad - \frac{A_{it} + A_{jt}}{n}I_{st} + 2\sum_{s=1}^{n} \frac{A_{is} + A_{sj}}{n^2}I_{st}) \\
&= Cov(A_{ij}I_{ij}, A_{st}I_{st}) + O(\frac{1}{n}) \\
&\to O(\frac{1}{n}),
\end{aligned}
$$

where we simplify $I(r(A_{ij}) < k)$ to I_{ij} and $I(r(A_{st}) < k)$ to I_{st}. Note that $Cov(A_{ij}I_{ij}, A_{st}I_{st}) \to 0$: when $i, j \notin \{s, t\}$, A_{ij} and A_{st} are independent, while I_{ij} and I_{st} are asymptotically independent. The latter follows from the fact that $I(r(A_{ij}) < k) \to I(A_{ij} < \epsilon_{A_j})$, where ϵ_{A_j} satisfies $\int_0^{\epsilon_{A_j}} A_{ij} dA_{ij} | X_j = \frac{k}{n} = p$ and is thus independent from ϵ_{A_t} for given p.

It is similar to see that when $i, j \neq s, t$, we have

$$
Cov(A_{ij}^H A_{ij}^H, A_{st}^H A_{st}^H) = O(\frac{1}{n})
$$

$$
Cov(A_{ij}^H A_{ij}^H I(r(A_{ij}) < k), A_{st}^H A_{st}^H I(r(A_{st}) < k)) \to O(\frac{1}{n}).
$$

5

When either i or j takes value in $\{s, t\}$, the covariance is always of $O(1)$. It follows that the variance of $dVar_k(\mathcal{X})$ is asymptotically $O(\frac{1}{n})$, such that

$$dVar_k(\mathcal{X}) \to E(dVar_k(\mathcal{X})) = E(A_{ij}^H A_{ij}^H I(A_{ij} < \epsilon_{A_j}))$$

by the weak law of large numbers.

The convergence proof is similar for $dVar_k(\mathcal{Y})$ and $dCov_{kl}(\mathcal{X}, \mathcal{Y})$.

3.

If X is degenerate, i.e., X is almost surely a constant, then clearly $A_{ij} = 0$ for all i, j such that $A_{ij}^H = 0$, and $dVar_k(\mathcal{X}) = 0$ for any $k = 1, \dots, n$.

Suppose for some k such that $dVar_k(\mathcal{X}) = 0$, then we must have $A_{ii}^H = 0$ for all i, i.e., $\bar{A}_{..} = 2\bar{A}_{i.}$ for all i. Since $\sum_{i=1}^{n} \bar{A}_{i.} = n\bar{A}_{..}$, it follows that $\bar{A}_{i.} = 0$ for all i. Thus $A_{ij} = 0$ for all i, j, such that X is degenerate.

The same holds for $dVar_l(\mathcal{Y})$ such that it is 0 if and only if Y is degenerate.

When X is non-degenerate, $dVar_k(\mathcal{X}) > 0$ always holds. However, as $dVar_k(\mathcal{X}) = O(p)$, $dVar_k(\mathcal{X}) \to 0$ if and only if $p \to 0$. A similar argument works for $dVar_l(\mathcal{Y})$ and $dCov_{kl}(\mathcal{X}, \mathcal{Y})$.

4.

When X and Y are non-degenerate, both distance variances are always positive by Proposition 3). It immediately follows that

$$dCorr_{kl}(\mathcal{X}, \mathcal{Y}) \to \frac{E(A_{ij}^H B_{ij}^H I(A_{ij} < \epsilon_{A_j}) I(B_{ij} < \epsilon_{B_j}))}{\sqrt{E(A_{ij}^H A_{ij}^H I(A_{ij} < \epsilon_{A_j})) E(B_{ij}^H B_{ij}^H I(B_{ij} < \epsilon_{B_j}))}}.$$

when neither p nor q converges to 0.

6

5. At $l = n$, it follows that

$$E(dCov_{kn}(\mathcal{X}, \mathcal{Y})) = E(A_{ij}^H B_{ij}^H I(A_{ij} < \epsilon_{A_j}))$$
$$= E(A_{ij}^H I(A_{ij} < \epsilon_{A_j})) E(B_{ij}^H)$$
$$= 0$$

for any $k = 1, \ldots, n$, because $E(B_{ij}^H) = 0$ and A is independent of B under the null hypothesis. Similarly $E(dCov_{nl}(\mathcal{X}, \mathcal{Y})) = 0$ for any $l = 1, \ldots, n$.

When $k, l \neq n$, we have

$$E(dCov_{kl}(\mathcal{X}, \mathcal{Y})) = E(A_{ij}^H B_{ij}^H I(A_{ij} < \epsilon_{A_j}) I(B_{ij} < \epsilon_{B_j}))$$
$$= E(A_{ij}^H B_{ij}^H I(A_{ij} < \epsilon_{A_j})) - E(A_{ij}^H B_{ij}^H I(A_{ij} < \epsilon_{A_j}) I(B_{ij} \geq \epsilon_{B_j}))$$
$$= -E(A_{ij}^H I(A_{ij} < \epsilon_{A_j})) E(B_{ij}^H I(B_{ij} \geq \epsilon_{B_j}))$$
$$> 0.$$

The last inequality follows because $E(A_{ij}^H I(A_{ij} < \epsilon_{A_j})) \leq 0$ with equality holds if X is degenerate or $p = 1$ (i.e., $\int_0^{\epsilon_{A_j}} A_{ij} | X_j = 1$ for all X_j in the support of X), by the fact that

$$E(A_{ij}^H) = E(A_{ij}^H I(A_{ij} < \epsilon_{A_j})) + E(A_{ij}^H I(A_{ij} \geq \epsilon_{A_j})) = 0,$$
$$E(A_{ij}^H I(A_{ij} < \epsilon_{A_j})) \leq E(A_{ij}^H I(A_{ij} \geq \epsilon_{A_j}));$$

and similarly $E(B_{ij}^H I(B_{ij} \geq \epsilon_{B_j})) \geq 0$ with equality holds if Y is degenerate or $q = 1$.

Then by Proposition 2) and 4), $dCov_{kl}$ and $dCorr_{kl}$ are asymptotically no less than 0 under the null hypothesis.

6. In the permutation test, the p-value of $dCorr_{kl}$ for given k, l is derived by comparing $dCorr_{kl}(\mathcal{X}, \mathcal{Y})$ to $\{dCorr_{kl}(\mathcal{X}, \mathcal{Y}P)\}$ containing all permutation matrices P.

Since $dVar_{kl}(\mathcal{Y}P) = dVar_{kl}(\mathcal{Y})$ for any permutation matrix P, the p-value is the same by comparing $dCov_{kl}(\mathcal{X}, \mathcal{Y})$ to $\{dCov_{kl}(\mathcal{X}, \mathcal{Y}P)\}$.

So the p-values are always the same between local distance covariance and local distance correlation; and it follows that the permutation test powers are also the same. □

Theorem 2. *Suppose $Y = cX$ for a non-zero scalar c, then for any n we always have*

$$\beta(dCorr_n) \geq \beta(dCorr_{kl}) \tag{2}$$

for all $k, l = 2, \ldots, n$, where β is the permutation test power at a given type 1 error α.

Thus local distance correlation is no better than global distance correlation under linear dependency.

Proof. Let us first suppose no two columns of $\mathcal{X}$ have the same magnitude, which holds with probability 1 if X is continuously distributed.

Then under linear dependency, it is clearly the case that

$$dCov_n(\mathcal{X}, \mathcal{Y}) = c \cdot dVar_n(\mathcal{X}) \geq dCov_n(\mathcal{X}, \mathcal{Y}P) \tag{3}$$

for all possible permutations P. The equality holds if and only if P is the identity permutation, such that the p-value of $dCov_n$ is always $\frac{1}{n!}$.

Since $dCov_{kl}(\mathcal{X}, \mathcal{Y}) = dCov_{kl}(\mathcal{X}, \mathcal{Y}I)$ for the identity permutation as well, the p-value of $dCov_{kl}$ is at least $\frac{1}{n!}$; thus the p-value of $dCov_{kl}$ is larger than or equal to the p-value of $dCov_n$ for all $(k, l) \neq (n, n)$.

The above result still holds if some or all sample points in $\mathcal{X}$ have the same magnitude. For example, if two columns of $\mathcal{X}$ coincide with each other, Equation 3 still holds for all permutation matrices, but the equality holds when P is either identity or permutes only the same two columns; it follows that the p-value of $dCov_n$ becomes $\frac{2}{n!}$, which is still no larger than the p-value of $dCov_{kl}$ for any $(k, l) \neq (n, n)$ by the same argument in the previous paragraph. A similar argument shows that the p-value of $dCov_n$ is always the smallest when there are more than two sample points that coincide with each other.

8

Thus for any given sample data, the p-value of $dCov_{kl}$ is always larger than or equal to the p-value of $dCov_n$. Then at any given type 1 error level α, either the alternative hypothesis is rejected by both test statistics when both p-values are higher than α, or it is always accepted when both p-values are lower than α, or it is accepted only by $dCov_n$ when only the p-value of $dCov_n$ is lower than α. Thus at any given α, we must have $\beta(dCov_n) \geq \beta(dCov_{kl})$ for all $(k, l) \neq (n, n)$.

By Proposition 6) on the equivalence of local distance covariance and local distance correlation in the permutation test, it follows that Equation 2 holds. So global distance correlation $dCorr_n$ cannot be worse than local distance correlation $\{dCorr_{kl}\}$ under linear dependency. □

Theorem 3. *There exists f_{XY}, n and α such that*

$$\beta(dCorr_n) > \beta(dCorr_{kl}) \tag{4}$$

for some $(k, l) \neq (n, n)$, where β is the permutation test power at the type 1 error α.

Thus local distance correlation can be better than global distance correlation under certain nonlinear dependency.

Proof. We give a simple discrete example of f_{XY} at $n = 6$ such that the p-value of $dCov_{kl}$ for $(k, l) \neq (n, n)$ is always lower than the p-value of $dCov_n$, which is equivalent to say the permutation test power $\beta(dCov_n)$ is smaller than $\beta(dCov_{kl})$ at an appropriate type 1 error level α.

Suppose under the alternative, f_{XY} is distributed as follows:

$$X \in \{-1, -0.6, -0.2, 0.2, 0.6, 1\} \text{ without replacement,}$$
$$Y = X^2,$$

9

which is a discrete version of the quadratic relationship in the simulations.

In this example we can easily consider all possibilities of A_{ij}^H and B_{ij}^H, and calculate $dCov_{kl}(\mathcal{X}, \mathcal{Y})$ and $\{dCov_{kl}(\mathcal{X}, \mathcal{Y}P)\}$ for all $k, l = 2, \ldots, 6$ and all permutation matrices. For example, $dCov_6(\mathcal{X}, \mathcal{Y}) = \frac{128}{75}$, $dCov_{26}(\mathcal{X}, \mathcal{Y}) = \frac{399}{212}$, and similarly all possible permuted test statistics. It follows that the p-value of $dCov_{26}$ is $\frac{4}{15}$, while the p-value of $dCov_6$ is $\frac{53}{90}$.

Note that although the p-value of $dCov_{kl}$ is less than $dCov_n$ in the above example, neither p-value are very significant. In fact, we can always consider equally spaced sample points in $[-1, 1]$ for X, increase n and reach the same conclusion with smaller p-values. Say at $n = 9$, the p-value of $dCov_{29}$ is 0.056 while the p-value of $dCov_9$ is 0.4264. But the computation of all possible permuted test statistics becomes much more time-consuming as n increases, and only approximate p-values can be derived with random permutations for larger n.

Therefore we have an example of f_{XY} and n such that $\beta(dCov_n) < \beta(dCov_{kl})$ for some $(k, l) \neq (n, n)$: by choosing $\alpha = 0.3$, at $n = 6$ we have $0 = \beta(dCov_6) < \beta(dCov_{26}) = 1$; by choosing $\alpha = 0.06$, at $n = 9$ we have $0 = \beta(dCov_9) < \beta(dCov_{29}) = 1$;

It follows that Equation 4 holds by Proposition 6), since local distance correlation is equivalent to local distance covariance in the permutation test. Thus local distance correlation $\{dCorr_{kl}\}$ can be better than the global distance correlation $dCorr_n$ under certain nonlinear dependency. $\square$

2 Simulations Set-Up

Here we list the exact dependencies for the 20 types of distributions used in the simulation section of the main paper. In the following, type 1-9 are based on the same function types from Simon and Tibshirani (2012), with slightly different coefficients for some and

an additional exponential dependency; type 10 corresponds to the uncorrelated binomial example on the Wikipedia page of uncorrelated random variable[1]; type 11-17 are exactly the same simulations used by Gorfine et al. (2012), which also appears on the Wikipedia page of correlation[2]; type 18-20 are from the distributions used in Szekely et al. (2007).

As a reminder, we have $\mathcal{X} = [X_1, \cdots, X_n] \in \mathcal{R}^{m_X \times n}$ and $\mathcal{Y} = [Y_1, \cdots, Y_n] \in \mathcal{R}^{m_Y \times n}$ with $(X_i, Y_i) \overset{i.i.d.}{\sim} f_{XY}$, where n is the sample size and m_X and m_Y are the dimensions. We considered two scenarios, a dimension 1 scenario and an increasing dimension scenario. For the dimension 1 scenario we always have $m_X = m_Y = 1$, while in the increasing dimension scenario we let m_X increase, with $m_Y = 1$ for type $1 - 16$ and $m_Y = m_X$ for type $17 - 20$.

For the purpose of increasing dimension scenario, we denote A as a size $1 \times m_X$ vector with $A_i = \frac{1}{i}$. Thus AX is a dimension 1 random variable that is a decaying weighted summation of X, which equals X when $m_X = 1$.

The noiseless distribution for each type is listed below, where we may use Z, U, V, Q to denote some auxiliary random variables.

1. Linear: $X \sim Uniform(-1, 1)^{m_X}$, $Y = AX$.
2. Quadratic: $X \sim Uniform(-1, 1)^{m_X}$, $Y = (AX)^2$.
3. Cubic: $X \sim Uniform(-1, 1)^{m_X}$, $Y = 128(AX - \frac{1}{3})^3 + 48(AX - \frac{1}{3})^2 - 12(AX - \frac{1}{3})$.
4. Sine Period 1/2: $X \sim Uniform(-1, 1)^{m_X}$, $Y = sin(4\pi \frac{AX}{\max(AX)})$.
5. Sine Period 1/8: $X \sim Uniform(-1, 1)^{m_X}$, $Y = sin(16\pi \frac{AX}{\max(AX)})$.
6. $X^{\frac{1}{4}}$: $X \sim Uniform(-1, 1)^{m_X}$, $Y = |AX|^{\frac{1}{4}}$.
7. Circle: $X \sim Uniform(-1, 1)^{m_X}$, $Z \sim Bernoulli(0.5)$, $Y = (2Z-1) \cdot |1 - (2\frac{AX}{\max(AX)} - 1)^2|^{\frac{1}{2}}$.
8. Step function: $X \sim Uniform(-1, 1)^{m_X}$, $Y = I(AX > E(AX))$, where $E(AX)$ is

---

[1]https://en.wikipedia.org/wiki/Uncorrelated_random_variables
[2]https://en.wikipedia.org/wiki/Correlation_and_dependence

11

the mean of AX and I is the indicator function. Note that in the simulation code the sample mean is used instead.

9. Exp(X): $X \sim Uniform(-1, 1)^{m_X}$, $Y = exp(AX)$.

10. Uncorrelated Binomial: $X \sim Bernoulli(0.5)^{m_X}$, $Z \sim Bernoulli(0.5)^{m_X}$, $Y = AX \cdot A(2Z - 1)$.

11. W: $X \sim Uniform(-1, 1)^{m_X}$, $Z \sim Uniform(-1, 1)^{m_X}$, $Y = 4(((AX)^2 - \frac{1}{2})^2 + AZ/500)$.

12. Square: Let $U \sim Uniform(-1, 1)^{m_X}$, $V \sim Uniform(-1, 1)^{m_X}$, $\theta = -\frac{\pi}{8}$. Then $X = AU\cos(\theta) + AV\sin(\theta), Y = -AU\sin(\theta) + AV\cos(\theta)$.

13. Diamond: Let $U \sim Uniform(-1, 1)^{m_X}$, $V \sim Uniform(-1, 1)^{m_X}$, $\theta = -\frac{\pi}{4}$. Then $X = AU\cos(\theta) + AV\sin(\theta), Y = -AU\sin(\theta) + AV\cos(\theta)$.

14. Parabola: $X \sim Uniform(-1, 1)^{m_X}$, $Q \sim Uniform(0, 1)$, $Y = ((AX)^2 + Q)/2$. Note that this is like a noisy version of type 2 quadratic.

15. Two Parabolas: $X \sim Uniform(-1, 1)^{m_X}$, $Q \sim Uniform(0, 1)$, $Z \sim Bernoulli(0.5)$, $Y = ((AX)^2 + Q) \cdot (Z - \frac{1}{2})$.

16. Circle 2: Let $U \sim Uniform(-1, 1)^{m_X}$, $Q \sim Normal(0, 1)$, $Z \sim Normal(0, 1)$. Then $X = \sin(AU\pi) + \frac{Q}{8}$, $Y = \cos(AU\pi) + \frac{Q}{8}$. Note that this is a different circle distribution from type 7: it is constructed based on trigonometry and already includes certain noise; although at dimension 1 they are quite similar, their dependencies in increasing dimension are of larger discrepancy.

17. Independent Clouds: Let $U \sim Normal(0, 1)^{m_X}$, $V \sim Normal(0, 1)^{m_Y}$, $Z \sim Bernoulli(0.5)$, $Q \sim Bernoulli(0.5)$. Then we set $m_X = m_Y$, and take $X = AU/3 + Z, Y = AV/3 + Q$.

18. Joint Normal: Set $m_X = m_Y$. Denote $\rho = \frac{1}{2m_X}$, I_{m_X} as the identity matrix of size

$m_X \times m_X$, J_{m_X} as the matrix of ones of size $m_X \times m_X$, and $\Sigma = \begin{bmatrix} I_{m_X} & \rho J_{m_X} \\ \rho J_{m_X} & I_{m_X} \end{bmatrix}$. Then let $(X, Y) \sim Normal(0, \Sigma)$.

19. $Log(X^2)$: $X \sim Normal(0, 1)^{m_X}$, $Y = log(X^2)$.

20. Multiplicative Noise: $X \sim Normal(0, 1)^{m_X}$, $Z \sim Normal(0, 1)$, $Y = X \cdot Z$.

For each distribution, X and Y are dependent under the alternative (except type 17). By generating the sample data and calculating the sample test statistic for sufficiently many Monte-Carlo replicates, we obtain the empirical distribution of the test statistic under the alternative. For the empirical distribution under the null, we first generate (X, Y) by the same joint distribution, then independently generate X' by the same marginal distribution, followed by calculating the test statistic between (X', Y). Since X' distributes the same as X, the independence test is an approximation of the permutation test.

Note that in the simulation code, appropriate level of white noise may be added to Y so that it is more difficult to test independence on the sample data sets. This is because in the absence of noise, certain dependency such as linear at dimension 1 is very easy to be detected: the testing power of all methods is 1 at very small n, and we cannot meaningfully compare the methods.

To see that local distance correlation are strongly dependent for adjacent neighbors, in Figure 1 we show how the powers of local distance correlation change with respect to increasing neighborhood k, by plotting $\max_{l=2,...,n}\{\beta(dCorr_{kl})\}$ as the y-axis for each k. They are plotted at a fixed sample size and a fixed dimension for the two scenarios in the main paper, using the same thresholds as the performance profiles figure there. Clearly for consecutive neighborhoods, the testing powers are very close to each other, which is also the case if we instead plot the powers with respect to neighborhood l, or look at the surface plot. This implies that it is usually fine when a sub-optimal neighborhood choice is picked

13

(say it is determined by minimizing the p-value from a permutation test), because such a choice is likely close to the optimal choice and still yields very good testing power.

For the performance profiles in the simulation, we have chosen two power thresholds 0.8 and 0.5 for both scenarios to illustrate the advantage of local distance correlation. The same advantage in fact holds for any other threshold choices as well, and in Figure 2 we plot the area under curve of performance profiles against the power threshold, i.e., the threshold 0.8 and 0.5 in both scenarios now take value from 0 to 1, and we calculate the performance profiles and take their AUC as the y-axis. This figure yields a similar interpretation as the performance profiles figure in the main paper, i.e., local distance correlation significantly improves over global distance correlation, and local modified distance correlation is the most reliable method.

(a)   (b)

Figure 1: Testing Power for 20 Simulations with respect to Increasing Neighborhoods

(a)   (b)

Figure 2: Area under Curve of Performance Profiles for 20 Simulations

References

Gorfine, M., R. Heller, and Y. Heller (2012). Comment on detecting novel associations in large data sets. *available at http://ie.technion.ac.il/ gorfinm/files/science6.pdf* .

Simon, N. and R. Tibshirani (2012). Comment on detecting novel associations in large data sets. *available at http://arxiv.org/abs/1401.7645*.

Szekely, G. and M. Rizzo (2013). The distance correlation t-test of independence in high dimension. *Journal of Multivariate Analysis 117*, 193–213.

Szekely, G., M. Rizzo, and N. Bakirov (2007). Measuring and testing independence by correlation of distances. *Annals of Statistics 35*(6), 2769–2794.