Laplacian smooth test statistic for two-sample testing

Alden Green

December 6, 2018

1 Goals

- Find an asymptotic null distribution.

2 Setup

We observe data $X_1, \ldots, X_n \sim P$ and $Y_1, \ldots, Y_m \sim Q$. Our goal is to test the hypothesis $H_0 : P = Q$ vs. the alternative $H_a : P \neq Q$.

Let $Z = (X_1, \ldots, X_n, Y_1, \ldots, Y_m)$. Define 1_X to be the $n + m$ length indicator vector for X

$$1_X[i] = \begin{cases} 1, i \in 0, 1, \ldots, n \\ 0 \text{ otherwise} \end{cases}$$

and similarly for 1_Y

$$1_Y[j] = \begin{cases} 1, j \in n + 1, \ldots, m \\ 0 \text{ otherwise} \end{cases}$$

and define $a = \frac{1_X}{m} - \frac{1_Y}{n}$.

Form an $n + m \times n + m$ similarity matrix A, where $A_{ij} = K(Z_i, Z_j)$ for some unspecified choice of K, and take $L = D - A$ to be Laplacian matrix of A (where D is a diagonal matrix with $D_{ii} = \sum_{j \in [n+m]} A_{ij}$).

We are ready to define our test statistic.

$$T_2^2 = \left(\max_{\theta : \theta^T L \theta \leq 1} a^T \theta. \right)^2$$

Spectral properties of L. Define the pseudo-inverse of L to be $L^\dagger$. In what follows, we will assume A defines a connected graph G. In this setting, it is well known that L has exactly one 0 eigenvalue, with corresponding eigenvector $\mathbf{1}$. Let $P_{1^\perp}$ be the projection onto the linear subspace orthogonal to this eigenvector.

Poissonization. For $p \in (0,1)$, draw $U_1, \ldots, U_N \sim Bern(p)$. Then, draw $Z_i \sim P_{2U_i - 1}$. Consider $a = (a_i)_{i=1}^N$ with $a_i = 2U_i - 1$. Let the null hypothesis be $H_0 : P_1 = P_2$ and the alternative be $H_a : P_1 \neq P_2$.

Distances between probability measures. For a function f, define its **Lipschitz norm** $\|f\|_L$ to be

$$\inf K : |f(x) - f(y)| \leq K \|x - y\| .$$

Define the **Wasserstein distance** between two measures μ and ν to be

$$\mathcal{W}(\mu, \nu) := \sup \left\{ \left| \int h \, d\mu - \int h \, d\nu \right| : h \text{ Lipschitz, with } \|h\|_L \leq 1 \right\} .$$

If the measures μ and ν have corresponding cumulative distribution functions F_μ and F_ν then we can define the **Kolmogorov-Smirnov distance** to be

$$\|F_\mu - F_\nu\|_\infty := \sup_t |F_\mu(t) - F_\nu(t)| .$$

The following lemma allows us to translate from an upper bound on Wasserstein distance to Kolmogorov distance.

Lemma 1 (Wasserstein to Kolmogorov distance)**.** For any probability measures μ, ν with corresponding cdfs F_μ and F_ν and any $\epsilon' > 0$, there exists some $\epsilon > 0$ such that

$$\mathcal{W}(\mu, \nu) < \epsilon \implies \sup_t |F_\mu(t) - F_\nu(t)| \leq \epsilon' .$$

3 Related Work

(Bhattacharya 2018) defines a general notion of 2-sample graph-based test statistics

Definition 3.1.

$$T(G) = \sum_{i=1}^{n} \sum_{j=n+1}^{n+m} 1(e_{ij} \in E)$$

and letting $a_0 = \frac{1}{2}(1_X - 1_Y)$, we can write

$$T(G) = a_0^T L a_0.$$

(Gretton 2012) considers the test statistic

$$T = a^T A a.$$

4 Conjectures

The following will be needed for Theorem 2.

Conjecture 1. There exists a sequence of scaling factors $(\rho_n)_{n=1}^{\infty}$ such that the **spectral measure** μ_n of $\rho_n L^{\dagger}$ converges weakly in probability

$$\mu_n(\rho_n L^{\dagger}) \xrightarrow{*} \nu_{\infty}.$$

where $V \sim \nu_{\infty}$ and $V_n \sim \mu_n$ are bounded almost surely for all n by some constant C.

Conjecture 2. For all $\epsilon > 0$, there exists N such that

$$\mathbb{P}\left(\max_{i \in [n]} \frac{1}{n} \left(\{\rho_n L^{\dagger}\}^2\right)_{ii} \leq \epsilon\right) \geq 1 - \epsilon$$

for all $n \geq N$.

5 Results

Central limit theorem for quadratic forms.

Theorem 1 (Chatterjee 08)**.** Let $a = (a_1, \ldots, a_n)$ be i.i.d random variables with with $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$. For some fixed real valued symmetric matrix $M = (M_{ij})_{1 \leq i,j \leq n}$, define

$$W = a^T M a.$$

with μ denoting the law of $(W - EW)/\sqrt{\mathrm{Var}(W)}$.

Then, letting $\mathcal{G}$ be the standard Gaussian measure

$$\mathcal{W}(\mu, \mathcal{G}) \leq \left(\frac{\mathrm{tr}(M^4)}{\mathrm{tr}(M^2)^2}\right)^{1/2} + \left(\frac{5 \max_i (M_{ii})^2}{\mathrm{tr}(M^2)}\right)^{1/2}. \tag{1}$$

Analytic form for T_2^2. The above result becomes obviously applicable thanks to the following expression for T_2^2.

Lemma 2.

$$T_2^2 = a^T L^\dagger a$$

Asymptotic null distribution for T_2.

Theorem 2. Denote the scaled version of the Laplacian smooth test statistic

$$W_n = \sqrt{\frac{2}{\mathrm{tr}((L^\dagger)^2)}}\Big(T_2^2 - 4\mathrm{tr}(L^\dagger)\Big).$$

If Conjectures 1 and 2 hold,

$$\lim_{n\to\infty} \sup_t |\mathbb{P}\left(W_n \le t\right) - \Phi(t)| = 0.$$

Proof. We will proceed by

1. Conditioning on the high-probability outcome that the Laplacian converges to a limiting object in the right sense.
2. Showing that, under such convergence of the Laplacian, both terms in Theorem 1 grow small with n.
3. Converting from Wasserstein distance to Kolmogorov distance.

Step 1. Fix $\epsilon > 0$. Throughout, let P_Z denote the distribution of Z, and likewise P_a denote the distribution of a.

For $V_n \sim \nu_n(\rho_n L^\dagger)$, and $V \sim \nu_\infty$ let

$$A_n = \left\{z \in \mathbb{R}^n : |EV_n^p - EV^p| \le \epsilon \text{ for } p = 1,2,4\right\} \bigcup \left\{z \in \mathbb{R}^n : \max_{i\in[n]} \frac{1}{n}\left(\{\rho_n L^\dagger\}^2\right)_{ii} \le \epsilon\right\}.$$

It is not hard to see that our Conjectures 1 and 2 imply A_n will eventually have high probability.

$$\begin{aligned}\mathbb{P}\left(A_n\right) &\ge \mathbb{P}\left(\left\{z \in \mathbb{R}^n : |EV_n^p - EV^p| \le \epsilon\right\}\right) + \mathbb{P}\left(\left\{z \in \mathbb{R}^n : \max_{i\in[n]} \frac{1}{n}\left(\{\rho_n L^\dagger\}^2\right)_{ii} \le \epsilon\right\}\right) \\ &\overset{(i)}{\ge} 1 - 2\epsilon \text{ for all } n \ge N. \qquad (2)\end{aligned}$$

where (i) follows from Conjecture 2 (for the second term), and Conjecture 1 (for the first term).

Writing $W_n := W_n(z,a)$ to emphasize that it is a function of z and a, we have by Tonelli's theorem that

$$
\begin{aligned}
\sup_t \left|\mathbb{P}\left(W_n \leq t\right) - \Phi(t)\right| &\overset{(i)}{=} \sup_t \left| \int_{\mathbb{R}^N} \left(\int_{\{-1,1\}^N} 1(W_n(z,a) \leq t) dP_a \right) dP_z - \Phi(t) \right| \\
&= \sup_t \left| \int_{\mathbb{R}^N} \left(\int_{\{-1,1\}^N} 1(W_n(z,a) \leq t) dP_a \right) - \Phi(t) dP_z \right| \\
&\leq \int_{\mathbb{R}^N} \sup_t \left| \left(\int_{\{-1,1\}^N} 1(W_n(z,a) \leq t) dP_a \right) - \Phi(t) \right| dP_z \\
&\overset{(ii)}{\leq} \int_{A_n} \sup_t \left| \left(\int_{\{-1,1\}^N} 1(W_n(z,a) \leq t) dP_a \right) - \Phi(t) \right| dP_z + 2\epsilon
\end{aligned} \tag{3}
$$

where (i) follows from Tonelli's theorem and (ii) from (2).

Step 2. Denote as

$$
F_{a|z}(z,t) := \left(\int_{\{-1,1\}^N} 1(W_n(z,a) \leq t) dP_a \right)
$$

and note that for any z this defines a measure over the Borel σ-algebra $\mathcal{B}(\mathbb{R})$, which we will call $\mu_{a|Z}(z)$.

We wish to upper bound $\mathcal{W}(\mu_{a|Z}(z), \mathcal{G})$. To do so, we will compute upper bounds for each present in (1). For the first term, we have

$$
\begin{aligned}
\frac{\operatorname{tr}(\{L^\dagger\}^4)}{\operatorname{tr}(\{L^\dagger\}^2)^2} &= \frac{1}{n} \frac{\frac{1}{n}\operatorname{tr}(\rho_n^4 \{L^\dagger\}^4)}{\frac{1}{n^2}\rho_n^4 \operatorname{tr}(\{L^\dagger\}^2)^2} \\
&\leq \frac{1}{n} \frac{\mathbb{E}\left[V^4\right] + \epsilon}{\mathbb{E}\left[V^2\right]^2 - \epsilon}.
\end{aligned}
$$

For the second term, we have

$$
\begin{aligned}
\frac{\max_i (\{L^\dagger\}^2)_{ii}}{\operatorname{tr}(\{L^\dagger\}^2)} &= \frac{\frac{\rho_n^2}{n}(\{L^\dagger\}^2)_{ii}}{\frac{\rho_n^2}{n}\operatorname{tr}(\{L^\dagger\}^2)} \\
&\leq \frac{\epsilon}{\mathbb{E}\left[V^2\right] - \epsilon}.
\end{aligned}
$$

By Theorem 1 we therefore have

$$
\mathcal{W}(\mu_{a|Z}(z), \mathcal{G}) \leq \frac{1}{n} \frac{\mathbb{E}\left[V^4\right] + \epsilon}{\mathbb{E}\left[V^2\right]^2 - \epsilon} + \left(\frac{\epsilon}{\mathbb{E}\left[V^2\right] - \epsilon} \right)^{1/2}. \tag{4}
$$

Step 3. Note that the right hand side of (4) converges to 0 with ϵ. Therefore, for any ϵ sufficiently small, by (4) and Lemma 1 we have

$$\left\|F_{Z|a} - \Phi\right\|_{\infty} \leq \epsilon'.$$

Combined with (3) we have

$$\sup_t |\mathbb{P}\left(()\, W_n \leq t\right) - \Phi(t)| \leq 2\epsilon + \epsilon'.$$

for all $n \geq n_0$.

$\square$

Asymptotic mean. Under the null hypothesis of the **related generative model**, we have

$$\begin{aligned}
\mathbb{E}\left[T^2\right] &= \mathbb{E}\left[a^T L^\dagger a\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[a^T L^\dagger a \mid Z_1^{n+m}\right]\right] \\
&\overset{(i)}{=} \mathbb{E}\left[\sum_{i=1}^{n+m}\sum_{j=1}^{n+m} \mathbb{E}\left[a_i a_j\right] L^\dagger_{ij}\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n+m} \frac{1}{p(1-p)} L^\dagger_{ii}\right] \\
&= \frac{1}{N^2 p(1-p)} \cdot \mathbb{E}\left[\sum_{i=1}^{n+m} L^\dagger_{ii}\right] \\
&= \frac{1}{N^2 p(1-p)} \cdot \mathbb{E}\left[\mathrm{tr}(L^\dagger)\right] \qquad (5)
\end{aligned}$$

where (i) comes from the independence of Z and a under H_0.

Asymptotic variance. We begin by computing $\mathbb{E}\left[(T^2)^2\right]$. We will need the following terms

$$\begin{aligned}
S_4 &:= \sum_{i,k} L^\dagger_{ii} L^\dagger_{kk} \\
S_5 &:= \sum_{i,j} L^\dagger_{ij} L^\dagger_{ij} \\
S_6 &:= \sum_{i} L^\dagger_{ii} L^\dagger_{ii}.
\end{aligned}$$

Then

$$
\begin{aligned}
\mathbb{E}\left[(T^2)^2\right] &= \mathbb{E}\left[\sum_{i,j,k,l} L^\dagger_{ij} L^\dagger_{kl} \mathbb{E}\left[a_i a_j a_k a_l\right]\right] \\
&= \mathbb{E}\left[\frac{1}{N^2 p^2 (1-p)^2}(S_4 + 2S_5 - 3S_6) + \frac{p^3 + (1-p)^3}{N^4 p^3 (1-p)^3} S_6\right] \\
&\overset{(i)}{=} \frac{1}{N^4 p^2 (1-p)^2}\left(\mathbb{E}\left[tr(L^\dagger)\right]^2 + 2\mathbb{E}\left[tr(L^\dagger L^\dagger)\right] - 3\mathbb{E}\left[S_6\right]\right) + \frac{p^3 + (1-p)^3}{N^4 p^3 (1-p)^3}\mathbb{E}\left[S_6\right]
\end{aligned}
$$

where (i) follows from Lemma 3. Along with (5) we therefore obtain

$$
\operatorname{Var}\left(T^2\right) = \frac{1}{N^4 p^2 (1-p)^2}\left(2\mathbb{E}\left[tr(L^\dagger L^\dagger)\right] - 3\mathbb{E}\left[S_6\right]\right) + \frac{p^3 + (1-p)^3}{N^4 p^3 (1-p)^3}\mathbb{E}\left[S_6\right]
$$

Lemma 3.

$$
\begin{aligned}
S_4 &= tr(L^\dagger)^2 \\
S_5 &= tr(L^\dagger L^\dagger)
\end{aligned}
$$

Proof.

$$
\begin{aligned}
S_4 &= \sum_{i,k} L^\dagger_{ii} L^\dagger_{kk} \\
&= \left(\sum_k L^\dagger_{kk}\right)\left(\sum_i L^\dagger_{ii}\right) \\
&= tr(L^\dagger)^2.
\end{aligned}
$$

$$
\begin{aligned}
S_5 &:= \sum_{i,j} L^\dagger_{ij} L^\dagger_{ij} \\
&= \left\|L^\dagger\right\|_F = \operatorname{tr}(L^\dagger L^\dagger).
\end{aligned}
$$

□

Lemma 4. Write the Laplacian in terms of its spectral decomposition

$$
L = \sum_{k=2}^{n} \lambda_k u_k u_k^T
$$

where $\lambda_1 = 0$. Then, if $\max_{k,i} u_k[i] \leq \frac{C}{\sqrt{N}}$, we have

$$
S_6 \leq C^4 \frac{tr(L^\dagger)^2}{N}.
$$

Proof.

$$\begin{aligned} L^{\dagger}_{ii} &= \sum_{k=1}^{N} \lambda_k (u_k[i])^2 \\ &\leq \sum_{k=1}^{N} \lambda_k \frac{C^2}{N} \\ &= C^2 \frac{tr(L^{\dagger})}{N}. \end{aligned}$$

Therefore

$$\begin{aligned} S_6 &= \sum_{i} L^{\dagger}_{ii} L^{\dagger}_{ii} \\ &\leq \sum_{i} C^4 \frac{tr(L^{\dagger})^2}{N^2} \\ &= C^4 \frac{tr(L^{\dagger})^2}{N}. \end{aligned}$$

□

Theorem 3. Assume Conjectures 1, **??**, and **??** hold. Then

$$\rho_n n^{3/2} (T_2^2) \xrightarrow{L} N(0, \frac{2}{p^2(1-p)^2} \mathbb{E}_{\nu_\infty} V^2)$$

Proof. Denote $W(n) = \rho_n n^{3/2}(T_2^2)$ and note that

$$W(n) = \frac{\rho_n}{\sqrt{n}} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} L^{\dagger}_{ij} a_i a_j \right).$$

By the independence of a_i and a_j, $\mathbb{E}\left[a_i a_j \mid a_j\right] = 0$ when $i \neq j$. For the diagonal terms, we have that

$$\begin{aligned} L^{\dagger}_{ii} \frac{\rho_n}{\sqrt{n}} &\overset{(i)}{=} \mathcal{O}(\frac{tr(L^{\dagger})\rho_n}{n} \frac{1}{\sqrt{n}}) \\ &\overset{n\to\infty}{=} O(\frac{1}{\sqrt{n}}) \mathbb{E}_{\nu_\infty}(V) = 0 \end{aligned}$$

where (i) follows from Conjecture **??** and Lemma 4. So, asymptotically, the conditional expectation $\mathbb{E}\left[L^{\dagger}_{ij} a_i a_j \mid a_j\right] = 0$ a.s.

The variance of $W(n)$ is calculated as

$$\begin{aligned}
Var(W(n)) &= \frac{N^4\rho_n^2}{N} Var(T_2^2) \\
&\overset{(i)}{=} \frac{\rho_n^2}{Np^2(1-p)^2}\left(2\mathbb{E}\left[tr(L^\dagger L^\dagger)\right] - 3\mathbb{E}\left[S_6\right]\right) + \frac{\rho_n\left(p^3+(1-p)^3\right)}{Np^3(1-p)^3}\mathbb{E}\left[S_6\right] \\
&\overset{(ii)}{=} \frac{1}{p^2(1-p)^2}(2\mathbb{E}_{v_\infty}(V^2)) + \mathcal{O}(\max_i (L_{ii}^\dagger)^2\rho_n^2) \\
&\overset{(iii)}{=} \frac{1}{p^2(1-p)^2}(2\mathbb{E}_{v_\infty}(V^2)).
\end{aligned}$$

where (i) follows from our calculation for asymptotic variance, (ii) follows from Conjecture 1 and the definition of S_6 and (iii) from Conjecture **??**.

We compute

$$\lim_{n\to\infty} \frac{\rho_n^4}{n^2} tr((L^\dagger)^4) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}_{\nu_\infty}(V^4) = 0.$$

and roughly this should imply the fourth moment is asymptotically 0.

Finally, we have

$$\begin{aligned}
\sum_{j=1}^n \sigma_{ij}^2 &= \sum_{j=1}^n \frac{\rho_n^2 (L_{ij}^\dagger)^2}{np^2(1-p)^2} \\
&= \frac{\rho_n^2}{np^2(1-p)^2}(L_i^\dagger)^T(L_i^\dagger) \\
&= \frac{\rho_n^2}{np^2(1-p)^2}(L^\dagger L^\dagger)_{ii} \\
&\overset{(i)}{\le} \frac{\rho_n^2}{np^2(1-p)^2} C^2 \frac{tr(L^\dagger L^\dagger)}{N} \\
&= \mathcal{O}(1/n).
\end{aligned}$$

By Theorem 2.1 of (de Jong 87), we therefore have

$$\rho_n n^{3/2}(T_2^2) \xrightarrow{L} N(0, \frac{2}{p^2(1-p)^2}\mathbb{E}_{\nu_\infty} V^2).$$

□

6 Proofs

Proof of Lemma 2. Take the Lagrangian

$$L(\theta, \lambda) = -a^T\theta + \lambda\theta^T L\theta$$

and let

$$\lambda^\star = \sqrt{a^T L^\dagger a}$$
$$\theta^\star = \frac{a^T L^\dagger}{\lambda^\star}$$

The KKT conditions tell us that if

$$\frac{\partial L}{\partial \theta}(\lambda^\star, \theta^\star) = 0 \tag{6}$$
$${\theta^\star}^T L \theta^\star = 1 \tag{7}$$
$$\lambda^\star \geq 0 \tag{8}$$

then $\theta^\star$ is a primal solution.

We can write

$$\frac{\partial L}{\partial \theta} = -a^T + \lambda \theta^T L$$

and plugging in our choice for $\theta^\star$ yields

$$\begin{aligned}
\frac{\partial L}{\partial \theta}(\lambda^\star, \theta^\star) &= -a^T + \lambda^\star a^T L^\dagger L \\
&\overset{(i)}{=} -a^T + a^T P_{1^\perp} \\
&\overset{(ii)}{=} -a^T + a^T = 0.
\end{aligned}$$

where (i) follows from the already stated fact that L has one 0 eigenvalue with constant eigenvector, and (ii) from the fact that $a \perp \mathbf{1}$. As a result, $(\lambda^\star, \theta^\star)$ satisfy (6).

Then, we have

$$\begin{aligned}
{\theta^\star}^T L \theta^\star &= \frac{a^T L^\dagger L L^\dagger a}{{\lambda^\star}^2} \\
&= \frac{a^T L^\dagger a}{{\lambda^\star}^2} = 1.
\end{aligned}$$

satisfying (7).

Finally, because $L^\dagger$ is a positive-definite matrix, for any vector v $v^T L^\dagger v > 0$, and therefore $\lambda^\star \geq 0$, verifying (8).

As a result, $\theta^\star$ minimizes $a^T \theta$ subject to the given constraint. Plugging in our expression for $\theta^\star$ yields

$$T = a^T \theta^\star = \frac{a^T L^\dagger a}{\lambda^\star} = \sqrt{a^T L^\dagger a}.$$

□