Hypothesis Testing with Pairwise Distances and Associated Kernels

Dino Sejdinovic[⋆], Arthur Gretton[⋆,†],
Bharath Sriperumbudur[⋆] and Kenji Fukumizu[‡]

[⋆]Gatsby Unit, CSML, UCL, [†]MPI for Intelligent Systems, Tübingen,
[‡]The Institute of Statistical Mathematics, Tokyo

ICML 2012, Edinburgh, UK

Two-sample and independence tests

- **Two-sample test**: Given $\{Z^{(i)}\}_{i=1}^{n_z} \overset{i.i.d.}{\sim} P$, and $\{W^{(i)}\}_{i=1}^{n_w} \overset{i.i.d.}{\sim} Q$,
  - H_0: $P = Q$
  - H_A: $P \neq Q$

Two-sample and independence tests

- **Two-sample test**: Given $\{Z^{(i)}\}_{i=1}^{n_z} \overset{i.i.d.}{\sim} P$, and $\{W^{(i)}\}_{i=1}^{n_w} \overset{i.i.d.}{\sim} Q$,
  - H_0: $P = Q$
  - H_A: $P \neq Q$
- **Independence test**: Given $\left\{\left(X^{(i)}, Y^{(i)}\right)\right\}_{i=1}^{m} \overset{i.i.d.}{\sim} P_{XY}$,
  - H_0: $P_{XY} = P_X P_Y$
  - H_A: $P_{XY} \neq P_X P_Y$

Energy distance and distance covariance

- **Energy distance**:

$$D_E(P, Q) = 2\mathbb{E}_{ZW} \|Z - W\|_2 - \mathbb{E}_{ZZ'} \|Z - Z'\|_2 - \mathbb{E}_{WW'} \|W - W'\|_2 ,$$

where $Z, Z' \overset{i.i.d.}{\sim} P$ and $W, W' \overset{i.i.d.}{\sim} Q$.

Energy distance and distance covariance

- **Energy distance**:

$$D_E(P,Q) = 2\mathbb{E}_{ZW}\|Z-W\|_2 - \mathbb{E}_{ZZ'}\|Z-Z'\|_2 - \mathbb{E}_{WW'}\|W-W'\|_2,$$

  where $Z, Z' \overset{i.i.d.}{\sim} P$ and $W, W' \overset{i.i.d.}{\sim} Q$.

- **Distance covariance** (weighted L_2-distance between characteristic functions):

$$\begin{aligned}\mathcal{V}^2(X,Y) = & \ \mathbb{E}_{XY}\mathbb{E}_{X'Y'}\|X-X'\|_2\|Y-Y'\|_2 \\ & + \mathbb{E}_X\mathbb{E}_{X'}\|X-X'\|_2\mathbb{E}_Y\mathbb{E}_{Y'}\|Y-Y'\|_2 \\ & - 2\mathbb{E}_{XY}\left[\mathbb{E}_{X'}\|X-X'\|_2\mathbb{E}_{Y'}\|Y-Y'\|_2\right],\end{aligned}$$

  where (X,Y) and (X',Y') are $\overset{i.i.d.}{\sim} P_{XY}$.

- Székely and Rizzo (2004, 2005); Székely, Rizzo and Bakirov (2007); Székely and Rizzo (2009), Lyons (2011)

MMD & HSIC

- $k : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ a kernel on $\mathcal{Z}$, with RKHS $\mathcal{H}_k$; P a probability measure on $\mathcal{Z}$; mean embedding of P is $\mu_P = \int k(\cdot, z)dP(z)$
- **Maximum Mean Discrepancy** between P and Q:

$$\begin{aligned}\gamma_k(P, Q) &= \|\mu_k(P) - \mu_k(Q)\|_{\mathcal{H}_k} \\ &= \left[\mathbb{E}_{ZZ'} k(Z, Z') + \mathbb{E}_{WW'} k(W, W') - 2\mathbb{E}_{ZW} k(Z, W)\right]^{1/2}\end{aligned}$$

MMD & HSIC

- $k : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ a kernel on $\mathcal{Z}$, with RKHS $\mathcal{H}_k$; P a probability measure on $\mathcal{Z}$; mean embedding of P is $\mu_P = \int k(\cdot, z) dP(z)$
- **Maximum Mean Discrepancy** between P and Q:

$$\begin{aligned}\gamma_k(P, Q) &= \|\mu_k(P) - \mu_k(Q)\|_{\mathcal{H}_k} \\ &= \left[\mathbb{E}_{ZZ'} k(Z, Z') + \mathbb{E}_{WW'} k(W, W') - 2\mathbb{E}_{ZW} k(Z, W)\right]^{1/2}\end{aligned}$$

- $k_{\mathcal{X}}$ a kernel on $\mathcal{X}$, $k_{\mathcal{Y}}$ a kernel on $\mathcal{Y}$, and $k = k_{\mathcal{X}} k_{\mathcal{Y}}$
- **Hilbert-Schmidt Independence Criterion** between X and Y:

$$HSIC(X, Y; k_{\mathcal{X}}, k_{\mathcal{Y}}) = \|\mu_k(P_{XY}) - \mu_k(P_X P_Y)\|_{\mathcal{H}_k}$$

- Gretton et al (2005, 2008); Smola et al (2007); Zhang et al (2011); Gretton et al (2012)

Beyond Euclidean metrics

- Lyons (2011) generalized energy distance and distance covariance to *metric spaces of negative type* $(\mathcal{Z}, \rho)$, s.t.

$$\sum_{i=1}^{n} \alpha_i = 0 \Rightarrow \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \rho(z_i, z_j) \leq 0.$$

Beyond Euclidean metrics

- Lyons (2011) generalized energy distance and distance covariance to *metric spaces of negative type* $(\mathcal{Z}, \rho)$, s.t.

$$\sum_{i=1}^{n} \alpha_i = 0 \Rightarrow \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \rho(z_i, z_j) \leq 0.$$

- If ρ is a (semi)metric of negative type, then $k(z, z') = \frac{1}{2}\left[\rho(z, z_0) + \rho(z', z_0) - \rho(z, z')\right]$ is a valid kernel (**distance kernel**)

Beyond Euclidean metrics

- Lyons (2011) generalized energy distance and distance covariance to *metric spaces of negative type* $(\mathcal{Z}, \rho)$, s.t.

$$\sum_{i=1}^{n} \alpha_i = 0 \Rightarrow \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \rho(z_i, z_j) \leq 0.$$

- If ρ is a (semi)metric of negative type, then $k(z, z') = \frac{1}{2}\left[\rho(z, z_0) + \rho(z', z_0) - \rho(z, z')\right]$ is a valid kernel (**distance kernel**)

- If k is a kernel, then $\rho(z, z') = \|k(\cdot, z) - k(\cdot, z')\|^2_{\mathcal{H}_k}$ is a semimetric of negative type (*generated* by k)

Main results

Theorem

Let $(\mathcal{Z}, \rho)$ be a semimetric space of negative type and let k be any kernel that generates ρ. Then,

$$D_{E,\rho}(P, Q) = 2\gamma_k^2(P, Q).$$

Main results

Theorem

Let $(\mathcal{Z}, \rho)$ be a semimetric space of negative type and let k be any kernel that generates ρ. Then,

$$D_{E,\rho}(P, Q) = 2\gamma_k^2(P, Q).$$

Theorem

Let $(\mathcal{X}, \rho_{\mathcal{X}})$ and $(\mathcal{Y}, \rho_{\mathcal{Y}})$ be semimetric spaces of negative type, and let $k_{\mathcal{X}}$ and $k_{\mathcal{Y}}$ be any two kernels on $\mathcal{X}$ and $\mathcal{Y}$ that generate $\rho_{\mathcal{X}}$ and $\rho_{\mathcal{Y}}$, respectively. Then,

$$\mathcal{V}^2_{\rho_{\mathcal{X}},\rho_{\mathcal{Y}}}(X, Y) = 4HSIC^2(X, Y; k_{\mathcal{X}}, k_{\mathcal{Y}}).$$

Conclusions

- Distance-based statistics of Szekely et al are a special case of the RKHS framework.
- Conversely, RKHS-based statistics have a clear interpretation in terms of implicitly imposing a (semi)metric onto the original space.
- For problem settings defined most naturally in terms of some given distances, and where these distances are of negative type, RKHS machinery can be brought to bear.