Kernels Based Tests with Non-asymptotic Bootstrap Approaches for Two-sample Problems

Magalie Fromont, Béatrice Laurent, Matthieu Lerasle, Patricia Reynaud-Bouret (COLT-2012)

Zoltán Szabó

Machine Learning Journal Club, Gatsby Unit

November 2, 2015

Task: two-sample problem

- Given: 2 independent samples

$$Z^{(1)} = \left\{Z_1^{(1)}, \ldots, Z_{(N_1)}^{(1)}\right\} \sim s_1,$$
$$Z^{(2)} = \left\{Z_1^{(2)}, \ldots, Z_{(N_2)}^{(2)}\right\} \sim s_2.$$

- We want to test
  - $H_0 : s_1 = s_2$, against
  - $H_1 : s_1 \neq s_2$.

Task: two-sample problem

- Given: 2 independent samples

$$Z^{(1)} = \left\{Z_1^{(1)}, \ldots, Z_{(N_1)}^{(1)}\right\} \sim s_1,$$
$$Z^{(2)} = \left\{Z_1^{(2)}, \ldots, Z_{(N_2)}^{(2)}\right\} \sim s_2.$$

- We want to test
  - $H_0 : s_1 = s_2$, against
  - $H_1 : s_1 \neq s_2$.
- [$\mathcal{Z}$: measureable space, e.g. 'nice' $\mathcal{Z} \subseteq \mathbb{R}^d$.]

Possible errors

- Error of
  - first kind: H_0 is true, but we reject it.
  - second kind: H_0 is false, but we accept it.

Possible errors

- Error of
  - first kind: H_0 is true, but we reject it.
  - second kind: H_0 is false, but we accept it.
- Test is called of level $\alpha \in (0, 1)$ if

$$\mathbb{P}(\text{first kind of error}) \leq \alpha.$$

- Difficulty: hard to find the *non-asymptotic* quantile for α-levelness! $\rightarrow$ small-sample domain.

Two-sample problem: density model – classical setup

- Examples = different $Z^{(i)}$ generating mechanisms.
- Here: $N_1 = n_1$, $N_2 = n_2$: fixed,
  - $Z^{(1)} = \left\{Z_1^{(1)}, \ldots, Z_{(n_1)}^{(1)}\right\} \sim s_1$
  - $Z^{(2)} = \left\{Z_1^{(2)}, \ldots, Z_{(n_2)}^{(2)}\right\} \sim s_2$

Two-sample problem: density model – classical setup

- Examples = different $Z^{(i)}$ generating mechanisms.
- Here: $N_1 = n_1$, $N_2 = n_2$: fixed,
  - $Z^{(1)} = \left\{Z_1^{(1)}, \ldots, Z_{(n_1)}^{(1)}\right\} \sim s_1[\ll \nu]$,
  - $Z^{(2)} = \left\{Z_1^{(2)}, \ldots, Z_{(n_2)}^{(2)}\right\} \sim s_2[\ll \nu]$,
  - [ν:
    - non-atomic: $z \in \mathcal{Z}$ s.t. $\nu(\{z\}) > 0$ can not happen.
    - σ-finite measure: $\mathcal{Z} = \dot\cup_{i \in I} P_i$, I: countable, $\nu(P_i) < \infty$.]
  - [$s_1, s_2 \in L^2(\mathcal{Z}, \nu)$.]

Two-sample problem: heteroscedastic regression model

- $N_1 = n_1$, $N_2 = n_2$: fixed,
- $Z^{(1)}$, $Z^{(2)}$ i.i.d.

$$Z_i^{(1)} = \left(X_i^{(1)}, Y_i^{(1)}\right) \qquad Y_i^{(1)} = s_1\left(X_i^{(1)}\right) + \sigma(X_i^{(1)})\xi_i^{(1)},$$
$$Z_i^{(2)} = \left(X_i^{(2)}, Y_i^{(2)}\right) \qquad Y_i^{(2)} = s_2\left(X_i^{(2)}\right) + \sigma(X_i^{(2)})\xi_i^{(2)},$$

$$\left(X_i^{(1)}, \xi_i^{(1)}\right) \stackrel{\text{distr}}{=} \left(X_i^{(2)}, \xi_i^{(2)}\right), \ \mathbb{E}\left[\xi_i^{(1)} | X_i^{(1)}\right] = 0, \ \mathbb{E}\left[\left(\xi_i^{(1)}\right)^2 | X_i^{(1)}\right] = 1,$$

Two-sample problem: heteroscedastic regression model

- $N_1 = n_1$, $N_2 = n_2$: fixed,
- $Z^{(1)}$, $Z^{(2)}$ i.i.d.

$$Z_i^{(1)} = \left(X_i^{(1)}, Y_i^{(1)}\right) \qquad Y_i^{(1)} = s_1\left(X_i^{(1)}\right) + \sigma(X_i^{(1)})\xi_i^{(1)},$$
$$Z_i^{(2)} = \left(X_i^{(2)}, Y_i^{(2)}\right) \qquad Y_i^{(2)} = s_2\left(X_i^{(2)}\right) + \sigma(X_i^{(2)})\xi_i^{(2)},$$

$$\left(X_i^{(1)}, \xi_i^{(1)}\right) \stackrel{\text{distr}}{=} \left(X_i^{(2)}, \xi_i^{(2)}\right),\ \mathbb{E}\left[\xi_i^{(1)}|X_i^{(1)}\right] = 0,\ \mathbb{E}\left[\left(\xi_i^{(1)}\right)^2|X_i^{(1)}\right] = 1,$$

- [$s_1, s_2, \sigma \in L^2(X, P_X)$, $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, $\mathcal{Y} \subseteq \mathbb{R}$.]

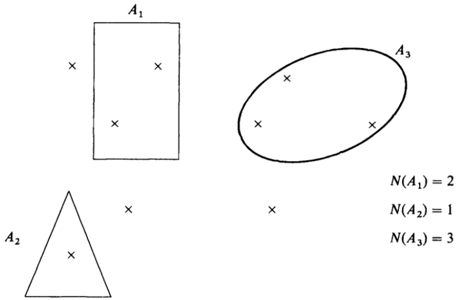
Two-sample problem: Poisson process model

- $X \sim \text{Poisson}[\lambda]: \mathbb{P}(X = n) = \frac{\lambda^k}{k!}e^{-\lambda},\ n = 0, 1, \ldots; \lambda \geq 0.$

Two-sample problem: Poisson process model

- $X \sim \text{Poisson}[\lambda] : \mathbb{P}(X = n) = \frac{\lambda^k}{k!} e^{-\lambda}, \; n = 0, 1, \ldots ; \lambda \geq 0.$
- # of raisins in a cake / stars on the sky:

Poisson process model



An N process producing random points in $(\mathcal{Z}, \mathcal{A}, m)$ is called PP if

1. $N(A) \sim \text{Poisson}[m(A)]$, for all $A \in \mathcal{A}$
2. $\{A_i\}_{i \in I}$ disjunct, I: countable $\Rightarrow \{N(A_i)\}_{i \in I}$ are independent.

Poisson process: it exists, construction

- If $\mathcal{Z}$ is finite ($m(\mathcal{Z}) < \infty$):
  - $Z_1, Z_2, \ldots$ i.i.d. , $\mathbb{P}(Z_i \in A) = \frac{m(A)}{m(\mathcal{Z})}$, $\forall A \in \mathcal{A}$,
  - $p \sim$ Poisson$[m(\mathcal{Z})]$, independent of X_i-s.
  - $N(A) = \sum_{i=1}^{p} \chi\{Z_i \in A\}$ is good!

Poisson process: it exists, construction

- If $\mathcal{Z}$ is finite ($m(\mathcal{Z}) < \infty$):
  - $Z_1, Z_2, \ldots$ i.i.d. , $\mathbb{P}(Z_i \in A) = \frac{m(A)}{m(\mathcal{Z})}$, $\forall A \in \mathcal{A}$,
  - $p \sim \text{Poisson}[m(\mathcal{Z})]$, independent of X_i-s.
  - $N(A) = \sum_{i=1}^{p} \chi\{Z_i \in A\}$ is good!
- If $\mathcal{Z}$ is σ-finite:

$$\mathcal{Z} = \dot{\cup}_{i \in I} P_i, \qquad P_i \leftrightarrow N_i, \qquad N = \sum_{i \in I} N_i.$$

Two-sample problem: Poisson process model

- $Z^{(1)} = \{Z_1^{(1)}, \ldots, Z_{(N_1)}^{(1)}\}$, $m := s_1$ intensity
- $Z^{(2)} = \{Z_1^{(2)}, \ldots, Z_{(N_2)}^{(2)}\}$, $m := s_2$
- N_1, N_2: Poisson r.v.-s.

Two-sample problem: Poisson process model

- $Z^{(1)} = \left\{Z_1^{(1)}, \ldots, Z_{(N_1)}^{(1)}\right\}$, $m := s_1$ intensity $[\ll \mu]$,
- $Z^{(2)} = \left\{Z_1^{(2)}, \ldots, Z_{(N_2)}^{(2)}\right\}$, $m := s_2$ $[\ll \mu]$,
- N_1, N_2: Poisson r.v.-s.
- [μ: non-atomic, σ-finite.]
- [$\mu = n\nu$; ν: non-atomic, σ-finite; ???]

After the examples, back to two-sampling (density model)

- Pooled samples: $Z = Z^{(1)} \cup Z^{(2)}$, $n = n_1 + n_2$. Test statistics:

$$T_K = \sum_{i \neq j \in \{1,\dots,n\}} K(Z_i, Z_j)\left(\epsilon_i^0 \epsilon_j^0 + c_{n_1,n_2}\right),$$

$$c_{n_1,n_2} = \frac{1}{n_1 n_2 (n_1 - 1 + n_2 - 1)},$$

$$a_{n_1,n_2} = \sqrt{\frac{1}{n_1(n_1-1)} - c_{n_1,n_2}},\ b_{n_1,n_2} = -a_{n_2,n_1},$$

$$\epsilon_i^0 = \begin{cases} a_{n_1,n_2} & Z_i \in Z^{(1)} \\ b_{n_1,n_2} & Z_i \in Z^{(2)}. \end{cases}$$

- Example: $K =$ reproducing kernel.

T_K makes sense

Let $(K \diamond p)(z) = \langle K(\cdot, z), p \rangle$, $\langle f, g \rangle = \int_{\mathcal{Z}} f(z) g(z) \mathrm{d}\nu(z)$. Unbiasedness:

$$\mathbb{E}_{s_1, s_2}[T_K] \overset{?}{=} \langle K \diamond (s_1 - s_2), s_1 - s_2 \rangle$$

T_K makes sense

Let $(K \diamond p)(z) = \langle K(\cdot, z), p \rangle$, $\langle f, g \rangle = \int_{\mathcal{Z}} f(z) g(z) \mathrm{d}\nu(z)$. Unbiasedness:

$$\mathbb{E}_{s_1, s_2}[T_K] \stackrel{?}{=} \langle K \diamond (s_1 - s_2), s_1 - s_2 \rangle$$

$$= \int_{\mathcal{Z}^2} K(z, z')(s_1 - s_2)(z)(s_1 - s_2)(z') \mathrm{d}\nu(z) \mathrm{d}\nu(z')$$

T_K makes sense

Let $(K \diamond p)(z) = \langle K(\cdot, z), p \rangle$, $\langle f, g \rangle = \int_{\mathcal{Z}} f(z) g(z) \mathrm{d}\nu(z)$. Unbiasedness:

$$\mathbb{E}_{s_1,s_2}[T_K] \stackrel{?}{=} \langle K \diamond (s_1 - s_2), s_1 - s_2 \rangle$$
$$= \int_{\mathcal{Z}^2} K(z, z')(s_1 - s_2)(z)(s_1 - s_2)(z') \mathrm{d}\nu(z) \mathrm{d}\nu(z')$$
$$= \sum_{\ell=1}^{2} \int_{\mathcal{Z}^2} K(z, z') s_\ell(z) s_\ell(z') \mathrm{d}\nu(z) \mathrm{d}\nu(z') - 2 \int_{\mathcal{Z}^2} K(z, z') s_1(z) s_2(z') \mathrm{d}\nu(z) \mathrm{d}\nu(z')$$

T_K makes sense

Let $(K \diamond p)(z) = \langle K(\cdot, z), p \rangle$, $\langle f, g \rangle = \int_{\mathcal{Z}} f(z) g(z) \mathrm{d}\nu(z)$. Unbiasedness:

$$\mathbb{E}_{s_1,s_2}[T_K] \stackrel{?}{=} \langle K \diamond (s_1 - s_2), s_1 - s_2 \rangle$$
$$= \int_{\mathcal{Z}^2} K(z, z')(s_1 - s_2)(z)(s_1 - s_2)(z') \mathrm{d}\nu(z) \mathrm{d}\nu(z')$$
$$= \sum_{\ell=1}^{2} \int_{\mathcal{Z}^2} K(z, z') s_\ell(z) s_\ell(z') \mathrm{d}\nu(z) \mathrm{d}\nu(z') - 2 \int_{\mathcal{Z}^2} K(z, z') s_1(z) s_2(z') \mathrm{d}\nu(z) \mathrm{d}\nu(z')$$
$$= \mathbb{E}_{s_1,s_2}\Big[\underbrace{\sum_{\ell=1}^{2} \sum_{i \neq j \in \{1,\ldots,n_\ell\}} K\left(Z_i^{(\ell)}, Z_j^{(\ell)}\right) c_{n_\ell} - 2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} K\left(Z_i^{(1)}, Z_j^{(2)}\right) \frac{1}{n_1 n_2}}_{=T_K} \Big],$$

where $c_{n_\ell} = \frac{1}{n_\ell (n_{\ell - 1})}$ $(\ell = 1, 2)$.

$\langle K \diamond (s_1 - s_2), s_1 - s_2 \rangle$ is finite

By CBS ($\|\cdot\| = \|\cdot\|_{L^2(\mathcal{Z},\nu)}$)

$$|\langle K \diamond (s_1 - s_2), s_1 - s_2 \rangle| \le \underbrace{\|K \diamond (s_1 - s_2)\|}_{=:(*)_1} \underbrace{\|s_1 - s_2\|}_{=:(*)_2},$$

where

- $(*)_1$: assumed to be finite,
- $(*)_2$: $< \infty$ since $s_1, s_2 \in L^2(\mathcal{Z}, \nu)$.

K-examples

If K is a reproducing kernel:

$$\mathbb{E}_{s_1,s_2}[T_K] = \int_{\mathcal{Z}^2} K(z,z')(s_1 - s_2)(z)(s_1 - s_2)(z')\mathrm{d}\nu(z)\mathrm{d}\nu(z')$$
$$= \left\| \int_{\mathcal{Z}} K(\cdot, z)(s_1 - s_2)(z)\mathrm{d}\nu(z) \right\|_{H(K)}^2$$
$$= \|\mu_{s_1} - \mu_{s_2}\|_{H(K)}^2, \quad \mu_s = \mu_{s,K}.$$

$\Rightarrow \mathbb{E}_{s_1,s_2}[T_K] = 0 \Leftrightarrow s_1 = s_2$, if K is characteristic.

K-examples: K is a projection kernel

Let $\{\phi_\lambda\}_{\lambda\in\Lambda}$ be an ONS.

$$\begin{aligned} K(z,z') &= \sum_{\lambda\in\Lambda}\phi_\lambda(z)\phi_\lambda(z'),\\ \int_{\mathcal{Z}} K(z,z')f(z')\mathrm{d}\nu(z') &= \mathrm{proj}_S(f)(z),\\ S &= \mathrm{span}\left(\{\phi_\lambda\}_{\lambda\in\Lambda}\right),\\ \langle K\diamond(s_1-s_2), s_1-s_2\rangle &= \left\|\mathrm{proj}_S(s_1-s_2)\right\|_{L^2(\mathcal{Z},\nu)}. \end{aligned}$$

Example: $\{\phi_\lambda\}_{\lambda\in\Lambda}$ Fourier/Haar basis.

K-examples: K is a convolution kernel

$\mathcal{Z} = \mathbb{R}^d$, $\nu =$Lebesgue measure, $k \in L^2\left(\mathbb{R}^d\right)$, $k(-x) = k(x)$, $h_i > 0$ $(\forall i)$.

$$K(z, z') = \frac{1}{\prod_{i=1}^d h_i} k\left(\frac{z_1 - z_1'}{h_1}, \ldots, \frac{z_d - z_d'}{h_d}\right),$$

$$\langle K \diamond (s_1 - s_2), s_1 - s_2 \rangle = \langle k_h * (s_1 - s_2), s_1 - s_2 \rangle_{L^2(\mathbb{R}^d)},$$

$$(f * g)(x) = \int_{\mathbb{R}^d} f(y) g(x - y) \mathrm{d}y,$$

$$k_h(u_1, \ldots, u_d) = \frac{1}{\prod_{i=1}^d h_i} k\left(\frac{u_1}{h_1}, \ldots, \frac{u_d}{h_d}\right).$$

Distribution of $T_K|Z$

- Recall: $T_K = \sum_{i \neq j \in \{1,\ldots,n\}} K(Z_i, Z_j) \left(\epsilon_i^0 \epsilon_j^0 + c_{n_1,n_2}\right)$,

$$\epsilon_i^0 = \begin{cases} a_{n_1,n_2} & i \in \{1, \ldots, n_1\} \\ b_{n_1,n_2} & i \in \{n_1 + 1, \ldots, n_1 + n_2\}. \end{cases}$$

- $R = (R_1, \ldots, R_n)$: rnd permutation, independent of Z.

$$\epsilon_i = \begin{cases} a_{n_1,n_2} & i \in \{R_1, \ldots, R_{n_1}\} \\ b_{n_1,n_2} & i \in \{R_{n_1+1}, \ldots, R_{n_1+n_2}\}. \end{cases},$$

$$T_K^{\epsilon} = \sum_{i \neq j \in \{1,\ldots,n\}} K(Z_i, Z_j) \left(\epsilon_i \epsilon_j + c_{n_1,n_2}\right).$$

Distribution of $T_K|Z$

- Recall: $T_K = \sum_{i\neq j\in\{1,\ldots,n\}} K(Z_i,Z_j)\left(\epsilon_i^0\epsilon_j^0 + c_{n_1,n_2}\right)$,

$$\epsilon_i^0 = \begin{cases} a_{n_1,n_2} & i\in\{1,\ldots,n_1\} \\ b_{n_1,n_2} & i\in\{n_1+1,\ldots,n_1+n_2\}. \end{cases}$$

- $R=(R_1,\ldots,R_n)$: rnd permutation, independent of Z.

$$\epsilon_i = \begin{cases} a_{n_1,n_2} & i\in\{R_1,\ldots,R_{n_1}\} \\ b_{n_1,n_2} & i\in\{R_{n_1+1},\ldots,R_{n_1+n_2}\}. \end{cases},$$

$$T_K^\epsilon = \sum_{i\neq j\in\{1,\ldots,n\}} K(Z_i,Z_j)\left(\epsilon_i\epsilon_j + c_{n_1,n_2}\right).$$

- $T_K|Z \overset{distr}{=} T_K^\epsilon|Z$.

Final test, level

- $q_{K,1-\alpha}^{(Z)}$: $(1-\alpha)-$quantile of $T_K^{\epsilon}|Z \stackrel{distr}{=} T_K|Z$.
- Reject H_0 if $T_K > q_{K,1-\alpha}^{(Z)}$.

Final test, level

- $q^{(Z)}_{K,1-\alpha}$: $(1-\alpha)-$quantile of $T^{\epsilon}_K|Z \overset{distr}{=} T_K|Z$.
- Reject H_0 if $T_K > q^{(Z)}_{K,1-\alpha}$.
- Given $\alpha \in (0,1)$, under H_0

$$\mathbb{P}_{s_1,s_2}\left(T_K > q^{(Z)}_{K,1-\alpha} \middle| Z \right) \leq \alpha.$$

Final test, level

- $q^{(Z)}_{K,1-\alpha}$: $(1-\alpha)-$quantile of $T^{\epsilon}_K|Z \stackrel{distr}{=} T_K|Z$.
- Reject H_0 if $T_K > q^{(Z)}_{K,1-\alpha}$.
- Given $\alpha \in (0,1)$, under H_0

$$\mathbb{P}_{s_1,s_2}\left(T_K > q^{(Z)}_{K,1-\alpha} \middle| Z \right) \leq \alpha.$$

- Taking expectation over Z, sup $[\mathbb{P}_{(H_0)}(A) = \sup_{(s_1,s_2),s_1=s_2} \mathbb{P}_{s_1,s_2}(A)]$:

$$\mathbb{P}_{(H_0)}(\Phi_{K,\alpha} = 1) \leq \alpha,$$

$$\Phi_{K,\alpha} = \chi\left\{ T_K > q^{(Z)}_{K,1-\alpha} \right\}.$$

Aggregated test

- Given:
  - $\{K_m\}_{m=1}^M$ kernels.
  - $\{w_m\}_{m=1}^M$ positive weights s.t. $\sum_m e^{-w_m} \leq 1$.

Aggregated test

- Given:
  - $\{K_m\}_{m=1}^M$ kernels.
  - $\{w_m\}_{m=1}^M$ positive weights s.t. $\sum_m e^{-w_m} \leq 1$.
- Notation: T_{K_m}, $T^{\epsilon}_{K_m}$; $q^{(Z)}_{m,1-u} = (1-u)$-quantile of $T^{\epsilon}_{K_m}|Z$.

Aggregated test

- Given:
  - $\{K_m\}_{m=1}^M$ kernels.
  - $\{w_m\}_{m=1}^M$ positive weights s.t. $\sum_m e^{-w_m} \le 1$.
- Notation: T_{K_m}, $T^{\epsilon}_{K_m}$; $q^{(Z)}_{m,1-u} = (1-u)$-quantile of $T^{\epsilon}_{K_m}|Z$.
- For $\alpha \in (0,1)$

$$u_{\alpha}^{(Z)} := \sup\left\{u > 0 : \mathbb{P}\Big(\underbrace{\sup_m \left(T^{\epsilon}_{K_m} - q^{(Z)}_{m,1-ue^{-w_m}}\right) > 0}_{\geq 1 \text{ test rejects}}\Big| Z\Big) \le \alpha\right\}.$$

Aggregated test

- Given:
  - $\{K_m\}_{m=1}^M$ kernels.
  - $\{w_m\}_{m=1}^M$ positive weights s.t. $\sum_m e^{-w_m} \leq 1$.
- Notation: T_{K_m}, $T_{K_m}^{\epsilon}$; $q_{m,1-u}^{(Z)} = (1-u)$-quantile of $T_{K_m}^{\epsilon}|Z$.
- For $\alpha \in (0,1)$

$$u_\alpha^{(Z)} := \sup\left\{u > 0 : \mathbb{P}\Big(\underbrace{\sup_m \left(T_{K_m}^{\epsilon} - q_{m,1-ue^{-w_m}}^{(Z)}\right) > 0}_{\geq 1 \text{ test rejects}}\Big|Z\Big) \leq \alpha\right\}.$$

- The test: $\exists m : T_{K_m}^{\epsilon} - q_{m,1-u_\alpha^{(Z)}e^{-w_m}}^{(Z)} > 0$; it is of level α.

Summary

- Focus: two-sample problem.
- Unbiasedness and exact quantile for the test statistics.
- Single kernel and aggregated tests of level α.