MMD Aggregated Two-Sample Test
KSD Aggregated Goodness-of-fit Test

University College London
Centre for Artificial Intelligence
Gatsby Computational Neuroscience Unit
Inria London Programme

Antonin Schrab

a.schrab@ucl.ac.uk
antoninschrab.github.io

MMD Aggregated Two-Sample Test






Antonin Schrab † ‡ §
Ilmun Kim ∗
Mlisande Albert ⋆
Batrice Laurent ⋆
Benjamin Guedj †§
Arthur Gretton ‡

† Centre for Artificial Intelligence, UCL
‡ Gatsby Computational Neuroscience Unit, UCL
§ Inria London Programme
∗ Department of Statistics & Data Science, Yonsei University
⋆ Institut de Mathmatiques, Universit de Toulouse

Two-sample problem

- samples $\mathbb{X}_m := (X_1, \ldots, X_m)$, $X_i \overset{\text{iid}}{\sim} p$ in $\mathbb{R}^d$
- samples $\mathbb{Y}_n := (Y_1, \ldots, Y_n)$, $Y_i \overset{\text{iid}}{\sim} q$ in $\mathbb{R}^d$

$$\begin{array}{ccc} \mathcal{H}_0\colon p = q & \text{against} & \mathcal{H}_a\colon p \neq q \\ \Delta(\mathbb{X}_m, \mathbb{Y}_n) = 1 & \iff & \text{reject } \mathcal{H}_0 \\ \Delta(\mathbb{X}_m, \mathbb{Y}_n) = 0 & \iff & \text{fail to reject } \mathcal{H}_0 \end{array}$$

Type I error: controlled by α by design

$$\mathbb{P}_{p\times p}(\Delta(\mathbb{X}_m, \mathbb{Y}_n) = 1) \ \leq \ \alpha$$

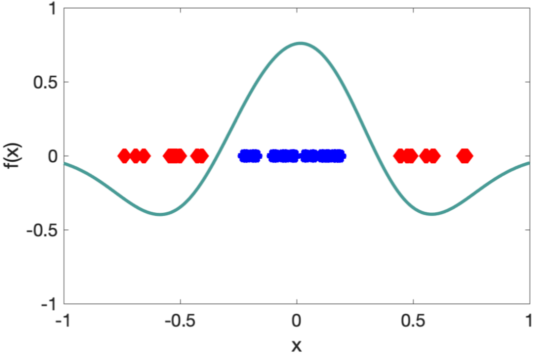
Type II error: find a condition on $\|p - q\|_2$ to control by β

$$\mathbb{P}_{p\times q}(\Delta(\mathbb{X}_m, \mathbb{Y}_n) = 0) \ \leq \ \beta$$

Two-sample test using the Maximum Mean Discrepancy

Kernel: $k_\lambda(x,y) := \prod_{i=1}^{d} \frac{1}{\lambda_i} K_i\left(\frac{x_i - y_i}{\lambda_i}\right)$ **Bandwidth:** $\lambda \in (0,\infty)^d$

$$\mathrm{MMD}_\lambda(p,q) := \sup_{f\in\mathcal{H}_\lambda : \|f\|_{\mathcal{H}_\lambda} \le 1} \left| \mathbb{E}_{X\sim p}[f(X)] - \mathbb{E}_{Y\sim q}[f(Y)] \right|$$



$p \neq q$

Our method: aggregate multiple tests with different **bandwidths**

Maximum Mean Discrepancy estimator

$$\mathrm{MMD}^2_\lambda(p,q) := \mathbb{E}_{p,p}[k_\lambda(X,X')] - 2\,\mathbb{E}_{p,q}[k_\lambda(X,Y)] + \mathbb{E}_{q,q}[k_\lambda(Y,Y')]$$

$$\widehat{\mathrm{MMD}}^2_\lambda(\mathbb{X}_m,\mathbb{Y}_n) := \frac{1}{m(m-1)}\sum_{1\le i\neq i'\le m} k_\lambda(X_i,X_{i'}) - \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} k_\lambda(X_i,Y_j) + \frac{1}{n(n-1)}\sum_{1\le j\neq j'\le n} k_\lambda(Y_j,Y_{j'})$$

MMD test for a fixed bandwidth λ

$$\Delta_\alpha^\lambda(\mathbb{X}_m, \mathbb{Y}_n) := \mathbb{1}\left(\widehat{\mathrm{MMD}}_\lambda^2(\mathbb{X}_m, \mathbb{Y}_n) > \widehat{q}_{1-\alpha}^\lambda\right)$$

Quantile: $\widehat{q}_{1-\alpha}^\lambda$ is the $\lceil (B+1)(1-\alpha) \rceil$-th largest value of $\widehat{\mathrm{MMD}}_\lambda^2(\mathbb{X}_m, \mathbb{Y}_n)$ and B permuted test statistics

$$\widehat{\mathrm{MMD}}_\lambda^2(\mathbb{X}_m^\sigma, \mathbb{Y}_n^\sigma) \quad \text{where} \quad (\mathbb{X}_m^\sigma, \mathbb{Y}_n^\sigma) = \sigma(\mathbb{X}_m \cup \mathbb{Y}_n)$$

Non-asymptotic level α

Time complexity:

$$\mathcal{O}\left(B\,(m+n)^2\right)$$

MMDAgg for a collection of bandwidths Λ

$$\Delta_\alpha^{\Lambda}(\mathbb{X}_m, \mathbb{Y}_n) := \mathbb{1}\left(\widehat{\mathrm{MMD}}^2_\lambda(\mathbb{X}_m, \mathbb{Y}_n) > \widehat{q}^{\lambda}_{1-u_\alpha w_\lambda} \text{ for some } \lambda \in \Lambda\right)$$

- positive weights $(w_\lambda)_{\lambda\in\Lambda}$ satisfying $\sum_{\lambda\in\Lambda} w_\lambda \leq 1$
- correction u_α defined as

$$\sup\left\{ u > 0 : \mathbb{P}_{p\times p}\left(\max_{\lambda\in\Lambda}\left(\widehat{\mathrm{MMD}}^2_\lambda(\mathbb{X}_m, \mathbb{Y}_n) - \widehat{q}^{\lambda}_{1-uw_\lambda}\right) > 0\right) \leq \alpha \right\}$$

Non-asymptotic level α

Time complexity:

$$\mathcal{O}\Big(|\Lambda|\,(B_1 + B_2)\,(m+n)^2\Big)$$

Minimax adaptivity over Sobolev balls

$$\mathcal{S}_d^s(R) := \left\{ f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d) : \int_{\mathbb{R}^d} \|\xi\|_2^{2s} |\widehat{f}(\xi)|^2 \,\mathrm{d}\xi \leq (2\pi)^d R^2 \right\}$$

Theorem

$$\Lambda^* := \left\{ 2^{-\ell} \mathbb{1}_d : \ell \in \left\{ 1, \dots, \left\lceil \frac{2}{d} \log_2 \left(\frac{m+n}{\ln(\ln(m+n))} \right) \right\rceil \right\} \right\}, \quad w_\lambda := \frac{6}{\pi^2 \ell^2}$$

Assuming $p - q \in \mathcal{S}_d^s(R)$, the condition

$$\|p - q\|_2 \geq C \left(\frac{m+n}{\ln(\ln(m+n))} \right)^{-2s/(4s+d)}$$

guarantees control over the probability of type II error of MMDAgg

$$\mathbb{P}_{p \times q}\left(\Delta_\alpha^{\Lambda^*}(\mathbb{X}_m, \mathbb{Y}_n) = 0 \right) \leq \beta.$$

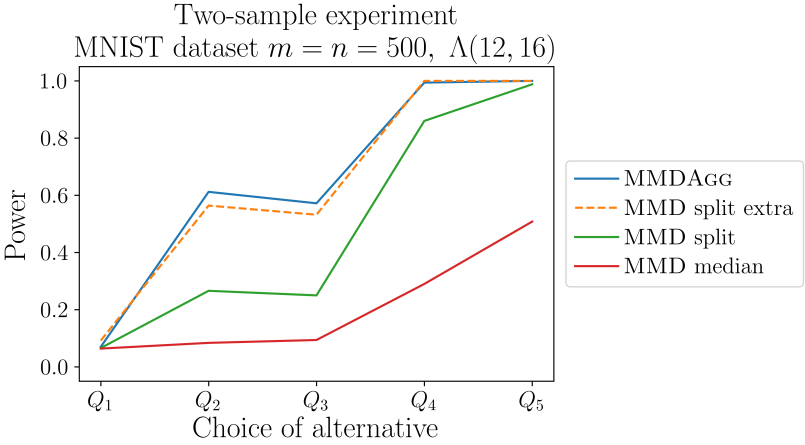
Minimax rate over Sobolev balls: $(m+n)^{-2s/(4s+d)}$

MMDAgg Experiment

$$\Lambda(\ell_-,\ell_+) := \left\{2^\ell \lambda_{med} : \ell \in \{\ell_-,\dots,\ell_+\}\right\} \qquad w_\lambda := 1 \,/\, |\Lambda|$$

$$\mathcal{P} := \{0,\dots,9\} \quad \mathcal{Q}_2 := \mathcal{P} \setminus \{8,6\} \quad \mathcal{Q}_4 := \mathcal{P} \setminus \{8,6,4,2\}$$

$$\mathcal{Q}_1 := \mathcal{P} \setminus \{8\} \quad \mathcal{Q}_3 := \mathcal{P} \setminus \{8,6,4\} \quad \mathcal{Q}_5 := \mathcal{P} \setminus \{8,6,4,2,0\}$$



KSD Aggregated Goodness-of-fit Test

Antonin Schrab
† ‡ §

Benjamin Guedj
†§

Arthur Gretton
‡

† Centre for Artificial Intelligence, UCL
‡ Gatsby Computational Neuroscience Unit, UCL
§ Inria London Programme

Goodness-of-fit problem & Kernel Stein Discrepancy

- model with probability density p or score function $\nabla \log p(z)$ on $\mathbb{R}^d$
- samples $\mathbb{Z}_n := (Z_1, \dots, Z_n)$, $Z_i \overset{\text{iid}}{\sim} q$ in $\mathbb{R}^d$

$$\mathcal{H}_0 \colon p = q \qquad \text{against} \qquad \mathcal{H}_a \colon p \neq q$$

Stein kernel: $h_{p,\lambda}(x, y)$ defined as

$$\big(\nabla \log p(x)^\top \nabla \log p(y)\big) k_\lambda(x, y) + \nabla \log p(y)^\top \nabla_1 k_\lambda(x, y) + \nabla \log p(x)^\top \nabla_2 k_\lambda(x, y) + \sum_{1 \leq i \leq d} \frac{\partial}{\partial x_i \, \partial y_i} k_\lambda(x, y)$$

Stein identity: $\mathbb{E}_p[h_{p,\lambda}(Z, \cdot)] = 0$

$$\mathrm{KSD}^2_{p,\lambda}(q) := \mathrm{MMD}^2_{h_{p,\lambda}}(p, q) = \mathbb{E}_{q,q}[h_{p,\lambda}(Z, Z')]$$

$$\widehat{\mathrm{KSD}}^2_{p,\lambda}(\mathbb{Z}_n) := \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} h_{p,\lambda}(Z_i, Z_j)$$

KSD test for a fixed bandwidth λ

$$\Delta_\alpha^\lambda(\mathbb{Z}_n) := \mathbb{1}\left(\widehat{\mathrm{KSD}}^2_{p,\lambda}(\mathbb{Z}_n) > \widehat{q}^{\lambda}_{1-\alpha}\right)$$

Quantile: $\widehat{q}^{\lambda}_{1-\alpha}$ is $\lceil B(1-\alpha)\rceil$-th largest of B bootstrap test statistics

Wild bootstrap: $\displaystyle\frac{1}{n(n-1)}\sum_{1\leq i\neq j\leq n} \epsilon_i\epsilon_j\, h_{p,\lambda}(Z_i, Z_j), \quad \epsilon_i \overset{\text{iid}}{\sim} \text{Unif}\{-1,1\}$

- asymptotic level α

Parametric bootstrap: $\displaystyle\frac{1}{N(N-1)}\sum_{1\leq i\neq j\leq N} h_{p,\lambda}(\widetilde{Z}_i, \widetilde{Z}_j), \qquad \widetilde{Z}_i \overset{\text{iid}}{\sim} p$

- non-asymptotic level α

Time complexity: $\mathcal{O}\left(Bn^2\right)$

KSDAgg for a collection of bandwidths Λ

$$\Delta_\alpha^{\Lambda}(\mathbb{Z}_n) := \mathbb{1}\left(\widehat{\mathrm{KSD}}^2_{p,\lambda}(\mathbb{Z}_n) > \widehat{q}^{\lambda}_{1-u_\alpha w_\lambda} \text{ for some } \lambda \in \Lambda\right)$$

- positive weights $(w_\lambda)_{\lambda\in\Lambda}$ satisfying $\sum_{\lambda\in\Lambda} w_\lambda \leq 1$
- correction u_α defined as

$$\sup\left\{u > 0 : \mathbb{P}_{p\times p}\left(\max_{\lambda\in\Lambda}\left(\widehat{\mathrm{KSD}}^2_{p,\lambda}(\mathbb{Z}_n) - \widehat{q}^{\lambda}_{1-uw_\lambda}\right) > 0\right) \leq \alpha\right\}$$

Wild bootstrap: asymptotic level α

Parametric bootstrap: non-asymptotic level α

Time complexity:

$$\mathcal{O}\left(|\Lambda|\,(B_1 + B_2)\, n^2\right)$$

Uniform separation rate

Integral transform: $(\kappa \diamond f)(y) := \int_{\mathbb{R}^d} \kappa(x,y) f(x) \,\mathrm{d}x$

Kernel assumption: $A_\lambda := \mathbb{E}_{q,q}\left[h_{p,\lambda}(Z,Z')^2\right] < \infty$

Theorem

The condition

$$\|p-q\|_2^2 \geq \min_{\lambda\in\Lambda}\left(\left\|(p-q) - h_{p,\lambda}\diamond(p-q)\right\|_2^2 + C \ln\left(\frac{1}{\alpha w_\lambda}\right)\frac{\sqrt{A_\lambda}}{\beta n}\right)$$

guarantees control over the probability of type II error of KSDAgg

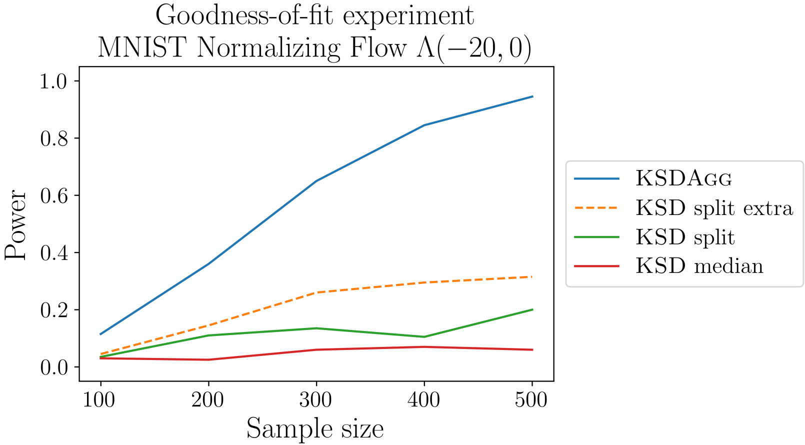
$$\mathbb{P}_q\left(\Delta^\Lambda_{\alpha,p}(\mathbb{Z}_n) = 0\right) \leq \beta.$$

KSDAgg Experiment

$$\Lambda(\ell_-, \ell_+) := \left\{ 2^\ell \lambda_{med} : \ell \in \{\ell_-, \ldots, \ell_+\} \right\} \qquad w_\lambda := 1 \,/\, |\Lambda|$$

model: Normalizing Flow density

samples: true MNIST digits



Conclusion: MMDAgg & KSDAgg

MMDAgg & KSDAgg tests:

- aggregate MMD/KSD tests with different kernel bandwidths (or kernels)
- avoids using arbitrary heuristics or data splitting
- wide range of kernels

MMDAgg theoretical results:

- optimal in the minimax sense (up to $\log(\log(m+n))$ term)
- adaptive test over Sobolev balls $\{\mathcal{S}_d^s(R) : s > 0, R > 0\}$
- quantile estimation: wild bootstrap or permutations

KSDAgg theoretical results:

- uniform separation rate upper bound
- quantile estimation: wild bootstrap or parametric bootstrap

MMDAgg & KSDAgg experimental results:

- outperforms state-of-the-art MMD/KSD adaptive tests

Thank you for your attention!

MMDAgg

SCAN ME

paper

SCAN ME

code

KSDAgg

SCAN ME

paper

SCAN ME

code