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 ∗	Mélisande Albert ⋆	Béatrice 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 Mathématiques, Université 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 $\mathbb{R}^d$

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

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:** $\lambda \in (0,\infty)^d$

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

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

Choice of **bandwidth** is **crucial** for test power!

Bandwidth selection methods: **median heuristic** & **data splitting**

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

MMDAgg for a collection of bandwidths Λ

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

- quantile $\widehat{q}^\lambda$ estimated using B_1 permuted test statistics
- positive weights $(w_\lambda)_{\lambda\in\Lambda}$ satisfying $\sum_{\lambda\in\Lambda} w_\lambda \leq 1$
- correction u_α defined as

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

- $\mathbb{P}_{p\times p}$ is estimated using B_2 permuted test statistics

Non-asymptotic level α

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

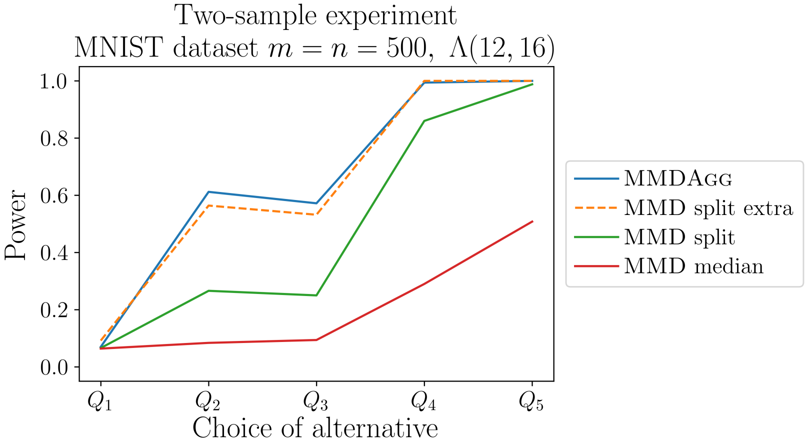
Power guarantees: minimax optimal & adaptive over Sobolev balls

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

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

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

$$\mathrm{KSD}^2_{p,\lambda}(q) := \mathrm{MMD}^2_{h_{p,\lambda}}(p, q) = \mathbb{E}_{q,q}[h_{p,\lambda}(Z, Z')]$$
$$\widehat{\mathrm{KSD}}^2_{p,\lambda}(\mathbb{Z}_n) := \frac{1}{n(n-1)} \sum_{1 \le i \neq j \le n} h_{p,\lambda}(Z_i, Z_j)$$

KSDAgg for a collection of bandwidths Λ

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

- quantile $\widehat{q}^{\lambda}$ estimated using B_1 bootstrapped test statistics
- positive weights $(w_\lambda)_{\lambda\in\Lambda}$ satisfying $\sum_{\lambda\in\Lambda} w_\lambda \leq 1$
- correction u_α defined as

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

- $\mathbb{P}_{p\times p}$ is estimated using B_2 bootstrapped test statistics

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

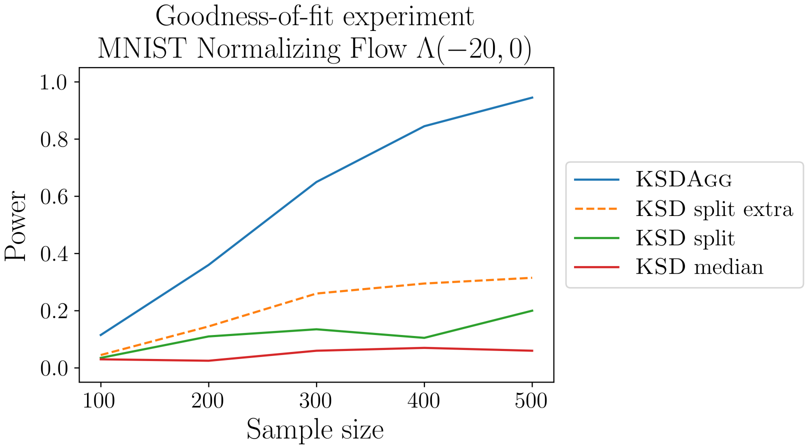
Power guarantees: upper bound on uniform separation rates

KSDAgg Experiment

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

model: Normalizing Flow density

samples: true MNIST digits



Thank you for your attention!

MMDAgg

SCAN ME

paper

SCAN ME

code

KSDAgg

SCAN ME

paper

SCAN ME

code