Efficient Aggregated Kernel Tests using Incomplete U-statistics

Antonin Schrab, Ilmun Kim,
Benjamin Guedj, Arthur Gretton

NEURAL INFORMATION PROCESSING SYSTEMS

University College London
Inria London
Yonsei University

Overview

Three unified testing frameworks:

- Two-sample testing
  - Given samples $X_1, \ldots, X_N \sim p$ & $Y_1, \ldots, Y_N \sim q$, is $p = q$?
  - **MMD:** Maximum Mean Discrepancy $\mathbb{E}\big[h_k^{\mathrm{MMD}}\big]$
- Independence testing
  - Given pairs of samples $\big(X_1, Y_1\big), \ldots, \big(X_N, Y_N\big) \sim r$, is $r = p \otimes q$?
  - **HSIC:** Hilbert Schmidt Independence Criterion $\mathbb{E}\big[h_{k,\ell}^{\mathrm{HSIC}}\big]$
- Goodness-of-fit testing
  - Given a model score $\nabla \log p(\cdot)$ & $Y_1, \ldots, Y_N \sim q$, is $p = q$?
  - **KSD:** Kernel Stein Discrepancy $\mathbb{E}\big[h_k^{\mathrm{KSD}}\big]$

Adaptive tests: aggregation over a large collection of kernels without paying a significant cost in power

Efficient tests: based on incomplete U-statistics to estimate $\mathbb{E}[h_k]$ by

$$\frac{1}{\big|\mathcal{D}_N\big|} \sum_{(i,j)\in\mathcal{D}_N} h_k(Z_i, Z_j), \qquad \mathcal{D}_N \subseteq \big\{(i,j)\big\}_{1\leq i\neq j\leq N^2}$$

Power guarantees

Sobolev ball: function space with unknown smoothness parameter s

Regularity assumption:

- **MMDAggInc:** $p - q$ lies in a Sobolev ball
- **HSICAggInc:** $r - p \otimes q$ lies in a Sobolev ball

Uniform separation rates: high test power is guaranteed for

- **MMDAggInc** when $\|p - q\|_2$ is larger than:
- **HSICAggInc** when $\|r - p \otimes q\|_2$ is larger than:

$$\left(\frac{|\mathcal{D}_N|/N}{\ln(\ln(|\mathcal{D}_N|/N))}\right)^{-2s/(4s+d)}$$

KSDAggInc: general power guarantees with no regularity assumption

Uniform separation rate analysis

$$\left(\frac{|\mathcal{D}_N|/N}{\ln(\ln(|\mathcal{D}_N|/N))}\right)^{-2s/(4s+d)}.$$

- Regime $|\mathcal{D}_N| \asymp N^2$:
  - quadratic-time test
  - uniform separation rate $\left(\frac{N}{\ln\ln N}\right)^{-2s/(4s+d)}$.
  - minimax rate over Sobolev balls, up to $\ln\ln N$ term
  - adaptive to unknown smoothness parameter s of Sobolev ball

- Regime $N \lesssim |\mathcal{D}_N| \lesssim N^2$:
  - trade-off between efficiency and rate of convergence
  - incur cost $(N^2/|\mathcal{D}_N|)^{2s/(4s+d)}$ in the minimax rate
  - rate deteriorates from quadratic (minimax) to linear (no guarantee)

- Regime $|\mathcal{D}_N| \lesssim N$:
  - no guarantee that rate converges to 0

AggInc: summary

- **Time complexity:** chosen by the user (trade-off with power)
- **Experiments:** consider linear-time AggInc tests
- **MMDAggInc & HSICAggInc:** outperform the current two-sample and independence state-of-the-art linear-time tests
- **KSDAggInc:** matches the power obtained by the linear-time state-of-the-art Cauchy RFF test of **Huggins and Mackey**

  Huggins and Mackey. **Random feature Stein discrepancies**. NeurIPS 2018.
- **Aggregated quadratic-time tests:**

  MMDAgg: Schrab et al. **MMD Aggregated Two-sample Test**. 2021.

  HSICAgg: Albert et al. **Adaptive test of independence based on HSIC measures**. The Annals of Statistics 2022.

  KSDAgg: Schrab et al. **KSD Aggregated Goodness-of-fit Test**. NeurIPS 2022.