Optimally-weighted estimators of the maximum mean discrepancy for likelihood-free inference

Ayush Bharti, Masha Naslidnyk, Oscar Key, Samuel Kaski, François-Xavier Briol

Introduction

- Simulators are mechanistic models of real-world phenomenon that are used widely in many domains of science, medicine and engineering. However, their likelihood functions are intractable.
- Likelihood-free inference (LFI) methods, such as approximate Bayesian computation and minimum distance estimation, rely on computing distances between observed and simulated data.
- Maximum mean discrepancy (MMD) is a popular choice of distance:
  - Desirable properties: leads to consistent and robust estimators
  - Applicable on any data type for which a kernel can be defined
  - Sample complexity of $\mathcal{O}(m^{-1/2})$, m being no. of simulated samples
- **Challenge in LFI:** Large computational cost of simulating data
- **Contribution:** MMD estimator with improved sample complexity

Methodology

Background

- Data $\{x_i\}_{i=1}^n \subseteq \mathcal{X} \subseteq \mathbb{R}^d$ denoted by empirical distribution $\mathbb{Q}^n$
- Simulator $\mathcal{P}_\theta = \{\mathbb{P}_\theta : \theta \in \Theta\}$, characterised by $(G_\theta, \mathbb{U})$
- To sample $y \sim \mathbb{P}_\theta$, (1) sample $u \sim \mathbb{U}$, from base space $\mathcal{U} \subset \mathbb{R}^s$, (2) apply the generator $y = G_\theta(u)$
- **Task:** Estimate θ given $\mathbb{Q}^n$
- The MMD between $\mathbb{P}$ and $\mathbb{Q}$ is the distance between their embeddings in RKHS $\mathcal{H}_k$ associated with kernel k:

$$\mathrm{MMD}_k^2(\mathbb{P}_\theta, \mathbb{Q}) = \mathbb{E}_{y,y' \sim \mathbb{P}_\theta}[k(y,y')] - 2\mathbb{E}_{y \sim \mathbb{P}_\theta, x \sim \mathbb{Q}}[k(x,y)] + \mathbb{E}_{x,x' \sim \mathbb{Q}}[k(x,x')]$$

Optimally-weighted MMD estimator

We use Bayesian quadrature weights to estimate integrals wrt simulator.

$$\mathrm{MMD}_k^2(\mathbb{P}_\theta^{m,w}, \mathbb{Q}^n) = \sum_{i,j=1}^{m} w_i w_j k(y_i, y_j) - \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} w_j k(x_i, y_j) + \frac{1}{n^2}\sum_{i,j=1}^{n} k(x_i, x_j)$$

Let $\mathbb{P}_\theta^{m,w} = \sum_{i=1}^m w_i \delta_{y_i} = \sum_{i=1}^m w_i \delta_{G_\theta(u_i)}$.

Using the reverse triangle inequality, we get

$$\left|\mathrm{MMD}_k(\mathbb{P}_\theta, \mathbb{Q}) - \mathrm{MMD}_k(\mathbb{P}_\theta^{m,w}, \mathbb{Q})\right| \leq \mathrm{MMD}_k(\mathbb{P}_\theta, \mathbb{P}_\theta^{m,w}).$$

Let $c : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ be a reproducing kernel of order ν_c s.t. $k(x, \cdot) \circ G_\theta \in \mathcal{H}_c$.

Then, the error upper bound can be written as:

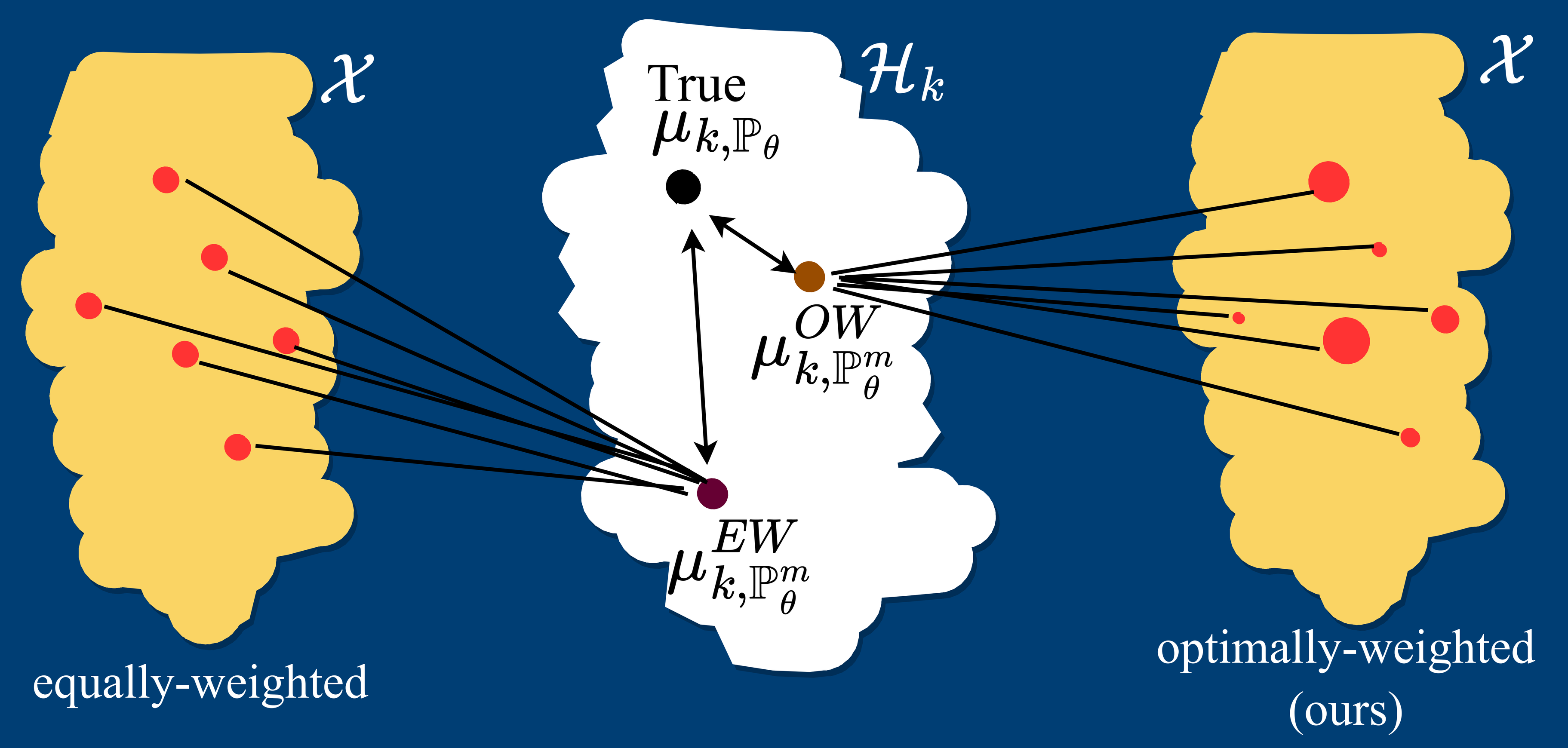
$$\mathrm{MMD}_k(\mathbb{P}_\theta, \mathbb{P}_\theta^{m,w}) = K \times \mathrm{MMD}_c\left(\mathbb{U}, \sum_{i=1}^m w_i \delta_{u_i}\right).$$

The weights minimising this upper bound are obtained in closed-form:

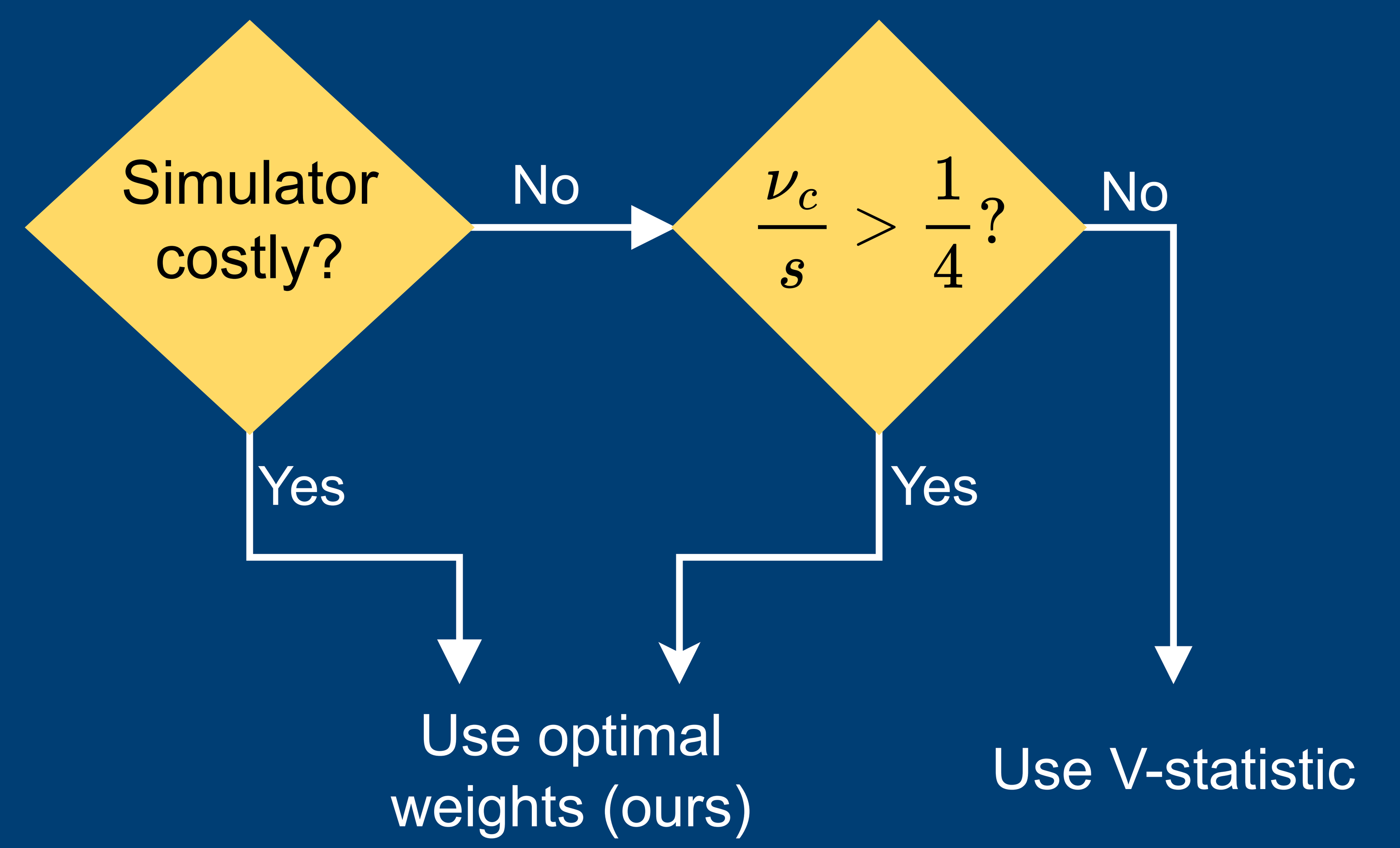
$$w^\star = \underset{w \in \mathbb{R}^m}{\arg\min}\, \mathrm{MMD}_c\left(\mathbb{U}, \sum_{i=1}^m w_i \delta_{u_i}\right) = \underset{w \in \mathbb{R}^m}{\arg\min}\, \mathrm{MMD}_c^2\left(\mathbb{U}, \sum_{i=1}^m w_i \delta_{u_i}\right)$$

Sample complexity: $\mathcal{O}(m^{-\nu_c/s - 1/2})$ $\Rightarrow$ improved rate for any ν_c and s!

We enable **scalable** inference for **computationally expensive** scientific **simulator-based models.**



When to use our method:

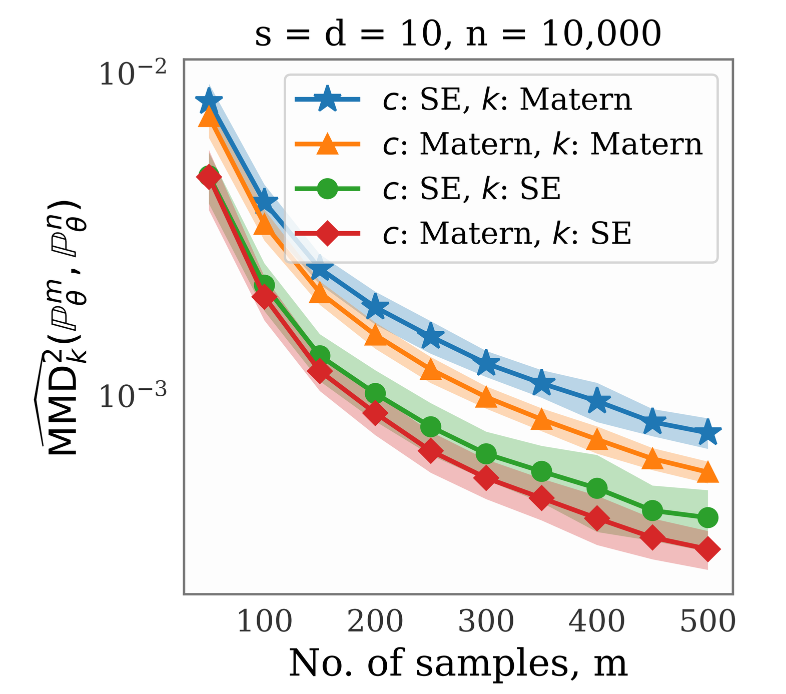


Take a picture to download the full paper

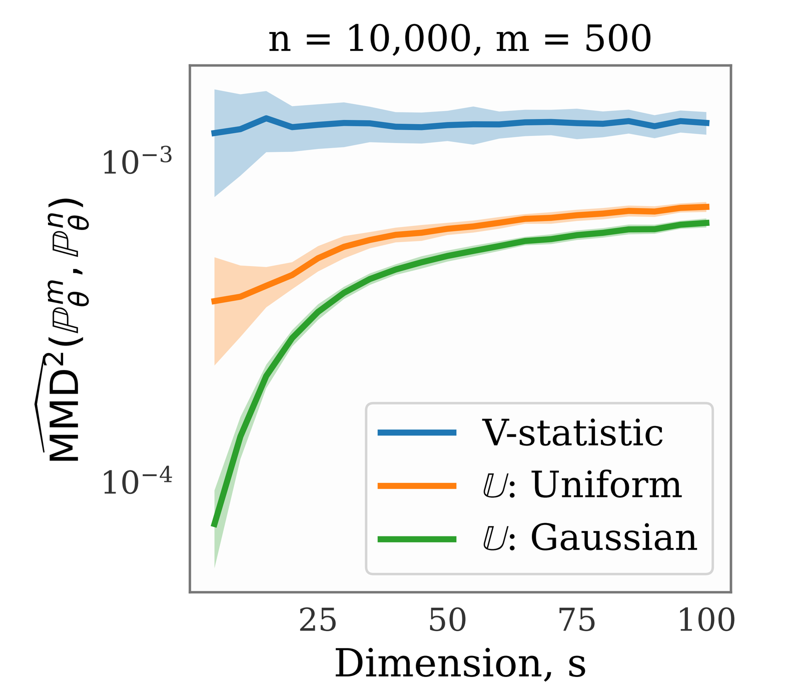
Benchmarking on popular simulators:

Model	s	d	IID V-stat	IID OW (ours)
g-and-k	1	1	2.25 (1.52)	**0.086** (0.049)
Two moons	2	2	2.36 (1.94)	**0.057** (0.054)
Bivariate Beta	5	2	2.13 (1.17)	**0.555** (0.227)
MA(2)	12	10	2.42 (0.796)	**0.705** (0.107)
M/G/1 queue	10	5	2.52 (1.19)	**1.71** (0.568)
Lotka-Volterra	600	2	2.13 (1.10)	**2.04** (0.956)

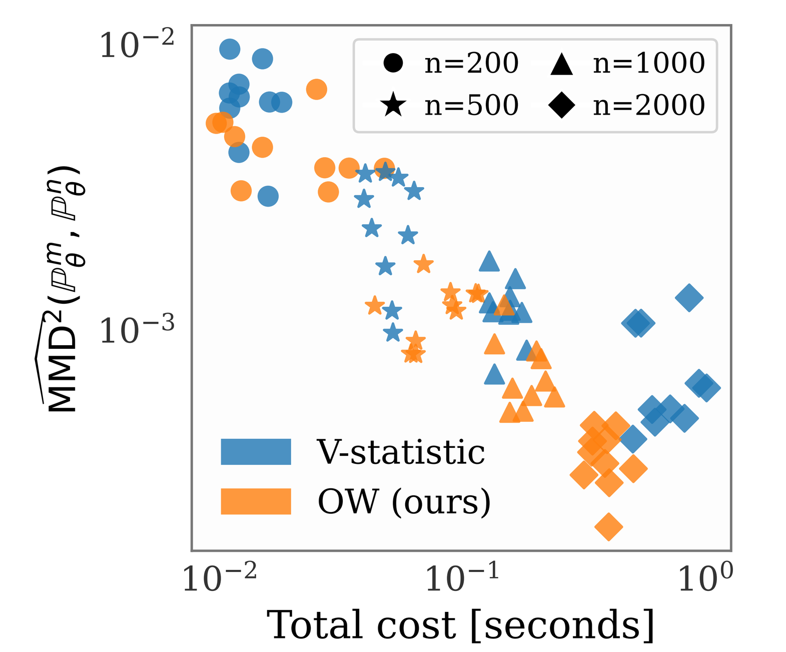
Varying choice of k and c:



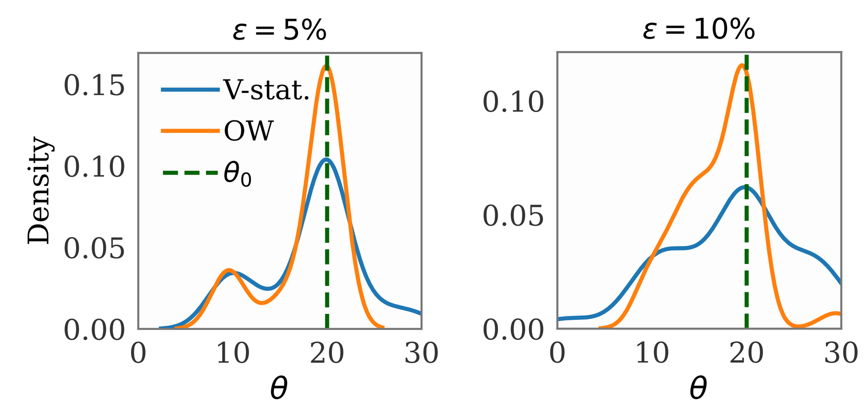
Varying dimensions s and d:



Performance vs. computational cost



Wind farm experiment:



Aalto University School of Science
UCL