Discrepancy-based Inference for Intractable Generative Models using Quasi-Monte Carlo

Ziang Niu, Johanna Meier, François-Xavier Briol

ziangniu@sas.upenn.edu

Penn
University *of* Pennsylvania

Intractable Generative Models

- Likelihood function is intractable.
- Sampling is possible.
- $u_i \sim \mathcal{U}([0,1]^s), G_\theta(u_i) \sim \mathbb{P}_\theta$.
- Examples include VAE, GANs.

Minimum Distance Estimators (MDE)

- $\{y_j\}_{j=1}^m \overset{IID}{\sim} \mathbb{Q} \in \mathcal{P}(\mathcal{X})$.
- one can construct an estimator through the framework of MDE:

$$\widehat{\theta}_m^D = \arg\min_{\theta \in \Theta} D(\mathbb{P}_\theta, \mathbb{Q}^m)$$

  where $\mathbb{Q}^m = \frac{1}{m} \sum_{j=1}^m \delta_{y_j}(x)$.
- $\mathbb{P}_\theta$ is unknown. Require a good approximation $D(\mathbb{P}_\theta^n, \mathbb{Q}^m)$.

UNIVERSITY of PENNSYLVANIA

Integral Probability Metrics (IPMs)

- An IPM is a probability metric which takes the form:

$$D_{\mathcal{F}}(\mathbb{P}, \mathbb{Q}) := \sup_{f \in \mathcal{F}} \left| \int_{\mathcal{X}} f(x)\mathbb{P}(\mathrm{d}x) - \int_{\mathcal{X}} f(x)\mathbb{Q}(\mathrm{d}x) \right|$$

- Popular metrics include:
  - Maximum Mean Discrepancy (MMD)
  - Wasserstein Distance
- Other popular divergences (not in IPMs) include:
  - Sinkhorn divergence $S_{c,p,\lambda}$
  - Sliced Wasserstein Distance.

Penn
UNIVERSITY of PENNSYLVANIA

Sample Complexity

Consider D is a metric.

- We want $D(\mathbb{P}_\theta, \mathbb{Q}^m)$ but what we can get is $D(\mathbb{P}_\theta^n, \mathbb{Q}^m)$.
- We could like make $\left|D(\mathbb{P}_\theta^n, \mathbb{Q}^m) - D(\mathbb{P}_\theta, \mathbb{Q})\right|$ to be as small as possible.
- By basic concentration inequality, we know

$$|D(\mathbb{P}_\theta^n, \mathbb{Q}^m) - D(\mathbb{P}_\theta, \mathbb{Q}^m)| \leq D(\mathbb{P}_\theta^n, \mathbb{P}_\theta)$$

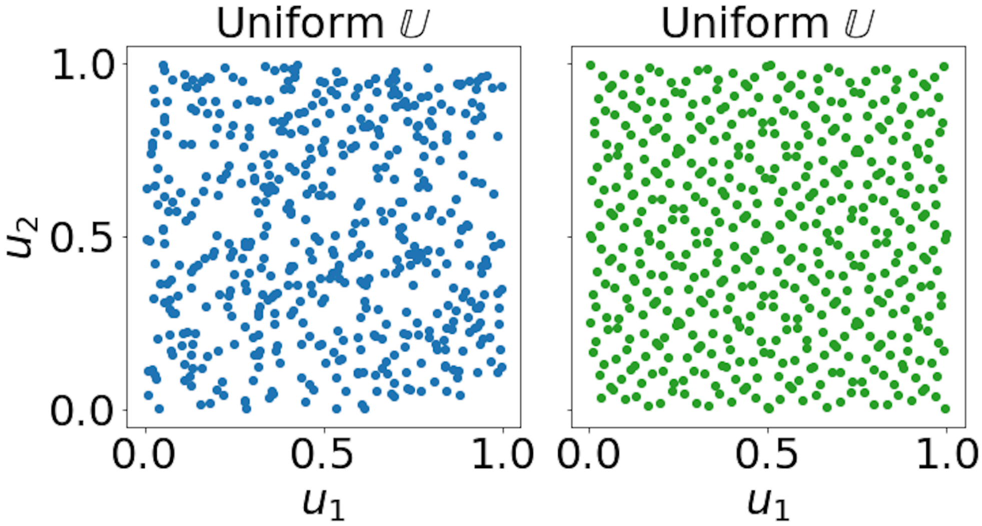
  Sample complexity $D(\mathbb{P}_\theta^n, \mathbb{P}_\theta)$ plays a key role here!

- Issue with Previous Method: Monte Carlo point set only guarantees $D(\mathbb{P}_\theta^n, \mathbb{P}_\theta) = O_p(n^{-1/2})$.

Penn
University *of* Pennsylvania

Enhancing Sample Complexity via Quasi-Monte Carlo

- QMC: generate a more ”diverse” set of samples from the model.



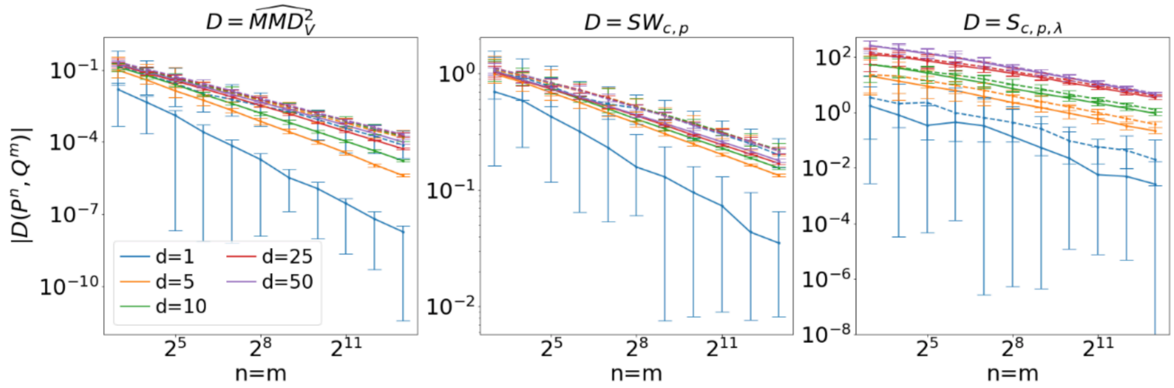
- **IDEA:** *Replace MC points to estimate discrepancies with QMC/RQMC points.*

Numerical Results: Inference for Multivariate g-and-k Models

The generator for g-and-k model is

$$G_\theta(u) := \theta_1 + \theta_2 \left(1 + 0.8\frac{(1-\exp(-\theta_3 z))}{(1+\exp(-\theta_3 z))}\right)(1+z^2)^{\theta_4} z$$

where $z = \Sigma^{\frac{1}{2}}\Phi^{-1}(u)^\top, u \sim \text{Unif}([0,1]^d)$. Σ is a symmetric Toepliz matrix with diagonal entries equal to 1 and subdiagonals equal to θ_5 and Φ^{-1} is the inverse CDF of Gaussian.



More Applications: generative neural network

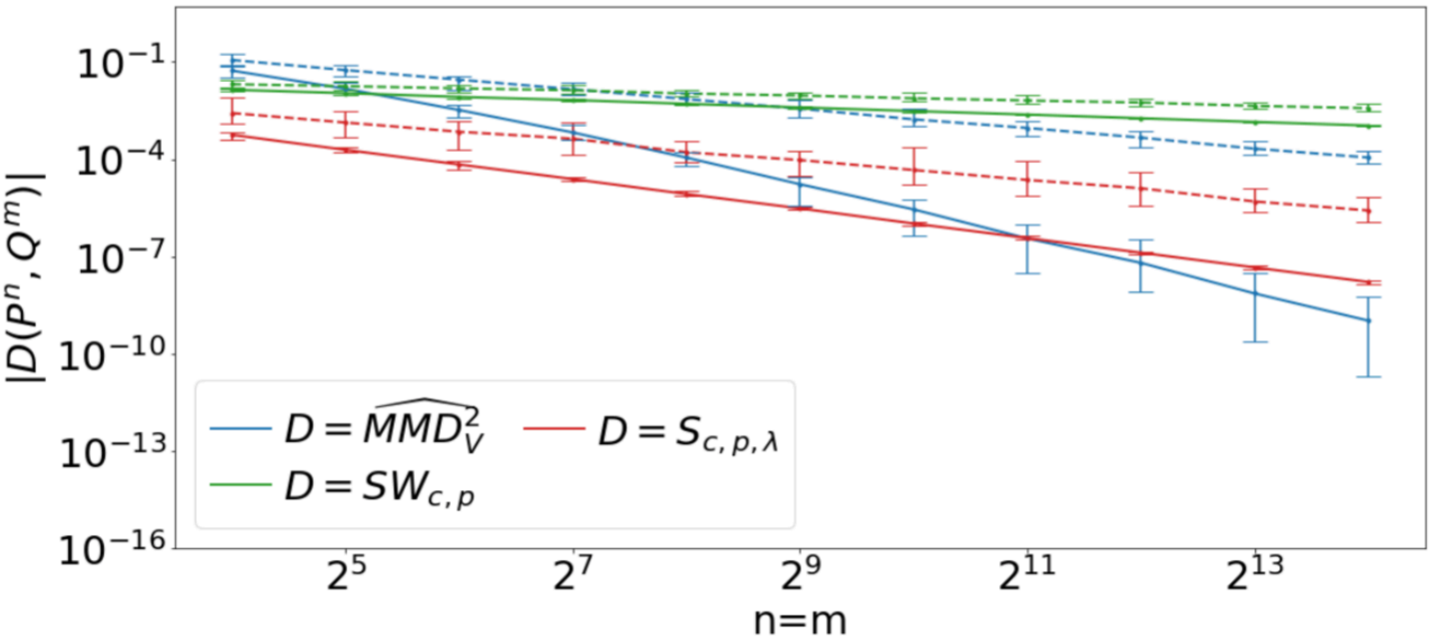
- Generative models widely used in modern machine learning are parametrized by neural network.
- Consider $G_\theta : \mathcal{U} \to \mathcal{X}$ with $\mathcal{U} = [0,1]^2$ and $\mathcal{X} = [0,1]^{784}$ (i.e. $s = 2$ and $d = 784$) of the form:

$$G_\theta(u) = \phi_2(\phi_1(\phi_1(u^\top W^1 + b^1)^\top W^2 + b^2)^\top W^3 + b^3)$$

where θ is a parameter vector containing all entries of the weight matrices and

- $\phi_1(x) = \log(\exp(x) + 1)$ (a softplus activation function)
- $\phi_2(x) = (1 + \exp(-x))^{-1}$ (a logistic activation function)

More Applications: generative neural network



Theory: Assumptions

Assumption 1(QMC points set)

Given a model $\mathbb{P}_\theta$ with generative process $(\text{Unif}([0,1]^s), G_\theta)$, we assume we have access to $x_i = G_\theta(u_i)$ for $i = 1, \ldots, n$ where $\{u_i\}_{i=1}^n \subset [0,1]^s$ form a QMC or RQMC point set for some $\alpha_s > 0$. Furthermore, we write $\mathbb{P}_\theta^n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$.

Assumption 2(Smoothness condition)

Assume that the domain $\mathcal{X} \subset \mathbb{R}^d$ is a compact space and that the generator is a map $G_\theta : [0,1]^s \to \mathcal{X}$ where:

1. $\partial^{(1,\ldots,1)}(G_\theta)_j \in \mathcal{C}([0,1]^s)$ for all $j = 1, \ldots, d$.
2. $\partial^v(G_\theta)_j(\cdot : 1_{-v}) \in L^{p_j}([0,1]^{|v|})$ for all $j = 1, \ldots, d$ and $v \in \{0,1\}^s \setminus (0, \ldots, 0)$, where $p_j \in [1, \infty]$ and $\sum_{j=1}^d p_j^{-1} \leq 1$.

UNIVERSITY of PENNSYLVANIA

Theory: Results

Theorem

Let $k \in \mathcal{C}^{s,s}(\mathcal{X})$, $\mathbb{P}_\theta \in \mathcal{P}_k$ and suppose Assumptions 1-2 hold. Then,

$$\mathbf{MMD}(\mathbb{P}_\theta, \mathbb{P}_\theta^n) = O(n^{-1}(\log n)^{\alpha_s}).$$

Corollary

Suppose the conditions in above Theorem hold. Then,

$$|\mathbf{MMD}(\mathbb{P}_\theta, \mathbb{Q}^m) - \mathbf{MMD}(\mathbb{P}_\theta^n, \mathbb{Q}^m)| = O(n^{-1}(\log n)^{\alpha_s}).$$

Penn
UNIVERSITY *of* PENNSYLVANIA

More Details

Discrepancy-based Inference for Intractable Generative Models using Quasi-Monte Carlo

Ziang Niu[1,*], Johanna Meier[2,*], François-Xavier Briol[3,†]

[1]Renmin University of China, [2]Leibniz Universität Hannover, [3]University College London,
[*]contributed equally, [†]corresponding author.

ArXiv Link: https://arxiv.org/pdf/2106.11561.pdf