# Differentially Private Equivalence Testing for Continuous Distributions and Applications

Daniel Omer     Or Sheffet
Bar-Ilan University

## Problem

Given a family of continues distribution and some $\alpha > 0$, and sample access to an two unknown continues distributions $p$ and $q$, **How many samples** $m(\alpha, \epsilon, k, \delta)$ are required to distinguish between $p = q$ (Case 1) versus $\|p - q\|_{\mathcal{A}_k} \geq \alpha$ (Case 2) while preserving $(\epsilon, \delta)$-DP

## Definition

we define the $\mathcal{A}_k$-distance between $p$ and $q$ by $\|\mathcal{P} - \mathcal{Q}\|_{\mathcal{A}_k} = \sup_{\mathcal{I}} \sum_{j=1}^{k} |\mathcal{P}(I_j) - \mathcal{Q}(I_j)|$ where $\mathcal{I}$ is a partition of the real-line $\mathbb{R}$ into $k$ intervals $I_1, I_2, ..., I_k$

# Differential Privacy

## Definition

(Differential Privacy [DMNS06]). A randomized algorithm $\mathcal{A}$ with domain $\mathcal{U}$ is $(\epsilon, \delta)-$ differentially private if for all $\mathcal{S} \subseteq \text{Range}(\mathcal{A})$, and for all $D, D' \in \mathcal{U}$ such that $D, D'$ differ on a single entry:

$$\Pr[\mathcal{A}(D) \in \mathcal{S}] \leq e^{\epsilon} \Pr[\mathcal{A}(D') \in \mathcal{S}] + \delta$$

If $\delta = 0$, we say that $\mathcal{A}$ is $\epsilon-$differentially private.

## Theorem (Non-private)[DKN15]

For the non-private case, the number of samples is

$$O\left(\frac{k^{4/5}}{\alpha^{6/5}} + \frac{\sqrt{k}}{\alpha^2}\right)$$

to test whether $\mathcal{P} = \mathcal{Q}$ or $\|\mathcal{P} - \mathcal{Q}\|_{\mathcal{A}_k} \geq \alpha$

# Our Contribution

Our result is to show an upper bound for the private case

**Theorem**

For the private case, the number of samples is

$$\tilde{O}\left(\max\left\{k^{4/5}/\alpha^{6/5}, k^{2/3}/\alpha\epsilon^{1/3}, k^{1/2}/\alpha^{2}, k^{1/3}/\alpha^{4/3}\epsilon^{2/3}, \sqrt{k}/\alpha\epsilon\right\}\right)$$

to a $(\epsilon, \delta) - DP$ algorithm test whether $\mathcal{P} = \mathcal{Q}$ or $\|\mathcal{P} - \mathcal{Q}\|_{\mathcal{A}_k} \geq \alpha$

# More Result

| **Distrib. Family** | **Num of Intervals** | **Private upper bound** |
|---|---|---|
| $t$-piecewise constant | $t$ | $\tilde{O}\left(\max\left\{\frac{t^{4/5}}{\alpha^{6/5}}, \frac{t^{2/3}}{\alpha\epsilon^{1/3}}, \frac{t^{1/2}}{\alpha^2}, \frac{t^{1/3}}{\alpha^{4/3}\epsilon^{2/3}}, \frac{\sqrt{t}}{\alpha\epsilon}\right\}\right)$ |
| $t$-piecewise degree-$d$ | $t(d+1)$ | $\tilde{O}\left(\max\left\{\frac{(t(d+1))^{4/5}}{\alpha^{6/5}}, \frac{(t(d+1))^{2/3}}{\alpha\epsilon^{1/3}}, \frac{(t(d+1))^{1/2}}{\alpha^2}, \frac{(t(d+1))^{1/3}}{\alpha^{4/3}\epsilon^{2/3}}, \frac{\sqrt{(t(d+1))}}{\alpha\epsilon}\right\}\right)$ |
| log-concave | $\frac{1}{\sqrt{\alpha}}$ | $\tilde{O}\left(\max\left\{\frac{1}{\alpha^{9/5}}, \frac{1}{\alpha^{4/3}\epsilon^{1/3}}, \frac{1}{\alpha^{3}\epsilon^{2/3}}, \frac{1}{\alpha^{5/4}\epsilon}\right\}\right)$ |
| $k$-mixture of log-concave | $\frac{k}{\sqrt{\alpha}}$ | $\tilde{O}\left(\max\left\{\frac{t^{4/5}}{\alpha^{8/5}}, \frac{k^{2/3}}{\alpha^{4/3}\epsilon^{1/3}}, \frac{k^{1/2}}{\alpha^{9/5}}, \frac{k^{1/3}}{\alpha^{3}\epsilon^{2/3}}, \frac{\sqrt{k}}{\alpha^{5/4}\epsilon}\right\}\right)$ |
| $t$-model over $[n]$ | $\frac{t\log(n)}{\alpha}$ | $\tilde{O}\left(\max\left\{\frac{(t\log(n))^{4/5}}{\alpha^2}, \frac{(t\log(n))^{2/3}}{\alpha^{5/2}\epsilon^{1/3}}, \frac{(t\log(n))^{1/2}}{\alpha^{5/2}}, \frac{(t\log(n))^{1/3}}{\alpha^{5/3}\epsilon^{2/3}}, \frac{\sqrt{t\log(n)}}{\alpha^{3/2}\epsilon}\right\}\right)$ |
| MHR over $[n]$ | $\frac{\log(n/\alpha)}{\alpha}$ | $\tilde{O}\left(\max\left\{\frac{(\log(n/\alpha))^{4/5}}{\alpha^2}, \frac{\log(n/\alpha)^{2/3}}{\alpha^{5/3}\epsilon^{1/3}}, \frac{(\log(n/\alpha))^{1/2}}{\alpha^{5/2}}, \frac{(\log(n/\alpha))^{1/3}}{\alpha^{5/3}\epsilon^{2/3}}, \frac{\sqrt{\log(n/\alpha)}}{\alpha^{3/2}\epsilon}\right\}\right)$ |

Ilias Diakonikolas, Daniel M Kane, and Vladimir Nikishkin.
Optimal algorithms and lower bounds for testing closeness of structured distributions.
In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1183–1202. IEEE, 2015.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith.
Calibrating noise to sensitivity in private data analysis.
In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pages 265–284. Springer, 2006.