The Structure of Optimal Private Tests of Simple Hypotheses

Audra McMillan

Northeastern and Boston University

joint work with
Clément Canonne (Stanford),
Gautam Kamath (U. Waterloo),
Adam Smith (BU),
Jonathan Ullman (Northeastern)

https://arxiv.org/abs/1811.11148
Hypothesis Testing

Do at least 50% of people support the Democrats?
Do at least 50% of people support the Democrats?

Given a space of distributions $\Delta(\Omega)$,

$\mathcal{H}_0 \subset \Delta(\Omega)$ and $\mathcal{H}_1 \subset \Delta(\Omega)$

null hypothesis alternate hypothesis

**Hypothesis Test**

Given $x_1, \ldots, x_n \sim R$, a hypothesis test determines with high probability whether $R \in \mathcal{H}_0$ or $R \in \mathcal{H}_1$. 
Hypothesis Testing

Do at least 50% of people support the Democrats?

Given a space of distributions $\Delta(\Omega)$,

$\mathcal{H}_0 \subset \Delta(\Omega)$ and $\mathcal{H}_1 \subset \Delta(\Omega)$

null hypothesis alternate hypothesis

Hypothesis Test

Given $x_1, \ldots, x_n \sim R$, a hypothesis test determines with high probability whether $R \in \mathcal{H}_0$ or $R \in \mathcal{H}_1$.

- Hypothesis tests are ubiquitous in statistical inference
- Formalise yes-or-no questions about an underlying population.
Do at least 50\% of people support the Democrats?
Simple Hypothesis Testing

\[ R = \text{Ber}(0.5) \text{ or } R = \text{Ber}(0.49) \]
Let $P$ and $Q$ be two distributions on the same domain $\Omega$. In a **simple hypothesis test**, 

$$\mathcal{H}_0 = \{P\} \text{ and } \mathcal{H}_1 = \{Q\}.$$ 

Many composite hypotheses can be reduced to simple hypotheses.

Lower bounds for simple hypothesis testing arise in lower bounds for estimation.
Differential privacy

Desirable: statistical tests to be stable under small changes in the data privacy and generalisability under adaptive data analysis

$\epsilon$-differential privacy

A test $T : \Omega^n \rightarrow \{\mathcal{H}_0, \mathcal{H}_1\}$ is $\epsilon$-differentially private (DP) if for all databases $x$ and $x'$ that differ on a single element, and all subsets $S \subset \{\mathcal{H}_0, \mathcal{H}_1\}$,

$$\mathbb{P}[T(x) \in S] \leq e^\epsilon \mathbb{P}[T(x') \in S]$$
Differential privacy

Desirable: statistical tests to be stable under small changes in the data privacy and generalisability under adaptive data analysis.

$\epsilon$-differential privacy

A test $T : \Omega^n \to \{\mathcal{H}_0, \mathcal{H}_1\}$ is $\epsilon$-differentially private (DP) if for all databases $x$ and $x'$ that differ on a single element, and all subsets $S \subset \{\mathcal{H}_0, \mathcal{H}_1\}$,

$$\mathbb{P}[T(x) \in S] \leq e^\epsilon \mathbb{P}[T(x') \in S]$$
Differential privacy

Desirable: statistical tests to be stable under small changes in the data privacy and generalisability under adaptive data analysis

$\epsilon$-differential privacy

A test $T : \Omega^n \rightarrow \{\mathcal{H}_0, \mathcal{H}_1\}$ is $\epsilon$-differentially private (DP) if for all databases $x$ and $x'$ that differ on a single element, and all subsets $S \subset \{\mathcal{H}_0, \mathcal{H}_1\}$,

$$\mathbb{P}[T(x) \in S] \leq e^\epsilon \mathbb{P}[T(x') \in S]$$

Other stability notions

[AcharyaSunZhang18] Sample complexity is asymptotically the same for:

- $\epsilon$-DP,
- $\epsilon$-Total-variation stability (requires $\mathbb{P}[T(x) \in S] \leq \mathbb{P}[T(x') \in S] + \epsilon$)
- stability notions in-between (KL-stability, “concentrated DP”)

Related work

- DP versions of popular statistical tests [Vu, Slavkovic '09, Uhler, Slavkovic, Feinberg '13, Wang, Lee, Kifer '15, Gaboardi, Lim, Rogers, Vadhan '16, Kifers, Rogers '17, Acharya, Sun, Zhang '18, Campbell, Bray, Ritz, Groce '18, Couch, Kazan, Shi, Bray, Groce '18a,b, Swanberg, Globus-Harris, Griffith, Ritz, Groce, Bray '18]
  - goodness-of-fit, closeness, independence
  - focus on small sample sizes

- Asymptotic sample complexity of private testing [Cai, Daskalakis, Kamath '17, Aliakbarpour, Diakonikolas, Rubinfeld '18, Acharya, Sun, Zhang '18, Acharya, Kamath, Sun, Zhang '18]

- "Local" model (e.g. randomized response) [Duchi, Jordan, Wainwright '13, '18, Sheffet '18, Acharya, Cannone, Freitag, Tyagi '18]

- Restricted class of algorithms where data points are randomized individually

- Our work: instance-specific sample complexity
  - First work to give an instance-specific characterisation of sample complexity in the central model.
  - [Duchi, Jordan, Wainwright '13]: instance-specific characterisation for the same problem in the local model.
Related work

- **DP versions of popular statistical tests** [Vu, Slavkovic '09, Uhler, Slavkovic, Feinberg '13, Wang, Lee, Kifer '15, Gaboardi, Lim, Rogers, Vadhan '16, Kifers, Rogers '17, Acharya, Sun, Zhang '18, Campbell, Bray, Ritz, Groce '18, Couch, Kazan, Shi, Bray, Groce '18a,b, Swanberg, Globus-Harris, Griffith, Ritz, Groce, Bray '18]
  - goodness-of-fit, closeness, independence
  - focus on small sample sizes

- **Asymptotic sample complexity of private testing** [Cai, Daskalakis, Kamath '17, Aliakbarpour, Diakonikolas, Rubinfeld '18, Acharya, Sun, Zhang '18, Acharya, Kamath, Sun, Zhang '18]
Related work

- **DP versions of popular statistical tests** [Vu, Slavkovic '09, Uhler, Slavkovic, Feinberg '13, Wang, Lee, Kifer '15, Gaboardi, Lim, Rogers, Vadhan '16, Kifers, Rogers '17, Acharya, Sun, Zhang '18, Campbell, Bray, Ritz, Groce '18, Couch, Kazan, Shi, Bray, Groce '18a,b, Swanberg, Globus-Harris, Griffith, Ritz, Groce, Bray '18]
  - goodness-of-fit, closeness, independence
  - focus on small sample sizes

- **Asymptotic sample complexity of private testing** [Cai, Daskalakis, Kamath '17, Aliakbarpour, Diakonikolas, Rubinfeld '18, Acharya, Sun, Zhang '18, Acharya, Kamath, Sun, Zhang '18]

- **“Local” model (e.g. randomized response)** [Duchi, Jordan, Wainwright '13, '18, Sheffet '18, Acharya, Cannone, Freitag, Tyagi '18]
  - Restricted class of algorithms where data points are randomized individually
Related work

- **DP versions of popular statistical tests** [Vu, Slavkovic '09, Uhler, Slavkovic, Feinberg '13, Wang, Lee, Kifer '15, Gaboardi, Lim, Rogers, Vadhan '16, Kifers, Rogers '17, Acharya, Sun, Zhang '18, Campbell, Bray, Ritz, Groce '18, Couch, Kazan, Shi, Bray, Groce '18a,b, Swanberg, Globus-Harris, Griffith, Ritz, Groce, Bray '18]
  - goodness-of-fit, closeness, independence
  - focus on small sample sizes

- **Asymptotic sample complexity of private testing** [Cai, Daskalakis, Kamath '17, Aliakbarpour, Diakonikolas, Rubinfeld '18, Acharya, Sun, Zhang '18, Acharya, Kamath, Sun, Zhang '18]

- **“Local” model (e.g. randomized response)** [Duchi, Jordan, Wainwright '13, '18, Sheffet '18, Acharya, Cannone, Freitag, Tyagi '18]
  - Restricted class of algorithms where data points are randomized individually
Related work

- **DP versions of popular statistical tests** [Vu, Slavkovic '09, Uhler, Slavkovic, Feinberg '13, Wang, Lee, Kifer '15, Gaboardi, Lim, Rogers, Vadhan '16, Kifers, Rogers '17, Acharya, Sun, Zhang '18, Campbell, Bray, Ritz, Groce '18, Couch, Kazan, Shi, Bray, Groce '18a,b, Swanberg, Globus-Harris, Griffith, Ritz, Groce, Bray '18]
  - goodness-of-fit, closeness, independence
  - focus on small sample sizes

- **Asymptotic sample complexity of private testing** [Cai, Daskalakis, Kamath '17, Aliakbarpour, Diakonikolas, Rubinfeld '18, Acharya, Sun, Zhang '18, Acharya, Kamath, Sun, Zhang '18]

- **“Local” model (e.g. randomized response)** [Duchi, Jordan, Wainwright '13, '18, Sheffet '18, Acharya, Cannone, Freitag, Tyagi '18]
  - Restricted class of algorithms where data points are randomized individually

**Our work: instance-specific sample complexity**

First work to give an instance-specific characterisation of sample complexity in the central model.
Related work

- **DP versions of popular statistical tests** [Vu, Slavkovic '09, Uhler, Slavkovic, Feinberg '13, Wang, Lee, Kifer '15, Gaboardi, Lim, Rogers, Vadhan '16, Kifers, Rogers '17, Acharya, Sun, Zhang '18, Campbell, Bray, Ritz, Groce '18, Couch, Kazan, Shi, Bray, Groce '18a,b, Swanberg, Globus-Harris, Griffith, Ritz, Groce, Bray '18]
  - goodness-of-fit, closeness, independence
  - focus on small sample sizes

- **Asymptotic sample complexity of private testing** [Cai, Daskalakis, Kamath '17, Aliakbarpour, Diakonikolas, Rubinfeld '18, Acharya, Sun, Zhang '18, Acharya, Kamath, Sun, Zhang '18]

- **“Local” model (e.g. randomized response)** [Duchi, Jordan, Wainwright '13, '18, Sheffet '18, Acharya, Cannone, Freitag, Tyagi '18]
  - Restricted class of algorithms where data points are randomized individually

**Our work: instance-specific sample complexity**

First work to give an instance-specific characterisation of sample complexity in the central model.

[Duchi, Jordan, Wainwright 13]: instance-specific characterisation for the same problem in the local model.
A hypothesis test $T : \Omega^n \rightarrow \{P, Q\}$ is an algorithm that attempts to determine whether $X \sim P^n$ or $X \sim Q^n$.

The test $T : \Omega^n \rightarrow \{P, Q\}$ distinguishes between $P$ and $Q$ with sample complexity $SC^{P,Q}(T)$ if for all $n \geq SC^{P,Q}(T)$:

1. $\mathbb{P}_{X \sim P^n}[T(X) = P] \geq 2/3$
2. $\mathbb{P}_{X \sim Q^n}[T(X) = Q] \geq 2/3$

A test $T$ has optimal sample complexity if for all tests $T'$,

$$SC^{P,Q}(T) = O(SC^{P,Q}(T')).$$
A \( \epsilon \)-DP hypothesis test \( T : \Omega^n \rightarrow \{ P, Q \} \) is an algorithm that attempts to determine whether \( X \sim P^n \) or \( X \sim Q^n \) while maintaining the privacy of elements of the database.

The test \( T : \Omega^n \rightarrow \{ P, Q \} \) distinguishes between \( P \) and \( Q \) with sample complexity \( SC_{\epsilon}^{P,Q}(T) \) if for all \( n \geq SC_{\epsilon}^{P,Q}(T) \):

1. \( \mathbb{P}_{X \sim P^n} [T(X) = P] \geq 2/3 \)
2. \( \mathbb{P}_{X \sim Q^n} [T(X) = Q] \geq 2/3 \)
3. \( T \) is \( \epsilon \)-DP

A test \( T \) has optimal sample complexity if for all \( \epsilon \)-DP tests \( T' \),

\[
SC_{\epsilon}^{P,Q}(T) = O(SC_{\epsilon}^{P,Q}(T')).
\]
Contributions of this work

The Optimal Sample Complexity

We characterise the optimal sample complexity for $\epsilon$-DP distinguishing between $P$ and $Q$, for any distributions $P$ and $Q$. An Optimal Efficient* Test

Give a specific efficient* test that achieves this sample complexity. Along the way:

Novel interpretation of the Hellinger distance, insight into the classical setting.

Applications: (In paper) Optimal private algorithms for $\epsilon$-DP change point detection, improving [Cummings, Krehbiel, Mei, Tuo, Zhang 2018].
Contributions of this work

The Optimal Sample Complexity

We characterise the optimal sample complexity for $\epsilon$-DP distinguishing between $P$ and $Q$, for any distributions $P$ and $Q$.

An Optimal Efficient* Test

Give a specific efficient* test that achieves this sample complexity.
Contribution of this work

The Optimal Sample Complexity
We characterise the optimal sample complexity for $\epsilon$-DP distinguishing between $P$ and $Q$, for any distributions $P$ and $Q$.

An Optimal Efficient* Test
Give a specific efficient* test that achieves this sample complexity.

Along the way: Novel interpretation of the Hellinger distance, insight into the classical setting.
Contributions of this work

**The Optimal Sample Complexity**

We characterise the optimal sample complexity for $\epsilon$-DP distinguishing between $P$ and $Q$, for any distributions $P$ and $Q$.

**An Optimal Efficient* Test**

Give a specific efficient* test that achieves this sample complexity.

**Along the way:** Novel interpretation of the Hellinger distance, insight into the classical setting.

**Applications:** (In paper) Optimal private algorithms for $\epsilon$-DP change point detection, improving [Cummings, Krehbiel, Mei, Tuo, Zhang 2018].
Classical Solution

\[
\text{LLR}(X) = \begin{cases} 
P & \text{if } P^n(X) \geq Q^n(X) \\
Q & \text{if } P^n(X) < Q^n(X)
\end{cases}
\]

Lemma (Neyman-Pearson (1933))

\text{LLR has exactly optimal sample complexity.}
Classical Solution

$$LLR(X) = \begin{cases} P & \text{if } P^n(X) \geq Q^n(X) \\ Q & \text{if } P^n(X) < Q^n(X) \end{cases}$$

Lemma (Neyman-Pearson (1933))

LLR has exactly optimal sample complexity.

The sample complexity of LLR is

$$\Theta \left( \frac{1}{H^2(P, Q)} \right)$$

where

$$H^2(P, Q) = \frac{1}{2} \int_{\Omega} (\sqrt{P(x)} - \sqrt{Q(x)})^2 dx$$
Classical Solution

\[ \text{LLR}(X) = \sum_{i=1}^{n} \ln \frac{P(x_i)}{Q(x_i)}, \quad \text{LLR}(X) = \begin{cases} P & \text{if LLR}(X) \geq 0 \\ Q & \text{if LLR}(X) < 0 \end{cases} \]

Lemma (Neyman-Pearson (1933))

\text{LLR} \ has \ exactly \ optimal \ sample \ complexity.

The sample complexity of LLR is

\[ \Theta \left( \frac{1}{H^2(P, Q)} \right) \]

where

\[ H^2(P, Q) = \frac{1}{2} \int_{\Omega} (\sqrt{P(x)} - \sqrt{Q(x)})^2 dx \]
Classical Solution

\[
\text{LLR}(X) = \sum_{i=1}^{n} \ln \frac{P(x_i)}{Q(x_i)}, \quad \text{LLR}(X) = \begin{cases} 
P & \text{if } \text{LLR}(X) \geq 0 \\
Q & \text{if } \text{LLR}(X) < 0
\end{cases}
\]

Lemma (Neyman-Pearson (1933))

LLR has exactly optimal sample complexity.

The sample complexity of LLR is

\[\Theta \left( \frac{1}{H^2(P, Q)} \right)\]

where

\[H^2(P, Q) = \frac{1}{2} \int_{\Omega} (\sqrt{P(x)} - \sqrt{Q(x)})^2 dx\]

Goal: Obtain an analogous simple characterisation for \(\epsilon\)-DP testing.
Statement of the main theorem: Instance-optimal test.

Key Insight: Novel interpretation of the Hellinger distance.

Upper Bound: Noisy clamped log-likelihood test

Lower Bound: Reduce to a “fictional test”
Statement of the main theorem: Instance-optimal test.

Key Insight: Novel interpretation of the Hellinger distance.

Upper Bound: Noisy clamped log-likelihood test

Lower Bound: Reduce to a “fictional test”
Previously known bounds

\[ SC_{\epsilon} P, Q \] lies somewhere in this range.

\[
\max \{ SC_{\epsilon} P, Q, \frac{1}{\epsilon TV(P, Q)} \} \quad \frac{1}{\epsilon} SC_{\epsilon} P, Q
\]
Previously known bounds

\[ SC_{\varepsilon}^{P, Q} \text{ lies somewhere in this range.} \]

\[ \max\{ SC_{\varepsilon}^{P, Q}, \frac{1}{\varepsilon TV(P,Q)} \} \quad \frac{1}{\varepsilon} SC_{\varepsilon}^{P, Q} \]

- For Bernoulli distributions, \( SC_{\varepsilon}^{P, Q} \) characterises the private sample complexity.
- **Phase transition:**
  - For \( \varepsilon \geq \frac{H^2(P,Q)}{TV(P,Q)} \), we get **privacy for free**.
  - When \( \varepsilon < \frac{H^2(P,Q)}{TV(P,Q)} \), we get an extra term but it’s dependence on \( P \) and \( Q \) is lower order since \( \frac{1}{H^2(P,Q)} \geq \frac{1}{TV(P,Q)} \).
This characterisation does not hold in general...
Example

$$SC^P, Q = \Theta \left( \frac{1}{\alpha^{3/2}} \right)$$
Example

\[ SC^{P,Q} = \Theta \left( \frac{1}{\alpha^{3/2}} \right) \]

<table>
<thead>
<tr>
<th>( \epsilon )</th>
<th>( SC^{P,Q}_\epsilon )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>( \sqrt{\alpha} )</td>
</tr>
<tr>
<td>( \Theta \left( \frac{1}{\epsilon \alpha^{3/2}} \right) )</td>
<td>( \frac{1}{\epsilon \alpha^{3/2}} )</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>( \sqrt{\alpha} )</td>
</tr>
<tr>
<td>( \Theta \left( \frac{1}{\epsilon \alpha^2} \right) )</td>
<td>( \frac{1}{\epsilon \alpha} )</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>( \sqrt{\alpha} )</td>
</tr>
<tr>
<td>( \Theta \left( \frac{1}{\epsilon \alpha} \right) )</td>
<td>( \frac{1}{\epsilon \alpha^{3/2}} )</td>
</tr>
</tbody>
</table>
Main Theorem: Clamped Log-likelihood Ratio

\[ \text{cLLR}(X) = \sum_{i=1}^{n} \left[ \ln \frac{P(x_i)}{Q(x_i)} \right]^\epsilon - \epsilon' \]
Main Theorem: Clamped Log-likelihood Ratio

\[ c\text{LLR}(X) = \sum_{i=1}^{n} \left[ \ln \frac{P(x_i)}{Q(x_i)} \right]^{\epsilon}_{-\epsilon'} \]
Main Theorem: Clamped Log-likelihood Ratio

\[ \text{cLLR}(X) = \sum_{i=1}^{n} \left( \ln \frac{P(x_i)}{Q(x_i)} \right)^{\epsilon}_{-\epsilon'} \]
Main Theorem: Two Optimal Private Tests

\[ c\text{LLR}(X) = \sum_{i=1}^{n} \left[ \ln \frac{P(x_i)}{Q(x_i)} \right]^\epsilon \]

\[ \epsilon' \leq \epsilon \]

**Soft Clamped Log-likelihood Test**

\[ \text{scLLR}(X) = \begin{cases} P & \text{w.p. } e^{\frac{1}{2} c\text{LLR}(X)} \frac{1}{1 + e^{\frac{1}{2} c\text{LLR}(X)}} \\ Q & \text{w.p. } \frac{1}{1 + e^{\frac{1}{2} c\text{LLR}(X)}} \end{cases} \]

**Noisy Clamped Log-likelihood Test**

\[ \text{ncLLR}(X) = \begin{cases} P & \text{if } c\text{LLR}(X) + \text{Lap}(2) \geq 0 \\ Q & \text{otherwise} \end{cases} \]
The Main Theorem: Optimal Private Sample Complexity

**Theorem**

The soft clamped log-likelihood test, scLLR, and the noisy clamped log-likelihood test, ncLLR, both have sample complexity

\[ SC_{\epsilon}^{P,Q} = \Theta \left( \frac{1}{\epsilon \tau + (1 - \tau) H^2(P', Q')} \right), \]

which is minimal (up to constants) among \( \epsilon \)-DP testing algorithms.

\[ P' = \frac{1}{1 - \tau} \min\{P, e^{\epsilon'} Q\} \quad \text{and} \quad Q' = \frac{1}{1 - \tau} \min\{Q, e^{\epsilon} P\} \]
The Main Theorem: Optimal Private Sample Complexity

**Theorem**

The soft clamped log-likelihood test, $\text{scLLR}$, and the noisy clamped log-likelihood test, $\text{ncLLR}$, both have sample complexity

$$SC_{\epsilon}^{P,Q} = \Theta \left( \frac{1}{\epsilon \tau + (1 - \tau) H^2(P', Q')} \right),$$

which is minimal (up to constants) among $\epsilon$-DP testing algorithms.

$$P' = \frac{1}{1 - \tau} \min\{P, e^{\epsilon'} Q\} \quad \text{and} \quad Q' = \frac{1}{1 - \tau} \min\{Q, e^{\epsilon} P\}$$
Comparison to previous bounds

Non-private SC

\[ SC_{\epsilon}^{P,Q} = \Omega \left( \frac{1}{H^2(P, Q)} \right) \]

For \( \epsilon < 1 \), using subsample-and-aggregate [NRS07],

\[ SC_{\epsilon}^{P,Q} = O \left( \frac{1}{\epsilon} SC_{P,Q} \right) . \]
Comparison to previous bounds

Non-private SC

$$SC_{\epsilon}^{P, Q} = \Omega \left( \frac{1}{H^2(P, Q)} \right)$$

For $\epsilon < 1$, using subsample-and-aggregate [NRS07],

$$SC_{\epsilon}^{P, Q} = O \left( \frac{1}{\epsilon} SC^{P, Q} \right).$$

Locally differentially private SC

$$SC_{\epsilon, \text{local}}^{P, Q} = \Theta \left( \frac{1}{\epsilon^2 TV(P, Q)^2} \right),$$

which is the optimal sample complexity for $\epsilon$-DP local testing algorithms [DuchiJordanWainwright13].
A fictional test

\[ P = (1 - \tau)P' + \tau P'' \text{ and } Q = (1 - \tau)Q' + \tau Q'' \]

where \( P'' \) and \( Q'' \) have disjoint support.
\[ P = (1 - \tau)P' + \tau P'' \quad \text{and} \quad Q = (1 - \tau)Q' + \tau Q'' \]

where \( P'' \) and \( Q'' \) have disjoint support.

\[ \{ \quad \tau n \text{ samples from } P'' \text{ or } Q'' \quad \} \]

\[ \{ \quad (1 - \tau)n \text{ samples from } P' \text{ or } Q' \quad \} \]
A fictional test

\[ P = (1 - \tau)P' + \tau P'' \] and \[ Q = (1 - \tau)Q' + \tau Q'' \]

where \( P'' \) and \( Q'' \) have disjoint support.

\( \tau n \) samples from \( P'' \) or \( Q'' \)

\( P'' \) and \( Q'' \) have disjoint support so this requires

\[ \tau n = \Theta \left( \frac{1}{\epsilon} \right) \]

\( (1 - \tau)n \) samples from \( P' \) or \( Q' \)
A fictional test

\[ P = (1 - \tau)P' + \tau P'' \] and \[ Q = (1 - \tau)Q' + \tau Q'' \]

where \( P'' \) and \( Q'' \) have disjoint support.

\[
\begin{align*}
\tau n & \text{ samples from } P'' \text{ or } Q'' \\
(1 - \tau) n & \text{ samples from } P' \text{ or } Q'
\end{align*}
\]

\( P'' \) and \( Q'' \) have disjoint support so this requires

\[
\tau n = \Theta \left( \frac{1}{\epsilon} \right)
\]

\( P' \) and \( Q' \) are designed so we get privacy for free so this requires

\[
(1 - \tau) n = \Theta \left( \frac{1}{H^2(P', Q')} \right)
\]
Proof Idea

Main Theorem

\[
SC^P _\epsilon Q = \Theta \left( \frac{1}{\epsilon \tau + (1 - \tau) H^2(P', Q')} \right)
\]

Key Insight: If \(-\epsilon \leq \ln \frac{P(x)}{Q(x)} \leq \epsilon\) then we get privacy for free.

Lower Bound: You must have enough samples for one of these tests to work.

Upper Bound: The noisy clamped LLR test “captures” both tests without labels.
Proof Idea

Main Theorem

$$SC_{\epsilon}^{P,Q} = \Theta \left( \min \left\{ \frac{1}{\epsilon \tau}, \frac{1}{(1-\tau)H^2(P', Q')} \right\} \right)$$

Key Insight: If $-\epsilon \leq \ln \frac{P(x)}{Q(x)} \leq \epsilon$ then we get privacy for free.

Lower Bound: You must have enough samples for one of these tests to work.

Upper Bound: The noisy clamped LLR test “captures” both tests without labels.
Statement of the main theorem: Instance-optimal test.

Key Insight: Novel interpretation of the Hellinger distance.

Upper Bound: Noisy clamped log-likelihood test

Lower Bound: Reduce to a “fictional test”
The Advantage of a Test

The Advantage of a test $T : \Omega^n \rightarrow \{P, Q\}$ is

$$\text{adv}_n(T) = \mathbb{P}_{X \sim P^n} [T(X) = P] - \mathbb{P}_{X \sim Q^n} [T(X) = P]$$

true positive \hspace{5cm} false positive

A test distinguishes $P$ and $Q$ if and only if $\text{adv}_n(T) \geq 2/3$. 
Step 1: Interpretation of the Hellinger Distance

\[ \text{LLR}(X) = \sum_{i=1}^{n} \ln \frac{P(x_i)}{Q(x_i)}, \]

\[ \text{LLR}(X) = \begin{cases} 
   P & \text{if } \text{LLR}(X) \geq 0 \\
   Q & \text{if } \text{LLR}(X) < 0 
\end{cases} \]
Step 1: Interpretation of the Hellinger Distance

\[ \text{LLR}(X) = \sum_{i=1}^{n} \ln \frac{P(x_i)}{Q(x_i)}, \quad \text{sLLR}(X) = \begin{cases} P & \text{w.p.} \frac{e^{\frac{1}{2} \text{LLR}(X)}}{1+e^{\frac{1}{2} \text{LLR}(X)}} \\ Q & \text{w.p.} \frac{1}{1+e^{\frac{1}{2} \text{LLR}(X)}} \end{cases} \]
Step 1: Interpretation of the Hellinger Distance

$$\text{LLR}(X) = \sum_{i=1}^{n} \ln \frac{P(x_i)}{Q(x_i)}, \quad \text{sLLR}(X) = \begin{cases} P & \text{w.p.} \frac{e^{\frac{1}{2} \text{LLR}(X)}}{1 + e^{\frac{1}{2} \text{LLR}(X)}} \\ Q & \text{w.p.} \frac{1}{1 + e^{\frac{1}{2} \text{LLR}(X)}} \end{cases}$$

**Theorem**

*The Hellinger distance $H^2(P^n, Q^n)$ is exactly the advantage of the soft log-likelihood test, sLLR.*
Proof

Note that for all \( x \in \mathcal{X}^n \), \( \mathbb{P} [\text{sLLR}(x) = P] = \frac{e^{\frac{1}{2} \text{LLR}(x)}}{1+e^{\frac{1}{2} \text{LLR}(x)}} = \frac{\sqrt{P^n(x)}}{\sqrt{Q^n(x)}+1} \)

so \( 2\mathbb{P} [\text{sLLR}(x) = P] - 1 = \frac{\sqrt{P^n(x)}-1}{\sqrt{P^n(x)}+1} \)
Proof

Note that for all $x \in \mathcal{X}^n$, $\mathbb{P}[\text{sLLR}(x) = P] = \frac{e^{\frac{1}{2} \text{LLR}(x)}}{1 + e^{\frac{1}{2} \text{LLR}(x)}} = \frac{\sqrt{\frac{P^n(x)}{Q^n(x)}}}{\sqrt{\frac{P^n(x)}{Q^n(x)}} + 1}$

so $2\mathbb{P}[\text{sLLR}(x) = P] - 1 = \frac{\sqrt{\frac{P^n(x)}{Q^n(x)}} - 1}{\sqrt{\frac{P^n(x)}{Q^n(x)}} + 1}$

and

$$\text{adv}_n(\text{sLLR}) = \mathbb{P}_{X \sim P^n}[T(X) = P] - \mathbb{P}_{X \sim Q^n}[T(X) = P]$$
Proof

Note that for all $x \in \mathcal{X}^n$, $\mathbb{P}[s\text{LLR}(x) = P] = \frac{e^{\frac{1}{2} \text{LLR}(x)}}{1 + e^{\frac{1}{2} \text{LLR}(x)}} = \frac{\sqrt{\frac{P^n(x)}{Q^n(x)}}}{\sqrt{\frac{P^n(x)}{Q^n(x)}} + 1}$

so $2\mathbb{P}[s\text{LLR}(x) = P] - 1 = \frac{\sqrt{\frac{P^n(x)}{Q^n(x)}} - 1}{\sqrt{\frac{P^n(x)}{Q^n(x)}} + 1}$

and

$$\text{adv}_n(s\text{LLR}) = \mathbb{P}_{X \sim P^n} [T(X) = P] - \mathbb{P}_{X \sim Q^n} [T(X) = P]$$

$$= \frac{1}{2} \left( 2 \mathbb{P}_{X \sim P^n} [T(X) = P] - 2 \mathbb{P}_{X \sim Q^n} [T(X) = P] \right)$$
Proof

Note that for all $x \in \mathcal{X}^n$, $\mathbb{P}[sLLR(x) = P] = \frac{e^{\frac{1}{2}LLR(x)}}{1 + e^{\frac{1}{2}LLR(x)}} = \frac{\sqrt{P^n(x)}}{\sqrt{Q^n(x)} + 1}$

so $2\mathbb{P}[sLLR(x) = P] - 1 = \frac{\sqrt{P^n(x)} - 1}{\sqrt{Q^n(x)} + 1}$

and

$$ad_{\mathcal{X}}(sLLR) = \mathbb{P}_{X \sim P^n}[T(X) = P] - \mathbb{P}_{X \sim Q^n}[T(X) = P]$$

$$= \frac{1}{2} \left( 2 \mathbb{P}_{X \sim P^n}[T(X) = P] - 2 \mathbb{P}_{X \sim Q^n}[T(X) = P] \right)$$

$$= \frac{1}{2} \int (P^n(X) - Q^n(X)) \frac{\sqrt{P^n(X)} - 1}{\sqrt{Q^n(X)} + 1} dX$$
Proof

Note that for all \( x \in \mathcal{X}^n \), \( \mathbb{P}[sLLR(x) = P] = \frac{e^{\frac{1}{2}LLR(x)}}{1 + e^{\frac{1}{2}LLR(x)}} = \frac{\sqrt{P^n(x)}}{\sqrt{Q^n(x)} + 1} \)

so \( 2\mathbb{P}[sLLR(x) = P] - 1 = \frac{\sqrt{Q^n(x)} - 1}{\sqrt{Q^n(x)} + 1} \)

and

\[
adv_n(sLLR) = \mathbb{P}_{X \sim P^n}[T(X) = P] - \mathbb{P}_{X \sim Q^n}[T(X) = P] \\
= \frac{1}{2} \left( 2 \mathbb{P}_{X \sim P^n}[T(X) = P] - 2 \mathbb{P}_{X \sim Q^n}[T(X) = P] \right) \\
= \frac{1}{2} \int (P^n(X) - Q^n(X)) \frac{\sqrt{P^n(X)} - 1}{\sqrt{P^n(X)} + 1} dX \\
= \frac{1}{2} \int (\sqrt{P^n(X)} - \sqrt{Q^n(X)})^2 dX
\]
Proof

Note that for all \( x \in \mathcal{X}^n \), \( \mathbb{P} \left[ \text{sLLR}(x) = P \right] = \frac{e^{\frac{1}{2} \text{LLR}(x)}}{1 + e^{\frac{1}{2} \text{LLR}(x)}} = \frac{\sqrt{\frac{P^n(x)}{Q^n(x)}}}{\sqrt{\frac{P^n(x)}{Q^n(x)} + 1}} \)

so \( 2 \mathbb{P} \left[ \text{sLLR}(x) = P \right] - 1 = \frac{\sqrt{\frac{P^n(x)}{Q^n(x)}} - 1}{\sqrt{\frac{P^n(x)}{Q^n(x)} + 1}} \)

and

\[
\text{adv}_n(\text{sLLR}) = \mathbb{P}_{X \sim P^n} \left[ T(X) = P \right] - \mathbb{P}_{X \sim Q^n} \left[ T(X) = P \right]
\]

\[
= \frac{1}{2} \left( 2 \mathbb{P}_{X \sim P^n} \left[ T(X) = P \right] - 2 \mathbb{P}_{X \sim Q^n} \left[ T(X) = P \right] \right)
\]

\[
= \frac{1}{2} \int (P^n(X) - Q^n(X)) \frac{\sqrt{\frac{P^n(X)}{Q^n(X)}} - 1}{\sqrt{\frac{P^n(X)}{Q^n(X)} + 1}} dX
\]

\[
= \frac{1}{2} \int (\sqrt{P^n(X)} - \sqrt{Q^n(X)})^2 dX
\]

\[
= H^2(P^n, Q^n)
\]
Theorem

The Hellinger distance \( H^2(P^n, Q^n) \) is exactly the advantage of the soft log-likelihood test, sLLR.
Theorem

The Hellinger distance $H^2(P^n, Q^n)$ is exactly the advantage of the soft log-likelihood test, sLLR.

Corollary

$SC^{P,Q}(sLLR) = O \left( \frac{1}{H^2(P,Q)} \right)$, which is optimal among non-private tests.

\[
\text{adv} = H^2(P^n, Q^n) = 1 - \langle \sqrt{P^n}, \sqrt{Q^n} \rangle \\
= 1 - \langle \sqrt{P}, \sqrt{Q} \rangle^n \\
= 1 - \left( 1 - H^2(P, Q) \right)^n \\
\approx nH^2(P, Q)
\]
Privacy for free!

The randomised tester $sLLR$ has optimal (nonprivate) sample complexity. 

\[
\text{If } -\epsilon \leq \ln \frac{P(x)}{Q(x)} \leq \epsilon \quad \text{then,}
\]

- $sLLR$ is exactly $scLLR$:

\[
LLR(X) = \sum_{i=1}^{n} \ln \frac{P(x_i)}{Q(x_i)} = \sum_{i=1}^{n} \left[ \ln \frac{P(x_i)}{Q(x_i)} \right]^\epsilon = cLLR(X)
\]

- $sLLR$ is $\epsilon$-DP.

- We get $\epsilon$-DP for free:

\[
SC^P_{\epsilon,Q} = \Theta(SC^{P,Q})
\]
A fictional test

\[ P = (1 - \tau)P' + \tau P'' \quad \text{and} \quad Q = (1 - \tau)Q' + \tau Q'' \]

where \( P'' \) and \( Q'' \) have disjoint support.

\begin{align*}
\tau n \text{ samples from } P'' \text{ or } Q'' \\
(1 - \tau) n \text{ samples from } P' \text{ or } Q'
\end{align*}

\( P'' \) and \( Q'' \) have disjoint support so this requires

\[ \tau n = \Theta \left( \frac{1}{\epsilon} \right) \]

\( P' \) and \( Q' \) satisfy

\[ -\epsilon \leq \ln \frac{P'(x)}{Q'(x)} \leq \epsilon \]

so this requires

\[ (1 - \tau) n = \Theta \left( \frac{1}{H^2(P', Q')} \right) \]
Statement of the main theorem: Instance-optimal test.

Key Insight: Novel interpretation of the Hellinger distance.

Upper Bound: Noisy clamped log-likelihood test

Lower Bound: Reduce to a “fictional test”
Upper Bound Proof Outline

\[
cLLR(X) = \sum_{i=1}^{n} \left[ \ln \frac{P(x_i)}{Q(x_i)} \right]^{\epsilon} = \sum_{x_i \sim P'' \text{ or } Q''} \epsilon + \sum_{x_i \sim P' \text{ or } Q'} \ln \frac{P'(x_i)}{Q'(x_i)}
\]

\[
P = (1 - \tau)P' + \tau P'' \quad \text{and} \quad Q = (1 - \tau)Q' + \tau Q''
\]
Upper Bound Proof Outline

\[
c\text{LLR}(X) = \sum_{i=1}^{n} \left[ \ln \frac{P(x_i)}{Q(x_i)} \right]^\epsilon = \sum_{x_i \sim P''} \epsilon + \sum_{x_i \sim P'} \ln \frac{P'(x_i)}{Q'(x_i)}
\]

If \( n \) is large enough that

\[
\sqrt{\text{Var}_{P^n}[c\text{LLR}(X)]} < \Delta_{\text{gap}}
\]

\[
\sqrt{\text{Var}_{Q^n}[c\text{LLR}(X)]} < \Delta_{\text{gap}}
\]

\( \Delta_{\text{gap}} > 1 \)

then

\[
SC_{\epsilon}^{P,Q}(nc\text{LLR}) = O(n).
\]
Upper Bound

cLLR(X) = \sum_{x_i \sim P''} \epsilon + \sum_{x_i \sim P'} \ln \frac{P'(x_i)}{Q'(x_i)}

\mathbb{E}_{P^n}[cLLR(X)] = \epsilon \tau n + (1 - \tau) n KL(P' \| Q')
\mathbb{E}_{Q^n}[cLLR(X)] = -\epsilon \tau n - (1 - \tau) n KL(Q' \| P')
\Delta_{gap} \geq 2n(\epsilon \tau + (1 - \tau) H^2(P', Q'))
Upper Bound

\[
c\text{LLR}(X) = \sum_{x_i \sim P''} \epsilon + \sum_{x_i \sim P'} \ln \frac{P'(x_i)}{Q'(x_i)}
\]

\[
\mathbb{E}_{Pn}[c\text{LLR}(X)] = \epsilon \tau n + (1 - \tau) n KL(P' || Q')
\]

\[
\mathbb{E}_{Qn}[c\text{LLR}(X)] = -\epsilon \tau n - (1 - \tau) n KL(Q' || P')
\]

\[
\Delta_{\text{gap}} \geq 2n(\epsilon \tau + (1 - \tau) H^2(P', Q'))
\]

\[
\text{Var}_{Pn}[c\text{LLR}(X)] \leq \mathbb{E}_{Pn}[c\text{LLR}(X)^2]
\]

\[
= n \left[ \tau \epsilon^2 + (1 - \tau) \int P'(x) \left( \ln \frac{P'(x)}{Q'(x)} \right)^2 \, dx \right]
\]

\[
\leq n[\tau \epsilon + (1 - \tau) H^2(P', Q')]
\]
Upper Bound

\[ c\text{LLR}(X) = \sum_{x_i \sim P''} \epsilon + \sum_{x_i \sim P'} \ln \frac{P'(x_i)}{Q'(x_i)} \]

\[ \Delta_{\text{gap}} \geq 2n(\epsilon \tau + (1 - \tau)H^2(P', Q')) \]

\[ \text{Var}_{P^n}[c\text{LLR}(X)] \leq n[\tau \epsilon + (1 - \tau)H^2(P', Q')] \]

\[ \text{Var}_{Q^n}[c\text{LLR}(X)] \leq n[\tau \epsilon + (1 - \tau)H^2(P', Q')] \]

If \[ n = \frac{C}{\epsilon \tau + (1 - \tau)H^2(P', Q')} \] then ncLLR distinguishes between \( P \) and \( Q \).
Statement of the main theorem: Instance-optimal test.

Key Insight: Novel interpretation of the Hellinger distance.

Upper Bound: Noisy clamped log-likelihood test

Lower Bound: Reduce to a “fictional test”
Lower Bound

\[ P = \tau P'' + (1 - \tau)P' \]
\[ R = \tau Q'' + (1 - \tau)P' \]
\[ Q = \tau Q'' + (1 - \tau)Q' \]

Claim 1

\[ SC_{\epsilon}^{P,Q} = \Omega \left( \min \left\{ SC_{\epsilon}^{P,R}, SC_{\epsilon}^{R,Q} \right\} \right) \]

Claim 2

\[ SC_{\epsilon}^{P,R} = \Omega \left( \frac{1}{\tau} SC_{\epsilon}^{P'',Q''} \right) = \Omega \left( \frac{1}{\epsilon \tau} \right) \]
\[ SC_{\epsilon}^{R,Q} = \Omega \left( \frac{1}{1 - \tau} SC_{\epsilon}^{P',Q'} \right) = \Omega \left( \frac{1}{(1 - \tau)H^2(P', Q')} \right) . \]
Lower Bound: \( SC_{\epsilon}^{P,Q} = \Omega \left( \min \{ SC_{\epsilon}^{P,R}, SC_{\epsilon}^{R,Q} \} \right) \)

Suppose \( T : \Omega^n \rightarrow \{ P, Q \} \) distinguishes between \( P \) and \( Q \) so

\[
\mathbb{P}_{P^n}[T(X) = P] - \mathbb{P}_{Q^n}[T(X) = P] \geq 2/3.
\]
Suppose $T : \Omega^n \rightarrow \{P, Q\}$ distinguishes between $P$ and $Q$ so

$$\mathbb{P}_P [T(X) = P] - \mathbb{P}_Q [T(X) = P] \geq 2/3.$$ 

So, either

$$\mathbb{P}_P [T(X) = P] - \mathbb{P}_R [T(X) = P] \geq 1/3$$

or

$$\mathbb{P}_R [T(X) = P] - \mathbb{P}_Q [T(X) = P] \geq 1/3$$
Lower Bound: $SC_{\epsilon}^{P,Q} = \Omega \left( \min \left\{ SC_{\epsilon}^{P,R}, SC_{\epsilon}^{R,Q} \right\} \right)$

Suppose $T : \Omega^n \rightarrow \{P, Q\}$ distinguishes between $P$ and $Q$ so

$$\mathbb{P}_P[T(X) = P] - \mathbb{P}_Q[T(X) = P] \geq 2/3.$$ 

So, either

$$\mathbb{P}_P[T(X) = P] - \mathbb{P}_R[T(X) = P] \geq 1/3$$

or

$$\mathbb{P}_R[T(X) = P] - \mathbb{P}_Q[T(X) = P] \geq 1/3$$

Suppose $\mathbb{P}_P[T(X) = P] - \mathbb{P}_R[T(X) = P] \geq 1/3$ then we can convert $T$ into a test for $P$ vs. $R$ with only a constant factor increase in sample complexity.
Lower Bound: Distinguishing $P$ and $R$

\[ P = \tau P'' + (1 - \tau)P' \quad \text{and} \quad R = \tau Q'' + (1 - \tau)P' \]

Suppose $T : \Omega^n \rightarrow \{P, R\}$ distinguishes between $P$ and $R$ with $n$ samples.
Lower Bound: Distinguishing \( P \) and \( R \)

\[
P = \tau P'' + (1 - \tau) P' \quad \text{and} \quad R = \tau Q'' + (1 - \tau) P'
\]

Suppose \( T : \Omega^n \rightarrow \{P, R\} \) distinguishes between \( P \) and \( R \) with \( n \) samples.

\( \tau n \) samples from \( P'' \) or \( Q'' \)
Lower Bound: Distinguishing $P$ and $R$

\[ P = \tau P'' + (1 - \tau)P' \quad \text{and} \quad R = \tau Q'' + (1 - \tau)P' \]

Suppose $T : \Omega^n \rightarrow \{P, R\}$ distinguishes between $P$ and $R$ with $n$ samples.

\[ \tau n \text{ samples from } P'' \text{ or } Q'' \]
\[ \rightarrow \]
\[ (1 - \tau) n \text{ samples from } P' \]
Lower Bound: Distinguishing $P$ and $R$

\[ P = \tau P'' + (1 - \tau)P' \quad \text{and} \quad R = \tau Q'' + (1 - \tau)P' \]

Suppose $T : \Omega^n \rightarrow \{P, R\}$ distinguishes between $P$ and $R$ with $n$ samples.

\[ \tau n \text{ samples from } P'' \text{ or } Q'' \]

\[ (1 - \tau)n \text{ samples from } P' \]

\[ T \rightarrow \begin{align*} P & \mapsto P'' \\ R & \mapsto Q'' \end{align*} \]
Lower Bound: Distinguishing $P$ and $R$

\[ P = \tau P'' + (1 - \tau)P' \quad \text{and} \quad R = \tau Q'' + (1 - \tau)P' \]

Suppose $T : \Omega^n \to \{P, R\}$ distinguishes between $P$ and $R$ with $n$ samples.

Then this modified algorithm distinguishes between $P''$ and $Q''$ with $\tau n$ samples.

\[ \tau n \geq SC_{\epsilon}^{P'', Q''} = \Omega \left( \frac{1}{\epsilon} \right) \]
Lower Bound

\[ SC^{P,R}_\epsilon = \Omega \left( \frac{1}{\epsilon \tau} \right) \]

and similarly,

\[ SC^{R,Q}_\epsilon = \Omega \left( \frac{1}{1 - \tau} SC^{P',Q'}_\epsilon \right) = \Omega \left( \frac{1}{(1 - \tau) H^2(P', Q')} \right). \]

Lower Bound

\[ SC^{P,Q}_\epsilon = \Omega \left( \min \left\{ SC^{P,R}_\epsilon, SC^{R,Q}_\epsilon \right\} \right) \]

\[ = \Omega \left( \min \left\{ \frac{1}{\epsilon \tau}, \frac{1}{(1 - \tau) H^2(P', Q')} \right\} \right) \]
Lower Bound

\[ SC_{\epsilon}^{P,R} = \Omega \left( \frac{1}{\epsilon \tau} \right) \]

and similarly,

\[ SC_{\epsilon}^{R,Q} = \Omega \left( \frac{1}{1 - \tau} SC_{\epsilon}^{P', Q'} \right) = \Omega \left( \frac{1}{(1 - \tau) H^2(P', Q')} \right). \]

In paper, we prove the bound within a more general “coupling” framework, in the style of [AcharyaSunZhang17].
### Nonprivate Solution

\[ \text{LLR}(X) = \sum_{i=1}^{n} \ln \frac{P(x_i)}{Q(x_i)} \]

\[ \text{LLR}(X) = \begin{cases} P & \text{if } \text{LLR}(X) \geq 0 \\ Q & \text{if } \text{LLR}(X) < 0 \end{cases} \]

Exactly optimal sample complexity.

\[ \text{SC} = \Theta \left( \frac{1}{H^2(P, Q)} \right) \]

### Private Solution

\[ \text{cLLR}(X) = \sum_{i=1}^{n} \left[ \ln \frac{P(x_i)}{Q(x_i)} \right]^\epsilon \]

\[ \text{scLLR}(X) = \begin{cases} P & \text{w.p. } \frac{e^{\frac{1}{2} \text{cLLR}(X)}}{1+e^{\frac{1}{2} \text{cLLR}(X)}} \\ Q & \text{w.p. } \frac{1}{1+e^{\frac{1}{2} \text{cLLR}(X)}} \end{cases} \]

Optimal sample complexity.

\[ \text{SC} = \Theta \left( \frac{1}{\epsilon \tau + (1-\tau)H^2(P', Q')} \right) \]
Conclusion

**Nonprivate Solution**

\[
\text{LLR}(X) = \sum_{i=1}^{n} \ln \frac{P(x_i)}{Q(x_i)}
\]

\[
\text{sLLR}(X) = \begin{cases} 
P & \text{w.p.} \frac{e^{\frac{1}{2} \text{LLR}(X)}}{1 + e^{\frac{1}{2} \text{LLR}(X)}} \\
Q & \text{w.p.} \frac{1}{1 + e^{\frac{1}{2} \text{LLR}(X)}} 
\end{cases}
\]

optimal sample complexity.

\[
\text{SC} = \Theta \left( \frac{1}{H^2(P, Q)} \right)
\]

**Private Solution**

\[
\text{cLLR}(X) = \sum_{i=1}^{n} \left[ \ln \frac{P(x_i)}{Q(x_i)} \right]^{-\epsilon}
\]

\[
\text{scLLR}(X) = \begin{cases} 
P & \text{w.p.} \frac{e^{\frac{1}{2} \text{cLLR}(X)}}{1 + e^{\frac{1}{2} \text{cLLR}(X)}} \\
Q & \text{w.p.} \frac{1}{1 + e^{\frac{1}{2} \text{cLLR}(X)}} 
\end{cases}
\]

Optimal sample complexity.

\[
\text{SC} = \Theta \left( \frac{1}{\epsilon \tau + (1-\tau)H^2(P', Q')} \right)
\]
**Conclusion**

<table>
<thead>
<tr>
<th>Nonprivate Solution</th>
<th>Private Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{LLR}(X) = \sum_{i=1}^{n} \ln \frac{P(x_i)}{Q(x_i)} )</td>
<td>( \text{cLLR}(X) = \sum_{i=1}^{n} \left[ \ln \frac{P(x_i)}{Q(x_i)} \right]^{\epsilon} )</td>
</tr>
<tr>
<td>( \text{sLLR}(X) = \begin{cases} P \quad \text{w.p.} \quad \frac{e^{\frac{1}{2} \text{LLR}(X)}}{1+e^{\frac{1}{2} \text{LLR}(X)}} \ Q \quad \text{w.p.} \quad \frac{1}{1+e^{\frac{1}{2} \text{LLR}(X)}} \end{cases} )</td>
<td>( \text{scLLR}(X) = \begin{cases} P \quad \text{w.p.} \quad \frac{e^{\frac{1}{2} \text{cLLR}(X)}}{1+e^{\frac{1}{2} \text{cLLR}(X)}} \ Q \quad \text{w.p.} \quad \frac{1}{1+e^{\frac{1}{2} \text{cLLR}(X)}} \end{cases} )</td>
</tr>
</tbody>
</table>

**optimal sample complexity.**

\( \text{SC} = \Theta \left( \frac{1}{H^2(P, Q)} \right) \)

**Optimal sample complexity.**

\( \text{SC} = \Theta \left( \frac{1}{\epsilon \tau + (1-\tau)H^2(P', Q')} \right) \)

---

**In-The-Paper Exclusive:** Optimal algorithms for private change point detection.
Some Open Problems

- How can the characterization be “lifted” to multiple testing problems?
Some Open Problems

- How can the characterization be “lifted” to multiple testing problems?

- What are good general methods for making statistical analyses differentially private?
  - Our result: simple transformation for simple tests
Some Open Problems

- How can the characterization be “lifted” to multiple testing problems?

- What are good general methods for making statistical analyses differentially private?
  - Our result: simple transformation for simple tests

- How should we design validation systems that answer adaptively posed hypothesis tests?
Some Open Problems

- How can the characterization be “lifted” to multiple testing problems?

- What are good general methods for making statistical analyses differentially private?
  - Our result: simple transformation for simple tests

- How should we design validation systems that answer adaptively posed hypothesis tests?

Thank you!