Simple Binary Hypothesis Testing: Locally Private and Communication-Efficient

Ankit Pensia

Algorithms Seminar,
Google

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Joint Work With

Amir Asadi

Varun Jog

Po-Ling Loh

Outline

- Motivation
- Problem Statement
- Our Results
- Proof Sketch
- Conclusion

Simple Hypothesis Testing: Centralized

- Let p and q be two known distributions over $\{1, \ldots, k\}$

Problem (Simple Hypothesis Testing):

Input: i.i.d. samples from either p or q

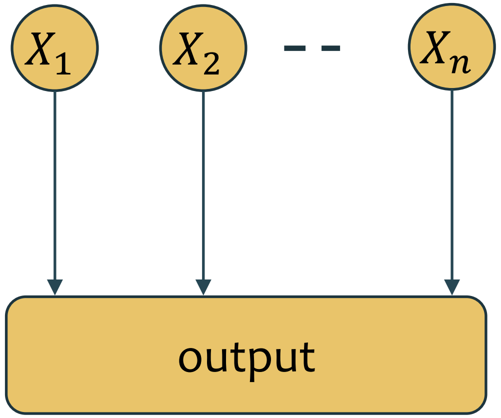
Simple Hypothesis Testing: Centralized

- Let p and q be two known distributions over $\{1, \ldots, k\}$

Problem (Simple Hypothesis Testing):

Input: i.i.d. samples from either p or q

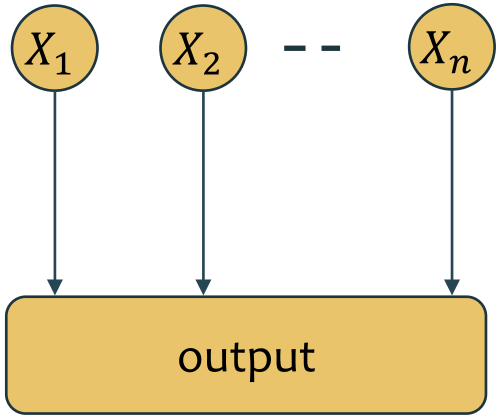
Output: whether they came from p or q



Simple Hypothesis Testing: Centralized

- Let p and q be two known distributions over $\{1, \dots, k\}$

Problem (Simple Hypothesis Testing):
Input: i.i.d. samples from either p or q
Output: whether they came from p or q



- Arguably, the most fundamental statistical problem
  - A natural building block
  - Optimal test: Likelihood ratio test

Simple Hypothesis Testing: Centralized

- Let p and q be two known distributions over $\{1, \ldots, k\}$

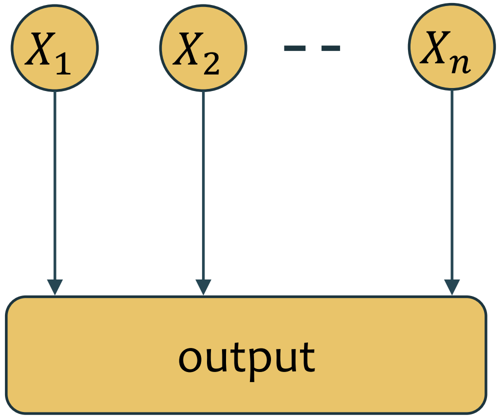
Problem (Simple Hypothesis Testing):

Input: i.i.d. samples from either p or q

Output: whether they came from p or q

- Arguably, the most fundamental statistical problem
  - A natural building block
  - Optimal test: Likelihood ratio test

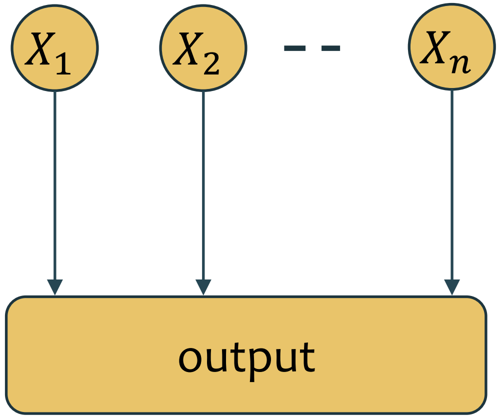
Requires access to X_i's



Simple Hypothesis Testing: Centralized

- Let p and q be two known distributions over $\{1, \ldots, k\}$

Problem (Simple Hypothesis Testing):
Input: i.i.d. samples from either p or q
Output: whether they came from p or q



- Arguably, the most fundamental statistical problem
  - A natural building block
  - Optimal test: Likelihood ratio test

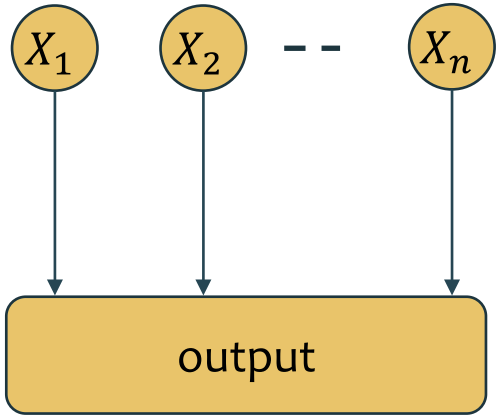
Requires access to X_i's

- Data is distributed these days
  - Limited communication bandwidth
  - Privacy concerns

Simple Hypothesis Testing: Centralized

- Let p and q be two known distributions over $\{1, \ldots, k\}$

Problem (Simple Hypothesis Testing):
Input: i.i.d. samples from either p or q
Output: whether they came from p or q



- Arguably, the most fundamental statistical problem
  - A natural building block
  - Optimal test: Likelihood ratio test
- Data is distributed these days
  - Limited communication bandwidth
  - Privacy concerns

Requires access to X_i's

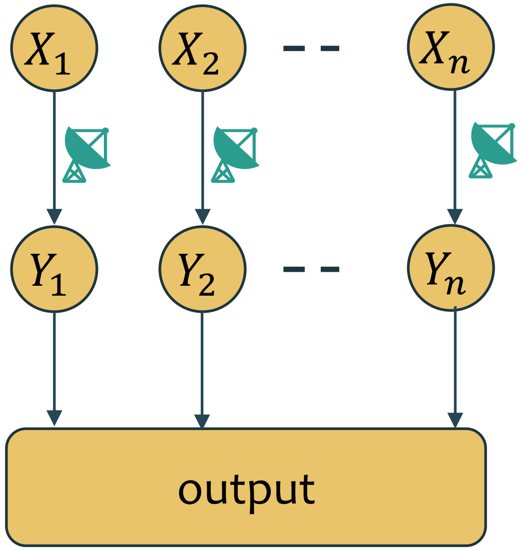
Requires quantizing/privatizing X_i's

Simple Hypothesis Testing: Decentralized

- Let p and q be two known distributions over $\{1, \ldots, k\}$

> Problem (Simple Hypothesis Testing):
>
> Input: i.i.d. samples from either p or q
>
> Output: whether they came from p or q

- : captures communication and/or privacy



Simple Hypothesis Testing: Decentralized

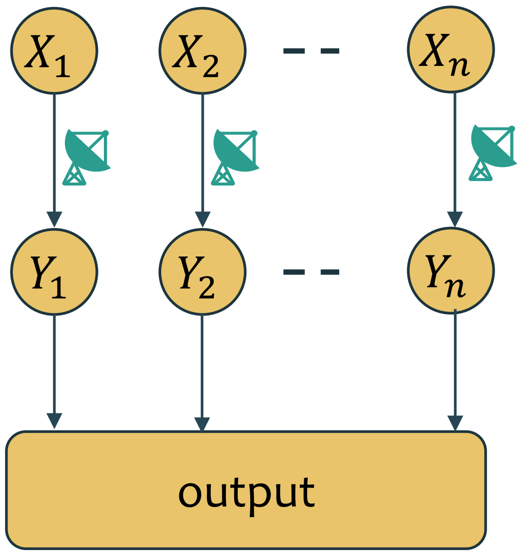
- Let p and q be two known distributions over $\{1, \dots, k\}$

Problem (Decentralized Simple Hypothesis Testing):

Input: modified samples from either p or q

Output: whether they came from p or q

- : captures communication and/or privacy

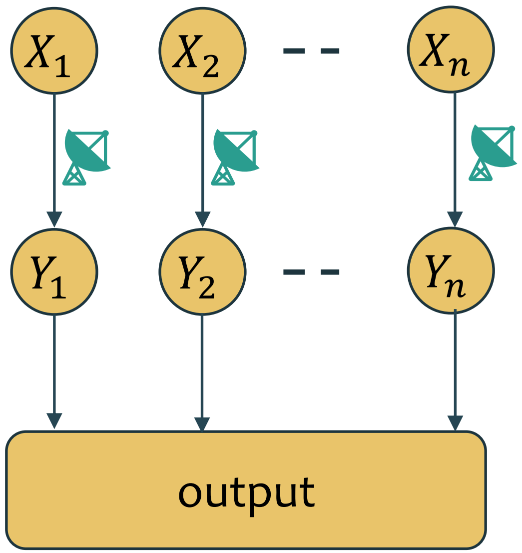


Simple Hypothesis Testing: Decentralized

- Let p and q be two known distributions over $\{1, \ldots, k\}$

Problem (Decentralized Simple Hypothesis Testing):
Input: modified samples from either p or q
Output: whether they came from p or q

- : captures communication and/or privacy

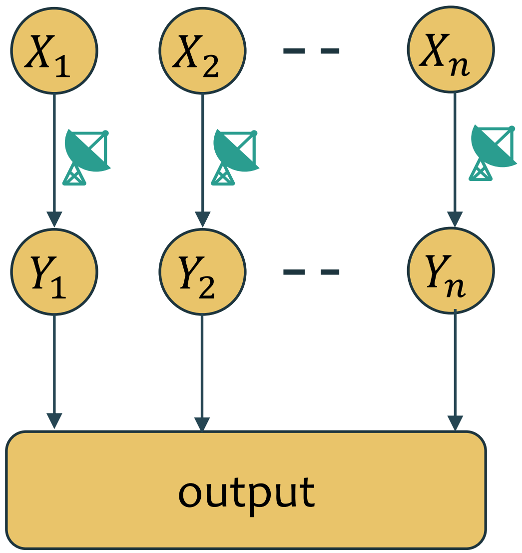


[Tsi93] J. Tsitsiklis. Decentralized Detection. 1993

Simple Hypothesis Testing: Decentralized

- Let p and q be two known distributions over $\{1, \dots, k\}$

Problem (Decentralized Simple Hypothesis Testing):
  Input: modified samples from either p or q
  Output: whether they came from p or q



- : captures communication and/or privacy

How do we perform decentralized hypothesis testing?

[Tsi93] J. Tsitsiklis. Decentralized Detection. 1993

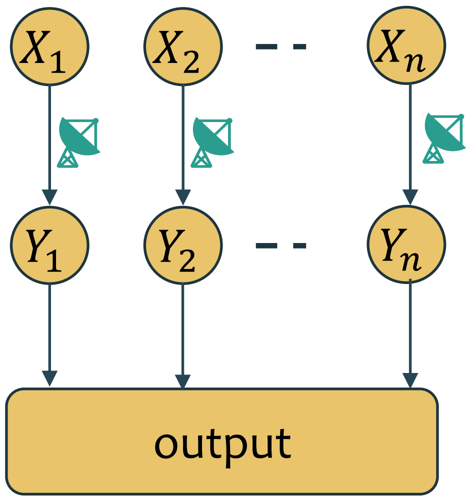
Outline

- Motivation
- **Problem Statement**
- Our Results
- Proof Sketch
- Conclusion

Privacy Model and Communication Constraints

- Local Differential Privacy (LDP)
  - Everyone releases a randomized version of data
  - Channel is ϵ-LDP if:

$$\frac{\mathbb{P}(Y_i = y \mid X_i = x)}{\mathbb{P}(Y_i = y \mid X_i = x')} \leq e^{\epsilon} \quad \text{for all } x, x', y$$

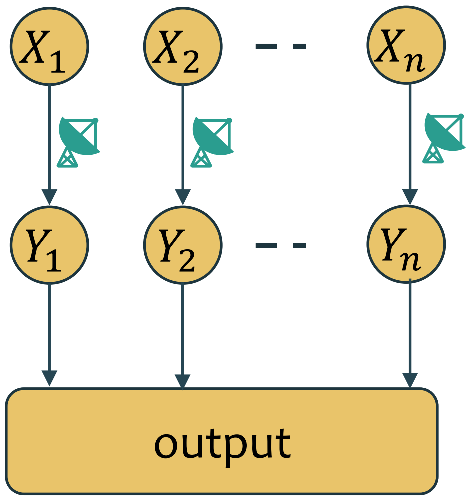


Privacy Model and Communication Constraints

- Local Differential Privacy (LDP)
  - Everyone releases a randomized version of data
  - Channel is ϵ-LDP if:

$$\frac{\mathbb{P}(Y_i = y \mid X_i = x)}{\mathbb{P}(Y_i = y \mid X_i = x')} \leq e^{\epsilon} \quad \text{for all } x, x', y$$

Can't reliably distinguish between x and x' using values of Y_i



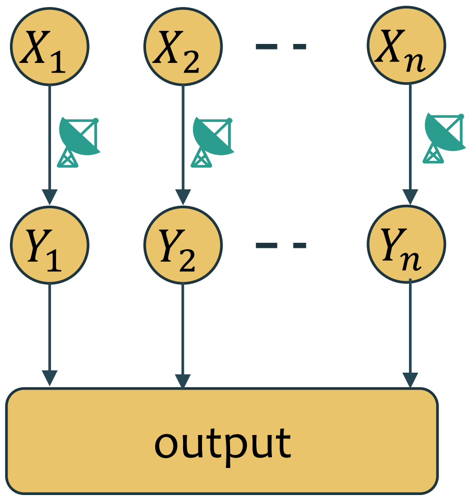
Privacy Model and Communication Constraints

- Local Differential Privacy (LDP)
  - Everyone releases a randomized version of data
  - Channel is ϵ-LDP if:

$$\frac{\mathbb{P}(Y_i = y \mid X_i = x)}{\mathbb{P}(Y_i = y \mid X_i = x')} \le e^{\epsilon} \quad \text{for all } x, x', y$$

Can't reliably distinguish between x and x' using values of Y_i

  - Non-interactive (private-coin): Y_i's are independent



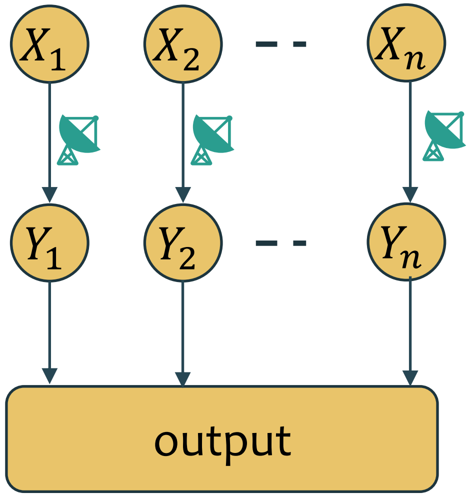
Privacy Model and Communication Constraints

- Local Differential Privacy (LDP)
  - Everyone releases a randomized version of data
  - Channel is ϵ-LDP if:

$$\frac{\mathbb{P}(Y_i = y \mid X_i = x)}{\mathbb{P}(Y_i = y \mid X_i = x')} \leq e^{\epsilon} \quad \text{for all } x, x', y$$

Can't reliably distinguish between x and x' using values of Y_i

  - Non-interactive (private-coin): Y_i's are independent

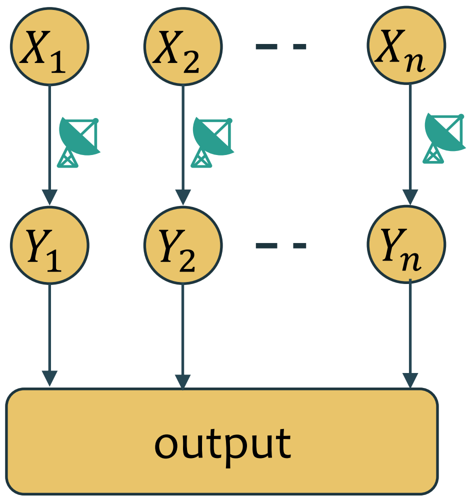


- Communication-constraints
  - $Y_i \in \{1, \ldots, \ell\}$ for some $\ell \ll k$

Questions of Interest

Problem (Decentralized Simple Hypothesis Testing):
Input: modified samples from either p or q
Output: whether they came from p or q

Goal: Design the test and channels so that the $\boxed{\text{probability of error} \leq 0.1}$

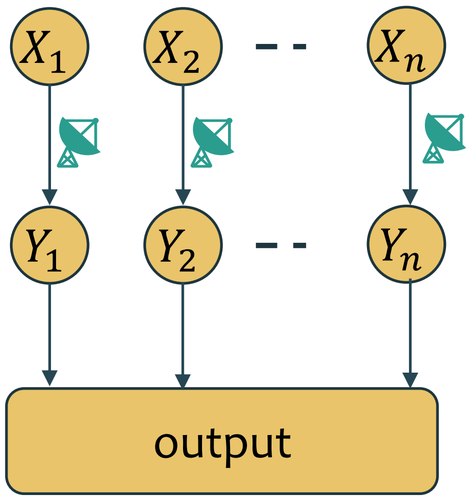


Problem (Decentralized Simple Hypothesis Testing):
Input: modified samples from either p or q
Output: whether they came from p or q

Questions of Interest

Goal: Design the test and channels so that the $\boxed{\text{probability of error} \leq 0.1}$

Sample Complexity: Minimum n to achieve above goal



Questions of Interest

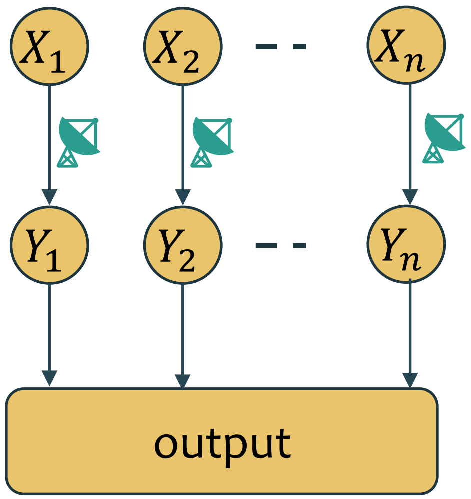
Problem (Decentralized Simple Hypothesis Testing):
Input: modified samples from either p or q
Output: whether they came from p or q

Goal: Design the test and channels so that the $\boxed{\text{probability of error} \leq 0.1}$

Sample Complexity: Minimum n to achieve above goal

n^*_{original} := Sample complexity (no constraints)

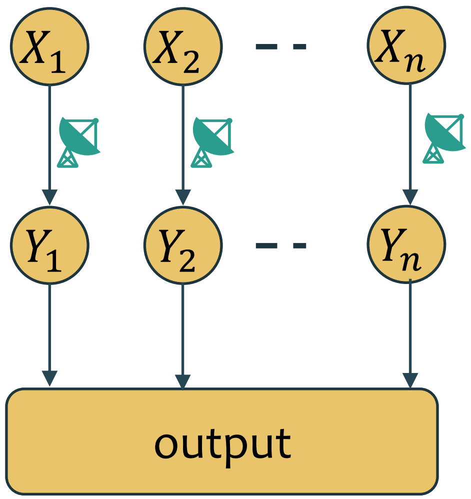
$n^*_{\text{constraints}}$:= Sample complexity with channels satisfying constraints



Problem (Decentralized Simple Hypothesis Testing):
Input: modified samples from either p or q
Output: whether they came from p or q

Questions of Interest

Goal: Design the test and channels so that the $\boxed{\text{probability of error} \leq 0.1}$



Sample Complexity: Minimum n to achieve above goal

n^*_{original} := Sample complexity (no constraints)

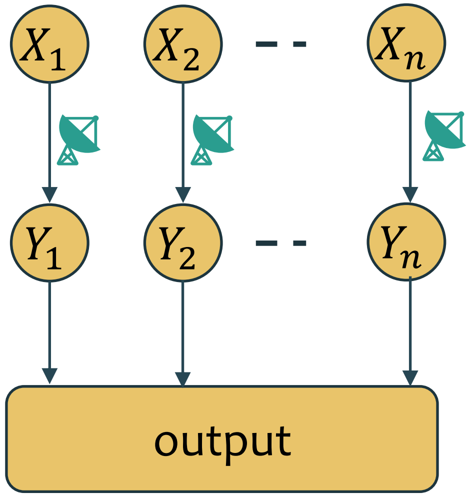
$n^*_{\text{constraints}}$:= Sample complexity with channels satisfying constraints

Questions:

Problem (Decentralized Simple Hypothesis Testing):
Input: modified samples from either p or q
Output: whether they came from p or q

Questions of Interest

Goal: Design the test and channels so that the $\boxed{\text{probability of error} \leq 0.1}$



Sample Complexity: Minimum n to achieve above goal

n^*_{original} := Sample complexity (no constraints)

$n^*_{\text{constraints}}$:= Sample complexity with channels satisfying constraints

Questions:

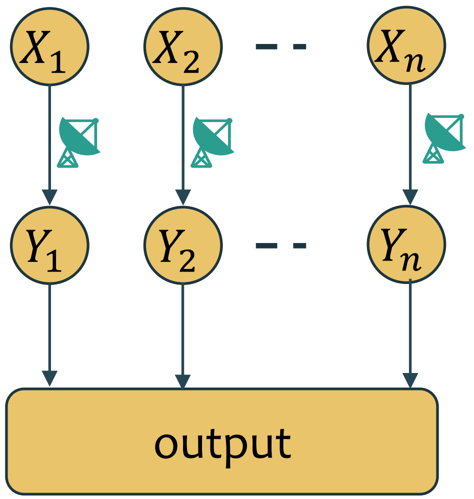
1. (Statistical) How much does sample complexity change? $\boxed{n^*_{\text{original}} \text{ vs. } n^*_{\text{constraints}}}$

Problem (Decentralized Simple Hypothesis Testing):
Input: modified samples from either p or q
Output: whether they came from p or q

Questions of Interest

Goal: Design the test and channels so that the $\boxed{\text{probability of error} \leq 0.1}$

Sample Complexity: Minimum n to achieve above goal



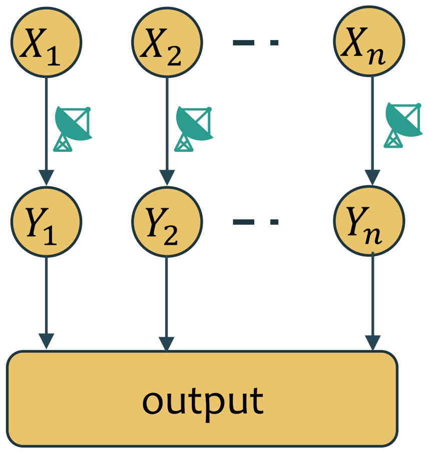
n^*_{original} := Sample complexity (no constraints)

$n^*_{\text{constraints}}$:= Sample complexity with channels satisfying constraints

Questions:

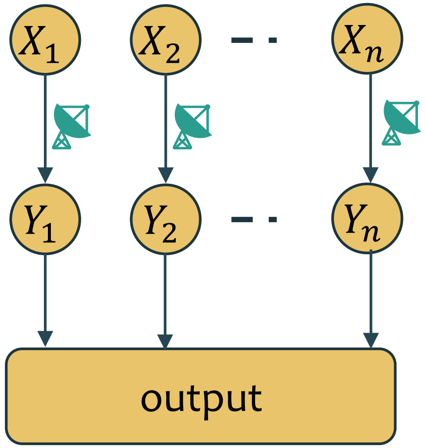
1. (Statistical) How much does sample complexity change? $\boxed{n^*_{\text{original}} \text{ vs. } n^*_{\text{constraints}}}$
2. (Computational) How to find (near)-optimal channels fast? $\boxed{\text{polynomial in support size}}$

Warmup: Scheffe's Test (Popular but Sub-optimal)



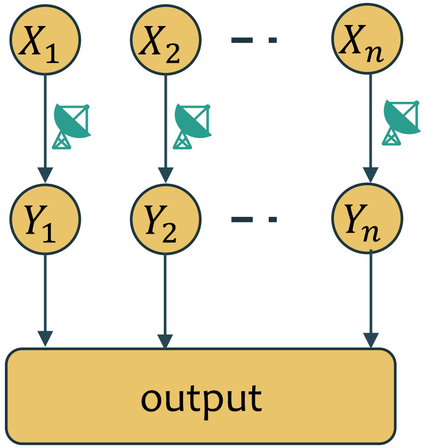
Warmup: Scheffe's Test (Popular but Sub-optimal)

- Scheffe's Test
  - Let $A \subset [k]$ be the set $\{ j : p_j \geq q_j \}$
  - Set $Y_i = 1$ if $X_i \in A$, else 0
  - Output p if $\sum_i Y_i$ is large enough, else q



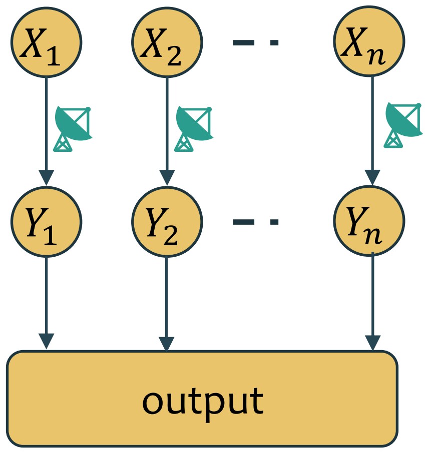
Warmup: Scheffe's Test (Popular but Sub-optimal)

- Scheffe's Test
  - Let $A \subset [k]$ be the set $\{ j : p_j \geq q_j \}$
  - Set $Y_i = 1$ if $X_i \in A$, else 0
  - Output p if $\sum_i Y_i$ is large enough, else q
- Pros: Simple, uses a single bit, and well-studied



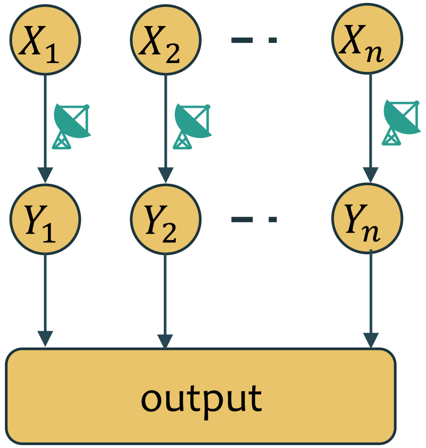
Warmup: Scheffe's Test (Popular but Sub-optimal)

- Scheffe's Test
  - Let $A \subset [k]$ be the set $\{ j : p_j \geq q_j \}$
  - Set $Y_i = 1$ if $X_i \in A$, else 0
  - Output p if $\sum_i Y_i$ is large enough, else q
- Pros: Simple, uses a single bit, and well-studied
- Cons: Sample complexity can increase quadratically!



Warmup: Scheffe's Test (Popular but Sub-optimal)

- Scheffe's Test
  - Let $A \subset [k]$ be the set $\{ j : p_j \geq q_j \}$
  - Set $Y_i = 1$ if $X_i \in A$, else 0
  - Output p if $\sum_i Y_i$ is large enough, else q
- Pros: Simple, uses a single bit, and well-studied
- Cons: Sample complexity can increase quadratically!

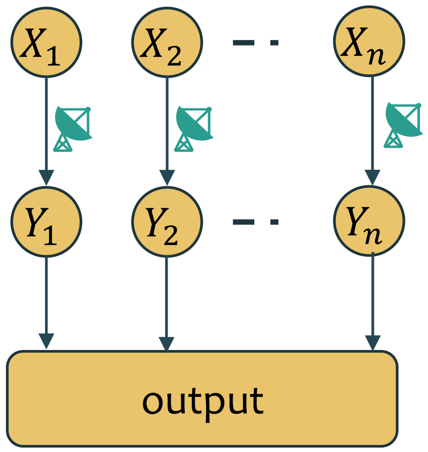


Example

$$p = \begin{pmatrix} 0.5 - 2\alpha \\ 0.5 + \alpha \\ \alpha \end{pmatrix} \qquad q = \begin{bmatrix} 0.5 \\ 0.5 \\ 0 \end{bmatrix}$$

Warmup: Scheffe's Test (Popular but Sub-optimal)

- Scheffe's Test
  - Let $A \subset [k]$ be the set $\{ j : p_j \geq q_j \}$
  - Set $Y_i = 1$ if $X_i \in A$, else 0
  - Output p if $\sum_i Y_i$ is large enough, else q
- Pros: Simple, uses a single bit, and well-studied
- Cons: Sample complexity can increase quadratically!



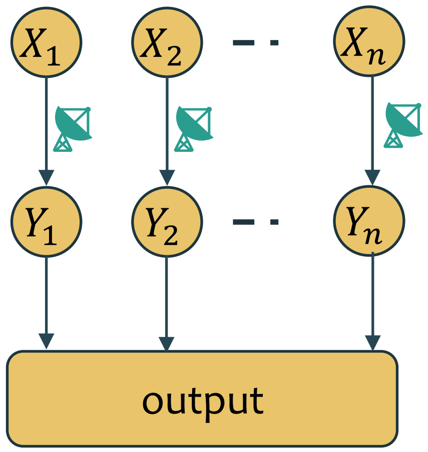
Example

$$p = \begin{pmatrix} 0.5 - 2\alpha \\ 0.5 + \alpha \\ \alpha \end{pmatrix} \quad q = \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix}$$

Needs only $1/\alpha$ samples

Warmup: Scheffe's Test (Popular but Sub-optimal)

- Scheffe's Test
  - Let $A \subset [k]$ be the set $\{ j : p_j \geq q_j \}$
  - Set $Y_i = 1$ if $X_i \in A$, else 0
  - Output p if $\sum_i Y_i$ is large enough, else q
- Pros: Simple, uses a single bit, and well-studied
- Cons: Sample complexity can increase quadratically!



Example

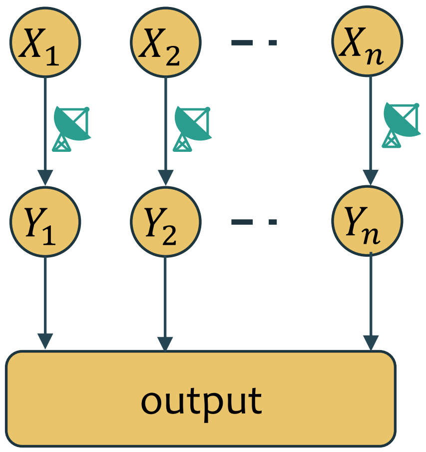
$$p = \begin{pmatrix} 0.5 - 2\alpha \\ 0.5 + \alpha \\ \alpha \end{pmatrix} \quad q = \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix}$$

$A = \{2,3\}$

Needs only $1/\alpha$ samples

Warmup: Scheffe's Test (Popular but Sub-optimal)

- Scheffe's Test
  - Let $A \subset [k]$ be the set $\{\, j : p_j \geq q_j \,\}$
  - Set $Y_i = 1$ if $X_i \in A$, else 0
  - Output p if $\sum_i Y_i$ is large enough, else q
- Pros: Simple, uses a single bit, and well-studied
- Cons: Sample complexity can increase quadratically!



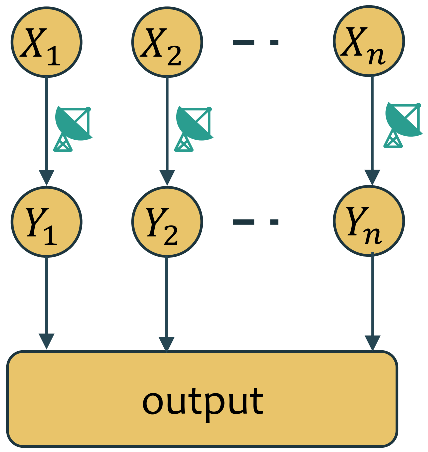
Example

$$p = \begin{pmatrix} 0.5 - 2\alpha \\ \boxed{0.5 + \alpha} \\ \boxed{\alpha} \end{pmatrix} \quad q = \begin{pmatrix} 0.5 \\ \boxed{0.5} \\ \boxed{0} \end{pmatrix} \quad A = \{2,3\} \quad p' = \begin{pmatrix} 0.5 - 2\alpha \\ 0.5 + 2\alpha \end{pmatrix} \quad q' = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

Needs only $1/\alpha$ samples

Warmup: Scheffe’s Test (Popular but Sub-optimal)

- Scheffe’s Test
  - Let $A \subset [k]$ be the set $\{ j: p_j \geq q_j \}$
  - Set $Y_i = 1$ if $X_i \in A$, else 0
  - Output p if $\sum_i Y_i$ is large enough, else q



- Pros: Simple, uses a single bit, and well-studied
- Cons: Sample complexity can increase quadratically!

Example

$$p = \begin{pmatrix} 0.5 - 2\alpha \\ 0.5 + \alpha \\ \alpha \end{pmatrix} \quad q = \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix}$$

$A = \{2,3\}$

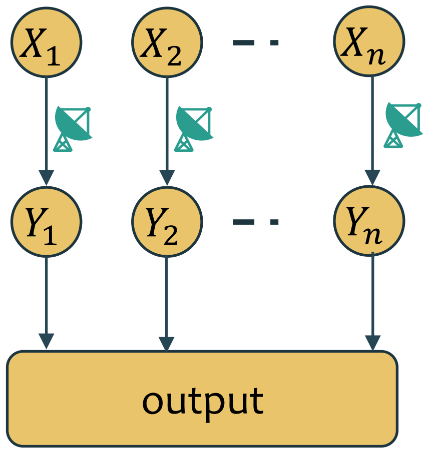
Needs only $1/\alpha$ samples

$$p' = \begin{pmatrix} 0.5 - 2\alpha \\ 0.5 + 2\alpha \end{pmatrix} \quad q' = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

Scheffe’s test needs $\mathbf{1/\alpha^2}$ samples

Warmup: Scheffe's Test (Popular but Sub-optimal)

- Scheffe's Test
  - Let $A \subset [k]$ be the set $\{ j : p_j \geq q_j \}$
  - Set $Y_i = 1$ if $X_i \in A$, else 0
  - Output p if $\sum_i Y_i$ is large enough, else q
- Pros: Simple, uses a single bit, and well-studied
- Cons: Sample complexity can increase quadratically!



Example

$p = \begin{pmatrix} 0.5 - 2\alpha \\ 0.5 + \alpha \\ \alpha \end{pmatrix}$ $q = \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix}$ $A = \{2,3\}$

Needs only $1/\alpha$ samples

$p' = \begin{pmatrix} 0.5 - 2\alpha \\ 0.5 + 2\alpha \end{pmatrix}$ $q' = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$

Scheffe's test needs $\mathbf{1/\alpha^2}$ samples

Is this quadratic blowup necessary?

Outline

- Motivation
- Problem Statement
- Our Results
  - Statistical
  - Computational
- Proof Sketch
- Conclusion

$n^*_{\text{Scheffe}} \lesssim (n^*)^2$

Our Results: Statistical Cost Of Communication Constraints

$n^* :=$ Sample complexity without constraints
$n^*_{\text{comm}}(\ell) :=$ Sample complexity with channels of ℓ messages

$n^*_{\text{Scheffe}} \lesssim (n^*)^2$

Our Results: Statistical Cost Of Communication Constraints

$n^* :=$ Sample complexity without constraints
$n^*_{\text{comm}}(\ell) :=$ Sample complexity with channels of ℓ messages

Theorem [PJL22] **(Statistical cost of communication constraints)** For $\ell \geq 2$,

$$n^*_{\text{comm}}(\ell) \lesssim n^* \left(1 + \frac{\log n^*}{\ell}\right)$$

- The sample complexity increases by at most a logarithmic factor

$n^*_{\text{Scheffe}} \lesssim (n^*)^2$

Our Results: Statistical Cost Of Communication Constraints

$n^* :=$ Sample complexity without constraints
$n^*_{\text{comm}}(\ell) :=$ Sample complexity with channels of ℓ messages

Theorem [PJL22] **(Statistical cost of communication constraints)** For $\ell \geq 2$,

$$n^*_{\text{comm}}(\ell) \lesssim n^* \left(1 + \frac{\log n^*}{\ell}\right)$$

Moreover, there exist cases where this is tight.

- The sample complexity increases by at most a logarithmic factor

$n^*_{\text{Scheffe}} \lesssim (n^*)^2$

Our Results: Statistical Cost Of Communication Constraints

n^* := Sample complexity without constraints

$n^*_{\text{comm}}(\ell)$:= Sample complexity with channels of ℓ messages

Theorem [PJL22] **(Statistical cost of communication constraints)** For $\ell \geq 2$,

$$n^*_{\text{comm}}(\ell) \lesssim n^* \left(1 + \frac{\log n^*}{\ell}\right)$$

Moreover, there exist cases where this is tight.

- The sample complexity increases by at most a logarithmic factor
- "Effective" domain size is $\log n^*$

$n^*_{\text{Scheffe}} \lesssim (n^*)^2$

Our Results: Statistical Cost Of Communication Constraints

n^* := Sample complexity without constraints

$n^*_{\text{comm}}(\ell)$:= Sample complexity with channels of ℓ messages

Theorem [PJL22] **(Statistical cost of communication constraints)** For $\ell \geq 2$,

$$n^*_{\text{comm}}(\ell) \lesssim n^* \left(1 + \frac{\log n^*}{\ell}\right)$$

Moreover, there exist cases where this is tight.

- The sample complexity increases by at most a logarithmic factor
- "Effective" domain size is $\log n^*$
- Also holds under additional constraints: robustness, privacy,…

Our Results: Statistical Cost Of Communication Constraints

n^* := Sample complexity without constraints
$n^*_{\mathrm{comm}}(\ell)$:= Sample complexity with channels of ℓ messages

Theorem [PJL22] **(Statistical cost of communication constraints)** For $\ell \geq 2$,

$$n^*_{\mathrm{comm}}(\ell) \lesssim n^* \left(1 + \frac{\log n^*}{\ell}\right)$$

Moreover, there exist cases where this is tight.

- The sample complexity increases by at most a logarithmic factor
- "Effective" domain size is $\log n^*$
- Also holds under additional constraints: robustness, privacy,…
- Closely related to preserving mutual information under quantization

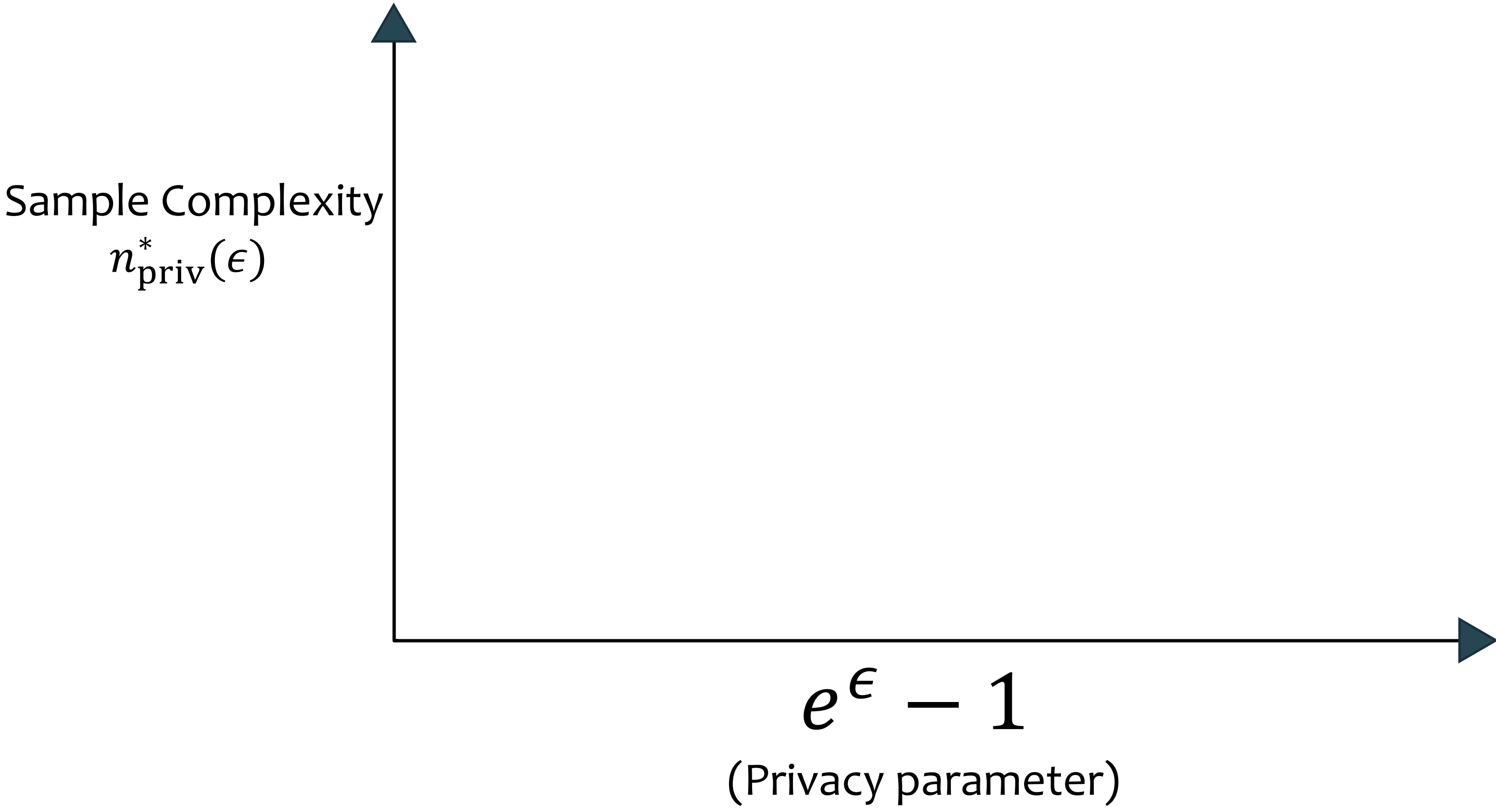
[BNOP21] A. Bhatt, B. Nazer, O. Ordentlich, Y. Polyanskiy. Information-Distilling Quantizers. 2021.

Statistical Cost of Privacy: Existing Results

$n^*_{\text{priv}}(\epsilon) :=$ Sample complexity with ϵ-LDP channels

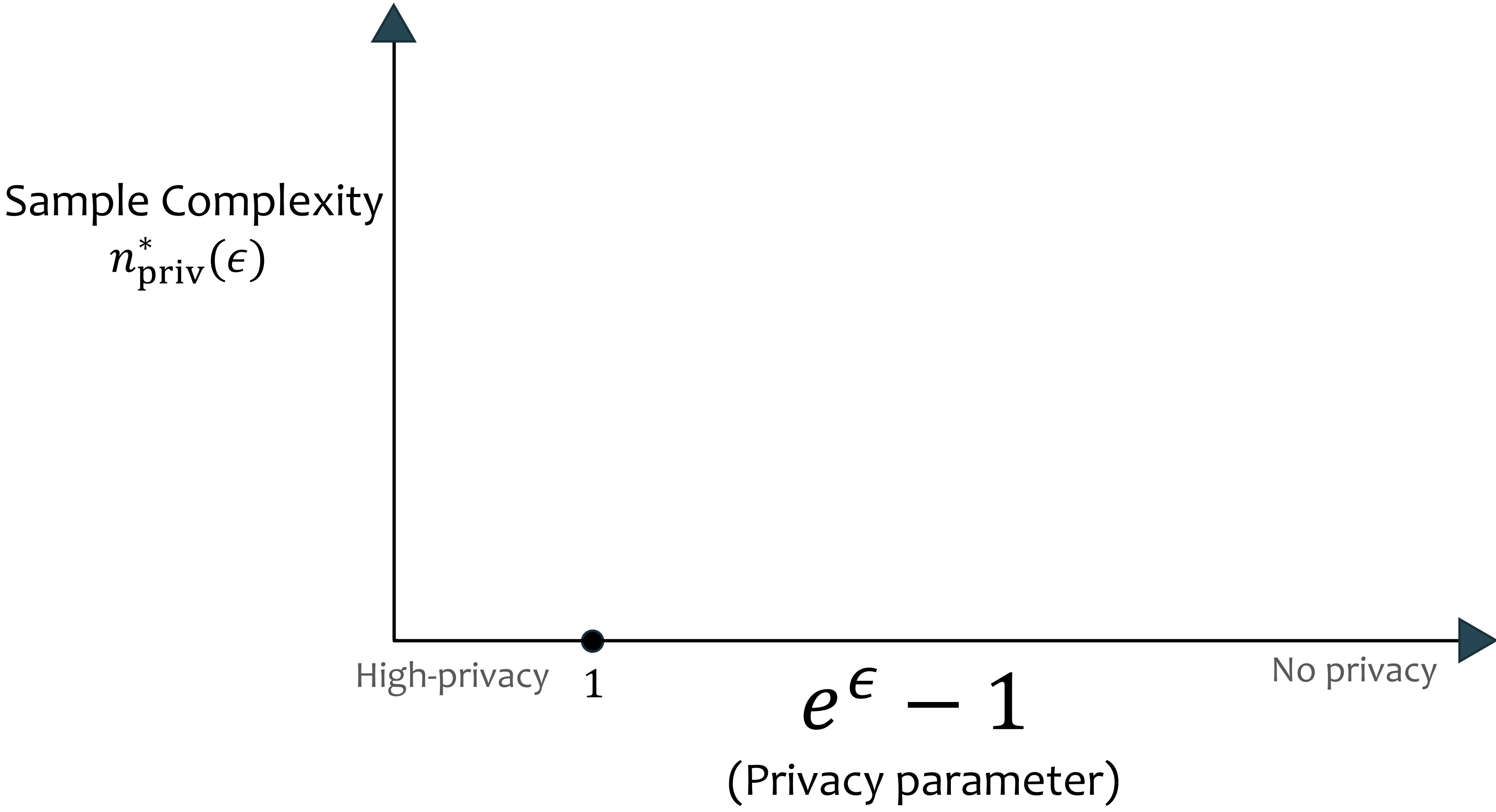
Statistical Cost of Privacy: Existing Results

$n^*_{\text{priv}}(\epsilon) :=$ Sample complexity with ϵ-LDP channels



Statistical Cost of Privacy: Existing Results

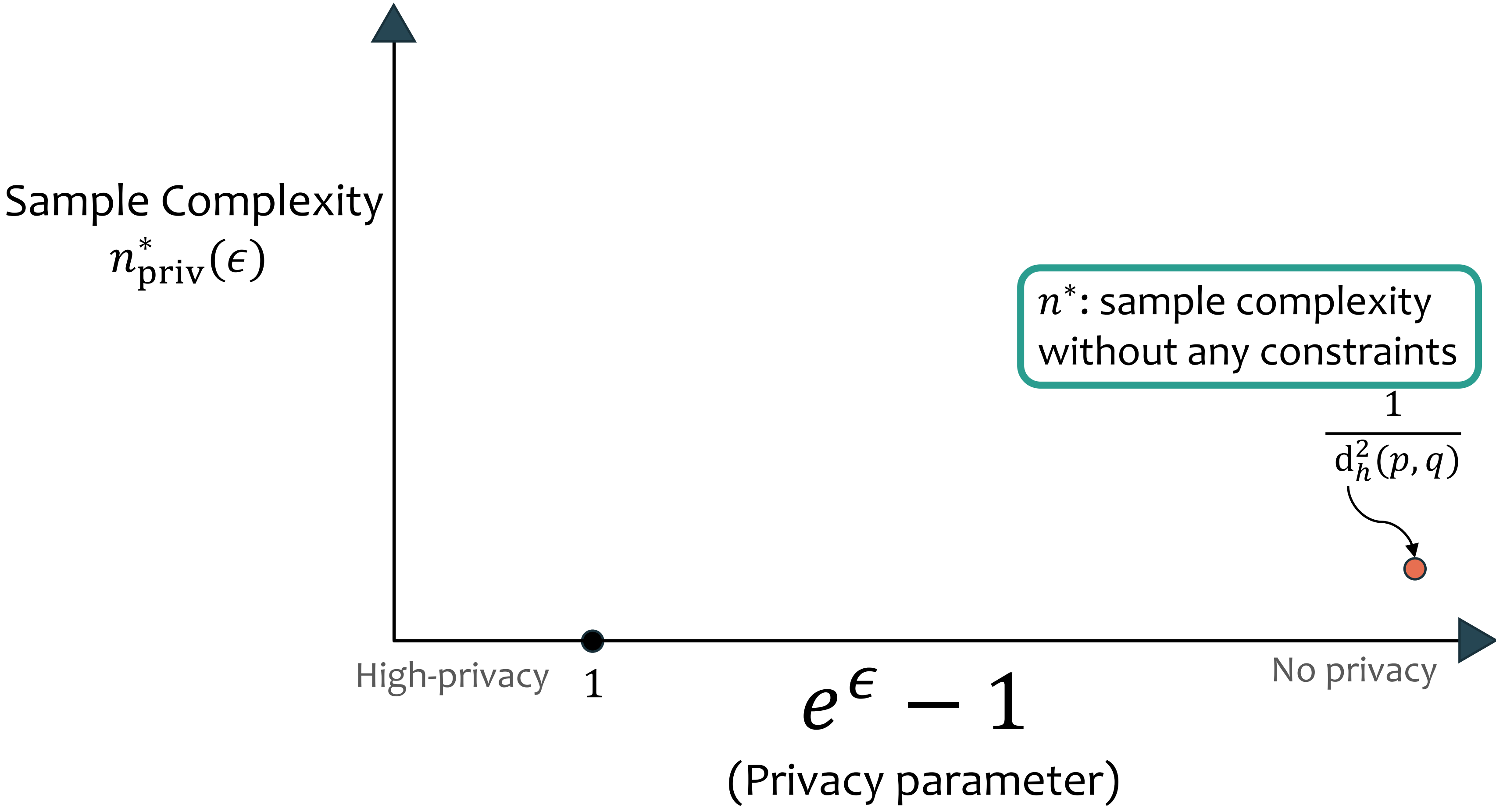
$n^*_{\text{priv}}(\epsilon)$:= Sample complexity with ϵ-LDP channels



d_h^2: Hellinger divergence
d_{TV}: Total variation distance

Statistical Cost of Privacy: Existing Results

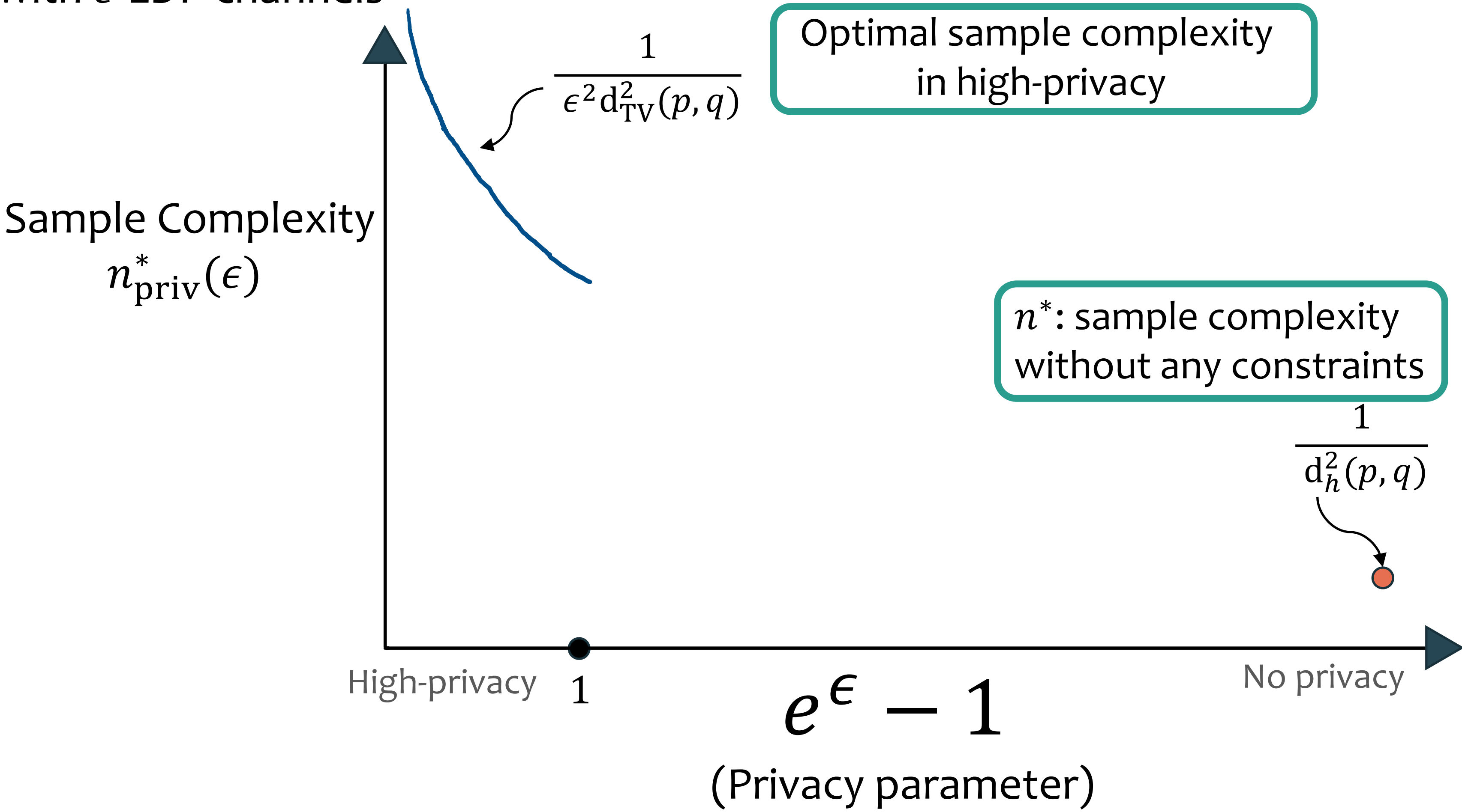
$n^*_{\mathrm{priv}}(\epsilon) :=$ Sample complexity with ϵ-LDP channels



Statistical Cost of Privacy: Existing Results

d_h^2: Hellinger divergence
d_{TV}: Total variation distance

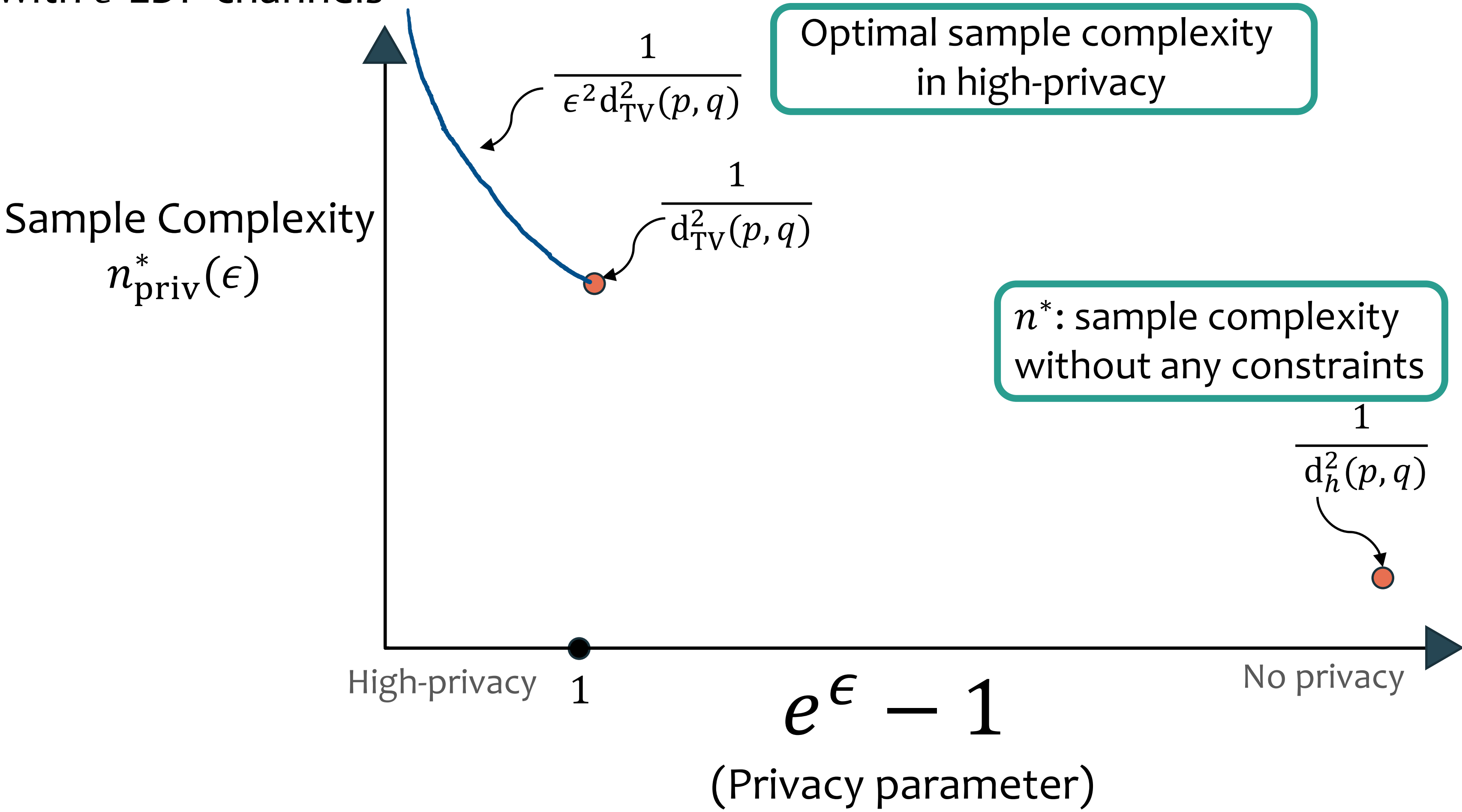
$n^*_{\text{priv}}(\epsilon)$:= Sample complexity with ϵ-LDP channels



[DJW13] J. Duchi, M. Wainwright, M.Jordan. Minimax Optimal Procedures for Locally Private Estimation. 2013.

Statistical Cost of Privacy: Existing Results

$n^*_{\mathrm{priv}}(\epsilon)$:= Sample complexity with ϵ-LDP channels

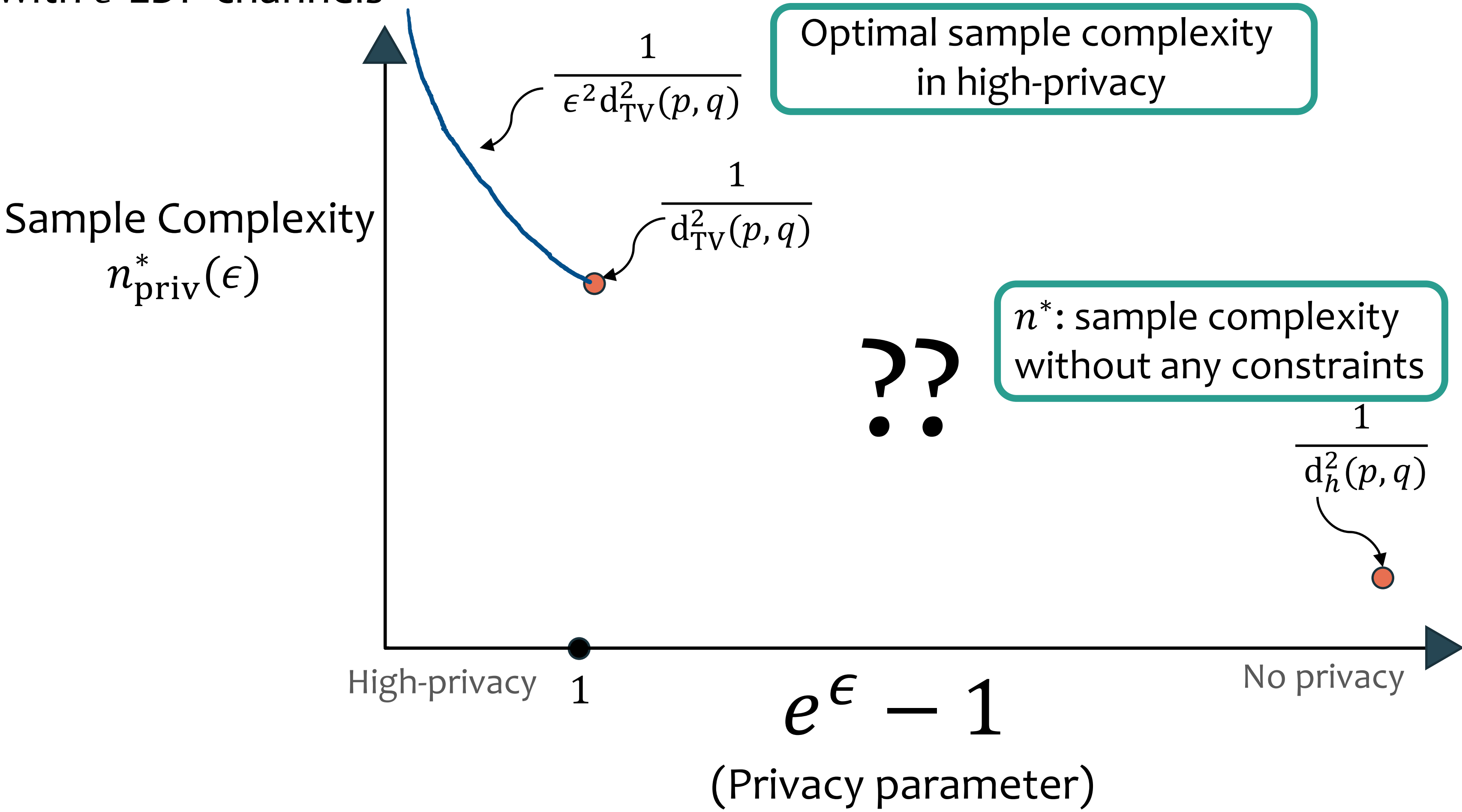


[DJW13] J. Duchi, M. Wainwright, M.Jordan. Minimax Optimal Procedures for Locally Private Estimation. 2013.

d_h^2: Hellinger divergence
d_{TV}: Total variation distance

Statistical Cost of Privacy: Existing Results

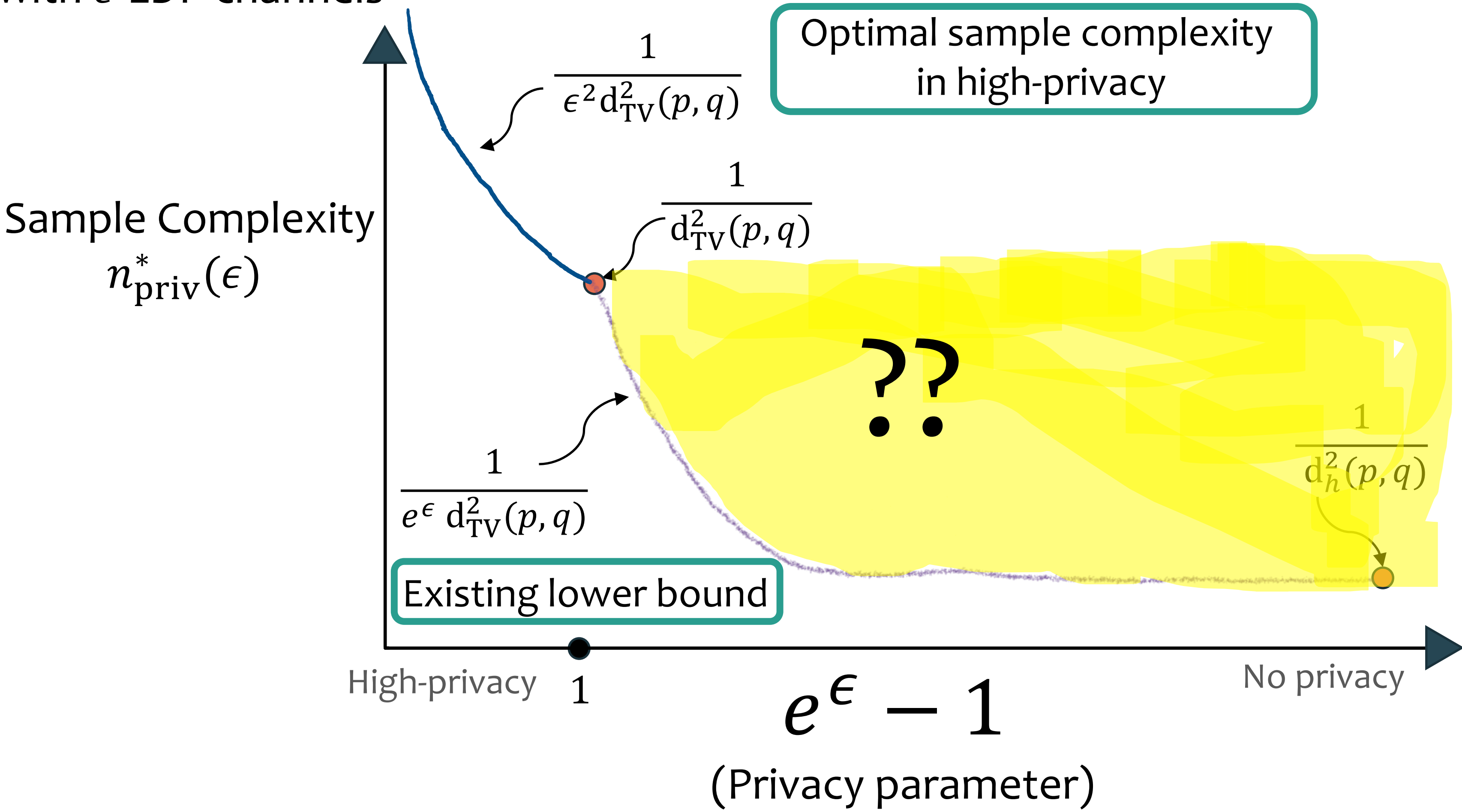
$n^*_{\mathrm{priv}}(\epsilon)$:= Sample complexity with ϵ-LDP channels



[DJW13] J. Duchi, M. Wainwright, M.Jordan. Minimax Optimal Procedures for Locally Private Estimation. 2013.

Statistical Cost of Privacy: Existing Results

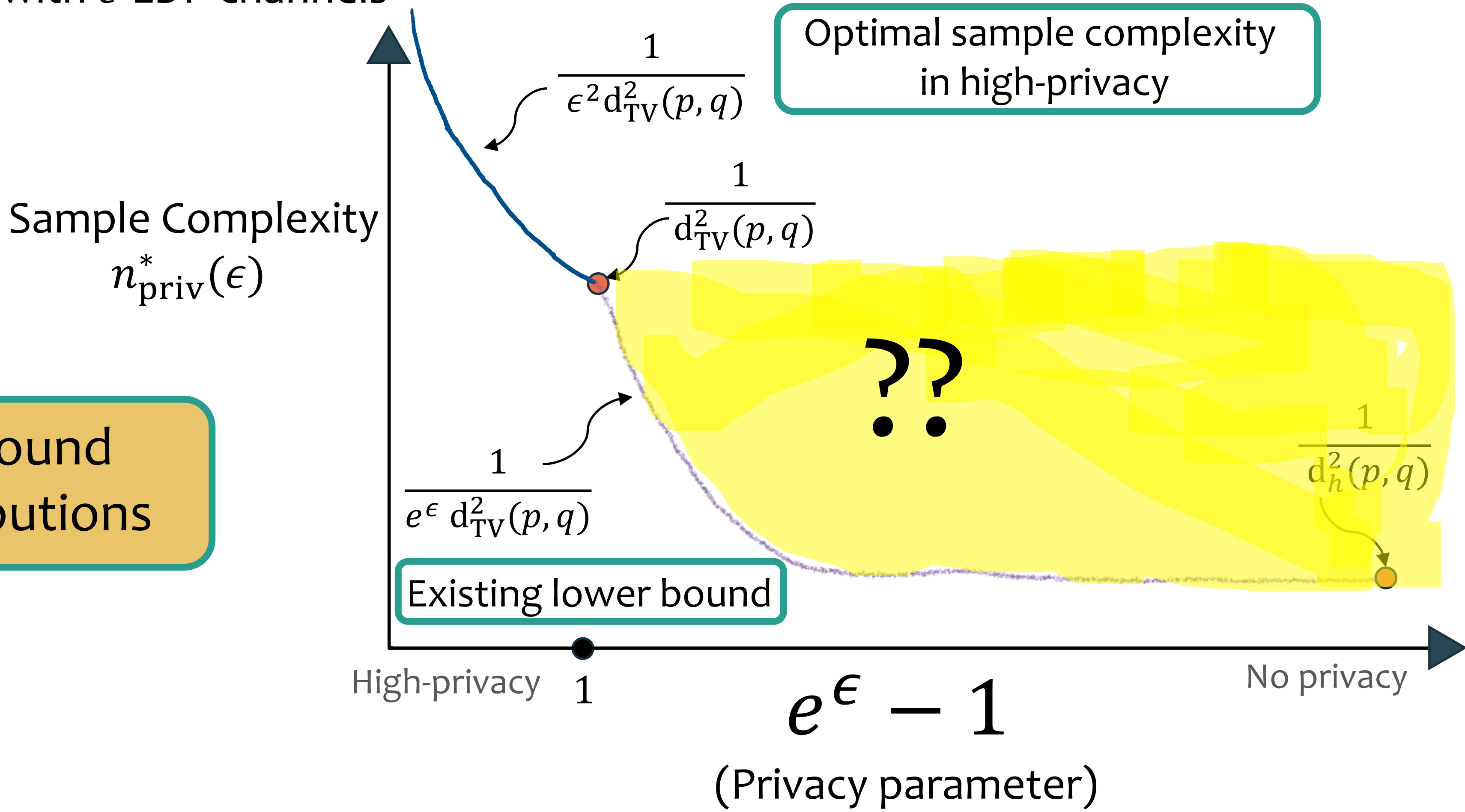
$n^*_{\mathrm{priv}}(\epsilon)$:= Sample complexity with ϵ-LDP channels



[DJW13] J. Duchi, M. Wainwright, M.Jordan. Minimax Optimal Procedures for Locally Private Estimation. 2013.
[AZ22] S. Asoodeh, H. Zhang. Contraction of Locally Private Mechanisms. 2022.

Statistical Cost of Privacy: Existing Results

$n^*_{\mathrm{priv}}(\epsilon)$:= Sample complexity with ϵ-LDP channels

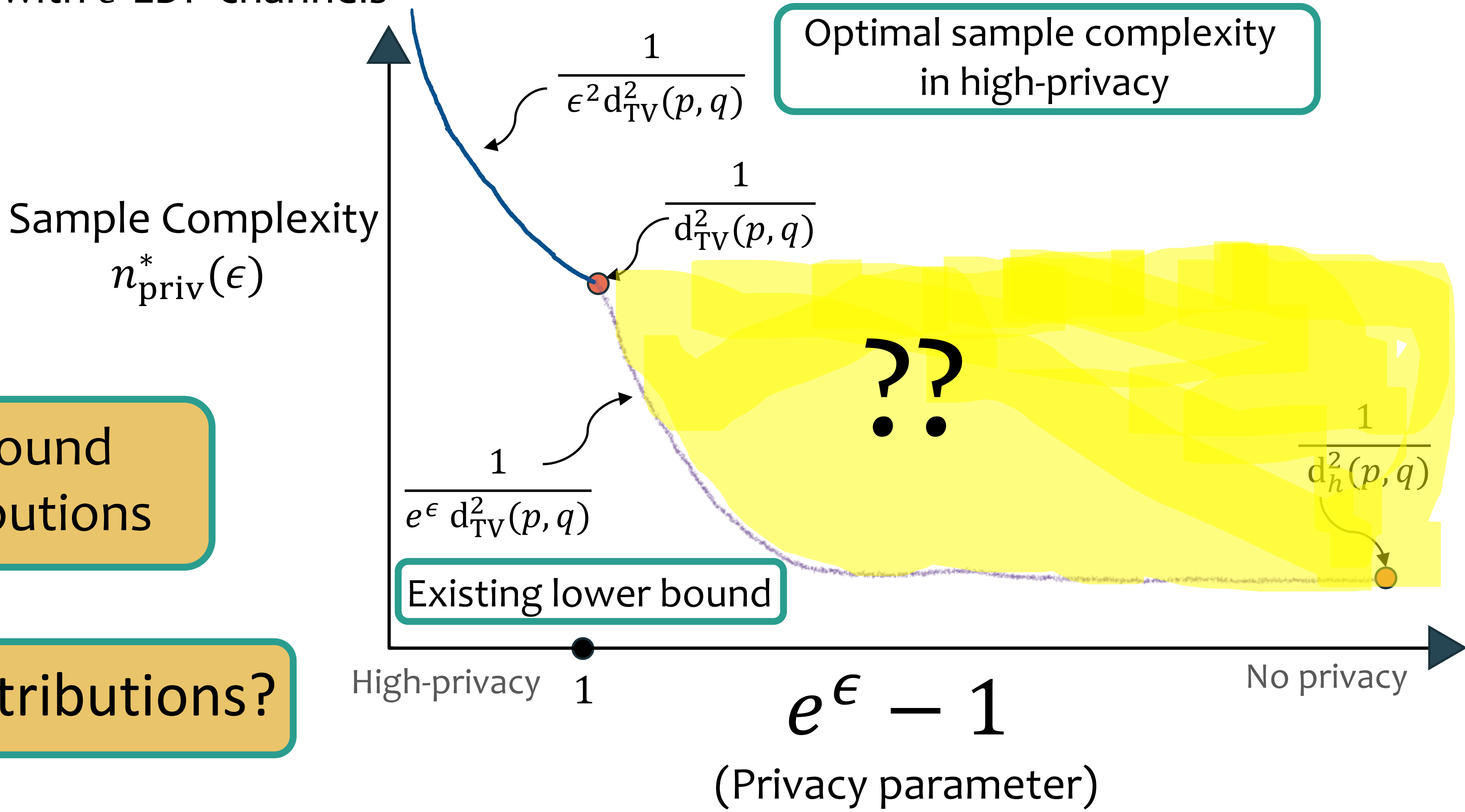


[PAJL23]: Existing lower bound is tight for Bernoulli distributions

[DJW13] J. Duchi, M. Wainwright, M.Jordan. Minimax Optimal Procedures for Locally Private Estimation. 2013.
[AZ22] S. Asoodeh, H. Zhang. Contraction of Locally Private Mechanisms. 2022.

Statistical Cost of Privacy: Existing Results

$n^*_{\mathrm{priv}}(\epsilon)$:= Sample complexity with ϵ-LDP channels



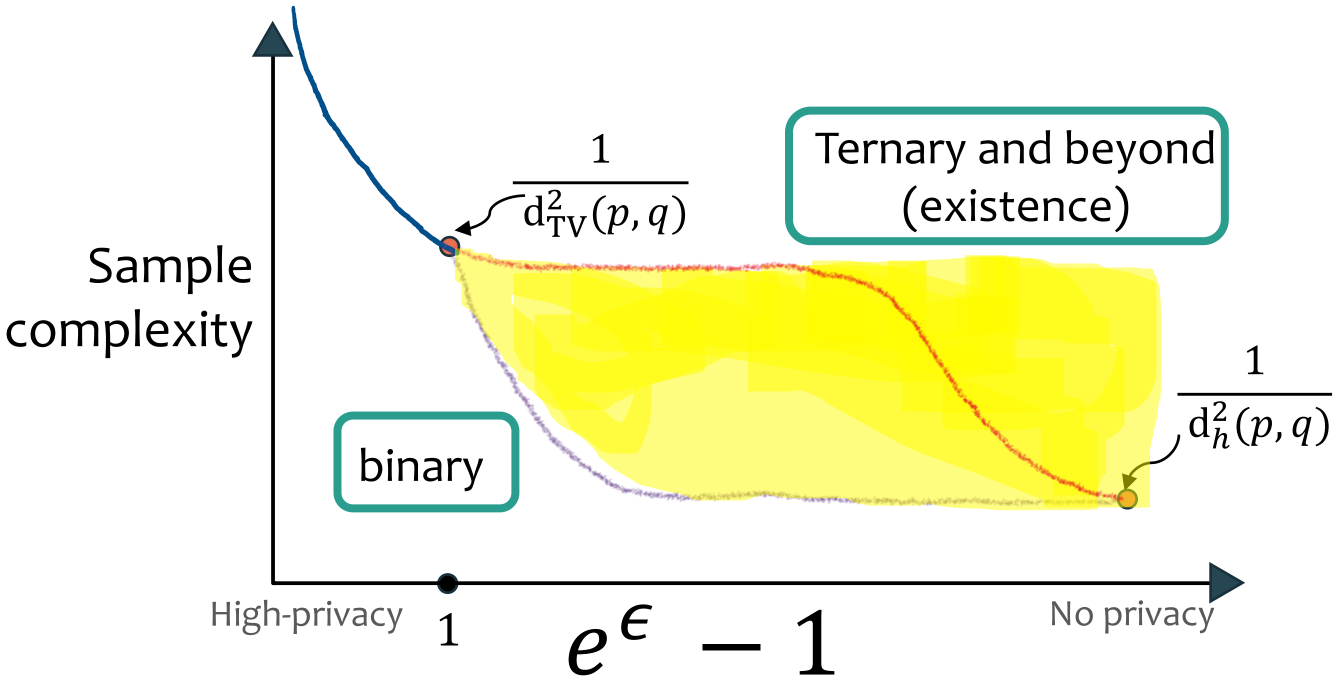
[PAJL23]: Existing lower bound is tight for Bernoulli distributions

What about general distributions?

[DJW13] J. Duchi, M. Wainwright, M.Jordan. Minimax Optimal Procedures for Locally Private Estimation. 2013.
[AZ22] S. Asoodeh, H. Zhang. Contraction of Locally Private Mechanisms. 2022.

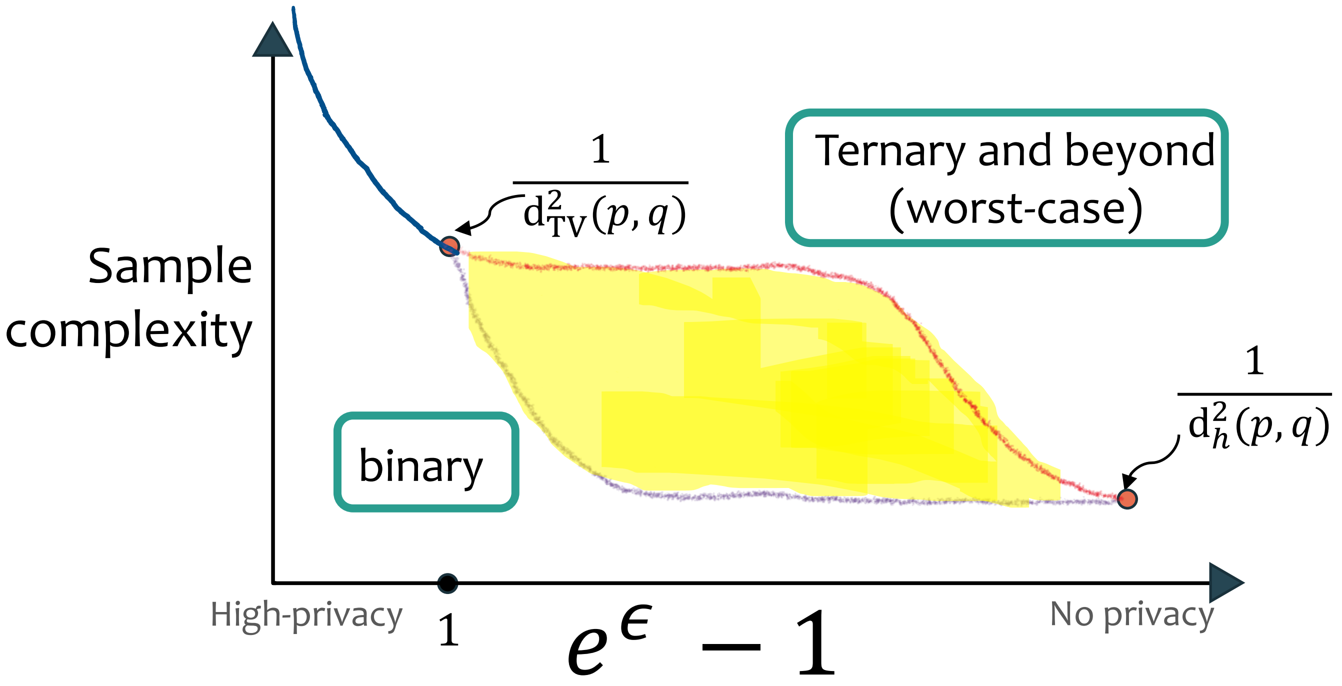
Our Results: Minimax Optimal Sample Complexity

Theorem[PAJL23] There exist ternary distributions p and q with larger sample complexities.



Our Results: Minimax Optimal Sample Complexity

Theorem[PAJL23] There exist ternary distributions p and q with larger sample complexities.

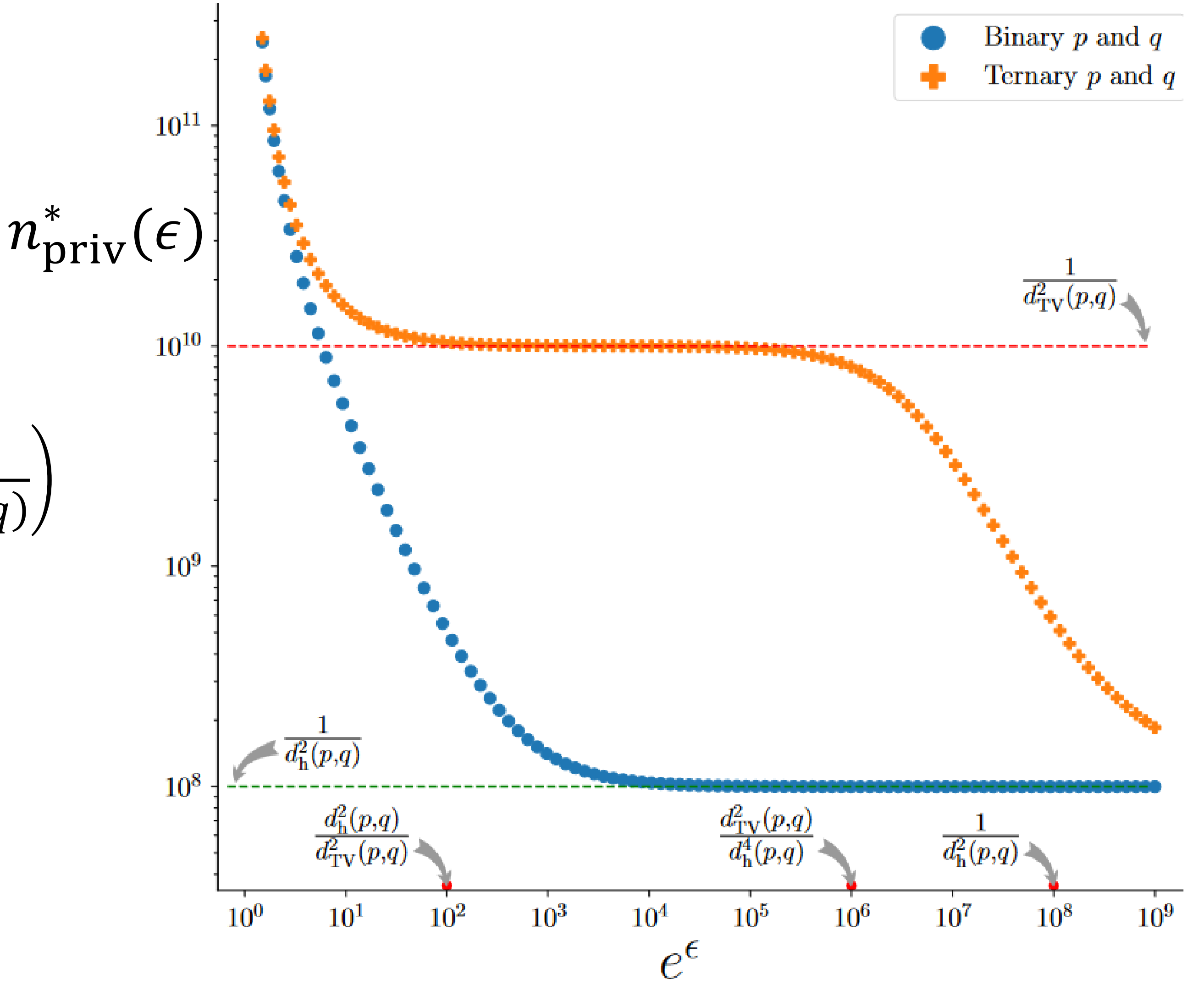


Theorem[PAJL23] There is an efficient algorithm with nearly-matching upper bounds for all distributions.

Exact Expressions and Simulations

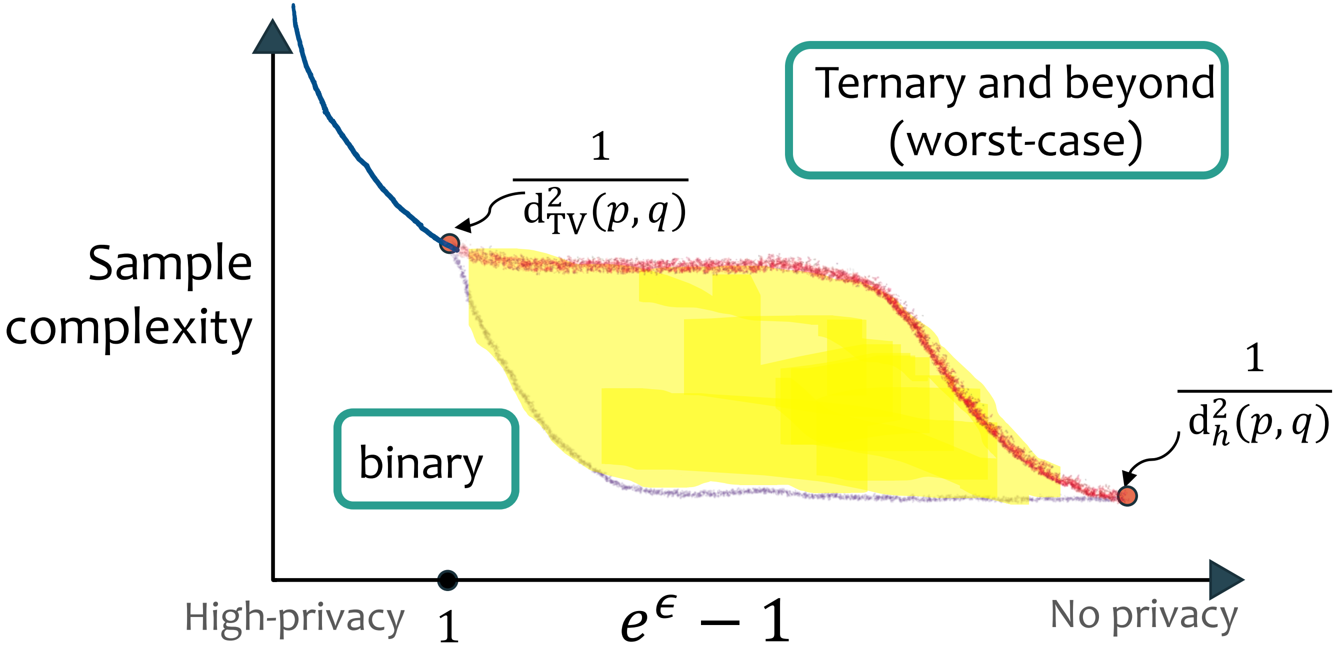
$$n^*_{\text{priv}}(\epsilon) \approx \begin{cases} \dfrac{1}{d^2_{\text{TV}}(p,q)} & \epsilon \in \left(1, \dfrac{d^2_{\text{TV}}(p,q)}{d^4_h(p,q)}\right) \\[2ex] \dfrac{1}{e^{\epsilon} d^4_h(p,q)} & \epsilon \in \left(\dfrac{d^2_{\text{TV}}(p,q)}{d^4_h(p,q)}, \dfrac{1}{d^2_h(p,q)}\right) \\[2ex] \dfrac{1}{d^2_{\text{h}}(p,q)} & \epsilon > \dfrac{1}{d^2_h(p,q)} \end{cases}$$

Minmax



Minimax Optimality and Looking Beyond

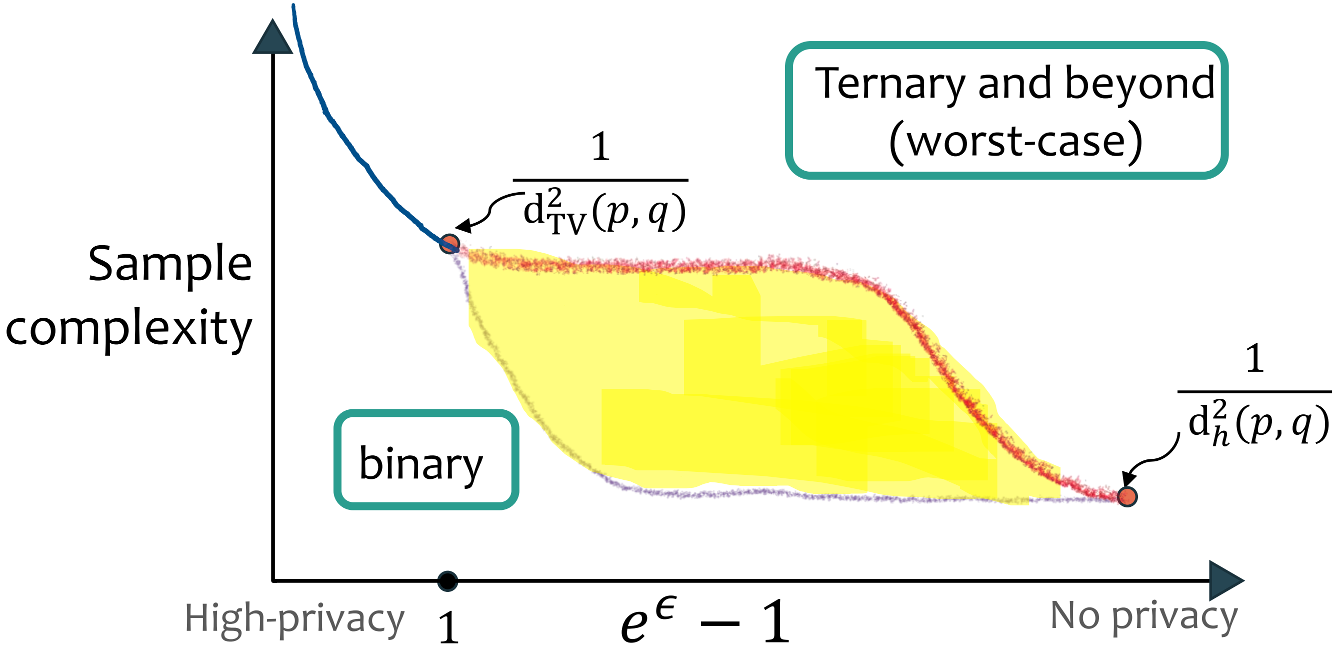
Theorem[PAJL23] Characterization of the minimax-optimal sample complexity over the class of distributions with certain total variation distance and Hellinger divergence.



Minimax Optimality and Looking Beyond

Theorem[PAJL23] Characterization of the minimax-optimal sample complexity over the class of distributions with certain total variation distance and Hellinger divergence.

Best-case: binary

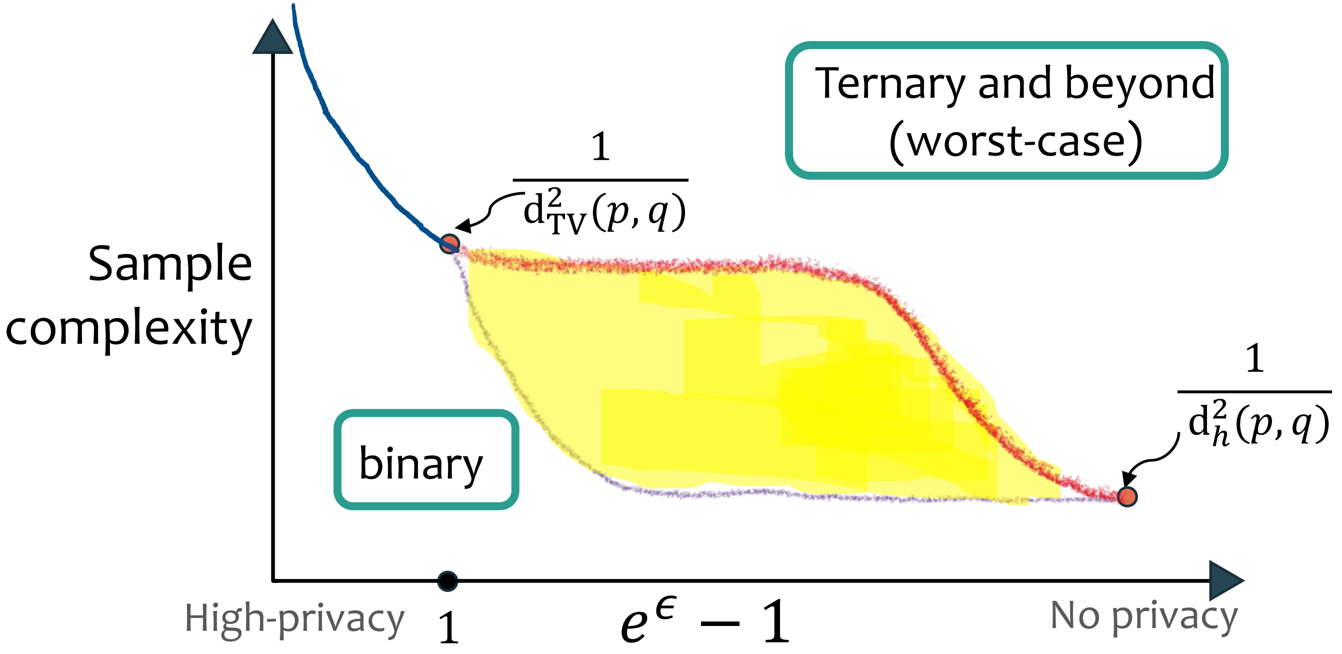


Minimax Optimality and Looking Beyond

Theorem[PAJL23] Characterization of the minimax-optimal sample complexity over the class of distributions with certain total variation distance and Hellinger divergence.

Best-case: binary

Worst-case: distributions from the lower bound



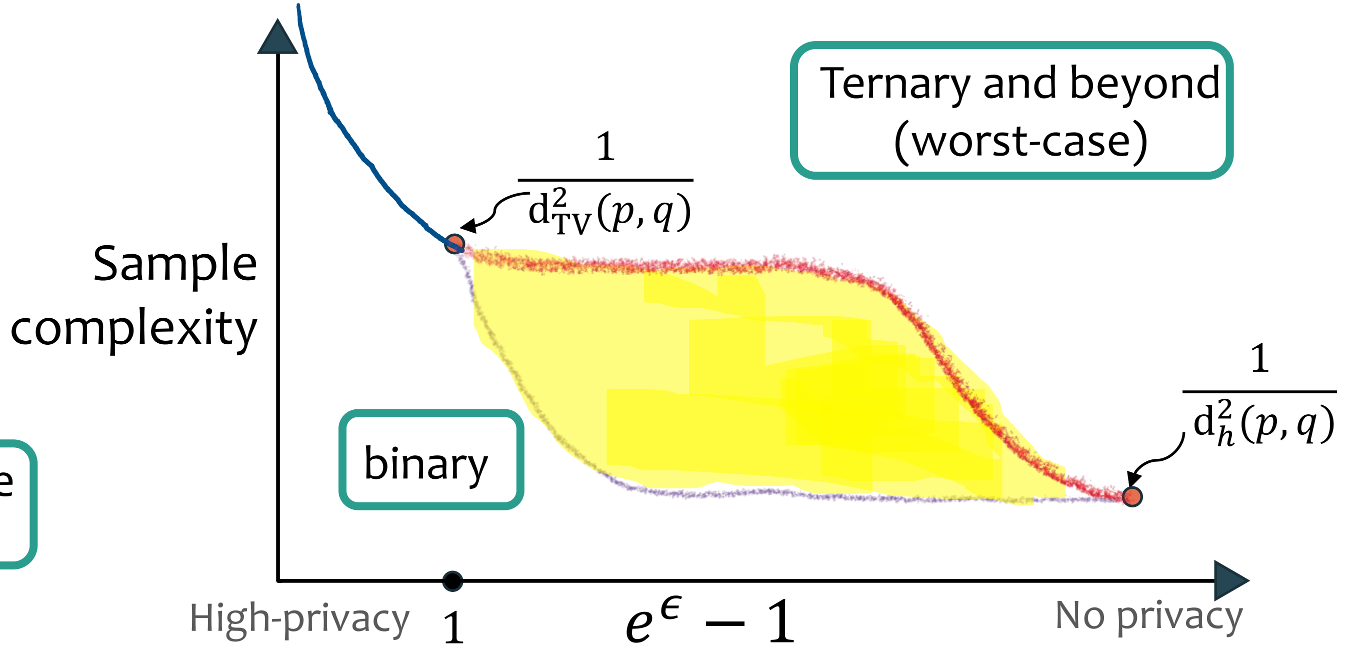
Minimax Optimality and Looking Beyond

Theorem[PAJL23] Characterization of the minimax-optimal sample complexity over the class of distributions with certain total variation distance and Hellinger divergence.

Best-case: binary

Worst-case: distributions from the lower bound

Real-life instances are neither the best-case nor the worst-case



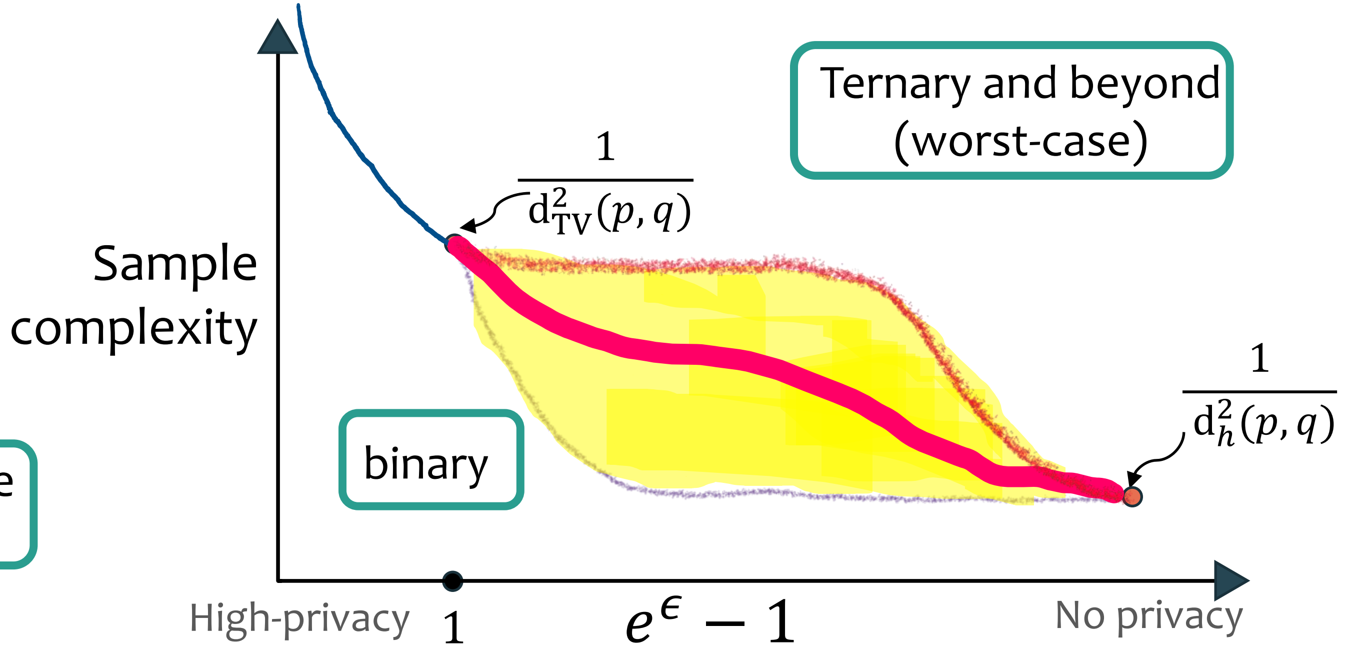
Minimax Optimality and Looking Beyond

Theorem[PAJL23] Characterization of the minimax-optimal sample complexity over the class of distributions with certain total variation distance and Hellinger divergence.

Best-case: binary

Worst-case: distributions from the lower bound

Real-life instances are neither the best-case nor the worst-case



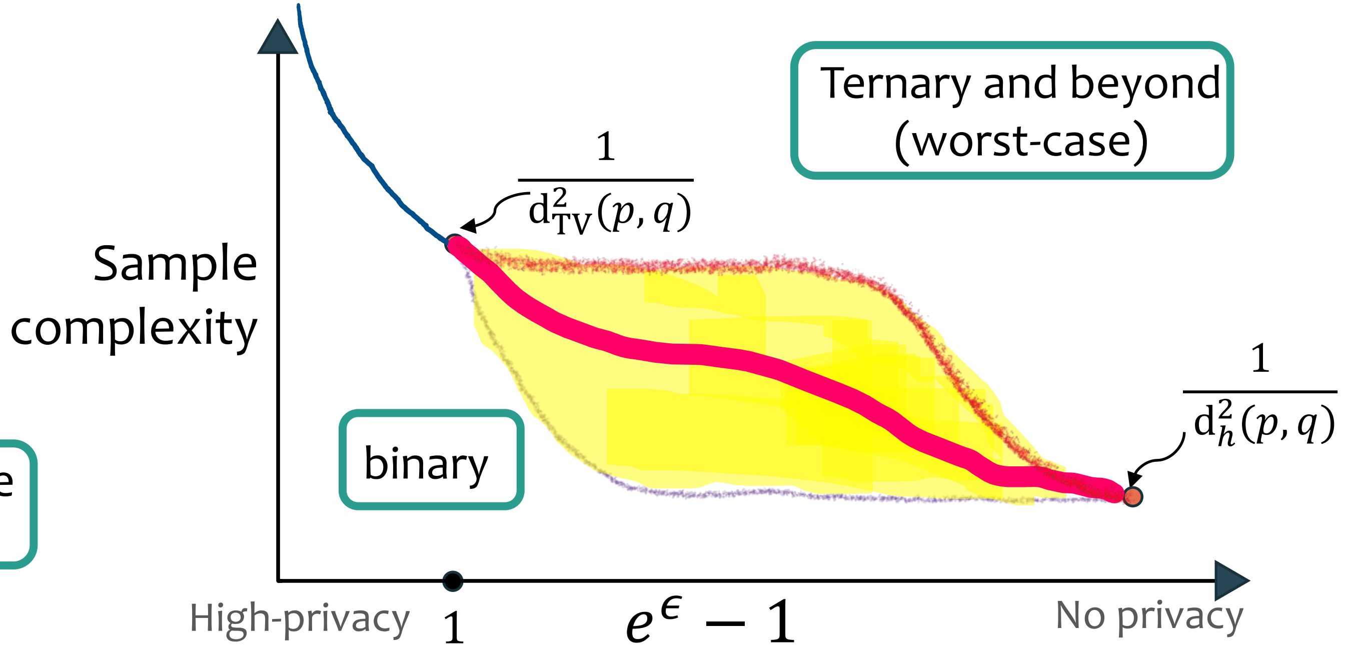
Minimax Optimality and Looking Beyond

Theorem[PAJL23] Characterization of the minimax-optimal sample complexity over the class of distributions with certain total variation distance and Hellinger divergence.

Best-case: binary

Worst-case: distributions from the lower bound

Real-life instances are neither the best-case nor the worst-case



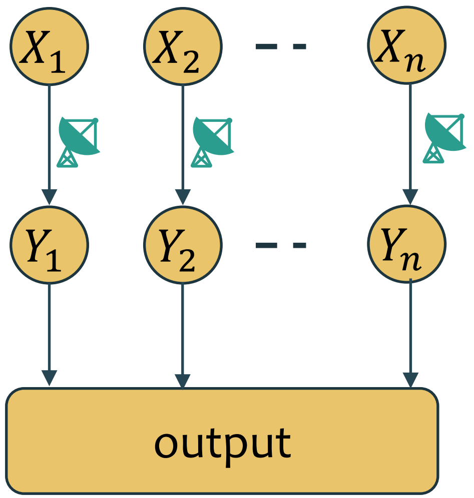
Are there efficient algorithms that adapt to the given instance?

Outline

- Motivation
- Problem Statement
- Our Results
  - Statistical
  - Computational
- Proof Sketch
- Conclusion

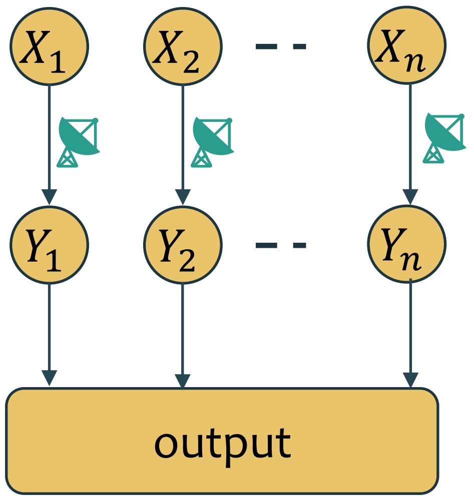
Computational Cost of Privacy

- Recall we need to map the original data $X_i \to Y_i$



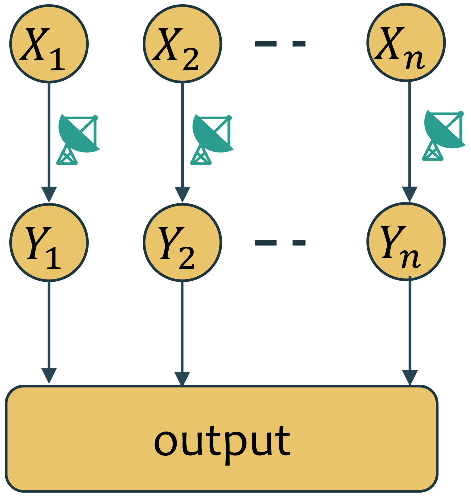
Computational Cost of Privacy

- Recall we need to map the original data $X_i \to Y_i$
- Performance depends on the channel



Computational Cost of Privacy

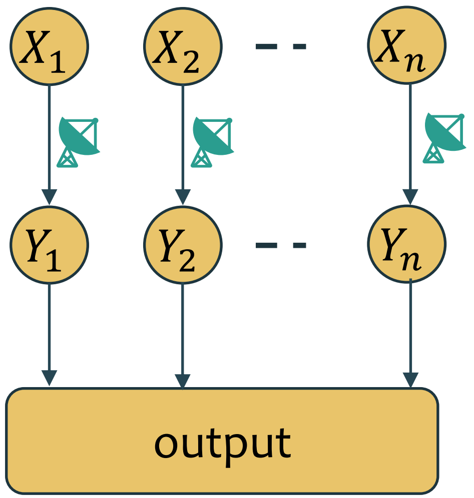
- Recall we need to map the original data $X_i \to Y_i$
- Performance depends on the channel 
  - Once the channel is fixed, perform likelihood ratio test



Computational Cost of Privacy

- Recall we need to map the original data $X_i \to Y_i$
- Performance depends on the channel
  - Once the channel is fixed, perform likelihood ratio test

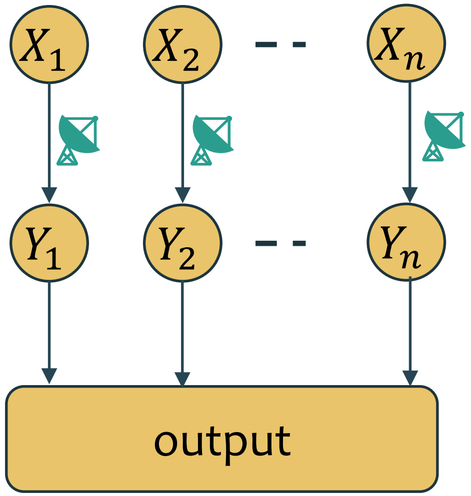
- Prior work on finding the optimal channel



Computational Cost of Privacy

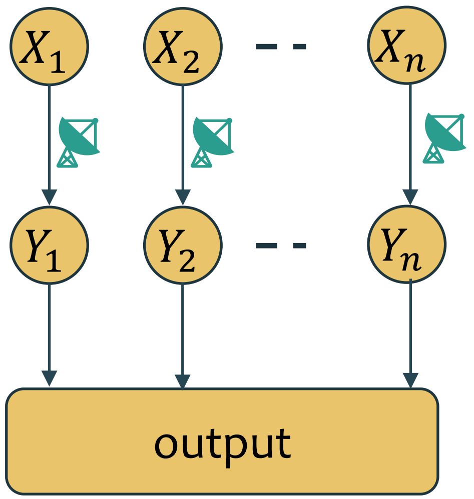
- Recall we need to map the original data $X_i \to Y_i$
- Performance depends on the channel 
  - Once the channel is fixed, perform likelihood ratio test

- Prior work on finding the optimal channel
  - $\epsilon \ll 1$: Well-understood



Computational Cost of Privacy

- Recall we need to map the original data $X_i \to Y_i$
- Performance depends on the channel
  - Once the channel is fixed, perform likelihood ratio test



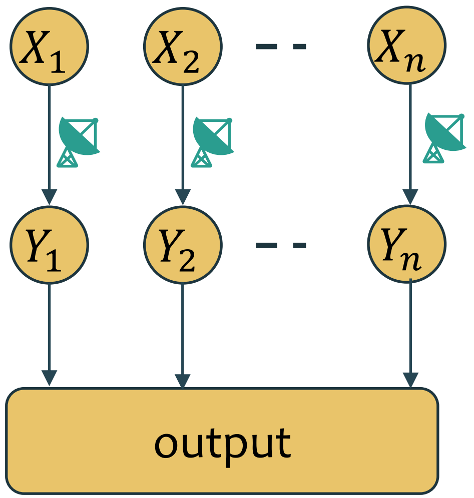
- Prior work on finding the optimal channel
  - $\epsilon \ll 1$: Well-understood
  - $\epsilon \gg 1$: No existing polynomial-time algorithm
    - Naïve algorithm would be 2^{k^2}
    - [KOV14] gave an exponential-time algorithm

[KOV14] P. Kairouz, S. Oh, P. Viswanath. Extremal Mechanisms for Local Differential Privacy. 2014

Computational Cost of Privacy

- Recall we need to map the original data $X_i \to Y_i$
- Performance depends on the channel
  - Once the channel is fixed, perform likelihood ratio test

- Prior work on finding the optimal channel
  - $\epsilon \ll 1$: Well-understood
  - $\epsilon \gg 1$: No existing polynomial-time algorithm
    - Naïve algorithm would be 2^{k^2}
    - [KOV14] gave an exponential-time algorithm



Can we efficiently find the (near)-optimal channel?

[KOV14] P. Kairouz, S. Oh, P. Viswanath. Extremal Mechanisms for Local Differential Privacy. 2014

Our Results: Computational Cost of Privacy

Theorem[**PAJL**23] Given any two distributions p and q on $[k]$ and ϵ,

Our Results: Computational Cost of Privacy

Theorem[**PAJL**23] Given any two distributions p and q on $[k]$ and ϵ, there is a **linear-time algorithm** to find an ϵ-LDP channel

Our Results: Computational Cost of Privacy

Theorem[**PAJL23**] Given any two distributions p and q on $[k]$ and ϵ,

there is a **linear-time algorithm** to find an ϵ-LDP channel

whose sample complexity is **near-optimal for p, q, and ϵ.**

Our Results: Computational Cost of Privacy

Theorem[PAJL23] Given any two distributions p and q on $[k]$ and ϵ, there is a **linear-time algorithm** to find an ϵ-LDP channel whose sample complexity is **near-optimal for $\boldsymbol{p}, \boldsymbol{q}$, and $\boldsymbol{\epsilon}$.**

- The channel uses only an output domain of size 2 (single bit)

Our Results: Computational Cost of Privacy

Theorem[PAJL23] Given any two distributions p and q on $[k]$ and ϵ,

there is a **linear-time algorithm** to find an ϵ-LDP channel

whose sample complexity is **near-optimal for $\boldsymbol{p, q}$, and $\boldsymbol{\epsilon}$.**

- The channel uses only an output domain of size 2 (single bit)
- Extends to other privacy notions: approximate DP, Renyi-DP, zero-concentrated DP

Our Results: Computational Cost of Privacy

Theorem[**PAJL**23] Given any two distributions p and q on $[k]$ and ϵ,
there is a **linear-time algorithm** to find an ϵ-LDP channel
whose sample complexity is **near-optimal for $\boldsymbol{p}, \boldsymbol{q}$, and $\boldsymbol{\epsilon}$.**

- The channel uses only an output domain of size 2 (single bit)
- Extends to other privacy notions: approximate DP, Renyi-DP, zero-concentrated DP
- Can be generalized to have a smooth tradeoff:

Our Results: Computational Cost of Privacy

Theorem[PAJL23] Given any two distributions p and q on $[k]$ and ϵ,

there is a **linear-time algorithm** to find an ϵ-LDP channel

whose sample complexity is **near-optimal for $\boldsymbol{p}, \boldsymbol{q}$, and $\boldsymbol{\epsilon}$.**

- The channel uses only an output domain of size 2 (single bit)
- Extends to other privacy notions: approximate DP, Renyi-DP, zero-concentrated DP
- Can be generalized to have a smooth tradeoff:
  - A $\text{poly}_\ell\left(k^{\ell^2}\right)$-time algorithm to an ℓ-output channel with sample complexity

$$\boxed{n^*_{\text{priv}}(\epsilon) \cdot \left(1 + \frac{\log n^*_{\text{priv}}(\epsilon)}{\ell}\right)}$$

Our Results: Computational Cost of Privacy, Generalized

- More broadly, consider the optimization problem

$$\max_{\in \mathcal{P}(\epsilon,\ell)} g(p, q)$$

$\mathcal{P}(\epsilon,\ell)$: All ϵ-LDP channels of output size ℓ

g: a (quasi)-convex objective

- Examples: f-divergences, Renyi Entropy, Wasserstein Norm
  - Maximal separation between p and q after privatization

Our Results: Computational Cost of Privacy, Generalized

- More broadly, consider the optimization problem

$$\max_{\in \mathcal{P}(\epsilon,\ell)} g(\, p, \; q)$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

g: a (quasi)-convex objective

- Examples: f-divergences, Renyi Entropy, Wasserstein Norm
  - Maximal separation between p and q after privatization

Recall: maximizing a convex objective is usually hard!

Our Results: Computational Cost of Privacy, Generalized

- More broadly, consider the optimization problem

$$\max_{\in \mathcal{P}(\epsilon,\ell)} g(\ p,\ \ q)$$

$\mathcal{P}(\epsilon,\ell)$: All ϵ-LDP channels of output size ℓ

g: a (quasi)-convex objective

- Examples: f-divergences, Renyi Entropy, Wasserstein Norm
  - Maximal separation between p and q after privatization

Recall: maximizing a convex objective is usually hard!

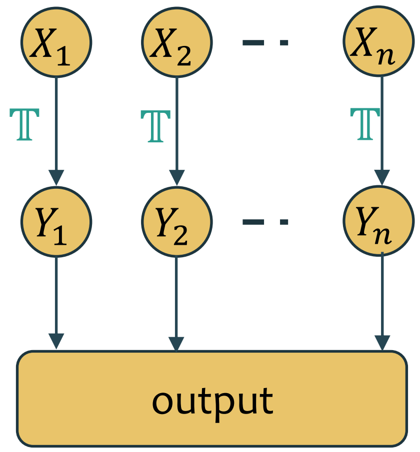
Theorem[PAJL23] There is a $\text{poly}_\ell\left(k^{\ell^2}\right)$-time algorithm to find the optimum.

Outline

- Motivation
- Problem Statement
- Our Results
- Proof Sketch
  - Statistical
  - Computational
- Conclusion

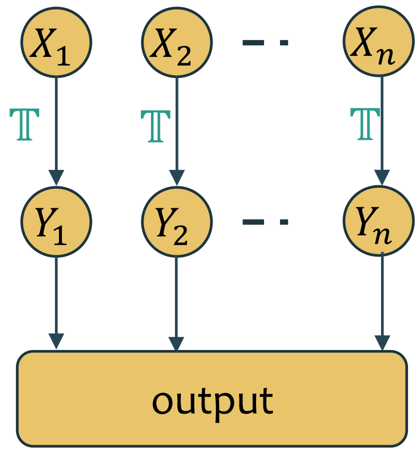
Proof Sketch: How to choose optimal $\mathbb{T}$?

- Suppose that every channel is fixed to be $\mathbb{T}$



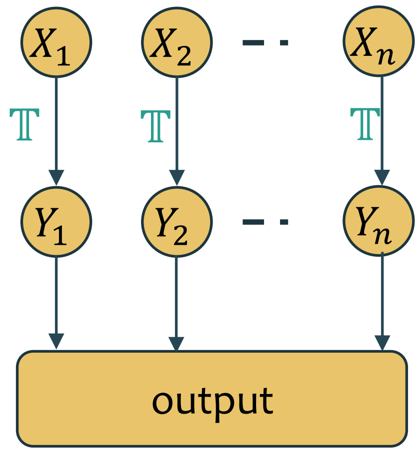
Proof Sketch: How to choose optimal $\mathbb{T}$?

- Suppose that every channel is fixed to be $\mathbb{T}$
- Then, each Y_i is either distributed as $\mathbb{T}p$ or as $\mathbb{T}q$

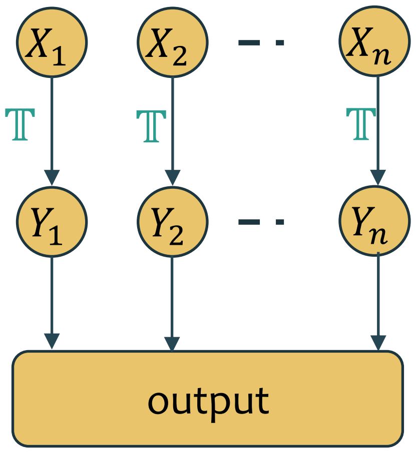


Proof Sketch: How to choose optimal $\mathbb{T}$?

- Suppose that every channel is fixed to be $\mathbb{T}$
- Then, each Y_i is either distributed as $\mathbb{T}p$ or as $\mathbb{T}q$
- We are effectively testing between $\mathbb{T}p$ and $\mathbb{T}q$

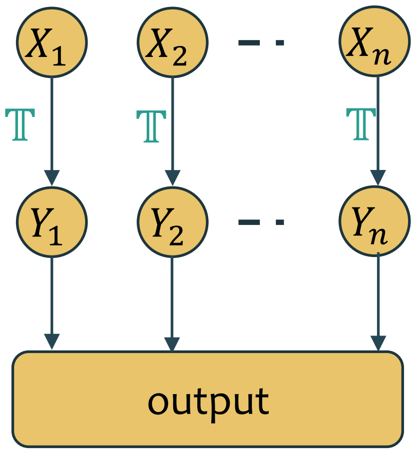


Proof Sketch: How to choose optimal $\mathbb{T}$?



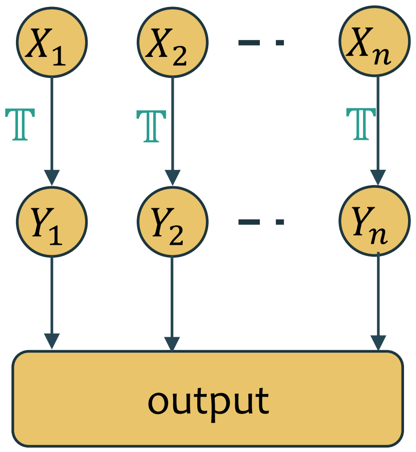
- Suppose that every channel is fixed to be $\mathbb{T}$
- Then, each Y_i is either distributed as $\mathbb{T}p$ or as $\mathbb{T}q$
- We are effectively testing between $\mathbb{T}p$ and $\mathbb{T}q$
- Thus, the sample complexity is $\frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}$

Proof Sketch: How to choose optimal $\mathbb{T}$?



- Suppose that every channel is fixed to be $\mathbb{T}$
- Then, each Y_i is either distributed as $\mathbb{T}p$ or as $\mathbb{T}q$
- We are effectively testing between $\mathbb{T}p$ and $\mathbb{T}q$
- Thus, the sample complexity is $\frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}$
- Leads to optimal choice of $\mathbb{T}$: $\boxed{\min_{\mathbb{T} \in \text{constraints}} \frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}}$

Proof Sketch: How to choose optimal $\mathbb{T}$?



- Suppose that every channel is fixed to be $\mathbb{T}$
- Then, each Y_i is either distributed as $\mathbb{T}p$ or as $\mathbb{T}q$
- We are effectively testing between $\mathbb{T}p$ and $\mathbb{T}q$
- Thus, the sample complexity is $\frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}$
- Leads to optimal choice of $\mathbb{T}$: $\boxed{\min_{\mathbb{T} \in \text{constraints}} \frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}}$

Statistical cost: Minimum value

Computational cost: time to find an approximate minimizer

Outline

- Motivation
- Problem Statement
- Our Results
- Proof Sketch
  - Statistical
  - Computational
- Conclusion

$$\min_{\mathbb{T} \in \text{constraints}} \frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}$$

Proof Sketch: Statistical Cost of Privacy

- Need to understand $\max_{\mathbb{T}:\epsilon-\text{LDP}} d_h^2(\mathbb{T}p, \mathbb{T}q)$

$$\min_{\mathbb{T} \in \text{constraints}} \frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}$$

Proof Sketch: Statistical Cost of Privacy

- Need to understand $\max_{\mathbb{T}:\epsilon-\text{LDP}} d_h^2(\mathbb{T}p, \mathbb{T}q)$
  - Data processing inequality implies $d_h^2(\mathbb{T}p, \mathbb{T}q)$ is smaller than $d_h^2(p, q)$

$$\min_{\mathbb{T} \in \text{constraints}} \frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}$$

Proof Sketch: Statistical Cost of Privacy

- Need to understand $\max_{\mathbb{T}:\epsilon-\mathrm{LDP}} d_h^2(\mathbb{T}p, \mathbb{T}q)$
  - Data processing inequality implies $d_h^2(\mathbb{T}p, \mathbb{T}q)$ is smaller than $d_h^2(p, q)$
- Privacy requires adding noise, which results in much smaller $d_h^2(\mathbb{T}p, \mathbb{T}q)$
  - Leads to "Strong data processing inequality"

$$\min_{\mathbb{T} \in \text{constraints}} \frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}$$

Proof Sketch: Statistical Cost of Privacy

- Need to understand $\max_{\mathbb{T}:\epsilon-\text{LDP}} d_h^2(\mathbb{T}p, \mathbb{T}q)$
  - Data processing inequality implies $d_h^2(\mathbb{T}p, \mathbb{T}q)$ is smaller than $d_h^2(p, q)$
- Privacy requires adding noise, which results in much smaller $d_h^2(\mathbb{T}p, \mathbb{T}q)$
  - Leads to "Strong data processing inequality"
- Analyzing the maximum requires knowing the optimal $\mathbb{T}$
  - Non-trivial in general but the binary setting is much easier (randomized-response)

$$\min_{\mathbb{T} \in \text{constraints}} \frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}$$

Proof Sketch: Statistical Cost of Privacy

- Need to understand $\max_{\mathbb{T}:\epsilon-\text{LDP}} d_h^2(\mathbb{T}p, \mathbb{T}q)$
  - Data processing inequality implies $d_h^2(\mathbb{T}p, \mathbb{T}q)$ is smaller than $d_h^2(p, q)$
- Privacy requires adding noise, which results in much smaller $d_h^2(\mathbb{T}p, \mathbb{T}q)$
  - Leads to "Strong data processing inequality"
- Analyzing the maximum requires knowing the optimal $\mathbb{T}$
  - Non-trivial in general but the binary setting is much easier (randomized-response)

Proposition [**PAJL23**] If p and q are Bernoulli distributions and $\epsilon \gg 1$, then

$$\min_{\mathbb{T} \in \text{constraints}} \frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}$$

Proof Sketch: Statistical Cost of Privacy

- Need to understand $\max_{\mathbb{T}:\epsilon-\mathrm{LDP}} d_h^2(\mathbb{T}p, \mathbb{T}q)$
  - Data processing inequality implies $d_h^2(\mathbb{T}p, \mathbb{T}q)$ is smaller than $d_h^2(p, q)$
- Privacy requires adding noise, which results in much smaller $d_h^2(\mathbb{T}p, \mathbb{T}q)$
  - Leads to "Strong data processing inequality"
- Analyzing the maximum requires knowing the optimal $\mathbb{T}$
  - Non-trivial in general but the binary setting is much easier (randomized-response)

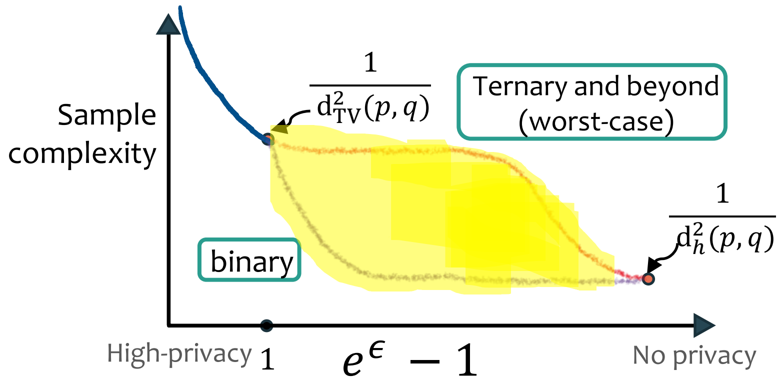
Proposition [PAJL23] If p and q are Bernoulli distributions and $\epsilon \gg 1$, then

$$\max_{\mathbb{T}:\epsilon-\mathrm{LDP}} d_h^2(\mathbb{T}p, \mathbb{T}q) \asymp \min(e^{\epsilon} d_{\mathrm{TV}}^2(p,q), d_h^2(p,q))$$

- The decrease (or the contraction) depends also on the total variation distance

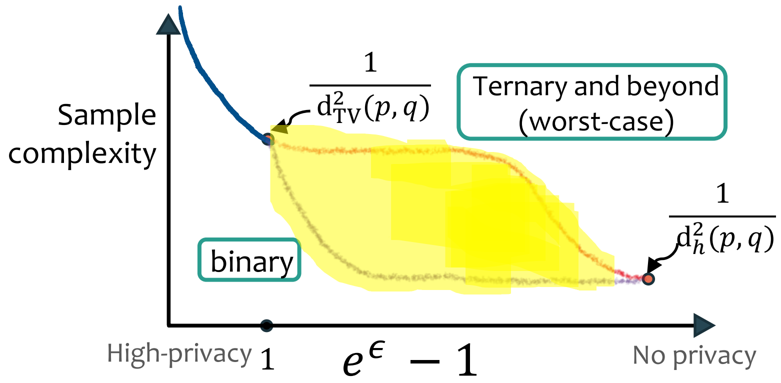
Proof Sketch: Why Is Ternary Much Harder?

- Suppose, we are interested in a binary private channel $\mathbb{T}$

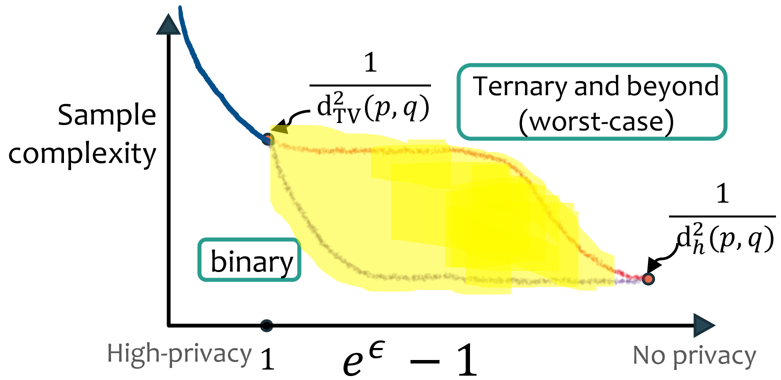


Proof Sketch: Why Is Ternary Much Harder?

- Suppose, we are interested in a binary private channel $\mathbb{T}$
- Can be shown that optimal $\mathbb{T}$ is of the form
  - First, a binary deterministic channel $\mathbb{T}'$
  - Then, the randomized-response to ensure privacy

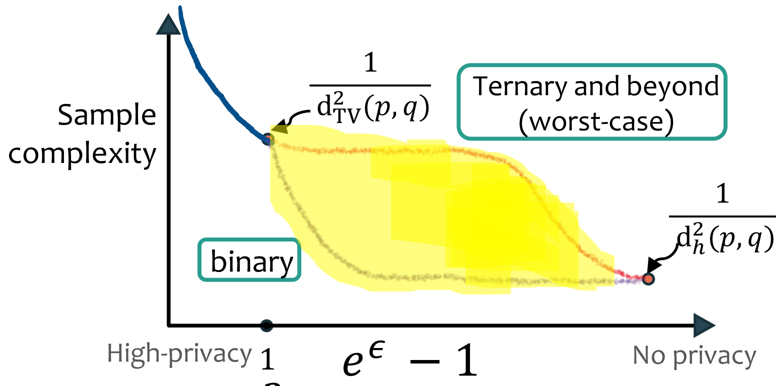


Proof Sketch: Why Is Ternary Much Harder?



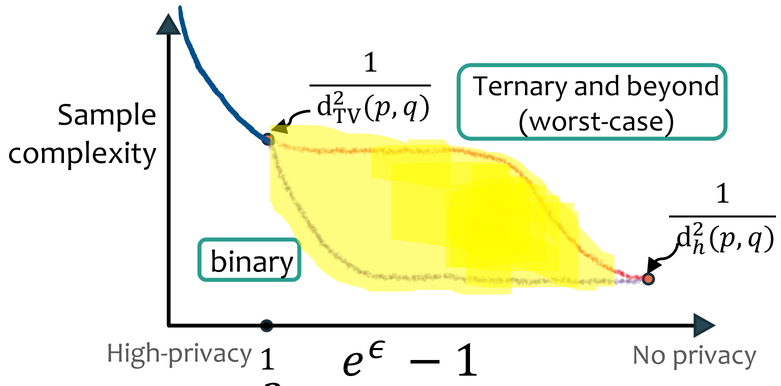
- Suppose, we are interested in a binary private channel $\mathbb{T}$
- Can be shown that optimal $\mathbb{T}$ is of the form
  - First, a binary deterministic channel $\mathbb{T}'$
  - Then, the randomized-response to ensure privacy
- Since the performance of randomized-response depends both on both d_{TV} and d_h^2
  - $\mathbb{T}'$ must try to preserve both d_{TV} and d_h^2

Proof Sketch: Why Is Ternary Much Harder?



- Suppose, we are interested in a binary private channel $\mathbb{T}$
- Can be shown that optimal $\mathbb{T}$ is of the form
    - First, a binary deterministic channel $\mathbb{T}'$
    - Then, the randomized-response to ensure privacy
- Since the performance of randomized-response depends both on both d_{TV} and d_h^2
    - $\mathbb{T}'$ must try to preserve both d_{TV} and d_h^2
    - Unfortunately, both can not be preserved always (see example)

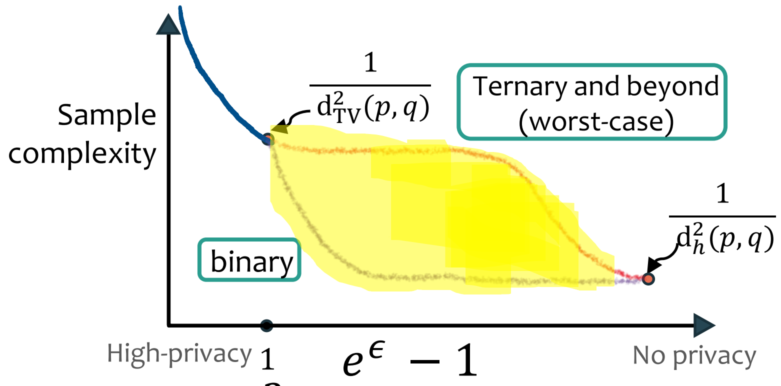
Proof Sketch: Why Is Ternary Much Harder?



- Suppose, we are interested in a binary private channel $\mathbb{T}$
- Can be shown that optimal $\mathbb{T}$ is of the form
  - First, a binary deterministic channel $\mathbb{T}'$
  - Then, the randomized-response to ensure privacy
- Since the performance of randomized-response depends both on both d_{TV} and d_h^2
  - $\mathbb{T}'$ must try to preserve both d_{TV} and d_h^2
  - Unfortunately, both can not be preserved always (see example)

$$p = \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix} \qquad q = \begin{pmatrix} 0.5 - \alpha - \gamma \\ 0.5 - \alpha + \gamma \\ 2\alpha \end{pmatrix}$$

Proof Sketch: Why Is Ternary Much Harder?



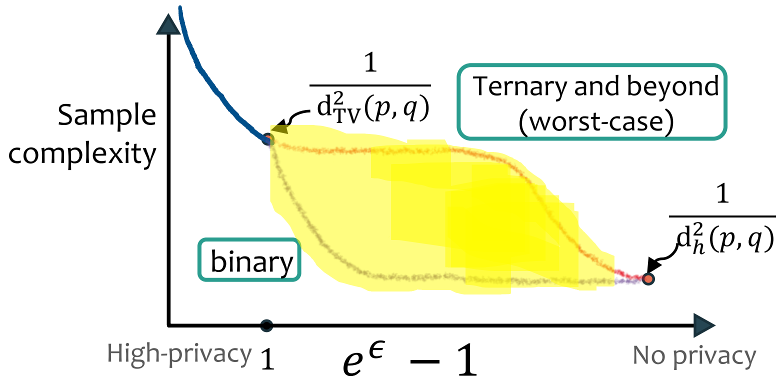
- Suppose, we are interested in a binary private channel $\mathbb{T}$
- Can be shown that optimal $\mathbb{T}$ is of the form
  - First, a binary deterministic channel $\mathbb{T}'$
  - Then, the randomized-response to ensure privacy
- Since the performance of randomized-response depends both on both d_{TV} and d_h^2
  - $\mathbb{T}'$ must try to preserve both d_{TV} and d_h^2
  - Unfortunately, both can not be preserved always (see example)

$$p = \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix} \qquad q = \begin{pmatrix} 0.5 - \alpha - \gamma \\ 0.5 - \alpha + \gamma \\ 2\alpha \end{pmatrix}$$

Dominant contribution to d_{TV}

Dominant contribution to d_h^2

Proof Sketch: Why Is Ternary Much Harder?



- Suppose, we are interested in a binary private channel $\mathbb{T}$
- Can be shown that optimal $\mathbb{T}$ is of the form
  - First, a binary deterministic channel $\mathbb{T}'$
  - Then, the randomized-response to ensure privacy
- Since the performance of randomized-response depends both on both d_{TV} and d_h^2
  - $\mathbb{T}'$ must try to preserve both d_{TV} and d_h^2
  - Unfortunately, both can not be preserved always (see example)

$$p = \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix} \qquad q = \begin{pmatrix} 0.5 - \alpha - \gamma \\ 0.5 - \alpha + \gamma \\ 2\alpha \end{pmatrix}$$

Dominant contribution to d_{TV}

Dominant contribution to d_h^2

  - If $\mathbb{T}'$ preserves Hellinger divergence, then the total variation decreases, and vice versa

Outline

- Motivation
- Problem Statement
- Our Results
- Proof Sketch
  - Statistical
  - Computational
- Conclusion

$$\min_{\mathbb{T} \in \text{constraints}} \frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}$$

Extreme points lead to optimal performance

- Recall the original objective

$$\max_{\mathbb{T} \in \mathcal{P}(\epsilon,\ell)} g(\mathbb{T}p, \mathbb{T}q)$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

g: a (quasi)-convex objective

$$\min_{\mathbb{T} \in \text{constraints}} \frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}$$

Extreme points lead to optimal performance

- Recall the original objective

$$\max_{\mathbb{T} \in \mathcal{P}(\epsilon,\ell)} \quad g(\mathbb{T}p, \mathbb{T}q)$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

g: a (quasi)-convex objective

- Let the joint range be $\mathcal{A} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$

$$\min_{\mathbb{T} \in \text{constraints}} \frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}$$

Extreme points lead to optimal performance

- Recall the original objective

$$\max_{\mathbb{T} \in \mathcal{P}(\epsilon,\ell)} g(\mathbb{T}p, \mathbb{T}q)$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

g: a (quasi)-convex objective

- Let the joint range be $\mathcal{A} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$

- By convexity of g and $\mathcal{A}$, the maximum value is attained at $\mathbb{T}$ only if $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}$

$$\min_{\mathbb{T} \in \text{constraints}} \frac{1}{d_h^2(\mathbb{T}p, \mathbb{T}q)}$$

Extreme points lead to optimal performance

- Recall the original objective

$$\max_{\mathbb{T} \in \mathcal{P}(\epsilon, \ell)} g(\mathbb{T}p, \mathbb{T}q)$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

g: a (quasi)-convex objective

- Let the joint range be $\mathcal{A} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$

- By convexity of g and $\mathcal{A}$, the maximum value is attained at $\mathbb{T}$ only if $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}$

What type of channels $\mathbb{T}$ lead to the extreme points of $\mathcal{A}$?

Extreme Points of the Joint Range: First Attempt

$$\mathcal{A} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

Theorem[**PAJL23**] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}$, then $\mathbb{T}$ can be decomposed as

Extreme Points of the Joint Range: First Attempt

$$\mathcal{A} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

Theorem[**PAJL**23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}$, then $\mathbb{T}$ can be decomposed as

- First, a deterministic channel from $[k]$ to $[2\ell^2]$

Extreme Points of the Joint Range: First Attempt

$$\mathcal{A} \coloneqq \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

Theorem[**PAJL**23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}$, then $\mathbb{T}$ can be decomposed as

- First, a deterministic channel from $[k]$ to $[2\ell^2]$
- Then, a (randomized) ϵ-LDP channel from $[2\ell^2]$ to $[\ell]$

Extreme Points of the Joint Range: First Attempt

$$\mathcal{A} \coloneqq \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

Theorem[**PAJL**23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}$, then $\mathbb{T}$ can be decomposed as

- First, a deterministic channel from $[k]$ to $[2\ell^2]$
- Then, a (randomized) ϵ-LDP channel from $[2\ell^2]$ to $[\ell]$

- The Good: Privacy step is independent of k

Extreme Points of the Joint Range: First Attempt

$$\mathcal{A} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

Theorem[**PAJL**23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}$, then $\mathbb{T}$ can be decomposed as

- First, a deterministic channel from $[k]$ to $[2\ell^2]$
- Then, a (randomized) ϵ-LDP channel from $[2\ell^2]$ to $[\ell]$

- The Good: Privacy step is independent of k
- The bad: The number of deterministic channels is ℓ^k

Extreme Points of the Joint Range: First Attempt

$$\mathcal{A} \coloneqq \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

Theorem[**PAJL**23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}$, then $\mathbb{T}$ can be decomposed as

- First, a deterministic channel from $[k]$ to $[2\ell^2]$
- Then, a (randomized) ϵ-LDP channel from $[2\ell^2]$ to $[\ell]$

- The Good: Privacy step is independent of k
- The bad: The number of deterministic channels is ℓ^k

Can we further reduce the search space in the first step?

Extreme Points of the Joint Range: Final

$$\mathcal{A} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

Theorem[**PAJL**23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}$, then $\mathbb{T}$ can be decomposed as

- First, a **threshold** deterministic channel from $[k]$ to $[2\ell^2]$
- Then, a (randomized) ϵ-LDP channel from $[2\ell^2]$ to $[\ell]$

Extreme Points of the Joint Range: Final

$$\mathcal{A} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

Theorem[**PAJL23**] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}$, then $\mathbb{T}$ can be decomposed as

- First, a **threshold** deterministic channel from $[k]$ to $[2\ell^2]$
- Then, a (randomized) ϵ-LDP channel from $[2\ell^2]$ to $[\ell]$

- The number of threshold channels is only polynomial, $k^{\mathrm{poly}(\ell)}$

Extreme Points of the Joint Range: Final

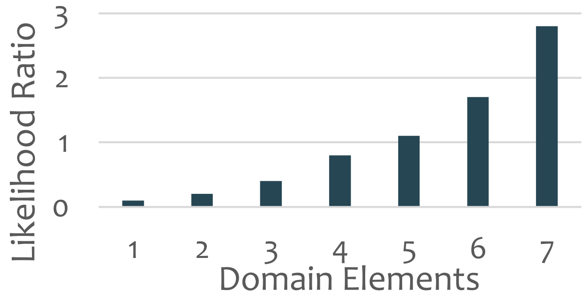
$$\mathcal{A} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

Theorem[**PAJL**23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}$, then $\mathbb{T}$ can be decomposed as

- First, a **threshold** deterministic channel from $[k]$ to $[2\ell^2]$
- Then, a (randomized) ϵ-LDP channel from $[2\ell^2]$ to $[\ell]$

- The number of threshold channels is only polynomial, $k^{\mathrm{poly}(\ell)}$



Extreme Points of the Joint Range: Final

$$\mathcal{A} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$$

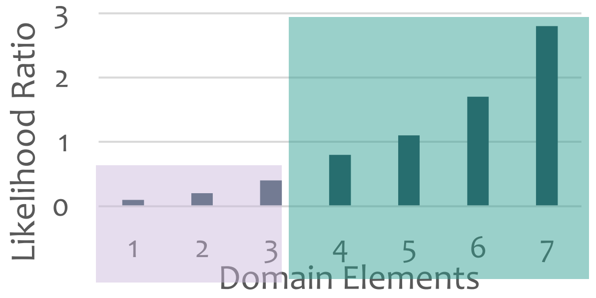
$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

Theorem[**PAJL23**] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}$, then $\mathbb{T}$ can be decomposed as

- First, a **threshold** deterministic channel from $[k]$ to $[2\ell^2]$
- Then, a (randomized) ϵ-LDP channel from $[2\ell^2]$ to $[\ell]$

- The number of threshold channels is only polynomial, $k^{\text{poly}(\ell)}$

Threshold Channel: A deterministic channel $\mathbb{T}$ is a threshold channel for p and q if $\mathbb{T}$ partitions the input domain by thresholding the likelihood ratios of p and q.



Extreme Points of the Joint Range: Final

$$\mathcal{A} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

Theorem[**PAJL23**] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}$, then $\mathbb{T}$ can be decomposed as

- First, a **threshold** deterministic channel from $[k]$ to $[2\ell^2]$
- Then, a (randomized) ϵ-LDP channel from $[2\ell^2]$ to $[\ell]$

- The number of threshold steps is only polynomial, $k^{\mathrm{poly}(\ell)}$

Extreme Points of the Joint Range: Final

$$\mathcal{A} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}(\epsilon, \ell)\}$$

$\mathcal{P}(\epsilon, \ell)$: All ϵ-LDP channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}$, then $\mathbb{T}$ can be decomposed as

- First, a **threshold** deterministic channel from $[k]$ to $[2\ell^2]$
- Then, a (randomized) ϵ-LDP channel from $[2\ell^2]$ to $[\ell]$

- The number of threshold steps is only polynomial, $k^{\mathrm{poly}(\ell)}$

Corollary [PAJL23]: $\mathrm{poly}_\ell\left(k^{\mathrm{poly}(\ell)}\right)$-time algorithms to maximize convex functions over $\mathcal{A}$.

Outline

- Motivation
- Problem Statement
- Our Results
- Proof Sketch
  - Statistical
  - Computational
  - Threshold Channels
- Conclusion

Structural Result: Optimality of Thresholds under Quantization

- For simplicity, let’s focus only on communication constraints

Structural Result: Optimality of Thresholds under Quantization

- For simplicity, let's focus only on communication constraints

$$\mathcal{A}_{\text{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\text{comm}}(\ell)\}$$

$\mathcal{P}_{\text{comm}}(\ell)$: All channels of output size ℓ

Structural Result: Optimality of Thresholds under Quantization

- For simplicity, let's focus only on communication constraints

$$\mathcal{A}_{\mathrm{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\mathrm{comm}}(\ell)\}$$

$\mathcal{P}_{\mathrm{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\mathrm{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Structural Result: Optimality of Thresholds under Quantization

- For simplicity, let's focus only on communication constraints

$$\mathcal{A}_{\mathrm{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\mathrm{comm}}(\ell)\}$$

$\mathcal{P}_{\mathrm{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\mathrm{comm}}$, then $\mathbb{T}$ must be a threshold channel.

- # of extreme points of $\mathcal{A}_{\mathrm{comm}}$, k^{ℓ}, is much smaller than that of $\mathcal{P}_{\mathrm{comm}}$, ℓ^{k}.

Structural Result: Optimality of Thresholds under Quantization

- For simplicity, let's focus only on communication constraints

$$\mathcal{A}_{\mathrm{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\mathrm{comm}}(\ell)\}$$

$\mathcal{P}_{\mathrm{comm}}(\ell)$: All channels of output size ℓ

Theorem[**PAJL23**] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\mathrm{comm}}$, then $\mathbb{T}$ must be a threshold channel.

- # of extreme points of $\mathcal{A}_{\mathrm{comm}}$, k^{ℓ}, is much smaller than that of $\mathcal{P}_{\mathrm{comm}}$, ℓ^{k}.

Corollary [**PAJL23**]: $\mathrm{poly}(k^{\ell})$-time algorithms to maximize convex functions over $\mathcal{A}_{\mathrm{comm}}$.

Proof Sketch: Optimality of Threshold Channels under Quantization

$$\mathcal{A}_{\mathrm{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\mathrm{comm}}(\ell)\}$$

$\mathcal{P}_{\mathrm{comm}}(\ell)$: All channels of output size ℓ

Proof Sketch: Optimality of Threshold Channels under Quantization

$$\mathcal{A}_{\mathrm{comm}} \coloneqq \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\mathrm{comm}}(\ell)\}$$

$\mathcal{P}_{\mathrm{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\mathrm{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof Sketch: Optimality of Threshold Channels under Quantization

$$\mathcal{A}_{\mathrm{comm}} \coloneqq \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\mathrm{comm}}(\ell)\}$$

$\mathcal{P}_{\mathrm{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\mathrm{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.

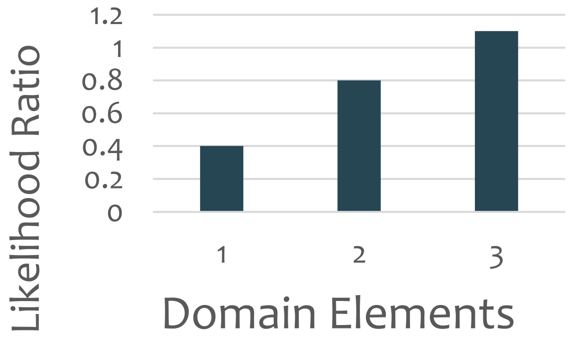
Proof Sketch: Optimality of Threshold Channels under Quantization

$$\mathcal{A}_{\mathrm{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\mathrm{comm}}(\ell)\}$$

$\mathcal{P}_{\mathrm{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\mathrm{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.



$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

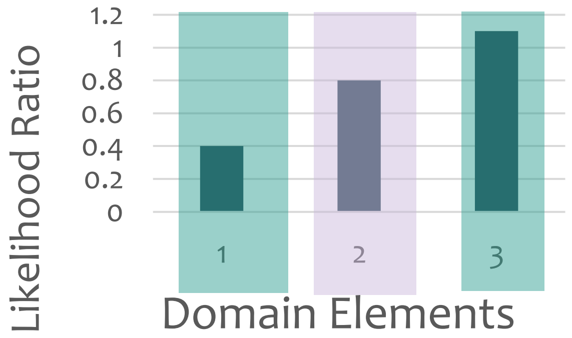
Proof Sketch: Optimality of Threshold Channels under Quantization

$$\mathcal{A}_{\mathrm{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\mathrm{comm}}(\ell)\}$$

$\mathcal{P}_{\mathrm{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\mathrm{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.



$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

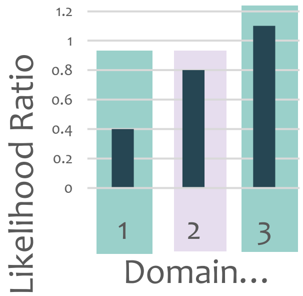
Proof Sketch: Optimality of Threshold Channels for Quantization

$$\mathcal{A}_{\mathrm{comm}} \coloneqq \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\mathrm{comm}}(\ell)\}$$

$\mathcal{P}_{\mathrm{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\mathrm{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.



$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

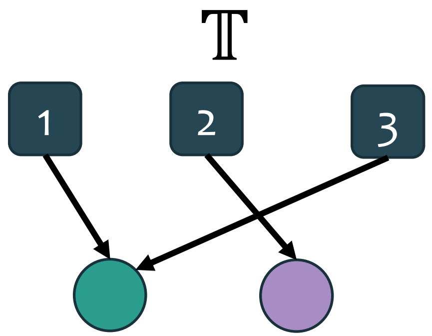
Proof Sketch: Optimality of Threshold Channels for Quantization

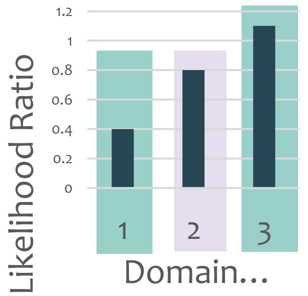
$$\mathcal{A}_{\text{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\text{comm}}(\ell)\}$$

$\mathcal{P}_{\text{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\text{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.





$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

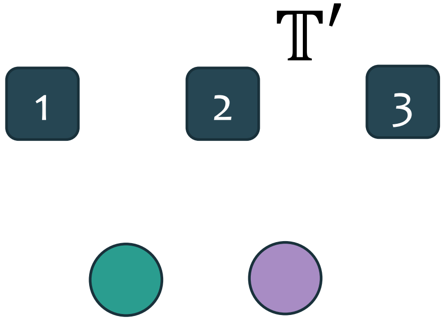
Proof Sketch: Optimality of Threshold Channels for Quantization

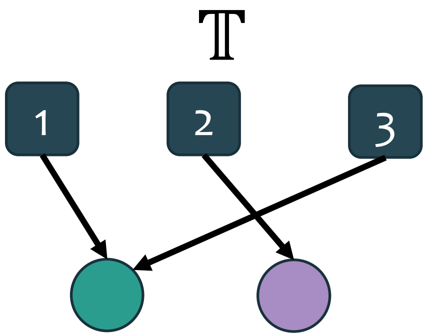
$\mathcal{P}_{\text{comm}}(\ell)$: All channels of output size ℓ

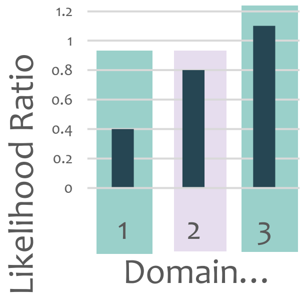
$$\mathcal{A}_{\text{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\text{comm}}(\ell)\}$$

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\text{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.







$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

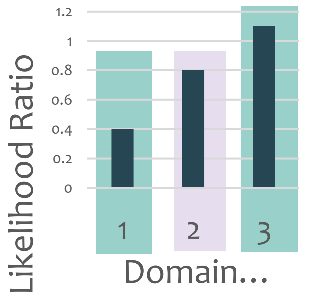
Proof Sketch: Optimality of Threshold Channels for Quantization

$$\mathcal{A}_{\mathrm{comm}} \coloneqq \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\mathrm{comm}}(\ell)\}$$

$\mathcal{P}_{\mathrm{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\mathrm{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.



$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

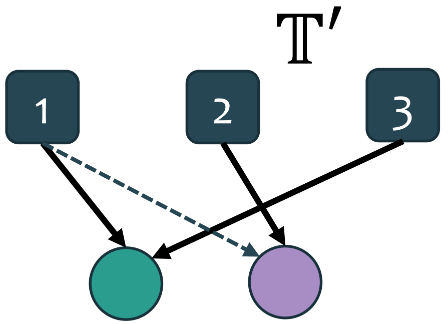
Proof Sketch: Optimality of Threshold Channels for Quantization

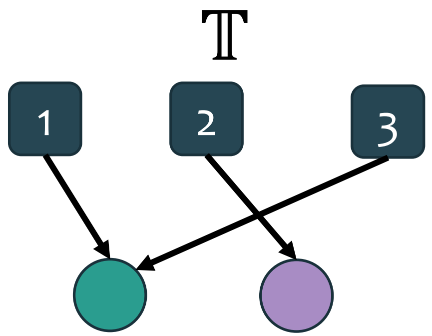
$$\mathcal{A}_{\text{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\text{comm}}(\ell)\}$$

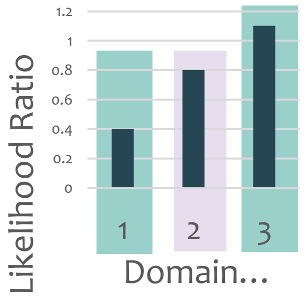
$\mathcal{P}_{\text{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\text{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.







$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

Proof Sketch: Optimality of Threshold Channels for Quantization

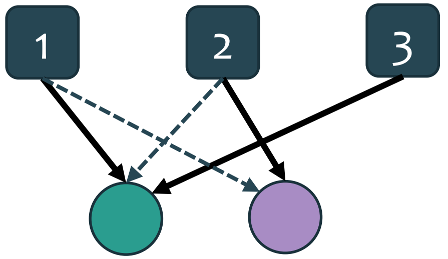
$$\mathcal{A}_{\mathrm{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\mathrm{comm}}(\ell)\}$$

$\mathcal{P}_{\mathrm{comm}}(\ell)$: All channels of output size ℓ

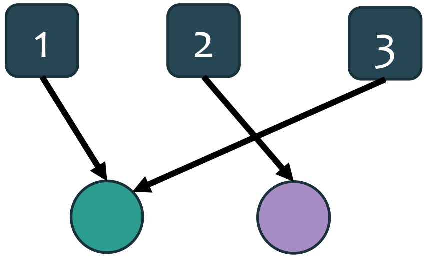
Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\mathrm{comm}}$, then $\mathbb{T}$ must be a threshold channel.

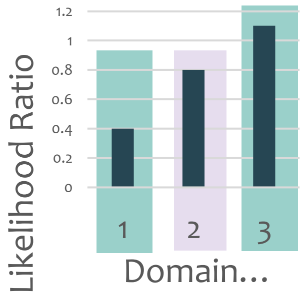
Proof: Suppose $\mathbb{T}$ is not a threshold channel.

$\mathbb{T}'$



$\mathbb{T}$





$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

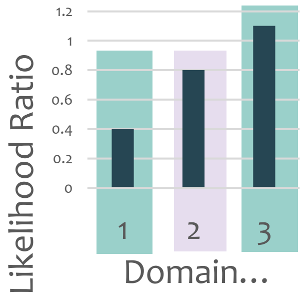
Proof Sketch: Optimality of Threshold Channels for Quantization

$$\mathcal{A}_{\text{comm}} \coloneqq \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\text{comm}}(\ell)\}$$

$\mathcal{P}_{\text{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\text{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.



$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

- Can choose the perturbations s.t. $\mathbb{T}'q = \mathbb{T}q$

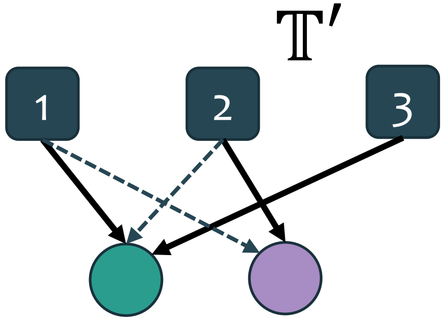
Proof Sketch: Optimality of Threshold Channels for Quantization

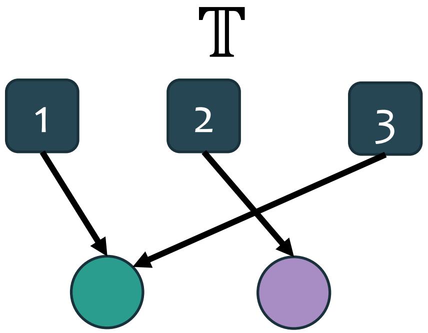
$$\mathcal{A}_{\text{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\text{comm}}(\ell)\}$$

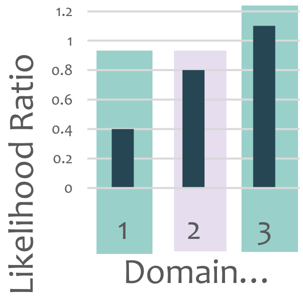
$\mathcal{P}_{\text{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\text{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.







$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

- Can choose the perturbations s.t. $\mathbb{T}'q = \mathbb{T}q$
- However, $\mathbb{T}'p$ puts more mass on ● than $\mathbb{T}p$ (2 has higher likelihood ratio)

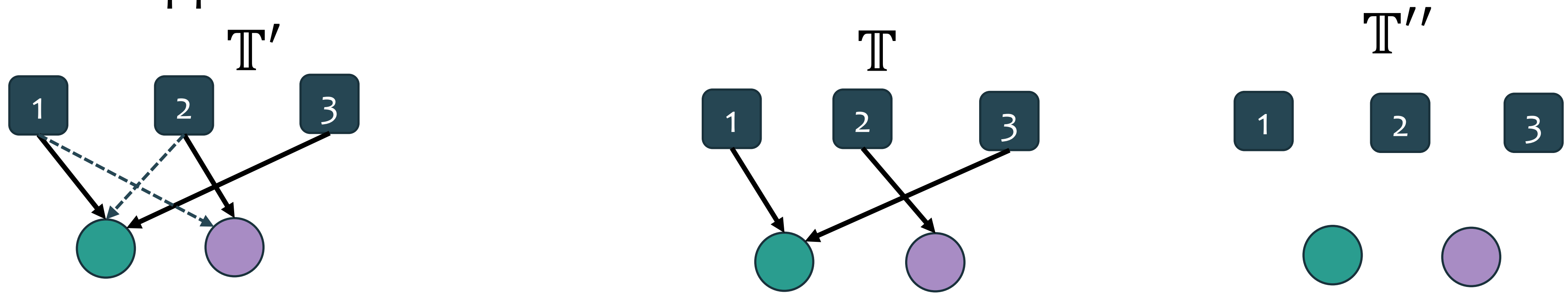
Proof Sketch: Optimality of Threshold Channels for Quantization

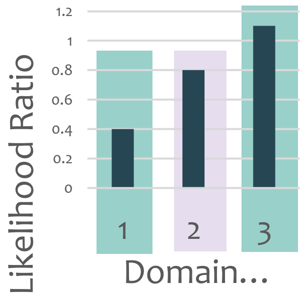
$$\mathcal{A}_{\text{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\text{comm}}(\ell)\}$$

$\mathcal{P}_{\text{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\text{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.





$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

- Can choose the perturbations s.t. $\mathbb{T}'q = \mathbb{T}q$
- However, $\mathbb{T}'p$ puts more mass on ● than $\mathbb{T}p$ (2 has higher likelihood ratio)

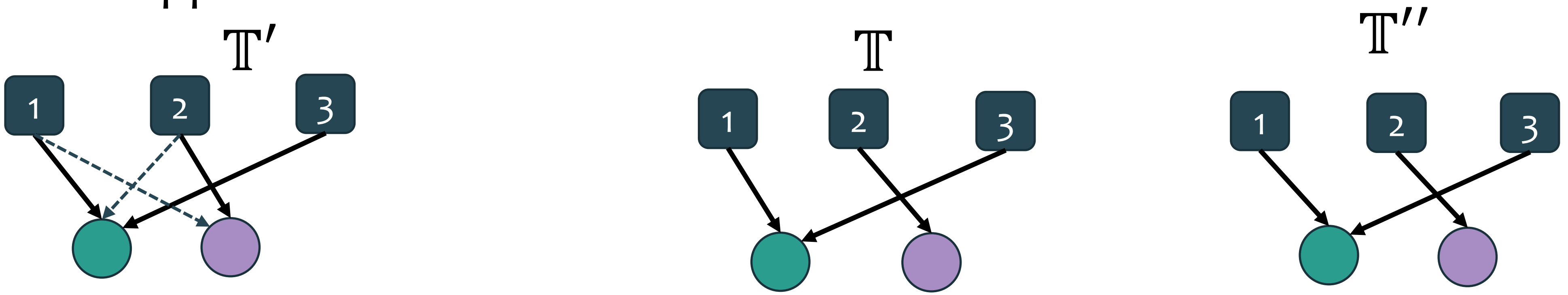
Proof Sketch: Optimality of Threshold Channels for Quantization

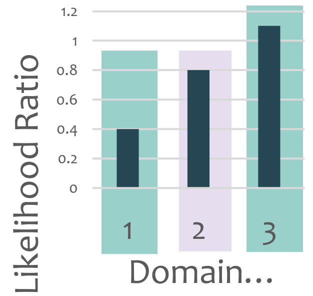
$$\mathcal{A}_{\text{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\text{comm}}(\ell)\}$$

$\mathcal{P}_{\text{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\text{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.





$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

- Can choose the perturbations s.t. $\mathbb{T}'q = \mathbb{T}q$
- However, $\mathbb{T}'p$ puts more mass on ● than $\mathbb{T}p$ (2 has higher likelihood ratio)

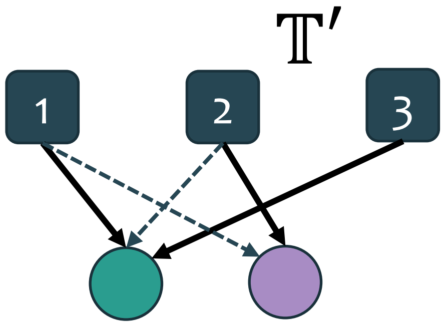
Proof Sketch: Optimality of Threshold Channels for Quantization

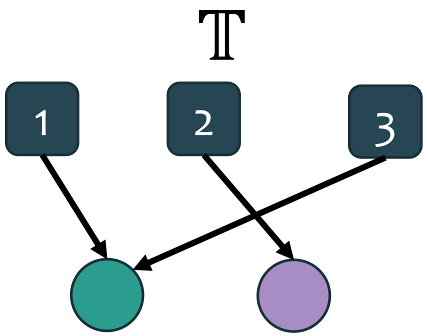
$$\mathcal{A}_{\mathrm{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\mathrm{comm}}(\ell)\}$$

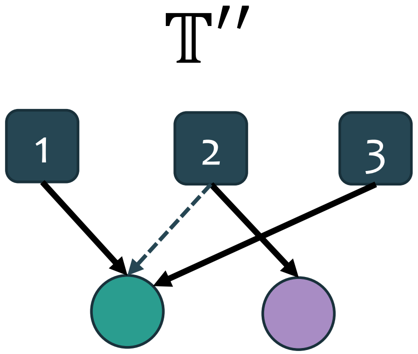
$\mathcal{P}_{\mathrm{comm}}(\ell)$: All channels of output size ℓ

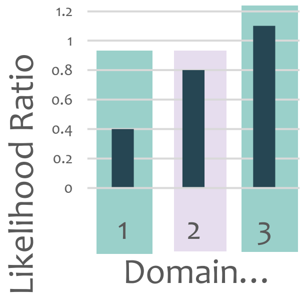
Theorem[**PAJL23**] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\mathrm{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.









$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

- Can choose the perturbations s.t. $\mathbb{T}'q = \mathbb{T}q$
- However, $\mathbb{T}'p$ puts more mass on ● than $\mathbb{T}p$ (2 has higher likelihood ratio)

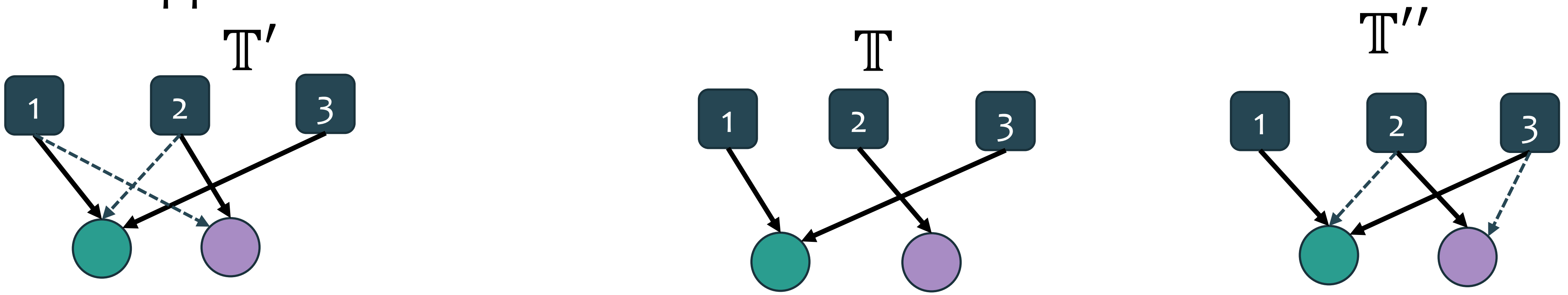
Proof Sketch: Optimality of Threshold Channels for Quantization

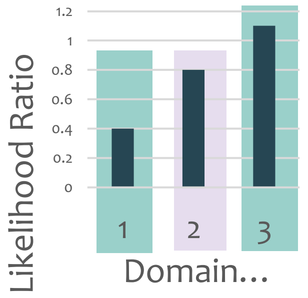
$$\mathcal{A}_{\text{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\text{comm}}(\ell)\}$$

$\mathcal{P}_{\text{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\text{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.





$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

- Can choose the perturbations s.t. $\mathbb{T}'q = \mathbb{T}q$
- However, $\mathbb{T}'p$ puts more mass on ● than $\mathbb{T}p$ (2 has higher likelihood ratio)

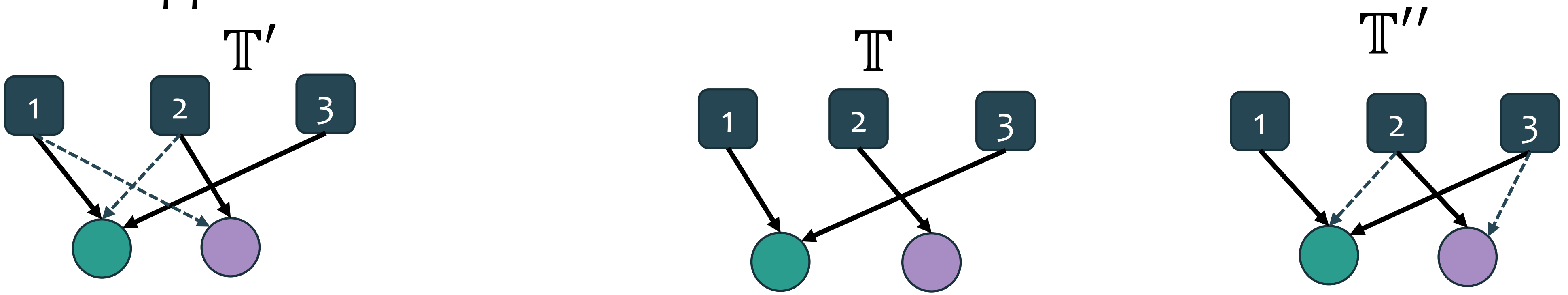
Proof Sketch: Optimality of Threshold Channels for Quantization

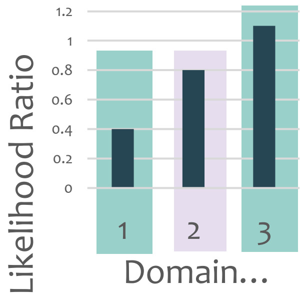
$$\mathcal{A}_{\mathrm{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\mathrm{comm}}(\ell)\}$$

$\mathcal{P}_{\mathrm{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\mathrm{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.





$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

- Can choose the perturbations s.t. $\mathbb{T}'q = \mathbb{T}q = \mathbb{T}''q$
- However, $\mathbb{T}'p$ puts more mass on than $\mathbb{T}p$ (2 has higher likelihood ratio)

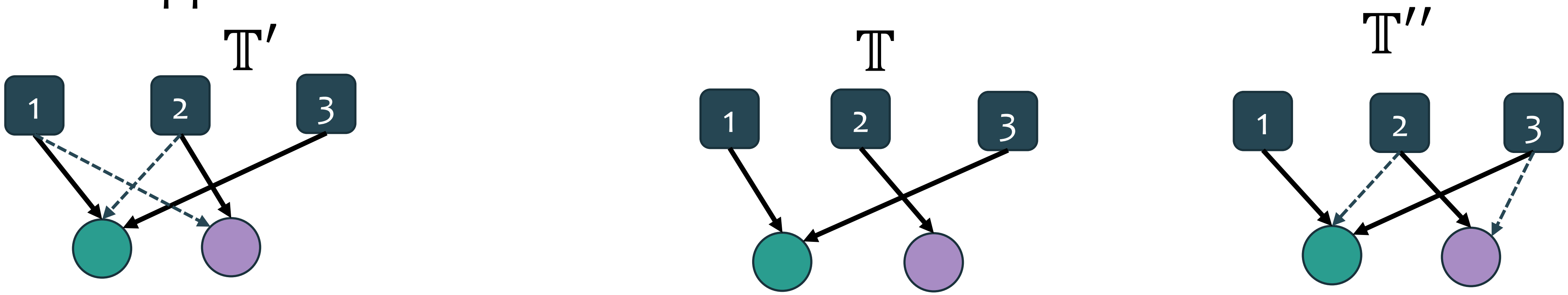
Proof Sketch: Optimality of Threshold Channels for Quantization

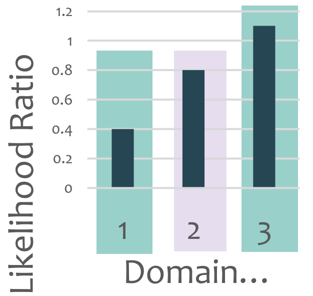
$$\mathcal{A}_{\text{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\text{comm}}(\ell)\}$$

$\mathcal{P}_{\text{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\text{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.





$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

- Can choose the perturbations s.t. $\mathbb{T}'q = \mathbb{T}q = \mathbb{T}''q$
- However, $\mathbb{T}'p$ puts more mass on ● than $\mathbb{T}p$ (2 has higher likelihood ratio)
- But, $\mathbb{T}''p$ puts less mass on ● than $\mathbb{T}p$

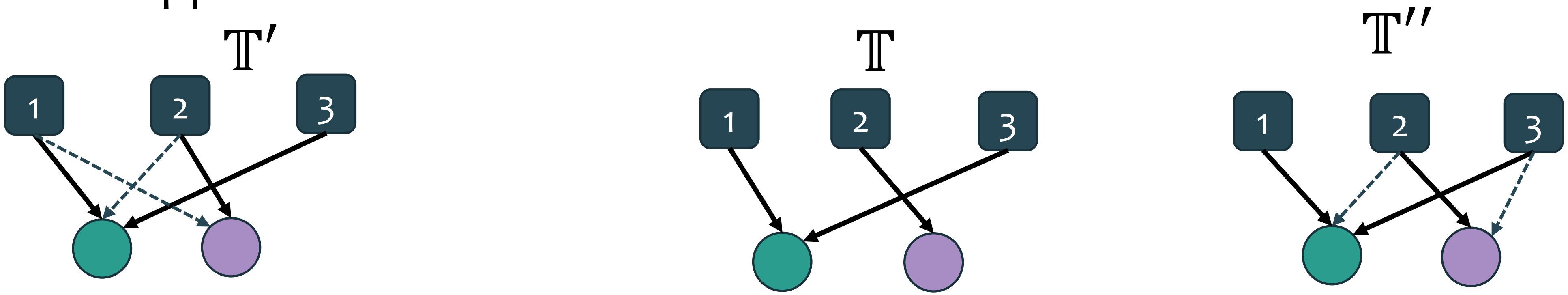
Proof Sketch: Optimality of Threshold Channels for Quantization

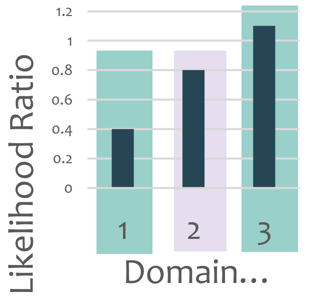
$$\mathcal{A}_{\text{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\text{comm}}(\ell)\}$$

$\mathcal{P}_{\text{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\text{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.





$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

- Can choose the perturbations s.t. $\mathbb{T}'q = \mathbb{T}q = \mathbb{T}''q$
- However, $\mathbb{T}'p$ puts more mass on ● than $\mathbb{T}p$ (2 has higher likelihood ratio)
- But, $\mathbb{T}''p$ puts less mass on ● than $\mathbb{T}p$
- Fluctuations in opposite directions → $(\mathbb{T}p, \mathbb{T}q)$ can't be an extreme point

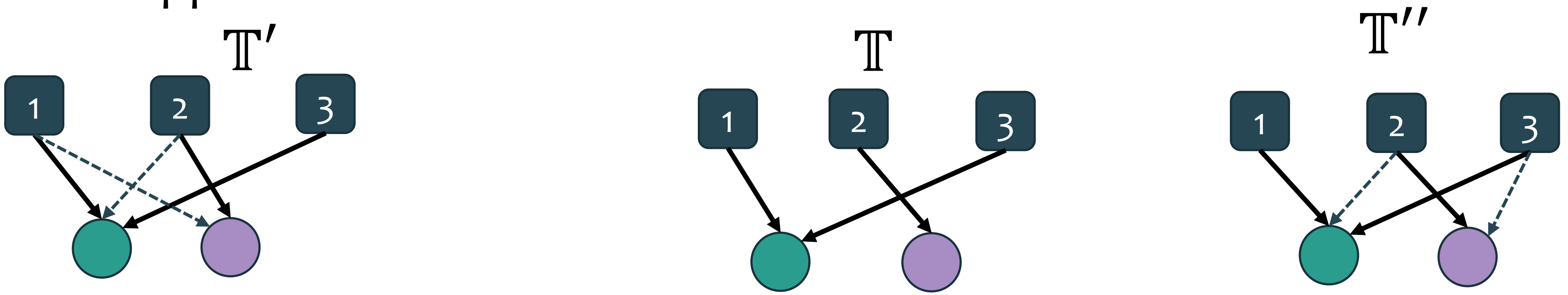
Proof Sketch: Optimality of Threshold Channels for Quantization

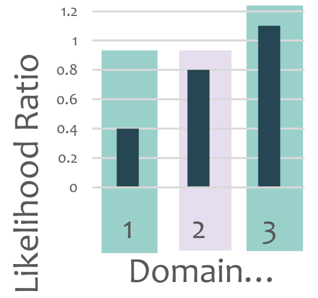
$$\mathcal{A}_{\text{comm}} := \{(\mathbb{T}p, \mathbb{T}q) : \mathbb{T} \in \mathcal{P}_{\text{comm}}(\ell)\}$$

$\mathcal{P}_{\text{comm}}(\ell)$: All channels of output size ℓ

Theorem[PAJL23] If $(\mathbb{T}p, \mathbb{T}q)$ is an extreme point of $\mathcal{A}_{\text{comm}}$, then $\mathbb{T}$ must be a threshold channel.

Proof: Suppose $\mathbb{T}$ is not a threshold channel.





$$\frac{p_1}{q_1} < \frac{p_2}{q_2} < \frac{p_3}{q_3}$$

- Can choose the perturbations s.t. $\mathbb{T}'q = \mathbb{T}q = \mathbb{T}''q$
- However, $\mathbb{T}'p$ puts more mass on than $\mathbb{T}p$ (2 has higher likelihood ratio)
- But, $\mathbb{T}''p$ puts less mass on than $\mathbb{T}p$
- Fluctuations in opposite directions → $(\mathbb{T}p, \mathbb{T}q)$ can't be an extreme point

Outline

- Motivation
- Problem Statement
- Our Results
- Proof Sketch
- **Conclusion**

Conclusion and Future Directions

- Derived minmax-optimal sample complexities under privacy
  - No longer depends only on TV distance and Hellinger

Conclusion and Future Directions

- Derived minmax-optimal sample complexities under privacy
  - No longer depends only on TV distance and Hellinger
- Computationally and Communication-efficient algorithms

Conclusion and Future Directions

- Derived minmax-optimal sample complexities under privacy
  - No longer depends only on TV distance and Hellinger
- Computationally and Communication-efficient algorithms

- Open problems:

Conclusion and Future Directions

- Derived minmax-optimal sample complexities under privacy
  - No longer depends only on TV distance and Hellinger
- Computationally and Communication-efficient algorithms

- Open problems:
  - Role of interactivity

Conclusion and Future Directions

- Derived minmax-optimal sample complexities under privacy
  - No longer depends only on TV distance and Hellinger
- Computationally and Communication-efficient algorithms

- Open problems:
  - Role of interactivity
  - Algorithms with better runtime dependence on ℓ --- the output size

Conclusion and Future Directions

- Derived minmax-optimal sample complexities under privacy
  - No longer depends only on TV distance and Hellinger
- Computationally and Communication-efficient algorithms

- Open problems:
  - Role of interactivity
  - Algorithms with better runtime dependence on ℓ --- the output size
  - Characterization of instance-optimal sample complexity
    - Looking beyond TV distance and Hellinger divergence

Conclusion and Future Directions

- Derived minmax-optimal sample complexities under privacy
  - No longer depends only on TV distance and Hellinger
- Computationally and Communication-efficient algorithms

- Open problems:
  - Role of interactivity
  - Algorithms with better runtime dependence on ℓ --- the output size
  - Characterization of instance-optimal sample complexity
    - Looking beyond TV distance and Hellinger divergence
  - M-ary hypothesis testing, optimally

Conclusion and Future Directions

- Derived minmax-optimal sample complexities under privacy
  - No longer depends only on TV distance and Hellinger
- Computationally and Communication-efficient algorithms

- Open problems:
  - Role of interactivity
  - Algorithms with better runtime dependence on ℓ --- the output size
  - Characterization of instance-optimal sample complexity
    - Looking beyond TV distance and Hellinger divergence
  - M-ary hypothesis testing, optimally

Thank you!