Optimal Piecewise-based Mechanism for Collecting Bounded Numerical Data under Local Differential Privacy

Authors: Ye Zheng, Sumita Mishra, Yidan Hu

Rochester Institute of Technology

LDP Mechanisms

- Randomization algorithm $\mathcal{M}: \mathcal{D} \rightarrow \widetilde{\mathcal{D}}$

  - quantifiable privacy for data $x \in \mathcal{D}$

$$\forall x_1, x_2 \in \mathcal{D}, \forall y \in \widetilde{\mathcal{D}} \qquad \max \frac{\Pr[\mathcal{M}(x_1) = y]}{\Pr[\mathcal{M}(x_2) = y]} \leq e^{\varepsilon}$$

Distinguishability between x_1 and x_2 (sensitive data)
from y (randomized data)

LDP Mechanisms

- Randomization algorithm $\mathcal{M}: \mathcal{D} \rightarrow \widetilde{\mathcal{D}}$

  \- quantifiable privacy for data $x \in \mathcal{D}$

$$\forall x_1, x_2 \in \mathcal{D}, \forall y \in \widetilde{\mathcal{D}} \qquad \max \frac{\Pr[\mathcal{M}(x_1) = y]}{\Pr[\mathcal{M}(x_2) = y]} \leq e^{\varepsilon}$$

Distinguishability between x_1 and x_2 (sensitive data)
from y (randomized data)

Privacy quantified by ε

$x_1 \rightarrow \mathcal{M} \rightarrow y$

Provable defense against
data inference attack

LDP Mechanisms

- Randomization algorithm $\mathcal{M}: \mathcal{D} \rightarrow \widetilde{\mathcal{D}}$

  - quantifiable privacy for data $x \in \mathcal{D}$

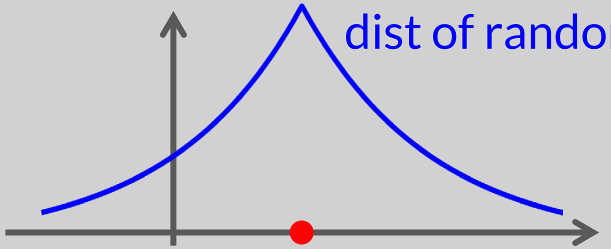
$$\forall x_1, x_2 \in \mathcal{D}, \forall y \in \widetilde{\mathcal{D}} \qquad \max \boxed{\frac{\Pr[\mathcal{M}(x_1) = y]}{\Pr[\mathcal{M}(x_2) = y]}} \leq e^{\varepsilon}$$

Privacy quantified by ε

$x_1 \rightarrow \mathcal{M} \rightarrow y$

Provable defense against data inference attack

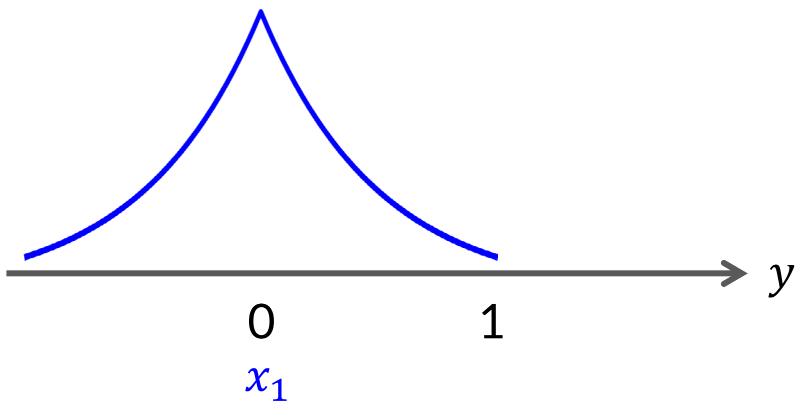
Data utility by expected errors



Sensitive $x_1 = 0.5$

LDP Mechanisms for Numerical Domain

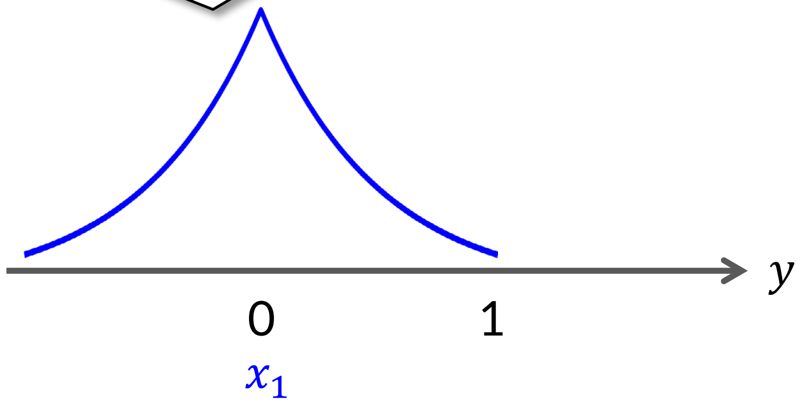
Laplace mechanism: $[0,1] \rightarrow (-\infty, +\infty)$



LDP Mechanisms for Numerical Domain

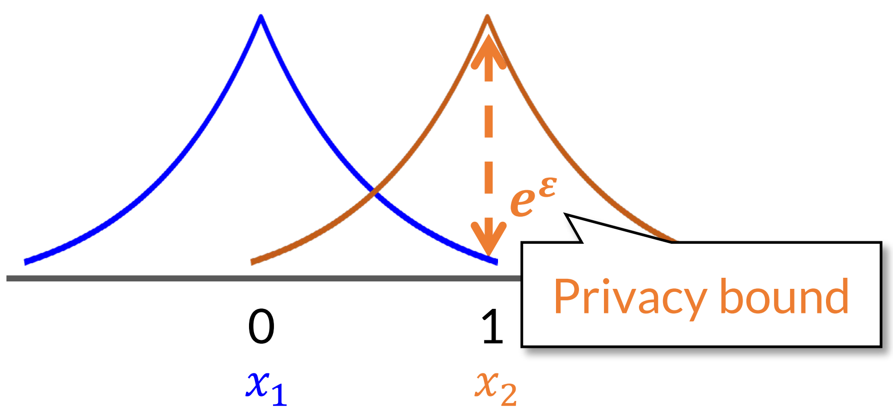
Symmetric (unbiased)
Useful for mean estimation
i.e. $(x_1 + x_2 + \cdots)/n$

La… $, +\infty)$



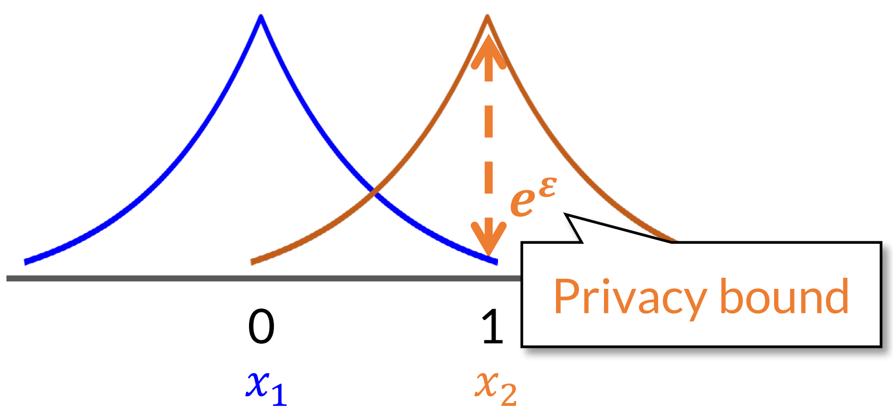
LDP Mechanisms for Numerical Domain

Laplace mechanism: $[0,1] \rightarrow (-\infty, +\infty)$

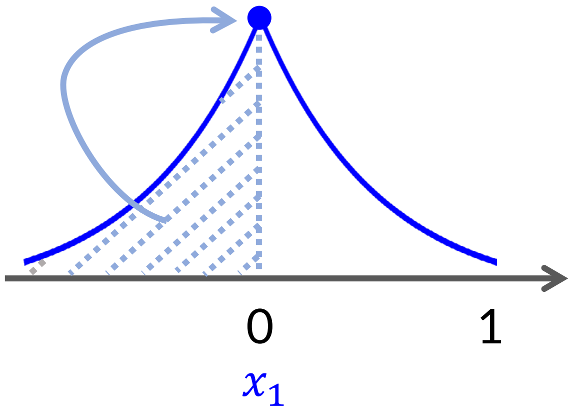


LDP Mechanisms for Numerical Domain

Laplace mechanism: $[0,1] \rightarrow (-\infty, +\infty)$

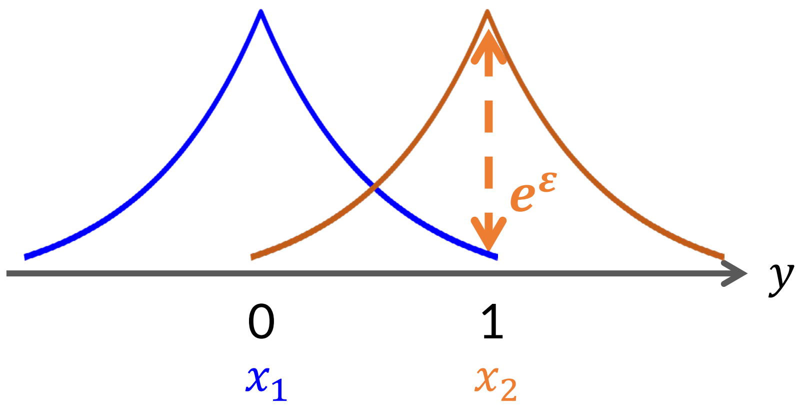


Laplace + truncation: $[0,1] \rightarrow [0,1]$

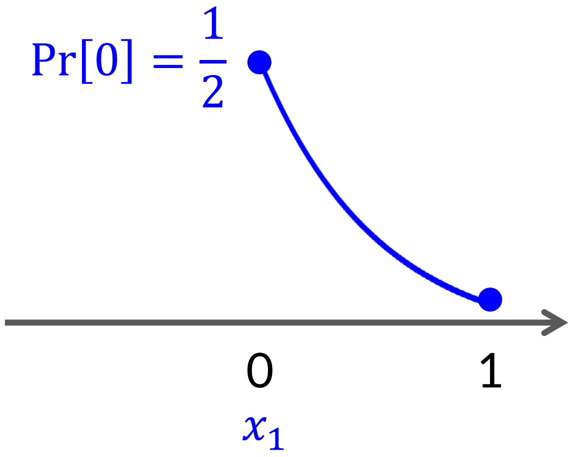


LDP Mechanisms for Numerical Domain

Laplace mechanism: $[0,1] \rightarrow (-\infty, +\infty)$

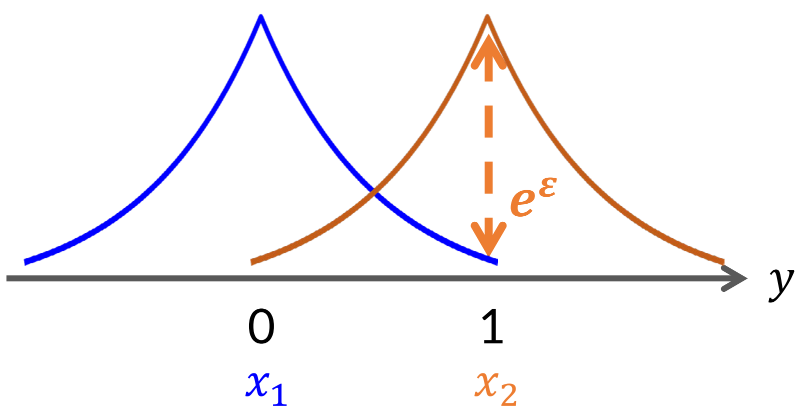


Laplace + truncation: $[0,1] \rightarrow [0,1]$



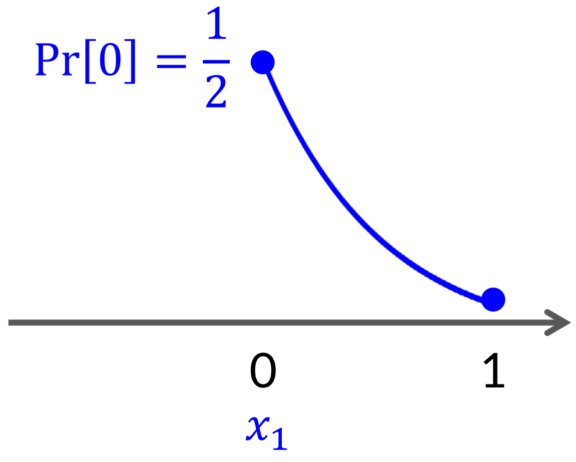
LDP Mechanisms for Numerical Domain

Laplace mechanism: $[0,1] \rightarrow (-\infty, +\infty)$



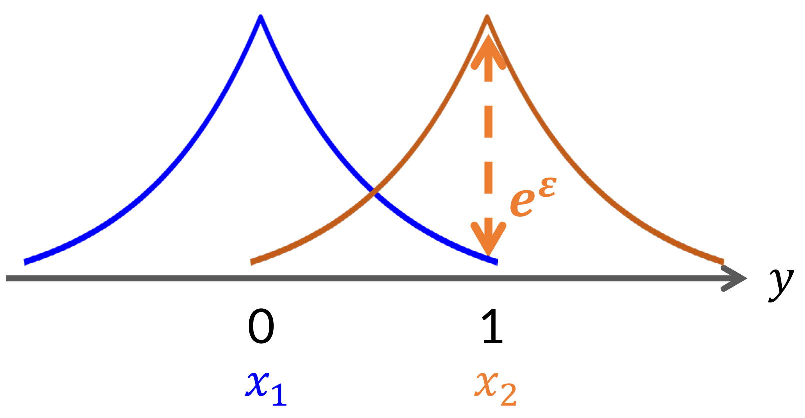
Laplace + truncation: $[0,1] \rightarrow [0,1]$

Smaller output
Useful for distribution estimation
i.e. $\text{dist}\{x_1, x_2, \cdots\}$

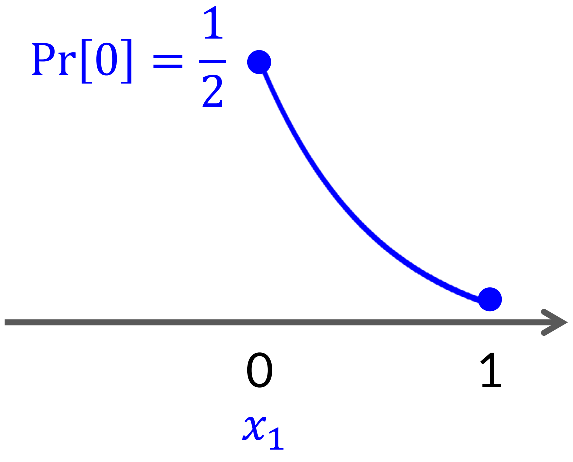


LDP Mechanisms for Numerical Domain

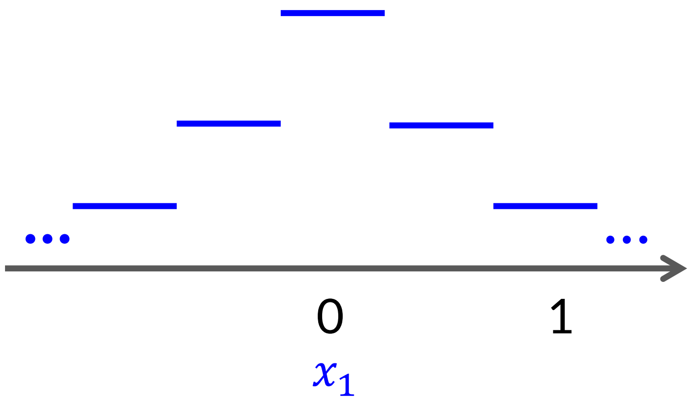
Laplace mechanism: $[0,1] \rightarrow (-\infty, +\infty)$



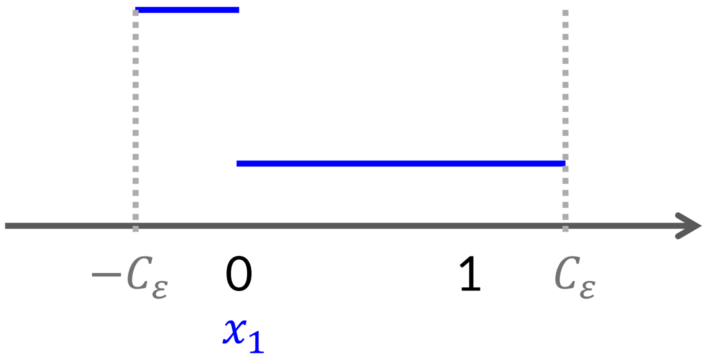
Laplace + truncation: $[0,1] \rightarrow [0,1]$



Staircase: $[0,1] \rightarrow (-\infty, +\infty)$

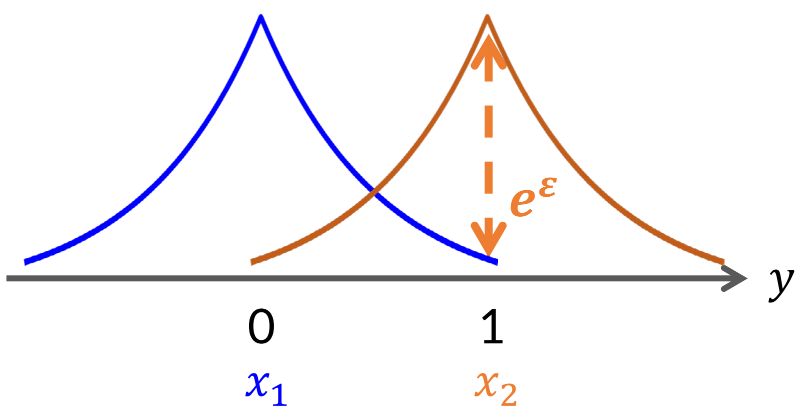


Piecewise mechanism: $[0,1] \rightarrow [-C_{\varepsilon}, C_{\varepsilon}]$

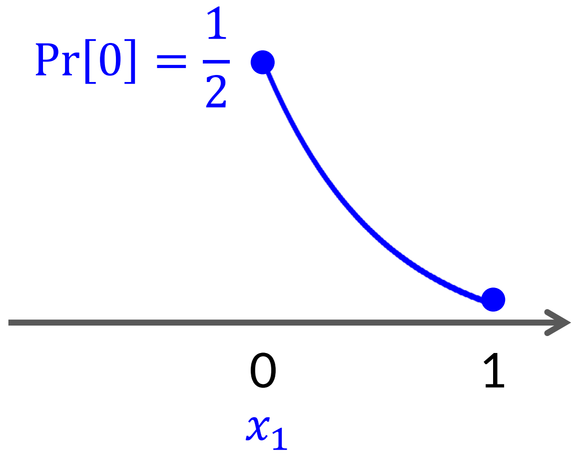


LDP Mechanisms for Numerical Domain

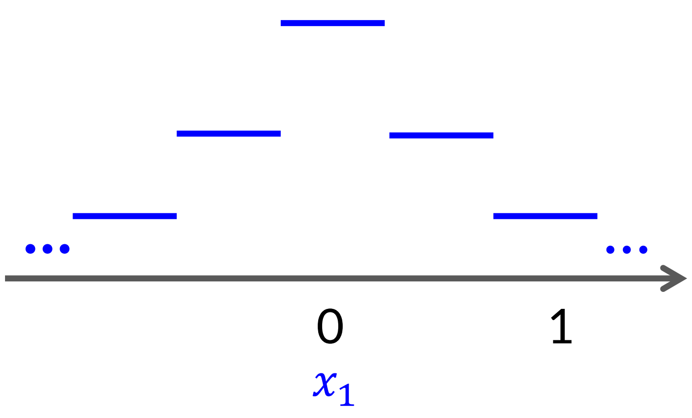
Laplace mechanism: $[0,1] \rightarrow (-\infty, +\infty)$



Laplace + truncation: $[0,1] \rightarrow [0,1]$



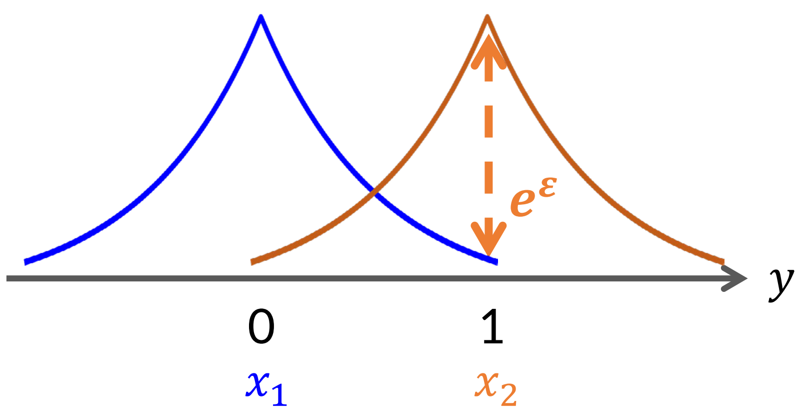
Staircase: $[0,1] \rightarrow (-\infty, +\infty)$



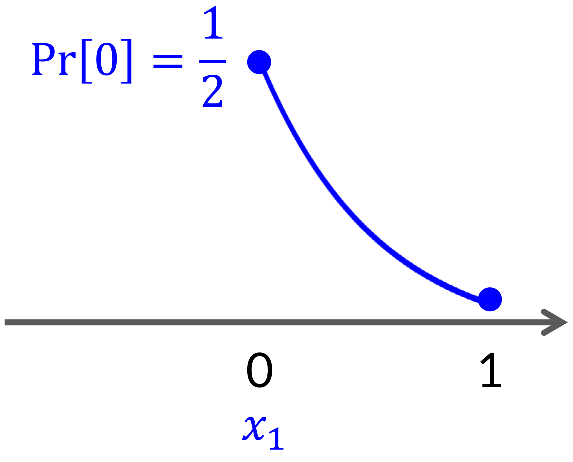
Piecewise mechanism: $[0,1] \rightarrow [-C_{\varepsilon}, C_{\varepsilon}]$

LDP Mechanisms for Numerical Domain

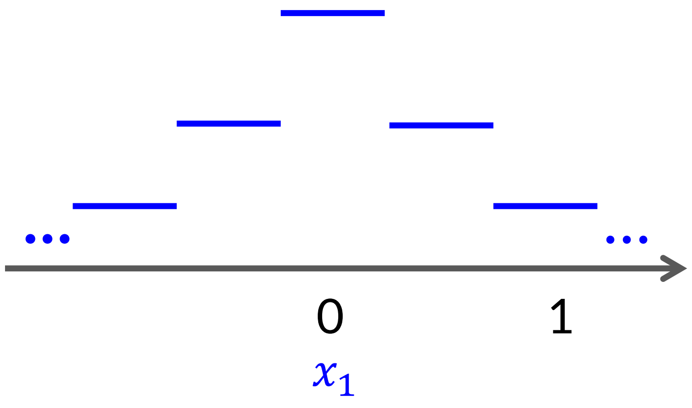
Laplace mechanism: $[0,1] \to (-\infty, +\infty)$



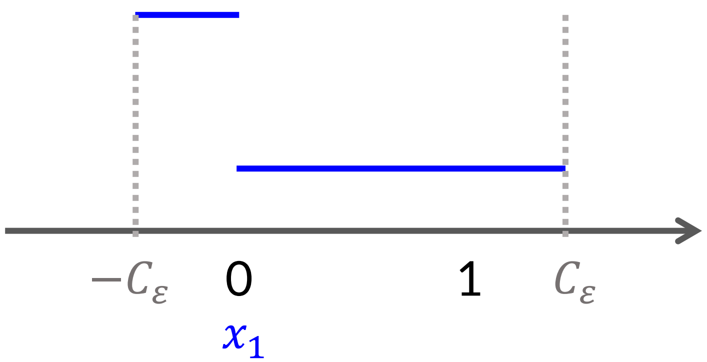
Laplace + truncation: $[0,1] \to [0,1]$



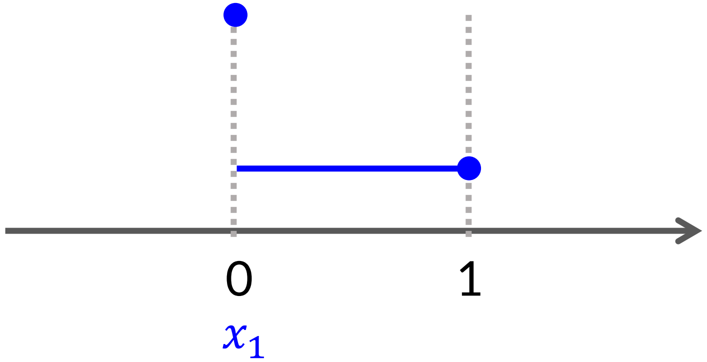
Staircase: $[0,1] \to (-\infty, +\infty)$



Piecewise mechanism: $[0,1] \to [-C_{\varepsilon}, C_{\varepsilon}]$



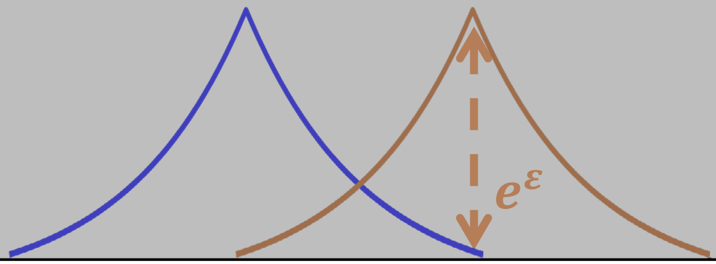
Piecewise + truncation: $[0,1] \to [0,1]$



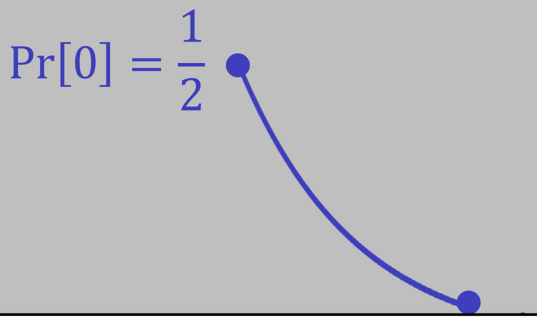
…

LDP Mechanisms for Numerical Domain

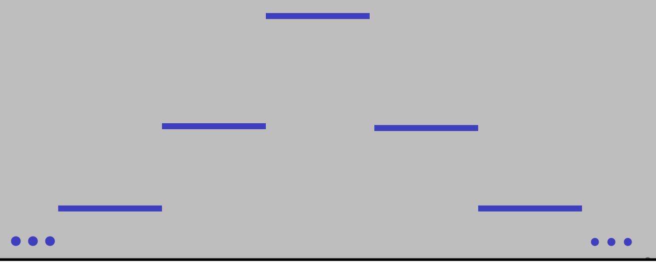
Laplace mechanism: $[0,1] \rightarrow (-\infty, +\infty)$



Laplace + truncation: $[0,1] \rightarrow [0,1]$

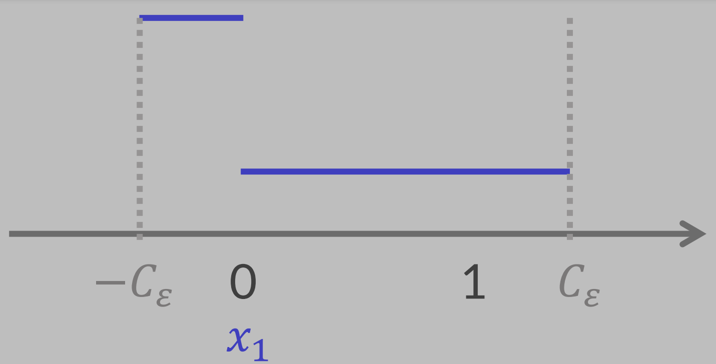


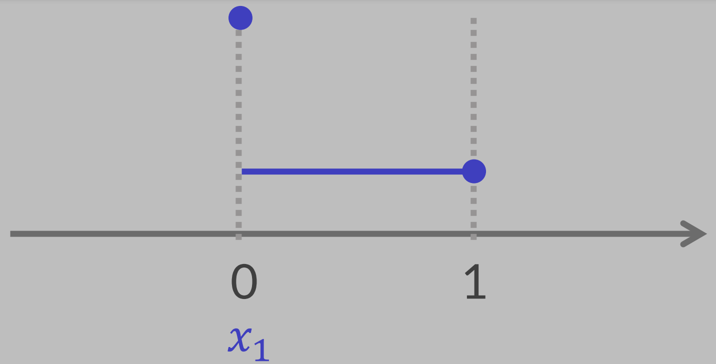
Staircase: $[0,1] \rightarrow (-\infty, +\infty)$



Privacy: LDP with the **same ε**

Utility: **Different errors**

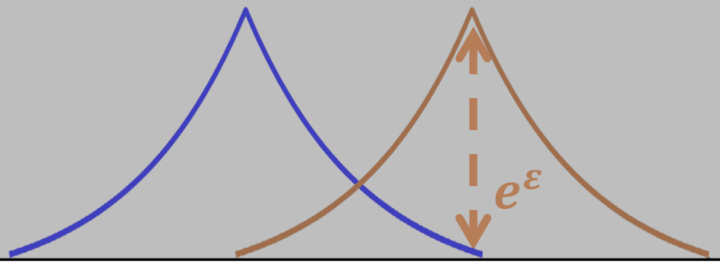




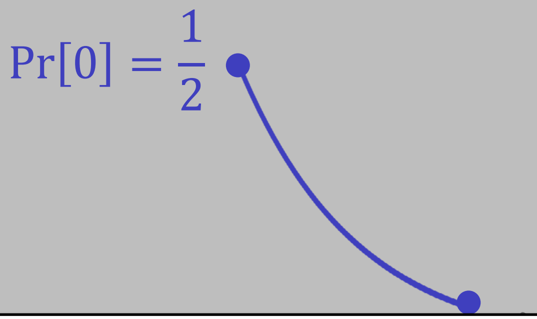
...

LDP Mechanisms for Numerical Domain

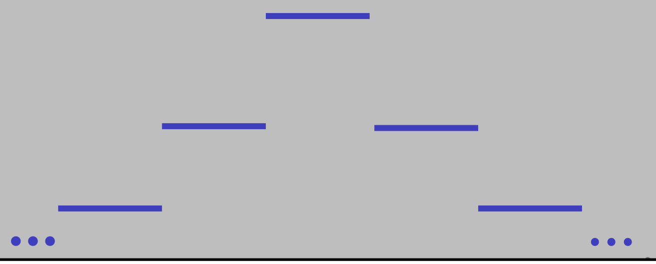
Laplace mechanism: $[0,1] \to (-\infty, +\infty)$



Laplace + truncation: $[0,1] \to [0,1]$



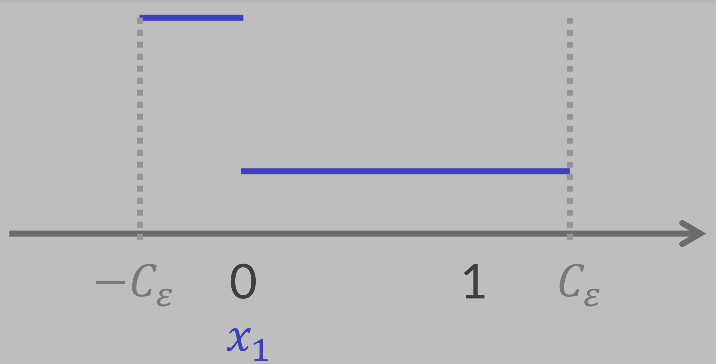
Staircase: $[0,1] \to (-\infty, +\infty)$

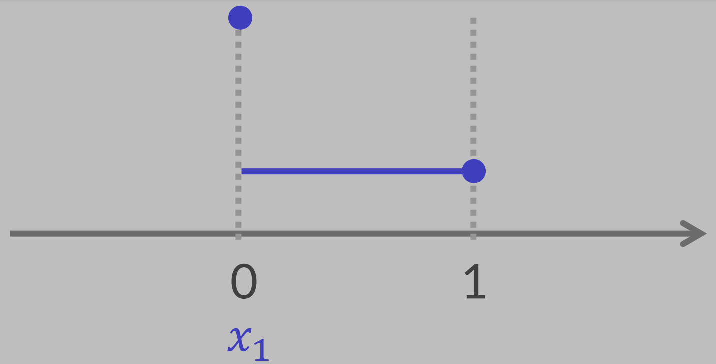


Privacy: LDP with the **same ε**

Utility: Different errors

Q: What is the **optimal** LDP mechanism?

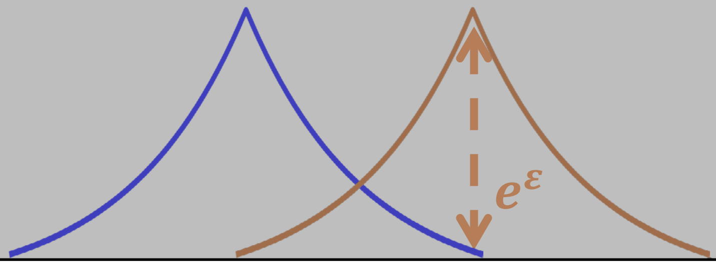




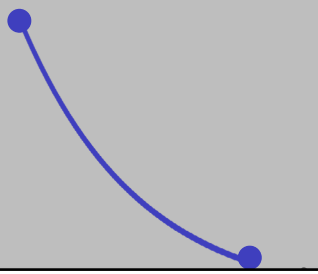
…

LDP Mechanisms for Numerical Domain

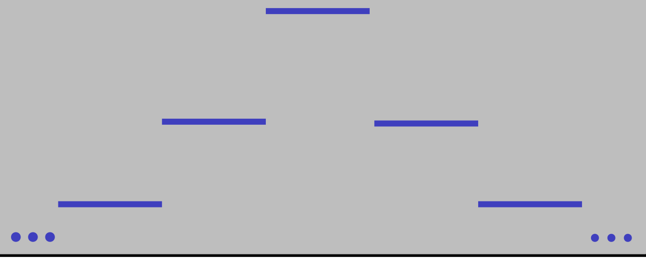
Laplace mechanism: $[0,1] \rightarrow (-\infty, +\infty)$



Laplace + truncation: $[0,1] \rightarrow [0,1]$



Staircase: $[0,1] \rightarrow (-\infty, +\infty)$

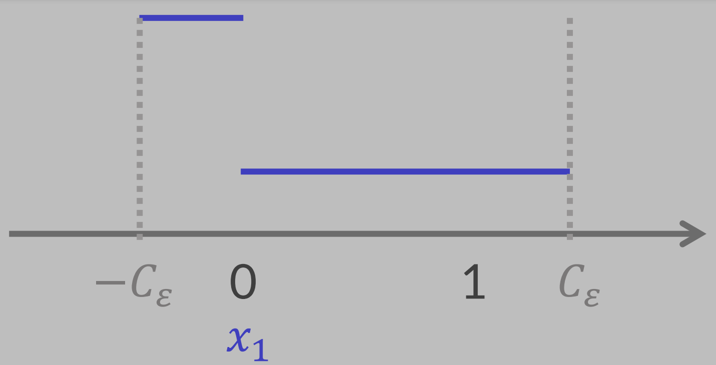


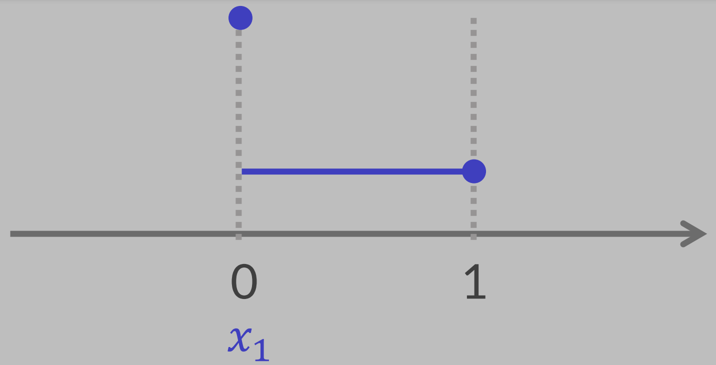
Privacy: LDP with the **same $\boldsymbol{\varepsilon}$**

Utility: Different errors

Q: What is the **optimal piecewise-based** mechanism?

?



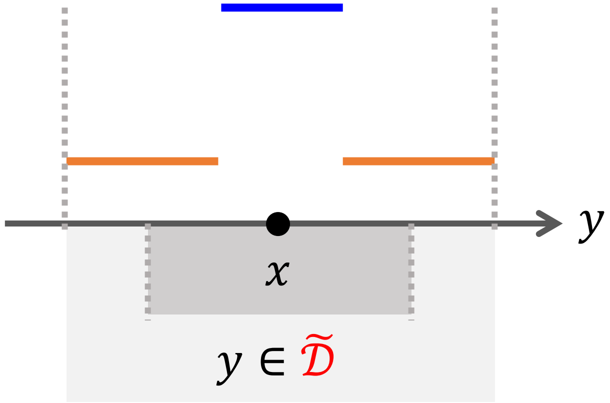


…

3-Piecewise Mechanism

- 3-piecewise distributions on bounded numerical domain $\mathcal{D} \to \widetilde{\mathcal{D}}$

  - given input x, sample output y from a distribution

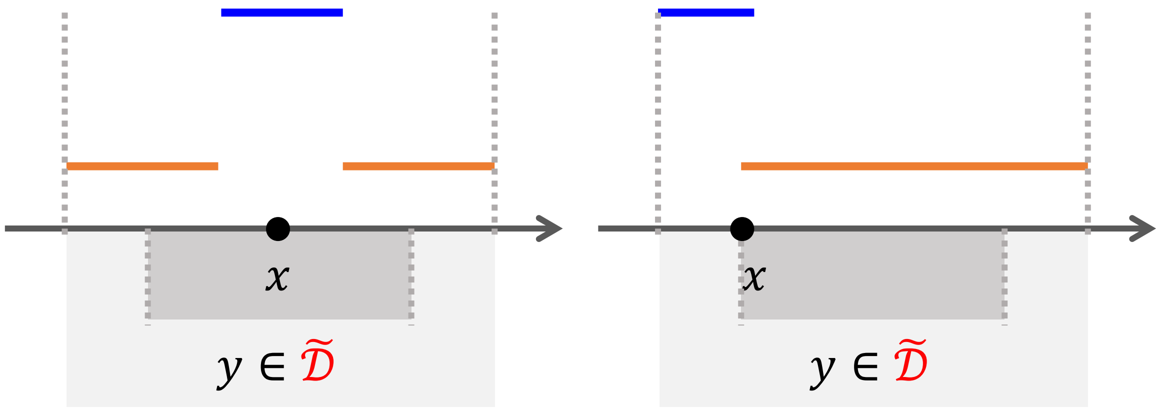


$$pdf[\mathcal{M}(x) = y] = \begin{cases} p_\varepsilon & \text{if } y \in [l_{x,\varepsilon}, r_{x,\varepsilon}] \\ \dfrac{p_\varepsilon}{\exp(\varepsilon)} & \text{if } y \in \widetilde{\mathcal{D}} \backslash [l_{x,\varepsilon}, r_{x,\varepsilon}] \end{cases}$$

3-Piecewise Mechanism

- 3-piecewise distributions on bounded numerical domain $\mathcal{D} \to \tilde{\mathcal{D}}$

  - given input x, sample output y from a distribution



Sampling probability depends on ε

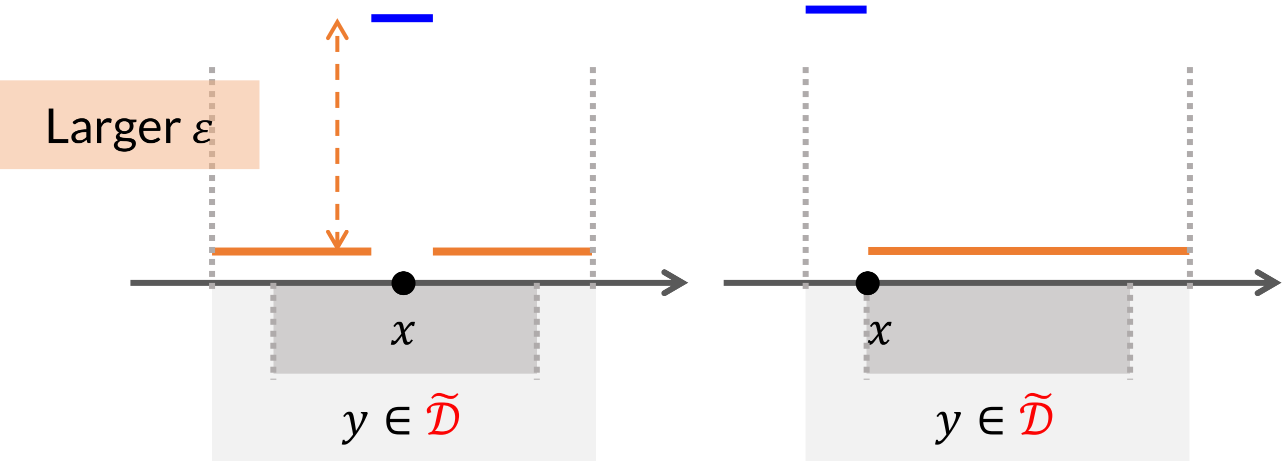
Sampling interval depends on x and ε

$$pdf[\mathcal{M}(x) = y] = \begin{cases} p_\varepsilon & \text{if } y \in [l_{x,\varepsilon}, r_{x,\varepsilon}] \\ \dfrac{p_\varepsilon}{\exp(\varepsilon)} & \text{if } y \in \tilde{\mathcal{D}} \setminus [l_{x,\varepsilon}, r_{x,\varepsilon}] \end{cases}$$

3-Piecewise Mechanism

- 3-piecewise distributions on bounded numerical domain $\mathcal{D} \rightarrow \widetilde{\mathcal{D}}$

  - given input x, sample output y from a distribution



Sampling probability depends on ε

Sampling interval depends on x and ε

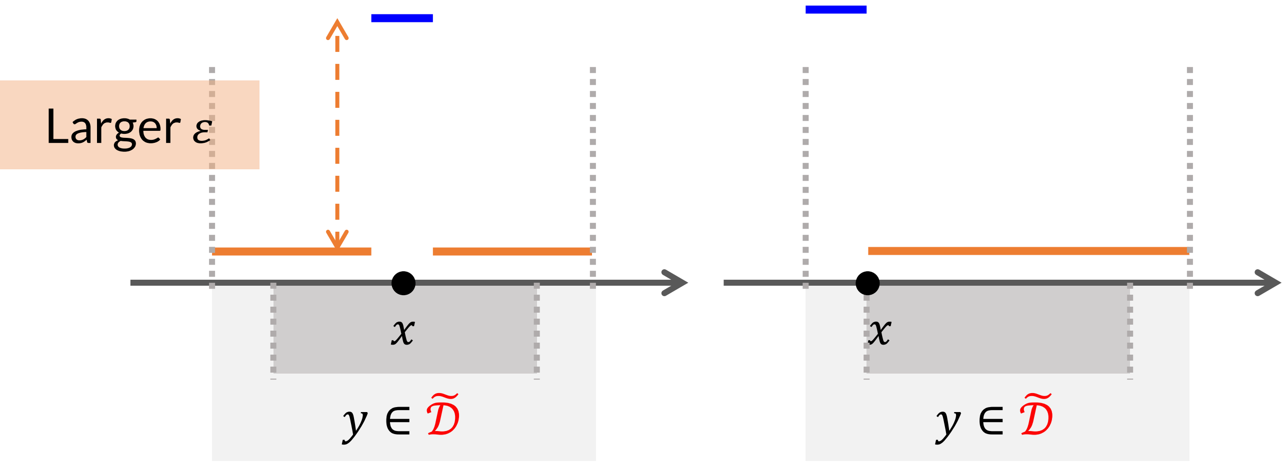
$$pdf[\mathcal{M}(x) = y] = \begin{cases} p_\varepsilon & \text{if } y \in [l_{x,\varepsilon}, r_{x,\varepsilon}] \\ \dfrac{p_\varepsilon}{\exp(\varepsilon)} & \text{if } y \in \widetilde{\mathcal{D}} \backslash [l_{x,\varepsilon}, r_{x,\varepsilon}] \end{cases}$$

- Instantiations: PM [ICDE'19], SW [SIGMOD'20], PTT [TMC'24] (design different $p_\varepsilon, l_{x,\varepsilon}, r_{x,\varepsilon}$)

3-Piecewise Mechanism

- 3-piecewise distributions on bounded numerical domain $\mathcal{D} \rightarrow \widetilde{\mathcal{D}}$

  - given input x, sample output y from a distribution



Sampling probability depends on ε

Sampling interval depends on x and ε

$$pdf[\mathcal{M}(x) = y] = \begin{cases} p_\varepsilon & \text{if } y \in [l_{x,\varepsilon}, r_{x,\varepsilon}] \\ \dfrac{p_\varepsilon}{\exp(\varepsilon)} & \text{if } y \in \widetilde{\mathcal{D}} \backslash [l_{x,\varepsilon}, r_{x,\varepsilon}] \end{cases}$$

- Instantiations: PM [ICDE'19], SW [SIGMOD'20], PTT [TMC'24] (design different $p_\varepsilon, l_{x,\varepsilon}, r_{x,\varepsilon}$)

  - different errors, but **without optimality**

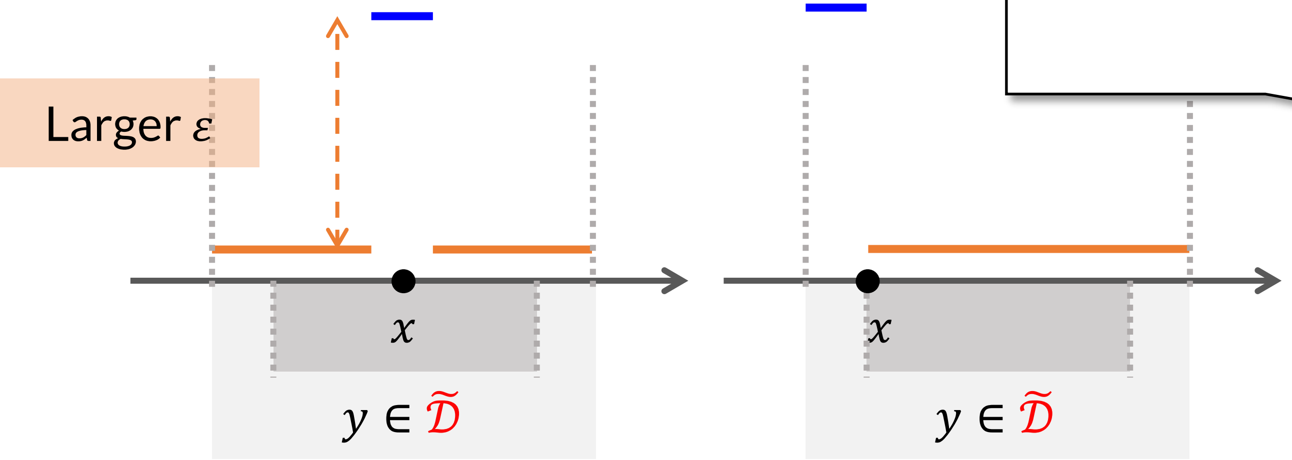
3-Piecewise Mechanism

- 3-piecewise distributions on boun

  - given input x, sample output y fr

NOT enough to study optimality of piecewise-based mechanism

- only 3 pieces, two probabilities



$$pdf[\mathcal{M}(x) = y] = \begin{cases} p_\varepsilon & \text{if } y \in [l_{x,\varepsilon}, r_{x,\varepsilon}] \\ \dfrac{p_\varepsilon}{\exp(\varepsilon)} & \text{if } y \in \widetilde{\mathcal{D}} \backslash [l_{x,\varepsilon}, r_{x,\varepsilon}] \end{cases}$$

- Instantiations: PM [ICDE'19], SW [SIGMOD'20], PTT [TMC'24] (design different $p_\varepsilon, l_{x,\varepsilon}, r_{x,\varepsilon}$)

  - different errors, but **without optimality**

3-Piecewise Mechanism

- 3-piecewise distributions on boun

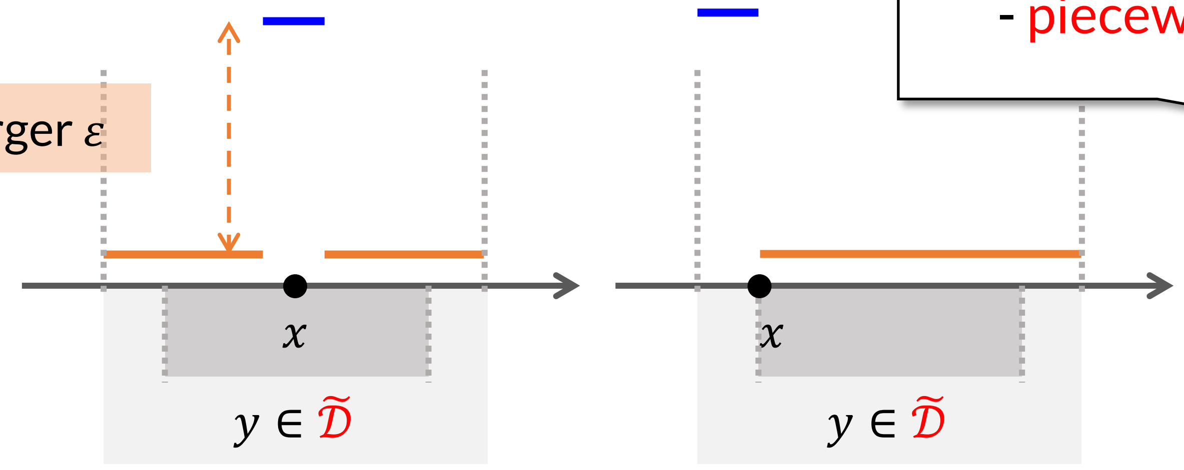
  - given input x, sample output y fr

NOT enough to study optimality of piecewise-based mechanism

- only 3 pieces, two probabilities

- piecewise distribution can have more pieces and probabilities

Larger ε



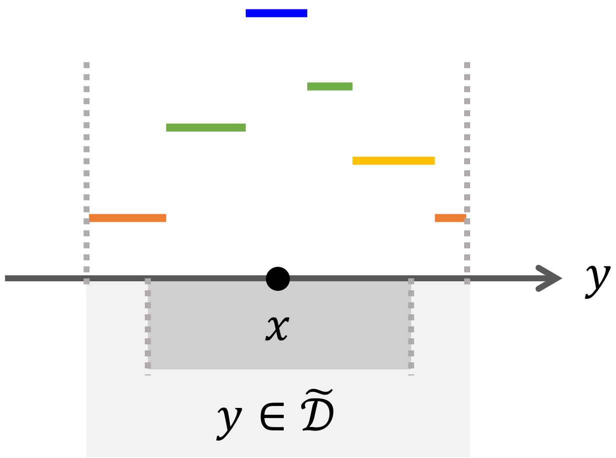
$$pdf[\mathcal{M}(x) = y] = \begin{cases} p_\varepsilon & \text{if } y \in [l_{x,\varepsilon}, r_{x,\varepsilon}] \\ \dfrac{p_\varepsilon}{\exp(\varepsilon)} & \text{if } y \in \widetilde{\mathcal{D}} \backslash [l_{x,\varepsilon}, r_{x,\varepsilon}] \end{cases}$$

- Instantiations: PM [ICDE'19], SW [SIGMOD'20], PTT [TMC'24] (design different $p_\varepsilon, l_{x,\varepsilon}, r_{x,\varepsilon}$)

  - different errors, but **without optimality**

Generalized Piecewise-based Mechanism

- **Most generalized version:** m-piecewise distribution

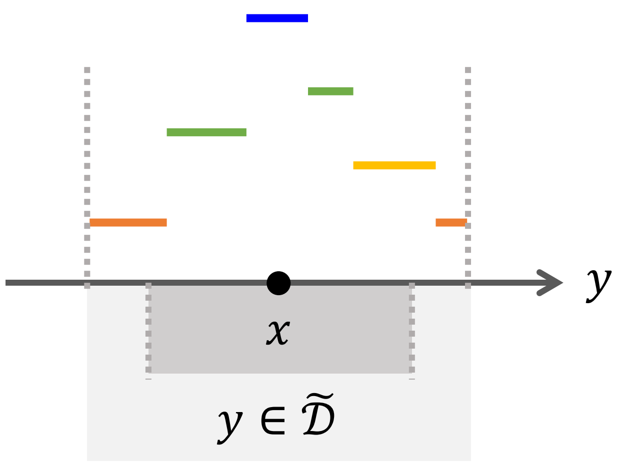


$$pdf[\mathcal{M}(x) = y] = \begin{cases} p_{1,\varepsilon} & \text{if } y \in [l_{1,x,\varepsilon}, r_{1,x,\varepsilon}] \\ p_{2,\varepsilon} & \text{if } y \in [l_{2,x,\varepsilon}, r_{2,x,\varepsilon}] \\ & \quad \dots \\ p_{m,\varepsilon} & \text{if } y \in [l_{m,x,\varepsilon}, r_{m,x,\varepsilon}] \end{cases}$$

$\frac{p_{i,\varepsilon}}{p_{j,\varepsilon}} \leq e^{\varepsilon}$ (LDP constraint)

Generalized Piecewise-based Mechanism

- **Most generalized version:** m-piecewise distribution



$$pdf[\mathcal{M}(x) = y] = \begin{cases} p_{1,\varepsilon} & \text{if } y \in [l_{1,x,\varepsilon}, r_{1,x,\varepsilon}] \\ p_{2,\varepsilon} & \text{if } y \in [l_{2,x,\varepsilon}, r_{2,x,\varepsilon}] \\ & \ldots \\ p_{m,\varepsilon} & \text{if } y \in [l_{m,x,\varepsilon}, r_{m,x,\varepsilon}] \end{cases}$$

$$\frac{p_{i,\varepsilon}}{p_{j,\varepsilon}} \leq e^{\varepsilon}$$ (LDP constraint)

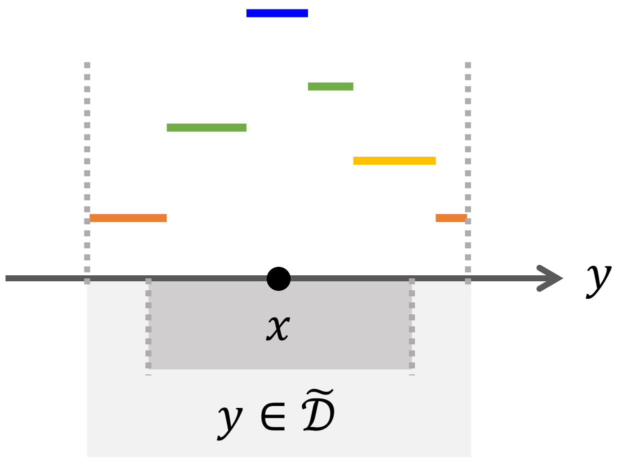
- Error (data utility):

$$\mathcal{L}(y, x)$$

$$\mathcal{L}(y, x) := |y - x|^p$$

Generalized Piecewise-based Mechanism

- **Most generalized version:** m-piecewise distribution



$$pdf[\mathcal{M}(x) = y] = \begin{cases} p_{1,\varepsilon} & \text{if } y \in [l_{1,x,\varepsilon}, r_{1,x,\varepsilon}] \\ p_{2,\varepsilon} & \text{if } y \in [l_{2,x,\varepsilon}, r_{2,x,\varepsilon}] \\ & \dots \\ p_{m,\varepsilon} & \text{if } y \in [l_{m,x,\varepsilon}, r_{m,x,\varepsilon}] \end{cases}$$

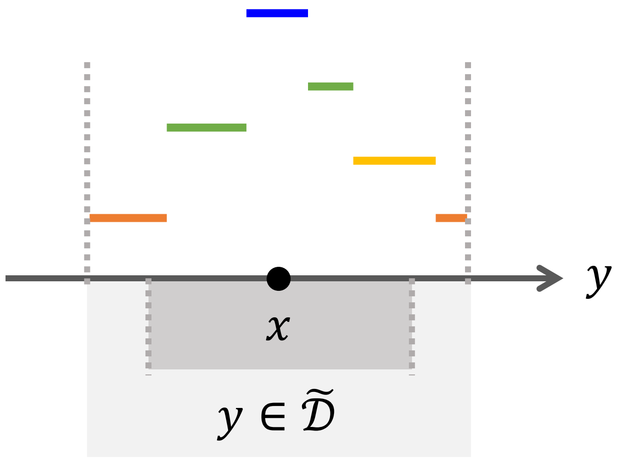
$\frac{p_{i,\varepsilon}}{p_{j,\varepsilon}} \leq e^{\varepsilon}$ (LDP constraint)

- Expected error:

$$\int_{\widetilde{\mathcal{D}}} \mathcal{L}(y, x) \cdot pdf[\mathcal{M}(x) = y] \mathrm{d}y$$

Optimal Piecewise-based Mechanism

- **Most generalized version:** m-piecewise distribution



$$pdf[\mathcal{M}(x) = y] = \begin{cases} p_{1,\varepsilon} & \text{if } y \in [l_{1,x,\varepsilon}, r_{1,x,\varepsilon}] \\ p_{2,\varepsilon} & \text{if } y \in [l_{2,x,\varepsilon}, r_{2,x,\varepsilon}] \\ & \quad \dots \\ p_{m,\varepsilon} & \text{if } y \in [l_{m,x,\varepsilon}, r_{m,x,\varepsilon}] \end{cases}$$

$$\frac{p_{i,\varepsilon}}{p_{j,\varepsilon}} \le e^{\varepsilon} \text{ (LDP constraint)}$$

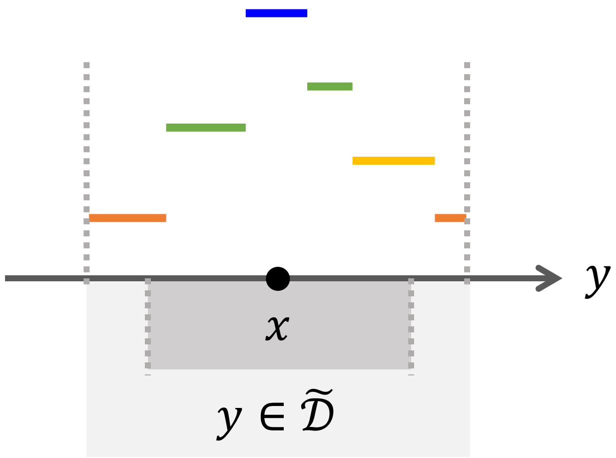
- Expected error:

$$\min_{\mathcal{M}:\, p_i, l_i, r_i} \int_{\widetilde{\mathcal{D}}} \mathcal{L}(y, x) \cdot pdf[\mathcal{M}(x) = y] \mathrm{d}y$$

Find $\mathcal{M}$ to minimize the error at x

Optimal Piecewise-based Mechanism

- **Most generalized version:** m-piecewise distribution



$$pdf[\mathcal{M}(x) = y] = \begin{cases} p_{1,\varepsilon} & \text{if } y \in [l_{1,x,\varepsilon}, r_{1,x,\varepsilon}] \\ p_{2,\varepsilon} & \text{if } y \in [l_{2,x,\varepsilon}, r_{2,x,\varepsilon}] \\ & \dots \\ p_{m,\varepsilon} & \text{if } y \in [l_{m,x,\varepsilon}, r_{m,x,\varepsilon}] \end{cases}$$

$\frac{p_{i,\varepsilon}}{p_{j,\varepsilon}} \leq e^{\varepsilon}$ (LDP constraint)

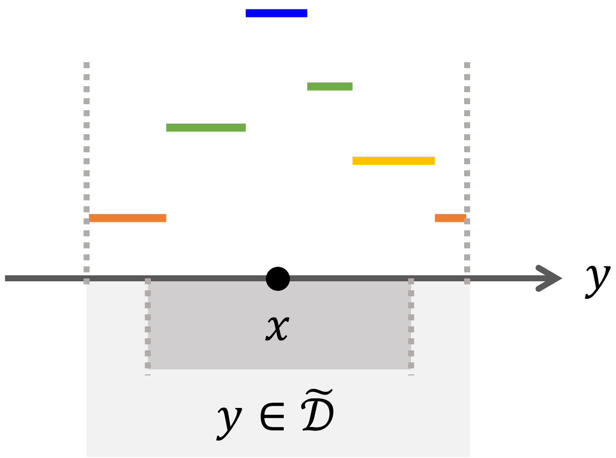
- Expected error:

$$\min_{\mathcal{M}:\, p_i, l_i, r_i} \max_{x \in \mathcal{D}} \int_{\widetilde{\mathcal{D}}} \mathcal{L}(y, x) \cdot pdf[\mathcal{M}(x) = y] \mathrm{d}y$$

Find $\mathcal{M}$ to minimize the worst-case error

Optimal Piecewise-based Mechanism

- **Most generalized version:** m-piecewise distribution



$$pdf[\mathcal{M}(x) = y] = \begin{cases} p_{1,\varepsilon} & \text{if } y \in [l_{1,x,\varepsilon}, r_{1,x,\varepsilon}] \\ p_{2,\varepsilon} & \text{if } y \in [l_{2,x,\varepsilon}, r_{2,x,\varepsilon}] \\ & \dots \\ p_{m,\varepsilon} & \text{if } y \in [l_{m,x,\varepsilon}, r_{m,x,\varepsilon}] \end{cases}$$

$\frac{p_{i,\varepsilon}}{p_{j,\varepsilon}} \leq e^{\varepsilon}$ (LDP constraint)

Solved $\mathcal{M}$ is
the optimal piecewise-based mechanism

Mathematically ≡ to find the **optimal piecewise distribution under the LDP constraint**

$$\min_{\mathcal{M}:\, p_i, l_i, r_i} \max_{x \in \mathcal{D}} \int_{\widetilde{\mathcal{D}}} \mathcal{L}(y, x) \cdot pdf[\mathcal{M}(x) = y] \mathrm{d}y$$

Find $\mathcal{M}$ to minimize the worst-case error

Challenges & Proofs

- Challenges
  1. min-max problem & multiple variables
  2. optimal results only for a specific m

$$\min_{\mathcal{M}:\, p_i, l_i, r_i} \max_{x \in \mathcal{D}} \int_{\widetilde{\mathcal{D}}} \mathcal{L}(y, x) \cdot pdf[\mathcal{M}(x) = y] \mathrm{d}y$$

2. $i \in [m]$

1. variables p_i, l_i, r_i

Challenges & Proofs

- Challenges
  1. min-max problem & multiple variables
  2. optimal results only for a specific m

$$\min_{\mathcal{M}:\, p_i, l_i, r_i} \max_{x \in \mathcal{D}} \int_{\tilde{\mathcal{D}}} \mathcal{L}(y, x) \cdot pdf[\mathcal{M}(x) = y] \mathrm{d}y$$

2. $i \in [m]$

1. variables p_i, l_i, r_i

1

Inner max
has a closed form

Worst-case from
$x =$ endpoints of $\mathcal{D}$

Reduced to:
minimization problem

Challenges & Proofs

- Challenges
  1. min-max problem & multiple variables
  2. optimal results only for a specific m

$$\min_{\mathcal{M}:\, p_i, l_i, r_i} \max_{x \in \mathcal{D}} \int_{\widetilde{\mathcal{D}}} \mathcal{L}(y, x) \cdot pdf[\mathcal{M}(x) = y] \mathrm{d}y$$

2. $i \in [m]$ 1. variables p_i, l_i, r_i

1

Inner max has a closed form

Worst-case from x = endpoints of $\mathcal{D}$

Reduced to:
minimization problem

2

The optimal #piece is m

If optimal m-piecewise = $(m+1)$-piecewise

Optimal m:
Check $m = 1, 2, \ldots$ until optimal

"Solvable"

"Solvable"

- **When $m \leq 3$: Analytical solvable → closed form $\mathcal{M}$**
- **When $m \geq 4$: Too many variables & non-linear**

$$\max_{x \in \{a,b\}} \sum_{i=1}^{m} p_i \int_{l_i}^{r_i} \mathcal{L}(y, x)\, \mathrm{d}y$$

- efficiently solved by off-the-shelf solvers, e.g. Gurobi

- limitation: needs given ε

- limitation: cannot provide closed-form $\mathcal{M}$: p_i, l_i, r_i (only optimal values)

- enough to analyze optimality

"Solvable"

- **When $m \leq 3$: Analytical solvable → closed form $\mathcal{M}$**
- **When $m \geq 4$: Too many variables & non-linear**

$$\max_{x \in \{a,b\}} \sum_{i=1}^{m} p_i \int_{l_i}^{r_i} \mathcal{L}(y,x)\, \mathrm{d}y$$

- efficiently solved by off-the-shelf solvers, e.g. Gurobi

- limitation: needs given ε

- limitation: cannot provide closed-form $\mathcal{M}$: p_i, l_i, r_i (only optimal values)

- enough to analyze optimality

Monte Carlo testing:
Optimality under 10^4 random ε

Hypothesis. For any domain $\mathcal{D} \rightarrow \mathcal{D}$, under error metrics $\mathcal{L}(y,x) \coloneqq |y-x|$ and $\mathcal{L}(y,x) \coloneqq (y-x)^2$, the optimal piecewise-based mechanism falls into 3-piecewise mechanism

different from existing instantiations

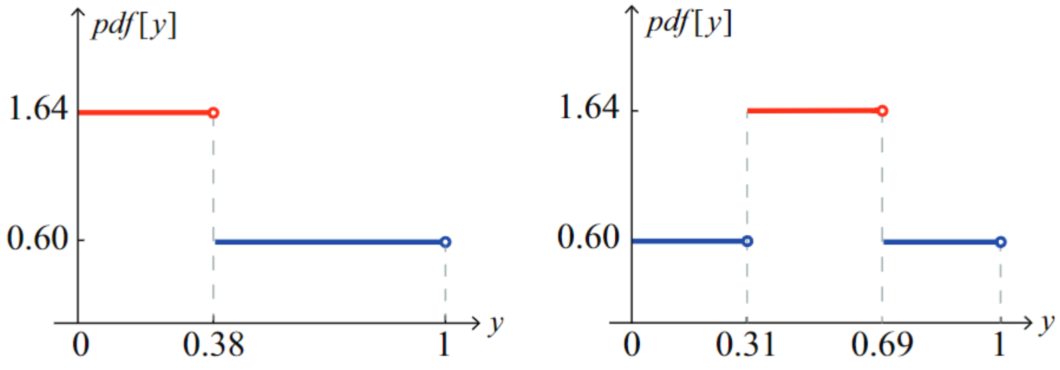
Optimal Closed-Form Mechanism

- Optimal $\mathcal{M}: [0,1) \to [0,1)$ under $\mathcal{L} \coloneqq |y - x|$

$$pdf[\mathcal{M}(x) = y] = \begin{cases} \exp\left(\frac{\varepsilon}{2}\right) & \text{if } y \in [l_{x,\varepsilon}, r_{x,\varepsilon}) \\ \exp\left(-\frac{\varepsilon}{2}\right) & \text{if } y \in [0,1) \backslash [l_{x,\varepsilon}, r_{x,\varepsilon}) \end{cases}$$

$$[l_{x,\varepsilon}, r_{x,\varepsilon}) = \begin{cases} [0,2C) & \text{if } x \in [0, C) \\ x + [-C, C) & \text{if } x \in [C, 1 - C) \\ [1 - 2C, 1) & \text{otherwise} \end{cases} \qquad C = \frac{\exp(\varepsilon/2) - 1}{2(\exp(\varepsilon) - 1)}$$

- When $\varepsilon = 1$:

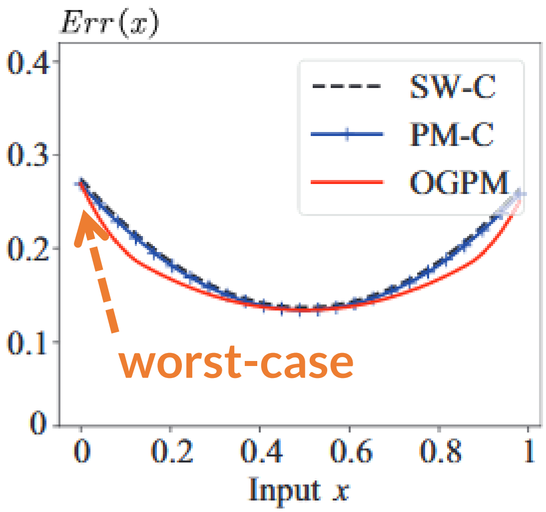


(a) $x = 0$. **(b)** $x = 0.5$.

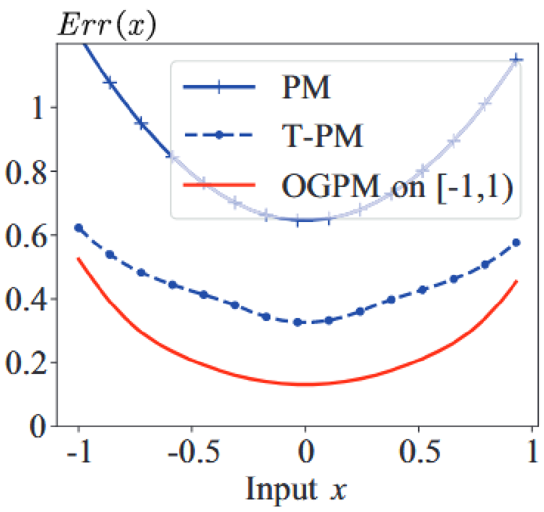
Comparison of Expected Errors

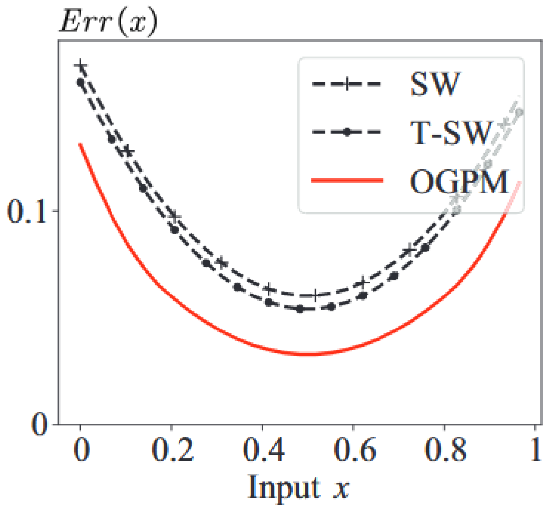
Whole-domain error (i.e. each point in $\mathcal{D}$) ($\varepsilon = 2$)

Compressed PM, SW



Original and **truncated** PM, SW

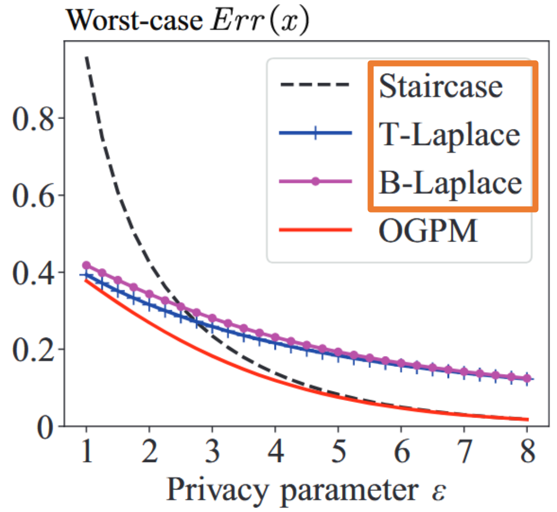




Lowest error

Worst-case error

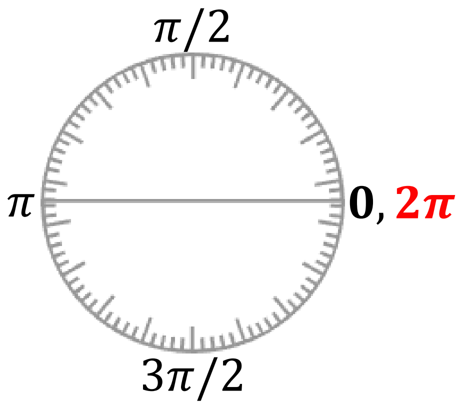
Non-piecewise-based



Lowest error

Circular Domain

- Different meaning of distance, e.g. $\mathrm{distance}(0, 2\pi) = 0$



$\mathcal{L} \rightarrow \mathcal{L}_{\mathrm{mod}}$

$\mathcal{L}_{\mathrm{mod}}(y, x) \coloneqq \min(\mathcal{L}(y, x), \mathcal{L}(y, 2\pi - x))$

Circular Domain

- Different meaning of distance, e.g. $\text{distance}(0, 2\pi) = 0$

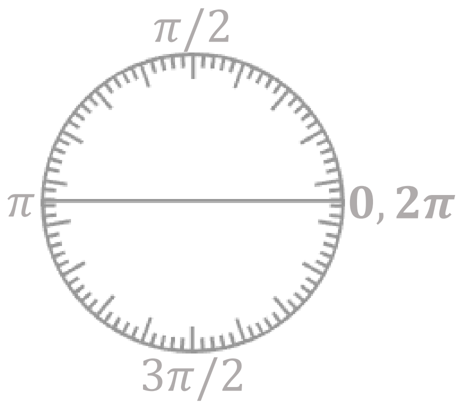
$$\mathcal{L} \to \mathcal{L}_{\text{mod}}$$

$$\mathcal{L}_{\text{mod}}(y, x) \coloneqq \min(\mathcal{L}(y, x), \mathcal{L}(y, 2\pi - x))$$

$$\min_{\mathcal{M}:\, p_i, l_i, r_i} \max_{x \in [0, 2\pi]} \int_{\widetilde{\mathcal{D}}} \mathcal{L}_{\text{mod}}(y, x) \cdot pdf[\mathcal{M}(x) = y] \mathrm{d}y$$

Circular Domain

- Different meaning of distance, e.g. distance$(0, 2\pi) = 0$



$$\mathcal{L} \rightarrow \mathcal{L}_{\mathrm{mod}}$$

$$\mathcal{L}_{\mathrm{mod}}(y, x) \coloneqq \min(\mathcal{L}(y, x), \mathcal{L}(y, 2\pi - x))$$

$$\min_{\mathcal{M}:\, p_i, l_i, r_i} \max_{x \in [0, 2\pi]} \int_{\widetilde{\mathcal{D}}} \mathcal{L}_{\mathrm{mod}}(y, x) \cdot pdf[\mathcal{M}(x) = y] \mathrm{d}y$$

- **Linking** to problems in the classical domain

min-max under $\mathcal{L}_{\mathrm{mod}}$	=	min under $\mathcal{L}$ at π		min under $\mathcal{L}_{\mathrm{mod}}$ at x	=	min under $\mathcal{L}$ at π + Transform

p_i $\qquad$ $l_{i,x}^{\mathrm{mod}}, r_{i,x}^{\mathrm{mod}}$

Optimal Closed-Form Mechanism

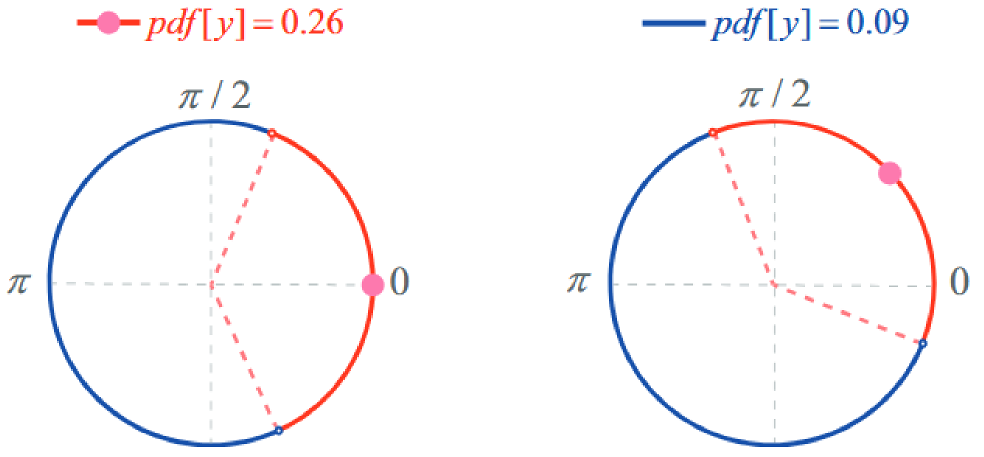
- Optimal $\mathcal{M}: [0,2\pi) \to [0,2\pi)$ under $\mathcal{L}_{\text{mod}}$

$$pdf[\mathcal{M}(x) = y] = \begin{cases} \frac{1}{2\pi}\exp\left(\frac{\varepsilon}{2}\right) & \text{if } y \in [l_{x,\varepsilon}^{\text{mod}}, r_{x,\varepsilon}^{\text{mod}}) \\ \frac{1}{2\pi}\exp\left(-\frac{\varepsilon}{2}\right) & \text{if } y \in [0,2\pi)\backslash[l_{x,\varepsilon}^{\text{mod}}, r_{x,\varepsilon}^{\text{mod}}) \end{cases}$$

$$[l_{x,\varepsilon}^{\text{mod}}, r_{x,\varepsilon}^{\text{mod}}) = [x - C, x + C) \mod 2\pi$$

$$C = \pi \frac{\exp(\varepsilon/2) - 1}{\exp(\varepsilon) - 1}$$

- When $\varepsilon = 1$:

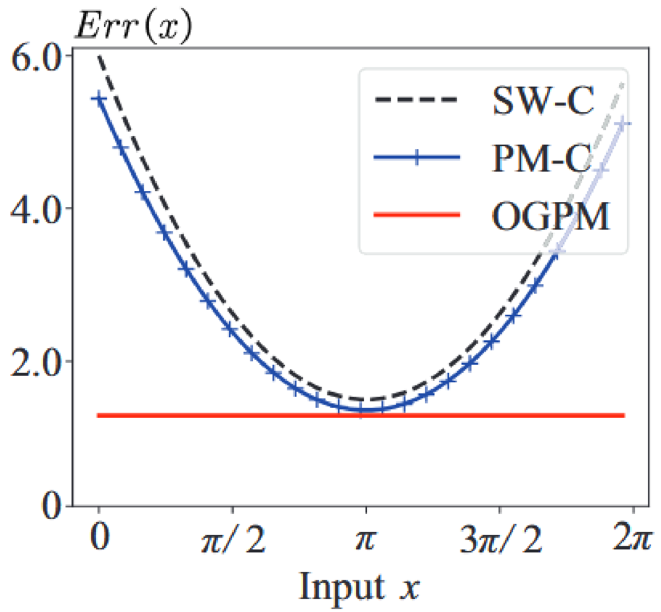


(a) $x = 0$. **(b)** $x = \pi/4$.

Comparison of Expected Errors

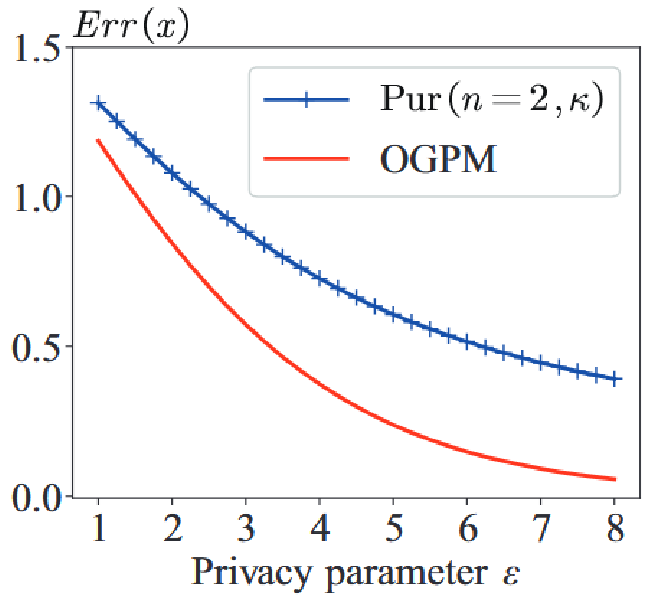
Whole-domain error ($\varepsilon = 2$)

PM, SW on the **flattened** domain



Worst-case error

Purkayastha mechanism [CCS'21]*



Lowest error

* Differential Privacy for Directional Data, *CCS'21*

Summary

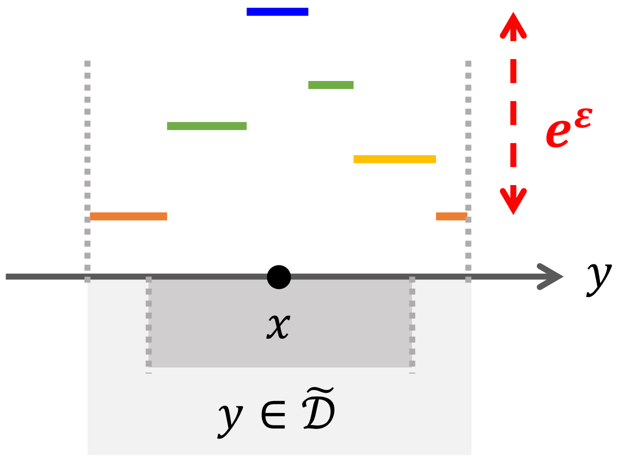
- **RQ:** What is the optimal piecewise-based mechanism?

- **Contributions**

  - solving framework for the optimality

  - closed-form mechanisms for the classical domain & circular domain

  - comparison with non-piecewise-based mechanisms



Optimal Piecewise-based Mechanism for Collecting Bounded Numerical Data under LDP

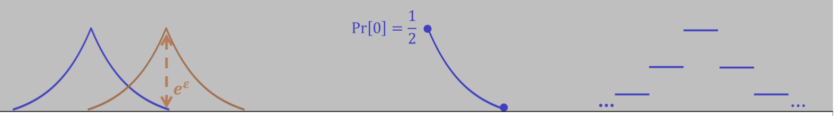
LDP Mechanisms for Numerical Domain

Laplace mechanism: $[0,1] \to (-\infty, +\infty)$

Laplace + truncation: $[0,1] \to [0,1]$

Staircase: $[0,1] \to (-\infty, +\infty)$

$\Pr[0] = \frac{1}{2}$

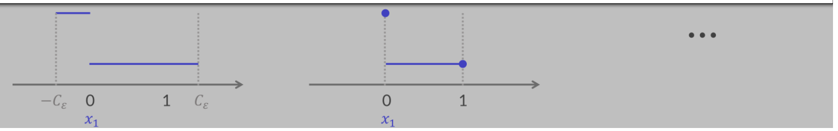


Privacy: LDP with the **same** ε

Utility: Different errors

Q: What is the **optimal piecewise-based** mechanism?

?



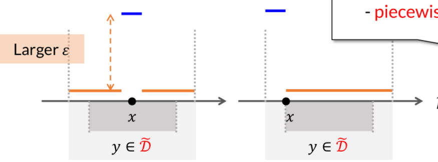
3-Piecewise Mechanism

- 3-piecewise distributions on boun[…]
  - given input x, sample output y fr[…]

NOT enough to study optimality of piecewise-based mechanism

- only 3 pieces, two probabilities
- piecewise distribution can have more pieces and probabilities

Larger ε

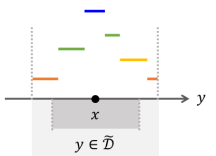


$$pdf[\mathcal{M}(x) = y] = \begin{cases} p_\varepsilon & \text{if } y \in [l_{x,\varepsilon}, r_{x,\varepsilon}] \\ \frac{p_\varepsilon}{\exp(\varepsilon)} & \text{if } y \in \tilde{\mathcal{D}} \setminus [l_{x,\varepsilon}, r_{x,\varepsilon}] \end{cases}$$

- Instantiations: PM [ICDE'19], SW [SIGMOD'20], PTT [TMC'24] (design different $p_\varepsilon, l_{x,\varepsilon}, r_{x,\varepsilon}$)
  - different errors, but **without optimality**

Optimal Piecewise-based Mechanism

- **Most generalized version:** m-piecewise distribution



$$pdf[\mathcal{M}(x) = y] = \begin{cases} p_{1,\varepsilon} & \text{if } y \in [l_{1,x,\varepsilon}, r_{1,x,\varepsilon}] \\ p_{2,\varepsilon} & \text{if } y \in [l_{2,x,\varepsilon}, r_{2,x,\varepsilon}] \\ \dots \\ p_{m,\varepsilon} & \text{if } y \in [l_{m,x,\varepsilon}, r_{m,x,\varepsilon}] \end{cases}$$

$\frac{p_{i,\varepsilon}}{p_{j,\varepsilon}} \le e^\varepsilon$ (LDP constraint)

Solved $\mathcal{M}$ is
the optimal piecewise-based mechanism

Mathematically $\equiv$ to find the optimal
piecewise distribution under the LDP constraint

$$\min_{\mathcal{M}: p_i, l_i, r_i} \max_{x \in \mathcal{D}} \int_{\tilde{\mathcal{D}}} \mathcal{L}(y, x) \cdot pdf[\mathcal{M}(x) = y] dy$$

Find $\mathcal{M}$ to minimize the worst-case error

Challenges & Proofs

- Challenges
  1. min-max problem & multiple variables
  2. optimal results only for a specific m

$$\min_{\mathcal{M}: p_i, l_i, r_i} \max_{x \in \mathcal{D}} \int_{\tilde{\mathcal{D}}} \mathcal{L}(y, x) \cdot pdf[\mathcal{M}(x) = y] dy$$

2. $i \in [m]$ 1. variables p_i, l_i, r_i

1 Inner max has a closed form
Worst-case from $x =$ endpoints of $\mathcal{D}$
Reduced to: minimization problem

2 The optimal #piece is m
If optimal m-piecewise = $(m+1)$-piecewise
Optimal m: Check $m = 1, 2, \dots$ until optimal

"Solvable"

Manually (Analytically) Solvable When $m = 3$

- When $m \ge 4$: Too many variables & non-linear
- Efficiently solved by off-the-shelf solvers, e.g. Gurobi
  - limitation: needs given ε
  - limitation: cannot provide closed-form $\mathcal{M}: p_i, l_i, r_i$
  - can be used to analyze optimality

$$\max_{x \in \{a,b\}} \sum_{i=1}^{m} p_i \int_{l_i}^{r_i} \mathcal{L}(y, x)\, dy$$

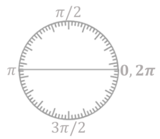
Monte Carlo testing:
Optimality under 10^4 random ε

Hypothesis. For any domain $\mathcal{D} \to \tilde{\mathcal{D}}$, under error metrics $\mathcal{L}(y,x) := |y - x|$ and $\mathcal{L}(y,x) := (y-x)^2$, the optimal piecewise-based mechanism falls into 3-piecewise mechanism.

different from existing instantiations
(closed-form $\mathcal{M}$ can be manually solved)

Circular Domain

- Different meaning of distance, e.g. distance$(0, 2\pi) = 0$



$\mathcal{L} \to \mathcal{L}_{\text{mod}}$

$\mathcal{L}_{\text{mod}}(y,x) := \min(\mathcal{L}(y,x), \mathcal{L}(y, 2\pi - x))$

$$\min_{\mathcal{M}: p_i, l_i, r_i} \max_{x \in [0, 2\pi]} \int_{\tilde{\mathcal{D}}} \mathcal{L}_{\text{mod}}(y, x) \cdot pdf[\mathcal{M}(x) = y] dy$$

- **Linking** to problems in the classical domain

min-max under $\mathcal{L}_{\text{mod}}$ = min under $\mathcal{L}$ at π — p_i

min under $\mathcal{L}_{\text{mod}}$ at x = min under $\mathcal{L}$ at π + Transform — $l_{i,x}^{\text{mod}}, r_{i,x}^{\text{mod}}$

Thank you!

Q & A

Optimality of LDP Mechanisms

- Optimal error (utility) under privacy level ε

  - many mechanisms are optimal in **order-of-magnitude**, e.g. $\Omega(\frac{1}{\sqrt{n}})$ for the counting query*

  - the staircase mechanism is optimal for **domain** $[0,1] \rightarrow (-\infty, +\infty)$†

  - the geometric mechanism is universally optimal if any **post-processing** is allowed, e.g. truncation††

* The Complexity of Differential Privacy, *book section of "Tutorials on the Foundations of Cryptography", 2017*

† The Staircase Mechanism in Differential Privacy, *journal version of ISIT'14*

†† Universally Utility-maximizing Privacy Mechanisms, *STOC'09*

Optimality of LDP Mechanisms

- Optimal error (utility) under privacy level ε

  - many mechanisms are optimal in **order-of-magnitude**, e.g. $\Omega(\frac{1}{\sqrt{n}})$ for the counting query*

  - the staircase mechanism is optimal for **domain** $[0,1] \rightarrow (-\infty, +\infty)^{\dagger}$

  - the geometric mechanism is universally optimal if any **post-processing** is allowed, e.g. truncation††

- **Specify the utility model** (conditions for optimality)

1 Error metric

$Err(\text{truth}, \text{rand})$

Err or $\boldsymbol{\Omega}(Err)$

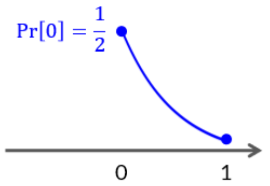
2 Data domain & type of mechanisms

Discrete / cont. $\mathcal{D} \rightarrow \widetilde{\mathcal{D}}$

Laplace-shape / piecewise

3 Post-processing

Laplace + truncation: $[0,1] \rightarrow [0,1]$



Proof Intuitions

- Worst-case error is achieved at endpoints

$$\max_{x\in\mathcal{D}} \int_{\widetilde{\mathcal{D}}} \mathcal{L}(y,x)\cdot pdf[\mathcal{M}(x)=y]\mathrm{d}y = \max_{x\in\mathcal{D}} \sum_{i=1}^{m} p_i \int_{l_i}^{r_i} \mathcal{L}(y,x)\,\mathrm{d}y \qquad (m\text{-piecewise distribution})$$

Proof Intuitions

- Worst-case error is achieved at endpoints

$$\max_{x\in\mathcal{D}} \int_{\widetilde{\mathcal{D}}} \mathcal{L}(y,x)\cdot pdf[\mathcal{M}(x)=y]\mathrm{d}y = \max_{x\in\mathcal{D}} \sum_{i=1}^{m} p_i \int_{l_i}^{r_i} \mathcal{L}(y,x)\,\mathrm{d}y \qquad (m\text{-piecewise distribution})$$

convex function w.r.t x

Proof Intuitions

- Worst-case error is achieved at endpoints

$$\max_{x\in\mathcal{D}} \int_{\widetilde{\mathcal{D}}} \mathcal{L}(y,x)\cdot pdf[\mathcal{M}(x)=y]\mathrm{d}y = \max_{x\in\mathcal{D}} \sum_{i=1}^{m} p_i \int_{l_i}^{r_i} \mathcal{L}(y,x)\,\mathrm{d}y \qquad (m\text{-piecewise distribution})$$

$$= \max_{x\in\{a,b\}} \sum_{i=1}^{m} p_i \int_{l_i}^{r_i} \mathcal{L}(y,x)\,\mathrm{d}y \qquad (\text{maximum principle})$$

Proof Intuitions

- Worst-case error is achieved at endpoints

$$\max_{x \in \mathcal{D}} \int_{\widetilde{\mathcal{D}}} \mathcal{L}(y,x) \cdot pdf[\mathcal{M}(x)] = \sum^{m} \int^{r_i}$$

After merging redundant pieces

- Optimal #piece is m **if** optimal m-piecewise = (m+1)-piecewise

Error from an arbitrary piece

(≥ 0 variable)

if: $$\min_{e_1,e_2,e_3} e_1 + e_2 + e_3 = \min_{e_1,e_2,e_3,e_4} e_1 + e_2 + e_3 + e_4$$

i.e. the error can't be lowered by arbitrary ≥ 0 variable

Proof Intuitions

- Worst-case error is achieved at endpoints

$$\max_{x\in\mathcal{D}} \int_{\tilde{\mathcal{D}}} \mathcal{L}(y,x)\cdot pdf[\mathcal{M}(x) =$$

After merging redundant pieces

- Optimal #piece is m **if** optimal m-piecewise = $(m+1)$-piecewise

Error from an arbitrary piece (≥ 0 variable)

if: $\min_{e_1,e_2,e_3} e_1+e_2+e_3 = \min_{e_1,e_2,e_3,e_4} e_1+e_2+e_3+e_4$

i.e. the error can't be lowered by arbitrary ≥ 0 variable

then: $= \min_{e_1,e_2,e_3,e_4,e_5} e_1+e_2+e_3+e_4+e_5$

otherwise, $e_4 \leftarrow e_4 + e_5$ can further lower the error