



Wetenschappelijk Onderzoek- en
Documentatiecentrum
Ministerie van Justitie en Veiligheid



An introduction to ϵ -differential privacy

Workshop at Surfnet

1 June 2023

Afternoon session

Mortaza S. Bargh

Outline

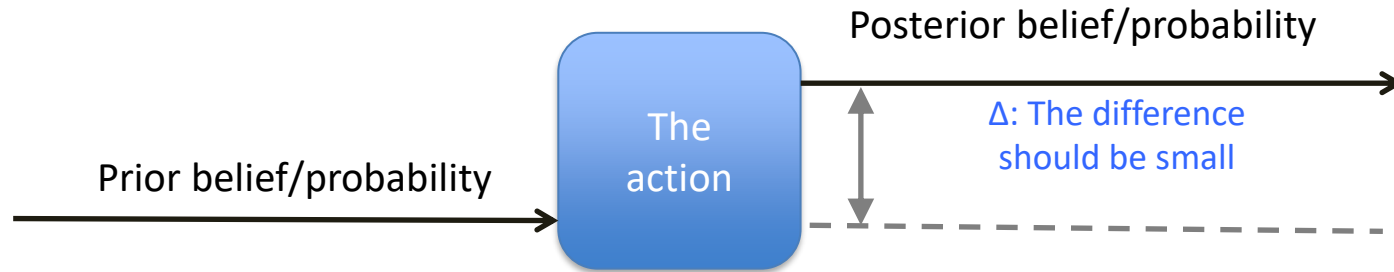
- ① Introduction
- ② ϵ -differential privacy: Interactive
- ③ Some exercises
- ④ ϵ -differential privacy: Non-interactive
- ⑤ Takeaways
- ⑥ References

1. Introduction

Introduction, ϵ -differential privacy – interactive, Some exercises, ϵ -differential privacy – non-interactive, Other relevant topics, Takeaways, References

Uninformative principle

- With respect to an information disclosure action



- What is the action?
 - Publishing a table with some info about person X
 - Including person X's record in a published table

Two paradigms

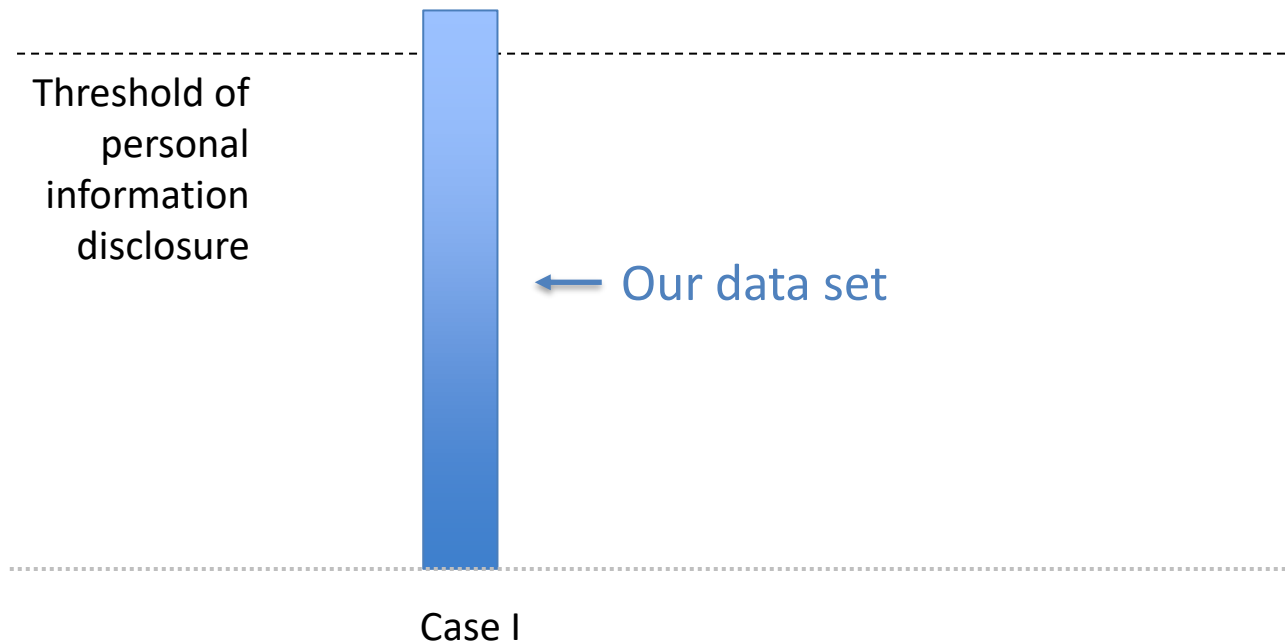
- Paradigm I (this morning)
 - To compare
 - Prior probability before accessing the dataset
 - Posterior probability after accessing the dataset
 - Considering background knowledge, **data extrinsic**
- Paradigm II (this afternoon)
 - For every data record, to compare the probability of
 - With the record (i.e., the subject's data) in the dataset
 - Without the record (i.e., the subject's_data) in the dataset
 - **No notion of background knowledge, data intrinsic**

Evolution of privacy definition

- From **normative** definition to **formal** definition
- **Normative notion of privacy** (paradigm I)
 - Underlying many privacy regulations (e.g., GDPR)
- **Example**
 - Our dataset contains personal data if it can reveal personal information when it is **combined with other datasets**

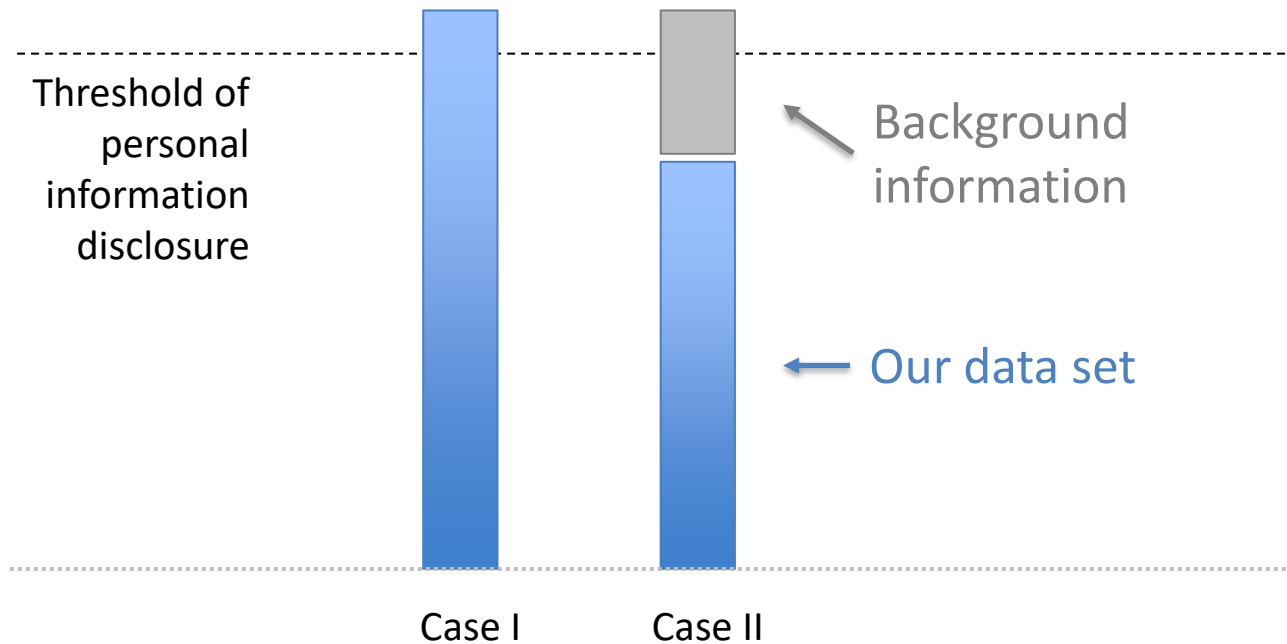
Evolution of privacy definition

- From normative definition to formal definition
- **Normative notion of privacy** (paradigm I)



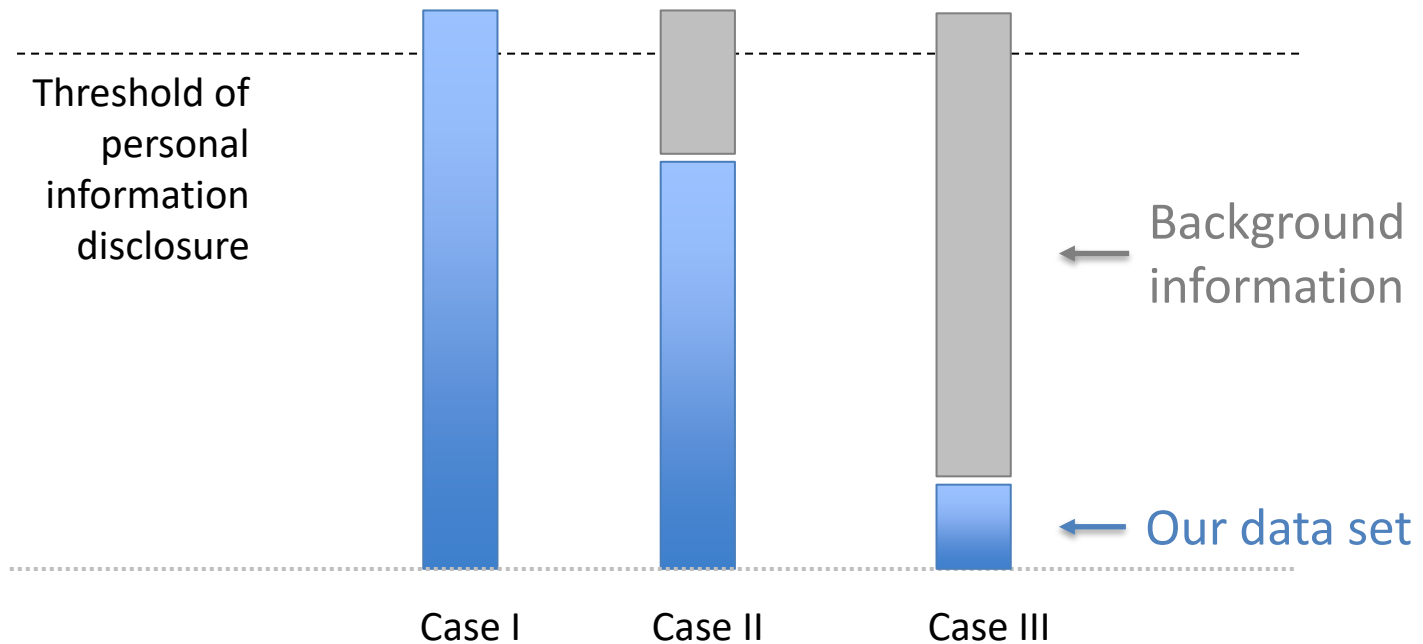
Evolution of privacy definition

- From normative definition to formal definition
- **Normative notion of privacy (paradigm I)**



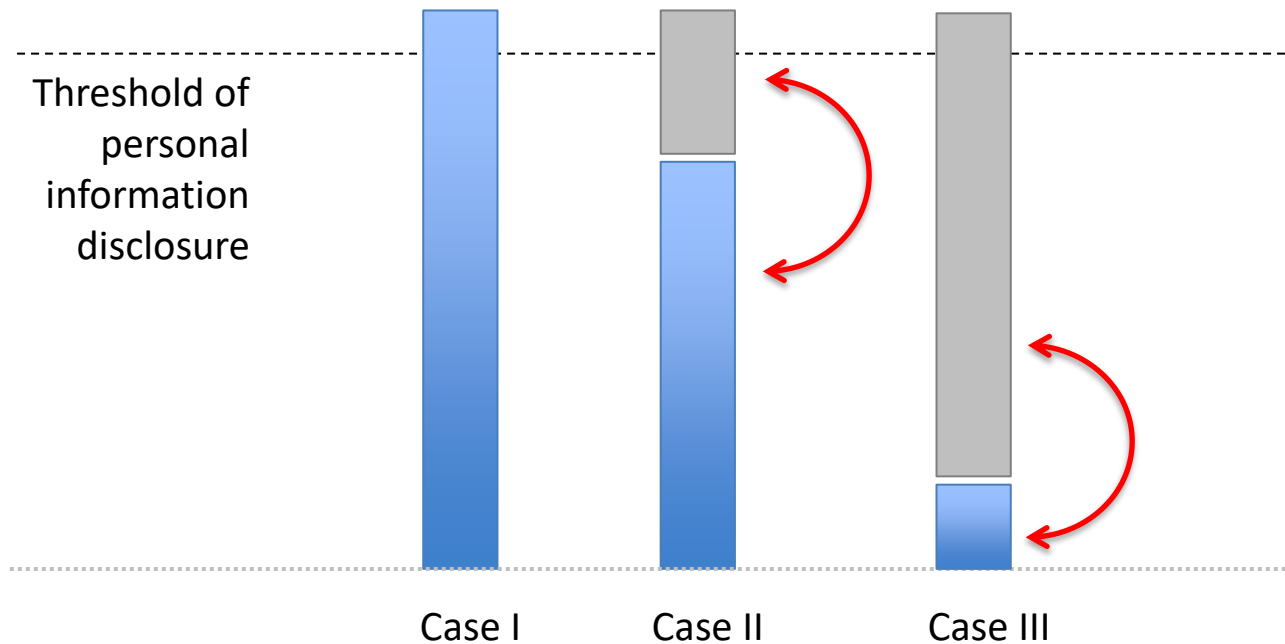
Evolution of privacy definition

- From normative definition to formal definition
- **Normative notion of privacy (paradigm I)**

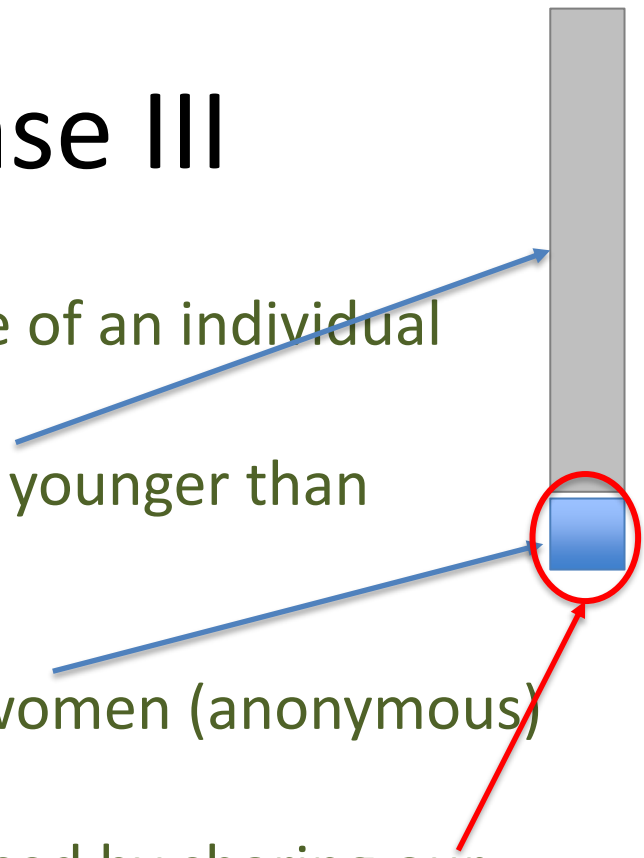


Evolution of privacy definition

- From normative definition to formal definition
- **Normative notion of privacy** (paradigm I)



An example for case III

- **Sensitive personal information:** The age of an individual
 - **Background knowledge:** Alice is 5 years younger than average American women
 - **Our data set:** The ages of all American women (anonymous)
 - **Question:** Is Alice's privacy is compromised by sharing our data set?
 - What if Alice is not American (i.e., Alice is not in data set D)
- 
- A vertical grey bar is positioned on the right side of the slide. At the bottom of this bar is a small blue square. This blue square is circled with a red circle. A blue arrow points from the text 'The age of an individual' to the top of the grey bar. Another blue arrow points from the text 'Alice is 5 years younger than average American women' to the blue square. A red arrow points from the text 'Is Alice's privacy is compromised by sharing our data set?' to the blue square.

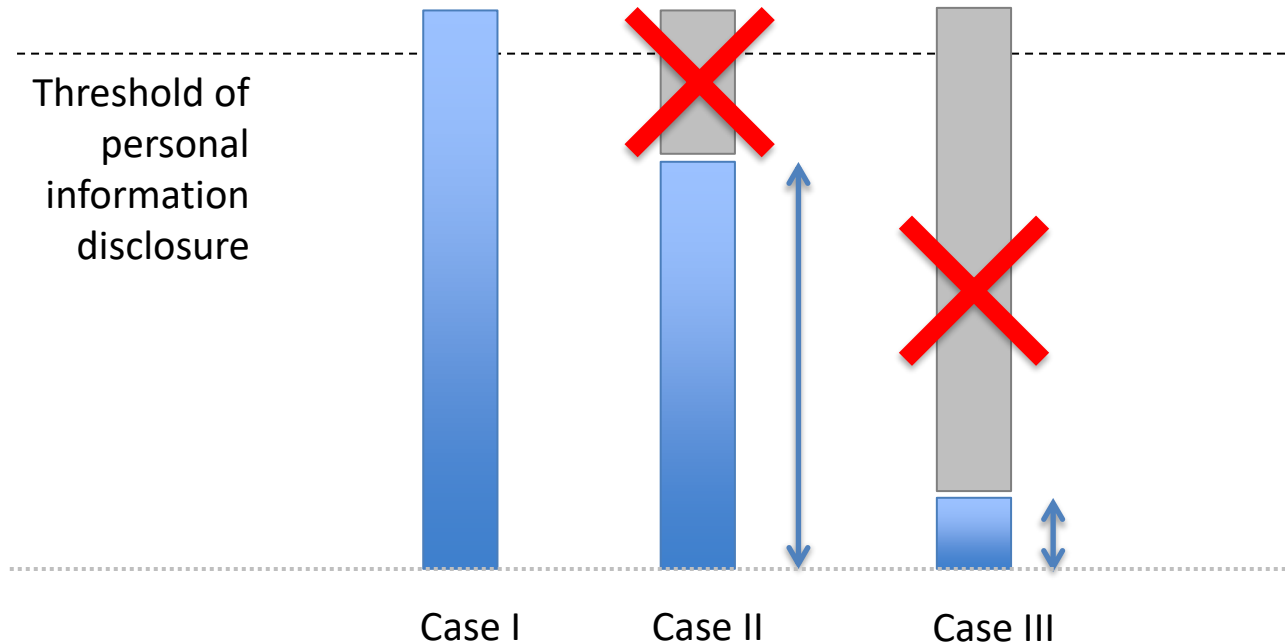
Evolution of privacy definition

- From normative definition to **formal** definition
- **Formal notion of privacy** (paradigm II)
 - Dwork et al. (2006) differential privacy
 - The **presence or absence of the data of an individual** in a dataset must not have an **observable impact** on the output of a computation over the data set
 - Already in use by Google, Apple, Uber, and the U.S. Census Bureau

Nice papers to read: (Nessim et al., 2018; 2019)

Evolution of privacy definition

- From **normative** definition to **formal** definition
- **Formal notion of privacy** (paradigm II)



Focusing on paradigm II

- Paradigm I: Normative (this morning)
 - To compare
 - Prior probability before accessing the dataset
 - Posterior probability after accessing the dataset
 - Considering background knowledge, **data extrinsic**
- Paradigm II: Formal (this afternoon)
 - For every data record, to compare the probability of
 - With the record (i.e., the subject's data) in the dataset
 - Without the record (i.e., the subject's_data) in the dataset
 - No notion of background knowledge, **data intrinsic**

Paradigm II: Formal

Data publication cases (in the following)

- Interactive
 - Reply to (multiple) queries
 - Statistical databases
- Non-interactive
 - Microdata: datasets about individuals

2. ϵ -differential privacy – interactive

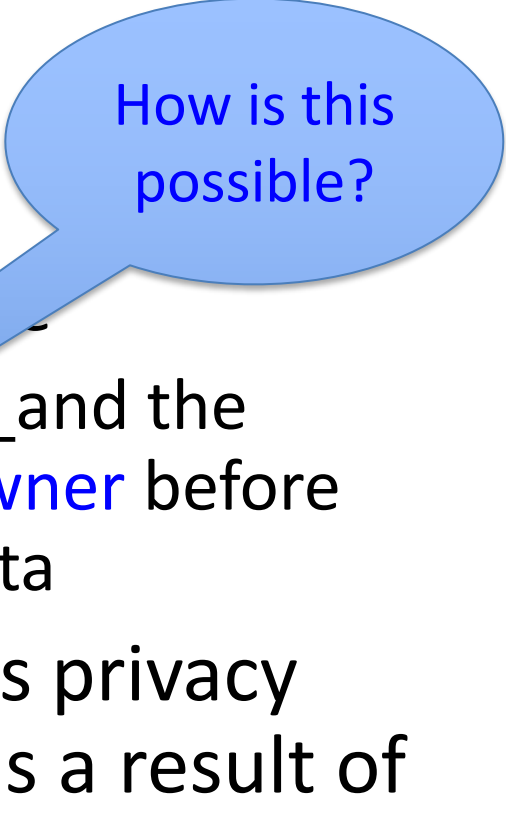
Introduction, **ϵ -differential privacy – interactive**, Some exercises, ϵ -differential privacy – non-interactive, Other relevant topics, Takeaways, References

Interactive data publishing

- Example queries (operations)
 - Mean, median, variance
 - Counts and marginal totals (# of people with glasses, etc.)
 - Correlation, regression coefficients
 - Histogram
 - A table derived from the microdata
- Goal: Achieving the **uninformative principle** with respect to an operation

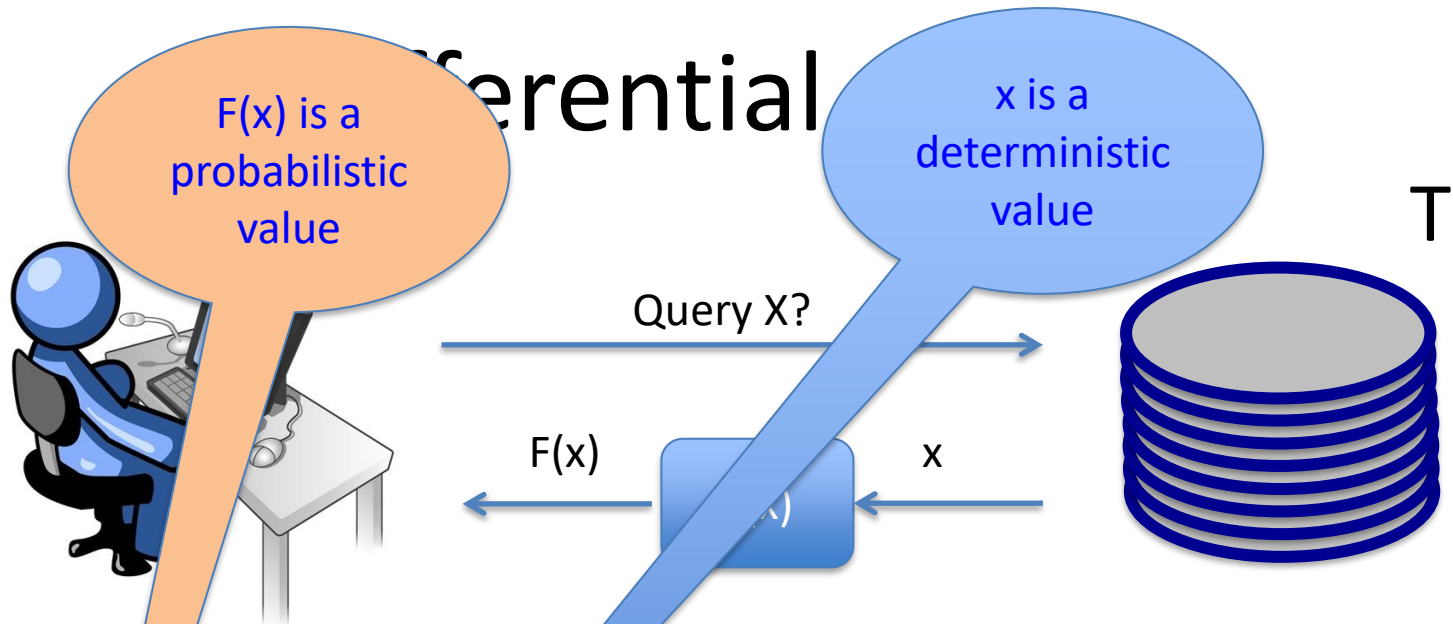
ϵ -Differential privacy

- Proposed by: Dwork [DWO'06]
- Motivation
 - To achieve the uninformative principle
 - Not to compare the prior probability and the posterior probability **about a data owner** before and after accessing the published data
- Dwork's: Risk to the record owner's privacy should not substantially increase as a result of participating in a statistical database



How is this possible?

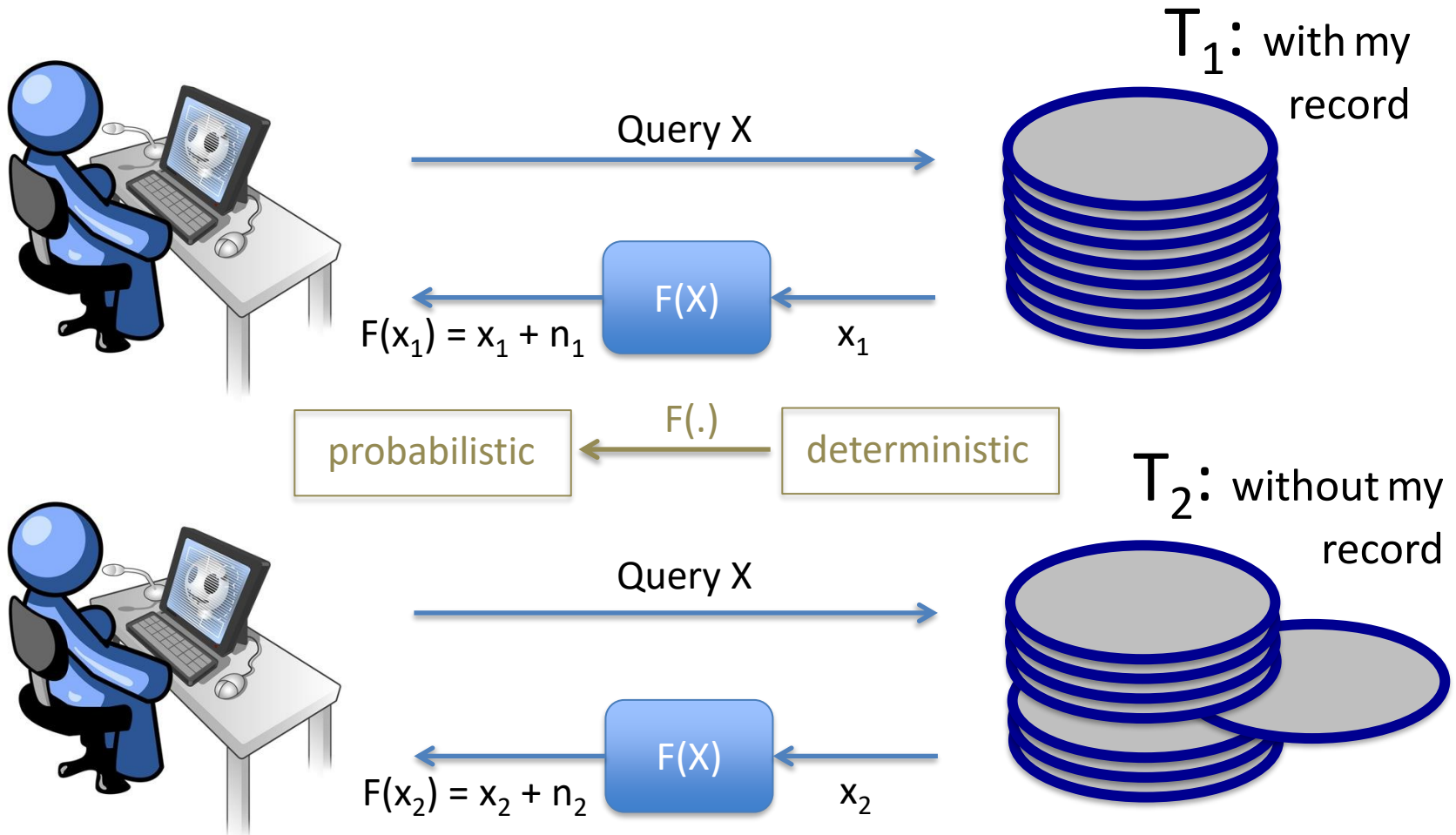
Differential



- Examples of query X: what is
 - Mean, median, variance
 - Counts and marginal totals
- Examples of Function $F(.)$
 - X = number of records with a specific property (like # of employees > 50 yrs.; # of people with glasses and gray hair)
 - $F(x) = x + n(\epsilon)$ where $n(\epsilon)$ is Laplace noise (a function of ϵ and sensitivity of X)

ϵ -Differential privacy

X: What is the number of people with glasses and gray hair?



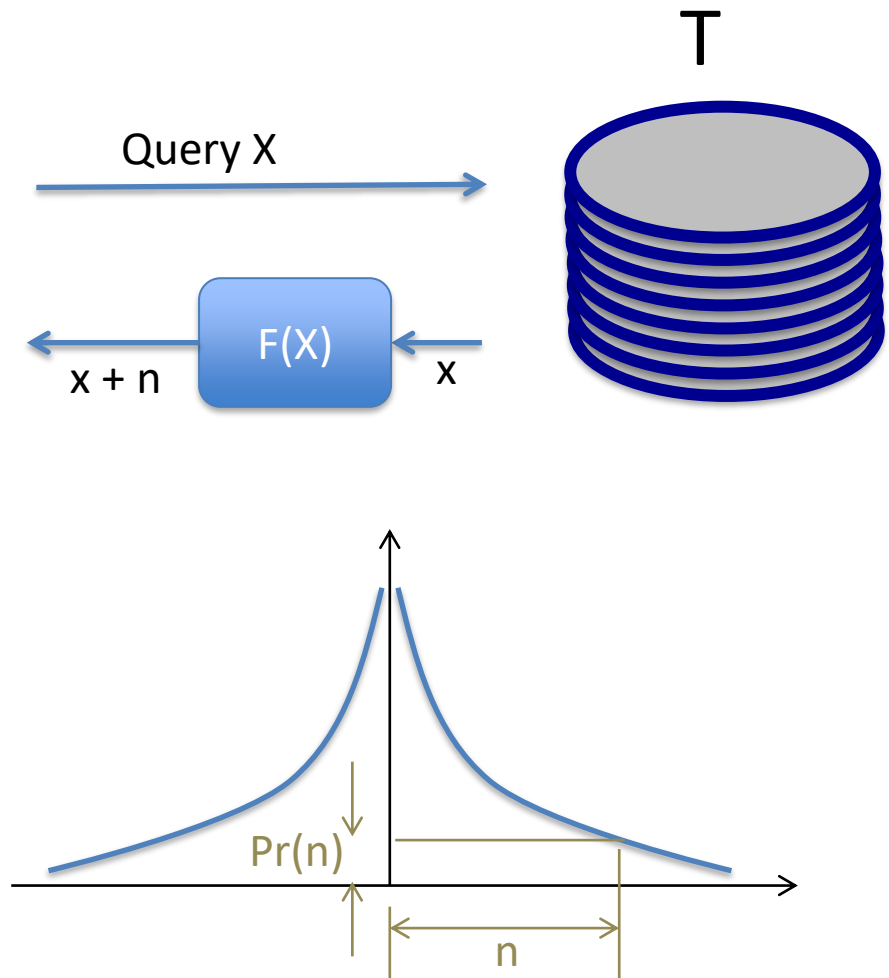
ϵ -Differential privacy: Adding noise

What does F add to the reply?

Laplace noise: Magnitude of n is determined with Laplace PDF (Probability Distribution Function)

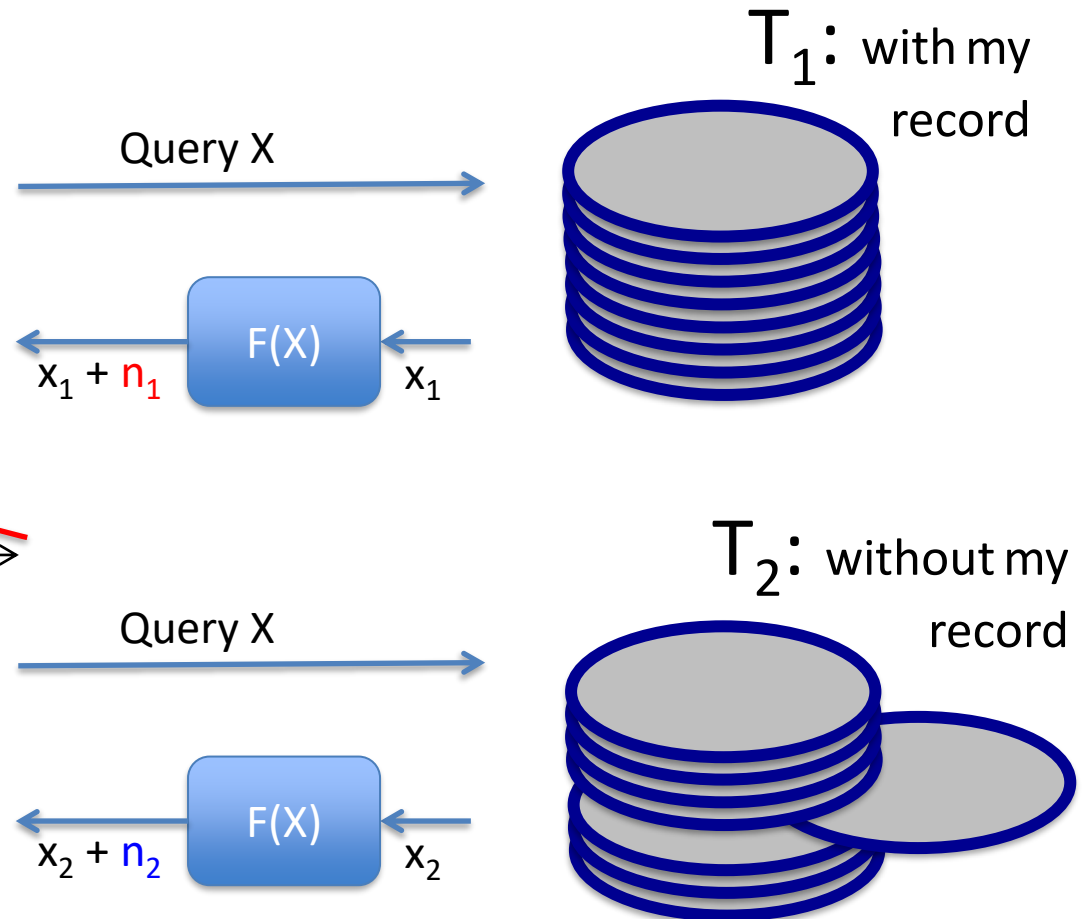
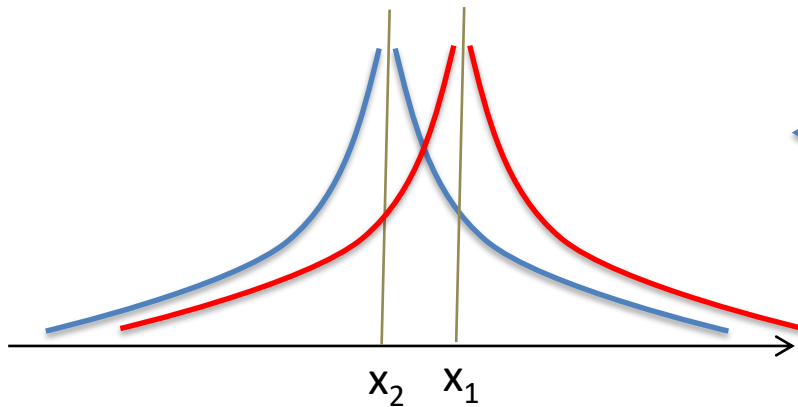
$$\Pr(\text{noise} = n) = \frac{1}{2\lambda} \exp(-|n|/\lambda)$$

$$\lambda = \text{Sen}(X) / \epsilon$$

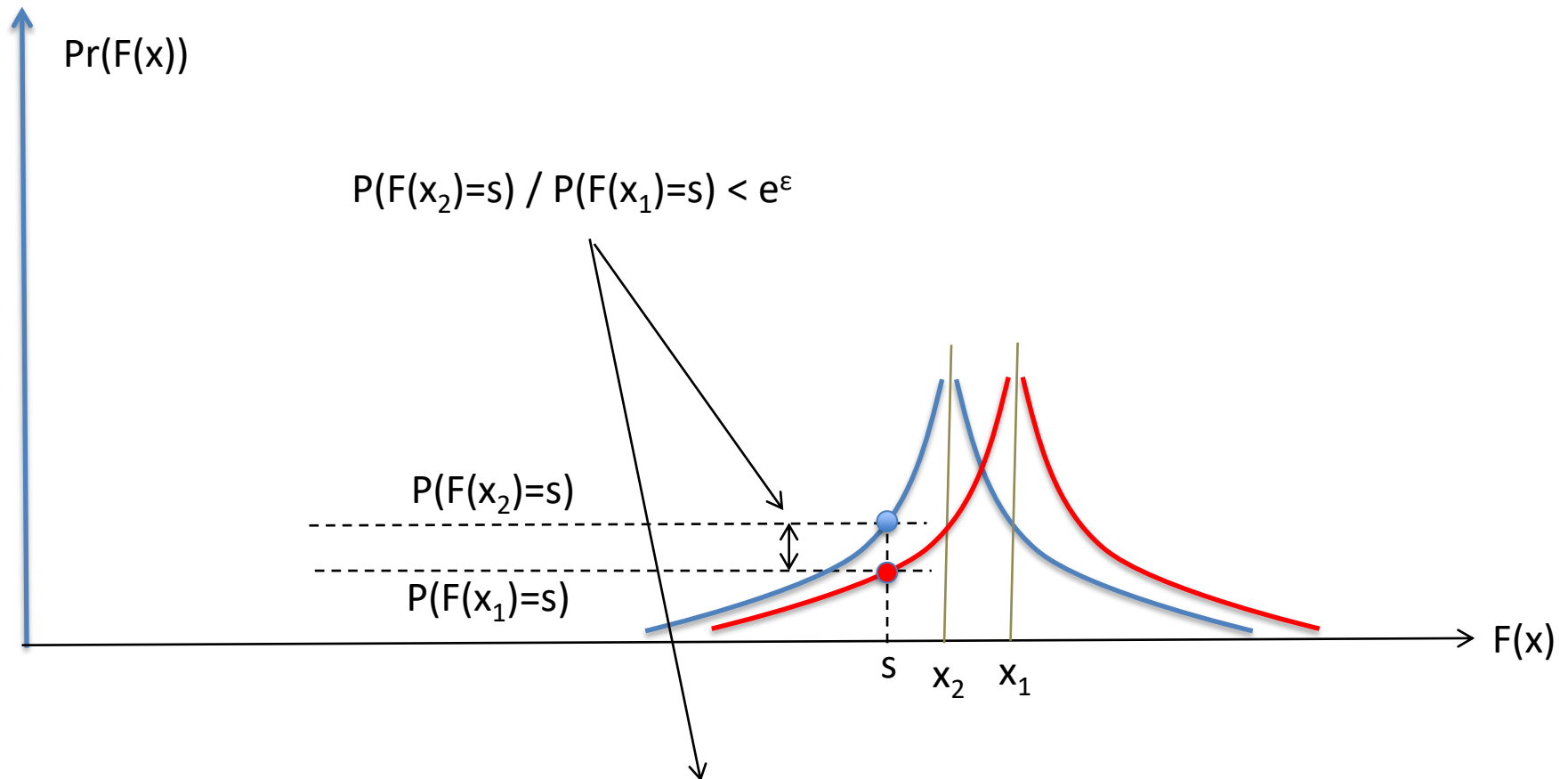


ϵ -Differential privacy: Adding noise

What does F add to the reply?



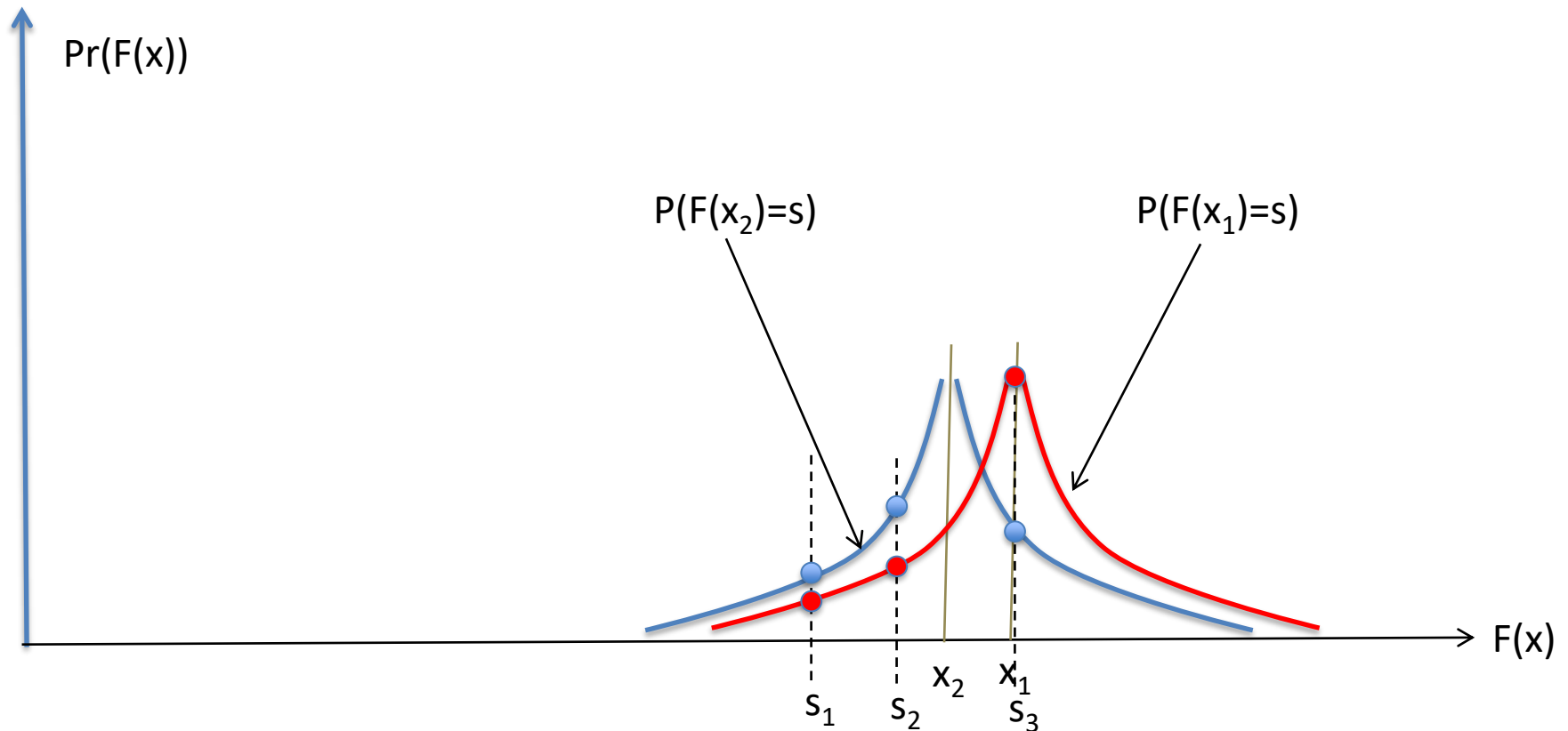
ϵ -Differential privacy: Impact of noise



More or less: $1 - \epsilon < P(F(x_2)=s) / P(F(x_1)=s) < 1 + \epsilon$

ϵ -Differential privacy: Impact of noise

X: What is the number of people with glasses and gray hair?

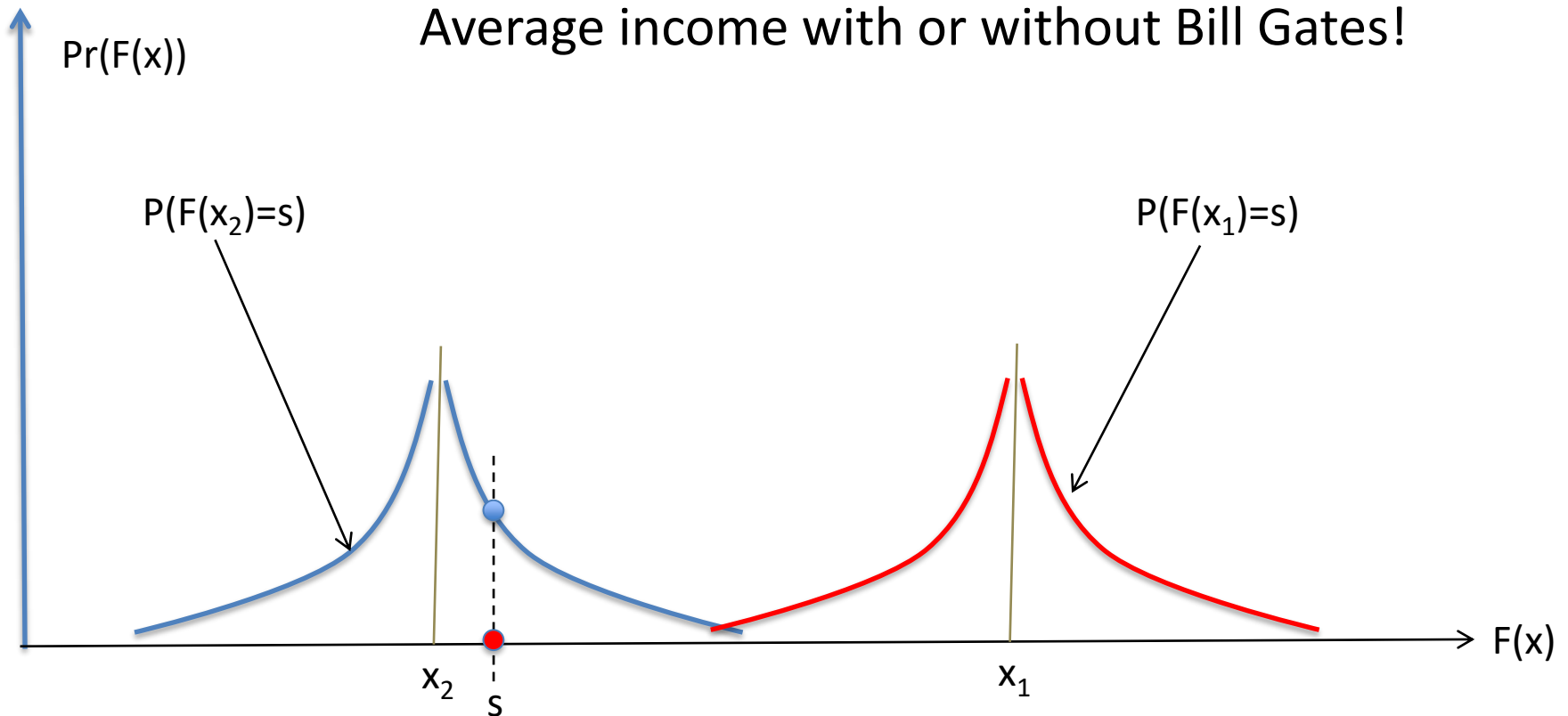


More or less: $1 - \epsilon < P(F(x_2)=s) / P(F(x_1)=s) < 1 + \epsilon$

ϵ -Differential privacy: Impact of noise

X: What is the average income among us (assume Bill Gates is here)?

What is the worst case?

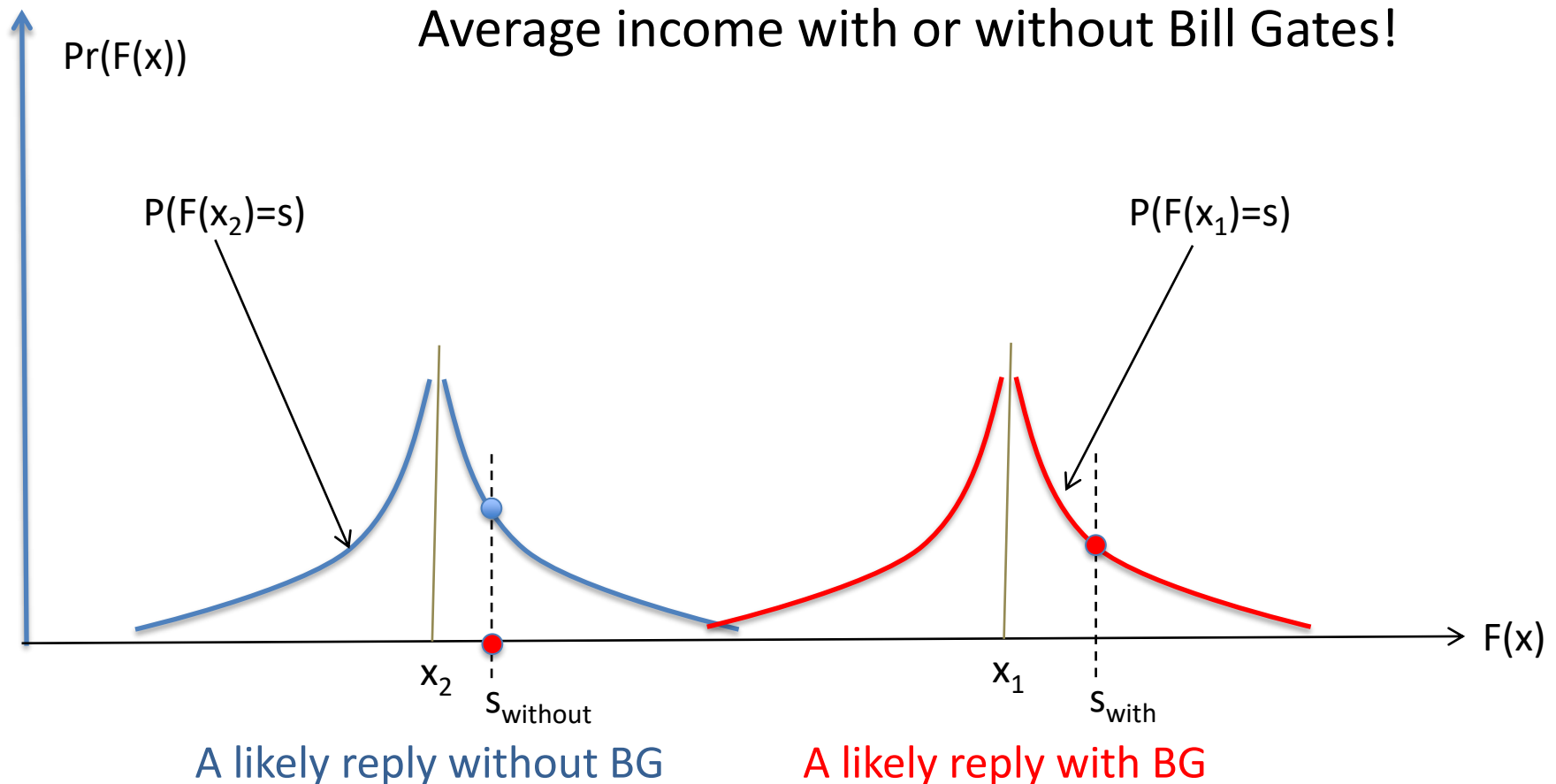


More or less: $1 - \epsilon < P(F(x_2)=s) / P(F(x_1)=s) < 1 + \epsilon$

ϵ -Differential privacy: Impact of noise

X: What is the average income among us (assume Bill Gates is here)?

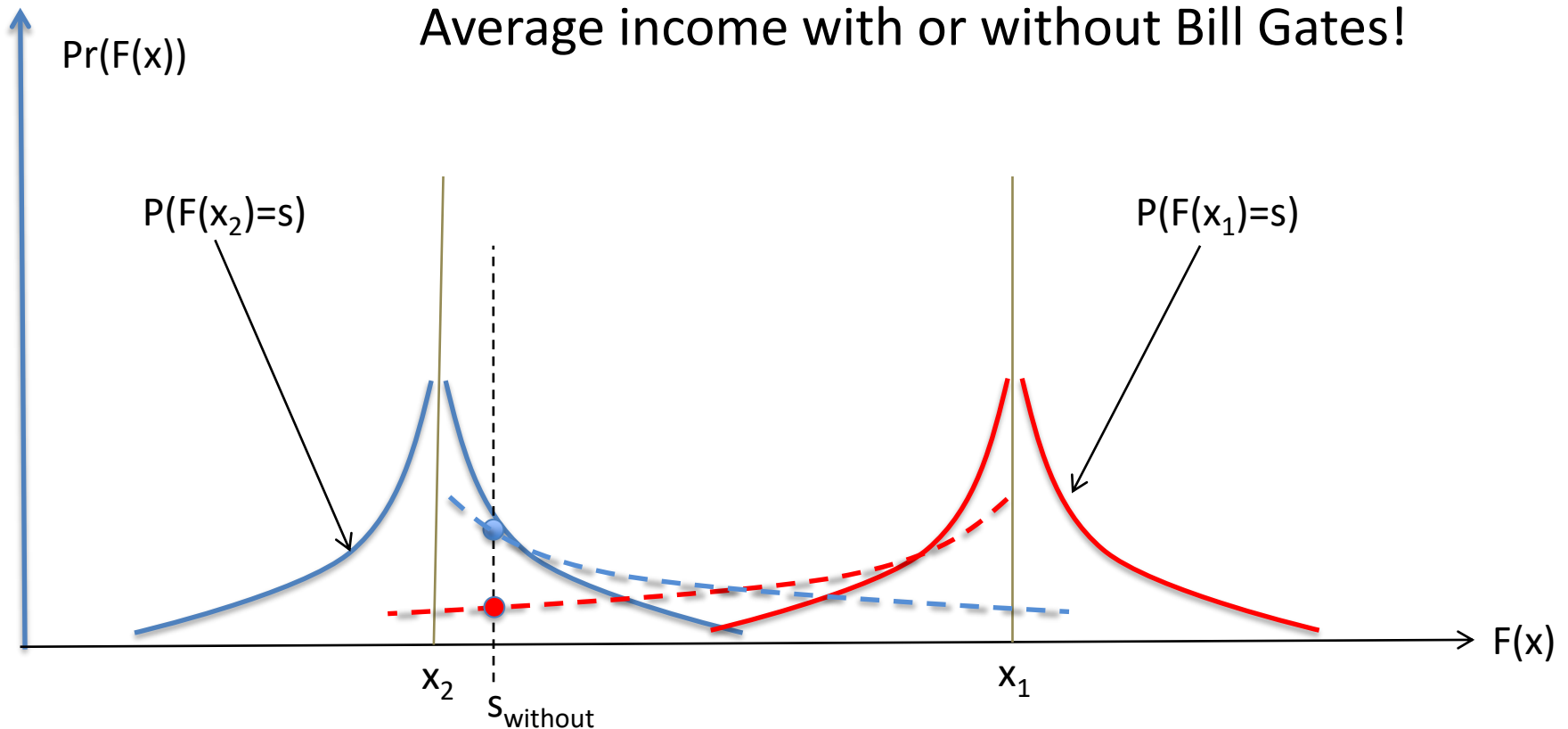
What is the worst case?



ϵ -Differential privacy: Impact of noise

X: What is the average income among us (assume Bill Gates is here)?

What is the worst case?

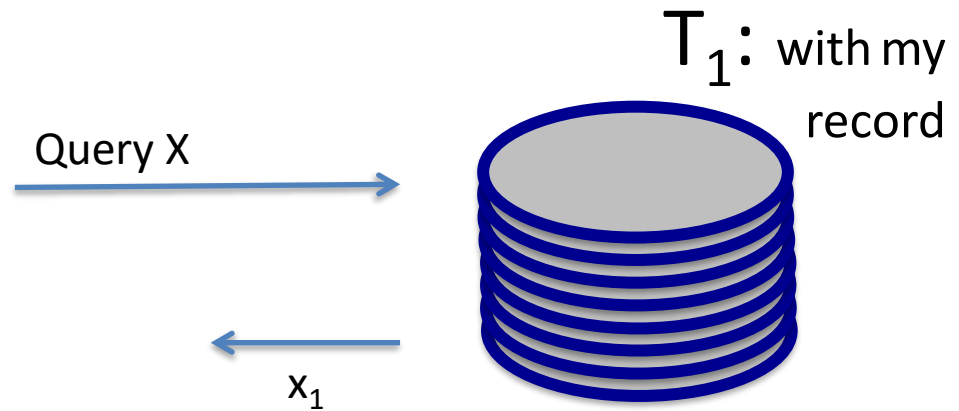


A likely reply without BG

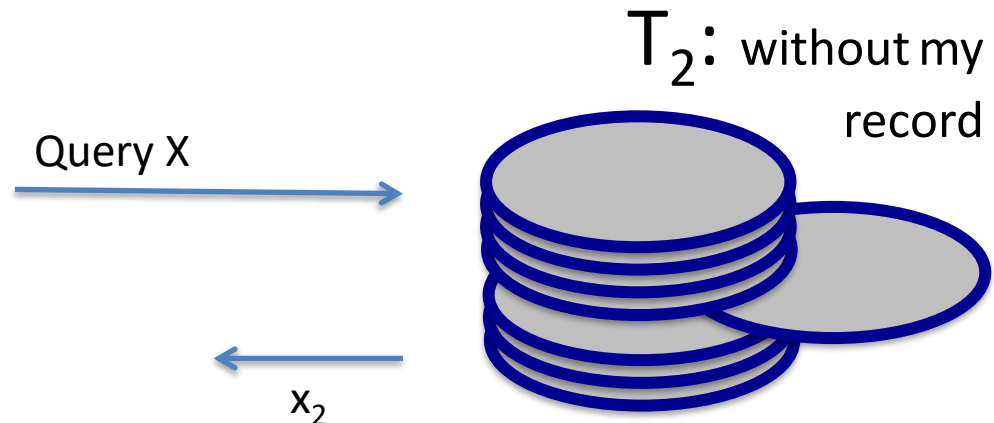
ϵ -D. privacy: How much noise to add?

X: What is the number of people with glasses and gray hair?

- Assume that there are N records in table T originally
- What could be the difference between x_1 and x_2 ?
- 1 (if I am with glasses and gray haired), 0 (otherwise)



- **Sensitivity** of query X is defined as the maximum of the possible answers above, considering all data records
 - $\text{Sen}(X) = \max |x_1 - x_2|$
- Sensitivity of a **count query** (like query X above) = 1



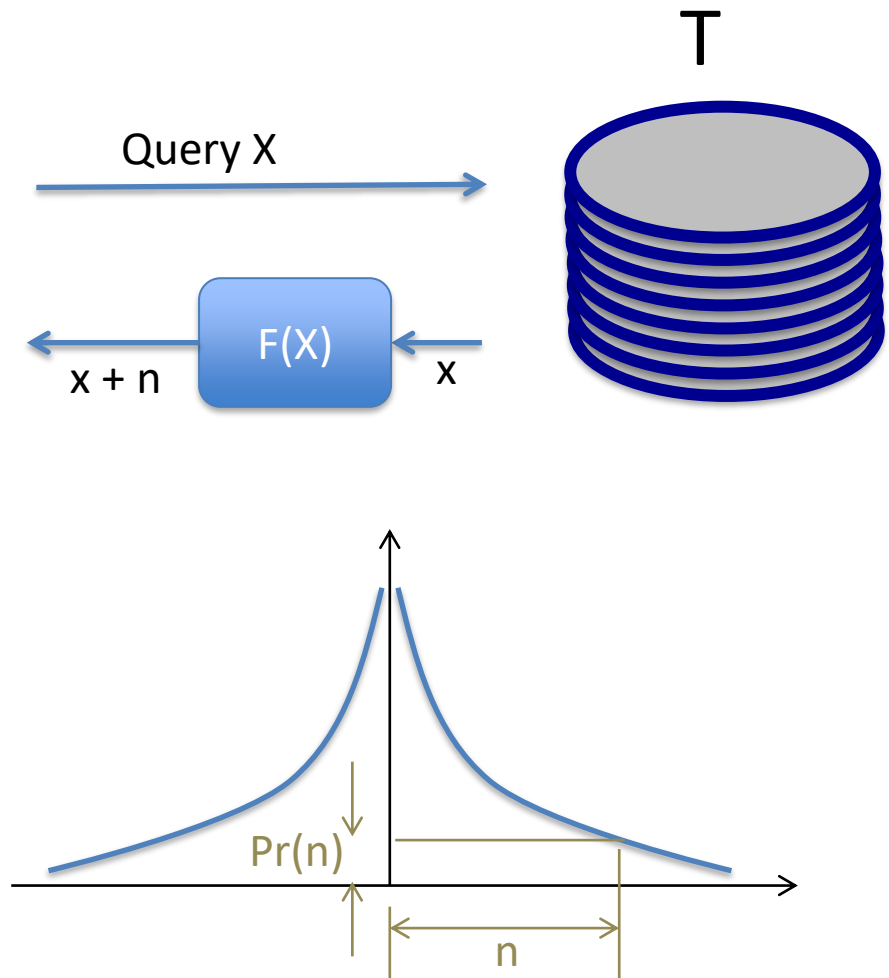
ϵ -Differential privacy: Sensitivity and ϵ

What does F add to the reply?

Laplace noise: Magnitude of n is determined with Laplace PDF (Probability Distribution Function)

$$\Pr(\text{noise} = n) = \frac{1}{2\lambda} \exp(-|n|/\lambda)$$

$$\lambda = \text{Sen}(X) / \epsilon$$



ϵ -D. privacy: Laplace distribution

- Laplace distribution $\text{Lap}(\lambda)$
 - $\Pr[X = x] = 1/2\lambda \exp(-|x|/\lambda)$
 - variance $2\lambda^2$
- Additive Laplace noise
 - $\lambda = \text{Sen}(X)/\epsilon$
 - $\Pr[X = x] = \epsilon/2\text{Sen}(X) \exp(-\epsilon|x|/\text{Sen}(X))$ or $\text{Lap}(\text{Sen}(X)/\epsilon)$
- This addition realizes differential privacy because
 - **Shifting the distribution** changes the probability by at most a **constant**
 - Max change due to someone being in the dataset: $x \rightarrow x + \text{Sen}(X)$

Proof: Assuming that $x + \text{Sen}(X) > 0$, we have

$$\begin{aligned} \Pr[X = x + \text{Sen}(X)] &= \epsilon/2\text{Sen}(X) \cdot \exp(-\epsilon|x + \text{Sen}(X)|/\text{Sen}(X)) \\ &= \exp(-\epsilon) \Pr[X = x] \end{aligned}$$

Or putting it differently: $\Pr[X = x] = \exp(\epsilon) \Pr[X = x + \text{Sen}(X)]$

Grasping the concept: Fixed sensitivity

- $\lambda = \text{Sen}(X)/\epsilon$ or $\epsilon = \text{Sen}(X)/\lambda$

Higher $\epsilon \rightarrow$ less privacy

- **Case I:** For the same $\text{Sen}(X)$

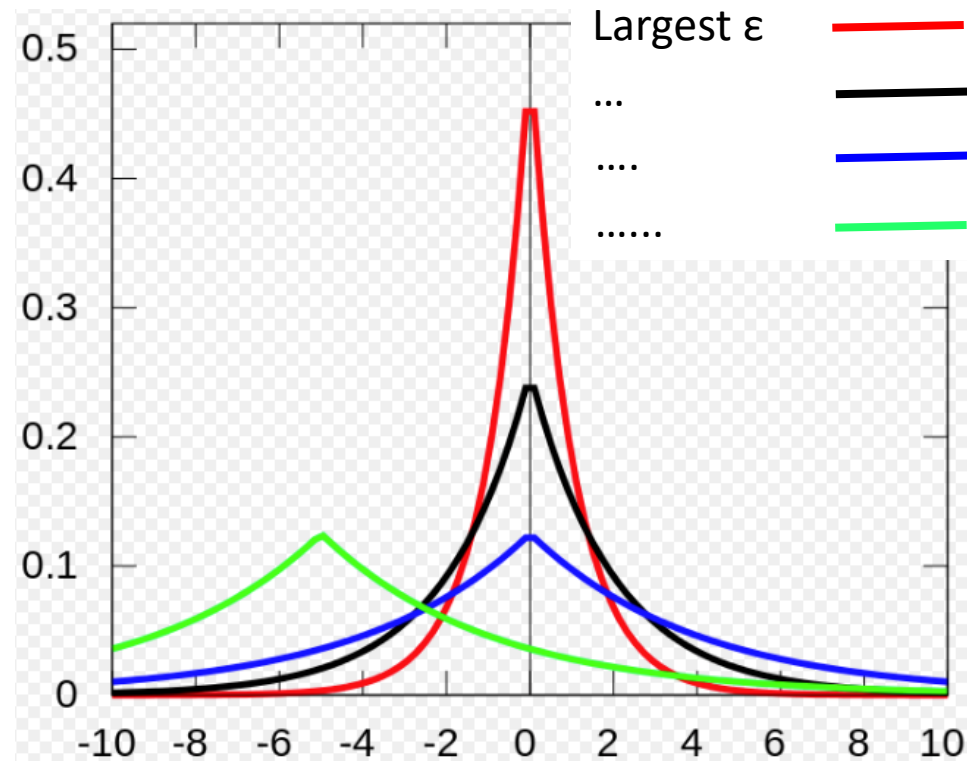
- The larger ϵ

- The smaller added Laplace noises (i.e., being more likely)

\rightarrow More utility

- The larger differences between the probabilities of having a record or not

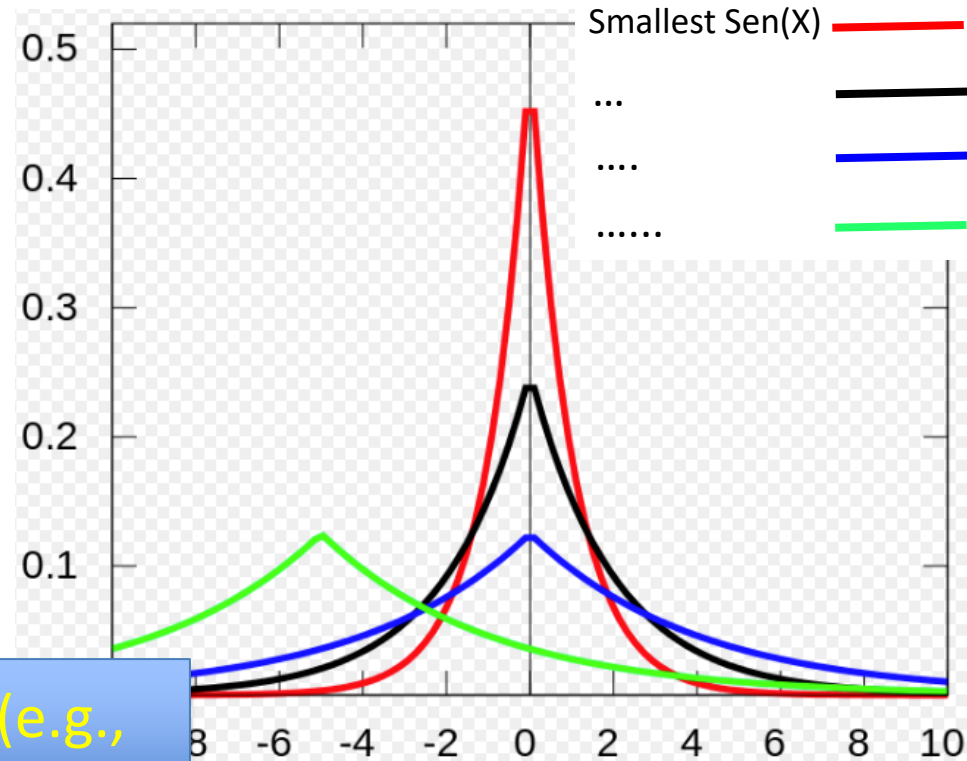
\rightarrow Less privacy



New topic: privacy-utility tradeoff

Grasping the concept: Fixed ϵ

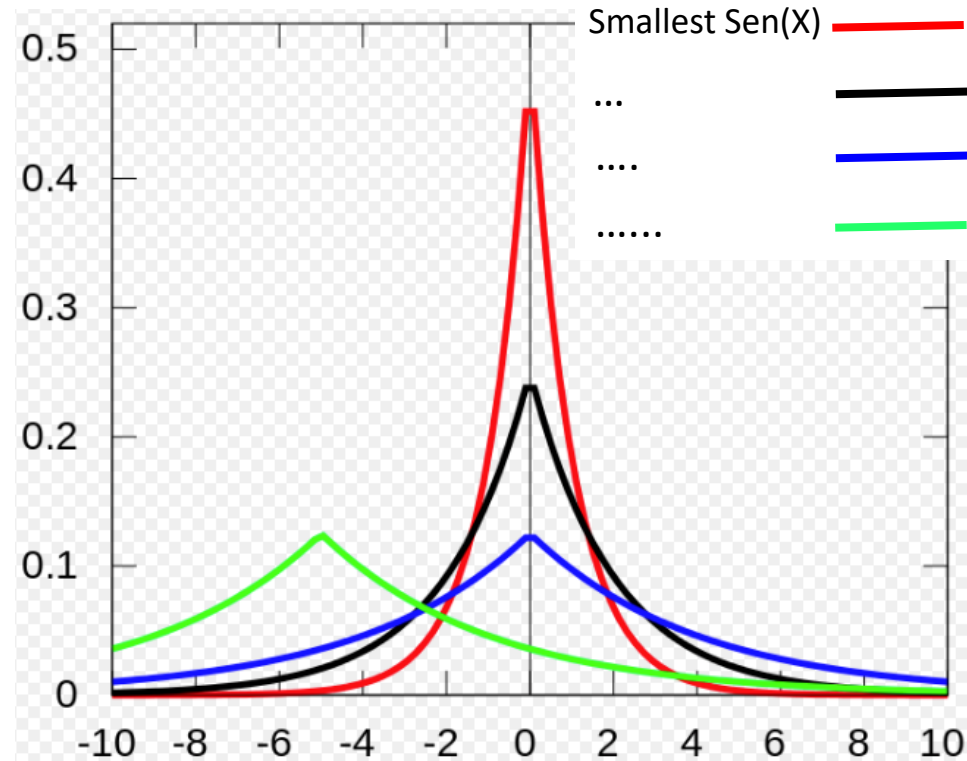
- $\lambda = \text{Sen}(X)/\epsilon$ or $\text{Sen}(X) = \epsilon \lambda$
- **Case II: For the same ϵ**
- The smaller sensitivity
 - The smaller added Laplace noises (i.e., being more likely)
 - No differences between probabilities (i.e., of having a record or not)



Delete outliers from the dataset (e.g., Bill Gates in case of the average income)

Grasping the concept: Fixed ϵ

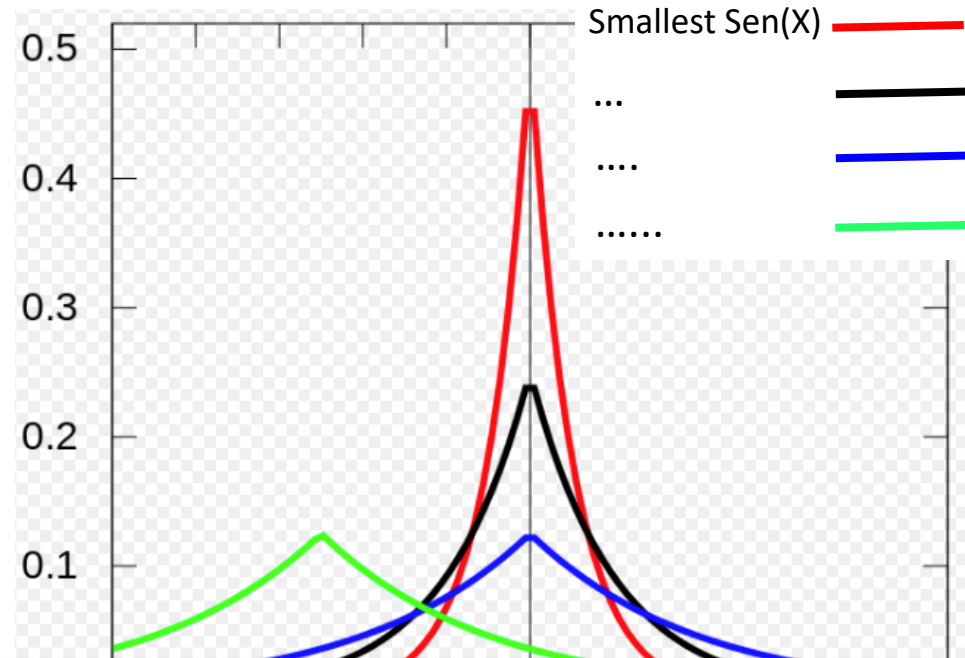
- $\lambda = \text{Sen}(X)/\epsilon$ or $\text{Sen}(X) = \epsilon \lambda$
- **Case II: For the same ϵ**
- The smaller sensitivity
 - The smaller added Laplace noises (i.e., being more likely)
→ More utility
 - No differences between probabilities (i.e., of having a record or not)
→ Same privacy



So no privacy-utility tradeoff, Eureka!

Grasping the concept: Fixed ϵ

- $\lambda = \text{Sen}(X)/\epsilon$ or $\text{Sen}(X) = \epsilon \lambda$
- **Case II: For the same ϵ**
- The smaller sensitivity
 - The smaller added Laplace noises (i.e., being more likely)
→ More utility
 - No differences between probabilities (i.e., of having a record or not)



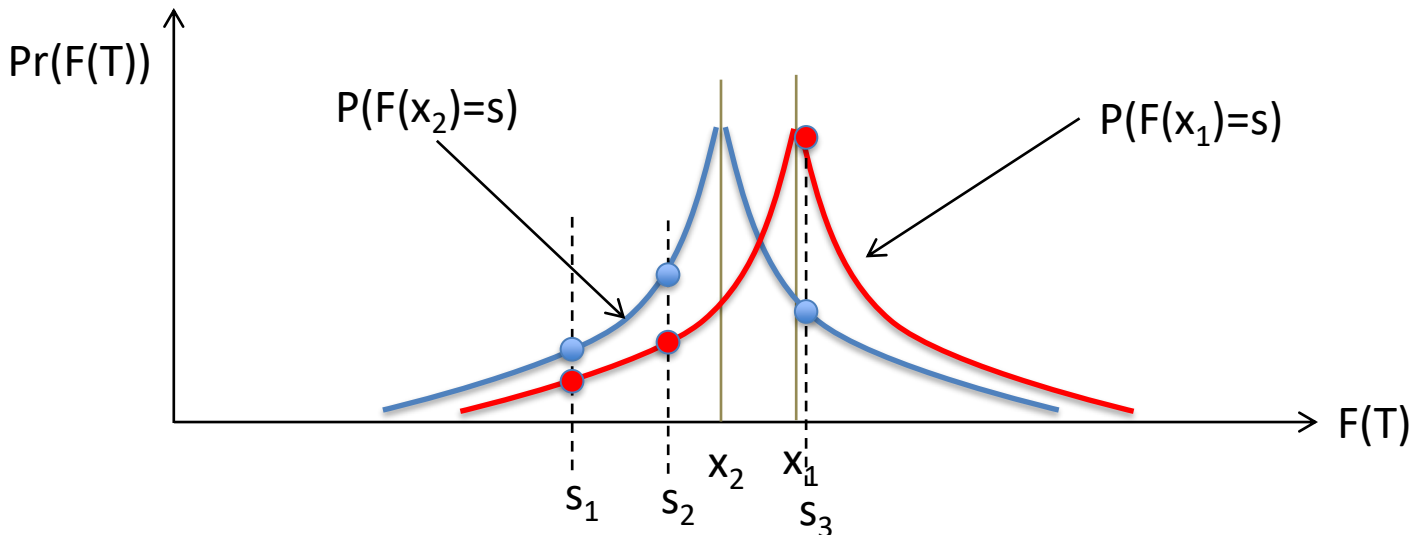
Making sensitivity smaller means distorting data, thus, reducing data utility (e.g., removing Bill Gates record in case of the average income)

ϵ -Differential privacy: Summary

- How
 - Assume tables T_1 and T_2 differ in one data record
 - F is a randomized function

$$\left| \ln \frac{P(F(T_1) = S)}{P(F(T_2) = S)} \right| \leq \epsilon \quad \forall S \in \text{Range}(F)$$

- Where $\text{Range}(F)$ is the set of possible outputs of the randomized function F



ϵ -Differential privacy: Summary

- How

- Assume tables T_1 and T_2 differ in one data record
- F is a randomized function

$$\left| \ln \frac{P(F(T_1) = S)}{P(F(T_2) = S)} \right| \leq \epsilon \quad \forall S \in \text{Range}(F)$$

- Where $\text{Range}(F)$ is the set of possible outputs of the randomized function F

- Note I: This is a **property of query X** (i.e., being specific to the query type because it depends on the sensitivity of query X)
- Note II: This may be a **property of the dataset** (i.e., being specific to the data records because it depends on the worst case data record)
 - Worst case ever (in this case does not depend on the dataset)
 - Worst case in that dataset

3. Some exercises

Introduction, ϵ -differential privacy – interactive, **Some exercises**, ϵ -differential privacy – non-interactive, Other relevant topics, Takeaways, References

Computing sensitivity of a query

- Assume: there are N records
- Sensitivity of counts
- Example query
 - How many people in this class have glasses? Answer: m
- What is the sensitivity?
- Remove/add someone without glasses: $m \rightarrow m$
- Add someone with glasses: $m \rightarrow m + 1$
- Remove someone with glasses: $m \rightarrow m - 1$
- What is the worst case change?
- Thus, the sensitivity is 1

Computing sensitivity of a query

- Assume: there are N records
- Sensitivity of sum
- Example query
 - What is total age of people < 25yrs in this class? Answer: m
- What is the sensitivity?
- Remove/add someone ≥ 25 yrs: $m \rightarrow m$
- Add someone < 25yrs: $m \rightarrow m + \text{age}(\text{someone})$
- Remove someone < 25yrs: $m \rightarrow m - \text{age}(\text{someone})$
- What is the worst case change?
- The sensitivity is: Max of age(someone here < 25yrs)

24 years, 365 days,

Computing sensitivity of a query

- Assume: there are N records
- Sensitivity of average
- Example query
 - What is the average number of people with glasses in this class? Answer: m
- What is the sensitivity?
- Remove/add someone without glasses: $m \rightarrow \approx m$
- Add someone with glasses: $m \rightarrow \approx m + 1/N$
- Remove someone with glasses: $m \rightarrow \approx m - 1/N$
- What the worst-case change?
- The sensitivity is: $1/N$

Sensitivity of a query

- Dataset: x_1, x_2, \dots, x_N
- Assume
 - N is odd
 - All x_n are real values in $[0, L]$ where L is a large number
 - $x_1 \leq x_2 \leq \dots \leq x_N$
 - Rank of the median = $m = (N+1)/2$
- What the worst-case sensitivity of the median?

Sensitivity of median: Worst case

- Dataset: x_1, x_2, \dots, x_N
- Assume
 - N is odd
 - All x_n are real values in $[0, L]$ where L is a large number
 - $x_1 \leq x_2 \leq \dots \leq x_N$
 - Rank of the median = $m = 0.5*(N+1)$
- What the worst-case sensitivity of the median?

x_1	x_2	...	x_{m-1}	x_m	x_{m+1}	...	x_{N-1}	x_N
0	0	0	0	0	L	L	L	L

- Median with all x_n is 0
- Median without x_n is
 - When $n > m$: 0 (when $n > m$)
 - when $n \leq m$: $0.5(x_{m-1} + x_{m+1}) = L/2$ (some would say $x_{m-1} + x_{m+1} = L$)
- Thus, the worst-case median (i.e., the global) sensitivity is $L/2$ (resp. L)

Sensitivity of median: Generic case

- Dataset: x_1, x_2, \dots, x_N where $x_1 \leq x_2 \leq \dots \leq x_N$ and $m = 0.5 * (N+1)$
- What the sensitivity of the median for a given dataset (i.e., a typical median)?

case-0	with all entries	x_1	x_2	...	x_{m-2}	x_{m-1}	x_m	x_{m+1}	x_{m+2}	...	x_{N-1}	x_N	
case-I	without x_m					*	×	*					
case-II	without $x_i, i < m$	←					*	*					
case-III	without $x_i, i > m$					*	*	→					

- case-I: median = $0.5(x_{m-1} + x_{m+1})$; med. sen. $MS_I = |x_m - 0.5(x_{m-1} + x_{m+1})|$
- case-II: median = $0.5(x_m + x_{m+1})$; med. sen. $MS_{II} = 0.5(x_{m+1} - x_m)$
- Case-III: median = $0.5(x_{m-1} + x_m)$; med. sen. $MS_{III} = 0.5(x_m - x_{m-1})$
- Median sensitivity = max of the three above

Danger: Inference attack!

Sensitivity of median: Inference attack

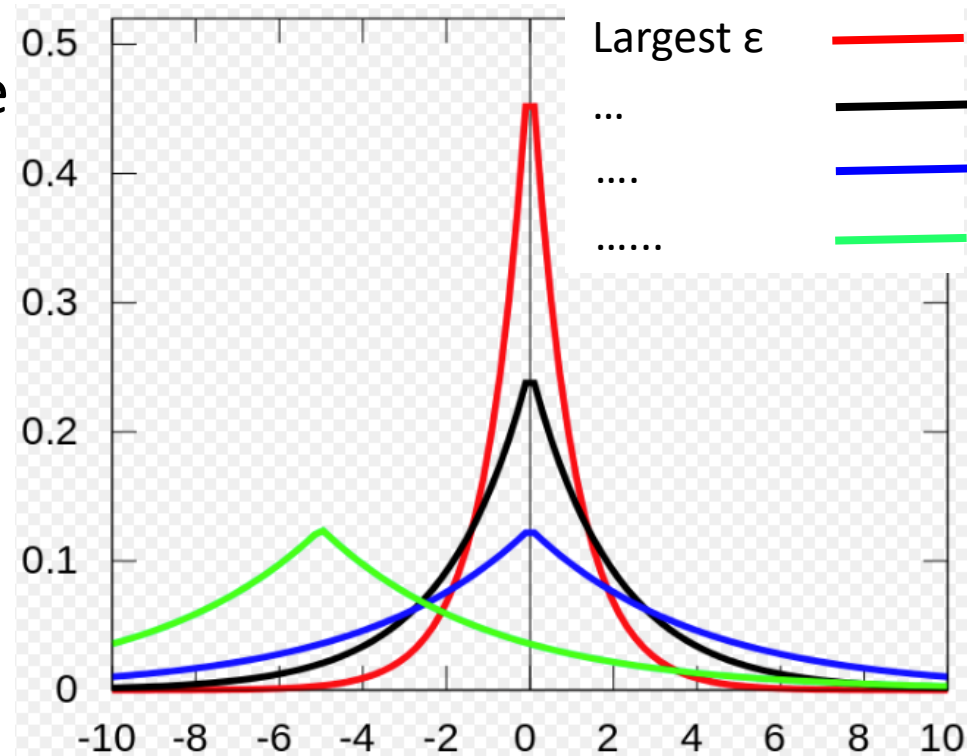
- Median sensitivity = $\max(0.5(x_m - x_{m-1}); 0.5(x_{m+1} - x_m))$
- Noise depends on dataset → noise level can leak some info about the dataset
- Dataset 1
 - $x_1 = \dots = x_m = x_{m+1} = 0$ and $x_{m+2} = x_{m+3} \dots = x_N = L$
 - Median sensitivity is 0
- Dataset 2
 - $x_1 = \dots = x_{m-1} = x_m = 0$ and $x_{m+1} = x_{m+2} \dots = x_N = L$
 - median sensitivity is $L/2$
- Smooth sensitivity: define an upper-bound (i.e., maximum) to sensitivity, see [NIS'07]

Getting some feeling on ϵ -Diff. privacy

- Search for the website: “Differential Privacy: The Basics”
- Link:
- <https://agkn.wordpress.com/2014/09/08/differential-privacy-the-basics/>
- Scenario of an inference attack:
 - “Suppose you have access to a database that allows you to compute the total income of all residents in a certain area.
 - If you knew that Mr. White was going to move to another area,
 - simply querying this database before and after his move would allow you to deduce his income.”

Recall: Effect of larger ϵ

- For the same $\text{Sen}(X)$
- The larger ϵ
 - The smaller added Laplace noises (i.e., being more likely)
 - The larger differences between the probabilities
- Chopping ϵ (budgeting)
$$\epsilon = \epsilon_1 + \epsilon_2 + \dots + \epsilon_m$$



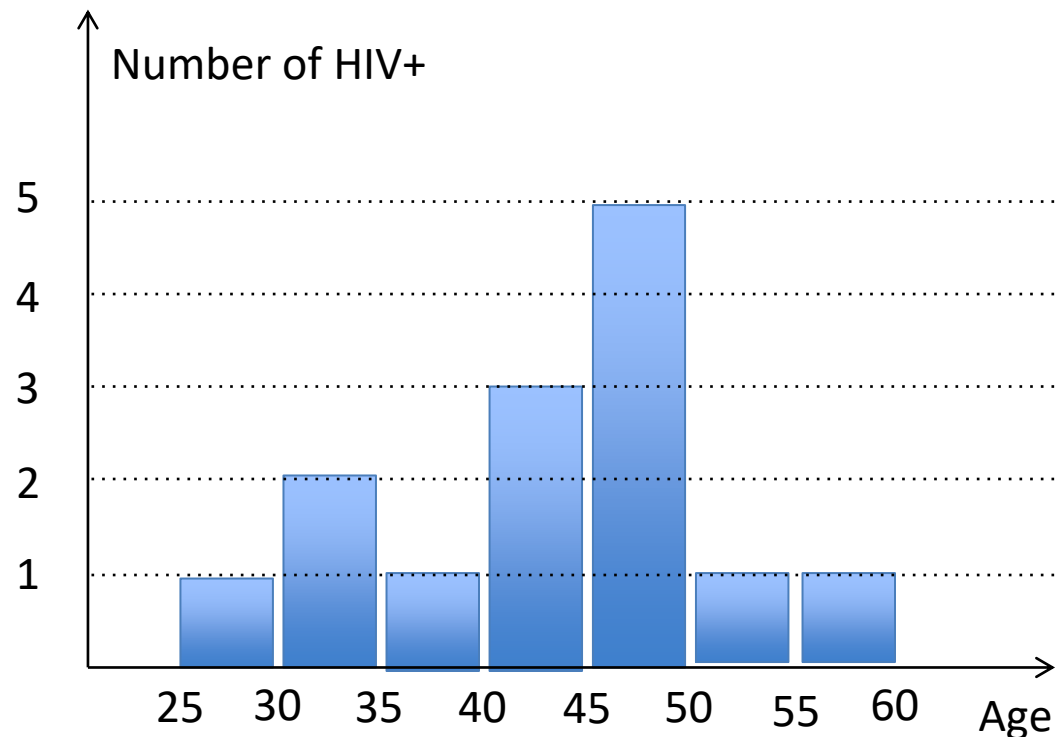
Multiple queries

- Limit on the number of queries (by the same analyst): m
- Budget $\varepsilon = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_m$
- How to make sure that at most m queries are made by a specific person?
- Access control
 - (Identification, authentication, and authorization)
- **Actually, this is a sort of usage control!**

Computing sensitivity of a query

- Assume: there are N records
- Sensitivity of histogram (/a distribution of values)
- Example query

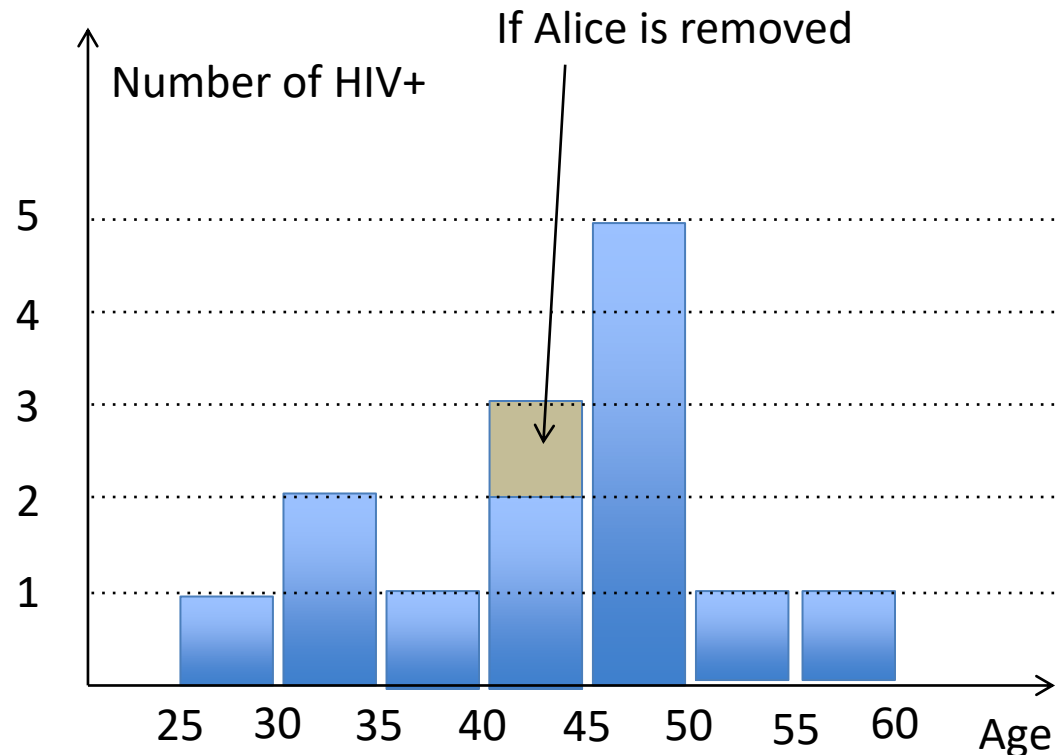
name	age	HIV+
Alice	42	yes
Bob	31	no
Carol	32	yes
Dave	36	yes
Ellen	45	yes
Frank	26	no
Grace	39	yes
...	...	



Computing sensitivity of a query

- Assume: there are N records
- Sensitivity of histogram (/a distribution of values)
- Example query

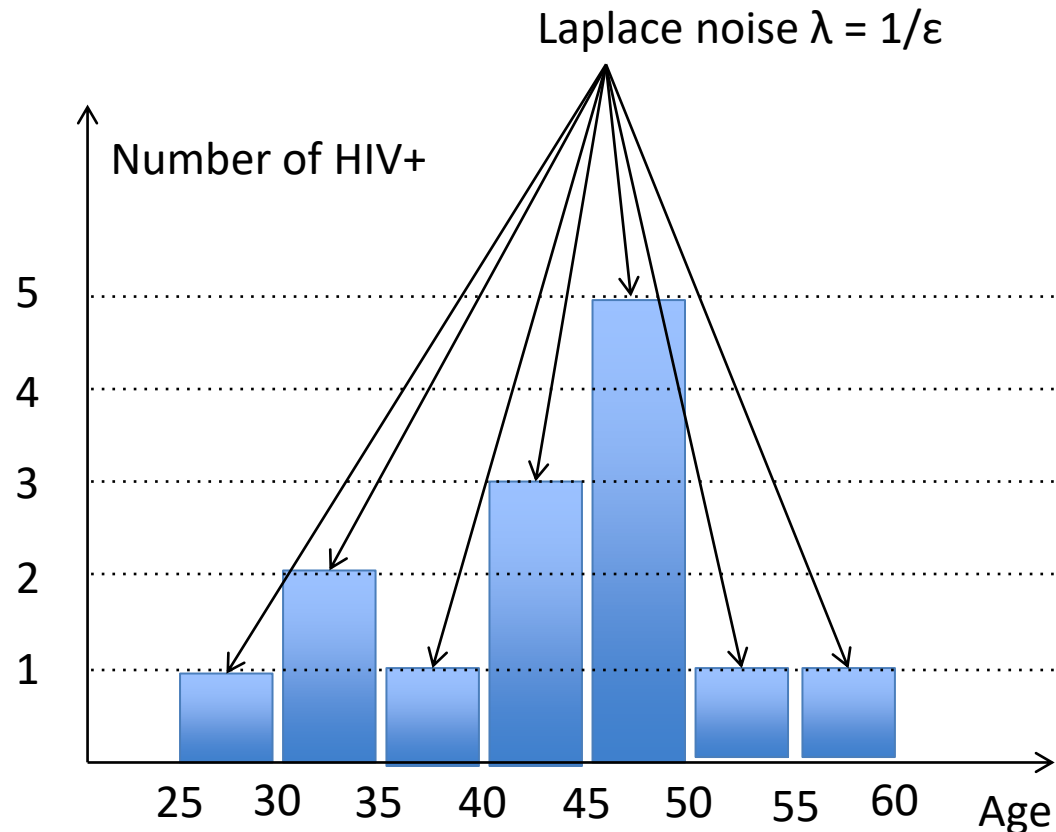
name	age	HIV+
Alice	42	yes
Bob	31	no
Carol	32	yes
Dave	36	yes
Ellen	45	yes
Frank	26	no
Grace	39	yes
...	...	



Computing sensitivity of a query

- Assume: there are N records
- Sensitivity of histogram (/a distribution of values)
- Sensitivity of counts?
- $\text{Sen}(\text{count per bin}) = 1$
- $\lambda = \text{Sen}(X)/\epsilon = 1/\epsilon$
 - Per bin

Objective: “The application of DP to such a histogram guarantees that changing or removing any record from the database has negligible impact on the output histogram” [XU'12]

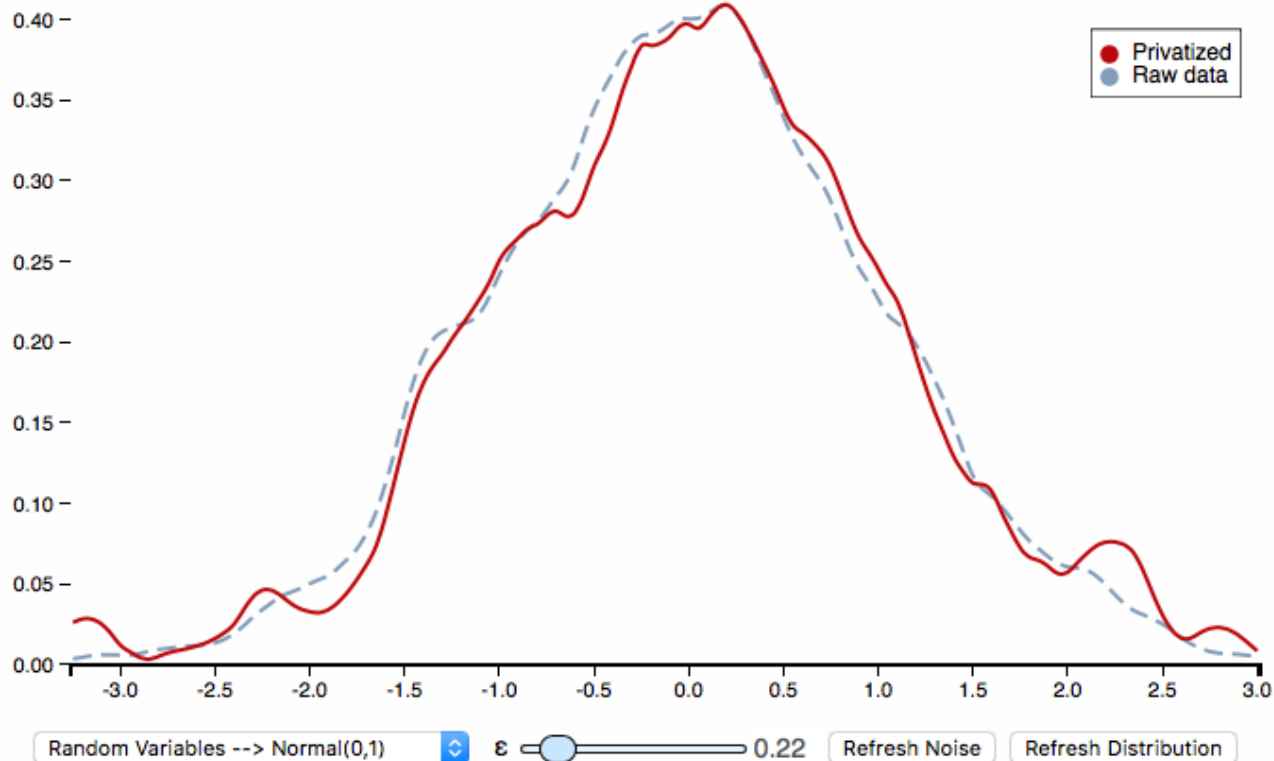


Getting some feeling on ϵ -Diff. privacy

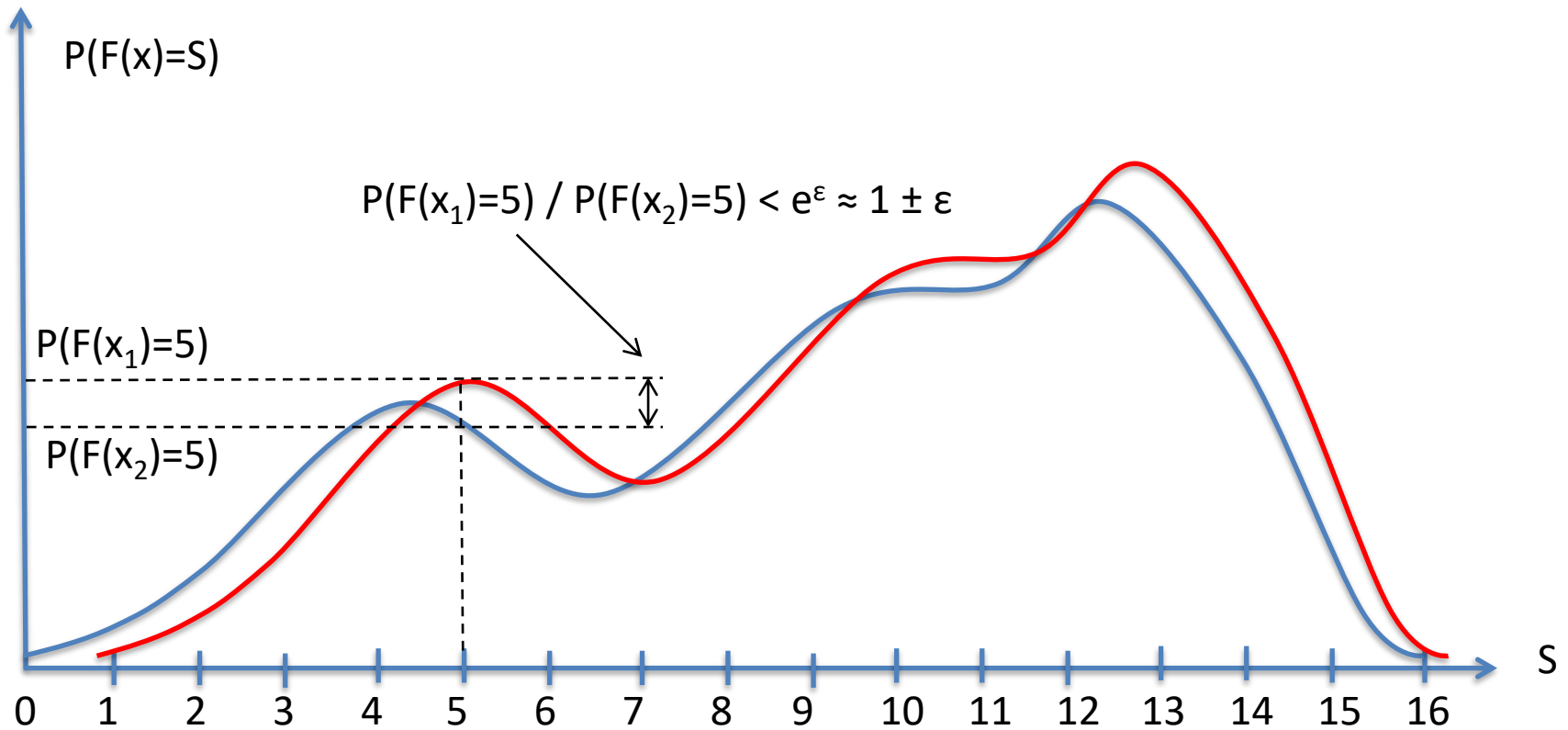
- Look at the interactive simulation
 - <http://content.research.neustar.biz/blog/differential-privacy/DensityWidget.html>
- Investigate the simulation
- Explain what happens when
 - ϵ is increased or decreased (use the “Step” button)

Getting some feeling on ϵ -Diff. privacy

- Look at the interactive simulation
 - <http://content.research.neustar.biz/blog/differential-privacy/DensityWidget.html>



ϵ -Differential privacy



4. ϵ -differential privacy – non-interactive

Introduction, ϵ -differential privacy – interactive, Some exercises, **ϵ -differential privacy – non-interactive**, Other relevant topics, Takeaways, References

Data publication cases

- Interactive
 - Reply to (multiple) queries
 - Statistical databases
- Non-interactive
 - Micro data: datasets about individuals

Formal protection of microdata

- How to protect this set with ϵ -Differential privacy?

name	job	sex	age	disease	Height
Bob	engineer	male	35	hepatitis	184
Fred	engineer	male	38	hepatitis	180
Doug	lawyer	male	38	HIV	210
Alice	writer	female	30	flu	172
Cathy	writer	female	33	HIV	170
Emily	dancer	female	31	HIV	169
Gladys	dancer	female	31	HIV	171

EID

QID

SAtt

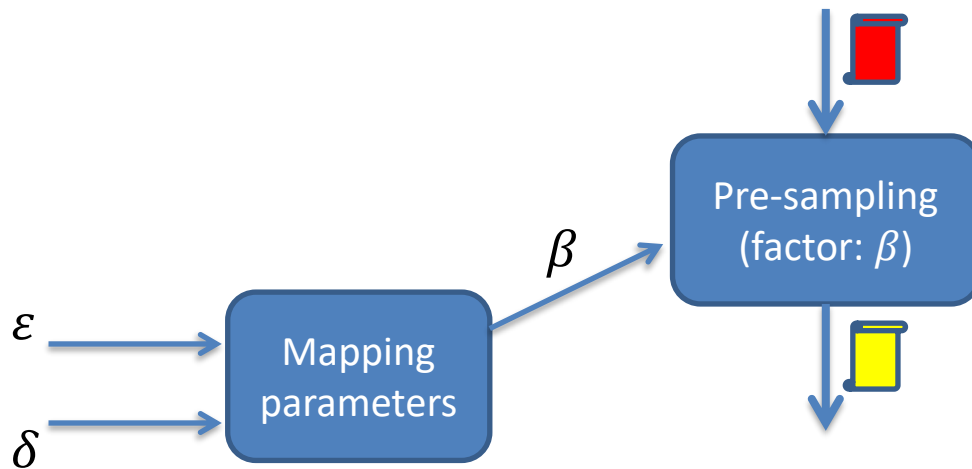
NSAtt

$$P(F(x_2)=s) / P(F(x_1)=s) < e^\epsilon$$

Protect microdata with ϵ -Differential

- See [BIL'18]
- Implemented in ARX

Protect microdata with ϵ -Differential



Formal protection of microdata

- How to protect this set with ϵ -Differential privacy?

name	job	sex	age	disease	Height
/***/n ₁	engineer	male	35	hepatitis	184
/***/n ₂	engineer	male	38	hepatitis	180
/***/n ₃	lawyer	male	38	HIV	210
/***/n ₄	writer	female	30	flu	172
/***/n ₅	writer	female	33	HIV	170
/***/n ₆	dancer	female	31	HIV	169
/***/n ₇	dancer	female	31	HIV	171

EID

QID

SAtt

NSAtt

Formal protection of microdata

- How to protect this set with ϵ -Differential privacy?

name	job	sex	age	disease	Height
/***/n ₁	engineer	male	35	hepatitis	184
/***/n ₃	lawyer	male	38	HIV	210
/***/n ₄	writer	female	30	flu	172
/***/n ₇	dancer	female	31	HIV	171

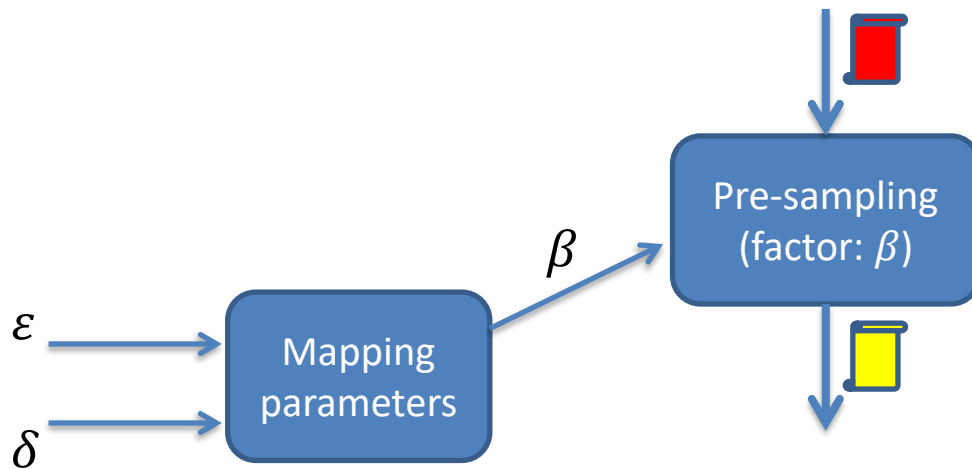
EID

QID

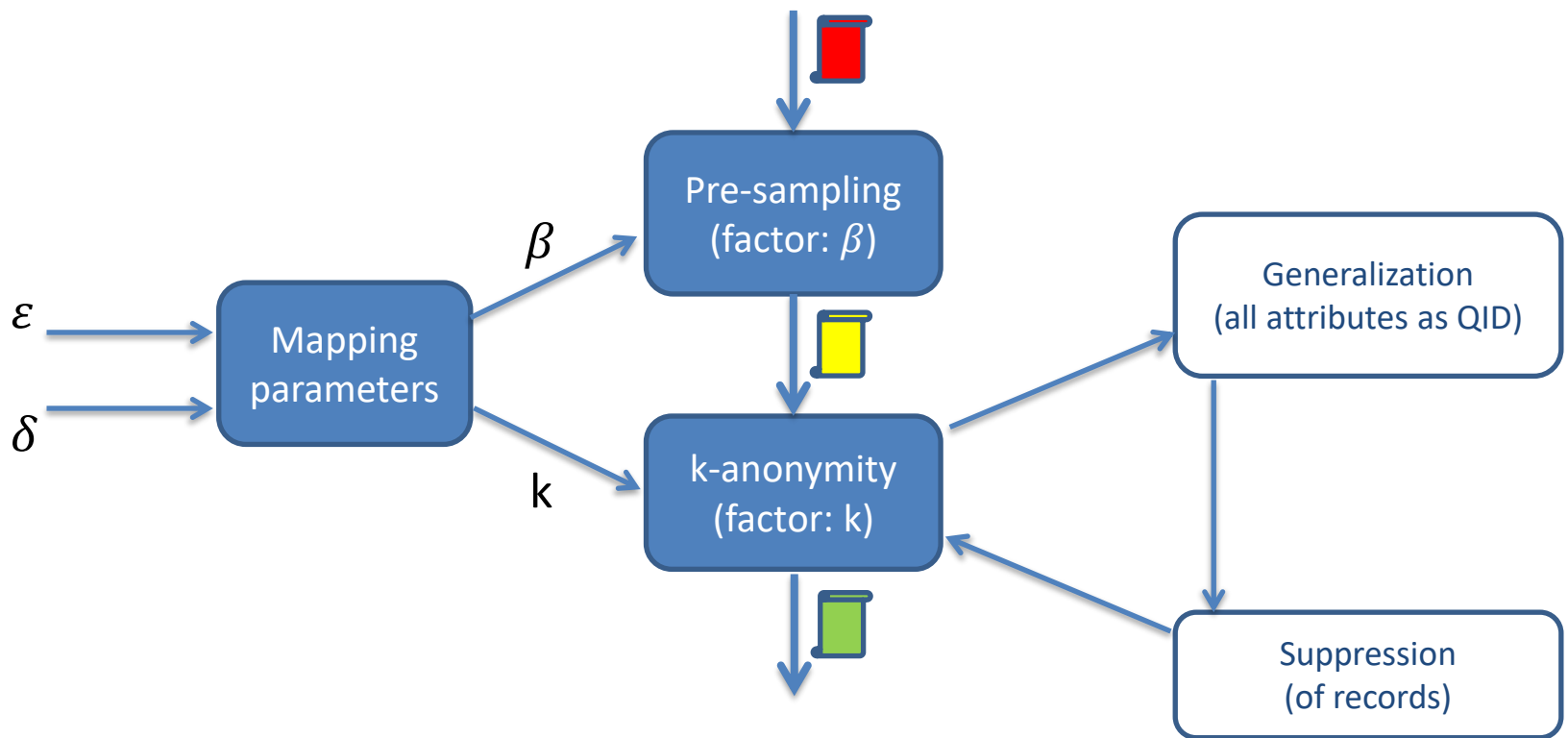
SAtt

NSAtt

Protect microdata with ϵ -Differential



Protect microdata with ϵ -Differential



Formal protection of microdata

- How to protect this set with ϵ -Differential privacy?

name	job	sex	age	disease	Height
/***/n ₁	engineer	male	35	hepatitis	184
/***/n ₃	lawyer	male	38	HIV	210
/***/n ₄	writer	female	30	flu	172
/***/n ₇	dancer	female	31	HIV	171

EID

QID

SAtt

NSAtt

Formal protection of microdata

- How to protect this set with ϵ -Differential privacy?

name	job	sex	age	disease	Height
/***/n ₁	engineer	male	35	hepatitis	184
/***/n ₃	lawyer	male	38	HIV	210
/***/n ₄	writer	female	30	flu	172
/***/n ₇	dancer	female	31	HIV	171

EID

QID

SAtt

NSAtt

Formal protection of microdata

- How to protect this set with ϵ -Differential privacy?

name	job	sex	age	disease	Height
/***/n ₁	Profes.	male	35-39	*	180-210
/***/n ₃	Profes.	male	35-39	*	180-210
/***/n ₄	Artist	female	30-34	*	170-190
/***/n ₇	Artist	female	30-34	*	170-190

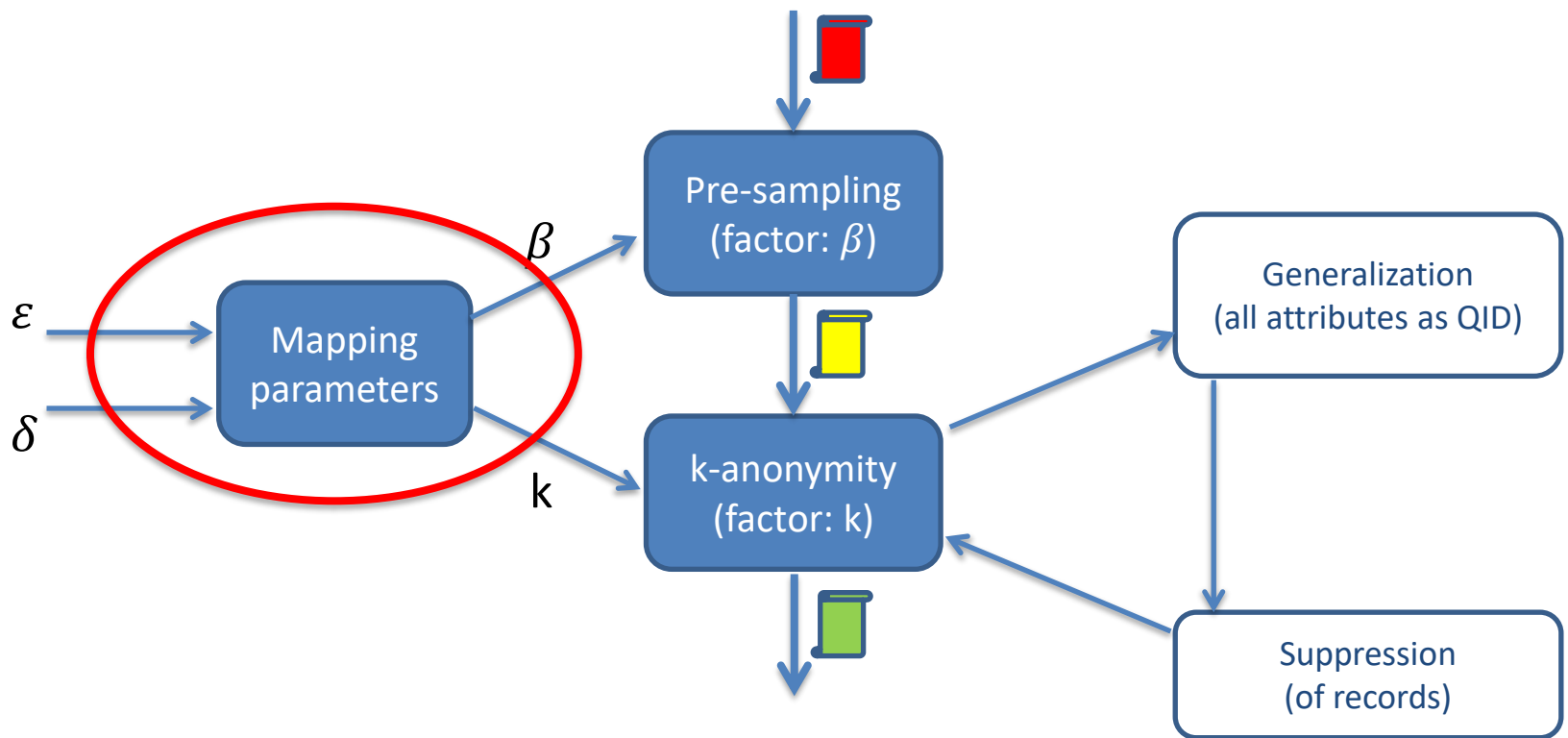
EID

QID

SAtt

NSAtt

Protect microdata with ϵ -Differential



5. Other relevant topics

Introduction, ϵ -differential privacy – interactive, Some exercises, ϵ -differential privacy – non-interactive, **Other relevant topics**, Takeaways, References

Further studies: On choosing ϵ

- Changing ϵ affects utility and privacy adversely (i.e., if one increases, the other decreases)
- Pessimistic ϵ values
 - Sometimes multiple queries are not concerned with the same records, why to apply a high ϵ on those records?
- Some rarely occurring records affect $\text{Sen}(X)$ more than the others
 - Why shouldn't we remove those rare records (like Bill Gates)?
 - Playing with ϵ and $\text{Sen}(X)$

Further studies: Other aspects

- Personalized differential privacy: Some records are more privacy sensitive than others
- How to choose the value of ϵ
 - Literature
 - Rule of thumb
 - Others: [HSU'14] [NAL'15]
 - Taking small steps in right direction
- Example applications: Deployed by USA Census
- Variants: $(\delta-\epsilon)$ -differential privacy
- Combining with secure multiparty computing

US Census 2020

The New York Times

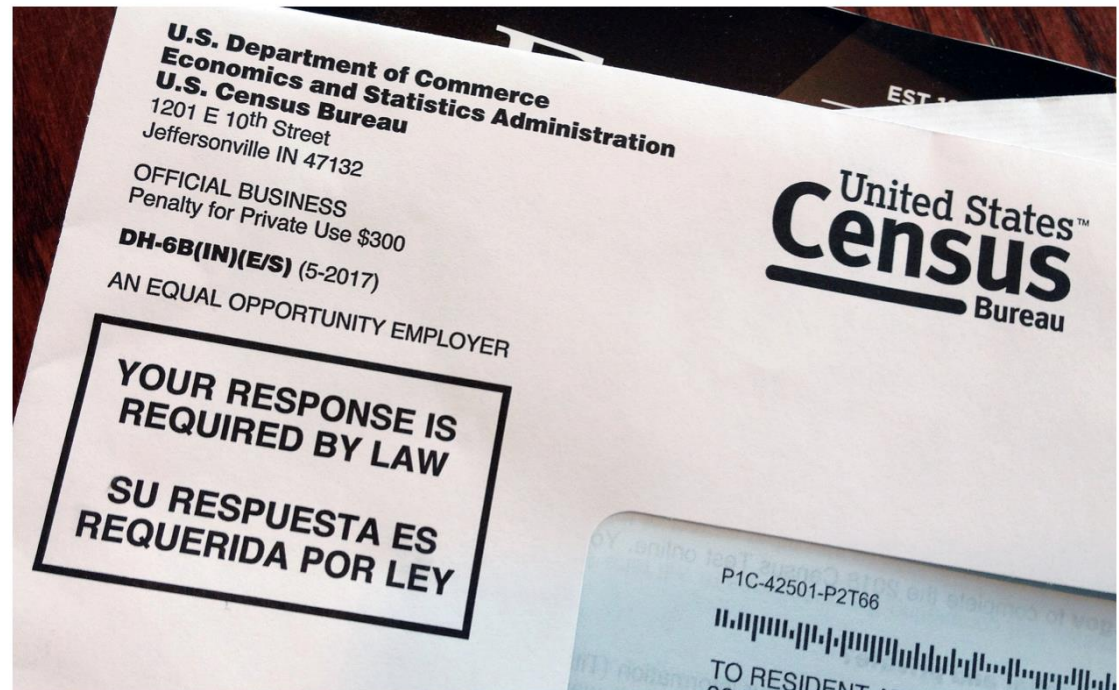
TheUpshot

Ref.: New York Times,
By Mark Hansen,
December 5, 2018

To Reduce Privacy Risks, the Census Plans to Report Less Accurate Data

Guaranteeing people's confidentiality has become more of a challenge, but some scholars worry that the new system will impede research.

Also in use by
Google,
Apple,
Uber



Further studies: Other aspects

- Adding noise with
 - Laplace distribution (already discussed)
 - Normal distribution
- Existing tools
 - RING
 - Rmind
 - PrivaDA
 - PINQ <https://www.microsoft.com/en-us/research/project/privacy-integrated-queries-pinq/>
 - Airavat http://www.cs.utexas.edu/~shmat/shmat_nsd10.pdf

5. Takeaways

Introduction, ϵ -differential privacy – interactive, Some exercises, ϵ -differential privacy – non-interactive, Other relevant topics, **Takeaways**, References

On ϵ -Differential privacy

- (-) Not about preventing record and attribute linking
- (+) Assuring record owners that they may submit their personal information to the database securely in the knowledge that (almost) nothing can be discovered from the database with their information that could not have been discovered without their information
- (+) Providing a guarantee against attackers with arbitrary background knowledge [DWO'06]
- (+) Applicable to both interactive and non-interactive query models [DWO'06]

On ϵ -Differential privacy

- What should the value of ϵ be?
- Is the definition of ϵ -differential privacy applicable to the case at hand?

6. References

Introduction, ϵ -differential privacy – interactive, Some exercises, ϵ -differential privacy – interactive, Other relevant topics, Takeaways, **References**

Note: The reference list is rather long. The most important references are marked in red-bold.

References

- **[BIL'18]: Bild, R, Kuhn, K.A. & Prasser, F. (2018). SafePub: A truthful data anonymization algorithm with Strong privacy guarantees, Proceedings on Privacy Enhancing Technologies ; (1):67–87.**
- [DAL'97]: Dalenius, T. (1977). Towards a methodology for statistical disclosure control, *Statistik Tidskrift* 15, pp. 429–444.
- [DWO'06]: Dwork, C. (2006). Differential privacy. In *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming*, pp. 1–12, doi 10.1007/11787006_1
- [DWO'08]: Dwork, C. (2008). Differential privacy: A survey of results. In *Proceedings of the 5th International Conference on Theory and Applications of Models of Computation (TAMC)*, pp. 1–19.
- [HSU'14]: Hsu, J., Gaboardi, M., Haeberlen, A., Khanna, S., NaraYan, A., Pierce, B.C., and Rotyh, A. (2014). Differential Privacy: An Economic Method for Choosing Epsilon.
- [LEE'11]: Lee, J. and Clifton, C. (2011). How much is enough? choosing ϵ for differential privacy. In *International Conference on Information Security*, pp. 325-340, Springer.

References

- [MAC'07]: Machanavajjhala, A., Kifer, D., Gehrke, J. and Venkatasubramanian, M. (2007). I-diversity: Privacy beyond k-anonymity. ACM Trans. Knowl. Discov. Data 1, 1.
- [NAL'15] Naldi, M., and D'Acquisto, G. (2015). Differential Privacy: An Estimation Theory-Based Method for Choosing Epsilon. arXiv preprint arXiv:1510.00917.
- **[NIS'18] Nissim, K., Steinke, T., Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., O'Brien, D.R. & Vadhan, S. (2018). Differential Privacy: A Primer for a Non-technical Audience. Vanderbilt Journal of Entertainment & Technology Law.**
- **[NIS'19]**
- [NIS'06]: Nissim, K., Raskhodnikova, S., & Smith, A. (2007). Smooth sensitivity and sampling in private data analysis. In Proceedings of the 39th ACM Symposium on Theory of Computing (STOC), pp. 75-84.
- [XU'12]: J. Xu, J., Zhang, Z., Xiao, X., Yang Y., and Yu, G. "Differentially private histogram publication," In proceedings of ICDE, 2012.
- [YAN'12]: Yang, Y., Zhang, Z., Miklau, G., Winslett, M., and Xiao, X. (2012). Differential privacy in data publication and analysis. In Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 601-606.
- [ZLI'16] Žliobaitė, I., & Custers, B. (2016). Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models. Artificial Intelligence and Law, pp. 1–19.