Data privacy: Introduction

Vicenç Torra

March 2025

Umeå University, Sweden

V. Torra (2022) A guide to data privacy, Springer

http://www.mdai.cat/dp

Outline

1. Introduction
   - A context
   - Privacy for machine learning and statistics
   - Two motivating examples
   - Concepts
2. Disclosure, risk measures, and privacy models
3. Privacy models
   - k-anonymity
   - differential privacy
4. Privacy for data: Microaggregation
5. Privacy for functions: DP
6. Summary

Introduction

A context:

Data-driven machine learning/statistical models

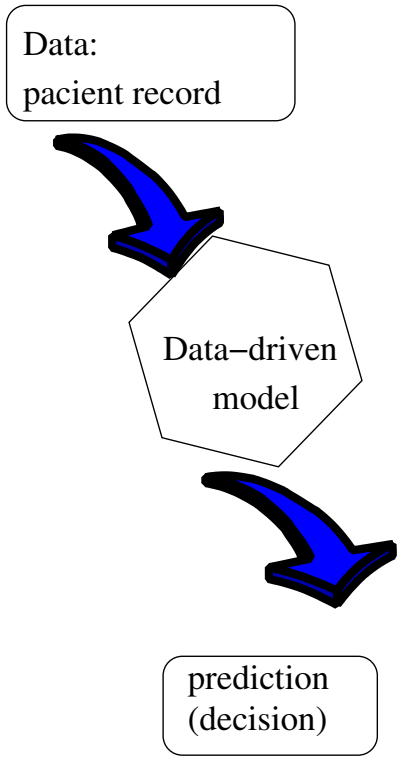
Prediction using (machine learning/statistical) models

- Data is collected to be used
  (otherwise, better not to collect them[1])

[1] Concept: Data minimization

Prediction using (machine learning/statistical) models

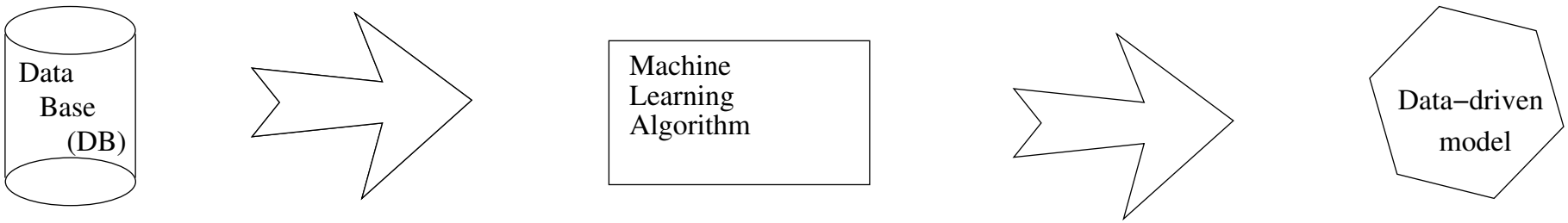
- Application of a model for decision making
  data $\Rightarrow$ prediction/decision



- Example: predict the length-of-stay at admission

Data-driven machine learning/statistical models

- From (huge) databases, build the "decision maker"
  - Use (logistic) regression, deep lerning, neural networks, . . . classification algorithms, decision trees, . . .



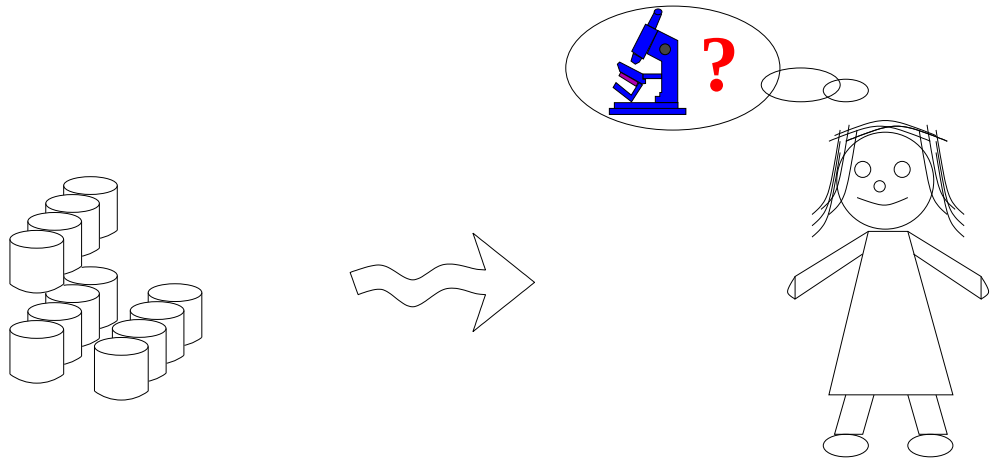
- Example: build a predictor from hospital historical data about length-of-stay at admission

Privacy for machine learning and statistics:

Data-driven machine learning/statistical models

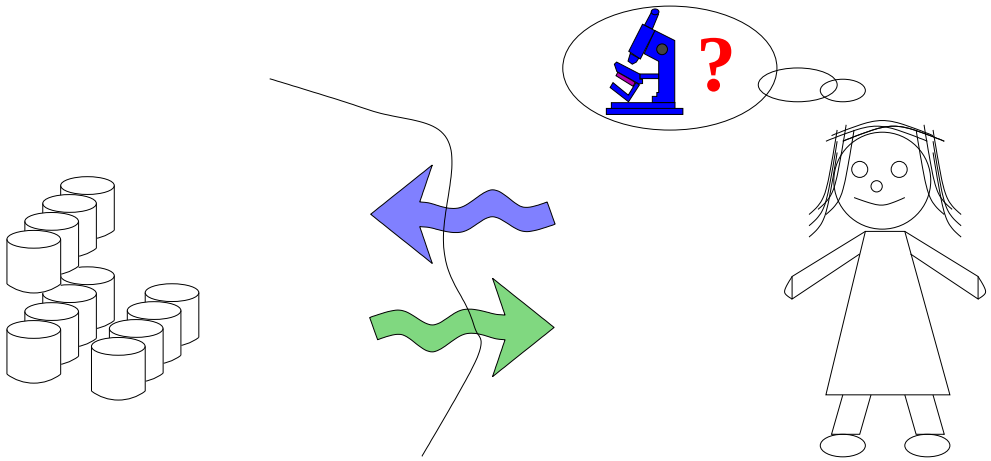
Data is sensitive

- Who/how is going to create this model (this "decision maker")?
- Case #1. Sharing (part of the data)



Data is sensitive

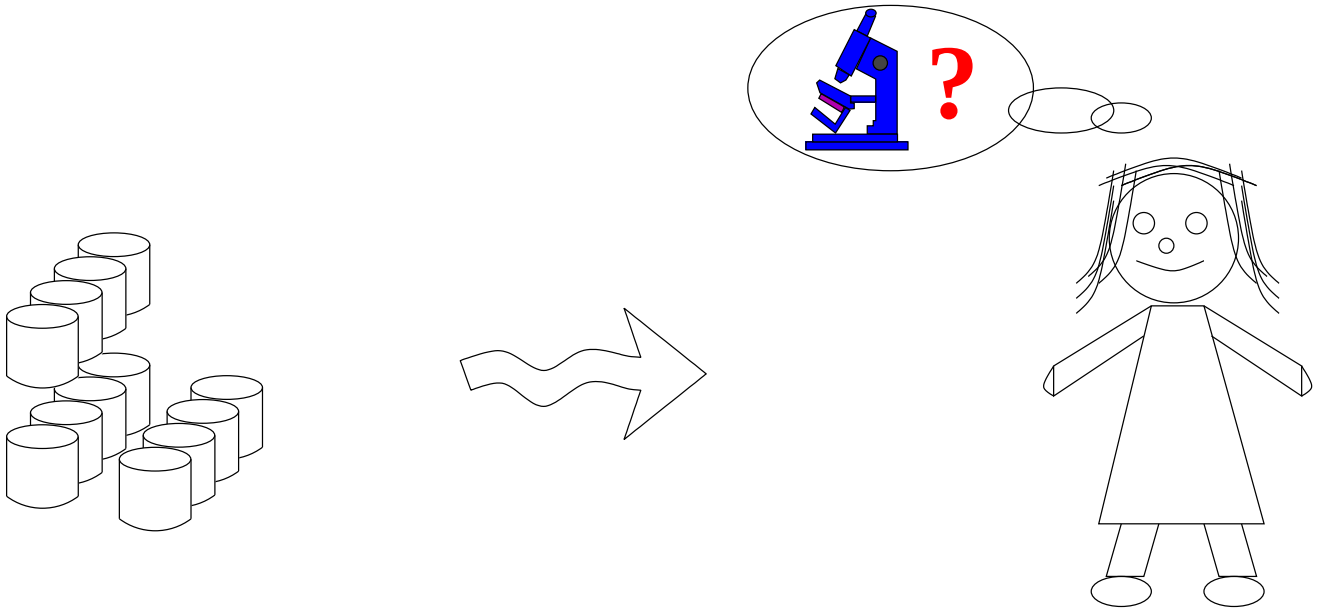
- Who/how is going to create this model (this "decision maker")?
- Case #2. Not sharing data, only querying data



Two motivating examples

Data is sensitive

- Data privacy: core
  - Someone needs to access to data to perform authorized analysis, but access to the data and the result of the analysis should avoid disclosure.



E.g., you are authorized to compute the average stay in a hospital, but maybe you are not authorized to see the length of stay of your neighbor.

Data is sensitive

- Case #1. Sharing (part of the data)
- Q: How different children ages and diagnoses affect this length of stay? Average length of stay is decreasing in the last years due to new hospital policies?
- Data: Existing database with previous admissions (2010–2019).
  To avoid disclosure a view of the DB restricting records to children born before 2019 and only providing for these records year of birth, town, year of admission, illness, and length of stay.

Data is sensitive

- Case #1. Sharing (part of the data)

Year birth	Year Admission	Town	Illness	Length stay (days)
2017	2019	Umeå	a	3
2015	2020	Umeå	b	2
2011	2020	Luleå	c	5
2017	2019	Luleå	a	2
2016	2020	Dorotea	b	4
2016	2020	Holmöns	d	2
2015	2019	Täfteå	e	4
2015	2019	Täfteå	e	4
2015	2018	Täfteå	e	4
2015	2018	Täfteå	e	4

- Is this data safe?

Data is sensitive

- Case #1. Sharing (part of the data)

Year birth	Year Admission	Town	Illness	Length stay (days)
2017	2019	Umeå	a	3
2015	2020	Umeå	b	2
2011	2020	Luleå	c	5
2017	2019	Luleå	a	2
2016	2020	Dorotea	b	4
2016	2020	Holmöns	d	2
2015	2019	Täfteå	e	4
2015	2019	Täfteå	e	4
2015	2018	Täfteå	e	4
2015	2018	Täfteå	e	4

- Is this data safe?

Holmöns 63, Täfteå 1383, Luleå 49123, Umeå 83249, Dorotea 2366

Data is sensitive: computation leads to disclosure

- Case #2. Sharing a computation.
  - Q: Mean income of admitted to hospital unit (e.g., psychiatric unit) for a given Town (Bunyola)?

Data is sensitive: computation leads to disclosure

- Case #2. Sharing a computation.
  - Q: Mean income of admitted to hospital unit (e.g., psychiatric unit) for a given Town (Bunyola)?
  - Mean income is not "personal data", is this ok ? NO!!:

Data is sensitive: computation leads to disclosure

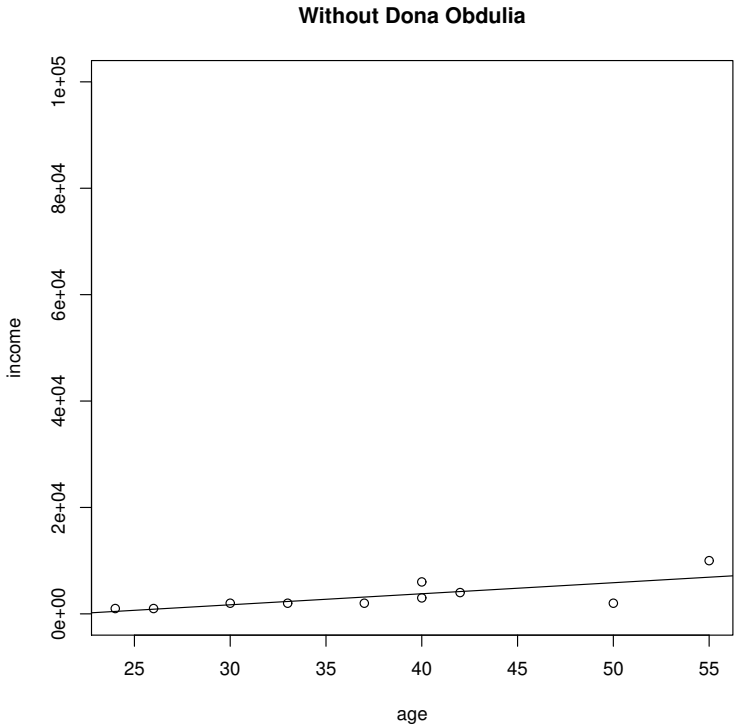
- Case #2. Sharing a computation.
  - Q: Mean income of admitted to hospital unit (e.g., psychiatric unit) for a given Town (Bunyola)?
  - Mean income is not "personal data", is this ok ? NO!!:
  - Example 1000 2000 3000 2000 1000 6000 2000 10000 2000 4000 $\Rightarrow$ mean = 3300

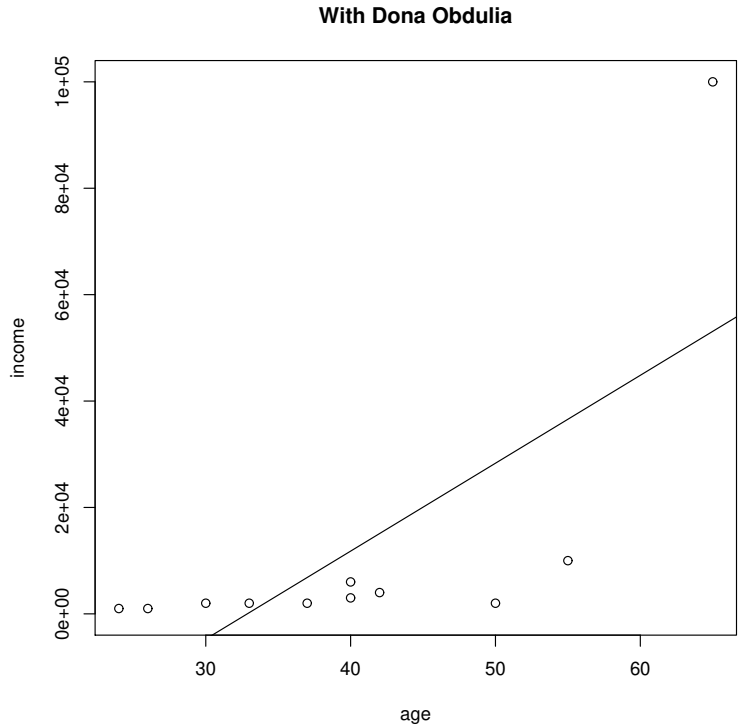
Data is sensitive: computation leads to disclosure

- Case #2. Sharing a computation.
  - Q: Mean income of admitted to hospital unit (e.g., psychiatric unit) for a given Town (Bunyola)?
  - Mean income is not "personal data", is this ok ? NO!!:
  - Example 1000 2000 3000 2000 1000 6000 2000 10000 2000 4000 $\Rightarrow$ mean = 3300
  - Adding Ms. Rich's salary 100,000 Eur/month: mean = 12090,90 !
    (a extremely high salary changes the mean significantly)
    $\Rightarrow$ We infer Ms. Rich from Town was attending the unit

Data is sensitive: computation leads to disclosure

- Case #2. Sharing a computation. Example 2





- Regression of income with respect to age with (right) and without (left) the record of Dona Obdúlia
  - income = -4524.2 + 207.5 age (without Ms. Rich = Dona Obdúlia)
  - income = -54307 + 1652 age (with Ms. Rich = Dona Obdúlia)

Data privacy: the core

Data is sensitive: Data privacy

- Data privacy
  - Information leakage/disclosure
    when there is some inference from the release
- Related fields / boundaries
  - Access control: who can access a database
    in general, we assume that access is granted
  - Security in communications
    in general, communications are protected
- So, data privacy: avoid disclosure because of inference

Data privacy: difficulties

Difficulties

- Difficulties: highly identifiable data
  - (Sweeney, 1997) on USA population
    - 87.1% (216 million/248 million) were likely made them unique based on
      5-digit ZIP, gender, date of birth,
    - 3.7% (9.1 million) had characteristics that were likely made them unique based on
      5-digit ZIP, gender, Month and year of birth.
  - Data from mobile devices / location data:
    - two positions can make you unique (home and working place)

Difficulties

- Difficulties: highly identifiable data and high dimensional data
  - AOL[2] and Netflix cases (search logs and movie ratings)
    $\Rightarrow$ User No. 4417749, hundreds of searches over a three-month period including queries 'landscapers in Lilburn, Ga'
    $\longrightarrow$ Thelma Arnold identified!
    $\Rightarrow$ individual users matched with film ratings on the Internet Movie Database.
  - Similar with credit card payments, shopping carts, ...

[2] http://www.nytimes.com/2006/08/09/technology/09aol.html

Difficulties

- Difficulties: highly identifiable data and high dimensional data
  - Data from people: In some cases, it is the combination of characteristics that make you unique. (Search logs, Market Basket Analysis)
    - Manga
    - K-pop music
    - Opera
    - IF Björklöven
    - ...

      $\rightarrow |\{ \text{Manga} \cap \text{k-pop} \cap \text{Opera} \cap \text{IF Björklöven} \}| = 1 = 1$

Concepts

Related fields

- Technical solutions from three different communities
  - Statistical disclosure control (SDC)
  - Privacy preserving data mining (PPDM)
  - Privacy enhancing technologies (PET)

Concepts

- Attacker, adversary, intruder
  - the set of entities working against some protection goal
  - increase their knowledge (e.g., facts, probabilities, . . .) on the items of interest (lol) (senders, receivers, messages, actions, etc.)

Concepts

- Disclosure. Attackers take advantage of observations to improve their knowledge on some confidential information about an IoI.
  $\Rightarrow$ SDC/PPDM: Observe DB, Δ knowledge of a particular subject (the respondent in a database)
  - Identity disclosure (entity disclosure). Linkability. Finding Mary in the database.
  - Attribute disclosure. Increase knowledge on Mary's salary.
    also: learning that someone is in the database, although not found.

Concepts

- Disclosure. Discussion.
  - Identity disclosure. Avoid.
  - Attribute disclosure. A more complex case. Some attribute disclosure is expected in data mining.

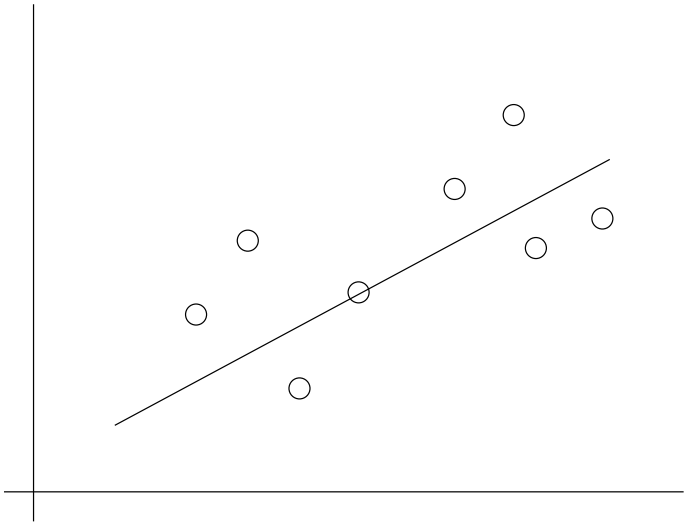
    At the other extreme, any improvement in our knowledge about an individual could be considered an intrusion. The latter is particularly likely to cause a problem for data mining, as the goal is to improve our knowledge. (J. Vaidya et al., 2006, p. 7.)

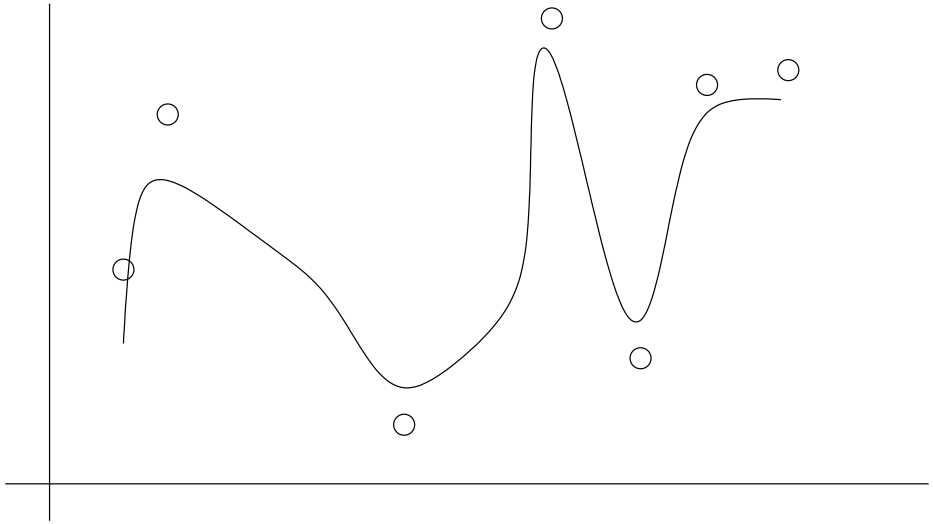
Concepts

- On the difficulty of measuring attribute disclosure:
  - In machine and statistical learning, models are expected to generalize data and avoid over-fitting. When a model generalizes correctly and there is no over-fitting, any inference for a particular individual x is due to general properties and not to its particularities. In contrast, bad generalization and over-fitting may imply that inferences are due to memorization and to learning particular features of certain records. When we require good data utility from a machine learning perspective, attribute disclosure should avoid detecting general information found in the data and focus on detecting these particular features of individuals.
  - This has connections with membership inference attacks, that, in short, try to detect records that are known to have been used in training a model, and they are detected because they are somehow distinguishable from more common ones.

Concepts

- If the model is a good generalization, (almost) coincidence may be ok.
- Compare:





Concepts

- Anonymity set. Anonymity of a subject means that the subject is not identifiable within a set of subjects, the anonymity set. Not distinguishable!

- Unlinkability. Unlinkability of two or more IoI, the attacker cannot sufficiently distinguish whether these IoIs are related or not.

Concepts

- Plausible deniability
  - I have nothing to do with this database, model, etc
  - Is this statement credible?
- For a database
  - at record level: This record is not mine!
  - at database level: I am not in this database!

Concepts

- Plausible deniability
  - I have nothing to do with this database, model, etc
  - Is this statement credible?
- For a database
  - at record level: This record is not mine!
  - at database level: I am not in this database!
- We will see that some privacy models provide guarantees for plausible deniability

Concepts

- Connections between plausible deniability and anonymity set
  - Plausible deniabilty: perspective of the individual
  - Anonymity set: perspective of the intruder

Concepts

- Transparency
  - DB is published: give details on how data has been produced. Description of any data protection process and parameters
  - Positive effect on data utility. Use information in data analysis.
  - Negative effect on risk. Intruders use the information to attack.

- The transparency principle in data privacy[3]

  Given a privacy model, a masking method should be compliant with this privacy model even if everything about the method is public knowledge. (Torra, 2017, p17)

[3] Similar to the Kerckhoffs's principle (Kerckhoffs, 1883) in cryptography: a cryptosystem should be secure even if everything about the system is public knowledge, except the key

Concepts

- Privacy by design (Ann Cavoukian, 2011)
  - Privacy "must ideally become an organization's default mode of operation" (Cavoukian, 2011) and thus, not something to be considered a posteriori. In this way, privacy requirements need to be specified, and then software and systems need to be engineered from the beginning taking these requirements into account.
  - *In the context of developing IT systems, this implies that privacy protection is a system requirement that must be treated like any other functional requirement. In particular, privacy protection (together with all other requirements) will determine the design and implementation of the system* (Hoepman, 2014)

Disclosure, risk measures, and privacy models

But first protection?

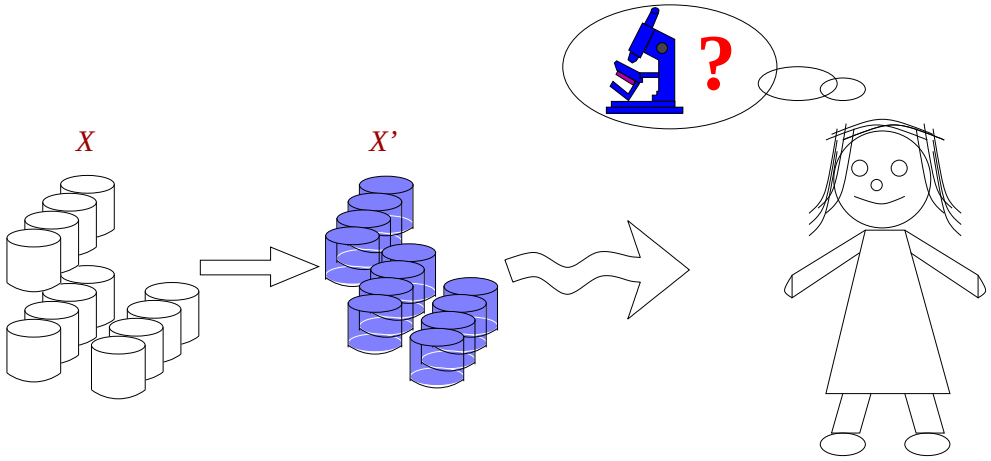
Disclosure, risk measures, and privacy models

- Disclosure, risk measures, and privacy models
- Protection mechanisms
  - Data protection mechanisms,
  - Privacy-preserving machine learning

Protection mechanisms need to be clearly disassociated of privacy models

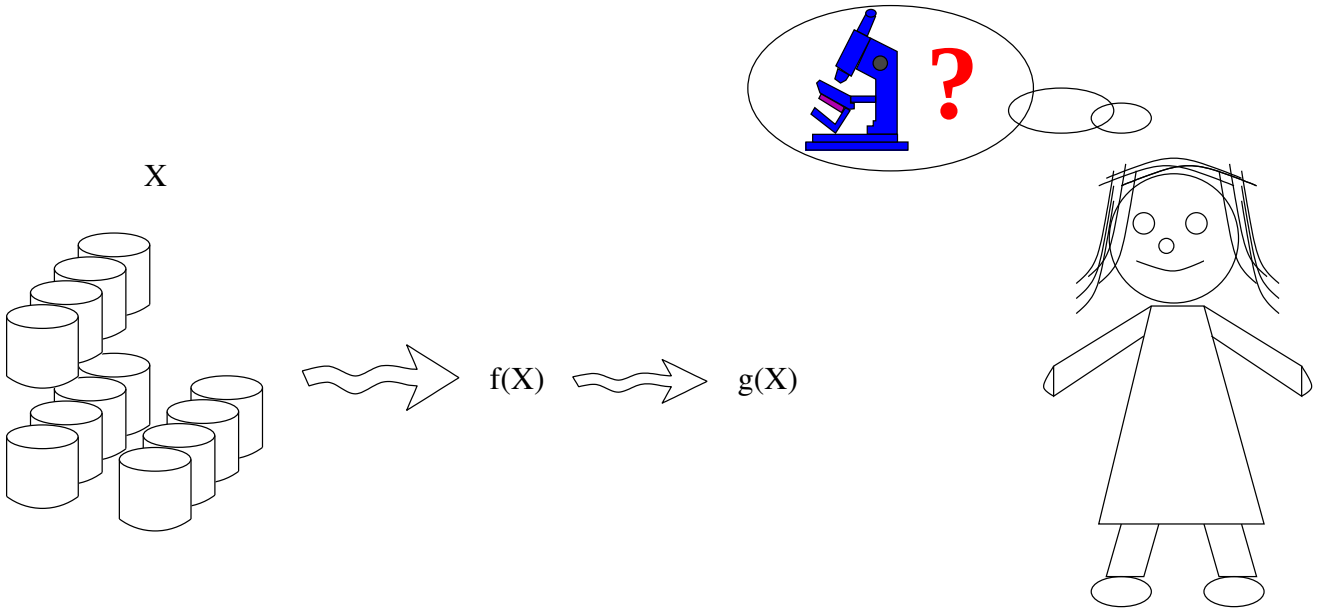
Disclosure, risk measures, and privacy models

- Privacy for data: data sharing, data publishing



Disclosure, risk measures, and privacy models

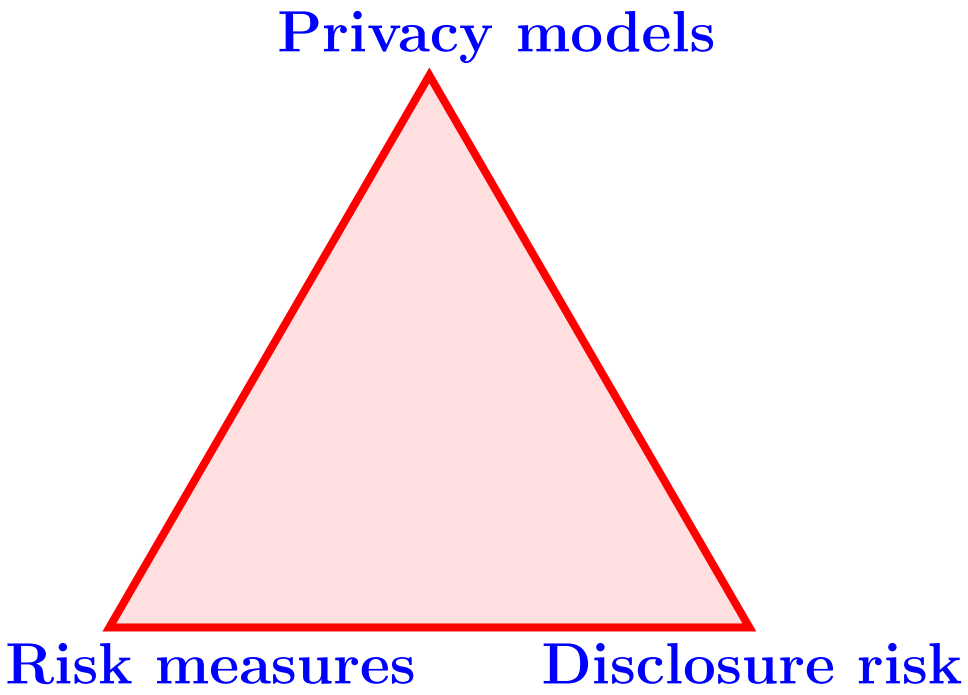
- Privacy for computations



Now, the risk et al. part

Disclosure, risk measures, and privacy models

- Three strongly related concepts
  (and as we will see linked to possible attacks)



Disclosure

Disclosure

- **Definition 3.1**
  - Disclosure takes place when intruders take advantage of the observation and analysis of a release to improve their knowledge on some item of interest.

- Release: data, statistics, data-driven machine learning model, output from a running model (e.g., an LLM)

Disclosure

- **Definition 3.1**
  - Disclosure takes place when intruders take advantage of the observation and analysis of a release to improve their knowledge on some item of interest.

- Release: data, statistics, data-driven machine learning model, output from a running model (e.g., an LLM)

- Intruder: the intruder attacks the release

Disclosure

- Disclosure. Intruders take advantage of observations to improve their knowledge
  - Identity disclosure (entity disclosure). Linkability.
    Finding Mary in the database.
  - Attribute disclosure. Increase knowledge.
    Learn Mary's salary. Increase precision on Mary's salary.

- Another dimension for disclosure: Boolean vs. measurable
  - Boolean: Disclosure either takes place or not.
    focus on one $performance$ measure. Minimize information loss
  - Measurable: Disclosure is a matter of degree that can be quantified.
    Some risk is permitted.
    multiobjetive optimization problem. Both performance and risk.

Disclosure

- Two dimensions. Privacy models / risk measures

	Attribute disclosure	Identity disclosure
Boolean	Differential privacy Result privacy Secure multiparty computation	k–Anonymity
Quantitative	Interval disclosure	Re–identification (record linkage) Uniqueness

Risk measures for attribute disclosure

Attribute disclosure: for a variable

- Original data X, and a data release $X' = \rho(X)$
  - Comparison between $V(x)$ and $V'(x)$?
  - Value associated to $V'(x)$ is too similar to $V(x)$?
  - Risk measure := proportion of too similar values

Attribute disclosure: Through Membership Inference Attacks

- For data-driven models m
  - Given x, was x in the training model of m?
    (Is my data used to build m?)
  - Idea:
    - We build a classifier mia

Attribute disclosure: Membership Inference Attacks

- For data-driven models m built from DB using A
- Notation:
  - $D^1, \ldots, D^k$: data sets for building shallow models (each D^i partitioned into training and testing D_i^{tr}, D_i^{te})
  - A: algorithm to build shallow models
- We train $sm_1, \ldots, sm_k$ (shallow models from $D^1, \ldots, D^k$)

Attribute disclosure: Membership Inference Attacks

Algorithm Model for membership inference attack: $mia(C_{tr}, A)$.

Data: D^i: data sets for building shallow models (each D^i partitioned into training and testing D_i^{tr}, D_i^{te}); A: algorithm to build shallow models

Result: Classifier for membership inference attack

begin

$sm_i = A(S_i^{tr})$ for all $i = 1, \ldots, k$

tuples $= \emptyset$

for $i = 1, \ldots, k$ **do**

forall the $x \in D_i^{tr}$ **do**

$tuples = tuples \cup \{(x, sm_i(x), training)\}$

end

forall the $x \in D_i^{te}$ **do**

$tuples = tuples \cup \{(x, sm_i(x), no - training)\}$

end

end

$mia =$ build-classifier($tuples$)

return mia

end

Attribute disclosure: Membership Inference Attacks

- Once we have the mia classifier, we define the membership inference attack attribute disclosure risk as

$$miaAR = \frac{|\{x|mia(x) = training\}|}{|X|}$$

(correctly identified members)

- In general, we can use performance measures (recall, precision, F1-score)

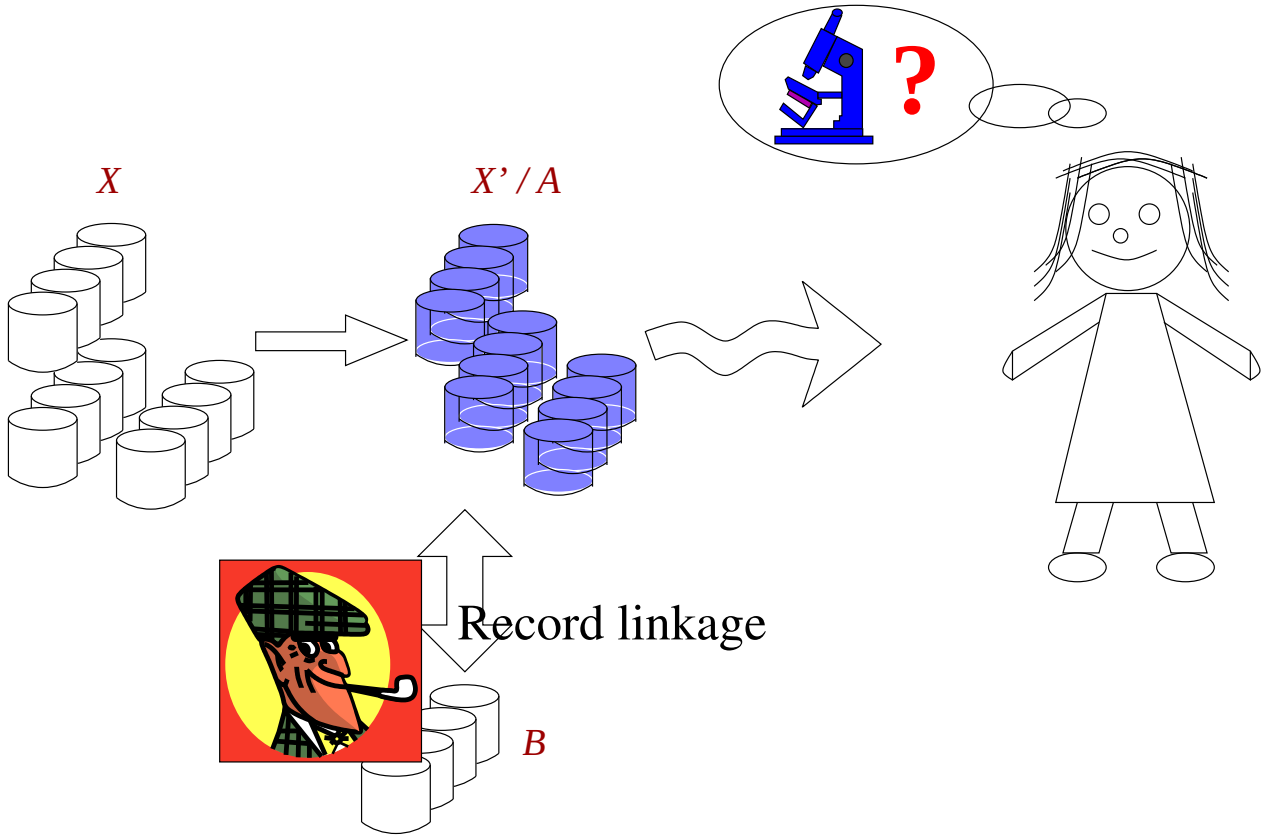
Risk measures for identity disclosure

Identity disclosure

- Re-identification. Estimation of correct re-identifications. Theoretically or empirically.
- Uniqueness. Probability that rare combinations in the protected data are also rare in the population.

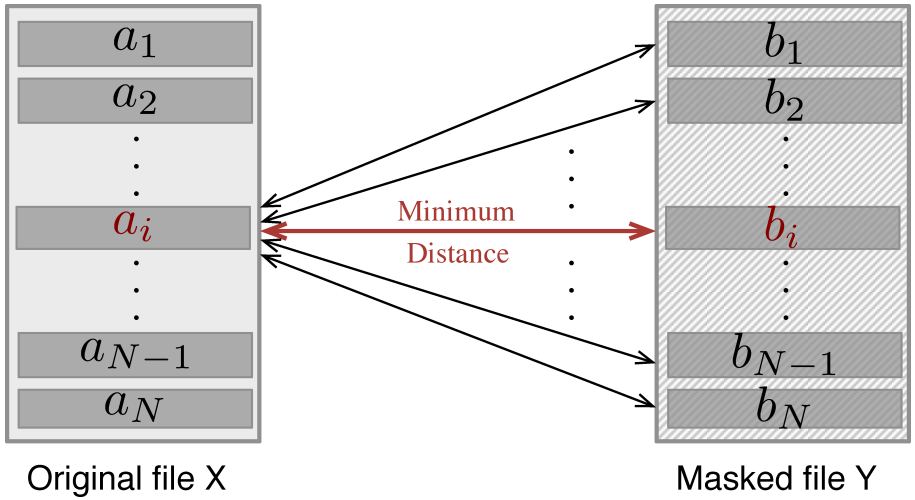
Identity disclosure

- **Privacy from re-identification**. Identity disclosure. Scenario:
  - A: File with the protected data set
  - B: File with the data from the intruder (subset of original X)



Identity disclosure

- Distance-based record linkage: $d(a, b)$ with $a \in A$ and $b \in B$.
  - Assign to the record at a minimum distance, ideally an intruder wants



- Compute and check
  - $b' = \arg\min_{b \in B} d(a, b)$
  - a_i linked to $b' = b_i$?

Identity disclosure

- **Re-identification**. Given $A = X' = \rho(X)$ and $B \subset X$, a measure:

$$Reid(B, A) = \frac{\sum_{b \in B} c(r(b), true(b))}{|B|}. \quad (1)$$

where

- $true : B \to A$, for each record b (of the intruder) returns the correct record for re-identification,
- $r : B \to A$, models the re-identification algorithm.
  Note: In order to make the definition general, we consider that r returns a probability distribution on A. That is, given a record b in B, it assigns to each record a in A a probability of matching.
- c a function, with $c(r(b), true(b))$ we evaluate the result for each record in $[0, 1]$.

Identity disclosure

A scenario for identity disclosure. Reidentification

- Flexible scenario for identity disclosure
  - A protected file using a masking method
  - B (intruder's) is a subset of the original file.

Identity disclosure

A scenario for identity disclosure. Reidentification

- Flexible scenario for identity disclosure
  - A protected file using a masking method
  - B (intruder's) is a subset of the original file.
    $\rightarrow$ intruder with information on only some individuals

Identity disclosure

A scenario for identity disclosure. Reidentification

- Flexible scenario for identity disclosure
  - A protected file using a masking method
  - B (intruder's) is a subset of the original file.
    $\rightarrow$ intruder with information on only some individuals
    $\rightarrow$ intruder with information on only some characteristics

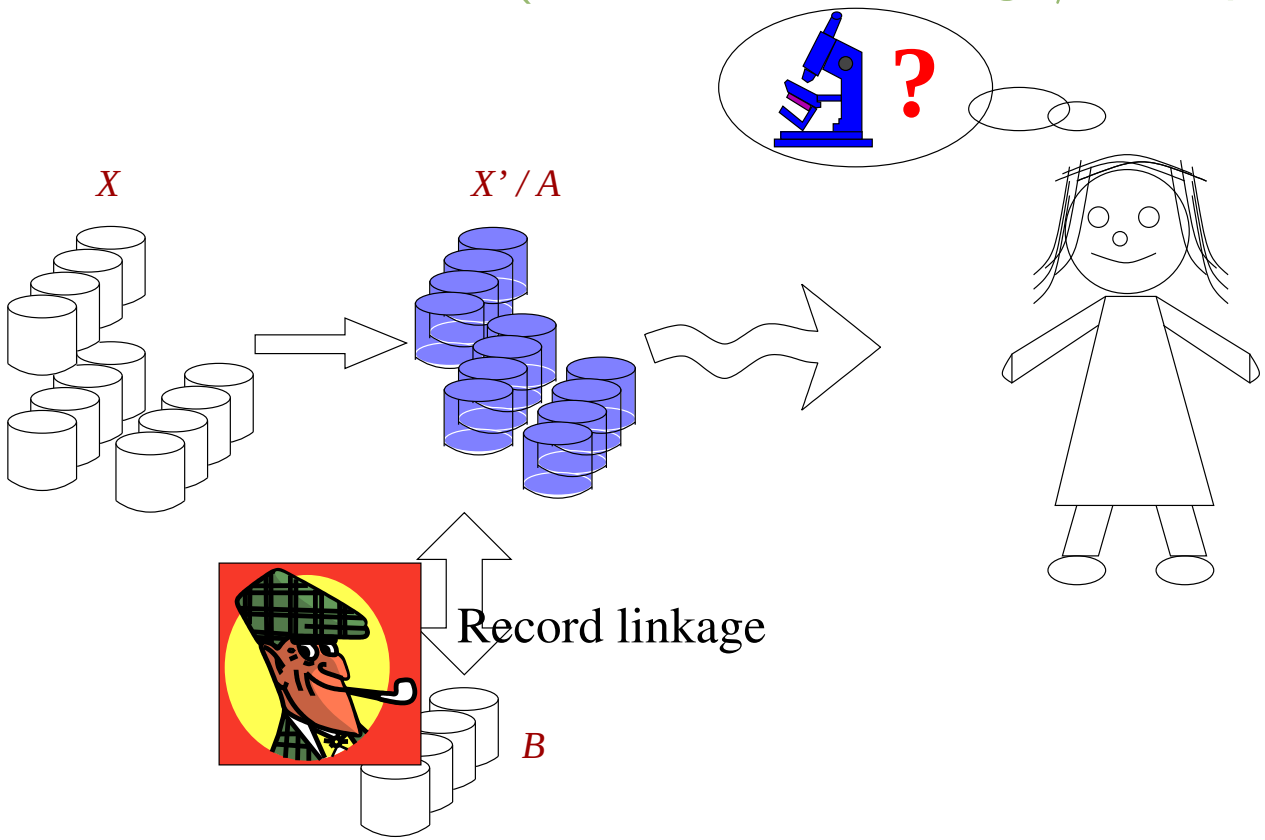
Identity disclosure

A scenario for identity disclosure. Reidentification

- Flexible scenario for identity disclosure
  - A protected file using a masking method
  - B (intruder's) is a subset of the original file.
    $\rightarrow$ intruder with information on only some individuals
    $\rightarrow$ intruder with information on only some characteristics
  - But also,
    - B with a schema different to the one of A (different attributes)
    - Other scenarios. E.g., synthetic data
    - Other type of data: graph data
      (reidentifying people in a social network)

Identity disclosure

- **Privacy from re-identification**. Worst-case scenario (maximum knowledge) to give upper bounds of risk:
  - transparency attacks (information on how data has been protected)
  - largest data set (original data)
  - best re-identification method (best record linkage/best parameters)



Identity disclosure

- **Privacy from re-identification**. Worst-case scenario.
  - ML for distance-based record linkage parameters. (A and B aligned)
  - Goal: as many correct reidentifications as possible.
  - Minimize K_i: minimize the number of records a_i that fail
- Formalization:

$$Minimize \sum_{i=1}^{N} K_i$$

$$Subject\ to:$$

$$\mathbb{C}_p(diff_1(a_i, b_j), \ldots, diff_n(a_i, b_j)) -$$

$$- \mathbb{C}_p(diff_1(a_i, b_i), \ldots, diff_n(a_i, b_i)) + CK_i > 0$$

$$K_i \in \{0, 1\}$$

Additional constraints according to $\mathbb{C}$

Privacy models

Definition

Privacy models

Definition

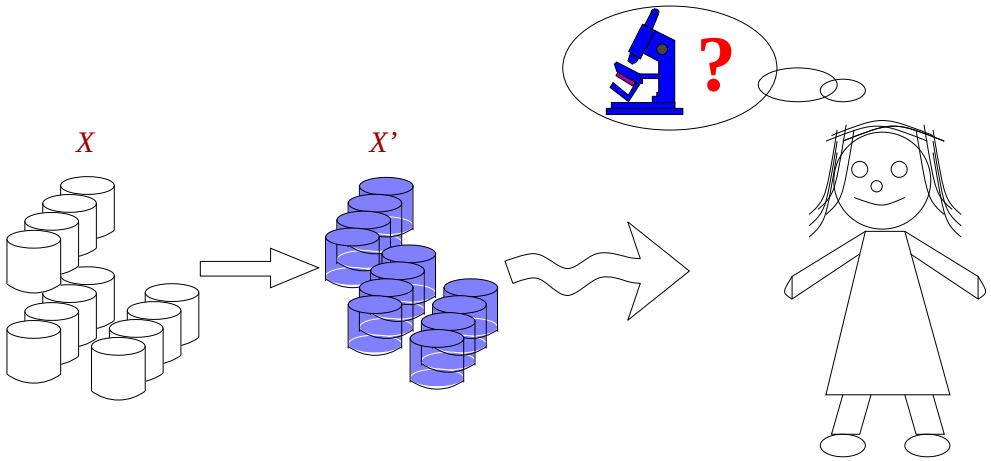
- A privacy model is a computational definition of privacy.

Summary of privacy models

Privacy models

Privacy models. Publish a DB

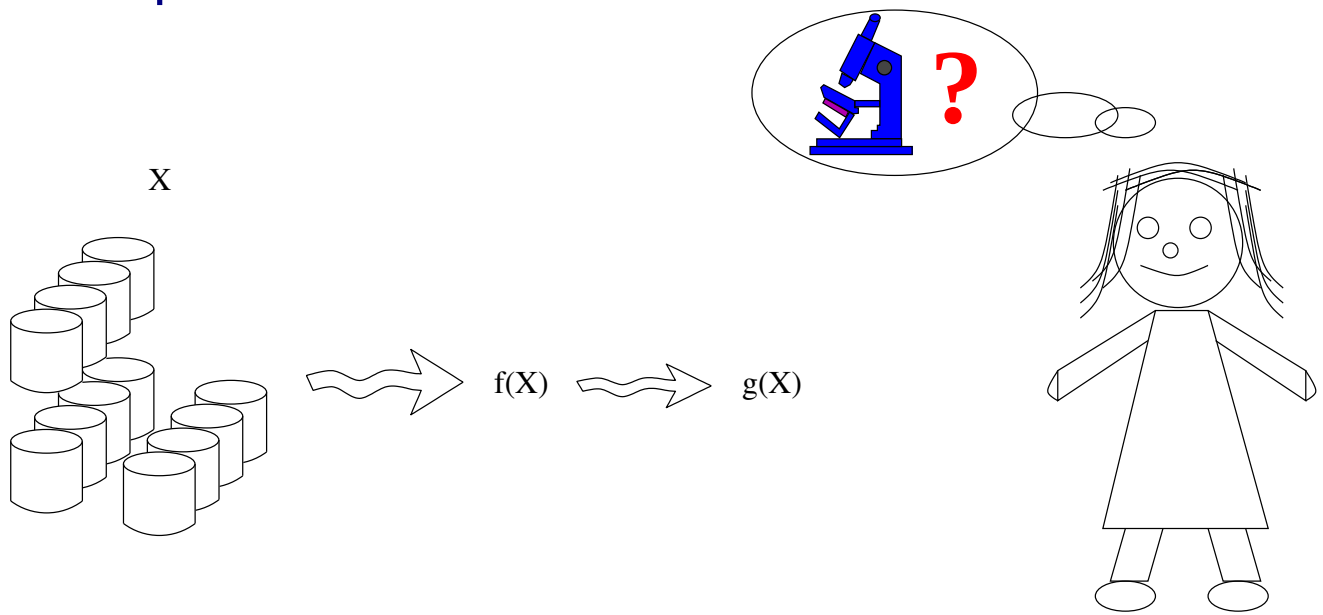
- **Reidentification privacy.** Avoid finding a record in a database.
- **k-Anonymity.** A record indistinguishable with $k-1$ other records.
- **k-Anonymity, l-diversity.** l possible categories
- **Interval disclosure.** The value for an attribute is outside an interval computed from the protected value: values different enough.
- **Result privacy.** We want to avoid some results when an algorithm is applied to a database.



Privacy models

Privacy models. Compute result

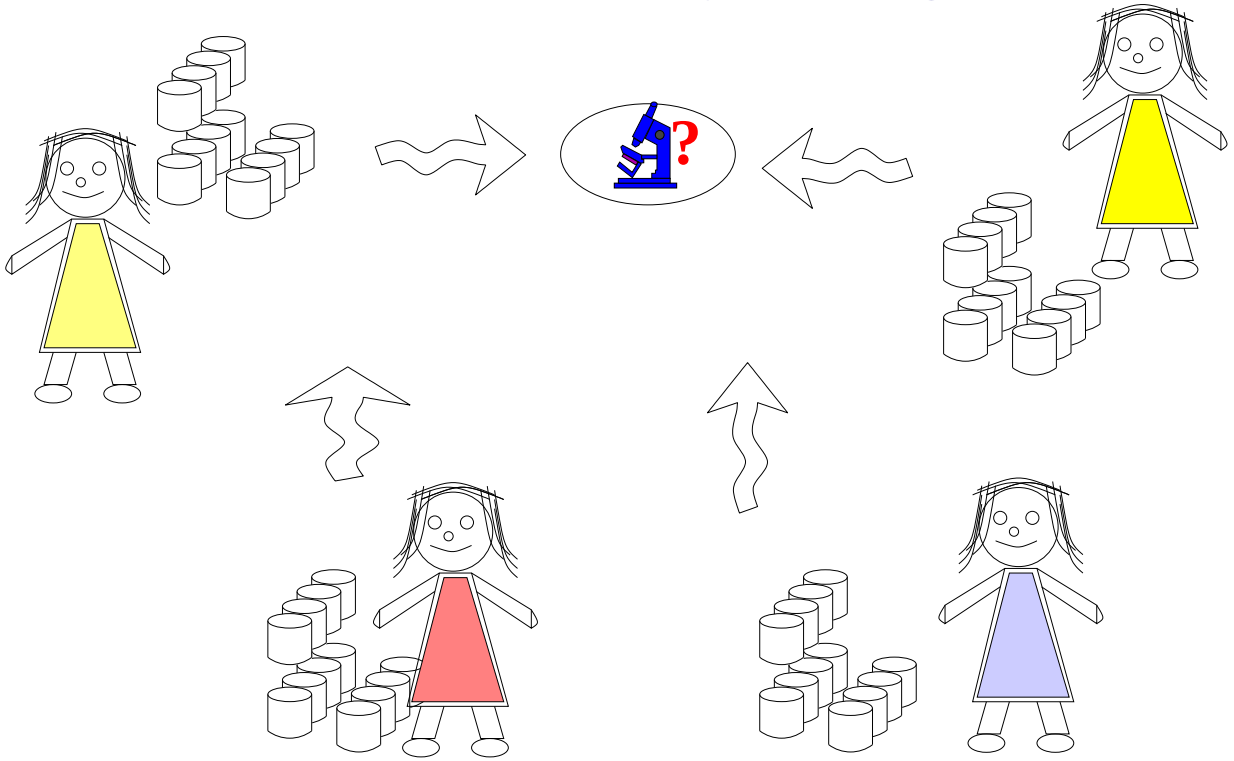
- **Differential privacy.** The output of a query to a database should not depend (much) on whether a record is in the database or not.
- **Integral privacy.** Inference on the databases. E.g., changes have been applied to a database.
- **Homomorphic encryption.** We want to avoid access to raw data and partial computations.



Privacy models

Privacy models. Compute / Share a result

- **Secure multiparty computation.** Several parties want to compute a function of their databases, but only sharing the result.



Privacy from re-identification

Privacy from re-identification

- A protected database A satisfies privacy from re-identification given intruder's knowledge B when

$$Reid(B, A) \leq r_{R1}$$

with a certain privacy level r_{R1} (e.g., $r_{R1} = 0.25$),

- or, alternatively (knows is correct, percentage)

$$KR.Reid(B, A) \leq (r_K, r_{R1})$$

with certain privacy levels r_K and r_{R1} (e.g., $r_K = 0$ and $r_{R1} = 0.5$).

k-Anonymity

k-Anonymity

Definition 3.4

- A database A satisfies k-anonymity with respect to a set of quasi-identifiers QI when the projection of A in this set QI results into a partition of DB in sets of at least k indistinguishable records.

City	Age	Illness
Barcelona	30	Cancer
Barcelona	30	Cancer
Tarragona	60	AIDS
Tarragona	60	AIDS

k-Anonymity

- Indistinguishability w.r.t. quasi-identifiers
- k-Anonymity and re-identification

$$KR.Reid(B, A) \leq (0, 1/k).$$

- Plausible deniability

k-Anonymity

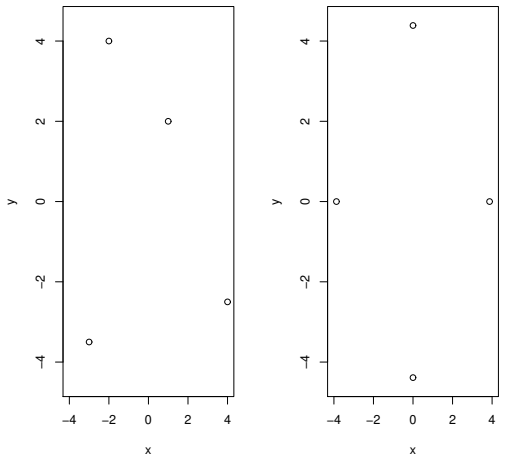
- Indistinguishability w.r.t. quasi-identifiers
- k-Anonymity and re-identification

$$KR.Reid(B, A) \leq (0, 1/k).$$

- Plausible deniability
  - at record level
  - but not at database level
- Records are independent

k-Anonymity

- k-confusion. Drop indistinguishability
  - Example
    - ▷ Original data: $X = \{(1,2),(-2,4),(4,-2),(-3,-4)\}$.
    - ▷ k-Anonymity: $X' = \{(0,0),(0,0),(0,0),(0,0)\}$.
    - ▷ k-Confusion: using $X'' = \{(x,0),(-x,0),(0,y),(0,-y)\}$, with standard deviations in X'' equal to the ones in X
      - $x = \sqrt{10}/\sqrt{2/3} = 3.872983$, $y = \sqrt{12.8333}/\sqrt{2/3} = 4.387476$



  - Discussion: k-confusion and re-identification

k-Anonymity

- Attacks
  - Homogeneity attack (external attack)
  - External knowledge attack (internal attack)
- These are attribute disclosure attacks
  - while k-anonymity is for identity disclosure
- Variations of k-anonymity to avoid attribute disclosure

k-Anonymity

- p-sensitive k-anonymity for $k > 1$ and $p \leq k$
  - if it satisfies k-anonymity and, for each group of records with the same combination of values for a set of quasi-identifiers, the number of distinct values for each confidential value is at least p (within the same group).

k-Anonymity

- p-sensitive k-anonymity for $k > 1$ and $p \leq k$
  - if it satisfies k-anonymity and, for each group of records with the same combination of values for a set of quasi-identifiers, the number of distinct values for each confidential value is at least p (within the same group).

- l-diversity
  - forces l different categories in each set. However, in this case, categories should have to be well-represented. Different meanings have been given to what well-represented means.

k-Anonymity

- t-closeness.
  - The distribution of the attribute in any k-anonymous subset of the database is similar to the one of the full database. Similarity: distance between the two distributions, distance below a given threshold t. The Earth Mover distance is used in the definition.

k-Anonymity

- k-anonymity and computational anonymity
  - Relaxation: not-all quasi-identifiers

    "We say that unconditional anonymity is theoretical anonymity. Computational anonymity is conditioned by the assumption that the adversary has some limitation. The limitations can be (...) restricted memory or knowledge." (Stokes (2012)).
  - A data set X satisfies (k, l)-anonymity if it is k-anonymous with respect to every subset of attributes of cardinality at most l.
    $\Rightarrow$ Intruder's knowledge limited to l attributes

- Example: (2,2)-anonymity

$$D = \{(a, b, e), (a, b, f), (c, d, e), (c, d, f),$$

$$(c, b, e), (c, b, f), (a, d, e), (a, d, f)\}.$$

Differential privacy

Differential privacy

- Computation-driven/single database
  - Privacy model: differential privacy[4]
  - We know the function/query to apply to the database: f
- Example:

  compute the mean of the attribute salary of the database for all those living in Town.

[4] There are other models as e.g. query auditing (determining if answering a query can lead to a privacy breach), and integral privacy

Differential privacy

- Differential privacy (Dwork, 2006).
  - Motivation:
    - ▷ the result of a query should not depend on the presence (or absence) of a particular individual
    - ▷ the impact of any individual in the output of the query is limited

  differential privacy ensures that the removal or addition of a single database item does not (substantially) affect the outcome of any analysis (Dwork, 2006)

Differential privacy

- Mathematical definition of differential privacy
  (in terms of a probability distribution on the range of the function/query)
  - A function K_q for a query q gives ϵ-differential privacy if for all data sets D_1 and D_2 differing in at most one element, and all $S \subseteq Range(K_q)$,

$$\frac{Pr[K_q(D_1) \in S]}{Pr[K_q(D_2) \in S]} \leq e^{\epsilon}.$$

  (with 0/0=1) or, equivalently,

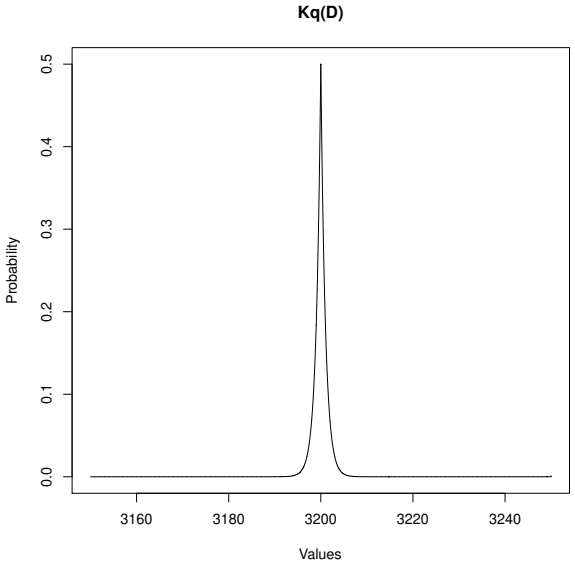
$$Pr[K_q(D_1) \in S] \leq e^{\epsilon} Pr[K_q(D_2) \in S].$$

- ϵ is the level of privacy required (privacy budget). The smaller the ϵ, the greater the privacy we have.

Differential privacy

- Differential privacy
  - A function K_q for a query q gives ϵ-differential privacy if . . .
    - $K_q(D)$ ~~is a constant. E.g.,~~ ~~$K_q(D) = 0$~~
    - $K_q(D)$ is a randomized version of $q(D)$:

$$K_q(D) = q(D) + and\ some\ appropriate\ noise$$



Differential privacy

- Properties
  - Plausible deniability: to an extend, in terms of ϵ

Differential privacy: Variations of differential privacy

- **Def. 3.17.** (ϵ, δ)-differential privacy (or δ-approximate ϵ-indistinguishability)
  - A function K_q for a query q gives (ϵ, δ)-differential privacy if for all data sets D_1 and D_2 differing in at most one element, and all $S \subseteq Range(K_q)$,

$$Pr[K_q(D_1) \in S] \leq e^{\epsilon} Pr[K_q(D_2) \in S] + \delta.$$

- Relaxes ϵ-DP, events with a probability smaller than δ for D_1 are still permited even if they do not occur in D_2.

Summary of privacy models

Summary

Privacy risk model/measure	Attribute disclosure	Identity disclosure	database release	query release	Boolean
Re-identification		X	X		Quantitative
Uniqueness		X	X		Quantitative
Result-driven	X		X		Boolean
k-Anonymity		X	X		Boolean
k-confusion		X	X		Boolean
k-concealment		X	X		Boolean
p-sensitive k-Anonymity	X	X	X		Boolean
k-Anonymity, l-diversity	X	X	X		Boolean
k-Anonymity, t-closeness	X	X	X		Boolean
Interval disclosure	X		X		Quantitative
Differential privacy	X			X	Boolean
Local differential privacy		X	X		Boolean
Integral privacy	X			X	Boolean
Homomorphic encryption	X			X	Boolean
Secure multiparty computation	X			X	Boolean

Privacy for data: Masking methods

Microaggregation

Microaggregation

- Informal definition. Small clusters are built for the data, and then each record is replaced by a representative.

Microaggregation

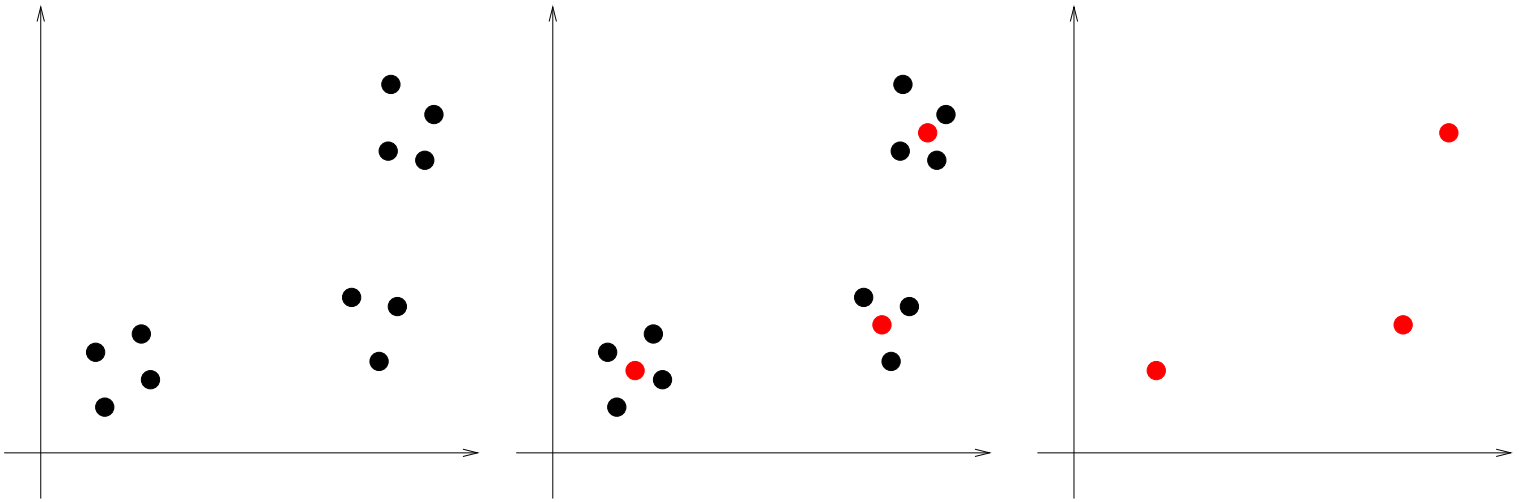
- Informal definition. Small clusters are built for the data, and then each record is replaced by a representative.
- Disclosure risk and information loss
  - Low disclosure is ensured requiring k records in each cluster
  - Low information loss is ensured as clusters are small

Microaggregation

- Operational definition. It is defined in terms of
  - **Partition.** Records are partitioned into several clusters, each of them consisting of at least k records.
  - **Aggregation.** For each of the clusters a representative (the centroid) is computed
  - **Replacement.** The original records are replaced by the representative of the cluster to which they belong to.

Microaggregation

- Graphical representation of the process.



Microaggregation

- Formalization. u_{ij} to describe the partition of the records in X. That is, $u_{ij} = 1$ if record j is assigned to the ith cluster. Let v_i be the representative of the ith cluster, then a general formulation of microaggregation with g clusters and a given k is as follows:

$$\begin{aligned} \text{Minimize} \quad & SSE = \sum_{i=1}^{g} \sum_{j=1}^{n} u_{ij} (d(x_j, v_i))^2 \\ \text{Subject to} \quad & \sum_{i=1}^{g} u_{ij} = 1 \text{ for all } j = 1, \ldots, n \\ & 2k \geq \sum_{j=1}^{n} u_{ij} \geq k \text{ for all } i = 1, \ldots, g \\ & u_{ij} \in \{0, 1\} \end{aligned}$$

Microaggregation

- Optimality
  - Polynomial solution when only one variable
  - Optimal solution is NP-hard for more than 2 variables
  - Heuristic methods have been developed:
    MDAV, Projected microaggregation

Microaggregation

- Heuristic approaches
  - usually follow the operational approach
    - ▷ **Build a partition.**
    - ▷ **Define an aggregation.** Mean of the records in the cluster
    - ▷ **Replacement.**

Microaggregation

- Multivariate case
  - When a file has several variables
    - ▷ Microaggregate all the variables at once
    - ▷ Microaggregate sets of variables
    - ▷ Microaggregate one variable at a time: individual ranking

Microaggregation

- Discussion and summary
  - The larger the k, the lower the risk, the larger the information loss
  - Microaggregation is related to k-anonymity:
    all variables together $\Rightarrow$ k-anonymity
  - It is easy to define microaggregation for any type of data
    - ▷ define distance,
    - ▷ define aggregation method (plurality rule - most frequent value)

    E.g, application to logs, sets of documents (via bags of words), graphs
    time series (different distances produce different effects)

Privacy for functions: Implementing differential privacy

Privacy model: definition

Differential privacy

- Recall: Mathematical definition of differential privacy (in terms of a probability distribution on the range of the function/query)
  - A function K_q for a query q gives ϵ-differential privacy if for all data sets D_1 and D_2 differing in at most one element, and all $S \subseteq Range(K_q)$,

$$\frac{Pr[K_q(D_1) \in S]}{Pr[K_q(D_2) \in S]} \leq e^{\epsilon}.$$

  (with 0/0=1) or, equivalently,

$$Pr[K_q(D_1) \in S] \leq e^{\epsilon} Pr[K_q(D_2) \in S].$$

- ϵ is the level of privacy required (privacy budget). The smaller the ϵ, the greater the privacy we have.

Differential privacy

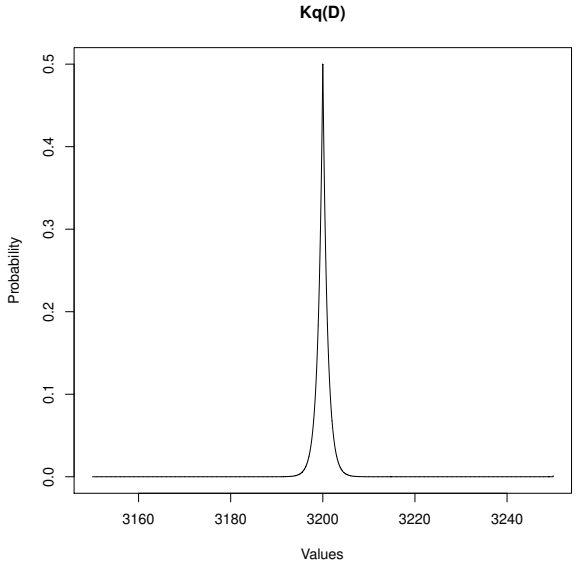
- Differential privacy: A KEY ELEMENT
  - for all data sets D_1 and D_2
    differing in at most one element

Understanding the definition:

Differential privacy for numerical data

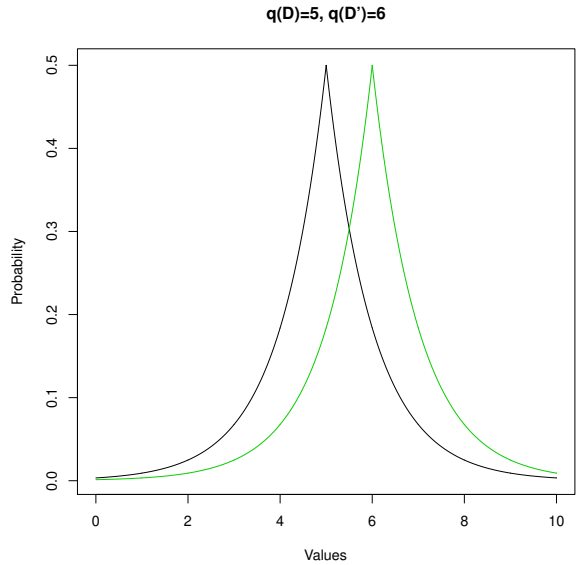
Differential privacy

- Differential privacy
  - A function K_q for a query q gives ϵ-differential privacy if . . .
    - $K_q(D)$ ~~is a constant. E.g.,~~
      $$~~K_q(D) = 0~~$$
    - $K_q(D)$ is a randomized version of $q(D)$:
      $$K_q(D) = q(D) + and\ some\ appropriate\ noise$$



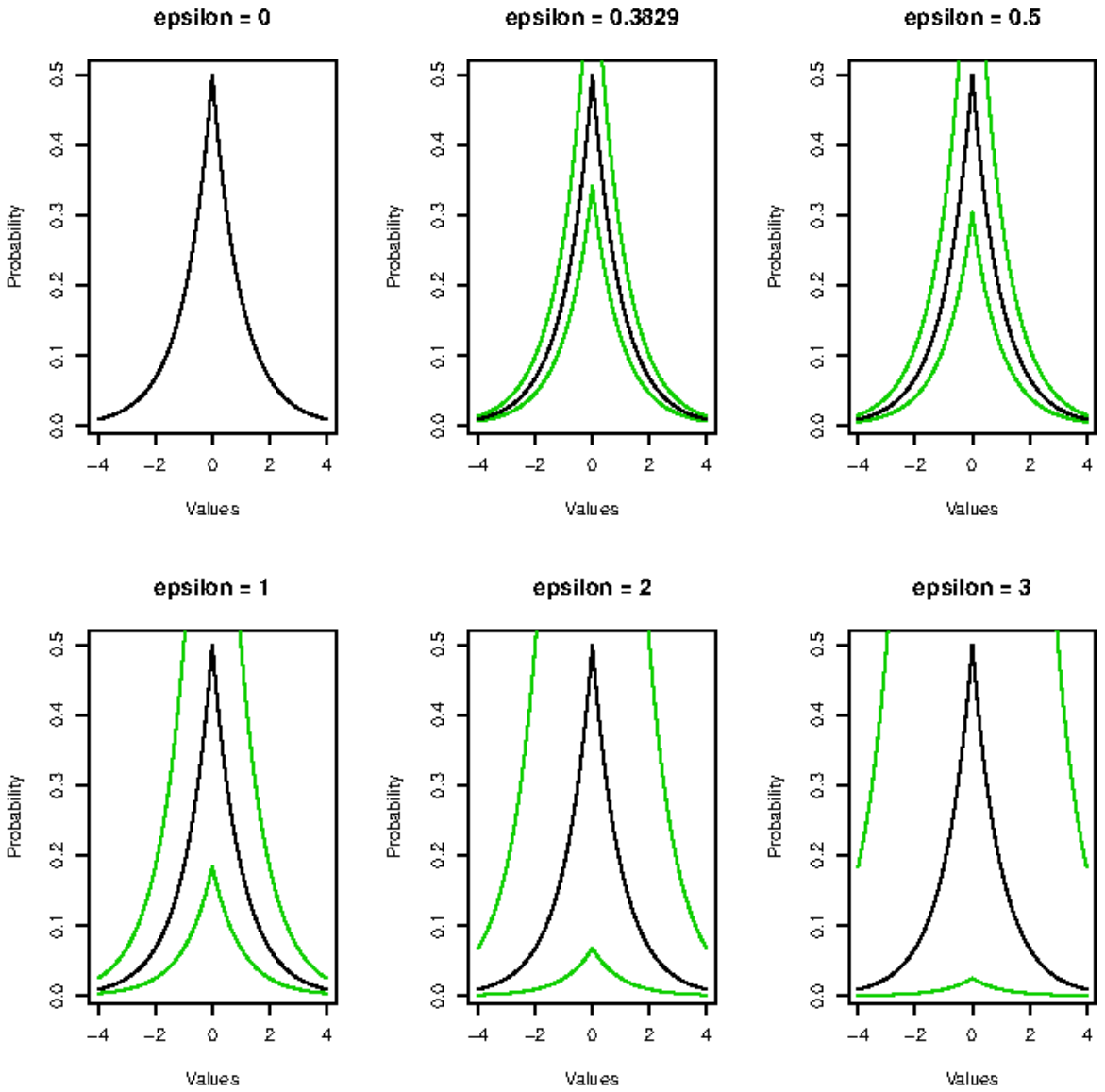
Differential privacy

- Differential privacy
  - $K_q(D)$ for a query q is a randomized version of $q(D)$
    - ▷ Given two neighbouring databases D and D'
      $K_q(D)$ and $K_q(D')$ should be similar enough . . .
  - Example with $q(D) = 5$ and $q(D') = 6$ and adding a Laplacian noise $L(0, 1)$



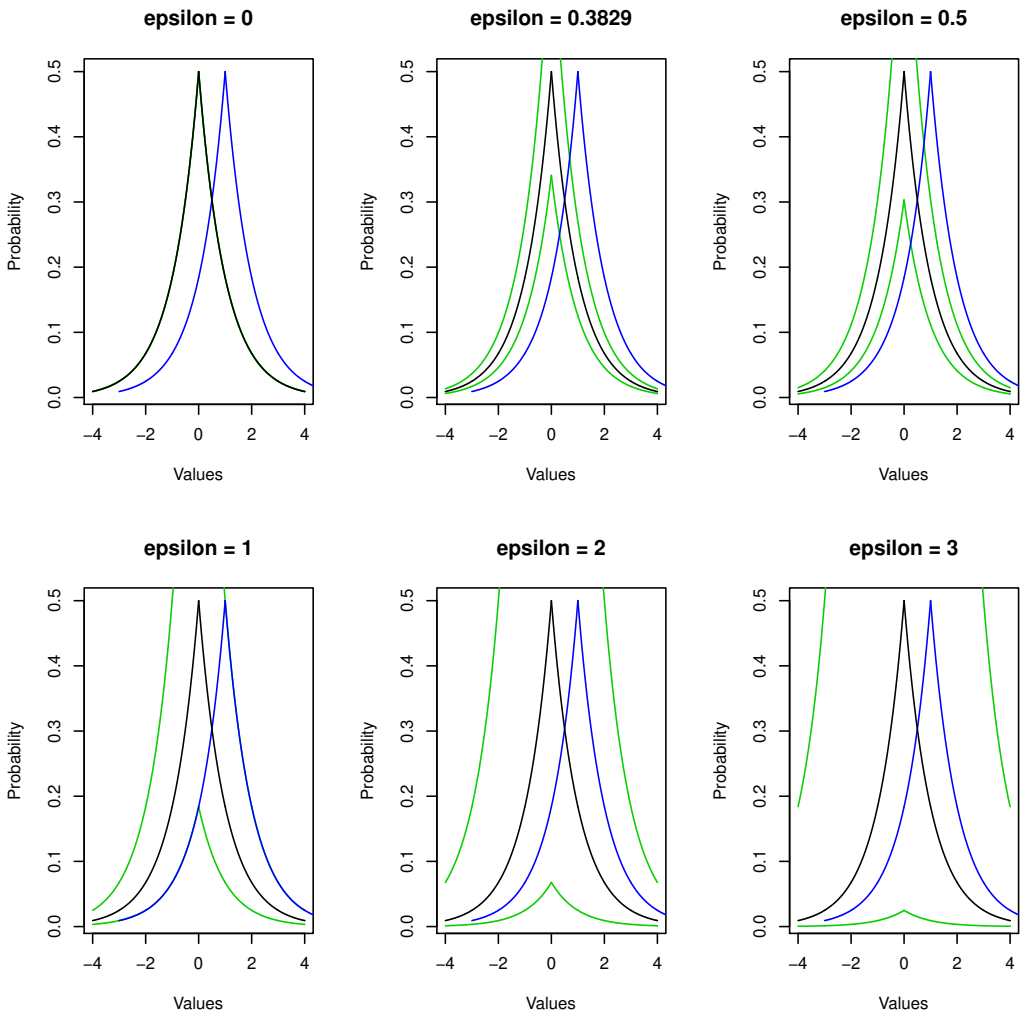
  - Let us compare different ϵ for noise following $L(0, 1)$ . . .

Differential privacy



Differential privacy

Is $0 + 1$ acceptable? I.e., are distributions L(0,1) L(1,1) similar enough?



Differential privacy

- These examples use the Laplace distribution $L(\mu, b)$.
  - I.e., probability density function:

$$f(x|\mu, b) = \frac{1}{2b} exp\left(-\frac{|x-\mu|}{b}\right)$$

    where
    - ▷ μ: location parameter
    - ▷ b: scale parameter (with $b > 0$)

- Properties
  - When $b = 1$, the function for $x > 0$ corresponds to the exponential distribution scaled by $1/2$.
  - Laplace has fatter tails than the normal distribution
  - When $\mu = 0$, for all translations $z \in \mathbb{R}$, $h(x+z)/h(x) \leq exp(|z|)$.

Differential privacy for numerical data: appropriate noise

Differential privacy

- Implementation of differential privacy for a numerical query.
  - $K_q(D)$ is a randomized version of $q(D)$:

$$K_q(D) = q(D) + and\ some\ appropriate\ noise$$

  - What is $and\ some\ appropriate\ noise$?

Differential privacy

- Implementation of differential privacy for a numerical query.
- Sensitivity of a query
  - Let $\mathcal{D}$ denote the space of all databases; let $q : \mathcal{D} \to \mathbb{R}^d$ be a query; then, the sensitivity of q is defined

$$\Delta_{\mathcal{D}}(q) = \max_{D,D' \in \mathcal{D}} ||q(D) - q(D')||_1.$$

where $||\cdot||_1$ is the L_1 norm, that is, $||(a_1, \ldots, a_d)||_1 = \sum_{i=1}^{d} |a_i|$.

Differential privacy

- Sensitivity of a query
  - Definition essentially meaningful when data has upper & lower bounds

An example: the case of the mean

Differential privacy

- Implementation of differential privacy: The case of the mean.
  - Sensitivity of the mean:

$$\Delta_{\mathcal{D}}(mean) = (max - min)/S$$

where $[min, max]$ is the range of the attribute, and S is the minimal cardinality of the set.

▷ If no assumption is made on the size of S: $\Delta_{\mathcal{D}}(mean) = (max - min)$

Differential privacy

- Implementation of differential privacy: The case of the mean.
  - Sensitivity of the mean:

$$\Delta_{\mathcal{D}}(mean) = (max - min)/S$$

where $[min, max]$ is the range of the attribute, and S is the minimal cardinality of the set.

▷ If no assumption is made on the size of S: $\Delta_{\mathcal{D}}(mean) = (max - min)$

  - **Proof.** Assume $S - 1$ values all in one extreme of the $[min, max]$ interval (say, max) and then we add one in the other extreme of the interval (say, min). Then, difference between the means is

$$\frac{(S-1)\cdot max}{S-1} - \frac{(S-1)\cdot max + min}{S} = max - \frac{(S-1)\cdot max}{S} - \frac{min}{S}$$

$$= \frac{S\cdot min - (S-1)\cdot max - min}{S} = \frac{max - min}{S}$$

Differential privacy

- Implementation of differential privacy for a numerical query.
  - Differential privacy via noise addition to the true response
  - Noise following a Laplace distribution $L(0, b)$ with mean equal to zero and scale parameter $b = \Delta(q)/\epsilon$.
    ($\Delta(q)$ is the sensitivity of the query)

- **Theorem.** The Laplace mechanism satisfies ϵ-differential privacy (proof Section 5.1.1, also later here)

Differential privacy

- Implementation of the Laplace mechanism

Algorithm Differential privacy for a numerical response $LM(D, q, \epsilon)$

Data: D: Database; q: query; ϵ: parameter of differential privacy

Result: Answer to the query q satisfying ϵ-differential privacy

begin

$a := q(D)$ with the original data

Compute $\Delta_{\mathcal{D}}(q)$, the sensitivity of the query for a space of databases D

Generate a random noise r from a $L(0, b)$ where $b = \Delta(q)/\epsilon$

return $a + r$

end

Differential privacy

- **Def. 3.17.** A function K_q for a query q gives (ϵ, δ)-differential privacy if for all data sets D_1 and D_2 differing in at most one element, and all $S \subseteq Range(K_q)$,

$$Pr[K_q(D_1) \in S] \leq e^{\epsilon} Pr[K_q(D_2) \in S] + \delta.$$

  - Interpretation. It relaxes ϵ-DP as events with a probability smaller than δ for D_1 are still permited even if they do not occur in D_2.

- **Prop. 5.1.** Algorithm 14 replacing the expression for b above by

$$b = \frac{\Delta(q)}{\epsilon - \log(1-\delta)}$$

satisfies (ϵ, δ)-differential privacy, for $\Delta(q)$ being the sensitivity of function q.

Differential privacy

- Example. $\Delta = 1$, $\epsilon = 0.5$
  - $\delta = 0.1$
    - $b = 1/(0.5 - log(0.9)) = 1.651908$
    - But, with $\epsilon - DP$, we only need $b = 1/0.5 = 2$

An example: the case of the mean

now with numbers

Differential privacy

- Implementation of differential privacy: The case of the mean.
  - Example[5]:
    - $D = \{1000, 2000, 3000, 2000, 1000, 6000, 2000, 10000, 2000, 4000\}$
      $\Rightarrow$ mean $= 3300$
    - Adding Ms. Rich's salary 100,000 Eur/month: mean $= 12090{,}90$!
      (a extremely high salary changes the mean significantly)
      $\Rightarrow$ We infer Ms. Rich from Town was attending the unit
  - $\Rightarrow$ Differential privacy to solve this problem

[5] Average wage in Ireland (2018): 38878 $\Rightarrow$ monthly 3239 Eur
https://www.frsrecruitment.com/blog/market-insights/average-wage-in-ireland/

Differential privacy

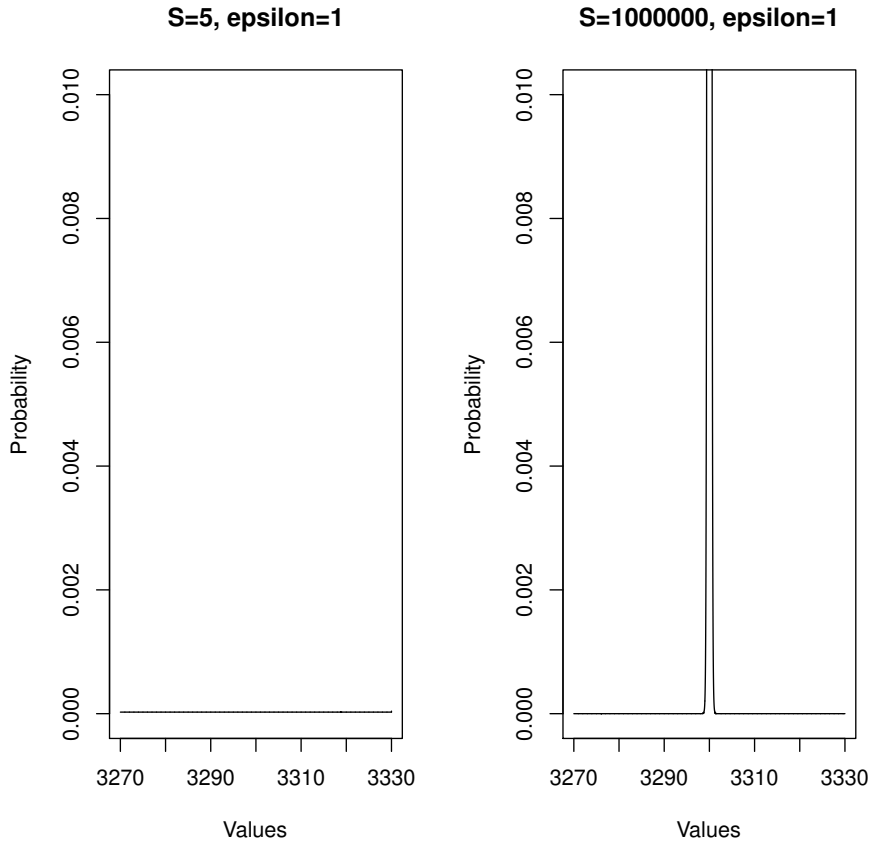
- Implementation of differential privacy: The case of the mean
  - Consider the mean salary
  - Range of salaries $[1000, 100000]$

Differential privacy

- Implementation of differential privacy: The case of the mean
  - Consider the mean salary
  - Range of salaries $[1000, 100000]$
- Compute for $\epsilon = 1$, assume that at least $S = 5$ records
  - sensitivity $\Delta_{\mathcal{D}}(q) = (max - min)/S = 19800$
  - scale parameter $b = 19800/1 = 19800$
  - For the database: (mean = 3300)
    D={1000, 2000, 3000, 2000, 1000, 6000, 2000, 10000, 2000, 4000}
  - Output: $K_{mean}(D) = 3300 + L(0, 19800)$
- Compute for $\epsilon = 1$, assume that at least $S = 10^6$ records
  - sensitivity $\Delta_{\mathcal{D}}(q) = (max - min)/S = 0.099$
  - scale parameter $b = 0.099/1 = 0.099$
  - For the database: (mean = 3300)
    D={1000, 2000, 3000, 2000, 1000, 6000, 2000, 10000, 2000, 4000}
  - Output: $K_{mean}(D) = 3300 + L(0, 0.099)$

Differential privacy

- Comparing
  - (i) $(S = 5, \epsilon = 1)$ $K_{mean}(D) = 3300 + L(0, 19800)$ and
  - (ii) $(S = 10^6, \epsilon = 1)$ $K_{mean}(D) = 3300 + L(0, 0.099)$



Differential privacy

- Laplace mechanism for differential privacy (numerical query)

$$K_q(D) = q(D) + L(0, \Delta(q)/\epsilon)$$

  - **Proposition.** For any function q, the Laplace mechanism satisfies ϵ-differential privacy.
    - **Proof.** Let $X \sim L(0, \Delta(q)/\epsilon)$, then the probability that the output is r for D is

$$\begin{aligned} Pr(K_q(D) = r) &= Pr(q(D) + X = r) = Pr(X = r - q(D)) \\ &= L(0, b)(r - q(D)) = \frac{1}{2b} exp\left(-\frac{|r - q(D)|}{b}\right) \end{aligned}$$

    - Similarly for D':

$$Pr(K_q(D') = r) = \cdots = \frac{1}{2b} exp\left(-\frac{|r - q(D')|}{b}\right)$$

- Now,

$$\frac{Pr(K_q(D)=r)}{Pr(K_q(D')=r)} = \frac{exp\left(-\frac{|r-q(D)|}{b}\right)}{exp\left(-\frac{|r-q(D')|}{b}\right)} = exp\left(\frac{|r-q(D')|-|r-q(D)|}{b}\right)$$

as $|a|-|b| \leq |a-b|$ (triangle inequality)

$$exp\left(\frac{|r-q(D')|-|r-q(D)|}{b}\right) \leq exp(\frac{|q(D)-q(D')|}{b}).$$

- As $\Delta_{\mathcal{D}}(q) = \max_{D,D' \in \mathcal{D}} ||q(D)-q(D')||_1$, then, $\Delta_{\mathcal{D}}(q) \geq ||q(D)-q(D')||_1$ for a pair of neighbouring D, D'. Therefore

$$exp\left(\frac{|q(D)-q(D')|}{b}\right) \leq exp\left(\frac{\Delta_{\mathcal{D}}(q)}{b}\right) = exp\left(\frac{\Delta_{\mathcal{D}}(q)}{\Delta_{\mathcal{D}}(q)/\epsilon}\right) = exp(\epsilon)$$

□

Means: Bounded and truncated

Differential privacy

- If the range of values is large, and sensitivity is large
  - Can we reduce the amount of noise?

Differential privacy: truncated mean

- Let us assume that the outcome is in the range $[mn, mx]$ (with $mn \neq mx$)
- Sensitivity is reduced: it is at most $[mx, mn]$
- Revisit the function mean, forcing output to be in $[mx, mn]$
  - We use

$$q'_{mn,mx}(x) = \begin{cases} mn & \text{if } x < mn \\ x & \text{if } mn \leq x \leq mx \\ mx & \text{if } mx < x \end{cases}$$

  - Then, define $q(D) = mean(D)$, and
  - $\tilde{q}(D) = q'_{mn,mx}(mean(D))$

Differential privacy

Algorithm truncated mean

Data: D: Database; S: minimum size of D; ϵ: parameter of differential privacy; mn, mx: real; max, min: real

Result: $truncated\text{-}mean$ satisfying ϵ-differential privacy and within the interval $[mn, mx]$

begin

$\Delta(mean) = \min((max - min)/S, mx - mn)$

$b = \Delta(mean)/\epsilon$

$m_0 = q'_{mn,mx}(mean(D))$ // A truncated mean

$m_1 = m_0 + L(0, b)$ // We add noise to the mean

$m_2 = q'(m_1)$ // Our output should also be in $[mn, mx]$

return (m_2)

end

Differential privacy

- Sensitivity of the truncated mean
  - If the range of the attribute is $[min, max]$,
  - S corresponds to the size of the database,
  - Sensitivity of the mean is:

$$\Delta_{\mathcal{D}}(mean) = (max - min)/S.$$

Differential privacy

- Sensitivity of the truncated mean
  - If the range of the attribute is $[min, max]$,
  - S corresponds to the size of the database,
  - Sensitivity of the mean is:

$$\Delta_{\mathcal{D}}(mean) = (max - min)/S.$$

  - But as truncated, and output in $[mn, mx]$

$$\Delta_{\mathcal{D}}(\tilde{q}) = \min((max - min)/S, (mx - mn)).$$

Differential privacy

- Sensitivity of the truncated mean. Example:
  - $[mn, mx] = [2000, 4000]$ and with $S = 5$
  - Sensitivity

$$\begin{aligned}\Delta_{\mathcal{D}}(\tilde{q}) &= \min((max - min)/5, (mx - mn)) \\ &= \min((1000000 - 1000)/5, (4000 - 2000)) = 2000.\end{aligned}$$

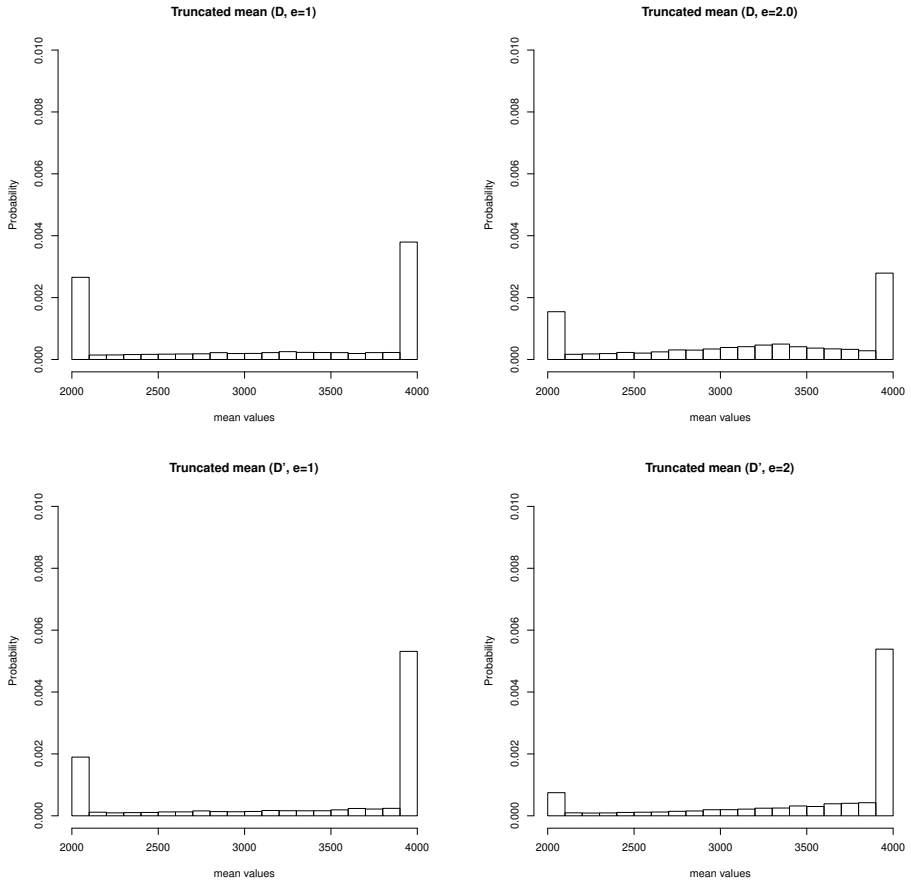
    When we apply $\tilde{q}$ to

    1000, 2000, 3000, 2000, 1000, 6000, 2000, 10000, 2000, 4000

    we have that the real mean is $mean = 3300$.
  - So, $L(0, b)$
    - For $\epsilon = 1$, we have $b = \Delta_{\mathcal{D}}/\epsilon = 2000/1 = 2000$
    - For $\epsilon' = 0.4$, we have $b' = \Delta_{\mathcal{D}}/\epsilon' = 2000/0.4 = 5000$,
    - For $\epsilon'' = 2$, we have $b'' = \Delta_{\mathcal{D}}/\epsilon'' = 2000/2 = 1000$.

Differential privacy

- Truncated mean, $\epsilon = 1$ and $\epsilon'' = 2$
  - D and $D' = D \cup \{\text{Dona Obdúlia's}\}$ with income 1000000,
  - Figures: applying $\tilde{q}$ 10000 times.



Differential privacy

- Alternatives
  - If output is in $D = [mn, mx]$, use a bounded Laplace distribution

$$L'(x; \mu, b) = \begin{cases} 0 & \text{if } x \notin D \\ \frac{1}{C_q}\frac{1}{2b}exp\left(-\frac{|x-\mu|}{b}\right) & \text{if } x \in D \end{cases}$$

    - ▷ ϵ-DP guaranteed for sensitivity $(mx - mn)$ and $b = \Delta(q)/\epsilon$. Not otherwise.
  - Use the Laplace mechanism and re-draw the mechanism until a value in D is obtained.

Composition theorems

Differential privacy: Composition theorems

- **Sequential composition.** $q_1, \ldots, q_n$ with $\epsilon_1, \ldots, \epsilon_n$ all applied to X provide $\epsilon = \sum_{i=1}^{n} \epsilon_i$ differential privacy
  - Example #1. Apply mean and variance to X
  - Example #2. Apply mean 5 times: $\epsilon = 5 \cdot \epsilon'$
- **Parallel composition.** $q_1, \ldots, q_n$ with $\epsilon_1, \ldots, \epsilon_n$ each applied to a disjoint X_i provide $\epsilon = \max_{i=1}^{n} \epsilon_i$ differential privacy
  - Max ϵ_i means smallest protection
  - Example: mean income of different towns
- **Post-processing.** q with ϵ applied to X, and q' applied to the result of q, then $q'(q(X))$ provides ϵ differential privacy
  - Example. Compute mean, then change currency

Differential privacy

- Properties of differential privacy
  - On the ϵ:
    - Small ϵ, more privacy, more noise into the solution
    - Large ϵ, less privacy, less noise into the solution
  - On the sensitivity:
    - Small sensitivity, less noise for achieving the same privacy
    - Large sensitivity, more noise for achieving the same privacy
  - Discussion here is for a single query (with privacy budget ϵ). Multiple queries (even multiple applications of the same query) need special treatment. E.g., additional privacy budget.
  - Randomness via e.g. Laplace means that any number can be selected. Including e.g. negative ones for salaries. Special treatment may be necessary.
  - Implementations for other type of functions
    - The exponential mechanism for non-numerical queries
    - Differential privacy for machine learning and statistical models

Computing histograms

Histograms

- Histogram: frequency of a set of items for a set of buckets (or bins).
- Differential privacy: Key aspects
  - Absolutely relevant whether the set of buckets is predefined or not.
    - Buckets are defined independently of the computation of the histogram, or they are built somehow from the data.
  - $B = \{b_1, \ldots, b_b\}$ be a set of b buckets,
  - D database
  - $c_D(b_i)$ counts for each bucket $i = 1, \ldots, b$
- Given D_1 and D_2 that differ in a single record,
  sensitivity of the two corresponding histograms is one

Histograms

- ϵ-DP histogram
  - $c'_D(b_i) = c_D(b_i) + r_i$ with r_i following $L(0, b)$
  - $b = \Delta(q)/\epsilon = 1/\epsilon$

Histograms

- Example: ϵ-DP histogram
  - Buckets: $b_1 = [1000, 1999]$, $b_2 = [2000, 2999]$, $b_3 = [3000, 3999]$
  - Data:
    $D = \{1234, 1300, 1233, 1250, 1284, 2000, 2300, 2044, 2573,$
    $2745, 2853, 2483, 3633, 3182, 3274, 3935\}$
  - Histogram$(D) = (5, 7, 4)$
  - Draw r_1, r_2, r_3 (independently) from $L(0, b)$ with $\epsilon = 1$, so, $b = 1$
  - Say, $r = (0.7534844, -0.6143575, -1.5725160)$
  - ϵ-DP histogram:
    $histogram(B, D) = (c'_D(b_1), c'_D(b_2), c'_D(b_3)) = (5, 7, 4) + r$
    $(5.753484, 6.385643, 2.427484)$

Histograms and composition

- Use composition to compute the mean from an histogram
  - D a database,
  - $B = \{b_1, \ldots, b_b\}$ buckets with range $b_i = [b_{in}, b_{ix})$
  - $c(b_i)$ counts
  - $m(b_i)$ the mean value of the interval.

- Approximate average as follows:
  - $mean(B, histogram(B, D)) = \frac{\sum_{i=1}^{b} c(b_i)m(b_i)}{\sum_{i=1}^{b} c(b_i)}$.

- Approximate average:
  the larger the buckets, the less acurate the mean

Histograms and composition

- Use composition to compute the mean from an histogram
  - D, $b_1 = [1000, 2000)$, $b_2 = [2000, 3000)$, $b_3 = [3000, 4000)$
  - histogram $(5, 7, 4)$
  - mean:
    $$mean(B, histogram(B, D)) = \frac{\sum_{i=1}^{b} c(b_i)m(b_i)}{\sum_{i=1}^{b} c(b_i)} = \frac{5 \cdot 1500 + 7 \cdot 2500 + 4 \cdot 3500}{5+7+4}$$
  - output: 2437.5

- Compare this result with the mean of D which is 2332.688 (i.e., non-DP)

- Other discretizations, another result!

- NOTE: We are here working with the histogram, but all computations are with non-private histograms this is not ϵ-differentially private

Histograms and composition

- Now, differentially private, using composition
  - **Step 1.** Compute $c = histogram(B, D)$
  - **Step 2.** Produce a differentially private histogram c'
  - **Step 3.** $mean(B, c') = \frac{\sum_{i=1}^{b} c'(b_i) m(b_i)}{\sum_{i=1}^{b} c'(b_i)}$
- If c' is ϵ-DP, mean(c') is also ϵ-DP (composition theorems)
- Example.
  - Using DP-histogram $(5.753484,\ 6.385643,\ 2.427484)$

$$\begin{aligned} mean(B, c') &= \frac{\sum_{i=1}^{b} c'(b_i) m(b_i)}{\sum_{i=1}^{b} c'(b_i)} \\ &= \frac{5.753484 \cdot 1500 + 6.385643 \cdot 2500 + 2.427484 \cdot 3500}{5.753484 + 6.385643 + 2.427484} \\ &= 2271.67 \end{aligned}$$

Histograms and composition

- Histograms and domain
  - We can apply this approach to compute the mean for any database.
  - We need buckets to span over the whole range of incomes.
    - So, if we consider the incomes in the range [1000, 100000] as when Dona Obdúlia was in the database, we need either a large bucket (with e.g., all incomes larger than 10000) or a large number of buckets.
  - This will have effects on the output.

Differential privacy: Categorical data

Categorical data

- Categorical output: $C = \{c_1, \ldots, c_c\}$
- Differential privacy, same definition applies
  - A function K_q for a query q gives ϵ-differential privacy if for all data sets D_1 and D_2 differing in at most one element, and all $S \subseteq Range(K_q)$,

$$\frac{Pr[K_q(D_1) \in S]}{Pr[K_q(D_2) \in S]} \leq e^{\epsilon}.$$

(with 0/0=1) or, equivalently,

$$Pr[K_q(D_1) \in S] \leq e^{\epsilon} Pr[K_q(D_2) \in S].$$

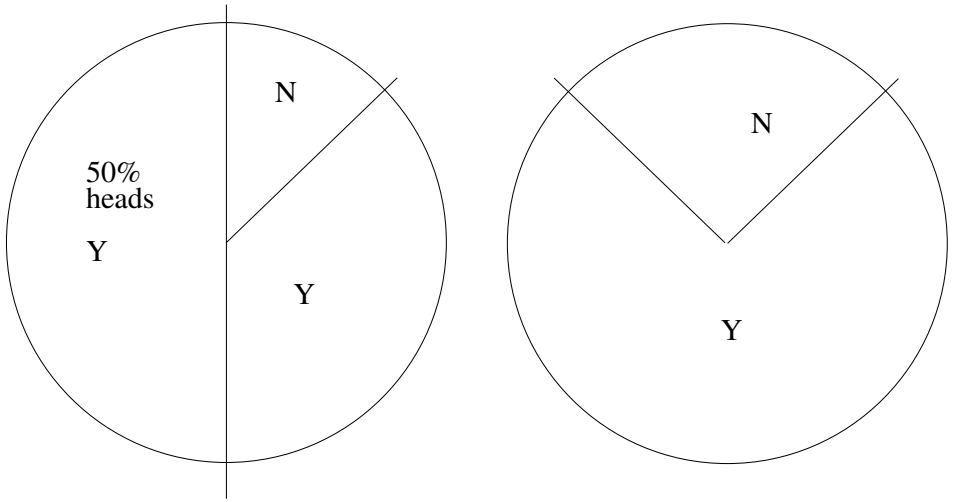
- Differential privacy using randomized response

Categorical data

- Randomized response:
  introduced for sensitive questions (Warner, 1965)
  - Categorical output, 2 outcomes: $C = \{Yes, No\}$
  - Example:
    - Have you consumed drugs this week? / Is your car now exceeding the speed limit?
  - Implementation:
    - toss a coin
    - if heads, return Yes
    - if tails, return the true answer

Categorical data

- Given all answers, we can estimate the true proportion
  - True proportion of Yes: p_N, True proportion of No: p_N.
    - Naturally, $p_Y = 1 - p_N$
    - r: proportion of answered No
    - Then, $p_N = 2 * r$, so, $p_Y = 1 - 2 * r$
  - Graphically



Categorical data: Example

- Given all answers, we can estimate the true proportion
- Example
  - We ask 100 people about their drug consumption
  - We get 45 Nos, and 55 Yes
- Answer

Categorical data: Example

- Given all answers, we can estimate the true proportion
- Example
  - We ask 100 people about their drug consumption
  - We get 45 Nos, and 55 Yes
- Answer
  - 50 answered Yes by default
  - so, $55 - 50 = 5$ answered a true yes
  - So, real yes was $2 * 5 = 10$
  - And true No was $r = 45$, So, total No is $r = 2 * 45 = 90$.

Categorical data

- In general,
  - probability p of returning right answer
  - probability p' of returning Y when false answer, $1 - p'$ of N

Categorical data

- **Algorithm** randomized response: $rr(f(X), p, p')$
  Data: $f(X)$: the true outcome of the query; p, p': probability in [0,1]

  Result: Randomized response for $f(X)$ with probabilities p, p'

  begin
    $r :=$ random number in [0,1] according to a uniform distribution
    if $r < p$ **then**
      return $f(X)$
    else
      $r' :=$ random number in [0,1] according to a uniform distribution
      if $r' < p'$ **then**
        return Y
      else
        return N
      end
    end
  end

Categorical data

- In general,
  - probability p of returning right answer
  - probability p' of returning Y when false answer, $1-p'$ of N
- π true proportion of Yes, o observed proportion of Yes

$$o = p * \pi + (1-p) * p'.$$

- So, given observed proportion o of Yes, we estimate π:

$$\hat{\pi} = (o - (1-p) * p')/p.$$

Categorical data

- In general, but with an example
  - probability $p = 0.5$ of returning right answer
  - probability $p' = 0.75$ of returning Y when false answer, $1 - p'$ of N
- We compute (assuming $\pi = 0.1$ Yes, as in the previous example)
  - π true proportion of Yes, o observed proportion of Yes

$$o = p * \pi + (1 - p) * p' \ = 1/2 * 0.1 + 1/2 * 3/4 = 0.425$$

  - So, given observed proportion o of Yes, we estimate π:

$$\hat{\pi} = (o - (1 - p) * p')/p \ = (0.425 - (1 - 0.5) * 3/4)/0.5 = 0.1$$

Categorical data

- In general, but with an example
  - probability $p = 0.5$ of returning right answer
  - probability $p' = 0.75$ of returning Y when false answer, $1 - p'$ of N
- We compute (assuming $\pi = 0.1$ Yes, as in the previous example)
  - π true proportion of Yes, o observed proportion of Yes

$$o = p * \pi + (1 - p) * p' \ = 1/2 * 0.1 + 1/2 * 3/4 = 0.425$$

  - So, given observed proportion o of Yes, we estimate π:

$$\hat{\pi} = (o - (1 - p) * p')/p \ = (0.425 - (1 - 0.5) * 3/4)/0.5 = 0.1$$

- However, in general, the larger the noise (i.e., p is very small), the more difficult to recover π: observe, $p = 0$. (Warner, 1965)

Differential privacy: general case with multiple categories

Categorical data

- General case: $C = \{c_1, \ldots, c_c\}$.
  - Randomized response, with a probability distribution for each c_i
  - from c_i to c_j

$$P(c_i, c_j) = P(X' = c_j | X = c_i).$$

Categorical data

- General case: $C = \{c_1, \ldots, c_c\}$.
  - Randomized response, with a probability distribution for each c_i
  - from c_i to c_j

$$P(c_i, c_j) = P(X' = c_j | X = c_i).$$

  - Naturally, for each c_i we have

$$P(X' = c_1 | X = c_i), \ldots, P(X' = c_c | X = c_i)$$

for all c_i it holds $\sum_j P(c_i, c_j) = \sum_j P(X' = c_j | X = c_i) = 1$.

Categorical data

- General case: $C = \{c_1, \ldots, c_c\}$.
  - Randomized response, with a probability distribution for each c_i
  - from c_i to c_j

$$P(c_i, c_j) = P(X' = c_j | X = c_i).$$

  - Naturally, for each c_i we have

$$P(X' = c_1 | X = c_i), \ldots, P(X' = c_c | X = c_i)$$

    for all c_i it holds $\sum_j P(c_i, c_j) = \sum_j P(X' = c_j | X = c_i) = 1$.
  - $P(c_i, c_j)$ a transition matrix P where the rows add to one
  - This is (like) PRAM

Categorical data

- General case: $C = \{c_1, \ldots, c_c\}$.

- **Algorithm** randomized response via PRAM: $rrPRAM(c, P)$
  Data: c: the true outcome of the query; P: transition matrix

  Result: Randomized response for c according to transition matrix P

  begin

  $r :=$ random number in [0,1] according to a uniform distribution
  Select k_0 in $\{1, \ldots, c\}$ such that
  $$\sum_{k=1}^{k_0-1} P(c' = c_i, |C = c) < r \leq \sum_{k=1}^{k_0} P(c' = c_i, |C = c)$$
  return c_{k_0}

  end

Categorical data

- General case: $C = \{c_1, \ldots, c_c\}$, true proportions?
  - After protection we observe: $o = (o_1, \ldots, o_c)$
  - but, the true response was $\pi = (\pi_1, \ldots, \pi_c)$
    here π_k is the proportion of respondents of class c_k
  - How to compute π from o?

Categorical data

- General case: $C = \{c_1, \ldots, c_c\}$, true proportions?
  - After protection we observe: $o = (o_1, \ldots, o_c)$
  - but, the true response was $\pi = (\pi_1, \ldots, \pi_c)$
    here π_k is the proportion of respondents of class c_k
  - How to compute π from o?
  - We know o from π:

$$o_j = \sum_{i=1}^{c} \pi_i P(X' = c_j | X = c_i)$$

    in matrix form:

$$o = P\pi$$

  - So, we can estimate

$$\hat{\pi} = P^{-1} o$$

Categorical data

- Randomized response = PRAM
  - The approach discussed here corresponds to PRAM
  - While PRAM assumes that we have the database available, Randomized response often considers local data being transmitted

Categorical data

- Randomized response = PRAM
  - The approach discussed here corresponds to PRAM
  - While PRAM assumes that we have the database available, Randomized response often considers local data being transmitted
  - i.e., local differential privacy

Appropriate noise: categorical data

Categorical data

- Local differential privacy, reminder, and rewriting
  - D_1 and D_2 are single records or categories

$$\frac{Pr[K_q(D_1) \in S]}{Pr[K_q(D_2) \in S]} \leq e^{\epsilon}.$$

Categorical data

- Local differential privacy, reminder, and rewriting
  - D_1 and D_2 are single records or categories

$$\frac{Pr[K_q(D_1) \in S]}{Pr[K_q(D_2) \in S]} \leq e^{\epsilon}.$$

  - with categories

$$\frac{Pr[K_q(c_i) = c_c]}{Pr[K_q(c_j) = c_c]} \leq e^{\epsilon}.$$

Categorical data

- Local differential privacy, reminder, and rewriting
  - D_1 and D_2 are single records or categories

$$\frac{Pr[K_q(D_1) \in S]}{Pr[K_q(D_2) \in S]} \leq e^{\epsilon}.$$

  - with categories

$$\frac{Pr[K_q(c_i) = c_c]}{Pr[K_q(c_j) = c_c]} \leq e^{\epsilon}.$$

  - and, in PRAM-like / randomized-response like

$$\frac{P(X' = c_c|c_i)}{P(X' = c_c|c_j)} \leq e^{\epsilon}.$$

Categorical data

- What is the appropriate noise ? (given ϵ)
- Assumptions on the matrix:
  - All categories same probability of being modified
    for all c_i, c_j we have $P(X' = c_i|c_i) = P(X' = c_j|c_j)$.
  - Non-diagonal values are all equal
    $P(X' = c_i|c_j) = P(X' = c_k|c_l)$ for all $i \neq j$, $k \neq l$
  - We assume $P(X' = c_i|c_i) > P(X' = c_j|c_i)$ for all $i \neq j$
- Summary, matrix of this form

$$\begin{pmatrix} q_d & q & \dots & q \\ q & q_d & \dots & q \\ \dots & & & \dots \\ q & q & \dots & q_d \end{pmatrix} \tag{2}$$

with $q_d = P(X' = c_i|c_i)$ for all i, and $q = P(X' = c_j|c_i)$ for $j \neq i$.

Categorical data

- What is the appropriate noise ? (given ϵ)
  - Probabilities after masking, we had c_i ?
    $$(P(X' = c_1|c_i), ..., P(X' = c_c|c_i)).$$
  - Probabilities after masking, we had c_j ?
    $$(P(X' = c_1|c_j), ..., P(X' = c_c|c_j)).$$

Categorical data

- What is the appropriate noise ? (given ϵ)
  - Probabilities after masking, we had c_i ?
    $$(P(X' = c_1|c_i), ..., P(X' = c_c|c_i)).$$
  - Probabilities after masking, we had c_j ?
    $$(P(X' = c_1|c_j), ..., P(X' = c_c|c_j)).$$
  - We require local ϵ-differential privacy, so we need
    $$P(X' = c_1|c_i)/P(X' = c_1|c_j) \leq e^\epsilon, \ldots, P(X' = c_c|c_i)/P(X' = c_c|c_j)) \leq e^\epsilon,$$
    This means
    $$\max_{k=1}^{c} P(X' = c_k|c_i)/P(X' = c_k|c_j) \leq e^\epsilon$$

Categorical data

- What is the appropriate noise ? (given ϵ)
  - Probabilities after masking, we had c_i ?
    $$(P(X' = c_1|c_i), ..., P(X' = c_c|c_i)).$$
  - Probabilities after masking, we had c_j ?
    $$(P(X' = c_1|c_j), ..., P(X' = c_c|c_j)).$$
  - We require local ϵ-differential privacy, so we need
    $$P(X' = c_1|c_i)/P(X' = c_1|c_j) \leq e^\epsilon, \ldots, P(X' = c_c|c_i)/P(X' = c_c|c_j)) \leq e^\epsilon,$$
    This means
    $$\max_{k=1}^{c} P(X' = c_k|c_i)/P(X' = c_k|c_j) \leq e^\epsilon$$
  - We assumed $P(X' = c_i|c_i)$ the largest value in a row, and all non-diagonal values are the same, so, maximum is obtained for $k = i$.

Categorical data

- What is the appropriate noise ? (given ϵ)
  - Probabilities after masking, we had c_i ?
    $$(P(X' = c_1|c_i), ..., P(X' = c_c|c_i)).$$
  - Probabilities after masking, we had c_j ?
    $$(P(X' = c_1|c_j), ..., P(X' = c_c|c_j)).$$
  - We require local ϵ-differential privacy, so we need
    $$P(X' = c_1|c_i)/P(X' = c_1|c_j) \leq e^\epsilon, \ldots, P(X' = c_c|c_i)/P(X' = c_c|c_j)) \leq e^\epsilon,$$
    This means
    $$\max_{k=1}^{c} P(X' = c_k|c_i)/P(X' = c_k|c_j) \leq e^\epsilon$$
  - We assumed $P(X' = c_i|c_i)$ the largest value in a row,
    and all non-diagonal values are the same,
    so, maximum is obtained for $k = i$.
    In order to get precisely ϵ privacy (and not $\epsilon_0 < \epsilon$ privacy) we require the equality to hold.
    $$P(X' = c_i|c_i)/P(X' = c_i|c_j) = e^\epsilon. \quad (3)$$

Categorical data

- What is the appropriate noise ? (given ϵ)
  - From these quotients, we compute the values, how?
    $$P(X' = c_i|c_i)/P(X' = c_i|c_j) = e^{\epsilon}.$$
  - each row needs to add to one, so
    $$P(X' = c_i|c_i) + (c-1)P(X' = c_i|c_j) = 1,$$
    or, equivalently, $P(X' = c_i|c_j) = (1 - P(X' = c_i|c_i))/(c-1)$.
  - Using this expression, we have that Equation 3 becomes
    $$P(X' = c_i|c_i)/((1 - P(X' = c_i|c_i))/(c-1)) = e^{\epsilon}.$$
  - This equality implies that $P(X' = c_i|c_i)$ is of the following form:
    $$P(X' = c_i|c_i) = e^{\epsilon}/(c - 1 + e^{\epsilon}),$$
  - and, therefore, $P(X' = c_i|c_j)$ for $i \neq j$ is
    $$P(X' = c_i|c_j) = 1/(c - 1 + e^{\epsilon}).$$

Categorical data

- Example: What is the appropriate noise? (given ϵ, and $c = 2$)
  - Our matrix will have this form

$$\begin{pmatrix} \frac{e^\epsilon}{1+e^\epsilon} & \frac{1}{1+e^\epsilon} \\ \frac{1}{1+e^\epsilon} & \frac{e^\epsilon}{1+e^\epsilon} \end{pmatrix} \tag{4}$$

Categorical data

- What is the appropriate noise? (given ϵ)
- Example. Maximum privacy $c = 2$ and $\epsilon = 0$,
  - the transition matrix contains only $1/2$.

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Categorical data

- What is the appropriate noise? (given ϵ, and $c = 2$)
  - Example. $c = 2$ and $\epsilon = 1$
    Answers: {I like this app, I do not like this app}

$$\begin{pmatrix} \frac{e^\epsilon}{1+e^\epsilon} = 0.73 & \frac{1}{1+e^\epsilon} = 0.27 \\ \frac{1}{1+e^\epsilon} = 0.27 & \frac{e^\epsilon}{1+e^\epsilon} = 0.73 \end{pmatrix}$$

Categorical data

- What is the appropriate noise? (given ϵ, and $c = 2$)
  - Example. $c = 2$ and $\epsilon = 10$
    Answers: {I like this app, I do not like this app}

$$\begin{pmatrix} \frac{e^\epsilon}{1+e^\epsilon} = 0.9999546 & \frac{1}{1+e^\epsilon} = 0.000123 \\ \frac{1}{1+e^\epsilon} = 0.000123 & \frac{e^\epsilon}{1+e^\epsilon} = 0.9999546 \end{pmatrix}$$

Categorical data

- What is the appropriate noise? (given ϵ, and $c = 2$)
  - Example. $c = 7$ and $\epsilon = 10$
  - Answers (Do you like this app): {Not-at-all, don't, ..., fantastic}

$$\begin{pmatrix} \frac{e^\epsilon}{c-1+e^\epsilon} = 0.9997277 & \dots & \frac{1}{c-1+e^\epsilon} = 0.0001233 & \frac{1}{c-1+e^\epsilon} = 0.000123 \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ \frac{1}{c-1+e^\epsilon} = 0.0001233 & \dots & \frac{1}{c-1+e^\epsilon} = 0.0001233 & \frac{e^\epsilon}{c-1+e^\epsilon} = 0.9997277 \end{pmatrix}$$

Categorical data

- Discussion based on Assumptions on the matrix (i, j, k as above)
  - $q_d = P(X' = c_i|c_i) = P(X' = c_j|c_j)$ (diagonal)
  - $q = P(X' = c_i|c_j) = P(X' = c_k|c_l)$
  - and $q_d = P(X' = c_i|c_i) > P(X' = c_j|c_i) = q$
- Nevertheless, we may have other assumptions

Categorical data

- Discussion based on Assumptions on the matrix (i, j, k as above)
  - $q_d = P(X' = c_i|c_i) = P(X' = c_j|c_j)$ (diagonal)
  - $q = P(X' = c_i|c_j) = P(X' = c_k|c_l)$
  - and $q_d = P(X' = c_i|c_i) > P(X' = c_j|c_i) = q$

- Nevertheless, we may have other assumptions
  - When $c = 2$, most general case (assume outputs 0 and 1)

$$\begin{pmatrix} p_{00} & p_{01} = 1 - p_{00} \\ p_{10} = 1 - p_{11} & p_{11} \end{pmatrix} \tag{5}$$

Categorical data

- Most general case $c = 2$
  - Matrix

$$\begin{pmatrix} p_{00} & p_{01} = 1 - p_{00} \\ p_{10} = 1 - p_{11} & p_{11} \end{pmatrix} \tag{6}$$

  - Region of feasibility (i.e., possible p_{00} and p_{11})

Categorical data

- Most general case $c = 2$
  - Matrix

$$\begin{pmatrix} p_{00} & p_{01} = 1 - p_{00} \\ p_{10} = 1 - p_{11} & p_{11} \end{pmatrix} \tag{6}$$

  - Region of feasibility (i.e., possible p_{00} and p_{11})
    - $p_{00} \leq (1 - p_{11})e^{\epsilon}$
    - $p_{11} \leq (1 - p_{00})e^{\epsilon}$
    - $(1 - p_{00}) \leq p_{11}e^{\epsilon}$
    - $(1 - p_{11}) \leq p_{00}e^{\epsilon}$

Neither categorical nor numerical:

Deep learning

- Deep learning is usually implemented with Stochastic Gradient Descent
  - Iterative process with (i is a sample)

$$w_{t+1} = w_t - \alpha_t \nabla g_i(w_t, x_i)$$

  - $\nabla g_i(w_t, x_i)$ is a vector of numbers, so we can just add noise

$$\nabla' g_i(w_t, x_i) = \nabla g_i(w_t, x_i) + Lap(\Delta(\nabla g)/\epsilon)$$

- Problems
  - The amount of noise is too high, and
  - We need to do multiple iterations with lots of samples x_i

- So, we need some variations
  - norm clipping: if the norm of the vector is too large, clip it

$$||g(x)||_2 = \begin{cases} ||g(x)||_2 & ||g(x)||_2 \leq C \\ C & ||g(x)||_2 > C \end{cases}$$

  - Grouping batches: compute average gradients of a batch $x_i \in I$

Some examples with more complex data

- Privacy for complex data and models[6]
  - Graphs
  - Smart grid data
  - Streaming data, multiple releases, etc. (temporal component)
  - Language models
  - Unlearning
- Privacy-preserving solutions in different environments
  - Federated learning
  - Privacy models for voting and decision making

[6] This relates to our own research:
https://www.umu.se/forskning/grupper/nausica-privacy-aware-transparent-decisions-group

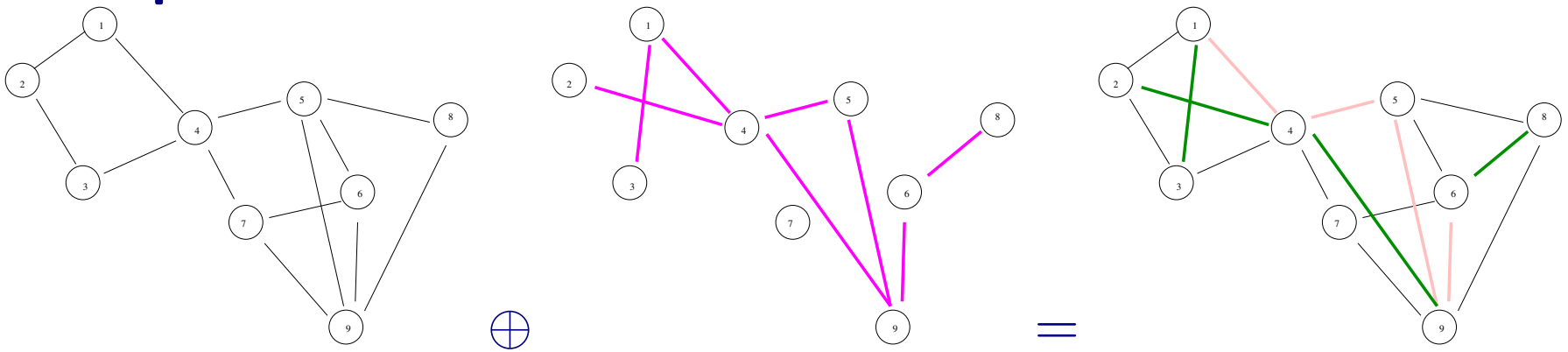
Privacy for graphs: An example

Graphs: All are graphs (data, recommendations, etc)

Noise addition for graphs: Similar idea but with graphs

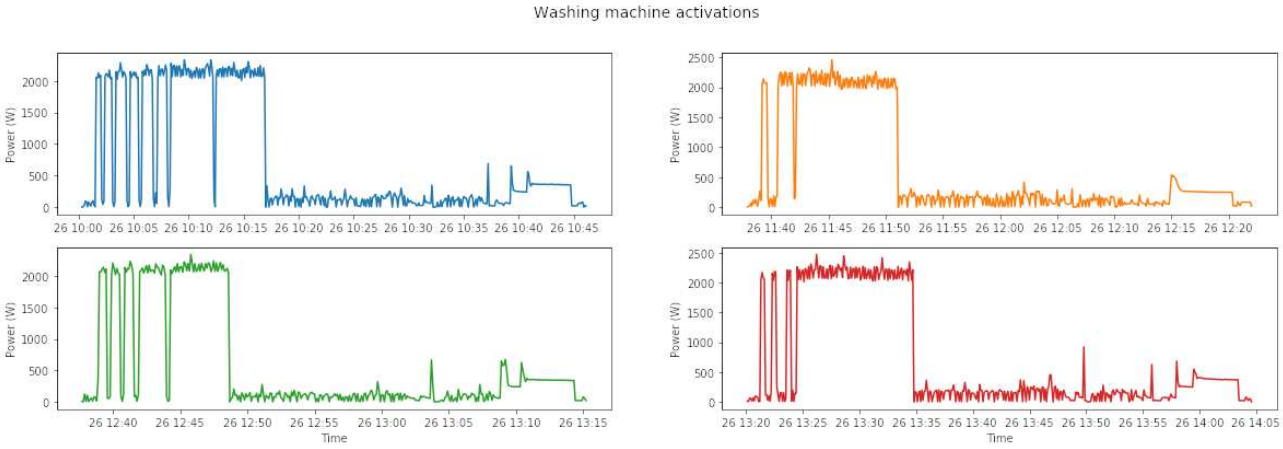
$$G' = G \oplus g$$

- **Example:**



Privacy for smart grid

- Smart grid: electric grid data
  - Data from households
- Sensitive data:
  - consumer habits,
  - Non-intrusive load monitoring (NILM): deduce types of appliances from aggregated energy consumption.



Federated learning

- Privacy in federated learning and trust
  - Local privacy: The agent does not trust the system
    Local-DP / k-anonymity / privacy for re-identification

Federated learning

- Privacy in federated learning and trust
  - Local privacy: The agent does not trust the system
    Local-DP / k-anonymity / privacy for re-identification
  - Global privacy: Only globally we can really protect individuals
    (global) DP

Federated learning

- Privacy in federated learning and trust
  - Local privacy: The agent does not trust the system
    Local-DP / k-anonymity / privacy for re-identification
  - Global privacy: Only globally we can really protect individuals
    (global) DP
  - Infrastructure: No one trusts any one
    Secure multiparty computation

Federated learning

- Privacy in federated learning and trust
  - Local privacy: The agent does not trust the system
    Local-DP / k-anonymity / privacy for re-identification
  - Global privacy: Only globally we can really protect individuals
    (global) DP
  - Infrastructure: No one trusts any one
    Secure multiparty computation
  - Data in the cloud
    Homomorphic encryption

Summary

Summary

- Concepts
  - What is data privacy?
  - Difficulties of data privacy: naive annonymization does not work
  - Concepts: anonymity set, plausible deniability, transparency
  - Disclosure: Identity and attribute
  - Privacy models: k-anonymity, differential privacy
  - Privacy for data: masking methods, microaggregation
  - Privacy for functions: laplacian noise

References

References

- http://www.mdai.cat/dp/
- Torra, V. (2022) Guide to data privacy, Springer.

- Cavoukian, A. (2011) Privacy by design. The 7 foundational principles in Privacy by Design. Strong privacy protection - now, and well into the future.
- D'Acquisto, G., Domingo-Ferrer, J., Kikiras, P., Torra, V., de Montjoye, Y.-A., Bourka, A. (2015) Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics, ENISA Report.
- Torra, V., Navarro-Arribas G. (2016) Big Data Privacy and Anonymization, Privacy and Identity Management 15-26. https://doi.org/10.1007/978-3-319-55783-0_2 (open access)
- Shokri, R., Stronati, M., Song, C., Shmatikov, V. (2017) Membership inference attacks against machine learning models, arXiv:1610.05820v2.