Differential Privacy: An Economic Method for Choosing Epsilon

Justin Hsu[1] Marco Gaboardi[2]
Andreas Haeberlen[1] Sanjeev Khanna[1] Arjun Narayan[1]
Benjamin C. Pierce[1] Aaron Roth[1]

[1]University of Pennsylvania

[2]University of Dundee

July 22, 2014

Why 'Anonymous' Data Sometimes Isn't

By Bruce Schneier 12.13.07

Last year, Netflix published 10 million movie rankings by 500,000 customers, as part of a challenge for people to come up with better recommendation systems than the one the company was using. The data was anonymized by removing personal details and replacing names with random numbers, to protect the privacy of the recommenders.

Arvind Narayanan and Vitaly Shmatikov, researchers at the University of Texas at Austin, de-anonymized some of the Netflix data by comparing rankings and timestamps with public information in the Internet Movie Database, or IMDb.

Their research (.pdf) illustrates some inherent security problems with anonymous data, but first it's important to explain what they did and did not do.

They did *not* reverse the anonymity of the entire Netflix dataset. What they did was reverse the

Problem: Privacy!

Why 'Anonymous' Data Sometimes Isn't

By Bruce Schneier 12

Last year, Netflix publi
for people to come up
data was anonymized b
protect the privacy of t

Arvind Narayanan and
anonymized some of th
in the Internet Movie I

Their research (.pdf) il
important to explain w

They did *not* reverse th

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

E-MAIL
PRINT
REPRINTS

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."

Problem: Privacy!

Why 'Anonymous' Data Sometimes Isn't

By Bruce Schneier 12

Last year, Netflix publi
for people to come up
data was anonymized b
protect the privacy of t

Arvind Narayanan and
anonymized some of th
in the Internet Movie I

Their research (.pdf) il
important to explain w

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

E-MAIL
PRINT
REPRINTS

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

eds of
eriod on
gers" to "60 single men" to
g."

A Practical Attack to De-Anonymize Social Network Users

Gilbert Wondracek, Thorsten Holz
Technical University Vienna, Austria
{gilbert,tho}@seclab.tuwien.ac.at

Engin Kirda
Institute Eurecom, Sophia Antipolis
kirda@eurecom.fr

Christopher Kruegel
University of California, Santa Barbara
chris@cs.ucsb.edu

***Abstract*—Social networking sites such as Facebook, LinkedIn, and Xing have been reporting exponential growth rates and have millions of registered users.**

interesting for attackers. Although social networking sites employ mechanisms to protect the privacy of their users,

Problem: Privacy!

Why 'Anonymous' Data Sometimes Isn't

By Bruce Schneier 12

Last year, Netflix publi
for people to come up
data was anonymized b
protect the privacy of t

Arvind Narayanan and
anonymized some of th
in the Internet Movie I

Their research (.pdf) il
important to explain w

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

E-MAIL
PRINT
REPRINTS

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

eds of
eriod on
gers" to "60 single men" to

A Practical Attack to De-Anonymize Social Network Users

Gilbert Wondracek, Thorsten Holz
Technical University Vienna, Austria
{gilbert,tho}@seclab.tuwien.ac.at

Engin Kirda
Institute Eurecom, Sophia Antipolis

Christopher Kruegel
University of California, Santa Barbara

***Abstract*—Social networking sites such as Facebook, LinkedIn, and Xing have been reporting exponential growth rates and have millions of registered users.**

Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan and Vitaly Shmatikov
The University of Texas at Austin

Abstract

We present a new class of statistical de-anonymization attacks against high-dimensional micro-data, such as individual preferences, recommendations, transaction records and so on. Our techniques

and sparsity. Each record contains many attributes (*i.e.*, columns in a database schema), which can be viewed as dimensions. Sparsity means that for the average record, there are no "similar" records in the multi-dimensional space defined by the attributes. This sparsity is empirically well-established [7, 4, 19] and related to the "fat

Differential privacy?

History

- Notion of privacy by Dwork, McSherry, Nissim, Smith
- Many algorithms satisfying differential privacy now known

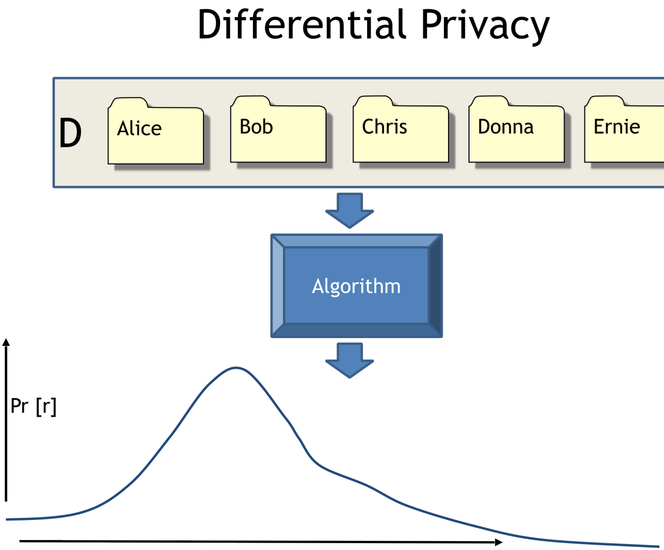
Differential privacy?

History

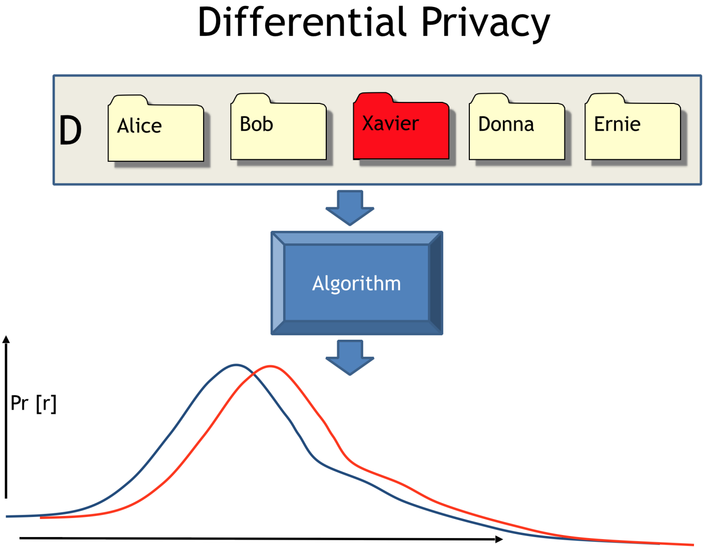
- Notion of privacy by Dwork, McSherry, Nissim, Smith
- Many algorithms satisfying differential privacy now known

Some key features

- Rigorous: differential privacy must be formally proved
- Randomized: property of a probabilistic algorithm
- Quantitative: numeric measure of “privacy loss”



Differential Privacy
D
Alice
Bob
Chris
Donna
Ernie
Algorithm
Pr [r]


Differential Privacy
D
Alice
Bob
Xavier
Donna
Ernie
Algorithm
Pr [r]

In words

The setting

- Database: multiset of records (one per individual)
- Neighboring databases D, D': databases differing in one record
- Randomized algorithm M mapping database to outputs $\mathcal{R}$

The setting

- Database: multiset of records (one per individual)
- Neighboring databases D, D': databases differing in one record
- Randomized algorithm M mapping database to outputs $\mathcal{R}$

Definition

Let $\varepsilon > 0$ be fixed. M is ε-differentially private if for all neighboring databases D, D' and sets of outputs $S \subseteq \mathcal{R}$,

$$\Pr[M(D) \in S] \leq e^{\varepsilon} \cdot \Pr[M(D') \in S].$$

But what about ε?

The equation

$$\Pr[M(D) \in S] \leq e^{\varepsilon} \cdot \Pr[M(D') \in S].$$

The equation

???

$$\Pr[M(D) \in S] \leq e^{\boxed{\varepsilon}} \cdot \Pr[M(D') \in S].$$

The challenge: How to set ε?

The equation

???

$$\Pr[M(D) \in S] \leq e^{\varepsilon} \cdot \Pr[M(D') \in S].$$

Why do we need to set ε?

- Many private algorithms work for a range of ε, but performance highly dependent on particular choice
- Experimental evaluations of private algorithms
- Real-world uses of private algorithms

An easy question?

Theorists say...

- Set ε to be small constant, like 2 or 3
- Proper setting of ε depends on society

An easy question?

Theorists say...

- Set ε to be small constant, like 2 or 3
- Proper setting of ε depends on society

Experimentalists say...

- Try a range of values
- Literature: $\varepsilon =$ 0.01 to 100

An easy question?

Theorists say...

- Set ε to be small constant, like 2 or 3
- Proper setting of ε depends on society

Experimentalists say...

- Try a range of values
- Literature: $\varepsilon = 0.01$ to 100

$e^{\varepsilon} \sim 1.01$

An easy question?

Theorists say...

- Set ε to be small constant, like 2 or 3
- Proper setting of ε depends on society

Experimentalists say...

- Try a range of values
- Literature: $\varepsilon =$ 0.01 to 100

$e^{\varepsilon} \sim 1.01$

$e^{\varepsilon} \sim 2.69 \cdot 10^{48}$

We say

Think about costs rather than privacy

- ε measures privacy, too abstract
- Monetary costs: more concrete way to measure privacy

We say

Think about costs rather than privacy

- ε measures privacy, too abstract
- Monetary costs: more concrete way to measure privacy

Add more parameters!(?)

- Break ε down into more manageable parameters
- More parameters, but more concrete
- Set ε as function of new parameters

The plan today

Model the central tradeoff

- Stronger privacy for smaller ε, weaker privacy for larger ε
- Better accuracy for larger ε, worse accuracy for smaller ε

The plan today

Model the central tradeoff

- Stronger privacy for smaller ε, weaker privacy for larger ε
- Better accuracy for larger ε, worse accuracy for smaller ε

Introduce parameters for two parties

- Individual: concerned about privacy
- Analyst: concerned about accuracy

The plan today

Model the central tradeoff

- Stronger privacy for smaller ε, weaker privacy for larger ε
- Better accuracy for larger ε, worse accuracy for smaller ε

Introduce parameters for two parties

- Individual: concerned about privacy
- Analyst: concerned about accuracy

Combine the parties

- Balance accuracy against privacy guarantee

The plan today

Model the central tradeoff

- Stronger privacy for smaller ε, weaker privacy for larger ε
- Better accuracy for larger ε, worse accuracy for smaller ε

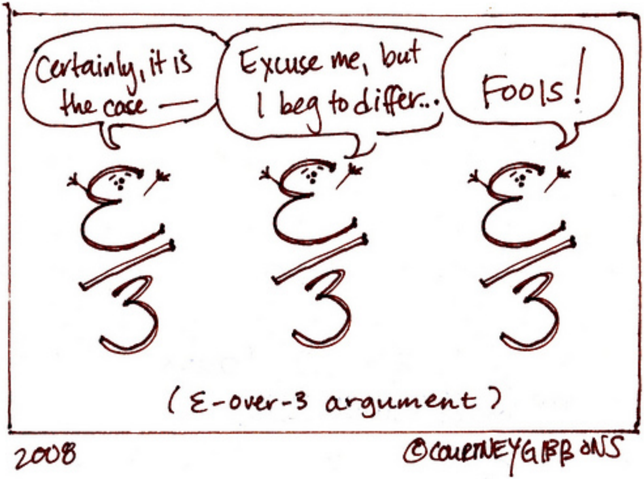
Introduce parameters for two parties

- Individual: concerned about privacy
- Analyst: concerned about accuracy

Combine the parties

- Balance accuracy against privacy guarantee

What does ε mean for privacy?



Participation

- Private algorithm M is a study
- Bob the individual has choice to participate in the study
- Study will happen regardless of Bob's choice

Interpreting ε

Participation

- Private algorithm M is a study
- Bob the individual has choice to participate in the study
- Study will happen regardless of Bob's choice

Bad events

- Set of real-world bad events $\mathcal{O}$
- Bob wants to avoid these events

Thought experiment: two possible worlds

- Identical, except Bob participates in first world and not in the second world
- Rest of database, all public information is identical
- All differences in two worlds due to the output of the study
- Every output $r \in \mathcal{R}$ leads to an event in $\mathcal{O}$ or not

For all sets of outputs S...

$$\Pr[M(\ D\ ) \in S] \leq e^{\varepsilon} \cdot \Pr[M(\ D'\ ) \in S].$$

For all sets of outputs S...

$$\Pr[M(\boxed{D}) \in S] \leq e^{\varepsilon} \cdot \Pr[M(\ D'\ ) \in S].$$

Participate

Outputs to events

For all sets of outputs S...

$$\Pr[M(\underbrace{D}_{\text{Participate}}) \in S] \leq e^{\varepsilon} \cdot \Pr[M(\overbrace{D'}^{\text{Don't participate}}) \in S].$$

For all sets of outputs S...

Don't participate

$$\Pr[M(D) \in S] \leq e^{\varepsilon} \cdot \Pr[M(D') \in S].$$

Participate

Bad events interpretation of ε

- Let S be set of outputs leading to events in $\mathcal{O}$
- Bob participating increases probability of bad event by at most e^{ε} factor

Bad events not equally bad

- Cost function on bad events $f : \mathcal{O} \to \mathbb{R}^+$ (non-negative)
- Insurance premiums, embarrassment, etc.

Bad events not equally bad

- Cost function on bad events $f : \mathcal{O} \rightarrow \mathbb{R}^+$ (non-negative)
- Insurance premiums, embarrassment, etc.

Our model

$\boxed{\text{Pay participants for their cost}}$

Marginal increase in cost

- Someone (society?) has decided the study is worth running
- Non-participants may feel cost, but are not paid
- Only pay participants for increase in expected cost

Marginal increase in cost

- Someone (society?) has decided the study is worth running
- Non-participants may feel cost, but are not paid
- Only pay participants for increase in expected cost

The cost of participation

- Can show: under ε-differential privacy, expected cost increase is at most e^{ε} factor when participating
- Non-participants: expected cost P
- Participants: expected cost at most $e^{\varepsilon} P$
- Compensate participants: $e^{\varepsilon} P - P$

Summing up: the individual model

Individuals

- have an expected cost P if they do not participate, determined by their cost function;
- can choose to participate in an ε-private study for fixed ε in exchange for fixed monetary payment;
- participate if payment is larger than their increase in expected cost for participating: $e^{\varepsilon}P - P$.

Summing up: the individual model

Individuals

- have an expected cost P if they do not participate, determined by their cost function;
- can choose to participate in an ε-private study for fixed ε in exchange for fixed monetary payment;
- participate if payment is larger than their increase in expected cost for participating: $e^{\varepsilon}P - P$.

How to set P?

- Depends on people's perception of privacy costs
- Derive empirically, surveys

Summing up: the individual model

Individuals

- have an expected cost P if they do not participate, determined by their cost function;
- can choose to participate in an ε-private study for fixed ε in exchange for fixed monetary payment;
- participate if payment is larger than their increase in expected cost for participating: $e^{\varepsilon} P - P$.

Bigger for bigger ε

How to set P?

- Depends on people's perception of privacy costs
- Derive empirically, surveys

The plan today

Model the central tradeoff

- Stronger privacy for smaller ε, weaker privacy for larger ε
- Better accuracy for larger ε, worse accuracy for smaller ε

Introduce parameters for two parties

- Individual: concerned about privacy
- Analyst: concerned about accuracy

Combine the parties

- Balance accuracy against privacy guarantee

The plan today

Model the central tradeoff

- Stronger privacy for smaller ε, weaker privacy for larger ε
- Better accuracy for larger ε, worse accuracy for smaller ε

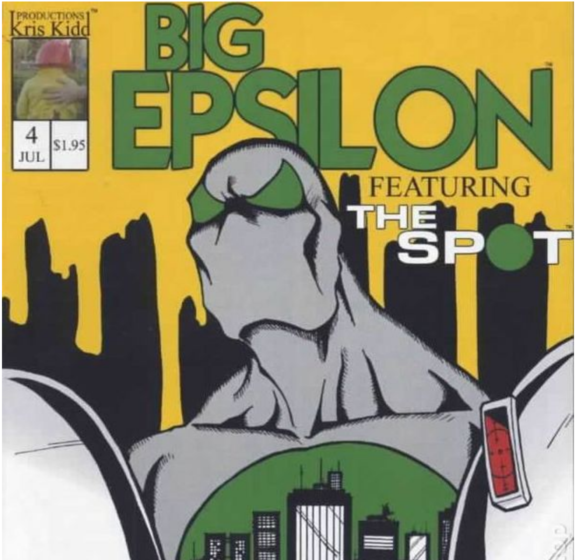
Introduce parameters for two parties

- Individual: concerned about privacy
- Analyst: concerned about accuracy

Combine the parties

- Balance accuracy against privacy guarantee

Why not just take ε small?



The other side

Accuracy?

- Study is run to learn some information; want useful results
- Setting ε small will be very private, but very inaccurate (?)

The other side

Accuracy?

- Study is run to learn some information; want useful results
- Setting ε small will be very private, but very inaccurate (?)

Another parameter: the study size N

- Natural parameter of the study, measures amount of data
- Typical studies: accuracy improves as N increases

Alice the analyst

- Has a private study M, works for range of ε and study size N
- Wants to set these two parameters
- Has numeric measure of accuracy for this study
- Wants to achieve set level of accuracy

Alice the analyst

- Has a private study M, works for range of ε and study size N
- Wants to set these two parameters
- Has numeric measure of accuracy for this study
- Wants to achieve set level of accuracy

What is accuracy?

Measure of accuracy

- Real number, depends on the study M, parameters ε and N
- Could be defined as:
  - Distance from true answer
  - Probability of exceeding error
  - Number of mistakes
  - ...

Level of accuracy

- Real number, maximum allowable accuracy
- Captures Alice's requirement for the study

Summing up: The analyst model

The analyst

- has an ε-private study M;
- has a numeric measure of accuracy $A_M(\varepsilon, N) : \mathbb{R}$;
- has a numeric accuracy level $T : \mathbb{R}$;
- wants $A_M(\varepsilon, N) \leq T$.

Summing up: The analyst model

The analyst

- has an ε-private study M;
- has a numeric measure of accuracy $A_M(\varepsilon, N) : \mathbb{R}$;
- has a numeric accuracy level $T : \mathbb{R}$;
- wants $A_M(\varepsilon, N) \leq T$.

How to set A_M?

- Theoretical accuracy guarantee for M from literature
- Empirical trials: measure accuracy of M on test data

Summing up: The analyst model

The analyst

- has an ε-private study M;
- has a numeric measure of accuracy $A_M(\varepsilon, N) : \mathbb{R}$;
- has a numeric accuracy level $T : \mathbb{R}$;
- wants $A_M(\varepsilon, N) \leq T$.

How to set A_M?

- Theoretical accuracy guarantee for M from literature
- Empirical trials: measure accuracy of M on test data

How to set T?

- Ask the analyst what accuracy is needed

The plan today

Model the central tradeoff

- Stronger privacy for smaller ε, weaker privacy for larger ε
- Better accuracy for larger ε, worse accuracy for smaller ε

Introduce parameters for two parties

- Individual: concerned about privacy
- Analyst: concerned about accuracy

Combine the parties

- Balance accuracy against privacy guarantee

The plan today

Model the central tradeoff

- Stronger privacy for smaller ε, weaker privacy for larger ε
- Better accuracy for larger ε, worse accuracy for smaller ε

Introduce parameters for two parties

- Individual: concerned about privacy
- Analyst: concerned about accuracy

Combine the parties

- Balance accuracy against privacy guarantee

Finally, how to set ε?

Combining the two parties

Budget

- Analyst has budget B (charge it to the grant!)
- Pays sufficient compensation to all N individuals

The goal: find ε and N such that

- Study is accurate enough
- Analayst has enough budget to pay all individuals

System of constraints

1. Accuracy constraint:

$$A_M(\varepsilon, N) \leq T$$

2. Budget constraint:

$$(e^{\varepsilon} P - P) \cdot N \leq B$$

System of constraints

1. Accuracy constraint:

$$A_M(\varepsilon, N) \leq T$$

2. Budget constraint:

$$(e^{\varepsilon} P - P) \cdot N \leq B$$

Variables

- Both sides want to find mutually agreeable setting of ε
- Analyst also wants to find appropriate study size N
- Study feasible $\Leftrightarrow$ constraints satisfiable

System of constraints

1. Accuracy constraint:

$$A_M(\varepsilon, N) \leq T$$

2. Budget constraint:

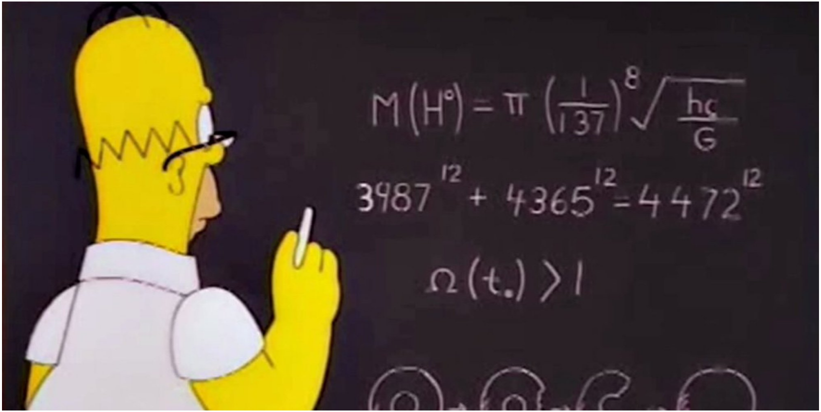
$$(e^{\varepsilon} P - P) \cdot N \leq B$$

Variables

- Both sides want to find mutually agreeable setting of ε
- Analyst also wants to find appropriate study size N
- Study feasible $\Leftrightarrow$ constraints satisfiable

Set ε (and N) to satisfy constraints

Case studies: See paper!



In the paper

- Handle (ε, δ)-privacy
- Add other constraints: limit size of study
- Rule out values of ε that aren't "intuitively" private

Extending the model

In the paper

- Handle (ε, δ)-privacy
- Add other constraints: limit size of study
- Rule out values of ε that aren't "intuitively" private

Further refinements?

- Handle collusion among participants
- Model large ε regime better

Where does that leave us?

Take-away points

- Parameter ε is too abstract
- Use economic cost as a measure of privacy
- Use more concrete parameters: costs, budgets, accuracy, etc.

Where does that leave us?

Take-away points

- Parameter ε is too abstract
- Use economic cost as a measure of privacy
- Use more concrete parameters: costs, budgets, accuracy, etc.

Going forward

- More empirical research: How do people perceive costs?
- Practical attacks on ε-differential privacy? For what ε?
  For what algorithms?

Differential Privacy: An Economic Method for Choosing Epsilon

Justin Hsu[1] Marco Gaboardi[2]
Andreas Haeberlen[1] Sanjeev Khanna[1] Arjun Narayan[1]
Benjamin C. Pierce[1] Aaron Roth[1]

[1]University of Pennsylvania

[2]University of Dundee

July 22, 2014

Which events are considered?

Key assumption: participation decision

- Bob's choice only visible via the output of the study
- Arbitrary side information may be public, as long as it is the same whether Bob participates or not
- Crucial for differential privacy to give a meaningful guarantee!

Which events are considered?

No "side-channels"

Key assumption: participation decision

- Bob's choice only visible via the output of the study
- Arbitrary side information may be public, as long as it is the same whether Bob participates or not
- Crucial for differential privacy to give a meaningful guarantee!

Which events are considered?

No “side-channels”

Key assumption: participation decision

- Bob’s choice only visible via the output of the study
- Arbitrary side information may be public, as long as it is the same whether Bob participates or not
- Crucial for differential privacy to give a meaningful guarantee!

Example: non-protected event

- Someone...
  - monitors Bob’s bank account and sees payment for study;
  - or sees Bob participating in the study;
- ...then uses output of study to break Bob’s privacy

Pitfalls

Individuals With Different Costs?

- Individuals may have different cost functions f
- But cost function may be private, correlated with private data
- Not clear how to compensate them differently, so pay each individual the same amount C

Sampling Bias

- Setting C too low can skew database towards people who don't have very high cost
- Ideal: C is the maximum increase in expected cost P

Setting: Bob the Individual

- Insurance companies don't know Bob smokes
- Bob is worried about his insurance premium increasing

Case Study: Estimating A Mean

Setting: Bob the Individual

- Insurance companies don't know Bob smokes
- Bob is worried about his insurance premium increasing

Setting: Alice the Analyst

- Alice conducting a study on medical records
- Goal: estimate the fraction of the patients who smoke
- Must work under ε-differential privacy

Standard Tool: The Laplace Mechanism

Adding Noise

- Want to compute fraction x, but privately
- Say x can differ by Δ on neighboring databases
- Draw noise ν from the Laplace distribution with scale Δ/ε
- Releasing $x + \nu$ is ε-differentially private

Standard Tool: The Laplace Mechanism

Adding Noise

- Want to compute fraction x, but privately
- Say x can differ by Δ on neighboring databases
- Draw noise ν from the Laplace distribution with scale Δ/ε
- Releasing $x + \nu$ is ε-differentially private

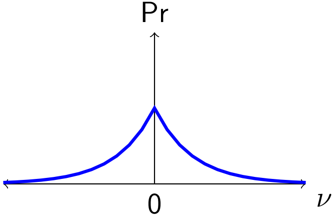


Figure: Laplace distribution

Standard Tool: The Laplace Mechanism

Adding Noise

- Want to compute fraction x, but privately
- Say x can differ by Δ on neighboring databases
- Draw noise ν from the Laplace distribution with scale Δ/ε
- Releasing $x + \nu$ is ε-differentially private

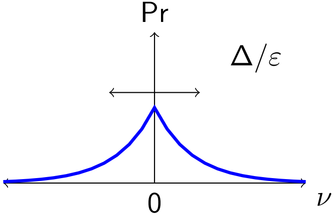


Figure: Laplace distribution

Standard Tool: The Laplace Mechanism

Adding Noise

- Want to compute fraction x, but privately
- Say x can differ by Δ on neighboring databases
- Draw noise ν from the Laplace distribution with scale Δ/ε
- Releasing $x + \nu$ is ε-differentially private

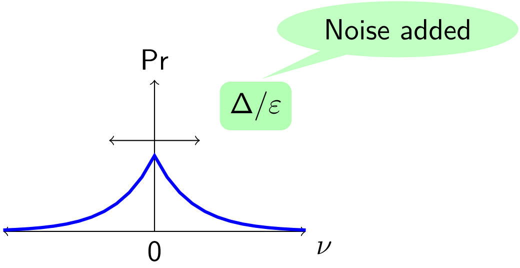


Figure: Laplace distribution

Instantiating the Individual: Estimating Cost

What is the Cost of Not Participating P?

- Correct way to estimate parameter: conduct surveys

What is the Cost of Not Participating P?

- Correct way to estimate parameter: conduct surveys
- Our bad event: health insurance premium increase ($1274)

What is the Cost of Not Participating P?

- Correct way to estimate parameter: conduct surveys
- Our bad event: health insurance premium increase ($1274)
- Bob estimates probability this happens even if he doesn't participate: 5%

Instantiating the Individual: Estimating Cost

What is the Cost of Not Participating P?

- Correct way to estimate parameter: conduct surveys
- Our bad event: health insurance premium increase (\$1274)
- Bob estimates probability this happens even if he doesn't participate: 5%
- Expected cost of non-participation: $P = 5\% \cdot \$1274 = \$63.7$

Instantiating the Individual: Estimating Cost

What is the Cost of Not Participating P?

- Correct way to estimate parameter: conduct surveys
- Our bad event: health insurance premium increase (\$1274)
- Bob estimates probability this happens even if he doesn't participate: 5%
- Expected cost of non-participation: $P = 5\% \cdot \$1274 = \$63.7$

$$\boxed{\text{Bob will participate if paid } 63.7 \cdot (e^{\varepsilon} - 1)}$$

Instantiating the Analyst: Estimating Accuracy

Measuring the Accuracy

- Alice wants fraction of smokers to within 0.05 error
- Measure of accuracy: $A_M(\varepsilon, N)$ is probability of exceeding this error, want probability to be small (at most 10% chance)

Instantiating the Analyst: Estimating Accuracy

Measuring the Accuracy

- Alice wants fraction of smokers to within 0.05 error
- Measure of accuracy: $A_M(\varepsilon, N)$ is probability of exceeding this error, want probability to be small (at most 10% chance)

Dependence on Database Size

- Changing one record changes μ by at most $1/N$
- As N grows, less noise needed for ε-privacy

The Budget Constraint

- Alice has $B = \$30{,}000$ to spend: constraint

$$63.7 \cdot (e^{\varepsilon} - 1) \cdot N \leq 30000$$

The Budget Constraint

- Alice has $B = \$30{,}000$ to spend: constraint

$$63.7 \cdot (e^{\varepsilon} - 1) \cdot N \leq 30000$$

The Accuracy Constraint

- Alice wants probability of exceeding error at most 10%
- Sets $T = 0.1$ and requires $A_M(\varepsilon, N) \leq T = 0.1$
- Can be shown via statistical tools, sufficient to have

$$2 \exp\left(-0.0002N\right) + \exp\left(-0.025N\varepsilon\right) \leq 0.1$$

The Budget Constraint

- Alice has $B = \$30{,}000$ to spend: constraint

$$63.7 \cdot (e^{\varepsilon} - 1) \cdot N \leq 30000$$

The Accuracy Constraint

- Alice wants probability of exceeding error at most 10%
- Sets $T = 0.1$ and requires $A_M(\varepsilon, N) \leq T = 0.1$
- Can be shown via statistical tools, sufficient to have

$$2 \exp\left(-0.0002N\right) + \exp\left(-0.025N\varepsilon\right) \leq 0.1$$

$$\boxed{\text{Study feasible} \Leftrightarrow \text{constraints satisfiable}}$$

Is the Study Feasible?

Is the Study Feasible?

Yes!

- $N = 15000, \varepsilon = 0.03$
- Bob is paid \$1.93

Yes!

- $N = 15000, \varepsilon = 0.03$
- Bob is paid \$1.93

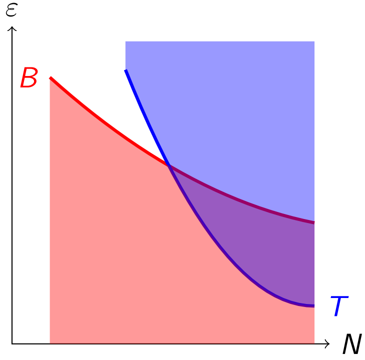


Figure: Feasible ε, N, for accuracy T and budget B.

Other Applications: The Cost of Privacy

Non-private Studies

- No privacy guarantee
- What if non-private studies had to pay extra for this risk?

Tradeoff

- Non-private study has better accuracy, need smaller study, but needs to pay more per person
- Private study has worse accuracy, needs bigger study, but pays less per person

Our Model

- Private study is sometimes cheaper than equivalent non-private study!