Learning and Privacy

Christos Dimitrakakis

February 14, 2019

Introduction

Privacy in databases

Bayesian inference and privacy

Just because they're the problem,
doesn't mean we aren't.

Privacy in statitical disclosure.

- Public analysis of sensitive data.
- Publication of “anonymised” data.

Not about cryptography

- Secure communication and computation.
- Authentication and verification.

An issue of trust

- Who to trust and how much.
- With what data to trust them.
- What you want out of the service.

Introduction

Privacy in databases

k-anonymity

Differential privacy

Bayesian inference and privacy

Anonymisation

Example 1 (Typical relational database in Tinder)

Birthday	Name	Height	Weight	Age	Postcode	Professio
06/07	Li Pu	190	80	60-70	1001	Politiciar
06/14	Sara Lee	185	110	70+	1001	Rentier
01/01	A. B. Student	170	70	40-60	6732	Time Tra

Anonymisation

Example 1 (Typical relational database in Tinder)

Birthday	Name	Height	Weight	Age	Postcode	Profession
06/07		190	80	60-70	1001	Politician
06/14		185	110	70+	1001	Rentier
01/01		170	70	40-60	6732	Time Traveller

The simple act of hiding or using random identifiers is called anonymisation.

Record linkage

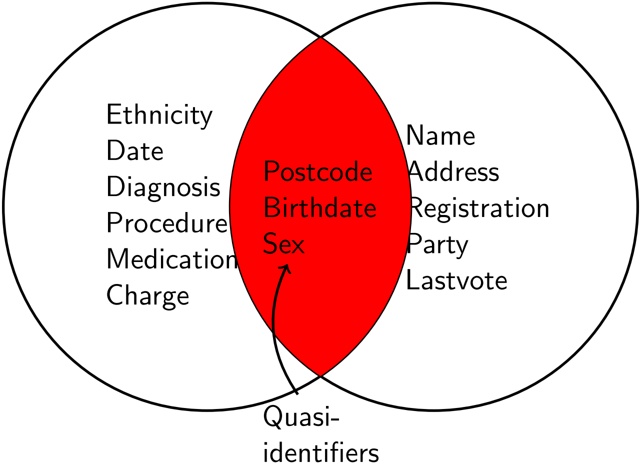


Figure: An example of two datasets, one containing sensitive and the other public information. The two datasets can be linked and individuals identified through the use of quasi-identifiers.

k-anonymity

(a) Samarati

(b) Sweeney

Definition 4 (*k*-anonymity)

A database provides *k*-anonymity if for every person in the database is indistinguishable from $k - 1$ persons with respect to *quasi-identifiers*.

It's the analyst's job to define quasi-identifiers

Birthday	Name	Height	Weight	Age	Postcode	Pr
06/07	Li Pu	190	80	60+	1001	Po
06/14	Sara Lee	185	110	60+	1001	Re
06/12	Nikos Papadopoulos	170	82	60+	1243	Po
01/01	A. B. Student	170	70	40-60	6732	Ti
05/08	Li Yang	175	72	30-40	6910	Ti

Table: 1-anonymity.

Birthday	Name	Height	Weight	Age	Postcode	Profession
06/07		190	80	60+	1001	Politician
06/14		185	110	60+	1001	Rentier
06/12		170	82	60+	1243	Politician
01/01		170	70	40-60	6732	Time Traveller
05/08		175	72	30-40	6910	Policeman

1-anonymity

Birthday	Name	Height	Weight	Age	Postcode	Profession
06/07		180-190	80+	60+	1*	
06/14		180-190	80+	60+	1*	
06/12		170-180	60+	60+	1*	
01/01		170-180	60-80	20-60	6*	
05/08		170-180	60-80	20-60	6*	

1-anonymity

Birthday	Name	Height	Weight	Age	Postcode	Profession
		180-190	80+	60+	1*	
		180-190	80+	60+	1*	
		170-180	60-80	69+	1*	
		170-180	60-80	20-60	6*	
		170-180	60-80	20-60	6*	

Table: 2-anonymity: the database can be partitioned in sets of at least 2 records

x_1

x

Figure: If two people contribute their data $x = (x_1, x_2)$ to a medical database, and an algorithm π computes some public output a from x, then it should be hard infer anything about the data from the public output.

Figure: If two people contribute their data $x = (x_1, x_2)$ to a medical database, and an algorithm π computes some public output a from x, then it should be hard infer anything about the data from the public output.

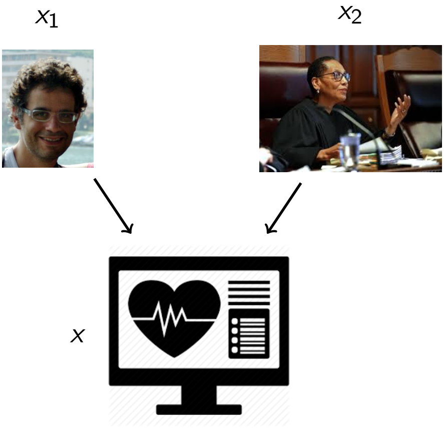


Figure: If two people contribute their data $x = (x_1, x_2)$ to a medical database, and an algorithm π computes some public output a from x, then it should be hard infer anything about the data from the public output.

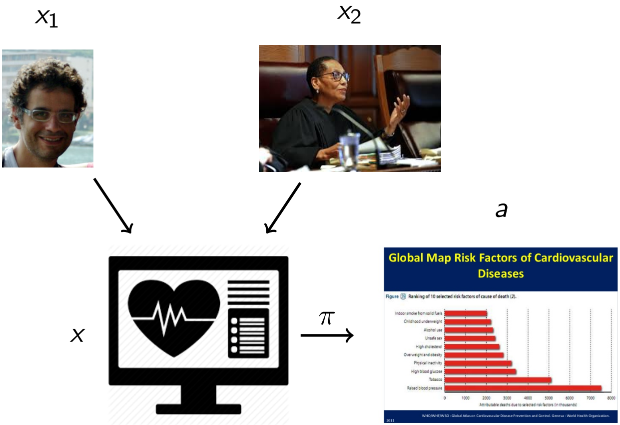


Figure: If two people contribute their data $x = (x_1, x_2)$ to a medical database, and an algorithm π computes some public output a from x, then it should be hard infer anything about the data from the public output.

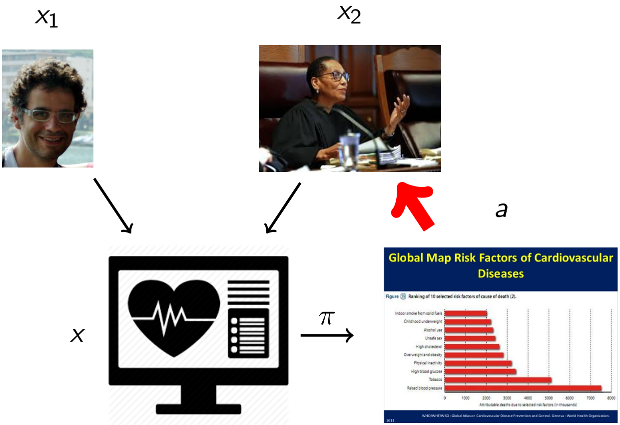


Figure: If two people contribute their data $x = (x_1, x_2)$ to a medical database, and an algorithm π computes some public output a from x, then it should be hard infer anything about the data from the public output.

Privacy desiderata

We wish to calculate something on some private data and publish a privacy-preserving, but useful, version of the result.

- Anonymity: Individual participation remains hidden.
- Secrecy: Individual data x_i is not revealed.
- Side-information: Linkage attacks are not possible.
- Utility: The calculation remains useful.

Example: The prevalence of drug use in sport

- n athletes
- Ask whether they have doped in the past year.
- Aim: calculate % of doping.
- How can we get truthful / accurate results?

Example: The prevalence of drug use in sport

- n athletes
- Ask whether they have doped in the past year.
- Aim: calculate % of doping.
- How can we get truthful / accurate results?

Algorithm for randomising responses about drug use

1. Flip a coin.
2. If it comes heads, respond truthfully.
3. Otherwise, flip another coin and respond `yes` if it comes heads and `no` otherwise.

The randomised response mechanism

Definition 5 (Randomised response)

The i-th user, whose data is $x_i \in \{0, 1\}$, responds with $a_i \in \{0, 1\}$ with probability

$$\pi(a_i = j \mid x_i = k) = p, \qquad \pi(a_i = k \mid x_i = k) = 1 - p,$$

where $j \neq k$.

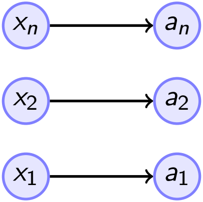


Figure: The local privacy model

The centralised privacy model

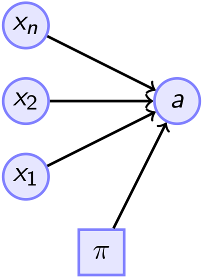


Figure: The centralised privacy model

Assumption 1

The data x is collected and the result a is published by a trusted curator

The centralised privacy model

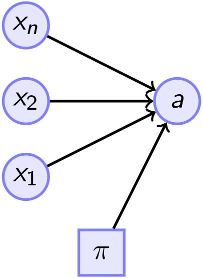


Figure: The centralised privacy model

Example 6

Calculate the ratio of people that take drugs

$$\mathbb{E}_{\pi}[a \mid \boldsymbol{x}] = \frac{1}{n} \sum_i x_i, \qquad \pi = \mathcal{L}aplace(\frac{1}{n} \sum_i x_i, \lambda)$$

Generalised queries

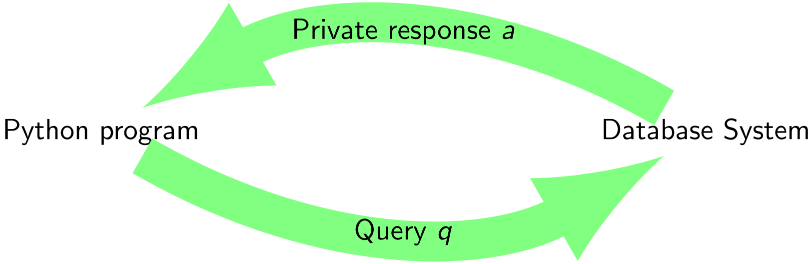


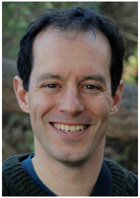
Figure: Private database access model

Response policy

The policy defines a distribution over responses

$$\pi(a \mid x, q)$$

Differential privacy.

Definition 7 (ϵ-Differential Privacy)

A stochastic algorithm $\pi : \mathcal{X} \to \mathcal{A}$, where $\mathcal{X}$ is endowed with a neighbourhood relation N, is said to be ϵ-differentially private if

$$\left| \ln \frac{\pi(a \mid x)}{\pi(a \mid x')} \right| \leq \epsilon, \qquad \forall x N x'. \tag{2.1}$$

Composition

Answering T queries with an ϵ-DP mechanism, loses ϵT privacy..

Defining neighbourhoods

Birthday	Name	Height	Weight
06/07	Li Pu	190	80
06/14	Sara Lee	185	110
06/12	Nikos Papadopoulos	170	82
01/01	A. B. Student	170	70
05/08	Li Yang	175	72

Table: Data x

Birthday	Name	Height	Weight
06/07	Li Pu	190	80
06/14	Sara Lee	185	110
01/01	A. B. Student	170	70
05/08	Li Yang	175	72

Table: 1-Neighbour x'

Defining neighbourhoods

Birthday	Name	Height	Weight
06/07	Li Pu	190	80
06/14	Sara Lee	185	110
06/12	Nikos Papadopoulos	170	82
01/01	A. B. Student	170	70
05/08	Li Yang	175	72

Table: Data x

Birthday	Name	Height	Weight
06/07	Li Pu	190	80
06/14	Sara Lee	185	110
06/12	Nikos Papadopoulos	180	80
01/01	A. B. Student	170	70
05/08	Li Yang	175	72

Table: 2-Neighbour x'

Answering any query with a ϵ-DP algorithm bounds the amount of information gained by any adversary, no matter their previous knowledge. This means they cannot even guess whether you are in the dataset.

The Exponential Mechanism.

Interactive queries

- System has data x.
- User asks query q.
- System responds with a
- We wish to maximise utility: $U : \mathcal{X}, \mathcal{A}, \mathcal{Q} \to \mathbb{R}$.

The Exponential Mechanism.

Interactive queries

- System has data x.
- User asks query q. —e.g. "what is the average of x"?
- System responds with a
- We wish to maximise utility: $U : \mathcal{X}, \mathcal{A}, \mathcal{Q} \to \mathbb{R}$.

The Exponential Mechanism.

Interactive queries

- System has data x.
- User asks query q.
- System responds with a —e.g. a noisy version of the average.
- We wish to maximise utility: $U : \mathcal{X}, \mathcal{A}, \mathcal{Q} \to \mathbb{R}$.

The Exponential Mechanism.

Interactive queries

- System has data x.
- User asks query q.
- System responds with a
- We wish to maximise utility: $U : \mathcal{X}, \mathcal{A}, \mathcal{Q} \to \mathbb{R}$. The utility is higher for responses closer to the correct response.

The Exponential Mechanism.

Interactive queries

- System has data x.
- User asks query q.
- System responds with a
- We wish to maximise utility: $U : \mathcal{X}, \mathcal{A}, \mathcal{Q} \to \mathbb{R}$.

Definition 8 (The Exponential mechanism)

For any utility function $U : \mathcal{Q} \times \mathcal{A} \times \mathcal{X} \to \mathbb{R}$, define the policy

$$\pi(a \mid x) \triangleq \frac{e^{\epsilon U(q,a,x)/\mathbb{L}(U(q))}}{\sum_{a'} e^{\epsilon U(q,a',x)/\mathbb{L}(U(q))}} \tag{2.2}$$

The Exponential Mechanism.

Interactive queries

- System has data x.
- User asks query q.
- System responds with a
- We wish to maximise utility: $U : \mathcal{X}, \mathcal{A}, \mathcal{Q} \to \mathbb{R}$.

Definition 8 (The Exponential mechanism)

For any utility function $U : \mathcal{Q} \times \mathcal{A} \times \mathcal{X} \to \mathbb{R}$, define the policy

$$\pi(a \mid x) \triangleq \frac{e^{\epsilon U(q,a,x)/\mathbb{L}(U(q))}}{\sum_{a'} e^{\epsilon U(q,a',x)/\mathbb{L}(U(q))}} \tag{2.2}$$

The $\mathbb{L}\,()$ term ensures the noise is calibrated to the privacy level we want

Theoretical foundations

A differentially private algorithm is intrinsically stable. This leads to a number of results.

- *Generalization in adaptive data analysis and holdout reuse*. Dwork et al, NIPS 2015.
- *Algorithmic stability for adaptive data analysis*. Bassily et al, STOC 2016.
- *Concentration Bounds for High Sensitivity Functions Through Differential Privacy*, Nissim and Stemmer, 2017.
- *Subgaussian Tail Bounds via Stability Arguments*, Steinke and Ullman, 2017.

Available privacy toolboxes

k-anonymity

- `https://github.com/qiyuangong/Mondrian` Mondrian k-anonymity

Differential privacy

- `https://github.com/bmcmenamin/thresholdOut-explorations` Threshold out
- `https://github.com/steven7woo/Accuracy-First-Differential-Privacy` Accuracy-constrained DP
- `https://github.com/menisadi/pydp` Various DP algorithms
- `https://github.com/haiphanNJIT/PrivateDeepLearning` Deep learning and DP

The Privacy Tools Project `https://privacytools.seas.harvard.edu/`

Introduction

Privacy in databases
- k-anonymity
- Differential privacy

Bayesian inference and privacy
- Setting
- Bayesian inference for privacy
- Robustness and privacy of the posterior distribution
- Posterior sampling query model
- Experiments

Bayesian inference and differential privacy

Bayesian estimation

- What are its robustness and privacy properties?
- How important is the selection of the prior?

Limiting the communication channel

- How should we communicate information about our posterior?
- How much can an adversary learn from our posterior?

Dramatis personae

- x – data.
- $\mathscr{B}$ – a (Bayesian) statistician.
- ξ – the statistician's prior belief.
- θ – a parameter
- $\mathscr{A}$ – an adversary. He knows ξ, should not learn x.

Dramatis personae

- x – data.
- $\mathscr{B}$ – a (Bayesian) statistician.
- ξ – the statistician's prior belief.
- θ – a parameter
- $\mathscr{A}$ – an adversary. He knows ξ, should not learn x.

The game

1. $\mathscr{B}$ selects a model family ($\mathcal{F}$) and a prior (ξ).
2. $\mathscr{B}$ observes data x and calculates the posterior $\xi(\theta|x)$.
3. $\mathscr{A}$ queries $\mathscr{B}$.
4. $\mathscr{B}$ responds with a function of the posterior $\xi(\theta|x)$.
5. Goto 3.

Bayesian inference

Estimating a coin's bias

A fair coin comes heads 50% of the time. We want to test an unknown coin, which we think may not be completely fair.

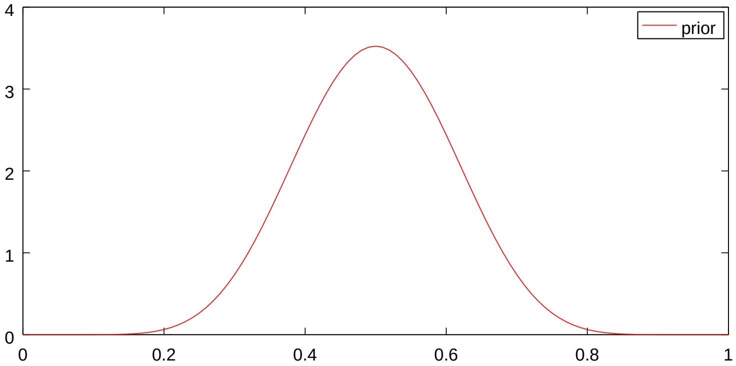


Figure: Prior belief ξ about the coin bias θ.

Bayesian inference

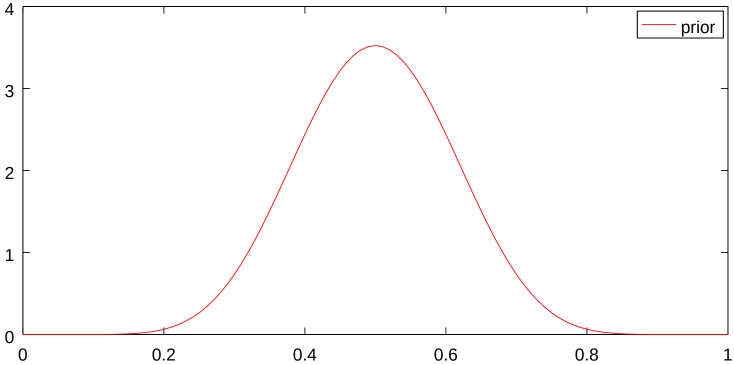


Figure: Prior belief ξ about the coin bias θ.

For a sequence of throws $x_t \in \{0, 1\}$,

$$P_\theta(x) \propto \prod_t \theta^{x_t}(1-\theta)^{1-x_t} = \theta^{\#\mathrm{Heads}}(1-\theta)^{\#\mathrm{Tails}}$$

Bayesian inference

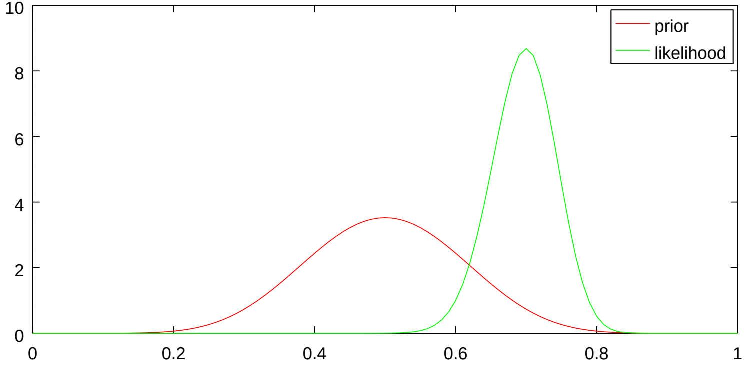


Figure: Prior belief ξ about the coin bias θ and likelihood of θ for the data.

Say we throw the coin 100 times and obtain 70 heads. Then we plot the likelihood $P_\theta(x)$ of different models.

Bayesian inference

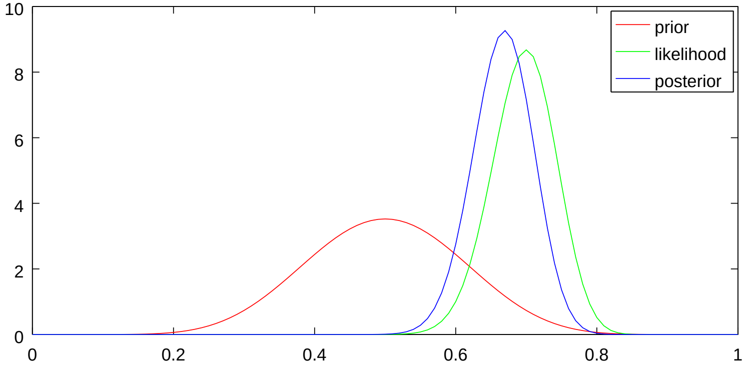


Figure: Prior belief $\xi(\theta)$ about the coin bias θ, likelihood of θ for the data, and posterior belief $\xi(\theta \mid x)$

From these, we calculate a posterior distribution over the correct models. This represents our conclusion given our prior and the data.

Bayesian inference

Setting

- Dataset space $\mathcal{S}$.
- Distribution family $\mathcal{F} \triangleq \{P_\theta \mid \theta \in \Theta\}$.
- Each P_θ is a distribution on $\mathcal{S}$.
- We wish to identify which θ generated the observed data x.
- Prior distribution ξ on Θ (i.e. initial belief)
- Posterior given data $x \in \mathcal{S}$ (i.e. conclusion)

$$\xi(\theta \mid x) = \frac{P_\theta(x)\xi(\theta)}{\phi(x)} \qquad \text{(posterior)}$$

$$\phi(x) \triangleq \sum_{\theta \in \Theta} P_\theta(x)\xi(\theta). \qquad \text{(marginal)}$$

Standard calculation that can be done exactly or approximately.

Introduction

Privacy in databases

Bayesian inference and privacy

- Bayesian inference for privacy
- Robustness and privacy of the posterior distribution
- Posterior sampling query model
- Experiments

What we want to show

- If we assume the family $\mathcal{F}$ is well-behaved ...
- ...or that the prior ξ is focused on the "nice" parts of $\mathcal{F}$

What we want to show

- If we assume the family $\mathcal{F}$ is well-behaved ...
- ...or that the prior ξ is focused on the "nice" parts of $\mathcal{F}$
- Inference is robust.
- Our knowledge is private.
- There are also well-known $\mathcal{F}$ satisfying our assumptions.

What we want to show

- If we assume the family $\mathcal{F}$ is well-behaved ...
- ...or that the prior ξ is focused on the "nice" parts of $\mathcal{F}$
- Inference is robust.
- Our knowledge is private.
- There are also well-known $\mathcal{F}$ satisfying our assumptions.

First, we must generalise differential privacy...

Differential privacy of conditional distribution $\xi(\cdot \mid x)$

Definition 9 ((ϵ, δ)-differential privacy)

$\xi(\cdot \mid x)$ is (ϵ, δ)-*differentially private* if, $\forall x \in \mathcal{S} = \mathcal{X}^n$, $B \subset \Theta$

$$\xi(B \mid x) \leq e^{\epsilon} \xi(B \mid y) + \delta,$$

for all y in the hamming-1 neighbourhood of x.

Differential privacy of conditional distribution $\xi(\cdot \mid x)$

Definition 9 ((ϵ, δ)-differential privacy)

$\xi(\cdot \mid x)$ is (ϵ, δ)-*differentially private* if, $\forall x \in \mathcal{S} = \mathcal{X}^n$, $B \subset \Theta$

$$\xi(B \mid x) \leq e^{\epsilon} \xi(B \mid y) + \delta,$$

for all y in the hamming-1 neighbourhood of x.

Definition 10 ((ϵ, δ)-differential privacy under ρ.)

$\xi(\cdot \mid x)$ is (ϵ, δ)-*differentially private under* a pseudo-metric $\rho : \mathcal{S} \times \mathcal{S} \to \mathbb{R}_+$ if, $\forall B \subset \Theta$ and $x \in \mathcal{S}$,

$$\xi(B \mid x) \leq e^{\epsilon \rho(x,y)} \xi(B \mid y) + \delta \rho(x, y), \qquad \forall y \in \mathcal{S}$$

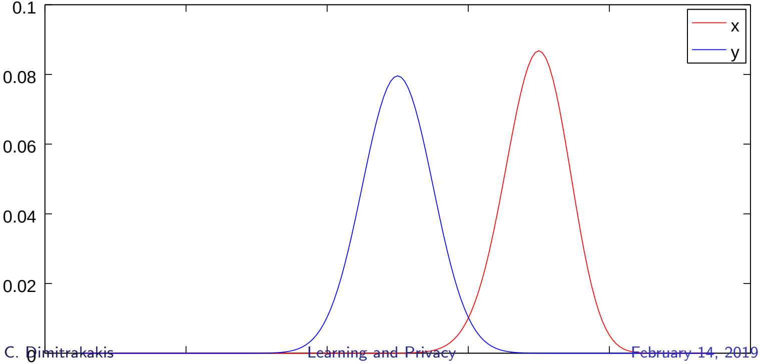
If two datasets x, y are close, then the distributions $\xi(\cdot \mid x)$ and $\xi(\cdot \mid y)$ are also close.

Sufficient conditions

Assumption 1 ($\mathcal{F}$ is Lipschitz)

For a given ρ *on* $\mathcal{S}$, $\exists L > 0$ *s.t.* $\forall \theta \in \Theta$:

$$\left| \ln \frac{P_\theta(x)}{P_\theta(y)} \right| \leq L\rho(x, y), \qquad \forall x, y \in \mathcal{S}, \tag{3.1}$$

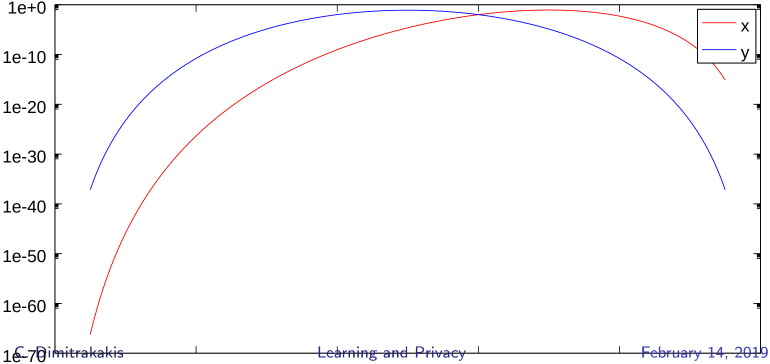


Sufficient conditions

Assumption 1 ($\mathcal{F}$ is Lipschitz)

For a given ρ on $\mathcal{S}$, $\exists L > 0$ s.t. $\forall \theta \in \Theta$:

$$\left| \ln \frac{P_\theta(x)}{P_\theta(y)} \right| \leq L\rho(x, y), \qquad \forall x, y \in \mathcal{S}, \tag{3.1}$$

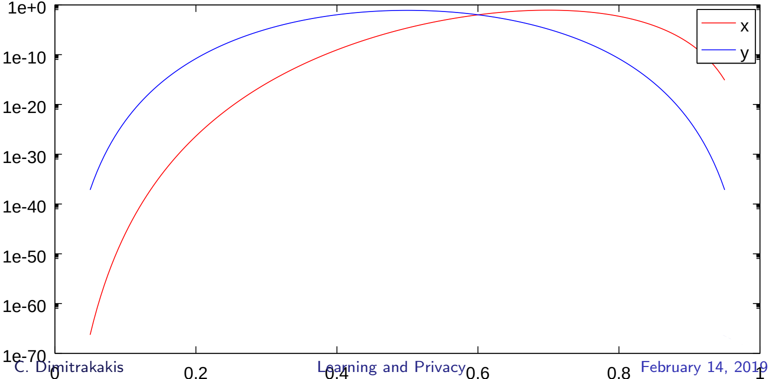


Stochastic Lipschitz condition

Assumption 2 (The prior is concentrated on nice parts of $\mathcal{F}$)

Let the set of L-Lipschitz parameters be Θ_L. Then $\exists c > 0$ s.t.

$$\xi(\Theta_L) \geq 1 - \exp(-cL), \forall L \tag{3.2}$$

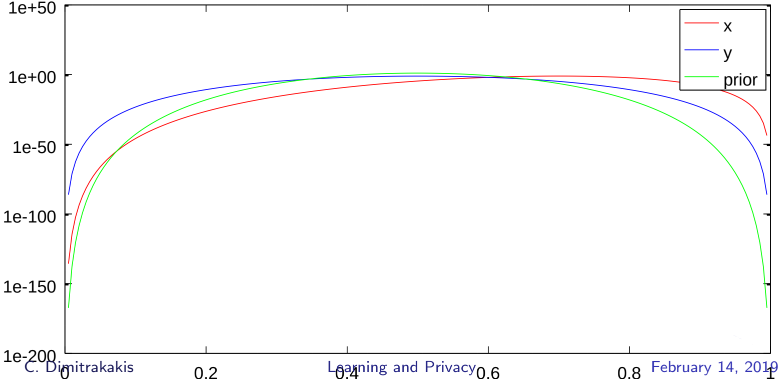


Stochastic Lipschitz condition

Assumption 2 (The prior is concentrated on nice parts of $\mathcal{F}$)

Let the set of L-Lipschitz parameters be Θ_L. Then $\exists c > 0$ s.t.

$$\xi(\Theta_L) \geq 1 - \exp(-cL), \forall L \tag{3.2}$$



Robustness of the posterior distribution

Definition 11 (KL divergence)

$$D\left(P \parallel Q\right) \triangleq \int \ln \frac{\mathrm{d}P}{\mathrm{d}Q} \, \mathrm{d}P. \tag{3.3}$$

Theorem 12

Robustness of the posterior distribution

Definition 11 (KL divergence)

$$D\left(P \parallel Q\right) \triangleq \int \ln \frac{\mathrm{d}P}{\mathrm{d}Q} \,\mathrm{d}P. \tag{3.3}$$

Theorem 12

(i) *Under Assumption 1,*

$$D\left(\xi(\cdot \mid x) \parallel \xi(\cdot \mid y)\right) \le 2L\rho(x, y) \tag{3.4}$$

Robustness of the posterior distribution

Definition 11 (KL divergence)

$$D\left(P \parallel Q\right) \triangleq \int \ln \frac{\mathrm{d}P}{\mathrm{d}Q} \,\mathrm{d}P. \tag{3.3}$$

Theorem 12

(i) *Under Assumption 1,*

$$D\left(\xi(\cdot \mid x) \parallel \xi(\cdot \mid y)\right) \leq 2L\rho(x, y) \tag{3.4}$$

(ii) *Under Assumption 2,*

$$D\left(\xi(\cdot \mid x) \parallel \xi(\cdot \mid y)\right) \leq \frac{\kappa C_\xi}{c} \cdot \rho(x, y) \tag{3.5}$$

Differential privacy of the posterior distribution

- Under Assumption 1, $B \in \sigma(\Theta)$:

$$\xi(B \mid x) \leq e^{2L\rho(x,y)}\xi(B \mid y) \tag{3.6}$$

  i.e. the posterior is $(2L, 0)$-DP under ρ.

- Under Assumption 2, for all $x, y \in \mathcal{S}$, $B \in \sigma(\Theta)$:

$$|\xi(B \mid x) - \xi(B \mid y)| \leq \sqrt{\frac{\kappa C_\xi}{2c}\rho(x,y)}$$

  i.e. the posterior is $\left(0, \sqrt{\frac{\kappa C_\xi}{2c}}\right)$-DP under $\sqrt{\rho}$.

Posterior sampling query model

- We select a prior ξ.
- We observe data x.
- We calculate a posterior $\xi(\cdot \mid x)$.
- An adversary has sampling-based access to the posterior.

Posterior sampling query model

- We select a prior ξ.
- We observe data x.
- We calculate a posterior $\xi(\cdot \mid x)$.
- An adversary has sampling-based access to the posterior.

First idea

At time t, the adversary observes a sample from the posterior:

$$\theta_t \sim \xi(\theta \mid x),$$

Posterior sampling query model

- We select a prior ξ.
- We observe data x.
- We calculate a posterior $\xi(\cdot \mid x)$.
- An adversary has sampling-based access to the posterior.

First idea

At time t, the adversary observes a sample from the posterior:

$$\theta_t \sim \xi(\theta \mid x),$$

$\mathscr{A}$ may instead query using a function $q : \Theta \to \mathcal{R}$, to obtain:

$$r_t = q(\theta_t)$$

Responding to queries via utilities

Posterior sampling

Given a prior ξ, data x and number of samples n,

$$\hat{\Theta} \sim \xi^n(\cdot \mid x).$$

Sample query response

For a query q_t and utility function $u_\theta : \mathcal{R} \times \mathcal{Q} \to [0,1]$, return:

$$r_t \in \arg\max_r \sum_{\theta \in \hat{\Theta}} u_\theta(r, q_t)$$

Theorem 13

If $\xi^\star$ is $\mathscr{A}$'s preferred prior, and we restrict it so $\xi(\Theta_L) = 1$:

(a) *The algorithm is $2Ln$-differentially private.*

(b) *$\mathscr{A}$'s regret is $O([1 - \xi^\star(\Theta_L)] + \sqrt{\ln(1/\delta)/n})$, w.p. $1 - \delta$.*

Another look at the exponential mechanism

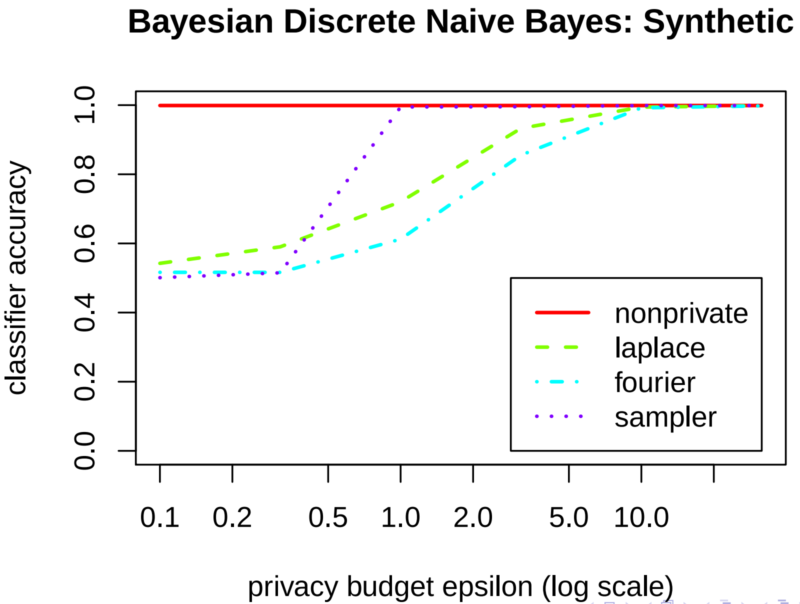
Define a utility function $u(x, r)$

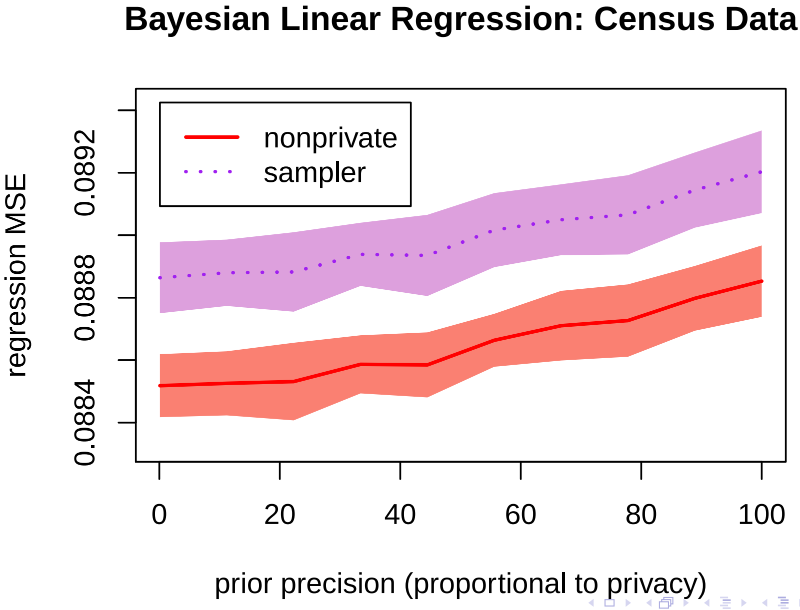
$$p(r) \propto e^{\epsilon u(x,r)} \mu(r).$$

Respond with r with probability $p(r)$.

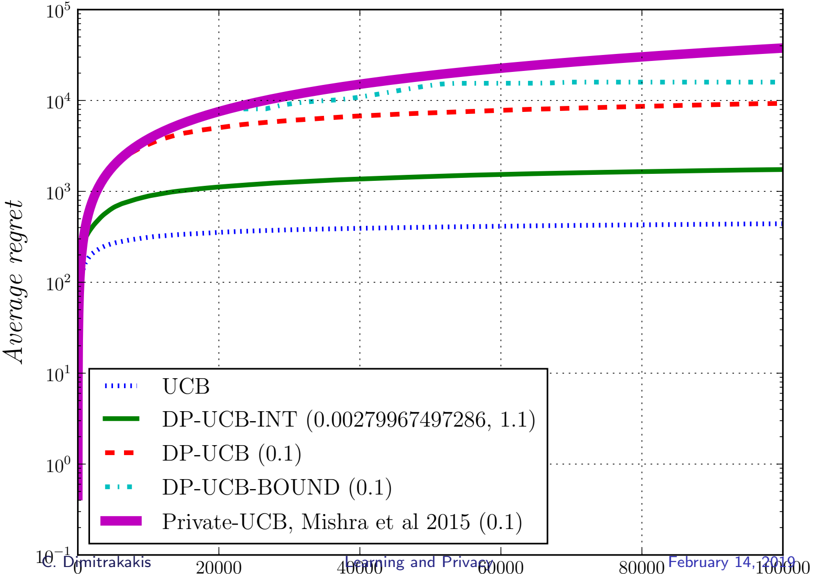
Connection to posterior mechanism

- Responses are parameters θ.
- Take $u(\theta, x) = \log P_\theta(x)$.
- Take $\mu(\theta) = \xi(\theta)$.
- Then $p(\theta) = \xi(\theta \mid x)$.
- Rather than tuning ϵ, we can tune
  - The prior ξ.
  - The number of samples n.


Bayesian Discrete Naive Bayes: Synthetic
classifier accuracy
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.5
1.0
2.0
5.0
10.0
privacy budget epsilon (log scale)
nonprivate
laplace
fourier
sampler


Bayesian Linear Regression: Census Data
nonprivate
sampler
regression MSE
0.0884
0.0888
0.0892
0
20
40
60
80
100
prior precision (proportional to privacy)

Multi-armed bandits



Conclusion

- Bayesian inference is inherently robust and private [hooray].
- Privacy is achieved by posterior sampling [Dimitrakakis et al].
- In certain cases by parameter noise [Zhang et al].
- DP also applicable to bandits [Tossou and Dimitrakakis] - Open problem: Thompson sampling.
- How to tune for unknown constants? (General problem in DP)

References

- C Dwork, F McSherry, K Nissim, A Smith, *Calibrating noise to sensitivity in private data analysis*, TCC 2006.
- C. Dimitrakakis, B. Nelson, A. Mitrokotsa, B. Rubinstein, *Robust and Private Bayesian Inference*, ALT 2014.
- A. Tossou, C. Dimitrakakis, *Algorithms for differentially private multi-armed bandits*, AAAI 2016.
- D. Mir, *Information-theoretic foundations of differential privacy*, EDBT/ICDT, 2012.
- YX. Wang, SE. Fienberg, A. Smola, *Privacy for Free: Posterior Sampling and Stochastic Gradient Monte Carlo*, ICML 2015.
- Z. Zhang, B. Rubinstein, C. Dimitrakakis, *On the Differential Privacy of Bayesian Inference*, AAAI 2016.