Differential privacy

From Bayesian inference to differential privacy and back

Christos Dimitrakakis

September 13, 2022

Overview

Example (Health insurance)

- We collect data x about treatments and patients.
- We disclose conclusions about treatment effectiveness.
- We want to hide individual patient information.
- Encryption does not help

The general problem

- We wish to estimate something from a dataset $x \in \mathcal{S}$.
- We wish to communicate what we learn to a third party.
- How much can they learn about x?

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?

Setting

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.

Setting

Dramatis personae

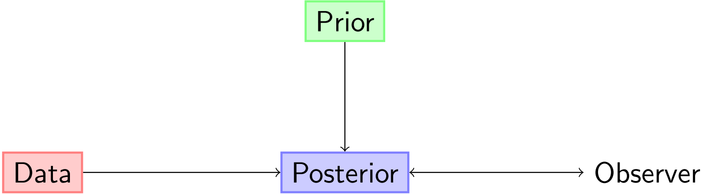
- x – data.
- $\mathscr{B}$ – a (Bayesian) statistician.
- ξ – the statistician's prior belief.
- θ – a parameter
- $\mathscr{A}$ – an adversary. He knows ξ, should not learn x.

The game

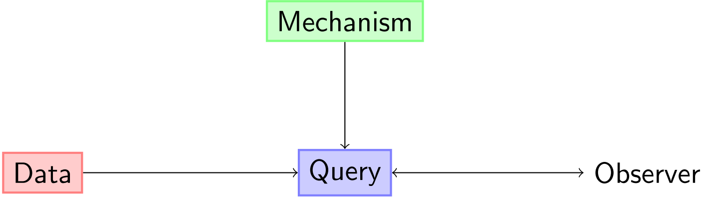
1. $\mathscr{B}$ selects a model family ($\mathcal{F}$) and a prior (ξ).
2. $\mathscr{B}$ observes data x and calculates the posterior $\xi(\theta|x)$.
3. $\mathscr{A}$ queries $\mathscr{B}$.
4. $\mathscr{B}$ responds with a function of the posterior $\xi(\theta|x)$.
5. Goto 3.

Two related problem viewpoints

Bayesian inference view



Mechanism design view



Differential privacy

A randomised mechanism π taking data x as input is basically a distribution condition on x. So we write:

Definition (ϵ-differential privacy)

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

$$\pi(B \mid x) \leq e^{\epsilon} \pi(B \mid y) \qquad ,$$

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

Differential privacy

A randomised mechanism π taking data x as input is basically a distribution condition on x. So we write:

Definition ((ϵ, δ)-differential privacy)

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

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

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

Differential privacy

A randomised mechanism π taking data x as input is basically a distribution condition on x. So we write:

Definition ((ϵ, δ)-differential privacy)

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

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

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

i.e. neighbouring datasets are statistically indistinguishable wrt the distribution induced by the mechanism.

Remark

A similar definition can be given for computationally indistinguishable distributions.

Differential privacy as hypothesis testing

- Assume an adversary wants to distinguish datasets x, y.
- We play a game where we emit a either from $\pi(a|x)$ or $\pi(a|y)$.
- The type I/II errors are bound by DP.

Bayesian properties of Differential privacy

If an adversary has a prior $\beta(x)$ on the data then, by Bayes:

$$\frac{\beta(x|a)}{\beta(x'|a)} = \frac{\pi(a|x)\beta(x)}{\pi(a|x')\beta(x')} \leq e^{\epsilon} \frac{\beta(x)}{\beta(x')}$$

so that, for the case where $\beta(x) = \beta(x')$,

$$\beta(x|a) \leq e^{\epsilon} \beta(x'|a)$$

The necessity of randomness

- Consider a deterministic mechanism $f : \mathcal{S} \to \{0, 1\}$.

The necessity of randomness

- Consider a deterministic mechanism $f : \mathcal{S} \to \{0, 1\}$.
- If there is at least one pair $x, y \in \mathcal{S}$ such that

$$f(x) = 0, \qquad f(y) = 1.$$

then:

The necessity of randomness

- Consider a deterministic mechanism $f : \mathcal{S} \to \{0, 1\}$.
- If there is at least one pair $x, y \in \mathcal{S}$ such that

$$f(x) = 0, \qquad f(y) = 1.$$

  then:
- An adversary $\mathscr{A}$ wants to guess the real data $x^\star$ and knows that $x^* \in \{ x, y \}$ can immediately discover the truth.

Responding to queries

- $\mathscr{B}$ normally responds to queries from $\mathscr{A}$.
- Queries can be defined equivalently as
  1. Additional inputs to the mechanism.
  2. A utility function submitted by $\mathscr{A}$ that $\mathscr{B}$ maximises.
  3. An function submitted by $\mathscr{A}$ that $\mathscr{B}$ evaluates.

Current differentially private mechanisms

Laplace mechanism

Add noise to responses to queries.

$$r = \underbrace{q(x)}_{\text{ideal response}} + \underbrace{\omega}_{\text{noise}}, \qquad \omega \sim \mathcal{L}aplace(\lambda)$$

Exponential mechanism

Define a utility function $u(x, r)$ maximised for $u(x, q(x))$

$$\underbrace{p(r)}_{\text{response probability}} \propto e^{\epsilon u(x,r)} \underbrace{\mu(r)}_{\text{base measure}} .$$

Other methods

- Subsample + aggregate
- Compressed sensing

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.

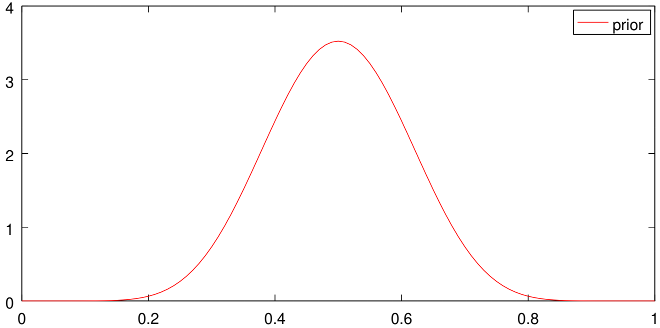


Figure: Prior belief ξ about the coin bias θ.

Bayesian inference

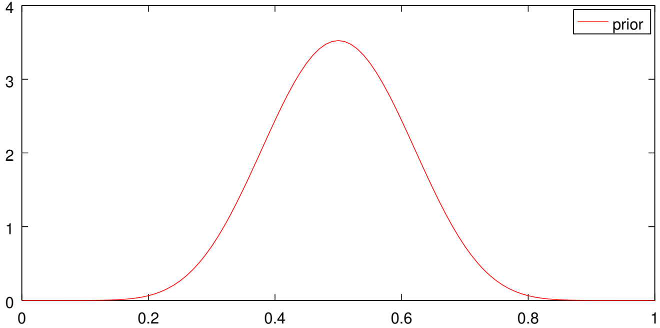


Figure: Prior belief ξ about the coin bias θ.

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

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

Bayesian inference

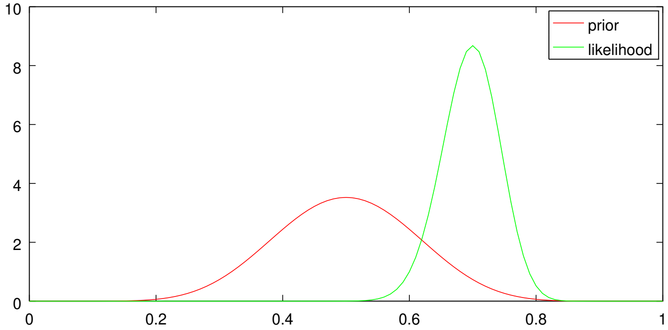


Figure: Prior belief ξ about the coin bias θ and likelihood of θ for the data.

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

Bayesian inference

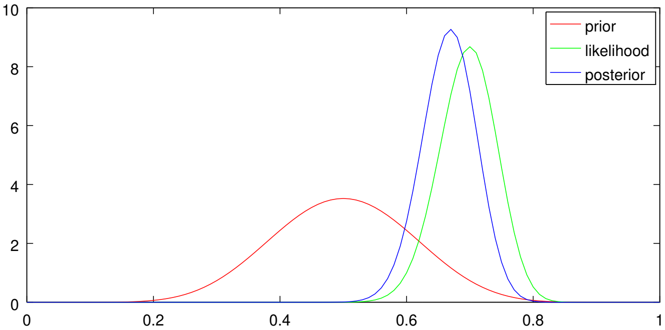


Figure: Prior belief $\xi(\theta)$ about the coin bias θ, likelihood of θ for the data, and posterior belief $\xi(\theta \mid x)$

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.

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.

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

Definition ((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.

We replace the neighbourhood with an apropriate pseudo-metric ρ:

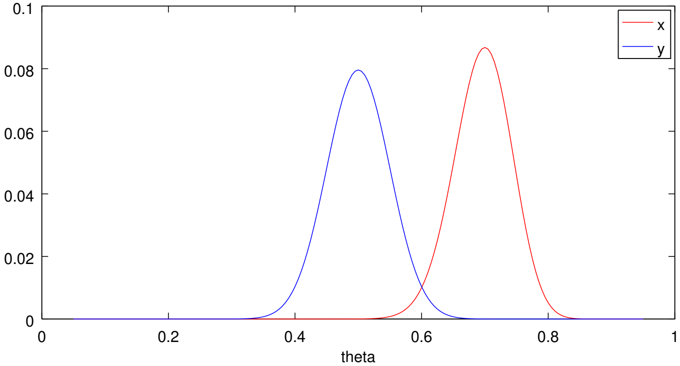
$$x \text{ neighbours } y \qquad \Leftrightarrow \qquad \rho(x, y) \leq 1$$

Sufficient conditions

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

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

$$\left| \ln \frac{P_\theta(x)}{P_\theta(y)} \right| \leq L\rho(x, y), \qquad \forall x, y \in \mathcal{S}, \tag{1}$$

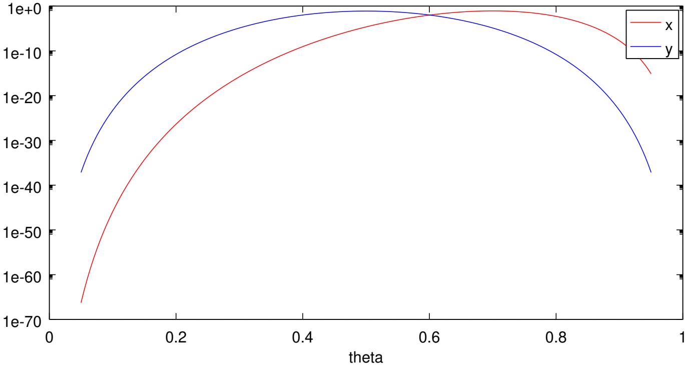


Sufficient conditions

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

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

$$\left| \ln \frac{P_\theta(x)}{P_\theta(y)} \right| \leq L\rho(x,y), \qquad \forall x, y \in \mathcal{S}, \tag{1}$$

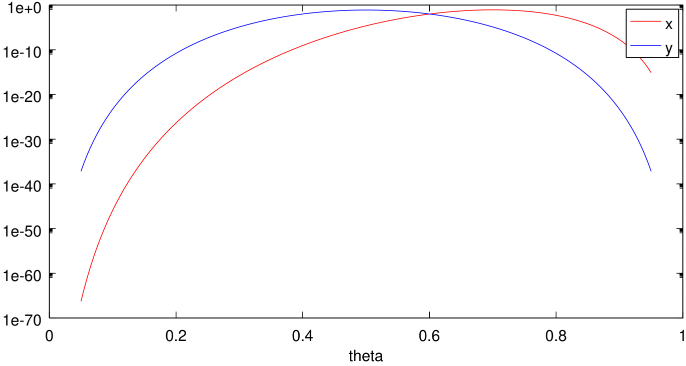


Stochastic Lipschitz condition

Assumption (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.

$$\xi(\Theta_L) \geq 1 - \exp(-cL), \forall L \tag{2}$$

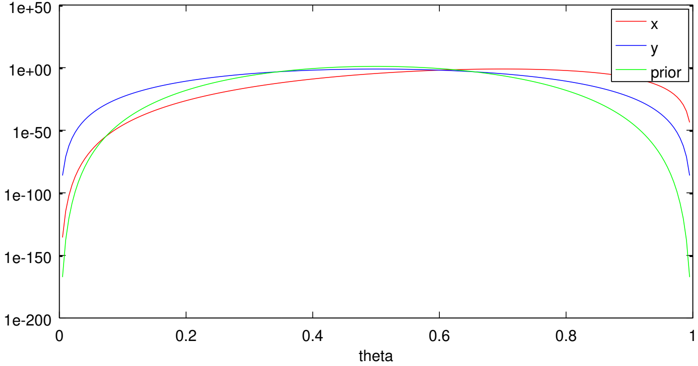


Stochastic Lipschitz condition

Assumption (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.

$$\xi(\Theta_L) \geq 1 - \exp(-cL), \forall L \tag{2}$$



Some properties of the posterior

Robustness of the posterior distribution

$$D\left(\xi(\cdot \mid x) \parallel \xi(\cdot \mid y)\right) \leq O(\rho(x, y)) \tag{3}$$

DP properties of the posterior

1. Assumption 1: the posterior is $(2L, 0)$-DP under ρ.

Some properties of the posterior

Robustness of the posterior distribution

$$D\left(\xi(\cdot \mid x) \parallel \xi(\cdot \mid y)\right) \leq O(\rho(x, y)) \tag{3}$$

DP properties of the posterior

1. Assumption 1: the posterior is $(2L, 0)$-DP under ρ.
2. Assumption 2: 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 calculate any query $q : \Theta \to \mathcal{R}$

$$r_t = q(\theta_t)$$

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 calculate any query $q : \Theta \to \mathcal{R}$

$$r_t = q(\theta_t)$$

Postprocessing: Because the sampling algorithm is DP, the query result is also DP.

Avoiding disclosure with multiple queries

First, release n samples from the posterior

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

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

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

Other mechanisms

Exponential mechanism

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

- 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.

Laplace mechanism

- Add noise to the sufficient statistics of Bayesian inference
- Release complete, noisy, posterior.

Inference under differential privacy

- x: Private data

Inference under differential privacy

- x: Private data
- θ: Latent variable of interest

Inference under differential privacy

- x: Private data
- θ: Latent variable of interest
- $\{P_\theta(x)\}$: Family.

Inference under differential privacy

- x: Private data
- θ: Latent variable of interest
- $\{P_\theta(x)\}$: Family.
- $\beta(\theta)$: Prior.

Inference under differential privacy

- x: Private data
- θ: Latent variable of interest
- $\{P_\theta(x)\}$: Family.
- $\beta(\theta)$: Prior.
- $a \sim \pi(a|x)$: Mechanism output.

Inference under differential privacy

- x: Private data
- θ: Latent variable of interest
- $\{P_\theta(x)\}$: Family.
- $\beta(\theta)$: Prior.
- $a \sim \pi(a|x)$: Mechanism output.

Inference under differential privacy

- x: Private data
- θ: Latent variable of interest
- $\{P_\theta(x)\}$: Family.
- $\beta(\theta)$: Prior.
- $a \sim \pi(a|x)$: Mechanism output.

Inferring θ in general: hard

Using knowledge of the mechanism:

$$\beta(\theta|a,\pi) \propto \beta(a|\theta,\pi)\beta(\theta) = \int_{\mathcal{X}} \pi(a|x)\,\mathrm{d}P_\theta(x) \quad \underbrace{\beta(\theta)}_{MonteCarlo} \tag{4}$$

Inference under differential privacy

- x: Private data
- θ: Latent variable of interest
- $\{P_\theta(x)\}$: Family.
- $\beta(\theta)$: Prior.
- $a \sim \pi(a|x)$: Mechanism output.

Inferring θ in general: hard

Using knowledge of the mechanism:

$$\beta(\theta|a,\pi) \propto \beta(a|\theta,\pi)\beta(\theta) = \int_{\mathcal{X}} \pi(a|x)\,\mathrm{d}P_\theta(x) \quad \underbrace{\beta(\theta)}_{MonteCarlo} \tag{4}$$

When $\pi(a|x)$ is posterior sampling: easy

For any one sample $a \in \Theta$, as long as $\beta = \pi$,

$$\beta(\theta|a,\pi) = \int_{\mathcal{X}} \beta(\theta|x)\,\underbrace{\mathrm{d}P_a(x)}_{MonteCarlo}\,. \tag{5}$$

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].
- Inference under DP generally an open problem.
- DP also applicable to bandits [Thakurta and Smith; Tossou and Dimitrakakis]

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, *Differential privacy for Bayesian inference through posterior sampling*, ALT 2014, JMLR 2017.
- A. Tossou, C. Dimitrakakis, *Algorithms for differentially private multi-armed bandits*, AAAI 2016.
- Z. Zhang, B. Rubinstein, C. Dimitrakakis, *On the Differential Privacy of Bayesian Inference*, AAAI 2016.
- D. Mir, *Information-theoretic foundations of differential privacy*, EDBT/ICDT, 2012.
- A. Thakurta, A. Smith *(nearly) optimal algorithms for private online learning in full-information and bandit settings.* NIPS 2013.
- 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.