Robust and private Bayesian inference[1]

Christos Dimitrakakis[1] Blaine Nelson[2,3]
Aikaterini Mitrokotsa[1] Benjamin Rubinstein[4]

[1]Chalmers university of Technology

[2]University of Potsdam

[3]Google

[4]University of Melbourne

23/10/2014

[1]Appeared at ALT 2014

Overview

- 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 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?

Example (Health insurance)

- We collect data about treatment and patients.
- Disclose treatment effectiveness, but not patient information.

Setting

Dramatis personae

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

Setting

Dramatis personae

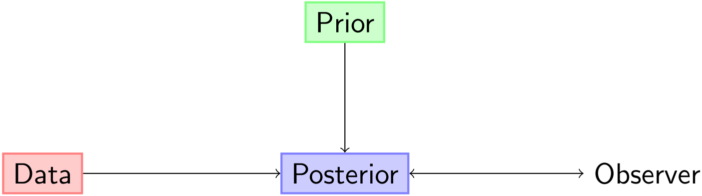
- x – data.
- $\mathscr{B}$ – a (Bayesian) statistician.
- ξ – the statistician's prior.
- θ – a parameter
- $\mathscr{A}$ – an adversary. He knows ξ, should not learn x.

The game

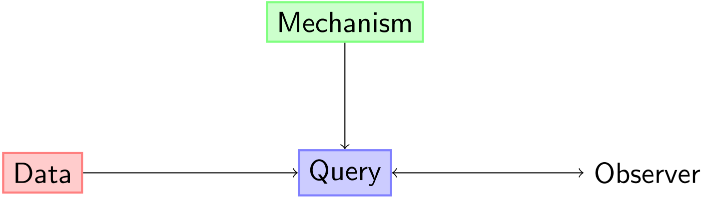
1. $\mathscr{B}$ selects a model family ($\mathcal{F}$) and a prior (ξ).
2. $\mathscr{B}$ observes data x and calculates the posterior $\xi(\theta|x)$.
3. $\mathscr{A}$ queries $\mathscr{B}$.
4. $\mathscr{B}$ responds with a function of the posterior $\xi(\theta|x)$.
5. Goto 3.

Two related problem viewpoints

Bayesian inference view



Mechanism design view



Privacy properties of Bayesian inference

- Bayesian inference
- Robustness of the posterior distribution
- A query model

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 $\mathcal{S}$.
- Prior distribution ξ on Θ.
- Posterior given data $x \in \mathcal{S}$:

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

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

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 ((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 ((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 ((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 ($\mathcal{F}$ is well-behaved)

i.e. all members of $\mathcal{F}$ are L-Lipschitz.

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

Sufficient conditions

Assumption ($\mathcal{F}$ is well-behaved)

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

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

i.e. all members of $\mathcal{F}$ *are L-Lipschitz.*

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

Sufficient conditions

Assumption ($\mathcal{F}$ is well-behaved)

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

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

i.e. all members of $\mathcal{F}$ are L-Lipschitz.

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. $\forall L \geq 0$:

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

When do these assumptions hold?

Example (Exponential families)

Family $\mathcal{F}$, with sufficient statistic T.

$$p_\theta(x) = h(x) \exp\left\{\eta_\theta^\top T(x) - A(\eta_\theta)\right\}$$

For a given $\theta \in \Theta$, we want to test if:

$$\left|\ln(h(x)/h(y)) + \eta_\theta^\top (T(x) - T(y))\right| \leq L\rho(x, y), \qquad \forall x, y \in \mathcal{X} \ .$$

Example (Exponential distribution)

Exponential prior with parameter $c > 0$, satisfies Assumption 2.

Example (Discrete Bayesian networks)

$$\left| \ln \frac{P_\theta(x)}{P_\theta(y)} \right| \leq \ln \frac{1}{\epsilon} \cdot \rho(x, y),$$

where: $\rho(x, y)$ is the number of edges and nodes influenced by the differences in x, y and ϵ the smallest P_θ-mass placed to any event.

Robustness of the posterior distribution

Definition (KL divergence)

$$D\left(P \parallel Q\right) \triangleq \int \ln \frac{\mathrm{d}P}{\mathrm{d}Q} \, \mathrm{d}P. \tag{3}$$

Theorem

Robustness of the posterior distribution

Definition (KL divergence)

$$D\left(P \parallel Q\right) \triangleq \int \ln \frac{\mathrm{d}P}{\mathrm{d}Q}\,\mathrm{d}P. \tag{3}$$

Theorem

(i) *Under Assumption 1,*

$$D\left(\xi(\cdot \mid x) \parallel \xi(\cdot \mid y)\right) \leq 2L\rho(x, y) \tag{4}$$

Robustness of the posterior distribution

Definition (KL divergence)

$$D\left(P \parallel Q\right) \triangleq \int \ln \frac{\mathrm{d}P}{\mathrm{d}Q}\,\mathrm{d}P. \tag{3}$$

Theorem

(i) *Under Assumption 1,*

$$D\left(\xi(\cdot \mid x) \parallel \xi(\cdot \mid y)\right) \leq 2L\rho(x, y) \tag{4}$$

(ii) *Under Assumption 2,*

$$D\left(\xi(\cdot \mid x) \parallel \xi(\cdot \mid y)\right) \leq \frac{\kappa}{c} \cdot \rho(x, y) \tag{5}$$

Differential privacy of the posterior

Theorem

1. *Under Assumption 1, $B \in \mathfrak{S}_\Theta$:*

$$\xi(B \mid x) \leq e^{2L\rho(x,y)} \xi(B \mid y) \tag{6}$$

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

Differential privacy of the posterior

Theorem

1. *Under Assumption 1, $B \in \mathfrak{S}_\Theta$:*

$$\xi(B \mid x) \leq e^{2L\rho(x,y)} \xi(B \mid y) \tag{6}$$

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

2. *Under Assumption 2, for all $x, y \in \mathcal{S}$, $B \in \mathfrak{S}_\Theta$:*

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

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

A query model

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

A query model

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

Access model

At time t, the adversary observes a sample from our posterior distribution.

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

A query model

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

Access model

At time t, the adversary observes a sample from our posterior distribution.

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

More generally, $\mathscr{A}$ can select a question $q : \Theta \to \mathcal{R}$, where $\mathcal{R}$ is a response space:

$$r_t = q(\theta_t)$$

Other mechanisms

Laplace mechanism

Add noise to responses to queries.

$$r = q(x) + \omega, \qquad \omega \sim \mathcal{L}aplace(\lambda)$$

Exponential mechanism

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

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

Subsampling

Perform inference on a random sample of the data.

Reduction of privacy to a testing problem

How fast can the adversary learn?

- Different datasets x, y give different $\xi(\cdot \mid x), \xi(\cdot \mid y)$.
- How many samples are needed to differentiate them?

Reduction of privacy to a testing problem

How fast can the adversary learn?

- Different datasets x, y give different $\xi(\cdot \mid x), \xi(\cdot \mid y)$.
- How many samples are needed to differentiate them?

Theorem

Under Assumption 1, the adversary can distinguish between data x, y with probability $1 - \delta$ if:

$$\rho(x, y) \geq \frac{3}{4Ln} \ln \frac{1}{\delta}. \tag{7}$$

Under Assumption 2, this becomes:

$$\rho(x, y) \geq \frac{3c}{2\kappa n} \ln \frac{1}{\delta}. \tag{8}$$

The Le Cam method for lower bounds

Idea: Use $\mathcal{S}$ for the "parameter" space of an estimator.

The Le Cam method for lower bounds

Idea: Use $\mathcal{S}$ for the "parameter" space of an estimator.

The family of posterior measures

$$\varXi \triangleq \{\, \xi(\cdot \mid x) \mid x \in \mathcal{S} \,\}, \tag{9}$$

The Le Cam method for lower bounds

Idea: Use $\mathcal{S}$ for the "parameter" space of an estimator.

The family of posterior measures

$$\Xi \triangleq \{\, \xi(\cdot \mid x) \mid x \in \mathcal{S} \,\}, \tag{9}$$

Lemma (Le Cam's method)

Let $\psi(\theta)$ be an estimator of x. Let $\mathcal{S}_1, \mathcal{S}_2$ be 2δ-separated and say $x \in S_i \Rightarrow \xi(\cdot \mid x) \in \Xi_i \subset \Xi$. Then:

$$\sup_{x \in \mathcal{S}} \mathbb{E}_{\xi}(\rho(\psi, x) \mid x) \geq \delta \sup_{\xi_i \in co(\Xi_i)} \|\xi_1 \wedge \xi_2\|. \tag{10}$$

The Le Cam method for lower bounds

Idea: Use $\mathcal{S}$ for the "parameter" space of an estimator.

The family of posterior measures

$$\Xi \triangleq \{ \xi(\cdot \mid x) \mid x \in \mathcal{S} \}, \tag{9}$$

Lemma (Le Cam's method)

Let $\psi(\theta)$ be an estimator of x. Let $\mathcal{S}_1, \mathcal{S}_2$ be 2δ-separated and say $x \in S_i \Rightarrow \xi(\cdot \mid x) \in \Xi_i \subset \Xi$. Then:

$$\sup_{x \in \mathcal{S}} \mathbb{E}_\xi(\rho(\psi, x) \mid x) \geq \delta \sup_{\xi_i \in co(\Xi_i)} \|\xi_1 \wedge \xi_2\|. \tag{10}$$

Expected distance between the real and guessed data:

$$\mathbb{E}_\xi(\rho(\psi, x) \mid x) = \int_\Theta \rho(\psi(\theta), x) \, \mathrm{d}\xi(\theta \mid x),$$

Conclusion

- Bayesian inference is inherently robust and private.
- It suffices to select the right prior. But how?
- In some cases, this is closed form.
- The general case is an open problem.
- Do we need to randomise at all?

Job ad

PhD student in differential privacy and distributed decision making.