A Short Tutorial on Differential Privacy

Borja Balle

Amazon Research Cambridge

The Alan Turing Institute — January 26, 2018

Outline

1. We Need Mathematics to Study Privacy? Seriously?
2. Differential Privacy: Definition, Properties and Basic Mechanisms
3. Differentially Private Machine Learning: ERM and Bayesian Learning
4. Variations on Differential Privacy: Concentrated DP and Local DP
5. Final Remarks

Outline

Anonymization Fiascos

Disturbing Headlines and Paper Titles

- "A Face Is Exposed for AOL Searcher No. 4417749" [Barbaro & Zeller '06]
- "Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset)" [Narayanan & Shmatikov '08]
- "Matching Known Patients to Health Records in Washington State Data" [Sweeney '13]
- "Harvard Professor Re-Identifies Anonymous Volunteers In DNA Study" [Sweeney et al. '13]
- ... and many others

In general, removing identifiers and applying anonymization heuristics is not always enough!

Why is Anonymization Hard?

- High-dimensional/high-resolution data is essentially unique:

office	*department*	*date joined*	*salary*	*d.o.b.*	*nationality*	*gender*
London	IT	Apr 2015	£####	May 1985	Portuguese	Female

- Lower dimension and lower resolution is more private, but less useful:

office	*department*	*date joined*	*salary*	*d.o.b.*	*nationality*	*gender*
UK	IT	2015	£####	1980-1985	—	Female

research

Why is Anonymization Hard?

- High-dimensional/high-resolution data is essentially unique:

office	*department*	*date joined*	*salary*	*d.o.b.*	*nationality*	*gender*
London	IT	Apr 2015	£####	May 1985	Portuguese	Female

- Lower dimension and lower resolution is more private, but less useful:

office	*department*	*date joined*	*salary*	*d.o.b.*	*nationality*	*gender*
UK	IT	2015	£####	1980-1985	—	Female

research

Managing Expectations

Unreasonable Privacy Expectations

- *Privacy for free?* No, privatizing requires removing information ($\Rightarrow$ accuracy loss)
- *Absolute privacy?* No, your neighbour's habits are correlated with your habits

Reasonable Privacy Expectations

- *Quantitative:* offer a knob to tune accuracy vs. privacy loss
- *Plausible deniability:* your presence in a database cannot be ascertained
- *Prevent targeted attacks:* limit information leaked even in the presence of side knowledge

The Promise of Differential Privacy

Quote from [Dwork and Roth, 2014]:

> *Differential privacy describes a promise, made by a data holder, or curator, to a data subject: "You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available."*

Quotes from the 2017 Gödel Prize citation awarded to Dwork, McSherry, Nissim and Smith:

> *Differential privacy was carefully constructed to avoid numerous and subtle pitfalls that other attempts at defining privacy have faced.*

> *The intellectual impact of differential privacy has been broad, with influence on the thinking about privacy being noticeable in a huge range of disciplines, ranging from traditional areas of computer science (databases, machine learning, networking, security) to economics and game theory, false discovery control, official statistics and econometrics, information theory, genomics and, recently, law and policy.*

Outline

Differential Privacy

Ingredients

- Input space X (with symmetric neighbouring relation $\simeq$)
- Output space Y (with σ-algebra of measurable events)
- Privacy parameter $\varepsilon \geqslant 0$

Differential Privacy [Dwork et al., 2006, Dwork, 2006]

A randomized mechanism $\mathcal{M} : X \to Y$ is ε-differentially private if for all neighbouring inputs $x \simeq x'$ and for all sets of outputs $E \subseteq Y$ we have

$$\mathbb{P}[\mathcal{M}(x) \in E] \leqslant e^{\varepsilon} \mathbb{P}[\mathcal{M}(x') \in E]$$

Intuitions behind the definition:

- The neighbouring relation $\simeq$ captures *what* is protected
- The probability bounds capture *how much* protection we get

Differential Privacy

Ingredients

- Input space X (with symmetric neighbouring relation $\simeq$)
- Output space Y (with σ-algebra of measurable events)
- Privacy parameter $\varepsilon \geqslant 0$

Differential Privacy **[Dwork et al., 2006, Dwork, 2006]**

A randomized mechanism $\mathcal{M} : X \to Y$ is ε-differentially private if for all neighbouring inputs $x \simeq x'$ and for all sets of outputs $E \subseteq Y$ we have

$$\mathbb{P}[\mathcal{M}(x) \in E] \leqslant e^{\varepsilon} \mathbb{P}[\mathcal{M}(x') \in E]$$

Intuitions behind the definition:

- The neighbouring relation $\simeq$ captures *what* is protected
- The probability bounds capture *how much* protection we get

research

Differential Privacy

Ingredients

- Input space X (with symmetric neighbouring relation $\simeq$)
- Output space Y (with σ-algebra of measurable events)
- Privacy parameter $\varepsilon \geqslant 0$

Differential Privacy **[Dwork et al., 2006, Dwork, 2006]**

A randomized mechanism $\mathcal{M} : X \to Y$ is ε-differentially private if for all neighbouring inputs $x \simeq x'$ and for all sets of outputs $E \subseteq Y$ we have

$$\mathbb{P}[\mathcal{M}(x) \in E] \leqslant e^{\varepsilon} \mathbb{P}[\mathcal{M}(x') \in E]$$

Intuitions behind the definition:

- The neighbouring relation $\simeq$ captures *what* is protected
- The probability bounds capture *how much* protection we get

research

Differential Privacy

Ingredients

- Input space X (with symmetric neighbouring relation $\simeq$)
- Output space Y (with σ-algebra of measurable events)
- Privacy parameter $\varepsilon \geqslant 0$

Differential Privacy [Dwork et al., 2006, Dwork, 2006]

A randomized mechanism $\mathcal{M} : X \to Y$ is ε-differentially private if for all neighbouring inputs $x \simeq x'$ and for all sets of outputs $E \subseteq Y$ we have

$$\mathbb{P}[\mathcal{M}(x) \in E] \leqslant e^{\varepsilon} \mathbb{P}[\mathcal{M}(x') \in E]$$

Intuitions behind the definition:

- The neighbouring relation $\simeq$ captures *what* is protected
- The probability bounds capture *how much* protection we get

research

Differential Privacy

Ingredients

- Input space X (with symmetric neighbouring relation $\simeq$)
- Output space Y (with σ-algebra of measurable events)
- Privacy parameter $\varepsilon \geqslant 0$

Differential Privacy **[Dwork et al., 2006, Dwork, 2006]**

A randomized mechanism $\mathcal{M} : X \to Y$ is ε-differentially private if for all neighbouring inputs $x \simeq x'$ and for all sets of outputs $E \subseteq Y$ we have

$$\mathbb{P}[\mathcal{M}(x) \in E] \leqslant e^{\varepsilon} \mathbb{P}[\mathcal{M}(x') \in E]$$

Intuitions behind the definition:

- The neighbouring relation $\simeq$ captures *what* is protected
- The probability bounds capture *how much* protection we get

research

Differential Privacy

Ingredients

- Input space X (with symmetric neighbouring relation $\simeq$)
- Output space Y (with σ-algebra of measurable events)
- Privacy parameter $\varepsilon \geqslant 0$

Differential Privacy **[Dwork et al., 2006, Dwork, 2006]**

A randomized mechanism $\mathcal{M} : X \to Y$ is ε-differentially private if for all neighbouring inputs $x \simeq x'$ and for all sets of outputs $E \subseteq Y$ we have

$$\mathbb{P}[\mathcal{M}(x) \in E] \leqslant e^{\varepsilon}\mathbb{P}[\mathcal{M}(x') \in E]$$

Intuitions behind the definition:

- The neighbouring relation $\simeq$ captures *what* is protected
- The probability bounds capture *how much* protection we get

Differential Privacy

Ingredients

- Input space X (with symmetric neighbouring relation $\simeq$)
- Output space Y (with σ-algebra of measurable events)
- Privacy parameter $\varepsilon \geqslant 0$

Differential Privacy **[Dwork et al., 2006, Dwork, 2006]**

A randomized mechanism $\mathcal{M} : X \to Y$ is ε-differentially private if for all neighbouring inputs $x \simeq x'$ and for all sets of outputs $E \subseteq Y$ we have

$$\mathbb{P}[\mathcal{M}(x) \in E] \leqslant e^{\varepsilon} \mathbb{P}[\mathcal{M}(x') \in E]$$

Intuitions behind the definition:

- The neighbouring relation $\simeq$ captures *what* is protected
- The probability bounds capture *how much* protection we get

research

DP before DP: Randomized Response

The Randomized Response Mechanism **[Warner, 1965]**

- n individuals answer a survey with one binary question
- The truthful answer for individual i is $x_i \in \{0, 1\}$
- Each individual answers truthfully ($y_i = x_i$) with probability $e^\varepsilon/(1 + e^\varepsilon)$ and falsely ($y_i = \bar{x}_i$) with probability $1/(1 + e^\varepsilon)$
- Let's denote the mechanism by $(y_1, \dots, y_n) = RR_\varepsilon(x_1, \dots, x_n)$

Intuition: Provides plausible deniability for each individual's answer

Claim: RR_ε is ε-DP *(free-range organic proof on the whiteboard)*

Utility: Averaging the (unbiased) answers $\tilde{y}_i$ from RR_ε satisfies w.h.p.

$$\left| \frac{1}{n} \sum_{i=1}^{n} x_i - \frac{1}{n} \sum_{i=1}^{n} \tilde{y}_i \right| \leqslant \mathcal{O}\left(\frac{1}{\varepsilon \sqrt{n}} \right)$$

research

DP before DP: Randomized Response

The Randomized Response Mechanism **[Warner, 1965]**

- n individuals answer a survey with one binary question
- The truthful answer for individual i is $x_i \in \{0, 1\}$
- Each individual answers truthfully ($y_i = x_i$) with probability $e^\varepsilon/(1 + e^\varepsilon)$ and falsely ($y_i = \bar{x}_i$) with probability $1/(1 + e^\varepsilon)$
- Let's denote the mechanism by $(y_1, \ldots, y_n) = RR_\varepsilon(x_1, \ldots, x_n)$

Intuition: Provides plausible deniability for each individual's answer

Claim: RR_ε is ε-DP *(free-range organic proof on the whiteboard)*

Utility: Averaging the (unbiased) answers $\tilde{y}_i$ from RR_ε satisfies w.h.p.

$$\left|\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} \tilde{y}_i\right| \leqslant \mathcal{O}\left(\frac{1}{\varepsilon\sqrt{n}}\right)$$

research

DP before DP: Randomized Response

The Randomized Response Mechanism **[Warner, 1965]**

- n individuals answer a survey with one binary question
- The truthful answer for individual i is $x_i \in \{0, 1\}$
- Each individual answers truthfully ($y_i = x_i$) with probability $e^{\varepsilon}/(1 + e^{\varepsilon})$ and falsely ($y_i = \bar{x}_i$) with probability $1/(1 + e^{\varepsilon})$
- Let's denote the mechanism by $(y_1, \ldots, y_n) = RR_{\varepsilon}(x_1, \ldots, x_n)$

Intuition: Provides plausible deniability for each individual's answer

Claim: RR_{ε} is ε-DP *(free-range organic proof on the whiteboard)*

Utility: Averaging the (unbiased) answers $\tilde{y}_i$ from RR_{ε} satisfies w.h.p.

$$\left| \frac{1}{n} \sum_{i=1}^{n} x_i - \frac{1}{n} \sum_{i=1}^{n} \tilde{y}_i \right| \leqslant \mathcal{O}\left(\frac{1}{\varepsilon \sqrt{n}} \right)$$

research

DP before DP: Randomized Response

The Randomized Response Mechanism **[Warner, 1965]**

- n individuals answer a survey with one binary question
- The truthful answer for individual i is $x_i \in \{0, 1\}$
- Each individual answers truthfully ($y_i = x_i$) with probability $e^{\varepsilon}/(1 + e^{\varepsilon})$ and falsely ($y_i = \bar{x}_i$) with probability $1/(1 + e^{\varepsilon})$
- Let's denote the mechanism by $(y_1, \ldots, y_n) = RR_{\varepsilon}(x_1, \ldots, x_n)$

Intuition: Provides plausible deniability for each individual's answer

Claim: RR_{ε} is ε-DP *(free-range organic proof on the whiteboard)*

Utility: Averaging the (unbiased) answers $\tilde{y}_i$ from RR_{ε} satisfies w.h.p.

$$\left| \frac{1}{n} \sum_{i=1}^{n} x_i - \frac{1}{n} \sum_{i=1}^{n} \tilde{y}_i \right| \leqslant \mathcal{O}\left(\frac{1}{\varepsilon \sqrt{n}} \right)$$

research

The Laplace Mechanism (for computing the mean)

Private Mean Computation

- A curator holds one bit $x_i \in \{0, 1\}$ for each of n individuals
- The curator proceeds by
  1. Computing the mean $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$,
  2. Sampling noise $Z \sim \mathrm{Lap}(\frac{1}{\varepsilon n})$, and
  3. Revealing the noisy mean $\tilde{\mu} = \mu + Z$
- Let's denote the mechanism by $\tilde{\mu} = \mathcal{M}_{\mathrm{Lap}}(x_1, \ldots, x_n)$

Claim: $\mathcal{M}_{\mathrm{Lap}}$ is ε-DP *(free-range organic proof on the whiteboard)*

Utility: The answer returned by the mechanism satisfies w.h.p.

$$|\mu - \tilde{\mu}| \leqslant \mathcal{O}\left(\frac{1}{\varepsilon n}\right)$$

research

The Laplace Mechanism (for computing the mean)

Private Mean Computation

- A curator holds one bit $x_i \in \{0, 1\}$ for each of n individuals
- The curator proceeds by
  1. Computing the mean $\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$,
  2. Sampling noise $Z \sim \mathrm{Lap}(\frac{1}{\varepsilon n})$, and
  3. Revealing the noisy mean $\tilde{\mu} = \mu + Z$
- Let's denote the mechanism by $\tilde{\mu} = \mathcal{M}_{\mathrm{Lap}}(x_1, \ldots, x_n)$

Claim: $\mathcal{M}_{\mathrm{Lap}}$ is ε-DP *(free-range organic proof on the whiteboard)*

Utility: The answer returned by the mechanism satisfies w.h.p.

$$|\mu - \tilde{\mu}| \leqslant \mathcal{O}\left(\frac{1}{\varepsilon n}\right)$$

research

The Laplace Mechanism (for computing the mean)

Private Mean Computation

- A curator holds one bit $x_i \in \{0, 1\}$ for each of n individuals
- The curator proceeds by
  1. Computing the mean $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$,
  2. Sampling noise $Z \sim \mathrm{Lap}(\frac{1}{\varepsilon n})$, and
  3. Revealing the noisy mean $\tilde{\mu} = \mu + Z$
- Let's denote the mechanism by $\tilde{\mu} = \mathcal{M}_{\mathrm{Lap}}(x_1, \ldots, x_n)$

Claim: $\mathcal{M}_{\mathrm{Lap}}$ is ε-DP *(free-range organic proof on the whiteboard)*

Utility: The answer returned by the mechanism satisfies w.h.p.

$$|\mu - \tilde{\mu}| \leqslant \mathcal{O}\left(\frac{1}{\varepsilon n}\right)$$

Approximate Differential Privacy

Ingredients

- Input space X (with symmetric neighbouring relation $\simeq$)
- Output space Y (with sigma-algebra of measurable events)
- Privacy parameters $\varepsilon \geqslant 0$, $\delta \in [0, 1]$

Approximate Differential Privacy

A randomized mechanism $\mathcal{M} : X \to Y$ is (ε, δ)-differentially private if for all neighbouring inputs $x \simeq x'$ and for all sets of outputs $E \subseteq Y$ we have

$$\mathbb{P}[\mathcal{M}(x) \in E] \leqslant e^{\varepsilon} \mathbb{P}[\mathcal{M}(x') \in E] + \delta$$

Interpretation

- δ accounts for "bad events" that might result in high privacy losses
- Mechanism $\mathcal{M}(x_1, \ldots, x_n) = x_{\mathrm{Unif}([n])}$ is $(0, 1/n)$-DP ($\Rightarrow$ should take $\delta \ll 1/n$)

research

Approximate Differential Privacy

Ingredients

- Input space X (with symmetric neighbouring relation $\simeq$)
- Output space Y (with sigma-algebra of measurable events)
- Privacy parameters $\varepsilon \geqslant 0$, $\delta \in [0, 1]$

Approximate Differential Privacy

A randomized mechanism $\mathcal{M} : X \to Y$ is (ε, δ)-differentially private if for all neighbouring inputs $x \simeq x'$ and for all sets of outputs $E \subseteq Y$ we have

$$\mathbb{P}[\mathcal{M}(x) \in E] \leqslant e^{\varepsilon} \mathbb{P}[\mathcal{M}(x') \in E] + \delta$$

Interpretation

- δ accounts for "bad events" that might result in high privacy losses
- Mechanism $\mathcal{M}(x_1, \ldots, x_n) = x_{\mathrm{Unif}([n])}$ is $(0, 1/n)$-DP ($\Rightarrow$ should take $\delta \ll 1/n$)

Approximate Differential Privacy

Ingredients

- Input space X (with symmetric neighbouring relation $\simeq$)
- Output space Y (with sigma-algebra of measurable events)
- Privacy parameters $\varepsilon \geqslant 0$, $\delta \in [0, 1]$

Approximate Differential Privacy

A randomized mechanism $\mathcal{M} : X \to Y$ is (ε, δ)-differentially private if for all neighbouring inputs $x \simeq x'$ and for all sets of outputs $E \subseteq Y$ we have

$$\mathbb{P}[\mathcal{M}(x) \in E] \leqslant e^{\varepsilon}\mathbb{P}[\mathcal{M}(x') \in E] + \delta$$

Interpretation

- δ accounts for "bad events" that might result in high privacy losses
- Mechanism $\mathcal{M}(x_1, \ldots, x_n) = x_{\mathrm{Unif}([n])}$ is $(0, 1/n)$-DP ($\Rightarrow$ should take $\delta \ll 1/n$)

research

Output Perturbation Mechanisms

The Laplace mechanism is an example of a more general class of mechanisms

Global Sensitivity: for any function $f : X \to \mathbb{R}^d$ define $\Delta_p = \sup_{x \simeq x'} \|f(x) - f(x')\|_p$

Output Perturbation (with Laplace and Gaussian noise)

- A curator holds one vector $x_i \in \mathbb{R}^d$ for each of n individuals
- The curator computes a function $f(x_1, \ldots, x_n)$ of the data,
- samples noise $Z \sim \mathrm{Lap}(\frac{\Delta_1}{\varepsilon})^d$ or $Z \sim \mathcal{N}(0, \sigma^2)^d$ with $\sigma = \frac{\Delta_2 \sqrt{C \log(1/\delta)}}{\varepsilon}$, and
- reveals the noisy value $f(x_1, \ldots, x_n) + Z$
- Let's denote the mechanisms $\mathcal{M}_{f,\mathrm{Lap}}$ and $\mathcal{M}_{f,\mathcal{N}}$ respectively
- Note the mechanism of the previous slide is $\mathcal{M}_{f,\mathrm{Lap}}$ for $f(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i$

Claim: $\mathcal{M}_{f,\mathrm{Lap}}$ is ε-DP and $\mathcal{M}_{f,\mathcal{N}}$ is (ε, δ)-DP

research

Output Perturbation Mechanisms

The Laplace mechanism is an example of a more general class of mechanisms

Global Sensitivity: for any function $f : X \to \mathbb{R}^d$ define $\Delta_p = \sup_{x \simeq x'} \|f(x) - f(x')\|_p$

Output Perturbation (with Laplace and Gaussian noise)

- A curator holds one vector $x_i \in \mathbb{R}^d$ for each of n individuals
- The curator computes a function $f(x_1, \ldots, x_n)$ of the data,
- samples noise $Z \sim \mathrm{Lap}(\frac{\Delta_1}{\varepsilon})^d$ or $Z \sim \mathcal{N}(0, \sigma^2)^d$ with $\sigma = \frac{\Delta_2 \sqrt{C \log(1/\delta)}}{\varepsilon}$, and
- reveals the noisy value $f(x_1, \ldots, x_n) + Z$
- Let's denote the mechanisms $\mathcal{M}_{f,\mathrm{Lap}}$ and $\mathcal{M}_{f,\mathcal{N}}$ respectively
- Note the mechanism of the previous slide is $\mathcal{M}_{f,\mathrm{Lap}}$ for $f(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i$

Claim: $\mathcal{M}_{f,\mathrm{Lap}}$ is ε-DP and $\mathcal{M}_{f,\mathcal{N}}$ is (ε, δ)-DP

research

Output Perturbation Mechanisms

The Laplace mechanism is an example of a more general class of mechanisms

Global Sensitivity: for any function $f : X \to \mathbb{R}^d$ define $\Delta_p = \sup_{x \simeq x'} \|f(x) - f(x')\|_p$

Output Perturbation (with Laplace and Gaussian noise)

- A curator holds one vector $x_i \in \mathbb{R}^d$ for each of n individuals
- The curator computes a function $f(x_1, \ldots, x_n)$ of the data,
- samples noise $Z \sim \mathrm{Lap}(\frac{\Delta_1}{\varepsilon})^d$ or $Z \sim \mathcal{N}(0, \sigma^2)^d$ with $\sigma = \frac{\Delta_2 \sqrt{C \log(1/\delta)}}{\varepsilon}$, and
- reveals the noisy value $f(x_1, \ldots, x_n) + Z$
- Let's denote the mechanisms $\mathcal{M}_{f,\mathrm{Lap}}$ and $\mathcal{M}_{f,\mathcal{N}}$ respectively
- Note the mechanism of the previous slide is $\mathcal{M}_{f,\mathrm{Lap}}$ for $f(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i$

Claim: $\mathcal{M}_{f,\mathrm{Lap}}$ is ε-DP and $\mathcal{M}_{f,\mathcal{N}}$ is (ε, δ)-DP

research

Fundamental Properties

- Robustness to post-processing: $\mathcal{M}$ is (ε, δ)-DP, then $F \circ \mathcal{M}$ is (ε, δ)-DP
- Composition: if $\mathcal{M}_j$, $j = 1, \ldots, k$, are $(\varepsilon_j, \delta_j)$-DP, then $\vec{x} \mapsto (\mathcal{M}_1(\vec{x}), \ldots, \mathcal{M}_k(\vec{x}))$ is $(\sum_j \varepsilon_j, \sum_j \delta_j)$-DP. In the homogeneous case this yields $(k\varepsilon, k\delta)$-DP
- Advanced composition: if $\mathcal{M}_j$, $j = 1, \ldots, k$, are (ε, δ)-DP, then $\vec{x} \mapsto (\mathcal{M}_1(\vec{x}), \ldots, \mathcal{M}_k(\vec{x}))$ is $(\varepsilon\sqrt{k \log(1/\delta')} + \varepsilon(e^\varepsilon - 1)k, k\delta + \delta')$-DP for any $\delta' > 0$
- Group privacy: if $\mathcal{M}$ is (ε, δ)-DP with respect to $x \simeq x'$, then $\mathcal{M}$ is $(t\varepsilon, t\delta)$ with respect to $x \simeq^t x'$ (ie. t changes)
- Protects against side knowledge: if attacker has prior $P^{x_i}_{prior}$ and computes $P^{x_i}_{posterior}$ after observing $\mathcal{M}(\vec{x})$ from ε-DP mechanism, then $\mathrm{dist}(P^{x_i}_{prior}, P^{x_i}_{posterior}) = \mathcal{O}(\varepsilon)$

research

Fundamental Properties

- Robustness to post-processing: $\mathcal{M}$ is (ε, δ)-DP, then $F \circ \mathcal{M}$ is (ε, δ)-DP
- Composition: if $\mathcal{M}_j$, $j = 1, \ldots, k$, are $(\varepsilon_j, \delta_j)$-DP, then $\vec{x} \mapsto (\mathcal{M}_1(\vec{x}), \ldots, \mathcal{M}_k(\vec{x}))$ is $(\sum_j \varepsilon_j, \sum_j \delta_j)$-DP. In the homogeneous case this yields $(k\varepsilon, k\delta)$-DP
- Advanced composition: if $\mathcal{M}_j$, $j = 1, \ldots, k$, are (ε, δ)-DP, then $\vec{x} \mapsto (\mathcal{M}_1(\vec{x}), \ldots, \mathcal{M}_k(\vec{x}))$ is $(\varepsilon\sqrt{k \log(1/\delta')} + \varepsilon(e^\varepsilon - 1)k, k\delta + \delta')$-DP for any $\delta' > 0$
- Group privacy: if $\mathcal{M}$ is (ε, δ)-DP with respect to $x \simeq x'$, then $\mathcal{M}$ is $(t\varepsilon, t\delta)$ with respect to $x \simeq^t x'$ (ie. t changes)
- Protects against side knowledge: if attacker has prior $P^{x_i}_{prior}$ and computes $P^{x_i}_{posterior}$ after observing $\mathcal{M}(\vec{x})$ from ε-DP mechanism, then $\mathrm{dist}(P^{x_i}_{prior}, P^{x_i}_{posterior}) = \mathcal{O}(\varepsilon)$

Fundamental Properties

- Robustness to post-processing: $\mathcal{M}$ is (ε, δ)-DP, then $F \circ \mathcal{M}$ is (ε, δ)-DP
- Composition: if $\mathcal{M}_j$, $j = 1, \ldots, k$, are $(\varepsilon_j, \delta_j)$-DP, then $\vec{x} \mapsto (\mathcal{M}_1(\vec{x}), \ldots, \mathcal{M}_k(\vec{x}))$ is $(\sum_j \varepsilon_j, \sum_j \delta_j)$-DP. In the homogeneous case this yields $(k\varepsilon, k\delta)$-DP
- Advanced composition: if $\mathcal{M}_j$, $j = 1, \ldots, k$, are (ε, δ)-DP, then $\vec{x} \mapsto (\mathcal{M}_1(\vec{x}), \ldots, \mathcal{M}_k(\vec{x}))$ is $(\varepsilon\sqrt{k \log(1/\delta')} + \varepsilon(e^\varepsilon - 1)k, k\delta + \delta')$-DP for any $\delta' > 0$
- Group privacy: if $\mathcal{M}$ is (ε, δ)-DP with respect to $x \simeq x'$, then $\mathcal{M}$ is $(t\varepsilon, t\delta)$ with respect to $x \simeq^t x'$ (ie. t changes)
- Protects against side knowledge: if attacker has prior $P^{x_i}_{prior}$ and computes $P^{x_i}_{posterior}$ after observing $\mathcal{M}(\vec{x})$ from ε-DP mechanism, then $\mathrm{dist}(P^{x_i}_{prior}, P^{x_i}_{posterior}) = \mathcal{O}(\varepsilon)$

Fundamental Properties

- Robustness to post-processing: $\mathcal{M}$ is (ε, δ)-DP, then $F \circ \mathcal{M}$ is (ε, δ)-DP
- Composition: if $\mathcal{M}_j$, $j = 1, \ldots, k$, are $(\varepsilon_j, \delta_j)$-DP, then $\vec{x} \mapsto (\mathcal{M}_1(\vec{x}), \ldots, \mathcal{M}_k(\vec{x}))$ is $(\sum_j \varepsilon_j, \sum_j \delta_j)$-DP. In the homogeneous case this yields $(k\varepsilon, k\delta)$-DP
- Advanced composition: if $\mathcal{M}_j$, $j = 1, \ldots, k$, are (ε, δ)-DP, then $\vec{x} \mapsto (\mathcal{M}_1(\vec{x}), \ldots, \mathcal{M}_k(\vec{x}))$ is $(\varepsilon\sqrt{k \log(1/\delta')} + \varepsilon(e^\varepsilon - 1)k, k\delta + \delta')$-DP for any $\delta' > 0$
- Group privacy: if $\mathcal{M}$ is (ε, δ)-DP with respect to $x \simeq x'$, then $\mathcal{M}$ is $(t\varepsilon, t\delta)$ with respect to $x \simeq^t x'$ (ie. t changes)
- Protects against side knowledge: if attacker has prior $P^{x_i}_{prior}$ and computes $P^{x_i}_{posterior}$ after observing $\mathcal{M}(\vec{x})$ from ε-DP mechanism, then $\mathrm{dist}(P^{x_i}_{prior}, P^{x_i}_{posterior}) = \mathcal{O}(\varepsilon)$

research

Fundamental Properties

- Robustness to post-processing: $\mathcal{M}$ is (ε, δ)-DP, then $F \circ \mathcal{M}$ is (ε, δ)-DP
- Composition: if $\mathcal{M}_j$, $j = 1, \dots, k$, are $(\varepsilon_j, \delta_j)$-DP, then $\vec{x} \mapsto (\mathcal{M}_1(\vec{x}), \dots, \mathcal{M}_k(\vec{x}))$ is $(\sum_j \varepsilon_j, \sum_j \delta_j)$-DP. In the homogeneous case this yields $(k\varepsilon, k\delta)$-DP
- Advanced composition: if $\mathcal{M}_j$, $j = 1, \dots, k$, are (ε, δ)-DP, then $\vec{x} \mapsto (\mathcal{M}_1(\vec{x}), \dots, \mathcal{M}_k(\vec{x}))$ is $(\varepsilon\sqrt{k \log(1/\delta')} + \varepsilon(e^{\varepsilon} - 1)k, k\delta + \delta')$-DP for any $\delta' > 0$
- Group privacy: if $\mathcal{M}$ is (ε, δ)-DP with respect to $x \simeq x'$, then $\mathcal{M}$ is $(t\varepsilon, t\delta)$ with respect to $x \simeq^t x'$ (ie. t changes)
- Protects against side knowledge: if attacker has prior $P^{x_i}_{prior}$ and computes $P^{x_i}_{posterior}$ after observing $\mathcal{M}(\vec{x})$ from ε-DP mechanism, then $\mathrm{dist}(P^{x_i}_{prior}, P^{x_i}_{posterior}) = \mathcal{O}(\varepsilon)$

The Exponential Mechanism

The Laplace and Gaussian mechanisms are examples of a more general class of mechanisms

Densities of output perturbation mechanisms

$$p_{\mathcal{M}_{f,\mathrm{Lap}}(x)}(y) \propto \exp\left(\frac{-\varepsilon\|y - f(x)\|_1}{\Delta_1}\right) \qquad p_{\mathcal{M}_{f,\mathcal{N}}(x)}(y) \propto \exp\left(\frac{-\varepsilon^2\|y - f(x)\|_2^2}{C\Delta_2^2 \log(1/\delta)}\right)$$

Exponential Mechanism

- Prior distribution over outputs with density π
- Scoring function $q : X \times Y \to \mathbb{R}_{\geqslant 0}$ provides scores for each output y w.r.t. input x
- The exponential mechanism $\mathcal{M}_{\pi,q}(x)$ outputs a sample from the distribution with density

$$p_{\pi,q}(y) \propto \pi(y) \exp\left(-\beta q(x, y)\right)$$

research

The Exponential Mechanism

The Laplace and Gaussian mechanisms are examples of a more general class of mechanisms

Densities of output perturbation mechanisms

$$p_{\mathcal{M}_{f,\mathrm{Lap}}(x)}(y) \propto \exp\left(\frac{-\varepsilon\|y - f(x)\|_1}{\Delta_1}\right) \qquad p_{\mathcal{M}_{f,\mathcal{N}}(x)}(y) \propto \exp\left(\frac{-\varepsilon^2\|y - f(x)\|_2^2}{C\Delta_2^2 \log(1/\delta)}\right)$$

Exponential Mechanism

- Prior distribution over outputs with density π
- Scoring function $q : X \times Y \to \mathbb{R}_{\geqslant 0}$ provides scores for each output y w.r.t. input x
- The exponential mechanism $\mathcal{M}_{\pi,q}(x)$ outputs a sample from the distribution with density

$$p_{\pi,q}(y) \propto \pi(y) \exp\left(-\beta q(x, y)\right)$$

research

The Exponential Mechanism

The Laplace and Gaussian mechanisms are examples of a more general class of mechanisms

Densities of output perturbation mechanisms

$$p_{\mathcal{M}_{f,\mathrm{Lap}}(x)}(y) \propto \exp\left(\frac{-\varepsilon\|y - f(x)\|_1}{\Delta_1}\right) \qquad p_{\mathcal{M}_{f,\mathcal{N}}(x)}(y) \propto \exp\left(\frac{-\varepsilon^2\|y - f(x)\|_2^2}{C\Delta_2^2 \log(1/\delta)}\right)$$

Exponential Mechanism

- Prior distribution over outputs with density π
- Scoring function $q : X \times Y \to \mathbb{R}_{\geqslant 0}$ provides scores for each output y w.r.t. input x
- The exponential mechanism $\mathcal{M}_{\pi,q}(x)$ outputs a sample from the distribution with density

$$p_{\pi,q}(y) \propto \pi(y) \exp\left(-\beta q(x, y)\right)$$

Calibrating The Exponential Mechanism

Properties of the Scoring Function

- Sensitivity: $\sup_{x \simeq x'} \sup_y |q(x,y) - q(x',y)| \leqslant \Delta$
- Lipschitz: $\sup_{x \simeq x'} |(q(x,y) - q(x',y)) - (q(x,y') - q(x',y'))| \leqslant L\|y - y'\|$

Properties of the Prior

- Strong log-concavity: $\pi(y) = e^{-W(y)}$ for some κ-strongly convex W

Privacy Guarantees for the Exponential Mechanism

Assumptions	β	Privacy	Reference
q bounded sensitivity	$\mathcal{O}\left(\frac{\varepsilon}{\Delta}\right)$	$(\varepsilon, 0)$	**[McSherry and Talwar, 2007]**
q Lipschitz + convex π strongly log-concave	$\mathcal{O}\left(\frac{\varepsilon\sqrt{\kappa}}{L\sqrt{\log(1/\delta)}}\right)$	(ε, δ)	**[Minami et al., 2016]**

research

Outline

Differentially Private Empirical Risk Minimization

Setup: A curator has features and labels $\vec{z} = ((x_1, y_1), \ldots, (x_n, y_n))$ about n individuals and wants to train a model by minimizing over $\theta \in \Theta$

$$L(\vec{z}, \theta) = \frac{1}{n} \sum_{i=1}^{n} \ell(x_i, y_i, \theta) + \frac{R(\theta)}{n}$$

Examples: logistic regression, SVM, linear regression, DNN, etc.

Private ERM Algorithms

- Output Perturbation: add some noise Z to $\hat{\theta} = \mathrm{argmin}_{\theta \in \Theta} L(\vec{z}, \theta)$
- Objective Perturbation: reveal the optimum of $L(\vec{z}, \theta) + \langle \theta, Z \rangle$ for some noise Z
- Gradient Perturbation: optimize $L(\vec{z}, \theta)$ using mini-batch SGD with noisy gradients

research

Differentially Private Empirical Risk Minimization

Setup: A curator has features and labels $\vec{z} = ((x_1, y_1), \ldots, (x_n, y_n))$ about n individuals and wants to train a model by minimizing over $\theta \in \Theta$

$$L(\vec{z}, \theta) = \frac{1}{n} \sum_{i=1}^{n} \ell(x_i, y_i, \theta) + \frac{R(\theta)}{n}$$

Examples: logistic regression, SVM, linear regression, DNN, etc.

Private ERM Algorithms

- Output Perturbation: add some noise Z to $\hat{\theta} = \mathrm{argmin}_{\theta \in \Theta} L(\vec{z}, \theta)$
- Objective Perturbation: reveal the optimum of $L(\vec{z}, \theta) + \langle \theta, Z \rangle$ for some noise Z
- Gradient Perturbation: optimize $L(\vec{z}, \theta)$ using mini-batch SGD with noisy gradients

DP-ERM: Method Comparison

Perturb	Optimization	Privacy	Assumptions	Excess Risk	Reference
Objective	Exact	(ε, δ)	linear model convexity	$\tilde{\mathcal{O}}\left(\frac{1}{\varepsilon\sqrt{n}}\right)$	[Jain and Thakurta, 2014]
Output	Exact	(ε, δ)	linear model convexity	$\mathcal{O}\left(\frac{1}{\varepsilon\sqrt{n}}\right)$	[Jain and Thakurta, 2014]
Output	SGD	ε	linear model convexity	$\mathcal{O}\left(\frac{d}{\varepsilon\sqrt{n}}\right)$	[Wu et al., 2016]
Output	SGD	ε	linear model strong convexity	$\mathcal{O}\left(\frac{d}{\varepsilon n}\right)$	[Wu et al., 2016]
Gradient	SGD	(ε, δ)	convexity	$\tilde{\mathcal{O}}\left(\frac{\sqrt{d}}{\varepsilon n}\right)$	[Bassily et al., 2014]
Gradient	SGD	(ε, δ)	strong convexity	$\tilde{\mathcal{O}}\left(\frac{d}{\varepsilon^2 n^2}\right)$	[Bassily et al., 2014]

See also **[Talwar et al., 2014, Abadi et al., 2016]**

research

Private Bayesian Learning

One-Posterior Sample (OPS) Mechanism **[Wang et al., 2015]**

- Curator has a prior $P_{prior}(\theta)$ and a model $P_{model}(x_i|\theta)$
- Given a dataset $\vec{x}$ the curators computes the posterior $P_{posterior}(\theta|\vec{x})$, and
- reveals a sample $\hat{\theta} \sim P_{posterior}(\theta|\vec{x})$

Claim: If the model satisfies $\sup_{x,x',\theta} |\log P_{model}(x|\theta) - \log P_{model}(x'|\theta)| \leqslant \varepsilon/2$ then OPS is ε-DP

See also: **[Wang et al., 2015, Foulds et al., 2016, Minami et al., 2016]** for DP with approximate inference, **[Park et al., 2016]** for DP with variational Bayes, and **[Zhang et al., 2016]** for Bayesian network mechanisms

Private Bayesian Learning

One-Posterior Sample (OPS) Mechanism **[Wang et al., 2015]**

- Curator has a prior $P_{prior}(\theta)$ and a model $P_{model}(x_i|\theta)$
- Given a dataset $\vec{x}$ the curators computes the posterior $P_{posterior}(\theta|\vec{x})$, and
- reveals a sample $\hat{\theta} \sim P_{posterior}(\theta|\vec{x})$

Claim: If the model satisfies $\sup_{x,x',\theta} |\log P_{model}(x|\theta) - \log P_{model}(x'|\theta)| \leqslant \varepsilon/2$ then OPS is ε-DP

See also: **[Wang et al., 2015, Foulds et al., 2016, Minami et al., 2016]** for DP with approximate inference, **[Park et al., 2016]** for DP with variational Bayes, and **[Zhang et al., 2016]** for Bayesian network mechanisms

Private Bayesian Learning

One-Posterior Sample (OPS) Mechanism **[Wang et al., 2015]**

- Curator has a prior $P_{prior}(\theta)$ and a model $P_{model}(x_i|\theta)$
- Given a dataset $\vec{x}$ the curators computes the posterior $P_{posterior}(\theta|\vec{x})$, and
- reveals a sample $\hat{\theta} \sim P_{posterior}(\theta|\vec{x})$

Claim: If the model satisfies $\sup_{x,x',\theta} |\log P_{model}(x|\theta) - \log P_{model}(x'|\theta)| \leqslant \varepsilon/2$ then OPS is ε-DP

See also: **[Wang et al., 2015, Foulds et al., 2016, Minami et al., 2016]** for DP with approximate inference, **[Park et al., 2016]** for DP with variational Bayes, and **[Zhang et al., 2016]** for Bayesian network mechanisms

Outline

Privacy Losses

Let $\mathcal{M} : X \to Y$ be a randomized mechanism with density function $p_{\mathcal{M}(x)}(y)$

Privacy Loss (function)

$$\mathcal{L}_{\mathcal{M},x,x'}(y) = \log\left(\frac{p_{\mathcal{M}(x)}(y)}{p_{\mathcal{M}(x')}(y)}\right)$$

Privacy Loss (random variable)

$$L_{\mathcal{M},x,x'} = \mathcal{L}_{\mathcal{M},x,x'}(\mathcal{M}(x))$$

Lemma (Sufficient Condition)
A mechanism $\mathcal{M} : X \to Y$ is (ε, δ)-DP if for any $x \simeq x'$ we have $\mathbb{P}[L_{\mathcal{M},x,x'} \geqslant \varepsilon] \leqslant \delta$

Privacy Losses

Let $\mathcal{M} : X \to Y$ be a randomized mechanism with density function $p_{\mathcal{M}(x)}(y)$

Privacy Loss (function)

$$\mathcal{L}_{\mathcal{M},x,x'}(y) = \log\left(\frac{p_{\mathcal{M}(x)}(y)}{p_{\mathcal{M}(x')}(y)}\right)$$

Privacy Loss (random variable)

$$L_{\mathcal{M},x,x'} = \mathcal{L}_{\mathcal{M},x,x'}(\mathcal{M}(x))$$

Lemma (Sufficient Condition)
A mechanism $\mathcal{M} : X \to Y$ is (ε, δ)-DP if for any $x \simeq x'$ we have $\mathbb{P}[L_{\mathcal{M},x,x'} \geqslant \varepsilon] \leqslant \delta$

research

Privacy Losses

Let $\mathcal{M} : X \to Y$ be a randomized mechanism with density function $p_{\mathcal{M}(x)}(y)$

Privacy Loss (function)

$$\mathcal{L}_{\mathcal{M},x,x'}(y) = \log\left(\frac{p_{\mathcal{M}(x)}(y)}{p_{\mathcal{M}(x')}(y)}\right)$$

Privacy Loss (random variable)

$$L_{\mathcal{M},x,x'} = \mathcal{L}_{\mathcal{M},x,x'}(\mathcal{M}(x))$$

Lemma (Sufficient Condition)
A mechanism $\mathcal{M} : X \to Y$ is (ε, δ)-DP if for any $x \simeq x'$ we have $\mathbb{P}[L_{\mathcal{M},x,x'} \geqslant \varepsilon] \leqslant \delta$

Privacy Losses

Let $\mathcal{M} : X \to Y$ be a randomized mechanism with density function $p_{\mathcal{M}(x)}(y)$

Privacy Loss (function)

$$\mathcal{L}_{\mathcal{M},x,x'}(y) = \log\left(\frac{p_{\mathcal{M}(x)}(y)}{p_{\mathcal{M}(x')}(y)}\right)$$

Privacy Loss (random variable)

$$L_{\mathcal{M},x,x'} = \mathcal{L}_{\mathcal{M},x,x'}(\mathcal{M}(x))$$

Lemma (Sufficient Condition)

A mechanism $\mathcal{M} : X \to Y$ is (ε, δ)-DP if for any $x \simeq x'$ we have $\mathbb{P}[L_{\mathcal{M},x,x'} \geqslant \varepsilon] \leqslant \delta$

Analysis of the Gaussian Mechanism

1. Setup: $\mathcal{M}(x) = f(x) + Z$ with $Z \sim \mathcal{N}(0, \sigma^2 I)$ with $\sigma = \frac{\Delta_2}{\varepsilon}\sqrt{C \log(1/\delta)}$ (for $\varepsilon \leqslant 1$)
2. Compute the distribution of the privacy loss random variable:

$$\mathcal{L}_{\mathcal{M},x,x'}(y) = \frac{\|y - f(x')\|_2^2 - \|y - f(x)\|_2^2}{2\sigma^2} = \frac{\|f(x) - f(x')\|_2^2}{2\sigma^2} + \frac{\langle y - f(x), f(x) - f(x')\rangle}{\sigma^2}$$

$$L_{\mathcal{M},x,x'} = \frac{\|f(x) - f(x')\|_2^2}{2\sigma^2} + \frac{\langle Z, f(x) - f(x')\rangle}{\sigma^2} \sim \mathcal{N}\left(\frac{\|f(x) - f(x')\|_2^2}{2\sigma^2}, \frac{\|f(x) - f(x')\|_2^2}{\sigma^2}\right)$$

3. Use a concentration bound for Gaussian random variables. With probability $\geqslant 1 - \delta$:

$$\mathcal{N}(\eta, 2\eta) \leqslant \eta + \sqrt{C_0 \eta \log(1/\delta)} \leqslant \varepsilon$$

4. Assuming $\varepsilon \leqslant 1$, a bit of algebra shows $\mathbb{P}[L_{\mathcal{M},x,x'} \geqslant \varepsilon] \leqslant \delta$ if:

$$\eta \leqslant \left(\sqrt{\varepsilon + C_1 \log(1/\delta)} - \sqrt{C_1 \log(1/\delta)}\right)^2 \leqslant \frac{\varepsilon^2}{C_2 \log(1/\delta)}$$

5. Substitute the definition of σ^2 and verify the condition is satisfied:

Analysis of the Gaussian Mechanism

1. Setup: $\mathcal{M}(x) = f(x) + Z$ with $Z \sim \mathcal{N}(0, \sigma^2 I)$ with $\sigma = \frac{\Delta_2}{\varepsilon}\sqrt{C \log(1/\delta)}$ (for $\varepsilon \leqslant 1$)
2. Compute the distribution of the privacy loss random variable:

$$\mathcal{L}_{\mathcal{M},x,x'}(y) = \frac{\|y - f(x')\|_2^2 - \|y - f(x)\|_2^2}{2\sigma^2} = \frac{\|f(x) - f(x')\|_2^2}{2\sigma^2} + \frac{\langle y - f(x), f(x) - f(x') \rangle}{\sigma^2}$$

$$L_{\mathcal{M},x,x'} = \frac{\|f(x) - f(x')\|_2^2}{2\sigma^2} + \frac{\langle Z, f(x) - f(x') \rangle}{\sigma^2} \sim \mathcal{N}\left(\frac{\|f(x) - f(x')\|_2^2}{2\sigma^2}, \frac{\|f(x) - f(x')\|_2^2}{\sigma^2}\right)$$

3. Use a concentration bound for Gaussian random variables. With probability $\geqslant 1 - \delta$:

$$\mathcal{N}(\eta, 2\eta) \leqslant \eta + \sqrt{C_0 \eta \log(1/\delta)} \leqslant \varepsilon$$

4. Assuming $\varepsilon \leqslant 1$, a bit of algebra shows $\mathbb{P}[L_{\mathcal{M},x,x'} \geqslant \varepsilon] \leqslant \delta$ if:

$$\eta \leqslant \left(\sqrt{\varepsilon + C_1 \log(1/\delta)} - \sqrt{C_1 \log(1/\delta)}\right)^2 \leqslant \frac{\varepsilon^2}{C_2 \log(1/\delta)}$$

5. Substitute the definition of σ^2 and verify the condition is satisfied:

Analysis of the Gaussian Mechanism

1. Setup: $\mathcal{M}(x) = f(x) + Z$ with $Z \sim \mathcal{N}(0, \sigma^2 I)$ with $\sigma = \frac{\Delta_2}{\varepsilon}\sqrt{C \log(1/\delta)}$ (for $\varepsilon \leqslant 1$)
2. Compute the distribution of the privacy loss random variable:

$$\mathcal{L}_{\mathcal{M},x,x'}(y) = \frac{\|y - f(x')\|_2^2 - \|y - f(x)\|_2^2}{2\sigma^2} = \frac{\|f(x) - f(x')\|_2^2}{2\sigma^2} + \frac{\langle y - f(x), f(x) - f(x')\rangle}{\sigma^2}$$

$$L_{\mathcal{M},x,x'} = \frac{\|f(x) - f(x')\|_2^2}{2\sigma^2} + \frac{\langle Z, f(x) - f(x')\rangle}{\sigma^2} \sim \mathcal{N}\left(\frac{\|f(x) - f(x')\|_2^2}{2\sigma^2}, \frac{\|f(x) - f(x')\|_2^2}{\sigma^2}\right)$$

3. Use a concentration bound for Gaussian random variables. With probability $\geqslant 1 - \delta$:

$$\mathcal{N}(\eta, 2\eta) \leqslant \eta + \sqrt{C_0 \eta \log(1/\delta)} \leqslant \varepsilon$$

4. Assuming $\varepsilon \leqslant 1$, a bit of algebra shows $\mathbb{P}[L_{\mathcal{M},x,x'} \geqslant \varepsilon] \leqslant \delta$ if:

$$\eta \leqslant \left(\sqrt{\varepsilon + C_1 \log(1/\delta)} - \sqrt{C_1 \log(1/\delta)}\right)^2 \leqslant \frac{\varepsilon^2}{C_2 \log(1/\delta)}$$

5. Substitute the definition of σ^2 and verify the condition is satisfied:

Analysis of the Gaussian Mechanism

1. Setup: $\mathcal{M}(x) = f(x) + Z$ with $Z \sim \mathcal{N}(0, \sigma^2 I)$ with $\sigma = \frac{\Delta_2}{\varepsilon}\sqrt{C \log(1/\delta)}$ (for $\varepsilon \leqslant 1$)
2. Compute the distribution of the privacy loss random variable:

$$L_{\mathcal{M},x,x'} \sim \mathcal{N}\left(\frac{\|f(x) - f(x')\|_2^2}{2\sigma^2}, \frac{\|f(x) - f(x')\|_2^2}{\sigma^2}\right) = \mathcal{N}(\eta, 2\eta)$$

3. Use a concentration bound for Gaussian random variables. With probability $\geqslant 1 - \delta$:

$$\mathcal{N}(\eta, 2\eta) \leqslant \eta + \sqrt{C_0 \eta \log(1/\delta)} \leqslant \varepsilon$$

4. Assuming $\varepsilon \leqslant 1$, a bit of algebra shows $\mathbb{P}[L_{\mathcal{M},x,x'} \geqslant \varepsilon] \leqslant \delta$ if:

$$\eta \leqslant \left(\sqrt{\varepsilon + C_1 \log(1/\delta)} - \sqrt{C_1 \log(1/\delta)}\right)^2 \leqslant \frac{\varepsilon^2}{C_2 \log(1/\delta)}$$

5. Substitute the definition of σ^2 and verify the condition is satisfied:

$$\eta = \frac{\|f(x) - f(x')\|_2^2}{2\sigma^2} = \frac{\varepsilon^2 \|f(x) - f(x')\|_2^2}{2\Delta_2^2 C \log(1/\delta)} \leqslant \frac{\varepsilon^2}{C_2 \log(1/\delta)}$$

research

Analysis of the Gaussian Mechanism

1. Setup: $\mathcal{M}(x) = f(x) + Z$ with $Z \sim \mathcal{N}(0, \sigma^2 I)$ with $\sigma = \frac{\Delta_2}{\varepsilon}\sqrt{C \log(1/\delta)}$ (for $\varepsilon \leqslant 1$)
2. Compute the distribution of the privacy loss random variable:

$$L_{\mathcal{M},x,x'} \sim \mathcal{N}\left(\frac{\|f(x) - f(x')\|_2^2}{2\sigma^2}, \frac{\|f(x) - f(x')\|_2^2}{\sigma^2}\right) = \mathcal{N}(\eta, 2\eta)$$

3. Use a concentration bound for Gaussian random variables. With probability $\geqslant 1 - \delta$:

$$\mathcal{N}(\eta, 2\eta) \leqslant \eta + \sqrt{C_0 \eta \log(1/\delta)} \leqslant \varepsilon$$

4. Assuming $\varepsilon \leqslant 1$, a bit of algebra shows $\mathbb{P}[L_{\mathcal{M},x,x'} \geqslant \varepsilon] \leqslant \delta$ if:

$$\eta \leqslant \left(\sqrt{\varepsilon + C_1 \log(1/\delta)} - \sqrt{C_1 \log(1/\delta)}\right)^2 \leqslant \frac{\varepsilon^2}{C_2 \log(1/\delta)}$$

5. Substitute the definition of σ^2 and verify the condition is satisfied:

$$\eta = \frac{\|f(x) - f(x')\|_2^2}{2\sigma^2} = \frac{\varepsilon^2 \|f(x) - f(x')\|_2^2}{2\Delta_2^2 C \log(1/\delta)} \leqslant \frac{\varepsilon^2}{C_2 \log(1/\delta)}$$

research

Analysis of the Gaussian Mechanism

1. Setup: $\mathcal{M}(x) = f(x) + Z$ with $Z \sim \mathcal{N}(0, \sigma^2 I)$ with $\sigma = \frac{\Delta_2}{\varepsilon}\sqrt{C \log(1/\delta)}$ (for $\varepsilon \leqslant 1$)
2. Compute the distribution of the privacy loss random variable:

$$L_{\mathcal{M},x,x'} \sim \mathcal{N}\left(\frac{\|f(x) - f(x')\|_2^2}{2\sigma^2}, \frac{\|f(x) - f(x')\|_2^2}{\sigma^2}\right) = \mathcal{N}(\eta, 2\eta)$$

3. Use a concentration bound for Gaussian random variables. With probability $\geqslant 1 - \delta$:

$$\mathcal{N}(\eta, 2\eta) \leqslant \eta + \sqrt{C_0 \eta \log(1/\delta)} \leqslant \varepsilon$$

4. Assuming $\varepsilon \leqslant 1$, a bit of algebra shows $\mathbb{P}[L_{\mathcal{M},x,x'} \geqslant \varepsilon] \leqslant \delta$ if:

$$\eta \leqslant \left(\sqrt{\varepsilon + C_1 \log(1/\delta)} - \sqrt{C_1 \log(1/\delta)}\right)^2 \leqslant \frac{\varepsilon^2}{C_2 \log(1/\delta)}$$

5. Substitute the definition of σ^2 and verify the condition is satisfied:

$$\eta = \frac{\|f(x) - f(x')\|_2^2}{2\sigma^2} = \frac{\varepsilon^2 \|f(x) - f(x')\|_2^2}{2\Delta_2^2 C \log(1/\delta)} \leqslant \frac{\varepsilon^2}{C_2 \log(1/\delta)}$$

research

Analysis of the Gaussian Mechanism

1. Setup: $\mathcal{M}(x) = f(x) + Z$ with $Z \sim \mathcal{N}(0, \sigma^2 I)$ with $\sigma = \frac{\Delta_2}{\varepsilon}\sqrt{C \log(1/\delta)}$ (for $\varepsilon \leqslant 1$)
2. Compute the distribution of the privacy loss random variable:

$$L_{\mathcal{M},x,x'} \sim \mathcal{N}\left(\frac{\|f(x) - f(x')\|_2^2}{2\sigma^2}, \frac{\|f(x) - f(x')\|_2^2}{\sigma^2}\right) = \mathcal{N}(\eta, 2\eta)$$

3. Use a concentration bound for Gaussian random variables. With probability $\geqslant 1 - \delta$:

$$\mathcal{N}(\eta, 2\eta) \leqslant \eta + \sqrt{C_0 \eta \log(1/\delta)} \leqslant \varepsilon$$

4. Assuming $\varepsilon \leqslant 1$, a bit of algebra shows $\mathbb{P}[L_{\mathcal{M},x,x'} \geqslant \varepsilon] \leqslant \delta$ if:

$$\eta \leqslant \left(\sqrt{\varepsilon + C_1 \log(1/\delta)} - \sqrt{C_1 \log(1/\delta)}\right)^2 \leqslant \frac{\varepsilon^2}{C_2 \log(1/\delta)}$$

5. Substitute the definition of σ^2 and verify the condition is satisfied:

$$\eta = \frac{\|f(x) - f(x')\|_2^2}{2\sigma^2} = \frac{\varepsilon^2 \|f(x) - f(x')\|_2^2}{2\Delta_2^2 C \log(1/\delta)} \leqslant \frac{\varepsilon^2}{C_2 \log(1/\delta)}$$

Differential Privacy as a Concentration Property

- Let $\mathcal{M} : X \to Y$ be a randomized mechanism with privacy loss r.v. $L_{\mathcal{M},x,x'}$
- Define the cumulant generating function of $\mathcal{M}$ as $\varphi_{\mathcal{M},x,x'}(s) = \log \mathbb{E}[e^{sL_{\mathcal{M},x,x'}}]$

Name	Definition	Reference
Concentrated DP (μ, τ)-CDP	$x \simeq x'$, $s > 0$ $\varphi_{\mathcal{M},x,x'}(s) \leqslant s\mu + \frac{s^2\tau^2}{2}$	[Dwork and Rothblum, 2016]
Zero-Concentrated DP (ξ, ρ)-zCDP	$x \simeq x'$, $s > 0$ $\varphi_{\mathcal{M},x,x'}(s) \leqslant s(\xi + \rho) + s^2\rho$	[Bun and Steinke, 2016]
Rényi DP $(\alpha + 1, \beta)$-RDP	$x \simeq x'$ $\varphi_{\mathcal{M},x,x'}(\alpha) \leqslant \alpha\beta$	[Mironov, 2017]

- Gaussian: For $L \sim \mathcal{N}(\eta, 2\eta)$ the c.g.f. is $\varphi(s) = s\eta + s^2\eta$, i.e. $(0, \eta)$-zCDP
- Markov: If $\exists s > 0$ such that $\sup_{x \simeq x'} \varphi_{\mathcal{M},x,x'}(s) + \log(1/\delta) \leqslant s\varepsilon$, then $\mathcal{M}$ is (ε, δ)-DP
- Moment accountant: Let $\varphi_i(s)$ be c.g.f. for mechanism $\mathcal{M}_i$. The mechanism $\mathcal{M}(x) = (\mathcal{M}_1(x), \ldots, \mathcal{M}_k(x))$ has c.g.f. $\varphi_{\mathcal{M}}(s) = \sum_{i=1}^{k} \varphi_i(s)$ [Abadi et al., 2016]

Differential Privacy as a Concentration Property

- Let $\mathcal{M} : X \to Y$ be a randomized mechanism with privacy loss r.v. $L_{\mathcal{M},x,x'}$
- Define the cumulant generating function of $\mathcal{M}$ as $\varphi_{\mathcal{M},x,x'}(s) = \log \mathbb{E}[e^{sL_{\mathcal{M},x,x'}}]$

Name	Definition	Reference
Concentrated DP (μ, τ)-CDP	$x \simeq x'$, $s > 0$ $\varphi_{\mathcal{M},x,x'}(s) \leqslant s\mu + \frac{s^2\tau^2}{2}$	**[Dwork and Rothblum, 2016]**
Zero-Concentrated DP (ξ, ρ)-zCDP	$x \simeq x'$, $s > 0$ $\varphi_{\mathcal{M},x,x'}(s) \leqslant s(\xi + \rho) + s^2\rho$	**[Bun and Steinke, 2016]**
Rényi DP $(\alpha + 1, \beta)$-RDP	$x \simeq x'$ $\varphi_{\mathcal{M},x,x'}(\alpha) \leqslant \alpha\beta$	**[Mironov, 2017]**

- Gaussian: For $L \sim \mathcal{N}(\eta, 2\eta)$ the c.g.f. is $\varphi(s) = s\eta + s^2\eta$, i.e. $(0, \eta)$-zCDP
- Markov: If $\exists s > 0$ such that $\sup_{x \simeq x'} \varphi_{\mathcal{M},x,x'}(s) + \log(1/\delta) \leqslant s\varepsilon$, then $\mathcal{M}$ is (ε, δ)-DP
- Moment accountant: Let $\varphi_i(s)$ be c.g.f. for mechanism $\mathcal{M}_i$. The mechanism $\mathcal{M}(x) = (\mathcal{M}_1(x), \ldots, \mathcal{M}_k(x))$ has c.g.f. $\varphi_{\mathcal{M}}(s) = \sum_{i=1}^{k} \varphi_i(s)$ **[Abadi et al., 2016]**

Differential Privacy as a Concentration Property

- Let $\mathcal{M} : X \to Y$ be a randomized mechanism with privacy loss r.v. $L_{\mathcal{M},x,x'}$
- Define the cumulant generating function of $\mathcal{M}$ as $\varphi_{\mathcal{M},x,x'}(s) = \log \mathbb{E}[e^{sL_{\mathcal{M},x,x'}}]$

Name	Definition	Reference
Concentrated DP (μ, τ)-CDP	$x \simeq x'$, $s > 0$ $\varphi_{\mathcal{M},x,x'}(s) \leqslant s\mu + \frac{s^2\tau^2}{2}$	**[Dwork and Rothblum, 2016]**
Zero-Concentrated DP (ξ, ρ)-zCDP	$x \simeq x'$, $s > 0$ $\varphi_{\mathcal{M},x,x'}(s) \leqslant s(\xi + \rho) + s^2\rho$	**[Bun and Steinke, 2016]**
Rényi DP $(\alpha + 1, \beta)$-RDP	$x \simeq x'$ $\varphi_{\mathcal{M},x,x'}(\alpha) \leqslant \alpha\beta$	**[Mironov, 2017]**

- Gaussian: For $L \sim \mathcal{N}(\eta, 2\eta)$ the c.g.f. is $\varphi(s) = s\eta + s^2\eta$, i.e. $(0, \eta)$-zCDP
- Markov: If $\exists s > 0$ such that $\sup_{x \simeq x'} \varphi_{\mathcal{M},x,x'}(s) + \log(1/\delta) \leqslant s\varepsilon$, then $\mathcal{M}$ is (ε, δ)-DP
- Moment accountant: Let $\varphi_i(s)$ be c.g.f. for mechanism $\mathcal{M}_i$. The mechanism $\mathcal{M}(x) = (\mathcal{M}_1(x), \ldots, \mathcal{M}_k(x))$ has c.g.f. $\varphi_{\mathcal{M}}(s) = \sum_{i=1}^{k} \varphi_i(s)$ **[Abadi et al., 2016]**

research

Differential Privacy as a Concentration Property

- Let $\mathcal{M} : X \to Y$ be a randomized mechanism with privacy loss r.v. $L_{\mathcal{M},x,x'}$
- Define the cumulant generating function of $\mathcal{M}$ as $\varphi_{\mathcal{M},x,x'}(s) = \log \mathbb{E}[e^{sL_{\mathcal{M},x,x'}}]$

Name	Definition	Reference
Concentrated DP (μ, τ)-CDP	$x \simeq x'$, $s > 0$ $\varphi_{\mathcal{M},x,x'}(s) \leqslant s\mu + \frac{s^2\tau^2}{2}$	[Dwork and Rothblum, 2016]
Zero-Concentrated DP (ξ, ρ)-zCDP	$x \simeq x'$, $s > 0$ $\varphi_{\mathcal{M},x,x'}(s) \leqslant s(\xi + \rho) + s^2\rho$	[Bun and Steinke, 2016]
Rényi DP $(\alpha + 1, \beta)$-RDP	$x \simeq x'$ $\varphi_{\mathcal{M},x,x'}(\alpha) \leqslant \alpha\beta$	[Mironov, 2017]

- Gaussian: For $L \sim \mathcal{N}(\eta, 2\eta)$ the c.g.f. is $\varphi(s) = s\eta + s^2\eta$, i.e. $(0, \eta)$-zCDP
- Markov: If $\exists s > 0$ such that $\sup_{x \simeq x'} \varphi_{\mathcal{M},x,x'}(s) + \log(1/\delta) \leqslant s\varepsilon$, then $\mathcal{M}$ is (ε, δ)-DP
- Moment accountant: Let $\varphi_i(s)$ be c.g.f. for mechanism $\mathcal{M}_i$. The mechanism $\mathcal{M}(x) = (\mathcal{M}_1(x), \ldots, \mathcal{M}_k(x))$ has c.g.f. $\varphi_{\mathcal{M}}(s) = \sum_{i=1}^{k} \varphi_i(s)$ [Abadi et al., 2016]

research

Differential Privacy as a Concentration Property

- Let $\mathcal{M} : X \to Y$ be a randomized mechanism with privacy loss r.v. $L_{\mathcal{M},x,x'}$
- Define the cumulant generating function of $\mathcal{M}$ as $\varphi_{\mathcal{M},x,x'}(s) = \log \mathbb{E}[e^{sL_{\mathcal{M},x,x'}}]$

Name	Definition	Reference
Concentrated DP (μ, τ)-CDP	$x \simeq x'$, $s > 0$ $\varphi_{\mathcal{M},x,x'}(s) \leqslant s\mu + \frac{s^2\tau^2}{2}$	**[Dwork and Rothblum, 2016]**
Zero-Concentrated DP (ξ, ρ)-zCDP	$x \simeq x'$, $s > 0$ $\varphi_{\mathcal{M},x,x'}(s) \leqslant s(\xi + \rho) + s^2\rho$	**[Bun and Steinke, 2016]**
Rényi DP $(\alpha + 1, \beta)$-RDP	$x \simeq x'$ $\varphi_{\mathcal{M},x,x'}(\alpha) \leqslant \alpha\beta$	**[Mironov, 2017]**

- Gaussian: For $L \sim \mathcal{N}(\eta, 2\eta)$ the c.g.f. is $\varphi(s) = s\eta + s^2\eta$, i.e. $(0, \eta)$-zCDP
- Markov: If $\exists s > 0$ such that $\sup_{x \simeq x'} \varphi_{\mathcal{M},x,x'}(s) + \log(1/\delta) \leqslant s\varepsilon$, then $\mathcal{M}$ is (ε, δ)-DP
- Moment accountant: Let $\varphi_i(s)$ be c.g.f. for mechanism $\mathcal{M}_i$. The mechanism $\mathcal{M}(x) = (\mathcal{M}_1(x), \ldots, \mathcal{M}_k(x))$ has c.g.f. $\varphi_{\mathcal{M}}(s) = \sum_{i=1}^{k} \varphi_i(s)$ **[Abadi et al., 2016]**

research

Differential Privacy Without a Trusted Curator

Issues with the Trusted Curator Assumption

- *Single point of failure:* a DP curator might have other security vulnerabilities
- *Conflicting incentives:* valuable the data provides incentives for the curator to misbehave
- *Requires agreement:* a large number of individuals need to agree on who to trust

Randomized response: recall in $(y_1, \ldots, y_n) = RR_\varepsilon(x_1, \ldots, x_n)$ each y_i depends only on x_i

Multi-Party and Local Differential Privacy

- Dataset x distributed among m parties, party i owns $\vec{x}_i$
- Analyst initiates randomized protocol $\Pi : X \to Y$ that interacts with the parties
- All the outputs produced by party i during $\Pi(x)$ determine a mechanism $\mathcal{M}_i(\vec{x}_i)$
- Π is *multi-party* (ε, δ)*-DP* if each $\mathcal{M}_i$ is (ε, δ)-DP
- When each $\vec{x}_i$ has size one we talk about *local DP*
- Utility loss: the difference between $\mathcal{O}(1/n)$ (Laplace) and $\mathcal{O}(1/\sqrt{n})$ (RR) is characteristic of local DP

Differential Privacy Without a Trusted Curator

Issues with the Trusted Curator Assumption

- *Single point of failure:* a DP curator might have other security vulnerabilities
- *Conflicting incentives:* valuable the data provides incentives for the curator to misbehave
- *Requires agreement:* a large number of individuals need to agree on who to trust

Randomized response: recall in $(y_1, \ldots, y_n) = RR_\varepsilon(x_1, \ldots, x_n)$ each y_i depends only on x_i

Multi-Party and Local Differential Privacy

- Dataset x distributed among m parties, party i owns $\vec{x}_i$
- Analyst initiates randomized protocol $\Pi : X \to Y$ that interacts with the parties
- All the outputs produced by party i during $\Pi(x)$ determine a mechanism $\mathcal{M}_i(\vec{x}_i)$
- Π is *multi-party* (ε, δ)*-DP* if each $\mathcal{M}_i$ is (ε, δ)-DP
- When each $\vec{x}_i$ has size one we talk about *local DP*
- Utility loss: the difference between $\mathcal{O}(1/n)$ (Laplace) and $\mathcal{O}(1/\sqrt{n})$ (RR) is characteristic of local DP

Differential Privacy Without a Trusted Curator

Issues with the Trusted Curator Assumption

- *Single point of failure:* a DP curator might have other security vulnerabilities
- *Conflicting incentives:* valuable the data provides incentives for the curator to misbehave
- *Requires agreement:* a large number of individuals need to agree on who to trust

Randomized response: recall in $(y_1, \ldots, y_n) = RR_\varepsilon(x_1, \ldots, x_n)$ each y_i depends only on x_i

Multi-Party and Local Differential Privacy

- Dataset x distributed among m parties, party i owns $\vec{x}_i$
- Analyst initiates randomized protocol $\Pi : X \to Y$ that interacts with the parties
- All the outputs produced by party i during $\Pi(x)$ determine a mechanism $\mathcal{M}_i(\vec{x}_i)$
- Π is *multi-party* (ε, δ)*-DP* if each $\mathcal{M}_i$ is (ε, δ)-DP
- When each $\vec{x}_i$ has size one we talk about *local DP*
- Utility loss: the difference between $\mathcal{O}(1/n)$ (Laplace) and $\mathcal{O}(1/\sqrt{n})$ (RR) is characteristic of local DP

Outline

1. We Need Mathematics to Study Privacy? Seriously?
2. Differential Privacy: Definition, Properties and Basic Mechanisms
3. Differentially Private Machine Learning: ERM and Bayesian Learning
4. Variations on Differential Privacy: Concentrated DP and Local DP
5. Final Remarks

Beyond This Tutorial...

Additional Results

- More basic mechanisms: sparse vector technique and other selection mechanisms, private data structures
- General theorems: everything is randomized response, lower bounds on utility, computational hardness, optimal mechanisms, connections to generalization
- Database perspective: answering multiple queries on the same data, adaptive vs. non-adaptive queries
- When global sensitivity is atypical: smoothed sensitivity, randomized DP
- Other privacy definitions: location privacy, pan DP, pufferfish privacy

Suggested Readings

- "The Algorithmic Foundations of Differential Privacy" **[Dwork and Roth, 2014]**
- "The Complexity of Differential Privacy" **[Vadhan, 2017]**

Some Open Research Directions

Bounds vs. Algorithms

- Few privacy analysis are tight: randomized response, Laplace mechanism, ε-DP exponential mechanism
- Most complex mechanisms add too much noise (constants in bounds matter!)
- Alternative: calibrate noise using "exact" numerical computations instead of bounds
- Challenges: concentration bounds vs. exact densities, compositions, sub-sampling and other mixtures, approximate sampling

Correctness and Attacks

- Given a mechanism, it is not possible to test empirically if it is DP
- We can only resort to mathematical proofs to establish correctness (can be automated?)
- But we should have sanity-check to tools to break DP of candidate implementations
- Challenge: from pseudo-code to implementation things can go wrong (floating-point $\neq \mathbb{R}$)

Conclusion

- Differential privacy provides a formal notion of privacy satisfying many desirable properties
  - Precise quantification of the privacy-utility trade-off
  - Robustness against powerful adversaries (eg. in the presence of side knowledge)
  - Applicable to a wide range of data analysis problems
- Mature research field with a rich toolbox of mechanism design strategies
- Natural starting point for application-specific privacy guarantees
- Several real-world deployments and open source tools
  - Google Chrome's RAPPOR
  - Apple's iOS 10
  - U.S. Census Bureau
  - GUPT, Microsoft's PINQ, Uber's FLEX

research

References I

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. (2016).
Deep learning with differential privacy.
In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 308–318. ACM.

Bassily, R., Smith, A. D., and Thakurta, A. (2014).
Private empirical risk minimization: Efficient algorithms and tight error bounds.
In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 464–473.

Bun, M. and Steinke, T. (2016).
Concentrated differential privacy: Simplifications, extensions, and lower bounds.
In *Theory of Cryptography Conference*, pages 635–658. Springer.

Dwork, C. (2006).
Differential privacy.
In *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II*, pages 1–12.

References II

Dwork, C., McSherry, F., Nissim, K., and Smith, A. D. (2006).
Calibrating noise to sensitivity in private data analysis.
In *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings*, pages 265–284.

Dwork, C. and Roth, A. (2014).
The algorithmic foundations of differential privacy.
Foundations and Trends® in Theoretical Computer Science, 9(3–4):211–407.

Dwork, C. and Rothblum, G. N. (2016).
Concentrated differential privacy.
arXiv preprint arXiv:1603.01887.

Foulds, J. R., Geumlek, J., Welling, M., and Chaudhuri, K. (2016).
On the theory and practice of privacy-preserving bayesian data analysis.
In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence, UAI 2016, June 25-29, 2016, New York City, NY, USA*.

References III

Jain, P. and Thakurta, A. G. (2014).
(near) dimension independent risk bounds for differentially private learning.
In *International Conference on Machine Learning*, pages 476–484.

McSherry, F. and Talwar, K. (2007).
Mechanism design via differential privacy.
In *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*, pages 94–103. IEEE.

Minami, K., Arai, H., Sato, I., and Nakagawa, H. (2016).
Differential privacy without sensitivity.
In *Advances in Neural Information Processing Systems*, pages 956–964.

Mironov, I. (2017).
Renyi differential privacy.
arXiv preprint arXiv:1702.07476.

research

References IV

Park, M., Foulds, J. R., Chaudhuri, K., and Welling, M. (2016).
Variational bayes in private settings (VIPS).
CoRR, abs/1611.00340.

Talwar, K., Thakurta, A., and Zhang, L. (2014).
Private empirical risk minimization beyond the worst case: The effect of the constraint set geometry.
CoRR, abs/1411.5417.

Vadhan, S. P. (2017).
The complexity of differential privacy.
In *Tutorials on the Foundations of Cryptography.*, pages 347–450.

Wang, Y., Fienberg, S. E., and Smola, A. J. (2015).
Privacy for free: Posterior sampling and stochastic gradient monte carlo.
In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, pages 2493–2502.

References V

Warner, S. L. (1965).
Randomized response: A survey technique for eliminating evasive answer bias.
Journal of the American Statistical Association, 60(309):63–69.

Wu, X., Kumar, A., Chaudhuri, K., Jha, S., and Naughton, J. F. (2016).
Differentially private stochastic gradient descent for in-rdbms analytics.
CoRR, abs/1606.04722.

Zhang, Z., Rubinstein, B. I. P., and Dimitrakakis, C. (2016).
On the differential privacy of bayesian inference.
In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA.*, pages 2365–2371.

A Short Tutorial on Differential Privacy

Borja Balle

Amazon Research Cambridge

The Alan Turing Institute — January 26, 2018