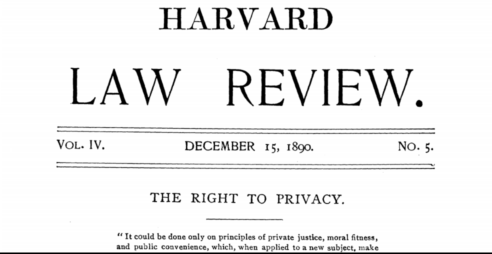
Privacy, Data Privacy and Differential Privacy

James Bailie, Ruobin Gong, Xiao-Li Meng and Jörg Drechsler

Tübingen

July 11, 2024


HARVARD

LAW REVIEW.

VOL. IV. DECEMBER 15, 1890. NO. 5.

THE RIGHT TO PRIVACY.

"It could be done only on principles of private justice, moral fitness, and public convenience, which, when applied to a new subject, make

The right to be ***let alone*****.**

Samuel D. Warren II

Louis Brandeis

Privacy – Can you define it?

- Law: **Privacy is the right to be let alone.**
  Warren & Brandeis (1890). The Right to Privacy. *Harvard Law Review.*

Privacy – Can you define it?

- Law: **Privacy is the right to be let alone.**
  Warren & Brandeis (1890). The Right to Privacy. *Harvard Law Review.*
- Econmics: **Privacy is the price of divulging information.**
  Acquisti et al. (2016) The Economics of Privacy.
  Journal of Economic Literature.

Privacy – Can you define it?

- Law: **Privacy is the right to be let alone.**
  Warren & Brandeis (1890). The Right to Privacy. *Harvard Law Review.*
- Econmics: **Privacy is the price of divulging information.**
  Acquisti et al. (2016) The Economics of Privacy.
  Journal of Economic Literature.
- Political Science: **The boundaries of power over the individual ascribe the rights of the individual to privacy.**
  Raab (2019). Political Science and Privacy. In *The Handbook of Privacy Studies: An Interdisciplinary Introduction.* Amsterdam University Press.

Privacy – Can you define it?

- Law: **Privacy is the right to be let alone.**
  Warren & Brandeis (1890). The Right to Privacy. *Harvard Law Review.*
- Econmics: **Privacy is the price of divulging information.**
  Acquisti et al. (2016) The Economics of Privacy.
  Journal of Economic Literature.
- Political Science: **The boundaries of power over the individual ascribe the rights of the individual to privacy.**
  Raab (2019). Political Science and Privacy. In *The Handbook of Privacy Studies: An Interdisciplinary Introduction.* Amsterdam University Press.
- Philosophy: "**Privacy . . . is a concept in disarray.** ... Currently privacy is a sweeping concept. . . . Philosophers . . . have frequently lamented the great difficulty in reaching a satisfying conception of privacy."
  Solove (2008) *Understanding Privacy.* Harvard University Press.

Data Privacy — What does that mean?

Data Privacy — What does that mean?

Data Content Privacy

Protect information that can be revealed by the recorded data values.

(Commonly known as “statistical disclosure control” or “disclosure avoidance”.)

Data Privacy — What does that mean?

Data Content Privacy

Protect information that can be revealed by the recorded data values.

(Commonly known as “statistical disclosure control” or “disclosure avoidance”.)

Metadata and Paradata Privacy

Protect the identities of the sender and the receiver, time of communication, etc.

Data Privacy — What does that mean?

Data Content Privacy

Protect information that can be revealed by the recorded data values.
(Commonly known as "statistical disclosure control" or "disclosure avoidance".)

Metadata and Paradata Privacy

Protect the identities of the sender and the receiver, time of communication, etc.

Right To Be Forgotten

Right to have personal data erased.

- But how do we operationalize *erasure*? Do we erasure all copies? All consequences?

Data Privacy — What does that mean?

Data Content Privacy

Protect information that can be revealed by the recorded data values.
(Commonly known as “statistical disclosure control” or “disclosure avoidance”.)

Metadata and Paradata Privacy

Protect the identities of the sender and the receiver, time of communication, etc.

Right To Be Forgotten

Right to have personal data erased.

- But how do we operationalize *erasure*? Do we erasure all copies? All consequences?

What is disclosure...

- Releasing statistics while maintaining privacy

$$\textit{A population} \xrightarrow{\textit{Data collection}} \textit{Dataset } \boldsymbol{X} \xrightarrow{\textit{Data release}} \textit{Statistics } T(\boldsymbol{X}, U)$$

What is disclosure...

- Releasing statistics while maintaining privacy

$$\textit{A population} \xrightarrow{\textit{Data collection}} \textit{Dataset } \boldsymbol{X} \overset{\textit{Data release}}{\rightsquigarrow} \textit{Statistics } T(\boldsymbol{X}, U)$$

- A long history

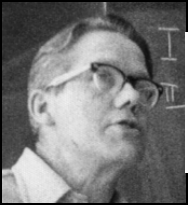
What is disclosure...

- Releasing statistics while maintaining privacy

$$\textit{A population} \xrightarrow{\textit{Data collection}} \textit{Dataset } \boldsymbol{X} \overset{\textit{Data release}}{\rightsquigarrow} \textit{Statistics } T(\boldsymbol{X}, U)$$

- A long history
  - Dalenius (1977), Duncan & Lambert (1986):

    If the release of the statistics T makes it possible to determine [a record X_i] more accurately than is possible without access to T, a disclosure has taken place.



Towards a methodology for statistical disclosure control

by Tore Dalenius[1]

What is disclosure... for a Bayesian?

> *If the release of the statistics T makes it possible to determine [a record X_i] more accurately than is possible without access to T, a disclosure has taken place.*

As Bayesians, can we formalise this?

What is disclosure... for a Bayesian?

> *If the release of the statistics T makes it possible to determine [a record X_i] more accurately than is possible without access to T, a disclosure has taken place.*

As Bayesians, can we formalise this?

- The attacker has a prior π on the record X_i.

What is disclosure... for a Bayesian?

> *If the release of the statistics T makes it possible to determine [a record X_i] more accurately than is possible without access to T, a disclosure has taken place.*

As Bayesians, can we formalise this?

- The attacker has a prior π on the record X_i.
- Without access to the statistics: $\pi(X_i)$.

What is disclosure... for a Bayesian?

> *If the release of the statistics T makes it possible to determine [a record X_i] more accurately than is possible without access to T, a disclosure has taken place.*

As Bayesians, can we formalise this?

- The attacker has a prior π on the record X_i.
- Without access to the statistics: $\pi(X_i)$.
- With the release of the statistics: $\pi(X_i \mid T)$.

What is disclosure... for a Bayesian?

> *If the release of the statistics T makes it possible to determine [a record X_i] more accurately than is possible without access to T, a disclosure has taken place.*

As Bayesians, can we formalise this?

- The attacker has a prior π on the record X_i.
- Without access to the statistics: $\pi(X_i)$.
- With the release of the statistics: $\pi(X_i \mid T)$.
- There is a disclosure if $\pi(X_i)$ and $\pi(X_i \mid T)$ differ.

What is disclosure... for a Bayesian?

> *If the release of the statistics T makes it possible to determine [a record X_i] more accurately than is possible without access to T, a disclosure has taken place.*

As Bayesians, can we formalise this?

- There is a disclosure if $\pi(X_i)$ and $\pi(X_i \mid T)$ differ.

What is disclosure... for a Bayesian?

> *If the release of the statistics T makes it possible to determine [a record X_i] more accurately than is possible without access to T, a disclosure has taken place.*

As Bayesians, can we formalise this?

- There is a disclosure if $\pi(X_i)$ and $\pi(X_i \mid T)$ differ.
- Dalenius (1977) recognised the impossibility of complete protection immediately:

  > *It may be argued that elimination of disclosure is possible only by elimination of statistics.*

What is disclosure... for a Bayesian?

> *If the release of the statistics T makes it possible to determine [a record X_i] more accurately than is possible without access to T, a disclosure has taken place.*

As Bayesians, can we formalise this?

- There is a disclosure if $\pi(X_i)$ and $\pi(X_i \mid T)$ differ.
- Dalenius (1977) recognised the impossibility of complete protection immediately:

  > *It may be argued that elimination of disclosure is possible only by elimination of statistics.*

- To produce useful statistics, we must allow for some (ideally small) amount of disclosure.

What is disclosure... for a Bayesian?

> *If the release of the statistics T makes it possible to determine [a record X_i] more accurately than is possible without access to T, a disclosure has taken place.*

As Bayesians, can we formalise this?

- There is a disclosure if $\pi(X_i)$ and $\pi(X_i \mid T)$ differ.
- Dalenius (1977) recognised the impossibility of complete protection immediately:

  > *It may be argued that elimination of disclosure is possible only by elimination of statistics.*

- To produce useful statistics, we must allow for some (ideally small) amount of disclosure.
- Measure "amount of disclosure" by how much $\pi(X_i)$ and $\pi(X_i \mid T)$ differ.

The *derivative* of differential privacy (DP)

The *derivative* of differential privacy (DP)

Thinking about T as a function of the dataset $\boldsymbol{x}$, its derivative is

$$\lim_{\boldsymbol{x}' \to \boldsymbol{x}} \frac{T(\boldsymbol{x}', U) - T(\boldsymbol{x}, U)}{\boldsymbol{x} - \boldsymbol{x}'}.$$

The *derivative* of differential privacy (DP)

Thinking about the distribution $\mathrm{P}_{\boldsymbol{x}}$ of T as a function of $\boldsymbol{x}$, its derivative is

$$\lim_{\boldsymbol{x}' \to \boldsymbol{x}} \frac{\mathrm{P}_{\boldsymbol{x}'}(T) - \mathrm{P}_{\boldsymbol{x}}(T)}{\boldsymbol{x}' - \boldsymbol{x}}.$$

The *derivative* of differential privacy (DP)

Thinking about the distribution $\mathrm{P}_{\boldsymbol{x}}$ of T as a function of $\boldsymbol{x}$, its derivative is

$$\lim_{\boldsymbol{x}' \to \boldsymbol{x}} \frac{d_{\mathrm{Pr}}(\mathrm{P}_{\boldsymbol{x}'}, \mathrm{P}_{\boldsymbol{x}})}{\boldsymbol{x}' - \boldsymbol{x}}.$$

The *derivative* of differential privacy (DP)

Thinking about the distribution $\mathrm{P}_{\boldsymbol{x}}$ of T as a function of $\boldsymbol{x}$, its derivative is

$$\lim_{\boldsymbol{x}' \to \boldsymbol{x}} \frac{d_{\mathrm{Pr}}(\mathrm{P}_{\boldsymbol{x}'}, \mathrm{P}_{\boldsymbol{x}})}{d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x})}.$$

The *derivative* of differential privacy (DP)

Thinking about the distribution $\mathrm{P}_{\boldsymbol{x}}$ of T as a function of $\boldsymbol{x}$, its derivative is

$$\cancel{\lim_{\boldsymbol{x}' \to \boldsymbol{x}}} \frac{d_{\mathrm{Pr}}(\mathrm{P}_{\boldsymbol{x}'}, \mathrm{P}_{\boldsymbol{x}})}{d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x})},$$

for all $\boldsymbol{x}, \boldsymbol{x}'$.

The *derivative* of differential privacy (DP)

Thinking about the distribution $\mathrm{P}_{\boldsymbol{x}}$ of T as a function of $\boldsymbol{x}$, its ~~derivative~~ Lipschitz constant is the smallest ε such that

$$d_{\mathrm{Pr}}(\mathrm{P}_{\boldsymbol{x}'}, \mathrm{P}_{\boldsymbol{x}}) \leq \varepsilon d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}),$$

for all $\boldsymbol{x}, \boldsymbol{x}'$.

The *derivative* of differential privacy (DP)

Thinking about the distribution $\mathrm{P}_{\boldsymbol{x}}$ of T as a function of $\boldsymbol{x}$, its ~~derivative~~ Lipschitz constant is the smallest ε such that

$$d_{\mathrm{Pr}}(\mathrm{P}_{\boldsymbol{x}'}, \mathrm{P}_{\boldsymbol{x}}) \leq \varepsilon d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}),$$

for all $\boldsymbol{x}, \boldsymbol{x}'$.

Definition: The statistic T is ε-differentially private if its Lipschitz constant is ε.

The *derivative* of differential privacy (DP)

Thinking about the distribution $\mathrm{P}_{\boldsymbol{x}}$ of T as a function of $\boldsymbol{x}$, its ~~derivative~~ Lipschitz constant is the smallest ε such that

$$d_{\mathrm{Pr}}(\mathrm{P}_{\boldsymbol{x}'}, \mathrm{P}_{\boldsymbol{x}}) \leq \varepsilon d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}),$$

for all $\boldsymbol{x}, \boldsymbol{x}'$.

Definition: The statistic T is ε-differentially private if its Lipschitz constant is ε.

- Recall that Lipschitz continuity $\approx$ differentiability.

The *derivative* of differential privacy (DP)

Thinking about the distribution $\mathrm{P}_{\boldsymbol{x}}$ of T as a function of $\boldsymbol{x}$, its ~~derivative~~ Lipschitz constant is the smallest ε such that

$$d_{\mathrm{Pr}}(\mathrm{P}_{\boldsymbol{x}'}, \mathrm{P}_{\boldsymbol{x}}) \leq \varepsilon d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}),$$

for all $\boldsymbol{x}, \boldsymbol{x}'$.

Definition: The statistic T is ε-differentially private if its Lipschitz constant is ε.

- Recall that Lipschitz continuity $\approx$ differentiability.
- Lipschitz constant is the supremum of the derivative.

The *derivative* of differential privacy (DP)

Thinking about the distribution $\mathrm{P}_{\boldsymbol{x}}$ of T as a function of $\boldsymbol{x}$, its ~~derivative~~ Lipschitz constant is the smallest ε such that

$$d_{\mathrm{Pr}}(\mathrm{P}_{\boldsymbol{x}'}, \mathrm{P}_{\boldsymbol{x}}) \leq \varepsilon d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}),$$

for all $\boldsymbol{x}, \boldsymbol{x}'$.

Definition: The statistic T is ε-differentially private if its Lipschitz constant is ε.

- Recall that Lipschitz continuity $\approx$ differentiability.
- Lipschitz constant is the supremum of the derivative.

Takeaway: Differential privacy is a “bound on the derivative” of T.

The *derivative* of differential privacy (DP)

Thinking about the distribution $\mathrm{P}_{\boldsymbol{x}}$ of T as a function of $\boldsymbol{x}$, its ~~derivative~~ Lipschitz constant is the smallest ε such that

$$d_{\mathrm{Pr}}(\mathrm{P}_{\boldsymbol{x}'}, \mathrm{P}_{\boldsymbol{x}}) \leq \varepsilon d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}),$$

for all $\boldsymbol{x}, \boldsymbol{x}'$.

Definition: The statistic T is ε-differentially private if its Lipschitz constant is ε.

- Recall that Lipschitz continuity $\approx$ differentiability.
- Lipschitz constant is the supremum of the derivative.

Takeaway: Differential privacy is a "bound on the derivative" of T.

- The choice of d_{Pr} and $d_{\mathcal{X}}$ determine the *flavour* of DP.

Some examples in the literature

The choice of d_{Pr} and $d_{\mathcal{X}}$ determine the *flavour* of DP:

Some examples in the literature

The choice of d_{Pr} and $d_{\mathcal{X}}$ determine the *flavour* of DP:

$\underline{d_{\mathrm{Pr}}}$: (ε, δ)-approximate DP (Dwork, Kenthapadi, et al., 2006) Rényi DP (Mironov, 2017) concentrated DP (Bun & Steinke, 2016) f-divergence privacy (Barber & Duchi, 2014; Barthe & Olmedo, 2013) f-DP (including Gaussian DP) (Dong et al., 2022)

$\underline{d_{\mathcal{X}}}$: $(\mathcal{R}, \varepsilon)$-generic DP (Kifer & Machanavajjhala, 2011a) edge vs node privacy (Hay et al., 2009; McSherry & Mahajan, 2010) d-metric DP (Chatzikokolakis et al., 2013) Blowfish privacy (He et al., 2014) element level DP (Asi et al., 2022) distributional privacy (Zhou et al., 2009) event-level vs user-level DP (Dwork et al., 2010)

$\underline{\mathscr{D}}$: privacy under invariants (Ashmead et al., 2019; Gong & Meng, 2020; Gao et al., 2022; Dharangutte et al., 2023) conditioned or empirical DP (J. M. Abowd et al., 2013; Charest & Hou, 2016) personalized DP (Ebadi et al., 2015; Jorgensen et al., 2015) individual DP (Soria-Comas et al., 2017; Feldman & Zrnic, 2022) bootstrap DP (O'Keefe & Charest, 2019) stratified DP (Bun et al., 2022) per-record DP (Seeman et al., 2023+) per-instance DP (Wang, 2018; Redberg & Wang, 2021)

$\underline{\mathcal{X}}$: DP for network data (Hay et al., 2009) for geospatial data (Andrés et al., 2013) Pufferfish DP (Kifer & Machanavajjhala, 2014) noiseless privacy (Bhaskar et al., 2011) privacy under partial knowledge (Seeman et al., 2022) privacy amplification (Beimel et al., 2010; Balle et al., 2020; Bun et al., 2022)

The *derivative* of differential privacy (DP)

Thinking about the distribution $\mathsf{P}_{\boldsymbol{x}}$ of T as a function of $\boldsymbol{x}$, its ~~derivative~~ Lipschitz constant is the smallest ε such that

$$d_{\mathsf{Pr}}(\mathsf{P}_{\boldsymbol{x}'}, \mathsf{P}_{\boldsymbol{x}}) \leq \varepsilon d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}).$$

Definition: The statistic T is ε-differentially private if its Lipschitz constant is ε.

- Recall that Lipschitz continuity $\approx$ differentiability.
- Lipschitz constant is the supremum of the derivative.

Takeaway: Differential privacy is a "bound on the derivative" of T.

- The choice of d_{Pr} and $d_{\mathcal{X}}$ determine the *flavour* of DP.

The *derivative* of differential privacy (DP)

Thinking about the distribution $\mathrm{P}_{\boldsymbol{x}}$ of T as a function of $\boldsymbol{x}$, its ~~derivative~~ Lipschitz constant is the smallest ε such that

$$d_{\mathrm{Pr}}(\mathrm{P}_{\boldsymbol{x}'}, \mathrm{P}_{\boldsymbol{x}}) \leq \varepsilon d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}).$$

Definition: The statistic T is ε-differentially private if its Lipschitz constant is ε.

- Recall that Lipschitz continuity $\approx$ differentiability.
- Lipschitz constant is the supremum of the derivative.

Takeaway: Differential privacy is a "bound on the derivative" of T.

- The choice of d_{Pr} and $d_{\mathcal{X}}$ determine the *flavour* of DP.

The classic choice: pure ε-DP (Dwork, McSherry, et al., 2006)

- d_{Pr} is the *max. log-likelihood ratio* $d_{\mathrm{MULT}}(\mathrm{P}_{\boldsymbol{x}}, \mathrm{P}_{\boldsymbol{x}'}) = \sup_t \left| \log \frac{\mathrm{p}_{\mathbf{x}}(T=t)}{\mathrm{p}_{\mathbf{x}'}(T=t)} \right|$
- $d_{\mathcal{X}}$ is the *Hamming distance*

Ex: randomized response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X_i = 1$: cheated; $X_i = 0$, not cheated.

Ex: randomized response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X_i = 1$: cheated; $X_i = 0$, not cheated.
- Each student i tosses a biased coin (with $p > 0.5$) secretly before answering. $U_i = 1$ if head, and $U_i = 0$ if tail.

Ex: randomized response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X_i = 1$: cheated; $X_i = 0$, not cheated.
- Each student i tosses a biased coin (with $p > 0.5$) secretly before answering. $U_i = 1$ if head, and $U_i = 0$ if tail.
- Report $T_i = 1$ if $X_i = U_i$, and otherwise report $T_i = 0$.

Ex: randomized response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X_i = 1$: cheated; $X_i = 0$, not cheated.
- Each student i tosses a biased coin (with $p > 0.5$) secretly before answering. $U_i = 1$ if head, and $U_i = 0$ if tail.
- Report $T_i = 1$ if $X_i = U_i$, and otherwise report $T_i = 0$.
- At the individual level, $T_i = 1$ can mean i is a cheater or not a cheater.

Ex: randomized response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X_i = 1$: cheated; $X_i = 0$, not cheated.
- Each student i tosses a biased coin (with $p > 0.5$) secretly before answering. $U_i = 1$ if head, and $U_i = 0$ if tail.
- Report $T_i = 1$ if $X_i = U_i$, and otherwise report $T_i = 0$.
- At the individual level, $T_i = 1$ can mean i is a cheater or not a cheater.
- But in aggregation:

$$p_T = \Pr(U = X) = p \times p_{\text{cheat}} + (1 - p) \times (1 - p_{\text{cheat}}).$$

Ex: randomized response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X_i = 1$: cheated; $X_i = 0$, not cheated.
- Each student i tosses a biased coin (with $p > 0.5$) secretly before answering. $U_i = 1$ if head, and $U_i = 0$ if tail.
- Report $T_i = 1$ if $X_i = U_i$, and otherwise report $T_i = 0$.
- At the individual level, $T_i = 1$ can mean i is a cheater or not a cheater.
- But in aggregation:

$$p_T = \Pr(U = X) = p \times p_{\text{cheat}} + (1 - p) \times (1 - p_{\text{cheat}}).$$

Recovering p_{cheat}:

$$p_{\text{cheat}} = \frac{p_T + p - 1}{2p - 1}$$

Ex: randomized response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X_i = 1$: cheated; $X_i = 0$, not cheated.
- Each student i tosses a biased coin (with $p > 0.5$) secretly before answering. $U_i = 1$ if head, and $U_i = 0$ if tail.
- Report $T_i = 1$ if $X_i = U_i$, and otherwise report $T_i = 0$.
- At the individual level, $T_i = 1$ can mean i is a cheater or not a cheater.
- But in aggregation:

$$p_T = \Pr(U = X) = p \times p_{\text{cheat}} + (1 - p) \times (1 - p_{\text{cheat}}).$$

Recovering p_{cheat}:

$$p_{\text{cheat}} = \frac{p_T + p - 1}{2p - 1}$$

Estimate

$$\hat{p}_{\text{cheat}} = \frac{\bar{T}_n + p - 1}{2p - 1}$$

Ex: randomized response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X_i = 1$: cheated; $X_i = 0$, not cheated.
- Each student i tosses a biased coin (with $p > 0.5$) secretly before answering. $U_i = 1$ if head, and $U_i = 0$ if tail.
- Report $T_i = 1$ if $X_i = U_i$, and otherwise report $T_i = 0$.
- At the individual level, $T_i = 1$ can mean i is a cheater or not a cheater.
- But in aggregation:

$$p_T = \Pr(U = X) = p \times p_{\text{cheat}} + (1 - p) \times (1 - p_{\text{cheat}}).$$

Recovering p_{cheat}:

$$p_{\text{cheat}} = \frac{p_T + p - 1}{2p - 1}$$

Estimate

$$\hat{p}_{\text{cheat}} = \frac{\bar{T}_n + p - 1}{2p - 1}$$

Ex: $T_n = 0.45$, $p = 0.6$

$$\hat{p}_{\text{cheat}} = \frac{0.45 + 0.6 - 1}{2 \times 0.6 - 1} = 0.25$$

What is the loss of information or the gain in privacy?

Increased variance

$$\mathrm{Var}(\hat{p}_{\mathrm{cheat}}) = \frac{1}{n}\frac{p_T(1-p_T)}{(2p-1)^2} \leq \frac{1}{16n}\frac{1}{(p-0.5)^2}$$

What is the loss of information or the gain in privacy?

Increased variance

$$\mathrm{Var}(\hat{p}_{\text{cheat}}) = \frac{1}{n}\frac{p_T(1-p_T)}{(2p-1)^2} \leq \frac{1}{16n}\frac{1}{(p-0.5)^2}$$

Control relative risk of disclosure via controlling the likelihood ratio

$$\frac{\pi(X_i=1|T_i)}{\pi(X_i=0|T_i)} = \frac{\Pr(T_i|X_i=1)}{\Pr(T_i|X_i=0)}\frac{\pi(X_i=1)}{\pi(X_i=0)}$$

What is the loss of information or the gain in privacy?

Increased variance

$$\mathrm{Var}(\hat{p}_{\mathrm{cheat}}) = \frac{1}{n}\frac{p_T(1-p_T)}{(2p-1)^2} \leq \frac{1}{16n}\frac{1}{(p-0.5)^2}$$

Control relative risk of disclosure via controlling the likelihood ratio

$$\frac{\pi(X_i=1|T_i)}{\pi(X_i=0|T_i)} = \frac{\Pr(T_i|X_i=1)}{\Pr(T_i|X_i=0)}\frac{\pi(X_i=1)}{\pi(X_i=0)} \Leftrightarrow \frac{\pi(X_i=1|T_i)}{\pi(X_i=1)} \Big/ \frac{\pi(X_i=0|T_i)}{\pi(X_i=0)} = \frac{\Pr(T_i|X_i=1)}{\Pr(T_i|X_i=0)}$$

What is the loss of information or the gain in privacy?

Increased variance

$$\mathrm{Var}(\hat{p}_{\text{cheat}}) = \frac{1}{n} \frac{p_T(1 - p_T)}{(2p - 1)^2} \leq \frac{1}{16n} \frac{1}{(p - 0.5)^2}$$

Control relative risk of disclosure via controlling the likelihood ratio

$$\frac{\pi(X_i\!=\!1|T_i)}{\pi(X_i\!=\!0|T_i)} = \frac{\Pr(T_i|X_i\!=\!1)}{\Pr(T_i|X_i\!=\!0)} \frac{\pi(X_i\!=\!1)}{\pi(X_i\!=\!0)} \Leftrightarrow \frac{\pi(X_i\!=\!1|T_i)}{\pi(X_i\!=\!1)} \bigg/ \frac{\pi(X_i\!=\!0|T_i)}{\pi(X_i\!=\!0)} = \frac{\Pr(T_i|X_i\!=\!1)}{\Pr(T_i|X_i\!=\!0)}$$

The “first” example of *differential privacy*

$$\frac{\Pr\left(T_i = 1 \mid X_i = 1\right)}{\Pr\left(T_i = 1 \mid X_i = 0\right)} = \frac{p}{1 - p} = e^{\varepsilon}, \quad \text{with } \varepsilon = \mathrm{logit}(p)$$

What is the loss of information or the gain in privacy?

Increased variance

$$\mathrm{Var}(\hat{p}_{\mathrm{cheat}}) = \frac{1}{n}\frac{p_T(1-p_T)}{(2p-1)^2} \leq \frac{1}{16n}\frac{1}{(p-0.5)^2}$$

Control relative risk of disclosure via controlling the likelihood ratio

$$\frac{\pi(X_i=1|T_i)}{\pi(X_i=0|T_i)} = \frac{\Pr(T_i|X_i=1)}{\Pr(T_i|X_i=0)}\frac{\pi(X_i=1)}{\pi(X_i=0)} \Leftrightarrow \frac{\pi(X_i=1|T_i)}{\pi(X_i=1)} \Big/ \frac{\pi(X_i=0|T_i)}{\pi(X_i=0)} = \frac{\Pr(T_i|X_i=1)}{\Pr(T_i|X_i=0)}$$

The “first” example of *differential privacy*

$$\frac{\Pr(T_i = 1 \mid X_i = 1)}{\Pr(T_i = 1 \mid X_i = 0)} = \frac{p}{1-p} = e^{\varepsilon}, \quad \text{with } \varepsilon = \mathrm{logit}(p)$$

$$\frac{\Pr(T_i = 0 \mid X_i = 1)}{\Pr(T_i = 0 \mid X_i = 0)} = \frac{1-p}{p} = e^{-\varepsilon}$$

What is the loss of information or the gain in privacy?

Increased variance

$$\mathrm{Var}(\hat{p}_{\mathrm{cheat}}) = \frac{1}{n}\frac{p_T(1-p_T)}{(2p-1)^2} \leq \frac{1}{16n}\frac{1}{(p-0.5)^2}$$

Control relative risk of disclosure via controlling the likelihood ratio

$$\frac{\pi(X_i=1|T_i)}{\pi(X_i=0|T_i)} = \frac{\Pr(T_i|X_i=1)}{\Pr(T_i|X_i=0)}\frac{\pi(X_i=1)}{\pi(X_i=0)} \Leftrightarrow \frac{\pi(X_i=1|T_i)}{\pi(X_i=1)}\Big/\frac{\pi(X_i=0|T_i)}{\pi(X_i=0)} = \frac{\Pr(T_i|X_i=1)}{\Pr(T_i|X_i=0)}$$

The "first" example of *differential privacy*

$$\frac{\Pr\left(T_i = 1 \mid X_i = 1\right)}{\Pr\left(T_i = 1 \mid X_i = 0\right)} = \frac{p}{1-p} = e^{\varepsilon}, \quad \text{with } \varepsilon = \mathrm{logit}(p)$$

$$\frac{\Pr\left(T_i = 0 \mid X_i = 1\right)}{\Pr\left(T_i = 0 \mid X_i = 0\right)} = \frac{1-p}{p} = e^{-\varepsilon}$$

$$e^{-\varepsilon} \leq \frac{\Pr\left(T_i = t \mid X_i = 1\right)}{\Pr\left(T_i = t \mid X_i = 0\right)} \leq e^{\varepsilon}, \quad \text{for } t = 0, 1.$$

What is the loss of information or the gain in privacy?

Increased variance

$$\mathrm{Var}(\hat{p}_{\text{cheat}}) = \frac{1}{n}\frac{p_T(1-p_T)}{(2p-1)^2} \leq \frac{1}{16n}\frac{1}{(p-0.5)^2}$$

Control relative risk of disclosure via controlling the likelihood ratio

$$\frac{\pi(X_i=1|T_i)}{\pi(X_i=0|T_i)} = \frac{\Pr(T_i|X_i=1)}{\Pr(T_i|X_i=0)}\frac{\pi(X_i=1)}{\pi(X_i=0)} \Leftrightarrow \frac{\pi(X_i=1|T_i)}{\pi(X_i=1)}\Big/\frac{\pi(X_i=0|T_i)}{\pi(X_i=0)} = \frac{\Pr(T_i|X_i=1)}{\Pr(T_i|X_i=0)}$$

The “first” example of *differential privacy*

$$\frac{\Pr(T_i = 1 \mid X_i = 1)}{\Pr(T_i = 1 \mid X_i = 0)} = \frac{p}{1-p} = e^{\varepsilon}, \quad \text{with } \varepsilon = \operatorname{logit}(p)$$

$$\frac{\Pr(T_i = 0 \mid X_i = 1)}{\Pr(T_i = 0 \mid X_i = 0)} = \frac{1-p}{p} = e^{-\varepsilon}$$

$$e^{-\varepsilon} \leq \frac{\Pr(T_i = t \mid X_i = 1)}{\Pr(T_i = t \mid X_i = 0)} \leq e^{\varepsilon}, \quad \text{for } t = 0, 1.$$

$$d_{\text{MULT}}(\mathrm{P}_{\boldsymbol{x}}, \mathrm{P}_{\boldsymbol{x}'}) \leq \varepsilon d_{\text{Ham}}(\boldsymbol{x}, \boldsymbol{x}'), \quad \text{for } \boldsymbol{x}, \boldsymbol{x}' \in \{0, 1\}^n.$$

Does pure ε-DP control disclosure?

Recall: Control disclosure $\Leftrightarrow$ control the "difference" between $\pi(X_i)$ and $\pi(X_i \mid T = t)$.

Does pure ε-DP control disclosure?

Recall: Control disclosure $\Leftrightarrow$ control the "difference" between $\pi(X_i)$ and $\pi(X_i \mid T = t)$.

The "strongest" attacker knows the values of $\boldsymbol{x}_{-i}$:

$$\pi(\boldsymbol{X} = \boldsymbol{x}) = \pi(X_i = x_i)\delta_{\boldsymbol{x}_{-i}=\boldsymbol{x}^*_{-i}}.$$

Does pure ε-DP control disclosure?

Recall: Control disclosure $\Leftrightarrow$ control the "difference" between $\pi(X_i)$ and $\pi(X_i \mid T = t)$.

The "strongest" attacker knows the values of $\boldsymbol{x}_{-i}$:

$$\pi(\boldsymbol{X} = \boldsymbol{x}) = \pi(X_i = x_i)\delta_{\boldsymbol{x}_{-i} = \boldsymbol{x}^*_{-i}}.$$

Then

$$\frac{\pi(X_i = x_i \mid T = t)}{\pi(X_i = x_i)} = \frac{\pi(X_i = x_i) \int \mathrm{p}_{\boldsymbol{x}}(T = t) d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)}{\pi(X_i = x_i) \int \mathrm{p}_{\boldsymbol{x}'}(T = t) d\pi(\boldsymbol{X} = \boldsymbol{x}')}$$

Does pure ε-DP control disclosure?

Recall: Control disclosure $\Leftrightarrow$ control the "difference" between $\pi(X_i)$ and $\pi(X_i \mid T = t)$.

The "strongest" attacker knows the values of $\boldsymbol{x}_{-i}$:

$$\pi(\boldsymbol{X} = \boldsymbol{x}) = \pi(X_i = x_i)\delta_{\boldsymbol{x}_{-i} = \boldsymbol{x}^*_{-i}}.$$

Then

$$\begin{aligned}\frac{\pi(X_i = x_i \mid T = t)}{\pi(X_i = x_i)} &= \frac{\pi(X_i = x_i) \int \mathrm{p}_{\boldsymbol{x}}(T = t) d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)}{\pi(X_i = x_i) \int \mathrm{p}_{\boldsymbol{x}'}(T = t) d\pi(\boldsymbol{X} = \boldsymbol{x}')} \\ &= \frac{\int \mathrm{p}_{\boldsymbol{x}}(T = t) d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)}{\int \mathrm{p}_{\boldsymbol{x}'}(T = t) d\pi(\boldsymbol{X} = \boldsymbol{x}')}\end{aligned}$$

Does pure ε-DP control disclosure?

Recall: Control disclosure $\Leftrightarrow$ control the "difference" between $\pi(X_i)$ and $\pi(X_i \mid T = t)$.

The "strongest" attacker knows the values of $\boldsymbol{x}_{-i}$:

$$\pi(\boldsymbol{X} = \boldsymbol{x}) = \pi(X_i = x_i)\delta_{\boldsymbol{x}_{-i} = \boldsymbol{x}^*_{-i}}.$$

Then

$$\begin{aligned}
\frac{\pi(X_i = x_i \mid T = t)}{\pi(X_i = x_i)} &= \frac{\pi(X_i = x_i) \int \mathrm{p}_{\boldsymbol{x}}(T = t) d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)}{\pi(X_i = x_i) \int \mathrm{p}_{\boldsymbol{x}'}(T = t) d\pi(\boldsymbol{X} = \boldsymbol{x}')} \\
&= \frac{\int \mathrm{p}_{\boldsymbol{x}}(T = t) d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)}{\int \mathrm{p}_{\boldsymbol{x}'}(T = t) d\pi(\boldsymbol{X} = \boldsymbol{x}')} \\
&= \frac{\mathrm{p}(T = t \mid X_i = x_i, \boldsymbol{X}_{-i} = \boldsymbol{x}^*_{-i})}{\int \mathrm{p}(T = t \mid X_i = x'_i, \boldsymbol{X}_{-i} = \boldsymbol{x}^*_{-i}) d\pi(X_i = x'_i)}
\end{aligned}$$

Does pure ε-DP control disclosure?

Recall: Control disclosure $\Leftrightarrow$ control the "difference" between $\pi(X_i)$ and $\pi(X_i \mid T = t)$.

The "strongest" attacker knows the values of $\boldsymbol{x}_{-i}$:

$$\pi(\boldsymbol{X} = \boldsymbol{x}) = \pi(X_i = x_i)\delta_{\boldsymbol{x}_{-i}=\boldsymbol{x}^*_{-i}}.$$

Then

$$\begin{aligned}
\frac{\pi(X_i = x_i \mid T = t)}{\pi(X_i = x_i)} &= \frac{\pi(X_i = x_i)\int \mathrm{p}_{\boldsymbol{x}}(T = t)d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)}{\pi(X_i = x_i)\int \mathrm{p}_{\boldsymbol{x}'}(T = t)d\pi(\boldsymbol{X} = \boldsymbol{x}')} \\
&= \frac{\int \mathrm{p}_{\boldsymbol{x}}(T = t)d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)}{\int \mathrm{p}_{\boldsymbol{x}'}(T = t)d\pi(\boldsymbol{X} = \boldsymbol{x}')} \\
&= \frac{\mathrm{p}(T = t \mid X_i = x_i, \boldsymbol{X}_{-i} = \boldsymbol{x}^*_{-i})}{\int \mathrm{p}(T = t \mid X_i = x_i', \boldsymbol{X}_{-i} = \boldsymbol{x}^*_{-i})d\pi(X_i = x_i')} \\
&\leq e^{\varepsilon}.
\end{aligned}$$

Ex: randomised response (cont.)

Recall $T_i = 1_{\{X_i = U_i\}}$.

Suppose an adversary's prior for X_1 is $\pi(X_1 = 1) = \theta$. Given $t \in \{0, 1\}$,

$$\begin{aligned} C_\theta(t) &:= \frac{\pi(X_1 = 1 | T_1 = t)}{\pi(X_1 = 1)} = \frac{\Pr(T_1 = t | X_1 = 1)}{\Pr(T_1 = t)} \\ &= \frac{LR(t)}{LR(t)\theta + (1 - \theta)}, \quad \text{where } LR(t) = \frac{\Pr\left(T_1 = t \mid X_1 = 1\right)}{\Pr\left(T_1 = t \mid X_1 = 0\right)} \end{aligned}$$

Ex: randomised response (cont.)

Recall $T_i = 1_{\{X_i = U_i\}}$.

Suppose an adversary's prior for X_1 is $\pi(X_1 = 1) = \theta$. Given $t \in \{0, 1\}$,

$$C_\theta(t) := \frac{\pi(X_1 = 1 | T_1 = t)}{\pi(X_1 = 1)} = \frac{\Pr(T_1 = t | X_1 = 1)}{\Pr(T_1 = t)}$$

$$= \frac{LR(t)}{LR(t)\theta + (1 - \theta)}, \quad \text{where } LR(t) = \frac{\Pr\left(T_1 = t \mid X_1 = 1\right)}{\Pr\left(T_1 = t \mid X_1 = 0\right)}$$

$$LR(t) \geq 1 \ \Rightarrow 1 \leq C_\theta(t) \leq LR(t)$$

$$\max_\theta C_\theta(t) = C_0(t) = LR(t)$$

$$\min_\theta C_\theta(t) = C_1(t) = 1$$

Ex: randomised response (cont.)

Recall $T_i = 1_{\{X_i = U_i\}}$.

Suppose an adversary's prior for X_1 is $\pi(X_1 = 1) = \theta$. Given $t \in \{0, 1\}$,

$$C_\theta(t) := \frac{\pi(X_1 = 1 | T_1 = t)}{\pi(X_1 = 1)} = \frac{\Pr(T_1 = t | X_1 = 1)}{\Pr(T_1 = t)}$$

$$= \frac{LR(t)}{LR(t)\theta + (1 - \theta)}, \quad \text{where } LR(t) = \frac{\Pr(T_1 = t \mid X_1 = 1)}{\Pr(T_1 = t \mid X_1 = 0)}$$

$$LR(t) \geq 1 \Rightarrow 1 \leq C_\theta(t) \leq LR(t) \qquad LR(t) \leq 1 \Rightarrow LR(t) \leq C_\theta(t) \leq 1$$

$$\max_\theta C_\theta(t) = C_0(t) = LR(t) \qquad \max_\theta C_\theta(t) = C_1(t) = 1$$

$$\min_\theta C_\theta(t) = C_1(t) = 1 \qquad \min_\theta C_\theta(t) = C_0(t) = LR(t)$$

Ex: randomised response (cont.)

Recall $T_i = 1_{\{X_i = U_i\}}$.

Suppose an adversary's prior for X_1 is $\pi(X_1 = 1) = \theta$. Given $t \in \{0, 1\}$,

$$C_\theta(t) := \frac{\pi(X_1 = 1 | T_1 = t)}{\pi(X_1 = 1)} = \frac{\Pr(T_1 = t | X_1 = 1)}{\Pr(T_1 = t)}$$

$$= \frac{LR(t)}{LR(t)\theta + (1-\theta)}, \quad \text{where } LR(t) = \frac{\Pr(T_1 = t \mid X_1 = 1)}{\Pr(T_1 = t \mid X_1 = 0)}$$

$$LR(t) \geq 1 \ \Rightarrow 1 \leq C_\theta(t) \leq LR(t) \qquad LR(t) \leq 1 \ \Rightarrow LR(t) \leq C_\theta(t) \leq 1$$

$$\max_\theta C_\theta(t) = C_0(t) = LR(t) \qquad \max_\theta C_\theta(t) = C_1(t) = 1$$

$$\min_\theta C_\theta(t) = C_1(t) = 1 \qquad \min_\theta C_\theta(t) = C_0(t) = LR(t)$$

The prior-to-posterior semantic for differential privacy:

$$e^{-\varepsilon} \leq C_\theta(t) \leq e^{\varepsilon} \quad \text{for all } \theta \text{ if and only if} \quad e^{-\varepsilon} \leq LR(t) \leq e^{\varepsilon} \quad \text{for all } t$$

However, what if X_1 and X_2 are *a priori* dependent?

Suppose our prior for (X_1, X_2) is $\pi(X_1 = a, X_2 = b) = \theta_{ab}$. Let

$$C_\pi(t_1, t_2) := \frac{\Pr(X_1 = 1 | T_1 = t_1, T_2 = t_2)}{\Pr(X_1 = 1)} = \frac{\Pr(T_1 = t_1, T_2 = t_2 | X_1 = 1)}{\Pr(T_1 = t_1, T_2 = t_2)}$$

Transferring the bound on likelihood ratio to posterior-to-prior ratio

$$C_\theta(t_1, t_2) = \frac{LR(t_1, t_2)}{LR(t_1, t_2)\theta_{1\cdot} + (1 - \theta_{1\cdot})}, \quad \theta_{1\cdot} = \pi(X_1 = 1) = \theta_{11} + \theta_{10}$$

$$LR(t_1, t_2) = \frac{\Pr(T_1 = t_1, T_2 = t_2 | X_1 = 1)}{\Pr(T_1 = t_1, T_2 = t_2 | X_1 = 0)}.$$

However, what if X_1 and X_2 are *a priori* dependent?

Suppose our prior for (X_1, X_2) is $\pi(X_1 = a, X_2 = b) = \theta_{ab}$. Let

$$C_\pi(t_1, t_2) := \frac{\Pr(X_1 = 1 | T_1 = t_1, T_2 = t_2)}{\Pr(X_1 = 1)} = \frac{\Pr(T_1 = t_1, T_2 = t_2 | X_1 = 1)}{\Pr(T_1 = t_1, T_2 = t_2)}$$

Transferring the bound on likelihood ratio to posterior-to-prior ratio

$$C_\theta(t_1, t_2) = \frac{LR(t_1, t_2)}{LR(t_1, t_2)\theta_{1\cdot} + (1 - \theta_{1\cdot})}, \quad \theta_{1\cdot} = \pi(X_1 = 1) = \theta_{11} + \theta_{10}$$

$$LR(t_1, t_2) = \frac{\Pr(T_1 = t_1, T_2 = t_2 | X_1 = 1)}{\Pr(T_1 = t_1, T_2 = t_2 | X_1 = 0)}.$$

Consider the case $t_1 = 1, t_2 = 1$, and recall $e^\varepsilon = p/(1 - p)$

$$LR(1, 1) = \frac{e^\varepsilon \frac{\theta_{11}}{\theta_{1\cdot}} + \frac{\theta_{10}}{\theta_{1\cdot}}}{\frac{\theta_{01}}{\theta_{0\cdot}} + e^{-\varepsilon} \frac{\theta_{00}}{\theta_{0\cdot}}}$$

The dependence is a big trouble maker

This means that when $\theta_{10} = \theta_{01} = 0$, $LR(1, 1) = e^{2\varepsilon} > e^{\varepsilon}$.

- But $\theta_{10} = \theta_{01} = 0$ means that $X_2 = X_1$, hence X_1 can be learned from the information for X_2. Consequently, the "individual information unit" for X_1 should be the pair $\{X_1, X_2\}$, not merely X_1.

The dependence is a big trouble maker

This means that when $\theta_{10} = \theta_{01} = 0$, $LR(1,1) = e^{2\varepsilon} > e^{\varepsilon}$.

- But $\theta_{10} = \theta_{01} = 0$ means that $X_2 = X_1$, hence X_1 can be learned from the information for X_2. Consequently, the "individual information unit" for X_1 should be the pair $\{X_1, X_2\}$, not merely X_1.
- In fact as soon as $\text{Cov}(X_1, X_2) > 0$, $LR(1,1) > e^{\varepsilon}$. This is because

$$LR(1,1) > e^{\varepsilon} \quad \Longleftrightarrow \quad \pi(X_2 = 1|X_1 = 1) > \pi(X_2 = 1|X_1 = 0)$$

But

$$\begin{aligned} \text{Cov}(X_1, X_2) &= \pi(X_1 = 1, X_2 = 1) - \pi(X_1 = 1)\Pr(X_2 = 1) \\ &= [\pi(X_2 = 1|X_1 = 1) - \pi(X_2 = 1|X_1 = 0)]\,\pi(X_1 = 0)\pi(X_1 = 1). \end{aligned}$$

Data are *accidental* representation, not *essential* information

Manipulating data values without considering their interdependence is not a legitimate information operation in general

Does pure ε-DP control disclosure?

For a general prior π,

$$\frac{\pi(X_i = x_i \mid T = t)}{\pi(X_i = x_i)} = \frac{\pi(X_i = x_i) \int \mathrm{p}_{\boldsymbol{x}}(T = t) d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)}{\pi(X_i = x_i) \int \mathrm{p}_{\boldsymbol{x}'}(T = t) d\pi(\boldsymbol{X} = \boldsymbol{x}')}$$

Does pure ε-DP control disclosure?

For a general prior π,

$$\begin{aligned}\frac{\pi(X_i = x_i \mid T = t)}{\pi(X_i = x_i)} &= \frac{\pi(X_i = x_i) \int \mathrm{p}_{\boldsymbol{x}}(T = t) d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)}{\pi(X_i = x_i) \int \mathrm{p}_{\boldsymbol{x}'}(T = t) d\pi(\boldsymbol{X} = \boldsymbol{x}')} \\ &= \frac{\int \mathrm{p}_{\boldsymbol{x}}(T = t) d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)}{\int \mathrm{p}_{\boldsymbol{x}'}(T = t) d\pi(\boldsymbol{X} = \boldsymbol{x}')}\end{aligned}$$

Does pure ε-DP control disclosure?

For a general prior π,

$$\begin{aligned}
\frac{\pi(X_i = x_i \mid T = t)}{\pi(X_i = x_i)} &= \frac{\pi(X_i = x_i) \int \mathrm{p}_{\boldsymbol{x}}(T = t) d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)}{\pi(X_i = x_i) \int \mathrm{p}_{\boldsymbol{x}'}(T = t) d\pi(\boldsymbol{X} = \boldsymbol{x}')} \\
&= \frac{\int \mathrm{p}_{\boldsymbol{x}}(T = t) d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)}{\int \mathrm{p}_{\boldsymbol{x}'}(T = t) d\pi(\boldsymbol{X} = \boldsymbol{x}')} \\
&= \int \frac{1}{\int \frac{\mathrm{p}_{\boldsymbol{x}'}(T=t)}{\mathrm{p}_{\boldsymbol{x}}(T=t)} d\pi(\boldsymbol{X} = \boldsymbol{x}')} d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)
\end{aligned}$$

Does pure ε-DP control disclosure?

For a general prior π,

$$
\begin{aligned}
\frac{\pi(X_i = x_i \mid T = t)}{\pi(X_i = x_i)} &= \frac{\pi(X_i = x_i) \int \mathrm{p}_{\boldsymbol{x}}(T = t) d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)}{\pi(X_i = x_i) \int \mathrm{p}_{\boldsymbol{x}'}(T = t) d\pi(\boldsymbol{X} = \boldsymbol{x}')} \\
&= \frac{\int \mathrm{p}_{\boldsymbol{x}}(T = t) d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i)}{\int \mathrm{p}_{\boldsymbol{x}'}(T = t) d\pi(\boldsymbol{X} = \boldsymbol{x}')} \\
&= \int \frac{1}{\int \frac{\mathrm{p}_{\boldsymbol{x}'}(T=t)}{\mathrm{p}_{\boldsymbol{x}}(T=t)} d\pi(\boldsymbol{X} = \boldsymbol{x}')} d\pi(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i} \mid X_i = x_i) \\
&\le e^{n\varepsilon},
\end{aligned}
$$

with equality as the records of $\boldsymbol{X}$ become totally dependent. (n is the number of records in $\boldsymbol{X}$.) (Dwork, McSherry, et al., 2006; Kifer & Machanavajjhala, 2011b)

What does DP actually protect?

What does DP actually protect?

- Does ε-DP guarantee the marginal prior-to-posterior ratio

$$e^{-\varepsilon} \leq \frac{\pi(X_i = x | T = t)}{\pi(X_i = x)} \leq e^{\varepsilon}, \quad \forall x, \forall t?$$ **No, not in general**

(Kifer & Machanavajjhala, 2011b, 2012; Tschantz et al., 2020)

What does DP actually protect?

- Does ε-DP guarantee the marginal prior-to-posterior ratio

$$e^{-\varepsilon} \leq \frac{\pi(X_i = x | T = t)}{\pi(X_i = x)} \leq e^{\varepsilon}, \quad \forall x, \forall t? \quad \textbf{No, not in general}$$

(Kifer & Machanavajjhala, 2011b, 2012; Tschantz et al., 2020)

- Does ε-DP guarantee the conditional prior-to-posterior ratio

$$e^{-\varepsilon} \leq \frac{\pi(X_i = x_i | T = t, \boldsymbol{X}_{-i})}{\pi(X_i = x | \boldsymbol{X}_{-i})} \leq e^{\varepsilon}? \quad \forall x, \forall t? \quad \textbf{Yes}$$

What does DP actually protect?

- Does ε-DP guarantee the marginal prior-to-posterior ratio
$$e^{-\varepsilon} \leq \frac{\pi(X_i = x | T = t)}{\pi(X_i = x)} \leq e^{\varepsilon}, \quad \forall x, \forall t?$$ **No, not in general**

(Kifer & Machanavajjhala, 2011b, 2012; Tschantz et al., 2020)

- Does ε-DP guarantee the conditional prior-to-posterior ratio
$$e^{-\varepsilon} \leq \frac{\pi(X_i = x_i | T = t, \boldsymbol{X}_{-i})}{\pi(X_i = x | \boldsymbol{X}_{-i})} \leq e^{\varepsilon}? \quad \forall x, \forall t?$$ **Yes**
- Thus the guaranteed limit e^{ε} is only for the **unique individual information**: variations unexplained by anyone else in the database or by knowledge on (and beyond) the database population.

A Bayesian characterisation of pure ε-DP (Bailie, Gong & Meng, 2024+)

A random statistic $T \in \mathbb{R}^d$ is ε-DP if and only if for every prior π on $\boldsymbol{X}$, every sub-σ field $\mathcal{F}$ of the corresponding full σ-field σ_π, every $B \in \mathscr{B}\left(\mathbb{R}^d\right)$, every i, and every $A \in \mathscr{B}(\Theta_i)$, where Θ_i is the state space of x_i, we have

$$e^{-c_i\varepsilon}\pi(X_i \in A \mid \mathcal{F}) \leq \pi\left(X_i \in A \mid T \in B; \mathcal{F}\right) \leq e^{c_i\varepsilon}\pi(X_i \in A \mid \mathcal{F}), \tag{1}$$

where c_i is the size of the *minimal information chamber* (MIC) for X_i.

A Bayesian characterisation of pure ε-DP (Bailie, Gong & Meng, 2024+)

A random statistic $T \in \mathbb{R}^d$ is ε-DP if and only if for every prior π on $\boldsymbol{X}$, every sub-σ field $\mathcal{F}$ of the corresponding full σ-field σ_π, every $B \in \mathscr{B}\left(\mathbb{R}^d\right)$, every i, and every $A \in \mathscr{B}(\Theta_i)$, where Θ_i is the state space of x_i, we have

$$e^{-c_i\varepsilon}\pi(X_i \in A \mid \mathcal{F}) \leq \pi\left(X_i \in A \mid T \in B; \mathcal{F}\right) \leq e^{c_i\varepsilon}\pi(X_i \in A \mid \mathcal{F}), \qquad (1)$$

where c_i is the size of the *minimal information chamber* (MIC) for X_i.

- *MIC* $= C_{-i} \cup \{X_i\}$: $C_{-i} \subset \boldsymbol{X}_{-i}$ is the *Markov boundary* for X_i, that is, the smallest subset of $\boldsymbol{X}_{-i}$ such that
$$\pi(X_i|\boldsymbol{X}_{-i}, \mathcal{F}) = \pi(X_i|C_{-i}, \mathcal{F}).$$
- MIC is the X_i's "information family" – knowing any one of them will provide information about X_i, in addition to public knowledge coded into $\mathcal{F}$.

A Bayesian characterisation of pure ε-DP (Bailie, Gong & Meng, 2024+)

A random statistic $T \in \mathbb{R}^d$ is ε-DP if and only if for every prior π on $\boldsymbol{X}$, every sub-σ field $\mathcal{F}$ of the corresponding full σ-field σ_π, every $B \in \mathscr{B}\left(\mathbb{R}^d\right)$, every i, and every $A \in \mathscr{B}(\Theta_i)$, where Θ_i is the state space of x_i, we have

$$e^{-c_i\varepsilon}\pi(X_i \in A \mid \mathcal{F}) \leq \pi\left(X_i \in A \mid T \in B; \mathcal{F}\right) \leq e^{c_i\varepsilon}\pi(X_i \in A \mid \mathcal{F}), \tag{1}$$

where c_i is the size of the *minimal information chamber* (MIC) for X_i.

- *MIC* $= C_{-i} \cup \{X_i\}$: $C_{-i} \subset \boldsymbol{X}_{-i}$ is the *Markov boundary* for X_i, that is, the smallest subset of $\boldsymbol{X}_{-i}$ such that
$$\pi(X_i|\boldsymbol{X}_{-i}, \mathcal{F}) = \pi(X_i|C_{-i}, \mathcal{F}).$$
- MIC is the X_i's "information family" – knowing any one of them will provide information about X_i, in addition to public knowledge coded into $\mathcal{F}$.
- Protecting *relative* risk against "strongest attacker" is the easiest — the more the attacker's prior information, the less left for protection.

What if the attacker knows something about U?

What if the attacker knows something about U?

Ex: privacy amplification by sampling

- Suppose $T(\boldsymbol{x}, U)$ is ε-DP and $\mathcal{S}(\boldsymbol{x}, U')$ randomly samples f fraction of $\boldsymbol{x}$.

What if the attacker knows something about U?

Ex: privacy amplification by sampling

- Suppose $T(\boldsymbol{x}, U)$ is ε-DP and $\mathcal{S}(\boldsymbol{x}, U')$ randomly samples f fraction of $\boldsymbol{x}$.
- $T'(\boldsymbol{x}, U, U') := T(\mathcal{S}(\boldsymbol{x}, U'), U)$ is ε'-DP with $\varepsilon' \approx f\varepsilon < \varepsilon$.

What if the attacker knows something about U?

Ex: privacy amplification by sampling

- Suppose $T(\boldsymbol{x}, U)$ is ε-DP and $\mathcal{S}(\boldsymbol{x}, U')$ randomly samples f fraction of $\boldsymbol{x}$.
- $T'(\boldsymbol{x}, U, U') := T(\mathcal{S}(\boldsymbol{x}, U'), U)$ is ε'-DP with $\varepsilon' \approx f\varepsilon < \varepsilon$.
- So the (conditional) prior-to-posterior ratio of T' should be in the interval $[e^{-\varepsilon'}, e^{\varepsilon'}]$.

What if the attacker knows something about U?

Ex: privacy amplification by sampling

- Suppose $T(\boldsymbol{x}, U)$ is ε-DP and $\mathcal{S}(\boldsymbol{x}, U')$ randomly samples f fraction of $\boldsymbol{x}$.
- $T'(\boldsymbol{x}, U, U') := T(\mathcal{S}(\boldsymbol{x}, U'), U)$ is ε'-DP with $\varepsilon' \approx f\varepsilon < \varepsilon$.
- So the (conditional) prior-to-posterior ratio of T' should be in the interval $[e^{-\varepsilon'}, e^{\varepsilon'}]$.

What if the attacker knows something about U?

Ex: privacy amplification by sampling

- Suppose $T(\boldsymbol{x}, U)$ is ε-DP and $\mathcal{S}(\boldsymbol{x}, U')$ randomly samples f fraction of $\boldsymbol{x}$.
- $T'(\boldsymbol{x}, U, U') := T(\mathcal{S}(\boldsymbol{x}, U'), U)$ is ε'-DP with $\varepsilon' \approx f\varepsilon < \varepsilon$.
- So the (conditional) prior-to-posterior ratio of T' should be in the interval $[e^{-\varepsilon'}, e^{\varepsilon'}]$.

Traditional statistical disclosure control attacker models

- The *nosy neighbor:* Knows that a record is in the sample.
- The *journalist:* Wants to learn about *any* record, so picks one in the sample.

What if the attacker knows something about U?

Ex: privacy amplification by sampling

- Suppose $T(\boldsymbol{x}, U)$ is ε-DP and $\mathcal{S}(\boldsymbol{x}, U')$ randomly samples f fraction of $\boldsymbol{x}$.
- $T'(\boldsymbol{x}, U, U') := T(\mathcal{S}(\boldsymbol{x}, U'), U)$ is ε'-DP with $\varepsilon' \approx f\varepsilon < \varepsilon$.
- So the (conditional) prior-to-posterior ratio of T' should be in the interval $[e^{-\varepsilon'}, e^{\varepsilon'}]$.

Traditional statistical disclosure control attacker models

- The *nosy neighbor:* Knows that a record is in the sample.
- The *journalist:* Wants to learn about *any* record, so picks one in the sample.

For these attackers, the (conditional) prior-to-posterior ratio of T' is in the interval $[e^{-\varepsilon}, e^{\varepsilon}]$, *not* the interval $[e^{-\varepsilon'}, e^{\varepsilon'}]$ (Bailie & Drechsler, 2024+).

The US Decennial Census

- In the 1990, 2000 and 2010 Censuses, *data swapping* – a traditional statistical disclosure method – was used.

The US Decennial Census

- In the 1990, 2000 and 2010 Censuses, *data swapping* – a traditional statistical disclosure method – was used.
- In the 2010s, the US Census Bureau determined that swapping did not provide sufficient privacy protection.

The US Decennial Census

- In the 1990, 2000 and 2010 Censuses, *data swapping* – a traditional statistical disclosure method – was used.
- In the 2010s, the US Census Bureau determined that swapping did not provide sufficient privacy protection.
- For the 2020 Census, disclosure avoidance was overhauled with the primary aim of satisfying *differential privacy*.

The US Decennial Census

- In the 1990, 2000 and 2010 Censuses, *data swapping* – a traditional statistical disclosure method – was used.
- In the 2010s, the US Census Bureau determined that swapping did not provide sufficient privacy protection.
- For the 2020 Census, disclosure avoidance was overhauled with the primary aim of satisfying *differential privacy*.
- They use two bespoke DP methods: the *TopDown Algorithm* (J. Abowd et al., 2022) and *SafeTabs* (Tumult Labs, 2022).

DP revisited: Five building blocks

A general DP specification (Bailie et al., 2024+)

A statistic T is ε-differentially private if its Lipschitz constant is ε.

DP revisited: Five building blocks

A general DP specification (Bailie et al., 2024+)

A statistic $T : \mathcal{X} \times \mathcal{U} \to \mathcal{T}$ is ε-differentially private if its Lipschitz constant is ε.

DP revisited: Five building blocks

A general DP specification (Bailie et al., 2024+)

A statistic $T : \mathcal{X} \times \mathcal{U} \to \mathcal{T}$ is ε-differentially private if

$$d_{\mathrm{Pr}}(\mathrm{P}_{\boldsymbol{x}'}, \mathrm{P}_{\boldsymbol{x}}) \leq \varepsilon d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}),$$

DP revisited: Five building blocks

A general DP specification (Bailie et al., 2024+)

A statistic $T : \mathcal{X} \times \mathcal{U} \to \mathcal{T}$ satisfies a *DP specification* $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, d_{\mathsf{Pr}}, \varepsilon_{\mathcal{D}})$ if

$$d_{\mathsf{Pr}}(\mathsf{P}_{\boldsymbol{x}'}, \mathsf{P}_{\boldsymbol{x}}) \leq \varepsilon d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}),$$

DP revisited: Five building blocks

A general DP specification (Bailie et al., 2024+)

A statistic $T : \mathcal{X} \times \mathcal{U} \to \mathcal{T}$ satisfies a *DP specification* $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, d_{\mathsf{Pr}}, \varepsilon_{\mathcal{D}})$ if

$$d_{\mathsf{Pr}}(\mathsf{P}_{\boldsymbol{x}'}, \mathsf{P}_{\boldsymbol{x}}) \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}),$$

for all datasets $\boldsymbol{x}, \boldsymbol{x}'$ in every data universe $\mathcal{D}$ in the data multiverse $\mathscr{D}$.

DP revisited: Five building blocks

A general DP specification (Bailie et al., 2024+)

A statistic $T : \mathcal{X} \times \mathcal{U} \to \mathcal{T}$ satisfies a *DP specification* $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, d_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}})$ if

$$d_{\mathrm{Pr}}(\mathrm{P}_{\boldsymbol{x}'}, \mathrm{P}_{\boldsymbol{x}}) \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}),$$

for all datasets $\boldsymbol{x}, \boldsymbol{x}'$ in every data universe $\mathcal{D}$ in the data multiverse $\mathscr{D}$.

- The **protection domain** (*what* can be protected?): dataset space $\mathcal{X}$;

DP revisited: Five building blocks

A general DP specification (Bailie et al., 2024+)

A statistic $T : \mathcal{X} \times \mathcal{U} \to \mathcal{T}$ satisfies a *DP specification* $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, d_{\mathsf{Pr}}, \varepsilon_{\mathcal{D}})$ if

$$d_{\mathsf{Pr}}(\mathsf{P}_{\boldsymbol{x}'}, \mathsf{P}_{\boldsymbol{x}}) \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}),$$

for all datasets $\boldsymbol{x}, \boldsymbol{x}'$ in every data universe $\mathcal{D}$ in the data multiverse $\mathscr{D}$.

- The **protection domain** (*what* can be protected?): dataset space $\mathcal{X}$;
- The **scope of protection** (*to where* does the protection extend?): data multiverse $\mathscr{D}$ (*essential*), a collection of data universes $\mathcal{D} \subset \mathcal{X}$ (*accidental*);

DP revisited: Five building blocks

A general DP specification (Bailie et al., 2024+)

A statistic $T : \mathcal{X} \times \mathcal{U} \to \mathcal{T}$ satisfies a *DP specification* $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, d_{\mathsf{Pr}}, \varepsilon_{\mathcal{D}})$ if

$$d_{\mathsf{Pr}}(\mathsf{P}_{\boldsymbol{x}'}, \mathsf{P}_{\boldsymbol{x}}) \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}),$$

for all datasets $\boldsymbol{x}, \boldsymbol{x}'$ in every data universe $\mathcal{D}$ in the data multiverse $\mathscr{D}$.

- The **protection domain** (*what* can be protected?): dataset space $\mathcal{X}$;
- The **scope of protection** (*to where* does the protection extend?): data multiverse $\mathscr{D}$ (*essential*), a collection of data universes $\mathcal{D} \subset \mathcal{X}$ (*accidental*);
- The **protection units** (*who* are the units of protection): the input divergence $d_{\mathcal{X}}$ on $\mathcal{X}$;

DP revisited: Five building blocks

A general DP specification (Bailie et al., 2024+)

A statistic $T : \mathcal{X} \times \mathcal{U} \to \mathcal{T}$ satisfies a *DP specification* $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, d_{\mathsf{Pr}}, \varepsilon_{\mathcal{D}})$ if

$$d_{\mathsf{Pr}}(\mathrm{P}_{\boldsymbol{x}'}, \mathrm{P}_{\boldsymbol{x}}) \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}),$$

for all datasets $\boldsymbol{x}, \boldsymbol{x}'$ in every data universe $\mathcal{D}$ in the data multiverse $\mathscr{D}$.

- The **protection domain** (*what* can be protected?): dataset space $\mathcal{X}$;
- The **scope of protection** (*to where* does the protection extend?): data multiverse $\mathscr{D}$ (*essential*), a collection of data universes $\mathcal{D} \subset \mathcal{X}$ (*accidental*);
- The **protection units** (*who* are the units of protection): the input divergence $d_{\mathcal{X}}$ on $\mathcal{X}$;
- The **standard of protection** (*how* to measure protection): the divergence d_{Pr} on probabilities;

DP revisited: Five building blocks

A general DP specification (Bailie et al., 2024+)

A statistic $T : \mathcal{X} \times \mathcal{U} \to \mathcal{T}$ satisfies a *DP specification* $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, d_{\mathsf{Pr}}, \varepsilon_{\mathcal{D}})$ if

$$d_{\mathsf{Pr}}(\mathsf{P}_{\boldsymbol{x}'}, \mathsf{P}_{\boldsymbol{x}}) \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{x}', \boldsymbol{x}),$$

for all datasets $\boldsymbol{x}, \boldsymbol{x}'$ in every data universe $\mathcal{D}$ in the data multiverse $\mathscr{D}$.

- The **protection domain** (*what* can be protected?): dataset space $\mathcal{X}$;
- The **scope of protection** (*to where* does the protection extend?): data multiverse $\mathscr{D}$ (*essential*), a collection of data universes $\mathcal{D} \subset \mathcal{X}$ (*accidental*);
- The **protection units** (*who* are the units of protection): the input divergence $d_{\mathcal{X}}$ on $\mathcal{X}$;
- The **standard of protection** (*how* to measure protection): the divergence d_{Pr} on probabilities;
- The **intensity of protection** (*how much* protection is afforded): privacy loss budget $\varepsilon_{\mathcal{D}} \in \mathbb{R}^{\geq 0}$, for each data universe $\mathcal{D}$.

Data swapping visualisation

State	Location	Number of adults	Number of children	Age1	Race1	$\cdots$
MA	Cambridge	2	2	45	White	$\cdots$
TX	Houston	1	0	28	Hispanic	$\cdots$
WA	Tacoma	5	0	67	Asian	$\cdots$
MA	Somerville	2	2	50	Black	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Data swapping visualisation

State	Location	Number of adults	Number of children	Age1	Race1	$\cdots$
MA	Cambridge	2	2	45	White	$\cdots$
TX	Houston	1	0	28	Hispanic	$\cdots$
WA	Tacoma	5	0	67	Asian	$\cdots$
MA	Somerville	2	2	50	Black	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

$\mathbf{V}_{\text{Stratify}}$

Data swapping visualisation

State	Location	Number of adults	Number of children	Age1	Race1	$\cdots$
MA	Cambridge	2	2	45	White	$\cdots$
TX	Houston	1	0	28	Hispanic	$\cdots$
WA	Tacoma	5	0	67	Asian	$\cdots$
MA	Somerville	2	2	50	Black	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

$\mathbf{V}_{\text{Stratify}}$

Data swapping visualisation

State	Location	Number of adults	Number of children	Age1	Race1	$\cdots$
MA	Cambridge	2	2	45	White	$\cdots$
TX	Houston	1	0	28	Hispanic	$\cdots$
WA	Tacoma	5	0	67	Asian	$\cdots$
MA	Somerville	2	2	50	Black	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

$\mathbf{V}_{\text{Stratify}}$

$\mathbf{V}_{\text{Swap}}$

Data swapping visualisation

State	Location	Number of adults	Number of children	Age1	Race1	$\cdots$
MA	Somerville	2	2	45	White	$\cdots$
TX	Houston	1	0	28	Hispanic	$\cdots$
WA	Tacoma	5	0	67	Asian	$\cdots$
MA	Cambridge	2	2	50	Black	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

$\mathbf{V}_{\text{Stratify}}$

$\mathbf{V}_{\text{Swap}}$

Data swapping visualisation

State	Location	Number of adults	Number of children	Age1	Race1	$\cdots$
MA	Somerville	2	2	45	White	$\cdots$
TX	Houston	1	0	28	Hispanic	$\cdots$
WA	Tacoma	5	0	67	Asian	$\cdots$
MA	Cambridge	2	2	50	Black	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

$\boldsymbol{V}_{\text{Stratify}}$

$\boldsymbol{V}_{\text{Swap}}$

$\boldsymbol{V}_{\text{Rest}}$

Data swapping visualisation

Massachusetts: Location by Race (head of household) Contingency Table

	White	Hispanic	Asian	Black	...
Boston					
Cambridge					
Brookline					
Somerville					
Watertown					
⋮					

Data swapping visualisation

Massachusetts: Location by Race (head of household) Contingency Table

	White	Hispanic	Asian	Black	$\cdots$
Boston					
Cambridge	-1			+1	
Brookline					
Somerville	+1			-1	
Watertown					
$\vdots$					

Data swapping visualisation

Massachusetts: Location by Race (head of household) Contingency Table

	White	Hispanic	Asian	Black	$\cdots$
Boston					
Cambridge	-1			+1	
Brookline					
Somerville	+1			-1	
Watertown					
$\vdots$					

Changes: Interior cells of $\mathbf{V}_{\text{Rest}} \times \mathbf{V}_{\text{Swap}}$.

Data swapping visualisation

Massachusetts: Location by Race (head of household) Contingency Table

	White	Hispanic	Asian	Black	$\cdots$
Boston					
Cambridge	-1			+1	
Brookline					
Somerville	+1			-1	
Watertown					
$\vdots$					

Changes: Interior cells of $\mathbf{V}_{\text{Rest}} \times \mathbf{V}_{\text{Swap}}$.

Invariants:

1. $\mathbf{V}_{\text{Stratify}} \times \mathbf{V}_{\text{Rest}}$
2. $\mathbf{V}_{\text{Stratify}} \times \mathbf{V}_{\text{Swap}}$

Swapping satisfies DP, subject to its invariants

Permutation swapping

Input: a dataset $\boldsymbol{x}$.
Define strata as groups of records which match on the swap key $\mathbf{V}_{\text{Stratify}}$.
Within each stratum:

1. Select each record independently with probability p (the swap rate).
2. Randomly permute swapping variable $\mathbf{V}_{\text{Swap}}$ of selected records.

Output: the *swapped* dataset $\boldsymbol{w}$.

Swapping satisfies DP, subject to its invariants

Permutation swapping

Input: a dataset $\boldsymbol{x}$.
Define strata as groups of records which match on the swap key $\mathbf{V}_{\text{Stratify}}$.
Within each stratum:

1. Select each record independently with probability p (the swap rate).
2. Randomly permute swapping variable $\mathbf{V}_{\text{Swap}}$ of selected records.

Output: the *swapped* dataset $\boldsymbol{w}$.

Permutation swapping is DP subject to its invariants, with input divergence $d_{\mathcal{X}} = d^{u}_{\text{Ham}}$, output divergence $d_{\text{Pr}} = d_{\text{MULT}}$ and budget

$$\varepsilon = \begin{cases} \ln(b+1) - \ln o & \text{if } 0 < p \leq 0.5, \\ \max\left\{\ln o, \ln(b+1) - \ln o\right\} & \text{if } 0.5 < p < 1, \end{cases}$$

where $o = p/(1-p)$ and b is the maximum stratum size.

Comparisons: US Decennial Censuses

	d_{Pr}	$d_{\mathcal{X}}$ (Unit)	Invariants	Privacy Loss Budget
TopDown*	D_{nor}	d^{p}_{Ham} (person)	Population (state) Total housing units (block) Occupied group quarters (block) Structural zeros	PL & DHC: $\rho^2 = 15.29$ $\varepsilon = 52.83$ ($\delta = 10^{-10}$)
SafeTab**	D_{nor}	d^{p}_{Ham} (person)	None	DDHC-A: $\rho^2 = 19.776$ DDHC-B & S-DHC: *TBD.*
Swapping	$d_{\textsc{Mult}}$	d^{h}_{Ham} (household)	Varies but greater than TDA	ε between 9.37-19.38

*(J. Abowd et al., 2022)

**(Tumult Labs, 2022)

- $\mathcal{X}$ is always the space of possible Census Edited Files, $\mathcal{X}_{\mathrm{CEF}}$.
- $D_{\mathsf{nor}}(P, Q) = \mathsf{sup}_{\alpha > 1} \frac{1}{\sqrt{\alpha}} \max \left[\sqrt{D_\alpha(P||Q)}, \sqrt{D_\alpha(Q||P)} \right]$ is the normalised Rényi metric [zero concentrated DP] (with D_α the Rényi divergence of order);
- $d_{\textsc{Mult}}(P, Q) = \mathsf{sup}_{S \in \mathcal{F}} \left| \ln \frac{P(S)}{Q(S)} \right|$ is the multiplicative distance (pure DP); and
- d^{u}_{Ham} is the Hamming distance (on units u).

The TopDown Algorithm (TDA) (J. Abowd et al., 2022)

Two-step procedure:

0. Start with a Census edited file $\boldsymbol{x} \in \mathcal{X}_{\text{CEF}}$.

The TopDown Algorithm (TDA) (J. Abowd et al., 2022)

Two-step procedure:

0. Start with a Census edited file $\boldsymbol{x} \in \mathcal{X}_{\mathrm{CEF}}$.
1. Add Gaussian noise to cells:

$$\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{q}(\boldsymbol{x}) + \boldsymbol{W},$$

where $\boldsymbol{W} \sim \mathcal{N}_{\mathbb{Z}}(0, \boldsymbol{\Sigma})$, so that $\boldsymbol{T}$ satisfies $\mathrm{DP}(\mathcal{X}_{\mathrm{CEF}}, \{\mathcal{X}_{\mathrm{CEF}}\}, d^{p}_{\mathrm{Ham}}, D_{\mathrm{nor}})$ with budget ρ_{TDA} (Canonne et al., 2022).

The TopDown Algorithm (TDA) (J. Abowd et al., 2022)

Two-step procedure:

0. Start with a Census edited file $\boldsymbol{x} \in \mathcal{X}_{\text{CEF}}$.
1. Add Gaussian noise to cells:

$$\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{q}(\boldsymbol{x}) + \boldsymbol{W},$$

   where $\boldsymbol{W} \sim \mathcal{N}_{\mathbb{Z}}(0, \boldsymbol{\Sigma})$, so that $\boldsymbol{T}$ satisfies $\text{DP}(\mathcal{X}_{\text{CEF}}, \{\mathcal{X}_{\text{CEF}}\}, d^p_{\text{Ham}}, D_{\text{nor}})$ with budget ρ_{TDA} (Canonne et al., 2022).

2. “Post-process”: find dataset $\boldsymbol{z}$ with $\boldsymbol{q}(\boldsymbol{z})$ close to $\boldsymbol{T}(\boldsymbol{x})$ such that $\boldsymbol{c}_{\text{TDA}}(\boldsymbol{z}) = \boldsymbol{c}_{\text{TDA}}(\boldsymbol{x})$.

The TopDown Algorithm (TDA) (J. Abowd et al., 2022)

Two-step procedure:

0. Start with a Census edited file $\boldsymbol{x} \in \mathcal{X}_{\text{CEF}}$.
1. Add Gaussian noise to cells:

$$\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{q}(\boldsymbol{x}) + \boldsymbol{W},$$

   where $\boldsymbol{W} \sim \mathcal{N}_{\mathbb{Z}}(0, \boldsymbol{\Sigma})$, so that $\boldsymbol{T}$ satisfies $\text{DP}(\mathcal{X}_{\text{CEF}}, \{\mathcal{X}_{\text{CEF}}\}, d^{p}_{\text{Ham}}, D_{\text{nor}})$ with budget ρ_{TDA} (Canonne et al., 2022).

2. "Post-process": find dataset $\boldsymbol{z}$ with $\boldsymbol{q}(\boldsymbol{z})$ close to $\boldsymbol{T}(\boldsymbol{x})$ such that $\boldsymbol{c}_{\text{TDA}}(\boldsymbol{z}) = \boldsymbol{c}_{\text{TDA}}(\boldsymbol{x})$.

The TopDown Algorithm (TDA) (J. Abowd et al., 2022)

Two-step procedure:

0. Start with a Census edited file $\boldsymbol{x} \in \mathcal{X}_{\mathrm{CEF}}$.
1. Add Gaussian noise to cells:

$$\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{q}(\boldsymbol{x}) + \boldsymbol{W},$$

   where $\boldsymbol{W} \sim \mathcal{N}_{\mathbb{Z}}(0, \boldsymbol{\Sigma})$, so that $\boldsymbol{T}$ satisfies $\mathrm{DP}(\mathcal{X}_{\mathrm{CEF}}, \{\mathcal{X}_{\mathrm{CEF}}\}, d_{\mathrm{Ham}}^{p}, D_{\mathrm{nor}})$ with budget ρ_{TDA} (Canonne et al., 2022).
2. "Post-process": find dataset $\boldsymbol{z}$ with $\boldsymbol{q}(\boldsymbol{z})$ close to $\boldsymbol{T}(\boldsymbol{x})$ such that $\boldsymbol{c}_{\mathrm{TDA}}(\boldsymbol{z}) = \boldsymbol{c}_{\mathrm{TDA}}(\boldsymbol{x})$.

TDA satisfies $\mathrm{DP}(\mathcal{X}_{\mathrm{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}}, d_{\mathrm{Ham}}^{p}, D_{\mathrm{nor}})$ with budget ρ_{TDA}.

Theorem: TDA satisfies DP, subject to its invariants

Let $\boldsymbol{c}_{\text{TDA}} : \mathcal{X}_{\text{CEF}} \to \mathbb{R}^l$ be the invariants of TDA and let $\mathscr{D}_{\boldsymbol{c}_{\text{TDA}}}$ be the induced data multiverse:

$$\mathscr{D}_{\boldsymbol{c}_{\text{TDA}}} = \{\mathcal{D} \subset \mathcal{X}_{\text{CEF}} \mid \boldsymbol{c}_{\text{TDA}}(\boldsymbol{x}) = \boldsymbol{c}_{\text{TDA}}(\boldsymbol{x}') \; \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}\}.$$

Theorem: TDA satisfies DP, subject to its invariants

Let $\boldsymbol{c}_{\mathrm{TDA}} : \mathcal{X}_{\mathrm{CEF}} \to \mathbb{R}^l$ be the invariants of TDA and let $\mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}}$ be the induced data multiverse:

$$\mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}} = \{\mathcal{D} \subset \mathcal{X}_{\mathrm{CEF}} \mid \boldsymbol{c}_{\mathrm{TDA}}(\boldsymbol{x}) = \boldsymbol{c}_{\mathrm{TDA}}(\boldsymbol{x}') \ \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}\}.$$

- TDA satisfies DP$(\mathcal{X}_{\mathrm{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}}, d^p_{\mathrm{Ham}}, D_{\mathrm{nor}})$ with privacy budget $\rho_{\mathrm{TDA}} = 2.63$ (for the PL Redistricting File) and $\rho_{\mathrm{TDA}} = 15.29$ (for the DHC).

Theorem: TDA satisfies DP, subject to its invariants

Let $\boldsymbol{c}_{\mathrm{TDA}} : \mathcal{X}_{\mathrm{CEF}} \to \mathbb{R}^{l}$ be the invariants of TDA and let $\mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}}$ be the induced data multiverse:

$$\mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}} = \{\mathcal{D} \subset \mathcal{X}_{\mathrm{CEF}} \mid \boldsymbol{c}_{\mathrm{TDA}}(\boldsymbol{x}) = \boldsymbol{c}_{\mathrm{TDA}}(\boldsymbol{x}') \ \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}\}.$$

- TDA satisfies $\mathrm{DP}(\mathcal{X}_{\mathrm{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}}, d^{p}_{\mathrm{Ham}}, D_{\mathrm{nor}})$ with privacy budget $\rho_{\mathrm{TDA}} = 2.63$ (for the PL Redistricting File) and $\rho_{\mathrm{TDA}} = 15.29$ (for the DHC).
- Let $\boldsymbol{c}'$ be any proper subset of TDA's invariants. TDA does not satisfy $\mathrm{DP}(\mathcal{X}_{\mathrm{CEF}}, \mathscr{D}_{\boldsymbol{c}'}, d_{\mathcal{X}}, D_{\mathrm{nor}})$ with any finite budget ρ.

Protecting Privacy via Randomized Response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X = 1$: cheated; $X = 0$, not cheated.

Protecting Privacy via Randomized Response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X = 1$: cheated; $X = 0$, not cheated.
- Each student tosses a biased coin (with $p > 0.5$) secretly before answering. $R = 1$ if head, and $R = 0$ if tail.

Protecting Privacy via Randomized Response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X = 1$: cheated; $X = 0$, not cheated.
- Each student tosses a biased coin (with $p > 0.5$) secretly before answering. $R = 1$ if head, and $R = 0$ if tail.
- Report $Y = 1$ if $X = R$, and otherwise report $Y = 0$.

Protecting Privacy via Randomized Response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X = 1$: cheated; $X = 0$, not cheated.
- Each student tosses a biased coin (with $p > 0.5$) secretly before answering. $R = 1$ if head, and $R = 0$ if tail.
- Report $Y = 1$ if $X = R$, and otherwise report $Y = 0$.
- At the individual level, $Y_i = 1$ can mean a cheater or not a cheater.

Protecting Privacy via Randomized Response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X = 1$: cheated; $X = 0$, not cheated.
- Each student tosses a biased coin (with $p > 0.5$) secretly before answering. $R = 1$ if head, and $R = 0$ if tail.
- Report $Y = 1$ if $X = R$, and otherwise report $Y = 0$.
- At the individual level, $Y_i = 1$ can mean a cheater or not a cheater.
- But in aggregation:

$$p_Y = \Pr(R = X) = p \times p_{\text{cheat}} + (1 - p) \times (1 - p_{\text{cheat}})$$

Protecting Privacy via Randomized Response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X = 1$: cheated; $X = 0$, not cheated.
- Each student tosses a biased coin (with $p > 0.5$) secretly before answering. $R = 1$ if head, and $R = 0$ if tail.
- Report $Y = 1$ if $X = R$, and otherwise report $Y = 0$.
- At the individual level, $Y_i = 1$ can mean a cheater or not a cheater.
- But in aggregation:

$$p_Y = \Pr(R = X) = p \times p_{\text{cheat}} + (1 - p) \times (1 - p_{\text{cheat}})$$

Recovering p_{cheat}:

$$p_{\text{cheat}} = \frac{p_Y + p - 1}{2p - 1}$$

Protecting Privacy via Randomized Response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X = 1$: cheated; $X = 0$, not cheated.
- Each student tosses a biased coin (with $p > 0.5$) secretly before answering. $R = 1$ if head, and $R = 0$ if tail.
- Report $Y = 1$ if $X = R$, and otherwise report $Y = 0$.
- At the individual level, $Y_i = 1$ can mean a cheater or not a cheater.
- But in aggregation:

$$p_Y = \Pr(R = X) = p \times p_{\text{cheat}} + (1 - p) \times (1 - p_{\text{cheat}})$$

Recovering p_{cheat}:

$$p_{\text{cheat}} = \frac{p_Y + p - 1}{2p - 1}$$

Estimate

$$\hat{p}_{\text{cheat}} = \frac{\bar{Y}_n + p - 1}{2p - 1}$$

Protecting Privacy via Randomized Response (Warner, 1965)

- Estimating exam cheating rate p_{cheat}. $X = 1$: cheated; $X = 0$, not cheated.
- Each student tosses a biased coin (with $p > 0.5$) secretly before answering. $R = 1$ if head, and $R = 0$ if tail.
- Report $Y = 1$ if $X = R$, and otherwise report $Y = 0$.
- At the individual level, $Y_i = 1$ can mean a cheater or not a cheater.
- But in aggregation:

$$p_Y = \Pr(R = X) = p \times p_{\text{cheat}} + (1 - p) \times (1 - p_{\text{cheat}})$$

Recovering p_{cheat}:

$$p_{\text{cheat}} = \frac{p_Y + p - 1}{2p - 1}$$

Estimate

$$\hat{p}_{\text{cheat}} = \frac{\bar{Y}_n + p - 1}{2p - 1}$$

Ex: $\bar{Y}_n = 0.45$, $p = 0.6$

$$\hat{p}_{\text{cheat}} = \frac{0.45 + 0.6 - 1}{2 \times 0.6 - 1} = 0.25$$

What is the loss of information or the gain in privacy?

Increased Variance

$$\mathrm{Var}(\hat{p}_{\text{cheat}}) = \frac{1}{n}\frac{p_Y(1-p_Y)}{(2p-1)^2} \leq \frac{1}{16n}\frac{1}{(p-0.5)^2}$$

What is the loss of information or the gain in privacy?

Increased Variance

$$\mathrm{Var}(\hat{p}_{\mathrm{cheat}}) = \frac{1}{n} \frac{p_Y(1 - p_Y)}{(2p - 1)^2} \leq \frac{1}{16n} \frac{1}{(p - 0.5)^2}$$

Control Relative Risk via Controlling Likelihood Ratio

$$\frac{\Pr(X_i = 1 | Y_i)}{\Pr(X_i = 0 | Y_i)} = \frac{\Pr(Y_i | X_i = 1)}{\Pr(Y_i | X_i = 0)} \frac{\Pr(X_i = 1)}{\Pr(X_i = 0)}$$

What is the loss of information or the gain in privacy?

Increased Variance

$$\mathrm{Var}(\hat{p}_{\mathrm{cheat}}) = \frac{1}{n}\frac{p_Y(1-p_Y)}{(2p-1)^2} \leq \frac{1}{16n}\frac{1}{(p-0.5)^2}$$

Control Relative Risk via Controlling Likelihood Ratio

$$\frac{\Pr(X_i=1|Y_i)}{\Pr(X_i=0|Y_i)} = \frac{\Pr(Y_i|X_i=1)}{\Pr(Y_i|X_i=0)}\frac{\Pr(X_i=1)}{\Pr(X_i=0)}$$

The "first" example of *differential privacy*

$$\frac{\Pr\left(Y_i=1 \mid X_i=1\right)}{\Pr\left(Y_i=1 \mid X_i=0\right)} = \frac{p}{1-p} = e^{\varepsilon}, \quad \text{with } \varepsilon = \mathrm{logit}(p)$$

What is the loss of information or the gain in privacy?

Increased Variance

$$\mathrm{Var}(\hat{p}_{\mathrm{cheat}}) = \frac{1}{n}\frac{p_Y(1-p_Y)}{(2p-1)^2} \leq \frac{1}{16n}\frac{1}{(p-0.5)^2}$$

Control Relative Risk via Controlling Likelihood Ratio

$$\frac{\Pr(X_i = 1|Y_i)}{\Pr(X_i = 0|Y_i)} = \frac{\Pr(Y_i|X_i = 1)}{\Pr(Y_i|X_i = 0)}\frac{\Pr(X_i = 1)}{\Pr(X_i = 0)}$$

The “first” example of *differential privacy*

$$\frac{\Pr\left(Y_i = 1 \mid X_i = 1\right)}{\Pr\left(Y_i = 1 \mid X_i = 0\right)} = \frac{p}{1-p} = e^{\varepsilon}, \quad \text{with } \varepsilon = \mathrm{logit}(p)$$

$$\frac{\Pr\left(Y_i = 0 \mid X_i = 1\right)}{\Pr\left(Y_i = 0 \mid X_i = 0\right)} = \frac{1-p}{p} = e^{-\varepsilon}$$

What is the loss of information or the gain in privacy?

Increased Variance

$$\mathrm{Var}(\hat{p}_{\mathrm{cheat}}) = \frac{1}{n}\frac{p_Y(1-p_Y)}{(2p-1)^2} \leq \frac{1}{16n}\frac{1}{(p-0.5)^2}$$

Control Relative Risk via Controlling Likelihood Ratio

$$\frac{\Pr(X_i = 1|Y_i)}{\Pr(X_i = 0|Y_i)} = \frac{\Pr(Y_i|X_i = 1)}{\Pr(Y_i|X_i = 0)}\frac{\Pr(X_i = 1)}{\Pr(X_i = 0)}$$

The "first" example of *differential privacy*

$$\frac{\Pr(Y_i = 1 \mid X_i = 1)}{\Pr(Y_i = 1 \mid X_i = 0)} = \frac{p}{1-p} = e^{\varepsilon}, \quad \text{with } \varepsilon = \mathrm{logit}(p)$$

$$\frac{\Pr(Y_i = 0 \mid X_i = 1)}{\Pr(Y_i = 0 \mid X_i = 0)} = \frac{1-p}{p} = e^{-\varepsilon}$$

$$e^{-\varepsilon} \leq \frac{\Pr(Y_i = y \mid X_i = 1)}{\Pr(Y_i = y \mid X_i = 0)} \leq e^{\varepsilon}, \quad \text{for } y = 0, 1$$

What is the loss of information or the gain in privacy?

Increased Variance

$$\operatorname{Var}(\hat{p}_{\text{cheat}}) = \frac{1}{n}\frac{p_Y(1-p_Y)}{(2p-1)^2} \le \frac{1}{16n}\frac{1}{(p-0.5)^2}$$

Control Relative Risk via Controlling Likelihood Ratio

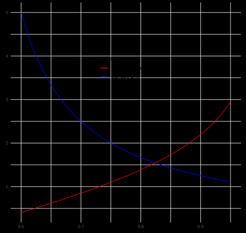
$$\frac{\Pr(X_i = 1|Y_i)}{\Pr(X_i = 0|Y_i)} = \frac{\Pr(Y_i|X_i = 1)}{\Pr(Y_i|X_i = 0)}\frac{\Pr(X_i = 1)}{\Pr(X_i = 0)}$$

The "first" example of *differential privacy*

$$\frac{\Pr\left(Y_i = 1 \mid X_i = 1\right)}{\Pr\left(Y_i = 1 \mid X_i = 0\right)} = \frac{p}{1-p} = e^{\varepsilon}, \quad \text{with } \varepsilon = \operatorname{logit}(p)$$

$$\frac{\Pr\left(Y_i = 0 \mid X_i = 1\right)}{\Pr\left(Y_i = 0 \mid X_i = 0\right)} = \frac{1-p}{p} = e^{-\varepsilon}$$

$$e^{-\varepsilon} \le \frac{\Pr\left(Y_i = y \mid X_i = 1\right)}{\Pr\left(Y_i = y \mid X_i = 0\right)} \le e^{\varepsilon}, \quad \text{for } y = 0, 1$$



Define *Pure* DP: Dwork et al. (2006) vs Dwork et al. (2016)

Let the database $\boldsymbol{X} = \{x_1, \ldots, x_n\}$ be a vector of n entries from some domain D, typically of the form $\{0, 1\}^d$ or $\mathbb{R}^d$. Let $T_{\mathcal{A}}$ be a random mechanism (map) from D^n to a state space $\mathcal{T}$, corresponding to a query from an adversary $\mathcal{A}$.

Define *Pure* DP: Dwork et al. (2006) vs Dwork et al. (2016)

Let the database $\boldsymbol{X} = \{x_1, \ldots, x_n\}$ be a vector of n entries from some domain D, typically of the form $\{0, 1\}^d$ or $\mathbb{R}^d$. Let $T_{\mathcal{A}}$ be a random mechanism (map) from D^n to a state space $\mathcal{T}$, corresponding to a query from an adversary $\mathcal{A}$.

Definition 1 of Dwork, McSherry, et al. (2006)

A mechanism is ε-indistinguishable if for all pairs $\boldsymbol{X}, \boldsymbol{X}' \in D^n$ which differ in only one entry, for all adversaries $\mathcal{A}$, and for all transcripts t:

$$\left| \ln \frac{\Pr(T_{\mathcal{A}}(\boldsymbol{X}) = t)}{\Pr(T_{\mathcal{A}}(\boldsymbol{X}') = t)} \right| \leq \varepsilon.$$

Define *Pure* DP: Dwork et al. (2006) vs Dwork et al. (2016)

Let the database $\boldsymbol{X} = \{x_1, \ldots, x_n\}$ be a vector of n entries from some domain D, typically of the form $\{0, 1\}^d$ or $\mathbb{R}^d$. Let $T_{\mathcal{A}}$ be a random mechanism (map) from D^n to a state space $\mathcal{T}$, corresponding to a query from an adversary $\mathcal{A}$.

Definition 1 of Dwork, McSherry, et al. (2006)

A mechanism is ε-indistinguishable if for all pairs $\boldsymbol{X}, \boldsymbol{X}' \in D^n$ which differ in only one entry, for all adversaries $\mathcal{A}$, and for all transcripts t:

$$\left| \ln \frac{\Pr\left(T_{\mathcal{A}}(\boldsymbol{X}) = t\right)}{\Pr\left(T_{\mathcal{A}}(\boldsymbol{X}') = t\right)} \right| \leq \varepsilon.$$

Define *Pure* DP: Dwork et al. (2006) vs Dwork et al. (2016)

Let the database $\boldsymbol{X} = \{x_1, \dots, x_n\}$ be a vector of n entries from some domain D, typically of the form $\{0, 1\}^d$ or $\mathbb{R}^d$. Let $T_{\mathcal{A}}$ be a random mechanism (map) from D^n to a state space $\mathcal{T}$, corresponding to a query from an adversary $\mathcal{A}$.

Definition 1 of Dwork, McSherry, et al. (2006)

A mechanism is ε-indistinguishable if for all pairs $\boldsymbol{X}, \boldsymbol{X}' \in D^n$ which differ in only one entry, for all adversaries $\mathcal{A}$, and for all transcripts t:

$$\left| \ln \frac{\Pr\left(T_{\mathcal{A}}(\boldsymbol{X}) = t\right)}{\Pr\left(T_{\mathcal{A}}(\boldsymbol{X}') = t\right)} \right| \leq \varepsilon.$$

Definition 2.1 of Dwork et al. (2016)

A noninteractive mechanism $\mathcal{M}$ is ε-differentially private (with respect to a given distance measure) if for all neighboring datasets $\boldsymbol{X}, \boldsymbol{X}' \in \mathbb{N}^{|D|}$, and for all events (measurable sets) S in the space of outputs of $\mathcal{M}$:

$$\Pr\left(M(\boldsymbol{X}) \in S\right) \leq e^{\varepsilon} \Pr\left(M(\boldsymbol{X}') \in S\right).$$

The probabilities are over the coin flips of M.

Differential Privacy for the 2020 U.S. Census: Can We Make Data Both Private and Useful?

Special Issue 2

FROM THE EDITORS

Harnessing the Known Unknowns: Differential Privacy and the 2020 Census

by Ruobin Gong, Erica L. Groshen, and Salil Vadhan

Published: Jun 24, 2022

Special Issue 2: Differential Privacy for the 2020 U.S. Census

CENSUS: IMPORTANCE, HISTORY, AND TECHNICAL CHANGES

Coming to Our Census: How Social Statistics Underpin Our Democracy (and Republic)

by Teresa A. Sullivan

Published: Jan 31, 2020

CONNECTIONS

Disclosure Protection in the Context of Statistical Agency Operations: Data Quality and Related Constraints

by John L. Eltinge

Published: Jun 24, 2022

Implementing Differential

Does DP control the posterior-to-prior ratio ?

Revisit the Random Response Mechanism: $Y_i = 1_{\{X_i = R_i\}}$.

Suppose an adversary's prior for X_1 is $\Pr(X_1 = 1) = \pi$.

$$
\begin{aligned}
C_\pi(y) &\equiv \frac{\Pr(X_1 = 1 | Y_1 = y)}{\Pr(X_1 = 1)} = \frac{\Pr(Y_1 = y | X_1 = 1)}{\Pr(Y_1 = y)} \\
&= \frac{LR(y)}{LR(y)\pi + (1-\pi)}, \quad \text{where } LR(y) = \frac{\Pr(Y_1 = y \mid X_1 = 1)}{\Pr(Y_1 = y \mid X_1 = 0)}
\end{aligned}
$$

Does DP control the posterior-to-prior ratio ?

Revisit the Random Response Mechanism: $Y_i = 1_{\{X_i = R_i\}}$.

Suppose an adversary's prior for X_1 is $\Pr(X_1 = 1) = \pi$.

$$C_\pi(y) \equiv \frac{\Pr(X_1 = 1 | Y_1 = y)}{\Pr(X_1 = 1)} = \frac{\Pr(Y_1 = y | X_1 = 1)}{\Pr(Y_1 = y)}$$

$$= \frac{LR(y)}{LR(y)\pi + (1 - \pi)}, \quad \text{where } LR(y) = \frac{\Pr\left(Y_1 = y \mid X_1 = 1\right)}{\Pr\left(Y_1 = y \mid X_1 = 0\right)}$$

$$LR(y) \geq 1 \ \Rightarrow 1 \leq C_\pi(y) \leq LR(y)$$

$$\max_\pi C_\pi(y) = C_0(y) = LR(y)$$

$$\min_\pi C_\pi(y) = C_1(y) = 1$$

Does DP control the posterior-to-prior ratio ?

Revisit the Random Response Mechanism: $Y_i = 1_{\{X_i = R_i\}}$.

Suppose an adversary's prior for X_1 is $\Pr(X_1 = 1) = \pi$.

$$C_\pi(y) \equiv \frac{\Pr(X_1 = 1 | Y_1 = y)}{\Pr(X_1 = 1)} = \frac{\Pr(Y_1 = y | X_1 = 1)}{\Pr(Y_1 = y)}$$

$$= \frac{LR(y)}{LR(y)\pi + (1 - \pi)}, \quad \text{where } LR(y) = \frac{\Pr(Y_1 = y \mid X_1 = 1)}{\Pr(Y_1 = y \mid X_1 = 0)}$$

$$LR(y) \geq 1 \Rightarrow 1 \leq C_\pi(y) \leq LR(y) \qquad LR(y) \leq 1 \Rightarrow LR(y) \leq C_\pi(y) \leq 1$$

$$\max_\pi C_\pi(y) = C_0(y) = LR(y) \qquad \max_\pi C_\pi(y) = C_1(y) = 1$$

$$\min_\pi C_\pi(y) = C_1(y) = 1 \qquad \min_\pi C_\pi(y) = C_0(y) = LR(y)$$

Does DP control the posterior-to-prior ratio ?

Revisit the Random Response Mechanism: $Y_i = 1_{\{X_i = R_i\}}$.

Suppose an adversary's prior for X_1 is $\Pr(X_1 = 1) = \pi$.

$$C_\pi(y) \equiv \frac{\Pr(X_1 = 1 | Y_1 = y)}{\Pr(X_1 = 1)} = \frac{\Pr(Y_1 = y | X_1 = 1)}{\Pr(Y_1 = y)}$$

$$= \frac{LR(y)}{LR(y)\pi + (1-\pi)}, \quad \text{where } LR(y) = \frac{\Pr(Y_1 = y \mid X_1 = 1)}{\Pr(Y_1 = y \mid X_1 = 0)}$$

$$LR(y) \geq 1 \;\Rightarrow 1 \leq C_\pi(y) \leq LR(y) \qquad LR(y) \leq 1 \;\Rightarrow LR(y) \leq C_\pi(y) \leq 1$$

$$\max_\pi C_\pi(y) = C_0(y) = LR(y) \qquad \max_\pi C_\pi(y) = C_1(y) = 1$$

$$\min_\pi C_\pi(y) = C_1(y) = 1 \qquad \min_\pi C_\pi(y) = C_0(y) = LR(y)$$

The prior-to-posterior semantic for differential privacy:

$$e^{-\varepsilon} \leq C_\pi(y) \leq e^{\varepsilon} \quad \text{for all } \pi \text{ if and only if } \quad e^{-\varepsilon} \leq LR(y) \leq e^{\varepsilon}$$

However, what if X_1 and X_2 are *a priori* dependent?

Suppose our prior for (X_1, X_2) is $\Pr(X_1 = a, X_2 = b) = \pi_{ab}$. Let

$$C_\pi(y_1, y_2) \equiv \frac{\Pr(X_1 = 1 | Y_1 = y_1, Y_2 = y_2)}{\Pr(X_1 = 1)} = \frac{\Pr(Y_1 = y_1, Y_2 = y_2 | X_1 = 1)}{\Pr(Y_1 = y_1, Y_2 = y_2)}$$

Transferring the bound on likelihood ratio to posterior-to-prior ratio

$$C_\pi(y_1, y_2) = \frac{LR(y_1, y_2)}{LR(y_1, y_2)\pi_{1\cdot} + (1 - \pi_{1\cdot})}, \quad \pi_{1\cdot} = \Pr(X_1 = 1) = \pi_{11} + \pi_{10}$$

$$LR(y_1, y_2) = \frac{\Pr(Y_1 = y_1, Y_2 = y_2 | X_1 = 1)}{\Pr(Y_1 = y_1, Y_2 = y_2 | X_1 = 0)}.$$

However, what if X_1 and X_2 are *a priori* dependent?

Suppose our prior for (X_1, X_2) is $\Pr(X_1 = a, X_2 = b) = \pi_{ab}$. Let

$$C_\pi(y_1, y_2) \equiv \frac{\Pr(X_1 = 1 | Y_1 = y_1, Y_2 = y_2)}{\Pr(X_1 = 1)} = \frac{\Pr(Y_1 = y_1, Y_2 = y_2 | X_1 = 1)}{\Pr(Y_1 = y_1, Y_2 = y_2)}$$

Transferring the bound on likelihood ratio to posterior-to-prior ratio

$$C_\pi(y_1, y_2) = \frac{LR(y_1, y_2)}{LR(y_1, y_2)\pi_{1\cdot} + (1 - \pi_{1\cdot})}, \quad \pi_{1\cdot} = \Pr(X_1 = 1) = \pi_{11} + \pi_{10}$$

$$LR(y_1, y_2) = \frac{\Pr(Y_1 = y_1, Y_2 = y_2 | X_1 = 1)}{\Pr(Y_1 = y_1, Y_2 = y_2 | X_1 = 0)}.$$

Consider the case $y_1 = 1, y_2 = 1$, and recall $e^\varepsilon = p/(1 - p)$

$$LR(1, 1) = \frac{e^\varepsilon \frac{\pi_{11}}{\pi_{1\cdot}} + \frac{\pi_{10}}{\pi_{1\cdot}}}{\frac{\pi_{01}}{\pi_{0\cdot}} + e^{-\varepsilon} \frac{\pi_{00}}{\pi_{0\cdot}}}$$

The dependence is a big trouble maker

This means that when $\pi_{10} = \pi_{01} = 0$, $LR(1,1) = e^{2\varepsilon} > e^{\varepsilon}$.

- But $\pi_{10} = \pi_{01} = 0$ means that $X_2 = X_1$, hence X_1 can be learned from the information for X_2. Consequently, the "individual information unit" for X_1 should be the pair $\{X_1, X_2\}$, not merely X_1.

The dependence is a big trouble maker

This means that when $\pi_{10} = \pi_{01} = 0$, $LR(1,1) = e^{2\varepsilon} > e^{\varepsilon}$.

- But $\pi_{10} = \pi_{01} = 0$ means that $X_2 = X_1$, hence X_1 can be learned from the information for X_2. Consequently, the "individual information unit" for X_1 should be the pair $\{X_1, X_2\}$, not merely X_1.
- In fact as soon as $\text{Cov}(X_1, X_2) > 0$, $LR(1,1) > e^{\varepsilon}$. This is because

$$LR(1,1) > e^{\varepsilon} \quad \Longleftrightarrow \quad \Pr(X_2 = 1 | X_1 = 1) > \Pr(X_2 = 1 | X_1 = 0)$$

But

$$\begin{aligned}\text{Cov}(X_1, X_2) &= \Pr(X_1 = 1, X_2 = 1) - \Pr(X_1 = 1)\Pr(X_2 = 1) \\ &= [\Pr(X_2 = 1 | X_1 = 1) - \Pr(X_2 = 1 | X_1 = 0)]\Pr(X_1 = 0)\Pr(X_1 = 1).\end{aligned}$$

Data are *accidental* representation, not *essential* information itself

Manipulating data values without considering their interdependence is not a legitimate information operation in general

In general, what does DP actual guarantee?

An attacker A is interested in learning about $\boldsymbol{X}_A = \{x_i, i \in I_A\}$ in a database $\boldsymbol{X} = \{X_i, i \in I\}$, where I_A could contain a single individual or everyone in I. Suppose the attacker has prior knowledge about the entire $\boldsymbol{X}$ in the form of $\pi(\boldsymbol{X})$.

In general, what does DP actual guarantee?

An attacker A is interested in learning about $\boldsymbol{X}_A = \{x_i, i \in I_A\}$ in a database $\boldsymbol{X} = \{X_i, i \in I\}$, where I_A could contain a single individual or everyone in I. Suppose the attacker has prior knowledge about the entire $\boldsymbol{X}$ in the form of $\pi(\boldsymbol{X})$.

Let $\pi_A(X_i)$ be the marginal prior, and $\pi_A(X_i|\boldsymbol{X}_{-i})$ be the conditional prior, conditioning on $\boldsymbol{X}_{-i} = \{X_j, j \neq i\}$. Upon learning $M = m$,

- Does ε-DP guarantees the marginal posterior-to-prior ratio

$$e^{-\varepsilon} \leq \frac{P_A(X_i = x|M = m)}{\pi_A(X_i = x)} \leq e^{\varepsilon}, \quad \forall x \in \mathcal{X}_i?$$ **No, not in general**

(Kifer & Machanavajjhala, 2011b, 2012; Tschantz et al., 2020)

In general, what does DP actual guarantee?

An attacker A is interested in learning about $\boldsymbol{X}_A = \{x_i, i \in I_A\}$ in a database $\boldsymbol{X} = \{X_i, i \in I\}$, where I_A could contain a single individual or everyone in I. Suppose the attacker has prior knowledge about the entire $\boldsymbol{X}$ in the form of $\pi(\boldsymbol{X})$.

Let $\pi_A(X_i)$ be the marginal prior, and $\pi_A(X_i|\boldsymbol{X}_{-i})$ be the conditional prior, conditioning on $\boldsymbol{X}_{-i} = \{X_j, j \neq i\}$. Upon learning $M = m$,

- Does ε-DP guarantees the marginal posterior-to-prior ratio

$$e^{-\varepsilon} \leq \frac{P_A(X_i = x|M = m)}{\pi_A(X_i = x)} \leq e^{\varepsilon}, \quad \forall x \in \mathcal{X}_i? \quad \textbf{No, not in general}$$

  (Kifer & Machanavajjhala, 2011b, 2012; Tschantz et al., 2020)

- Does ε-DP guarantees the conditional posterior-to-prior ratio

$$e^{-\varepsilon} \leq \frac{P_A(X_i = x|M = m, \boldsymbol{X}_{-i})}{\pi_A(X_i = x|\boldsymbol{X}_{-i})} \leq e^{\varepsilon}? \quad \forall x \in \mathcal{X}_i? \quad \textbf{Yes}$$

In general, what does DP actual guarantee?

An attacker A is interested in learning about $\boldsymbol{X}_A = \{x_i, i \in I_A\}$ in a database $\boldsymbol{X} = \{X_i, i \in I\}$, where I_A could contain a single individual or everyone in I. Suppose the attacker has prior knowledge about the entire $\boldsymbol{X}$ in the form of $\pi(\boldsymbol{X})$.

Let $\pi_A(X_i)$ be the marginal prior, and $\pi_A(X_i|\boldsymbol{X}_{-i})$ be the conditional prior, conditioning on $\boldsymbol{X}_{-i} = \{X_j, j \neq i\}$. Upon learning $M = m$,

- Does ε-DP guarantees the marginal posterior-to-prior ratio

$$e^{-\varepsilon} \leq \frac{P_A(X_i = x|M = m)}{\pi_A(X_i = x)} \leq e^{\varepsilon}, \quad \forall x \in \mathcal{X}_i?$$

No, not in general

(Kifer & Machanavajjhala, 2011b, 2012; Tschantz et al., 2020)

- Does ε-DP guarantees the conditional posterior-to-prior ratio

$$e^{-\varepsilon} \leq \frac{P_A(X_i = x|M = m, \boldsymbol{X}_{-i})}{\pi_A(X_i = x|\boldsymbol{X}_{-i})} \leq e^{\varepsilon}? \quad \forall x \in \mathcal{X}_i?$$

Yes

- Thus the guaranteed limit e^{ε} is only for the **unique individual information**: variations unexplained by anyone else in the database or by knowledge on (and beyond) the database population.

Theorem (Bailie, Gong & Meng, 2023)

A random map M delivers ε-DP under Hamming distance if and only if for every prior π on $\mathcal{D}$, every sub-σ field $\mathcal{F}$ of the corresponding full σ-field $\sigma_\pi(\mathcal{X})$, every $B \in \mathscr{B}\left(\mathbb{R}^d\right)$, every i, and every $A \in \mathscr{B}(\Theta_i)$, where Θ_i is the state space of x_i, we have

$$e^{-c_i\varepsilon}\pi(X_i \in A \mid \mathcal{F}) \leq \Pr\left(X_i \in A \mid M \in B; \mathcal{F}\right) \leq e^{c_i\varepsilon}\pi(x_i \in A \mid \mathcal{F}), \tag{2}$$

where $\pi(x_i|\mathcal{F})$ is the marginal prior for X_i (conditional on $\mathcal{F}$), $\Pr$ is the marginal posterior for X_i, and c_i is the size of the minimal information chamber (MIC) for X_i.

Theorem (Bailie, Gong & Meng, 2023)

A random map M delivers ε-DP under Hamming distance if and only if for every prior π on $\mathcal{D}$, every sub-σ field $\mathcal{F}$ of the corresponding full σ-field $\sigma_\pi(\mathcal{X})$, every $B \in \mathscr{B}\left(\mathbb{R}^d\right)$, every i, and every $A \in \mathscr{B}(\Theta_i)$, where Θ_i is the state space of $\mathbf{x}_i$, we have

$$e^{-c_i\varepsilon}\pi(X_i \in A \mid \mathcal{F}) \leq \Pr\left(X_i \in A \mid M \in B; \mathcal{F}\right) \leq e^{c_i\varepsilon}\pi(x_i \in A \mid \mathcal{F}), \tag{2}$$

where $\pi(x_i|\mathcal{F})$ is the marginal prior for X_i (conditional on $\mathcal{F}$), $\Pr$ *is the marginal posterior for X_i, and c_i is the size of the minimal information chamber (MIC) for X_i.*

- *MIC* $= C_{-i} \cup \{X_i\}$: $C_{-i} \subset \boldsymbol{X}_{-i}$ is the *Markov boundary* for X_i, that is, the smallest subset of $\boldsymbol{X}_{-i}$ such that
$$\pi(X_i|\boldsymbol{X}_{-i}, \mathcal{F}) = \pi(X_i|C_{-i}, \mathcal{F}).$$
- MIC is the X_i's "information family" – knowing any one of them will provide information about X_i, in addition to public knowledge coded into $\mathcal{F}$.

Theorem (Bailie, Gong & Meng, 2023)

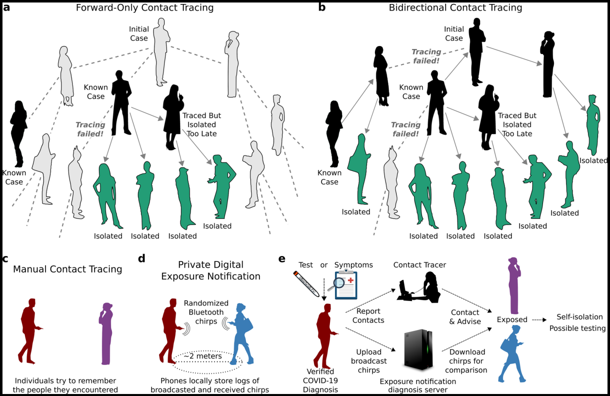
A random map M delivers ε-DP under Hamming distance if and only if for every prior π on $\mathcal{D}$, every sub-σ field $\mathcal{F}$ of the corresponding full σ-field $\sigma_\pi(\mathcal{X})$, every $B \in \mathscr{B}\left(\mathbb{R}^d\right)$, every i, and every $A \in \mathscr{B}(\Theta_i)$, where Θ_i is the state space of $\mathbf{x}_i$, we have

$$e^{-c_i\varepsilon}\pi(X_i \in A \mid \mathcal{F}) \leq \Pr\left(X_i \in A \mid M \in B; \mathcal{F}\right) \leq e^{c_i\varepsilon}\pi(x_i \in A \mid \mathcal{F}), \tag{2}$$

where $\pi(x_i|\mathcal{F})$ is the marginal prior for X_i (conditional on $\mathcal{F}$), Pr *is the marginal posterior for X_i, and c_i is the size of the minimal information chamber (MIC) for X_i.*

- *MIC* $= C_{-i} \cup \{X_i\}$: $C_{-i} \subset \boldsymbol{X}_{-i}$ is the *Markov boundary* for X_i, that is, the smallest subset of $\boldsymbol{X}_{-i}$ such that
$$\pi(X_i|\boldsymbol{X}_{-i}, \mathcal{F}) = \pi(X_i|C_{-i}, \mathcal{F}).$$
- MIC is the X_i's "information family" – knowing any one of them will provide information about X_i, in addition to public knowledge coded into $\mathcal{F}$.
- Protecting *relative* risk against "strong adversary" is the easiest — the more the adversary's prior information, the less left for protection.

Information spreads like a virus — we need to quarantine not only the infected individual but also everyone they've come into contact with.



Why is it called "Differential Privacy"?

Let the probability space for $M(\boldsymbol{X})$ be $\{\mathcal{M}, \mathcal{F}, P_{\boldsymbol{X}}\}$ (with $P_{\boldsymbol{X}}(S) = \Pr(M(\boldsymbol{X}) \in S|\boldsymbol{X})$)

Why is it called "Differential Privacy"?

Let the probability space for $M(\boldsymbol{X})$ be $\{\mathcal{M}, \mathcal{F}, P_{\boldsymbol{X}}\}$ (with $P_{\boldsymbol{X}}(S) = \Pr(M(\boldsymbol{X}) \in S|\boldsymbol{X})$)

"Differential" comes from "derivative", essential for studying *changes*

For log-likelihood $\ell(\boldsymbol{X}|S) = \ln \Pr(M(\boldsymbol{X}) \in S|\boldsymbol{X})$, pure DP is equivalent to requiring

$$\frac{\sup_{S\in\mathcal{F}} |\ell(\boldsymbol{X}|S) - \ell(\boldsymbol{X}'|S)|}{d_{\mathcal{X}}(\boldsymbol{X}, \boldsymbol{X}')} \leq \varepsilon, \quad \text{for all } \boldsymbol{X}, \boldsymbol{X}',$$

because "divergence" $d_{\mathcal{X}}(\boldsymbol{X}, \boldsymbol{X}') = 1$ for "neighboring" pair $\{\boldsymbol{X}, \boldsymbol{X}'\}$.

Why is it called "Differential Privacy"?

Let the probability space for $M(\boldsymbol{X})$ be $\{\mathcal{M}, \mathcal{F}, P_{\boldsymbol{X}}\}$ (with $P_{\boldsymbol{X}}(S) = \Pr(M(\boldsymbol{X}) \in S|\boldsymbol{X})$)

"Differential" comes from "derivative", essential for studying *changes*

For log-likelihood $\ell(\boldsymbol{X}|S) = \ln \Pr(M(\boldsymbol{X}) \in S|\boldsymbol{X})$, pure DP is equivalent to requiring

$$\frac{\sup_{S\in\mathcal{F}} |\ell(\boldsymbol{X}|S) - \ell(\boldsymbol{X}'|S)|}{d_{\mathcal{X}}(\boldsymbol{X}, \boldsymbol{X}')} \le \varepsilon, \quad \text{for all } \boldsymbol{X}, \boldsymbol{X}',$$

because "divergence" $d_{\mathcal{X}}(\boldsymbol{X}, \boldsymbol{X}') = 1$ for "neighboring" pair $\{\boldsymbol{X}, \boldsymbol{X}'\}$.

A general DP Specification (Bailie et al., 2023)

A data-release mechanism $M : \mathcal{X} \to \mathcal{M}$ satisfies a *DP specification* $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, d_{\Pr}, \varepsilon_{\mathcal{D}})$ if

$$d_{\Pr}\left[P_{\boldsymbol{X}}, P_{\boldsymbol{X}'}\right] \le \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{X}, \boldsymbol{X}'), \tag{3}$$

for all $\boldsymbol{X}, \boldsymbol{X}'$ in every data universe $\mathcal{D}$ in the data multiverse $\mathscr{D}$.

Five Building Blocks

A general DP Specification (Bailie et al., 2023)

A data-release mechanism $M : \mathcal{X} \to \mathcal{M}$ satisfies a *DP specification* $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, d_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}})$ if

$$d_{\mathrm{Pr}}\left[P_{\boldsymbol{X}}, P_{\boldsymbol{X}'}\right] \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{X}, \boldsymbol{X}'), \tag{4}$$

for all $\boldsymbol{X}, \boldsymbol{X}'$ in every data universe $\mathcal{D}$ in the data multiverse $\mathscr{D}$.

Five Building Blocks

A general DP Specification (Bailie et al., 2023)

A data-release mechanism $M : \mathcal{X} \to \mathcal{M}$ satisfies a *DP specification* $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, d_{\mathsf{Pr}}, \varepsilon_{\mathcal{D}})$ if

$$d_{\mathsf{Pr}}\left[P_{\boldsymbol{X}}, P_{\boldsymbol{X}'}\right] \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{X}, \boldsymbol{X}'), \tag{4}$$

for all $\boldsymbol{X}, \boldsymbol{X}'$ in every data universe $\mathcal{D}$ in the data multiverse $\mathscr{D}$.

- The **protection domain** (*what* can be protected?): dataset space $\mathcal{X}$;

Five Building Blocks

A general DP Specification (Bailie et al., 2023)

A data-release mechanism $M : \mathcal{X} \to \mathcal{M}$ satisfies a *DP specification* $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, d_{\mathsf{Pr}}, \varepsilon_{\mathcal{D}})$ if

$$d_{\mathsf{Pr}}\left[P_{\boldsymbol{X}}, P_{\boldsymbol{X}'}\right] \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{X}, \boldsymbol{X}'), \tag{4}$$

for all $\boldsymbol{X}, \boldsymbol{X}'$ in every data universe $\mathcal{D}$ in the data multiverse $\mathscr{D}$.

- The **protection domain** (*what* can be protected?): dataset space $\mathcal{X}$;
- The **scope of protection** (*to where* does the protection extend?): data multiverse $\mathscr{D}$ (*essential*), a collection of data universes $\mathcal{D} \subset \mathcal{X}$ (*accidental*);

Five Building Blocks

A general DP Specification (Bailie et al., 2023)

A data-release mechanism $M : \mathcal{X} \to \mathcal{M}$ satisfies a *DP specification* $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, d_{\mathsf{Pr}}, \varepsilon_{\mathcal{D}})$ if

$$d_{\mathsf{Pr}}\left[P_{\boldsymbol{X}}, P_{\boldsymbol{X}'}\right] \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{X}, \boldsymbol{X}'), \tag{4}$$

for all $\boldsymbol{X}, \boldsymbol{X}'$ in every data universe $\mathcal{D}$ in the data multiverse $\mathscr{D}$.

- The **protection domain** (*what* can be protected?): dataset space $\mathcal{X}$;
- The **scope of protection** (*to where* does the protection extend?): data multiverse $\mathscr{D}$ (*essential*), a collection of data universes $\mathcal{D} \subset \mathcal{X}$ (*accidental*);
- The **protection units** (*who* are the units of protection): the input divergence $d_{\mathcal{X}}$ on $\mathcal{X}$;

Five Building Blocks

A general DP Specification (Bailie et al., 2023)

A data-release mechanism $M : \mathcal{X} \to \mathcal{M}$ satisfies a *DP specification* $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, d_{\mathrm{Pr}}, \varepsilon_{\mathcal{D}})$ if

$$d_{\mathrm{Pr}}\left[P_{\boldsymbol{X}}, P_{\boldsymbol{X}'}\right] \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{X}, \boldsymbol{X}'), \tag{4}$$

for all $\boldsymbol{X}, \boldsymbol{X}'$ in every data universe $\mathcal{D}$ in the data multiverse $\mathscr{D}$.

- The **protection domain** (*what* can be protected?): dataset space $\mathcal{X}$;
- The **scope of protection** (*to where* does the protection extend?): data multiverse $\mathscr{D}$ (*essential*), a collection of data universes $\mathcal{D} \subset \mathcal{X}$ (*accidental*);
- The **protection units** (*who* are the units of protection): the input divergence $d_{\mathcal{X}}$ on $\mathcal{X}$;
- The **standard of protection** (*how* to measure protection): the divergence d_{Pr} on probabilities;

Five Building Blocks

A general DP Specification (Bailie et al., 2023)

A data-release mechanism $M : \mathcal{X} \to \mathcal{M}$ satisfies a *DP specification* $(\mathcal{X}, \mathscr{D}, d_{\mathcal{X}}, d_{\mathsf{Pr}}, \varepsilon_{\mathcal{D}})$ if

$$d_{\mathsf{Pr}}\left[P_{\boldsymbol{X}}, P_{\boldsymbol{X}'}\right] \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\boldsymbol{X}, \boldsymbol{X}'), \tag{4}$$

for all $\boldsymbol{X}, \boldsymbol{X}'$ in every data universe $\mathcal{D}$ in the data multiverse $\mathscr{D}$.

- The **protection domain** (*what* can be protected?): dataset space $\mathcal{X}$;
- The **scope of protection** (*to where* does the protection extend?): data multiverse $\mathscr{D}$ (*essential*), a collection of data universes $\mathcal{D} \subset \mathcal{X}$ (*accidental*);
- The **protection units** (*who* are the units of protection): the input divergence $d_{\mathcal{X}}$ on $\mathcal{X}$;
- The **standard of protection** (*how* to measure protection): the divergence d_{Pr} on probabilities;
- The **intensity of protection** (*how much* protection is afforded): privacy loss budget $\varepsilon_{\mathcal{D}} \in \mathbb{R}^{\geq 0}$, for each data universe $\mathcal{D}$.

Examples in the Literature

4. d_{Pr}: (ε, δ)-approximate DP (Dwork, Kenthapadi, et al., 2006) Rényi DP (Mironov, 2017) concentrated DP (Bun & Steinke, 2016) f-divergence privacy (Barber & Duchi, 2014; Barthe & Olmedo, 2013) f-DP (including Gaussian DP) (Dong et al., 2022).

Examples in the Literature

4. d_{Pr}: (ε, δ)-approximate DP (Dwork, Kenthapadi, et al., 2006) Rényi DP (Mironov, 2017) concentrated DP (Bun & Steinke, 2016) f-divergence privacy (Barber & Duchi, 2014; Barthe & Olmedo, 2013) f-DP (including Gaussian DP) (Dong et al., 2022).

3. $d_{\mathcal{X}}$: $(\mathcal{R}, \varepsilon)$-generic DP (Kifer & Machanavajjhala, 2011a) edge vs node privacy (Hay et al., 2009; McSherry & Mahajan, 2010) d-metric DP (Chatzikokolakis et al., 2013) Blowfish privacy (He et al., 2014) element level DP (Asi et al., 2022) distributional privacy (Zhou et al., 2009) event-level vs user-level DP (Dwork et al., 2010).

Examples in the Literature

4. d_{Pr}: (ε, δ)-approximate DP (Dwork, Kenthapadi, et al., 2006) Rényi DP (Mironov, 2017) concentrated DP (Bun & Steinke, 2016) f-divergence privacy (Barber & Duchi, 2014; Barthe & Olmedo, 2013) f-DP (including Gaussian DP) (Dong et al., 2022).

3. $d_{\mathcal{X}}$: $(\mathcal{R}, \varepsilon)$-generic DP (Kifer & Machanavajjhala, 2011a) edge vs node privacy (Hay et al., 2009; McSherry & Mahajan, 2010) d-metric DP (Chatzikokolakis et al., 2013) Blowfish privacy (He et al., 2014) element level DP (Asi et al., 2022) distributional privacy (Zhou et al., 2009) event-level vs user-level DP (Dwork et al., 2010).

2. $\mathscr{D}$: privacy under invariants (Ashmead et al., 2019; Gong & Meng, 2020; Gao et al., 2022; Dharangutte et al., 2023) conditioned or empirical DP (J. M. Abowd et al., 2013; Charest & Hou, 2016) personalized DP (Ebadi et al., 2015; Jorgensen et al., 2015) individual DP (Soria-Comas et al., 2017; Feldman & Zrnic, 2022) bootstrap DP (O'Keefe & Charest, 2019) stratified DP (Bun et al., 2022) per-record DP (Seeman et al., 2023+) per-instance DP (Wang, 2018; Redberg & Wang, 2021).

Examples in the Literature

4. d_{Pr}: (ε, δ)-approximate DP (Dwork, Kenthapadi, et al., 2006) Rényi DP (Mironov, 2017) concentrated DP (Bun & Steinke, 2016) f-divergence privacy (Barber & Duchi, 2014; Barthe & Olmedo, 2013) f-DP (including Gaussian DP) (Dong et al., 2022).

3. $d_{\mathcal{X}}$: $(\mathcal{R}, \varepsilon)$-generic DP (Kifer & Machanavajjhala, 2011a) edge vs node privacy (Hay et al., 2009; McSherry & Mahajan, 2010) d-metric DP (Chatzikokolakis et al., 2013) Blowfish privacy (He et al., 2014) element level DP (Asi et al., 2022) distributional privacy (Zhou et al., 2009) event-level vs user-level DP (Dwork et al., 2010).

2. $\mathscr{D}$: privacy under invariants (Ashmead et al., 2019; Gong & Meng, 2020; Gao et al., 2022; Dharangutte et al., 2023) conditioned or empirical DP (J. M. Abowd et al., 2013; Charest & Hou, 2016) personalized DP (Ebadi et al., 2015; Jorgensen et al., 2015) individual DP (Soria-Comas et al., 2017; Feldman & Zrnic, 2022) bootstrap DP (O'Keefe & Charest, 2019) stratified DP (Bun et al., 2022) per-record DP (Seeman et al., 2023+) per-instance DP (Wang, 2018; Redberg & Wang, 2021).

1. $\mathscr{X}$: DP for network data (Hay et al., 2009) for geospatial data (Andrés et al., 2013) Pufferfish DP (Kifer & Machanavajjhala, 2014) noiseless privacy (Bhaskar et al., 2011) privacy under partial knowledge (Seeman et al., 2022) privacy amplification (Beimel et al., 2010; Balle et al., 2020; Bun et al., 2022).

Examples from the US Decennial Censuses

	d_{Pr}	$d_{\mathcal{X}}$ (Unit)	Invariants	Privacy Loss Budget
TopDown*	D_{nor}	d^{p}_{Ham} (person)	Population (state) Total housing units (block) Occupied group quarters (block) Structural zeros	PL & DHC: $\rho^2 = 15.29$ $\varepsilon = 52.83$ ($\delta = 10^{-10}$)
SafeTab**	D_{nor}	d^{p}_{Ham} (person)	None	DDHC-A: $\rho^2 = 19.776$ DDHC-B & S-DHC: *TBD.*
Swapping	$d_{\textsc{Mult}}$	d^{h}_{Ham} (household)	Varies but greater than TDA	ε between 9.37-19.38

*(J. Abowd et al., 2022)

**(Tumult Labs, 2022)

- $\mathcal{X}$ is always the space of possible Census Edited Files, $\mathcal{X}_{\text{CEF}}$.
- $D_{\text{nor}}(P, Q) = \sup_{\alpha > 1} \frac{1}{\sqrt{\alpha}} \max \left[\sqrt{D_\alpha(P||Q)}, \sqrt{D_\alpha(Q||P)} \right]$ is the normalised Rényi metric [zero concentrated DP] (with D_α the Rényi divergence of order);
- $d_{\textsc{Mult}}(P, Q) = \sup_{S \in \mathcal{F}} \left| \ln \frac{P(S)}{Q(S)} \right|$ is the multiplicative distance (pure DP); and
- d^{u}_{Ham} is the Hamming distance (on units u).

Swapping Satisfies DP, Subject to its Invariants

Permutation Swapping

Input: a dataset $\boldsymbol{x}$.
Define strata as groups of records which match on the swap key $\mathbf{V}_{\text{Stratify}}$.
Within each stratum:

1. Select each record independently with probability p (the swap rate).
2. Randomly derange swapping variable $\mathbf{V}_{\text{Swap}}$ of selected records.

Output: the *swapped* dataset $\boldsymbol{w}$.

Swapping Satisfies DP, Subject to its Invariants

Permutation Swapping

Input: a dataset $\boldsymbol{x}$.
Define strata as groups of records which match on the swap key $\mathbf{V}_{\text{Stratify}}$.
Within each stratum:

1. Select each record independently with probability p (the swap rate).
2. Randomly derange swapping variable $\mathbf{V}_{\text{Swap}}$ of selected records.

Output: the *swapped* dataset $\boldsymbol{w}$.

Permutation Swapping is DP subject to its invariants, with input divergence $d_{\mathcal{X}} = d^{u}_{\text{Ham}}$, output divergence $d_{\mathsf{Pr}} = d_{\text{MULT}}$ and budget

$$\varepsilon = \begin{cases} \ln(b+1) - \ln o & \text{if } 0 < p \leq 0.5, \\ \max\left\{\ln o, \ln(b+1) - \ln o\right\} & \text{if } 0.5 < p < 1, \end{cases}$$

where $o = p/(1-p)$ and b is the maximum stratum size.

The TopDown Algorithm (TDA) (J. Abowd et al., 2022)

Two-step procedure:

0. Start with a Census edited file $\boldsymbol{x} \in \mathcal{X}_{\mathrm{CEF}}$.

The TopDown Algorithm (TDA) (J. Abowd et al., 2022)

Two-step procedure:

0. Start with a Census edited file $\boldsymbol{x} \in \mathcal{X}_{\mathrm{CEF}}$.
1. Add Gaussian noise to cells:

$$\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{q}(\boldsymbol{x}) + \boldsymbol{W},$$

where $\boldsymbol{W} \sim \mathcal{N}_{\mathbb{Z}}(0, \boldsymbol{\Sigma})$, so that $\boldsymbol{T}$ satisfies $\mathrm{DP}(\mathcal{X}_{\mathrm{CEF}}, \{\mathcal{X}_{\mathrm{CEF}}\}, d^{p}_{\mathrm{Ham}}, D_{\mathrm{nor}})$ with budget ρ_{TDA} (Canonne et al., 2022).

The TopDown Algorithm (TDA) (J. Abowd et al., 2022)

Two-step procedure:

0. Start with a Census edited file $\boldsymbol{x} \in \mathcal{X}_{\text{CEF}}$.
1. Add Gaussian noise to cells:

$$\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{q}(\boldsymbol{x}) + \boldsymbol{W},$$

   where $\boldsymbol{W} \sim \mathcal{N}_{\mathbb{Z}}(0, \boldsymbol{\Sigma})$, so that $\boldsymbol{T}$ satisfies $\text{DP}(\mathcal{X}_{\text{CEF}}, \{\mathcal{X}_{\text{CEF}}\}, d^p_{\text{Ham}}, D_{\text{nor}})$ with budget ρ_{TDA} (Canonne et al., 2022).

2. "Post-process": find dataset $\boldsymbol{z}$ with $\boldsymbol{q}(\boldsymbol{z})$ close to $\boldsymbol{T}(\boldsymbol{x})$ such that $\boldsymbol{c}_{\text{TDA}}(\boldsymbol{z}) = \boldsymbol{c}_{\text{TDA}}(\boldsymbol{x})$.

The TopDown Algorithm (TDA) (J. Abowd et al., 2022)

Two-step procedure:

0. Start with a Census edited file $\boldsymbol{x} \in \mathcal{X}_{\text{CEF}}$.
1. Add Gaussian noise to cells:

$$\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{q}(\boldsymbol{x}) + \boldsymbol{W},$$

   where $\boldsymbol{W} \sim \mathcal{N}_{\mathbb{Z}}(0, \boldsymbol{\Sigma})$, so that $\boldsymbol{T}$ satisfies $\mathrm{DP}(\mathcal{X}_{\text{CEF}}, \{\mathcal{X}_{\text{CEF}}\}, d^{p}_{\text{Ham}}, D_{\text{nor}})$ with budget ρ_{TDA} (Canonne et al., 2022).

2. "Post-process": find dataset $\boldsymbol{z}$ with $\boldsymbol{q}(\boldsymbol{z})$ close to $\boldsymbol{T}(\boldsymbol{x})$ such that $\boldsymbol{c}_{\text{TDA}}(\boldsymbol{z}) = \boldsymbol{c}_{\text{TDA}}(\boldsymbol{x})$.

The TopDown Algorithm (TDA) (J. Abowd et al., 2022)

Two-step procedure:

0. Start with a Census edited file $\boldsymbol{x} \in \mathcal{X}_{\text{CEF}}$.
1. Add Gaussian noise to cells:

$$\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{q}(\boldsymbol{x}) + \boldsymbol{W},$$

   where $\boldsymbol{W} \sim \mathcal{N}_{\mathbb{Z}}(0, \boldsymbol{\Sigma})$, so that $\boldsymbol{T}$ satisfies $\text{DP}(\mathcal{X}_{\text{CEF}}, \{\mathcal{X}_{\text{CEF}}\}, d^p_{\text{Ham}}, D_{\text{nor}})$ with budget ρ_{TDA} (Canonne et al., 2022).
2. "Post-process": find dataset $\boldsymbol{z}$ with $\boldsymbol{q}(\boldsymbol{z})$ close to $\boldsymbol{T}(\boldsymbol{x})$ such that $\boldsymbol{c}_{\text{TDA}}(\boldsymbol{z}) = \boldsymbol{c}_{\text{TDA}}(\boldsymbol{x})$.

TDA satisfies $\text{DP}(\mathcal{X}_{\text{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\text{TDA}}}, d^p_{\text{Ham}}, D_{\text{nor}})$ with budget ρ_{TDA}.

Theorem: TDA Satisfies DP, Subject to its Invariants

Let $\boldsymbol{c}_{\mathrm{TDA}} : \mathcal{X}_{\mathrm{CEF}} \to \mathbb{R}^l$ be the invariants of TDA and let $\mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}}$ be the induced data multiverse:

$$\mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}} = \{\mathcal{D} \subset \mathcal{X}_{\mathrm{CEF}} \mid \boldsymbol{c}_{\mathrm{TDA}}(\boldsymbol{x}) = \boldsymbol{c}_{\mathrm{TDA}}(\boldsymbol{x}') \; \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}\}.$$

Theorem: TDA Satisfies DP, Subject to its Invariants

Let $\boldsymbol{c}_{\mathrm{TDA}} : \mathcal{X}_{\mathrm{CEF}} \to \mathbb{R}^l$ be the invariants of TDA and let $\mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}}$ be the induced data multiverse:

$$\mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}} = \{\mathcal{D} \subset \mathcal{X}_{\mathrm{CEF}} \mid \boldsymbol{c}_{\mathrm{TDA}}(\boldsymbol{x}) = \boldsymbol{c}_{\mathrm{TDA}}(\boldsymbol{x}') \; \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}\}.$$

- TDA satisfies DP$(\mathcal{X}_{\mathrm{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\mathrm{TDA}}}, d^p_{\mathrm{Ham}}, D_{\mathrm{nor}})$ with privacy budget $\rho_{\mathrm{TDA}} = 2.63$ (for the PL Redistricting File) and $\rho_{\mathrm{TDA}} = 15.29$ (for the DHC).

Theorem: TDA Satisfies DP, Subject to its Invariants

Let $\boldsymbol{c}_{\text{TDA}} : \mathcal{X}_{\text{CEF}} \to \mathbb{R}^l$ be the invariants of TDA and let $\mathscr{D}_{\boldsymbol{c}_{\text{TDA}}}$ be the induced data multiverse:

$$\mathscr{D}_{\boldsymbol{c}_{\text{TDA}}} = \{\mathcal{D} \subset \mathcal{X}_{\text{CEF}} \mid \boldsymbol{c}_{\text{TDA}}(\boldsymbol{x}) = \boldsymbol{c}_{\text{TDA}}(\boldsymbol{x}') \; \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}\}.$$

- TDA satisfies $\text{DP}(\mathcal{X}_{\text{CEF}}, \mathscr{D}_{\boldsymbol{c}_{\text{TDA}}}, d^p_{\text{Ham}}, D_{\text{nor}})$ with privacy budget $\rho_{\text{TDA}} = 2.63$ (for the PL Redistricting File) and $\rho_{\text{TDA}} = 15.29$ (for the DHC).
- Let $\boldsymbol{c}'$ be any proper subset of TDA's invariants. TDA does not satisfy $\text{DP}(\mathcal{X}_{\text{CEF}}, \mathscr{D}_{\boldsymbol{c}'}, d_{\mathcal{X}}, D_{\text{nor}})$ with any finite budget ρ.

References I

Abowd, J., Ashmead, R., Cumings-Menon, R., Garfinkel, S., Heineck, M., Heiss, C., … Zhuravlev, P. (2022, June). The 2020 Census disclosure avoidance system TopDown algorithm. *Harvard Data Science Review*(Special Issue 2). doi: 10.1162/99608f92.529e3cb9

Abowd, J. M., Schneider, M. J., & Vilhuber, L. (2013). Differential privacy applications to Bayesian and linear mixed model estimation. *Journal of Privacy and Confidentiality*, *5*(1).

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*, *54*(2), 442–92.

Andrés, M. E., Bordenabe, N. E., Chatzikokolakis, K., & Palamidessi, C. (2013, November). Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security* (pp. 901–914). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/2508859.2516735

References II

Ashmead, R., Kifer, D., Leclerc, P., Machanavajjhala, A., & Sexton, W. (2019). *Effective privacy after adjusting for invariants with applications to the 2020 Census* (Tech. Rep.). https://github.com/uscensusbureau/census2020-das-e2e/blob/master/doc/20190711_0941_Effective_Privacy_after_Adjusting_for_Constraints__With_applications_to_the_2020_Census.pdf.

Asi, H., Duchi, J. C., & Javidbakht, O. (2022). Element level differential privacy: The right granularity of privacy. In *AAAI Workshop on Privacy-Preserving Artificial Intelligence.* Association for the Advancement of Artificial Intelligence.

Bailie, J., Gong, R., & Meng, X.-L. (2023). Can swapping be differentially private? A refreshment stirred, not shaken. *In preparation for Harvard Data Science Review.*

Balle, B., Barthe, G., & Gaboardi, M. (2020, January). Privacy profiles and amplification by subsampling. *Journal of Privacy and Confidentiality*, *10*(1). doi: 10.29012/jpc.726

References III

Barber, R. F., & Duchi, J. C. (2014, December). *Privacy and statistical risk: Formalisms and minimax bounds* (No. arXiv:1412.4451). `http://arxiv.org/abs/1412.4451`. arXiv. doi: 10.48550/arXiv.1412.4451

Barthe, G., & Olmedo, F. (2013). Beyond differential privacy: Composition theorems and relational logic for f-divergences between probabilistic programs. In F. V. Fomin, R. Freivalds, M. Kwiatkowska, & D. Peleg (Eds.), *Automata, languages, and programming* (pp. 49–60). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-39212-2_8

Beimel, A., Kasiviswanathan, S. P., & Nissim, K. (2010, February). Bounds on the sample complexity for private learning and private data release. In D. Micciancio (Ed.), *Proceedings of the 7th theory of cryptography conference, TCC 2010, Zurich, Switzerland* (pp. 437–454). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-11799-2_26

References IV

Bhaskar, R., Bhowmick, A., Goyal, V., Laxman, S., & Thakurta, A. (2011). Noiseless database privacy. In D. H. Lee & X. Wang (Eds.), *Advances in cryptology – ASIACRYPT 2011* (pp. 215–232). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-25385-0_12

Bun, M., Drechsler, J., Gaboardi, M., McMillan, A., & Sarathy, J. (2022, June). Controlling privacy loss in sampling schemes: An analysis of stratified and cluster sampling. In *Foundations of Responsible Computing (FORC 2022)* (p. 24).

Bun, M., & Steinke, T. (2016). Concentrated differential privacy: Simplifications, extensions, and lower bounds. In M. Hirt & A. Smith (Eds.), *Theory of cryptography* (pp. 635–658). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-662-53641-4_24

Canonne, C., Kamath, G., & Steinke, T. (2022, July). The discrete Gaussian for differential privacy. *Journal of Privacy and Confidentiality*, *12*(1). doi: 10.29012/jpc.784

References V

Charest, A.-S., & Hou, Y. (2016). On the meaning and limits of empirical differential privacy. *Journal of Privacy and Confidentiality*, *7*(3), 53–66.

Chatzikokolakis, K., Andrés, M. E., Bordenabe, N. E., & Palamidessi, C. (2013). Broadening the Scope of Differential Privacy Using Metrics. In E. De Cristofaro & M. Wright (Eds.), *Privacy Enhancing Technologies* (pp. 82–102). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-39077-7_5

Dharangutte, P., Gao, J., Gong, R., & Yu, F.-Y. (2023). Integer subspace differential privacy. In *Proceedings of the aaai conference on artificial intelligence (aaai-23).*

Dong, J., Roth, A., & Su, W. J. (2022). Gaussian differential privacy. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *84*(1), 3–37. doi: 10.1111/rssb.12454

References VI

Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., & Naor, M. (2006). Our data, ourselves: Privacy via distributed noise generation. In S. Vaudenay (Ed.), *Advances in cryptology - EUROCRYPT 2006* (pp. 486–503). Berlin, Heidelberg: Springer. doi: 10.1007/11761679_29

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference* (pp. 265–284).

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2016). Calibrating noise to sensitivity in private data analysis. *Journal of Privacy and Confidentiality*, *7*(3), 17–51.

References VII

Dwork, C., Naor, M., Pitassi, T., & Rothblum, G. N. (2010, June). Differential privacy under continual observation. In *Proceedings of the forty-second ACM symposium on Theory of computing* (pp. 715–724). New York, NY, USA: Association for Computing Machinery. (https://dl.acm.org/doi/10.1145/1806689.1806787) doi: 10.1145/1806689.1806787

Ebadi, H., Sands, D., & Schneider, G. (2015, January). Differential Privacy: Now it's Getting Personal. *ACM SIGPLAN Notices*, *50*(1), 69–81. doi: 10.1145/2775051.2677005

Feldman, V., & Zrnic, T. (2022, January). *Individual privacy accounting via a Rényi filter* (No. arXiv:2008.11193). http://arxiv.org/abs/2008.11193. arXiv.

References VIII

Gao, J., Gong, R., & Yu, F.-Y. (2022, June). Subspace differential privacy. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 36, pp. 3986–3995). doi: 10.1609/aaai.v36i4.20315

Gong, R., & Meng, X.-L. (2020). Congenial differential privacy under mandated disclosure. In *Proceedings of the 2020 acm-ims on foundations of data science conference* (pp. 59–70).

Hay, M., Li, C., Miklau, G., & Jensen, D. (2009, December). Accurate estimation of the degree distribution of private networks. In *2009 Ninth IEEE International Conference on Data Mining* (pp. 169–178). doi: 10.1109/ICDM.2009.11

He, X., Machanavajjhala, A., & Ding, B. (2014). Blowfish privacy: Tuning privacy-utility trade-offs using policies. In *Proceedings of the 2014 acm sigmod international conference on management of data* (pp. 1447–1458).

References IX

Jorgensen, Z., Yu, T., & Cormode, G. (2015, April). Conservative or liberal? Personalized differential privacy. In *2015 IEEE 31st International Conference on Data Engineering* (pp. 1023–1034). (`https://ieeexplore.ieee.org/document/7113353`) doi: 10.1109/ICDE.2015.7113353

Kifer, D., & Machanavajjhala, A. (2011a). No free lunch in data privacy. In *Proceedings of the 2011 international conference on Management of data - SIGMOD '11* (pp. 193–204). Athens, Greece: ACM Press. doi: 10.1145/1989323.1989345

Kifer, D., & Machanavajjhala, A. (2011b). No free lunch in data privacy. In *Proceedings of the 2011 acm sigmod international conference on management of data* (pp. 193–204).

Kifer, D., & Machanavajjhala, A. (2012). A rigorous and customizable framework for privacy. In *Proceedings of the 31st acm sigmod-sigact-sigai symposium on principles of database systems* (pp. 77–88).

References X

Kifer, D., & Machanavajjhala, A. (2014). Pufferfish: A framework for mathematical privacy definitions. *ACM Transactions on Database Systems (TODS)*, *39*(1), 1–36.

McSherry, F., & Mahajan, R. (2010, August). Differentially-private network trace analysis. In *Proceedings of the ACM SIGCOMM 2010 conference* (pp. 123–134). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/1851182.1851199

Mironov, I. (2017, August). Rényi differential privacy. *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, 263–275. doi: 10.1109/CSF.2017.11

O'Keefe, C. M., & Charest, A.-S. (2019). Bootstrap differential privacy. *Transactions on Data Privacy*, *12*, 1–28.

Raab, C. (2019). Political science and privacy. *The handbook of privacy studies: An interdisciplinary introduction*, 257.

References XI

Redberg, R., & Wang, Y.-X. (2021). Privately publishable per-instance privacy. In *Advances in Neural Information Processing Systems* (Vol. 34, pp. 17335–17346). Curran Associates, Inc.

Seeman, J., Reimherr, M., & Slavkovic, A. (2022, May). *Formal privacy for partially private data* (No. arXiv:2204.01102). `http://arxiv.org/abs/2204.01102`. arXiv.

Seeman, J., Sexton, W., Pujol, D., & Machanavajjhala, A. (2023+). Per-record differential privacy: Modeling dependence between individual privacy loss and confidential records.

Solove, D. J. (2008). *Understanding Privacy*. Cambridge, MA: Harvard University Press.

References XII

Soria-Comas, J., Domingo-Ferrer, J., Sánchez, D., & Megías, D. (2017, June). Individual differential privacy: A utility-preserving formulation of differential privacy guarantees. *IEEE Transactions on Information Forensics and Security*, *12*(6), 1418–1429. doi: 10.1109/TIFS.2017.2663337

Tschantz, M. C., Sen, S., & Datta, A. (2020). SoK: Differential privacy as a causal property. In *2020 ieee symposium on security and privacy (sp)* (pp. 354–371).

Tumult Labs. (2022, March). *SafeTab: DP algorithms for 2020 Census Detailed DHC Race & Ethnicity* (Tech. Rep.). https://www2.census.gov/about/partners/cac/sac/meetings/2022-03/dhc-attachment-1-safetab-dp-algorithms.pdf.

Wang, Y.-X. (2018, November). *Per-instance Differential Privacy* (No. arXiv:1707.07708). `http://arxiv.org/abs/1707.07708`. arXiv.

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

References XIII

Warren, S., & Brandeis, L. (1890). The right to privacy. *Harvard Law Review*, *4*(5), 193–220.

Zhou, S., Ligett, K., & Wasserman, L. (2009, June). Differential privacy with compression. In *Proceedings of the 2009 IEEE international conference on Symposium on Information Theory - Volume 4* (pp. 2718–2722). Coex, Seoul, Korea: IEEE Press.