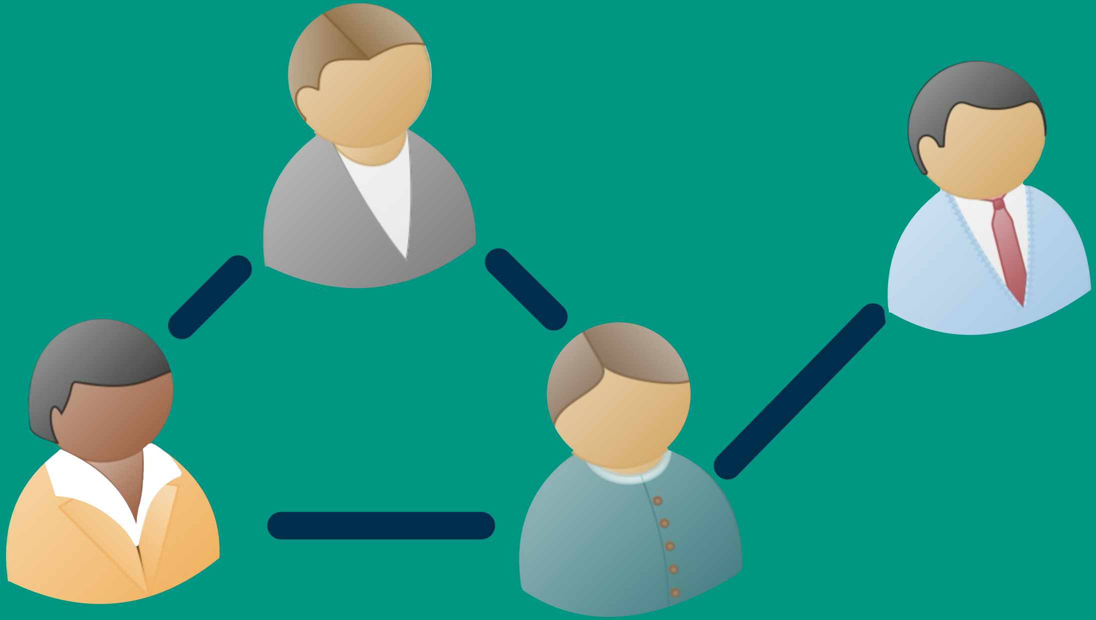


From DP to BDP: Noise Recalibration for Correlation-Resilient Privacy Guarantees

XIX RECSI, March 18th, 2026

Patricia Guerra-Balboa, Martin Lange, Javier Parra-Arnau, Thorsten Strufe

PRIVACY AND SECURITY

KASTEL

Motivation

Differential Privacy fails to measure privacy leakage under correlation

Theoretically exposed

Empirically confirmed

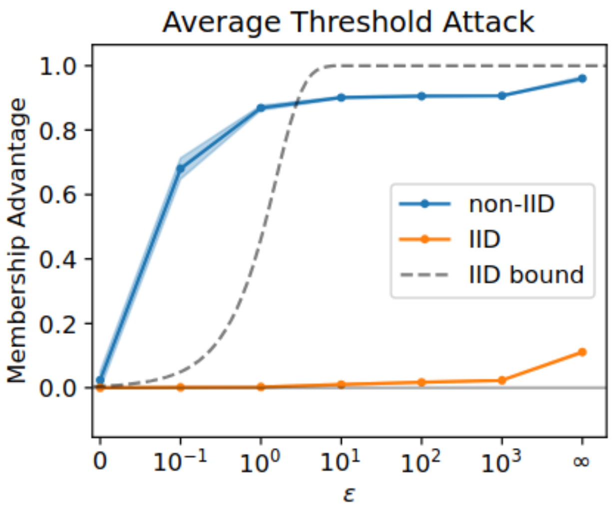


Figure: Humphries et al. 2023 MIA attack breaks DP guarantees.

Motivation

Differential Privacy fails to measure privacy leakage under correlation

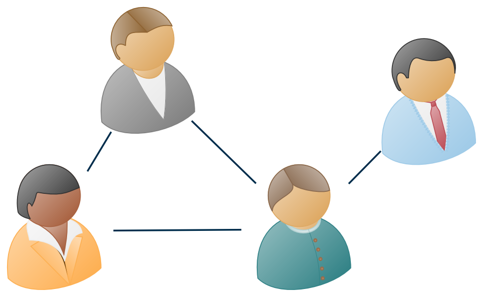
Theoretically exposed

Empirically confirmed

Dependencies among data records are present **in most of real world scenarios**.

Motivation

Dependencies among data records are present **in most of real world scenarios**.



Differential Privacy fails to measure privacy leakage under correlation

Theoretically exposed

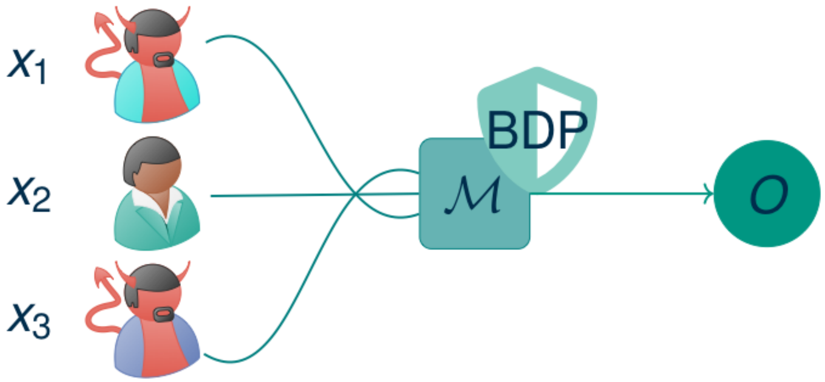
Empirically confirmed

New enhanced notion: Bayesian Differential Privacy

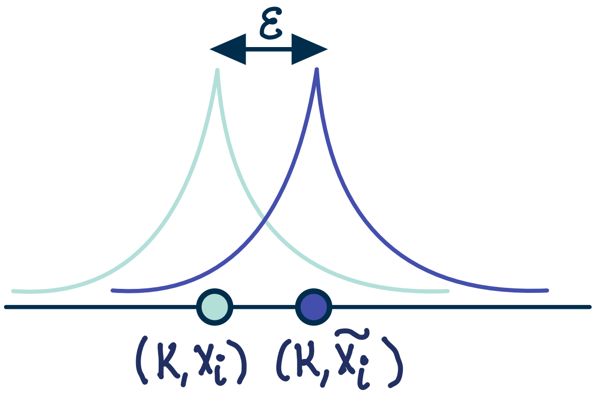
Bayesian Differential Privacy (BDP)

Bayesian DP leakage

$$\mathrm{BDPL}_{(K,i)} = \sup_{x_i, x_i', \mathbf{x}_K, S} \ln \frac{\Pr_{\mathcal{M}}[Y \in S \mid \mathbf{X}_K = \mathbf{x}_K, X_i = x_i]}{\Pr_{\mathcal{M}}[Y \in S \mid \mathbf{X}_K = \mathbf{x}_K, X_i = x_i']}, \text{ then } \varepsilon = \sup_{K,i} \mathrm{BDPL}_{(K,i)}.$$



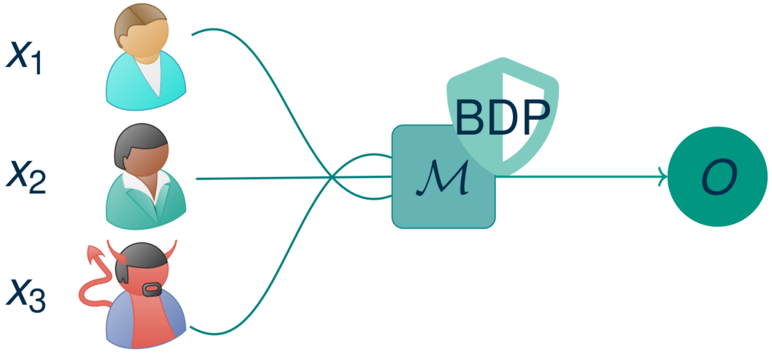
$K = \{1, 3\}, i = 2$



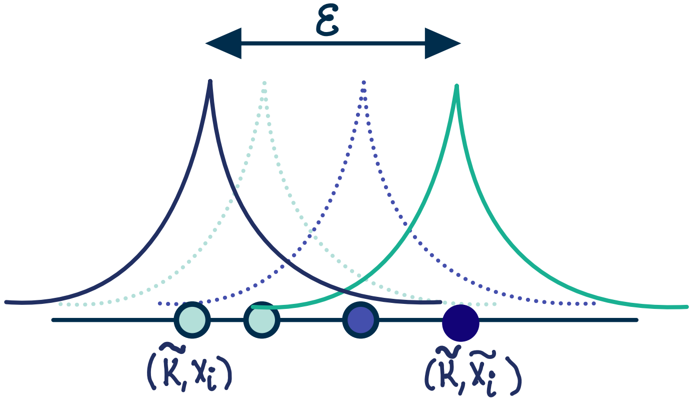
Bayesian Differential Privacy (BDP)

Bayesian DP leakage

$$\mathrm{BDPL}_{(K,i)} = \sup_{x_i, x_i', \mathbf{x}_K, S} \ln \frac{\Pr_{\mathcal{M}}[Y \in S \mid \mathbf{X}_K = \mathbf{x}_K, X_i = x_i]}{\Pr_{\mathcal{M}}[Y \in S \mid \mathbf{X}_K = \mathbf{x}_K, X_i = x_i']}, \text{ then } \varepsilon = \sup_{K,i} \mathrm{BDPL}_{(K,i)}.$$



$K = \{3\}, i = 2$



Bayesian Differential Privacy (BDP)

Bayesian DP leakage

$$\mathrm{BDPL}_{(K,i)} = \sup_{x_i, x_i', \mathbf{x}_K, S} \ln \frac{\Pr_{\mathcal{M}}[Y \in S \mid \mathbf{X}_K = \mathbf{x}_K, X_i = x_i]}{\Pr_{\mathcal{M}}[Y \in S \mid \mathbf{X}_K = \mathbf{x}_K, X_i = x_i']}, \text{ then } \varepsilon = \sup_{K,i} \mathrm{BDPL}_{(K,i)}.$$

Privacy

- ✔ Effective measure and resistance to correlation-based attacks.
- ✔ Instance of Pufferfish framework.
- ✔ Good properties: post-processing & (some) composition.

Utility

- ✘ Computationally intractable methods (computing the Wasserstein distance).
- ✘ Poor utility (methods based on group privacy).
- ✘ Limited applicability (lazy, binary, stationary Markov chains).

Our Research Question

Can we reduce utility loss while still retaining the privacy guarantees of BDP?

Our methodology: Understanding how DP leakage relates to BDP leakage:

$$\varepsilon\text{-DP} \Rightarrow \text{??-BDP}.$$

Against arbitrary correlations it is impossible

Kifer and Machanavajjhala 2014:

$$\begin{array}{c}\text{Pufferfish (including BDP)} \\ \wedge \\ \text{arbitrary correlation}\end{array} \implies \text{Free-lunch Privacy} \implies \text{No utility.}$$

We express this in term of (α, β)-accuracy: $0 \leq \beta < \frac{1}{e^{\varepsilon}+1}$ and any target query f, then $\alpha > \frac{1}{2} \max_{D,D'} |f(D) - f(D')|$.

$1 - \beta$ = Confidence
α = Error, interval radius with confidence $1 - \beta$.

Against arbitrary correlations it is impossible

Kifer and Machanavajjhala 2014:

Pufferfish (including BDP) $\wedge$ arbitrary correlation $\Rightarrow$ Free-lunch Privacy $\Rightarrow$ No utility.

We express this in term of (α, β)-accuracy: $0 \leq \beta < \frac{1}{e^{\varepsilon}+1}$ and any target query f, then $\alpha > \frac{1}{2} \max_{D,D'} |f(D) - f(D')|$.

$1 - \beta$ = Confidence
α = Error, interval radius with confidence $1 - \beta$.

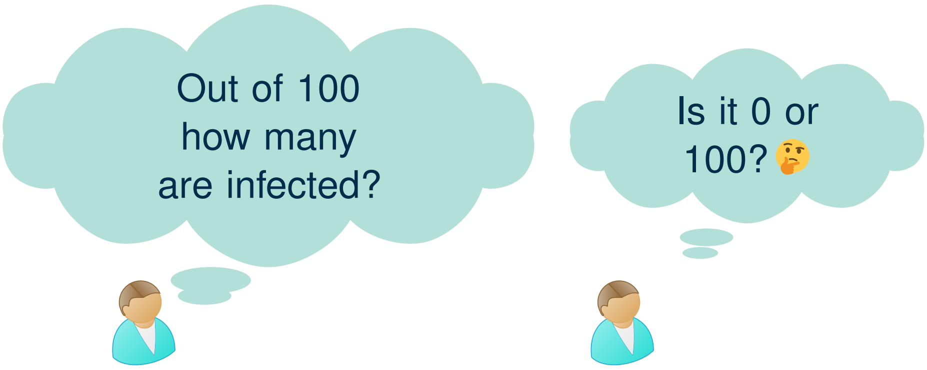
Against arbitrary correlations it is impossible

Kifer and Machanavajjhala 2014:

Pufferfish (including BDP) $\wedge$ arbitrary correlation $\Rightarrow$ Free-lunch Privacy $\Rightarrow$ No utility.

We express this in term of (α, β)-accuracy: $0 \leq \beta < \frac{1}{e^{\varepsilon}+1}$ and any target query f, then $\alpha > \frac{1}{2} \max_{D,D'} |f(D) - f(D')|$.

$1 - \beta$ = Confidence
α = Error, interval radius with confidence $1 - \beta$.



Few Correlated Records, Same Disaster

Our result (informal)

Privacy decreases linearly proportional to number of correlated records:

$$\varepsilon\text{-DP} \Rightarrow m\varepsilon\text{-BDP}$$

How does it impact utility?

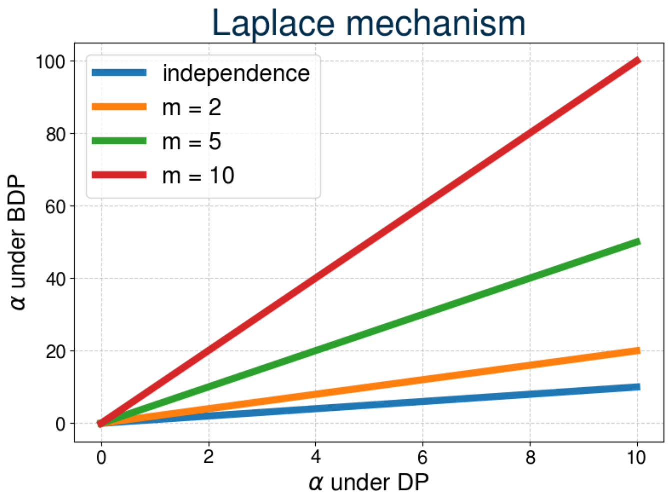


Figure: For the same confidence level, the upper bound on the query error α increases sharply.

Few Correlated Records, Same Disaster

Our result (informal)

Privacy decreases linearly proportional to number of correlated records:

$$\varepsilon\text{-DP} \Rightarrow m\varepsilon\text{-BDP}$$

This result is tight! Even excluding edge-cases.

How does it impact utility?

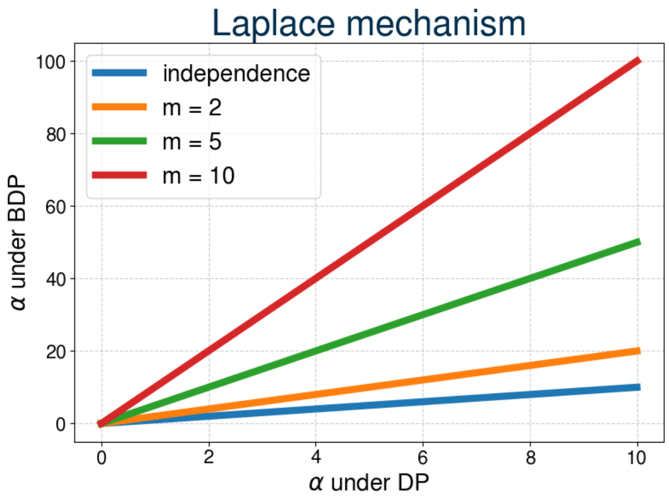


Figure: For the same confidence level, the upper bound on the query error α increases sharply.

Few Correlated Records, Same Disaster

Our result (informal)

Privacy decreases linearly proportional to number of correlated records:

$$\varepsilon\text{-DP} \Rightarrow m\varepsilon\text{-BDP}$$

This result is tight! Even excluding edge-cases.

Conclusion:

We need to target specific correlation models π to obtain utility

How does it impact utility?

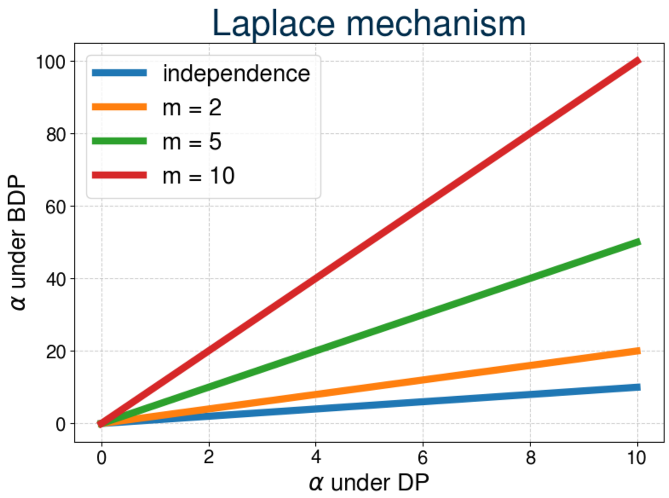


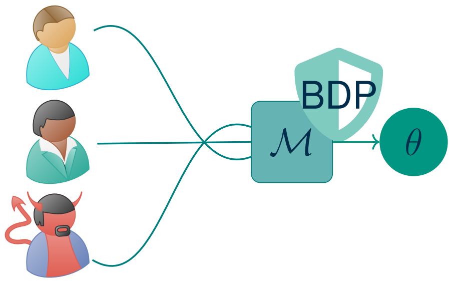
Figure: For the same confidence level, the upper bound on the query error α increases sharply.

New Strategy

Adjust the noise of DP mechanisms to obtain useful BDP mechanisms targeting specific priors π.

Assumptions:

- Global setting: All data is collected by a trusted data curator that applies the mechanism.
- The attacker does not have more knowledge about π than the data curator.

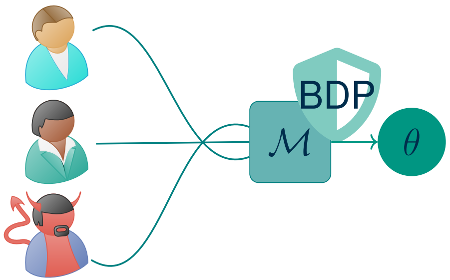


New Strategy

Adjust the noise of DP mechanisms to obtain useful BDP mechanisms targeting specific priors π.

Assumptions:

- Global setting: All data is collected by a trusted data curator that applies the mechanism.
- The attacker does not have more knowledge about π than the data curator.

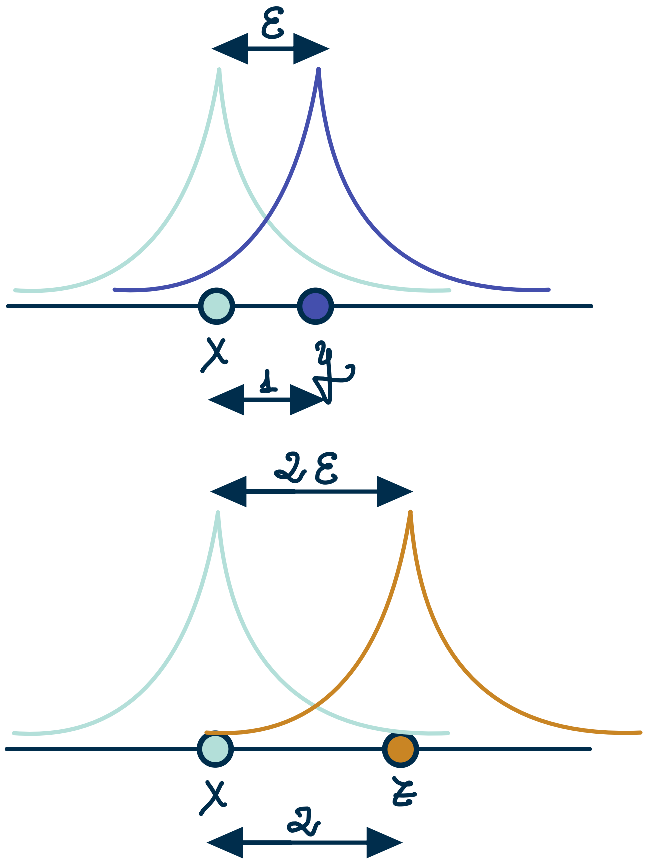


Multivariate Gaussian | **Markov Chains**

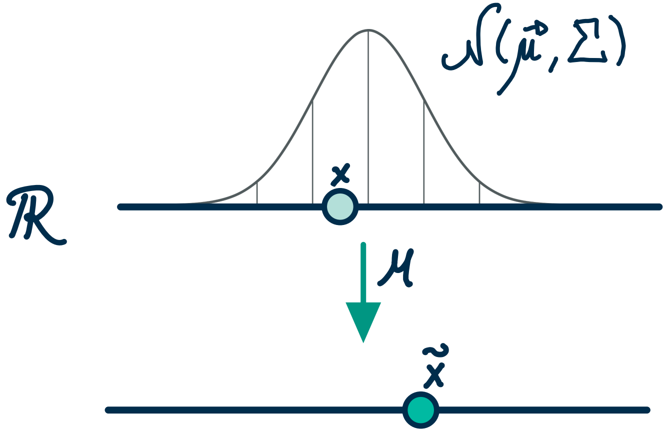
Multivariate Gaussian Correlation

Main Result (Informal)

- Let $\mathcal{M}$ be an $\varepsilon\ell_1$-private mechanism,



Multivariate Gaussian Correlation



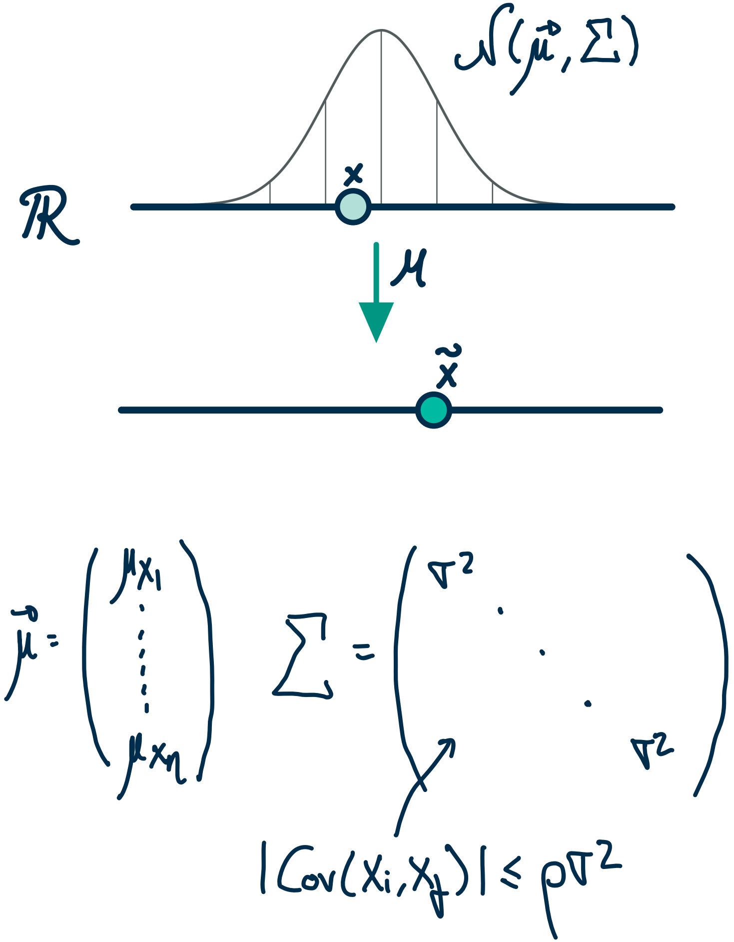
Main Result (Informal)

- Let $\mathcal{M}$ be an $\varepsilon\ell_1$-private mechanism,
- input data drawn from a multivariate Gaussian distribution

Multivariate Gaussian Correlation

Main Result (Informal)

- Let $\mathcal{M}$ be an $\varepsilon\ell_1$-private mechanism,
- input data drawn from a multivariate Gaussian distribution
- $\rho(m-2) < 1$ is the maximum correlation coefficient.

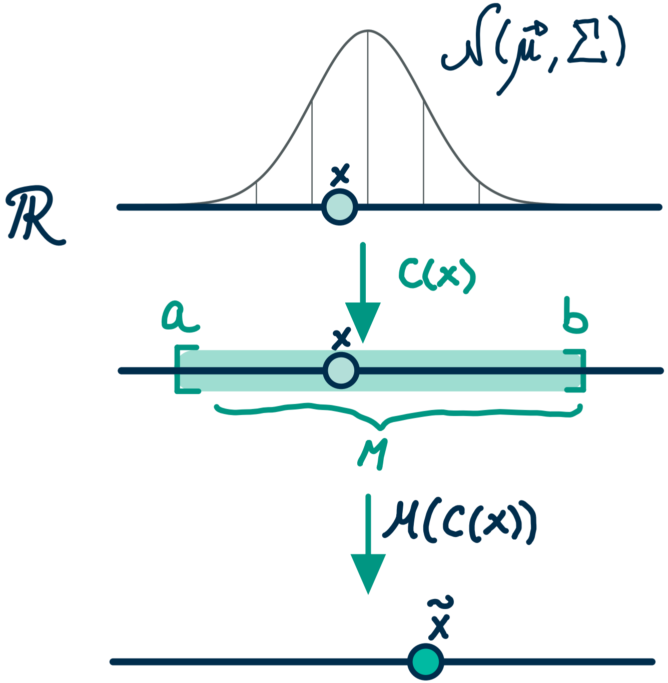


Multivariate Gaussian Correlation

Main Result (Informal)

- Let $\mathcal{M}$ be an $\varepsilon\ell_1$-private mechanism,
- input data drawn from a multivariate Gaussian distribution
- $\rho(m-2) < 1$ is the maximum correlation coefficient.

Then, using clipping as preprocessing step, $c_l(D)_i = \max(a, \min(b, D_i))$, we obtain $\mathcal{M}_l$ satisfying



Multivariate Gaussian Correlation

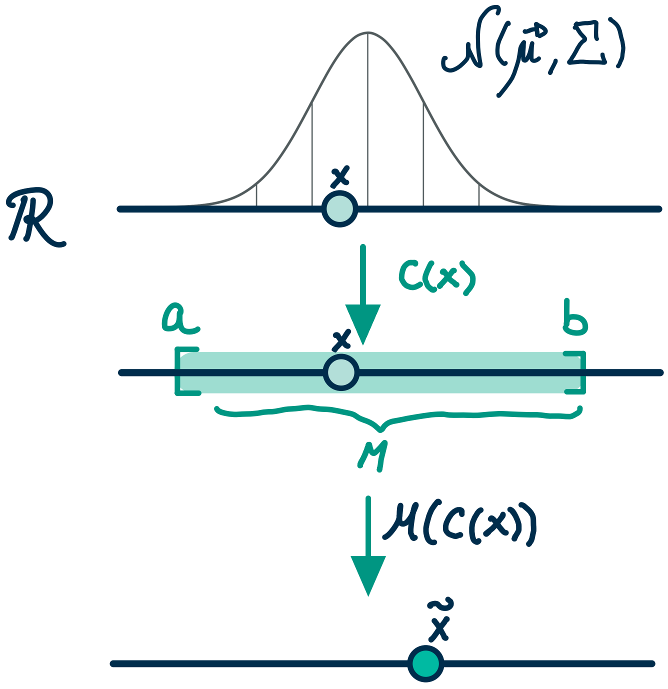
Main Result (Informal)

- Let $\mathcal{M}$ be an $\varepsilon\ell_1$-private mechanism,
- input data drawn from a multivariate Gaussian distribution
- $\rho(m-2) < 1$ is the maximum correlation coefficient.

Then, using clipping as preprocessing step, $c_I(D)_i = \max(a, \min(b, D_i))$, we obtain $\mathcal{M}_I$ satisfying

$$\mathrm{BDPL}(\mathcal{M}_I) \leq \left(\frac{m^2}{4(\frac{1}{\rho} - m + 2)} + 1 \right) M\varepsilon.$$

where M is the diameter of the interval $I = [a, b]$



Multivariate Gaussian Correlation (Impact on Real Databases)

- **Use-case:** Sum queries with Laplace mechanism. $\theta = f(D) + Z$ with $Z \sim \mathrm{Lap(b)}$.

Multivariate Gaussian Correlation (Impact on Real Databases)

- **Use-case:** Sum queries with Laplace mechanism. $\theta = f(D) + Z$ with $Z \sim \mathrm{Lap}(b)$.
- **Strategy:** We calibrate b to obtain BDP using our theorem.

Multivariate Gaussian Correlation (Impact on Real Databases)

- **Use-case:** Sum queries with Laplace mechanism. $\theta = f(D) + Z$ with $Z \sim \mathrm{Lap}(b)$.
- **Strategy:** We calibrate b to obtain BDP using our theorem.
- **Utility metric:** We set $\beta = 0.05$ (i.e., 95% confidence) and measure (α, β)-accuracy, both theoretically ($-$) and empirically ($\times$).

Multivariate Gaussian Correlation (Impact on Real Databases)

- **Use-case:** Sum queries with Laplace mechanism. $\theta = f(D) + Z$ with $Z \sim \mathrm{Lap}(b)$.
- **Strategy:** We calibrate *b* to obtain BDP using our theorem.
- **Utility metric:** We set $\beta = 0.05$ (i.e., 95% confidence) and measure (α, β)-accuracy, both theoretically (—) and empirically ($\times$).

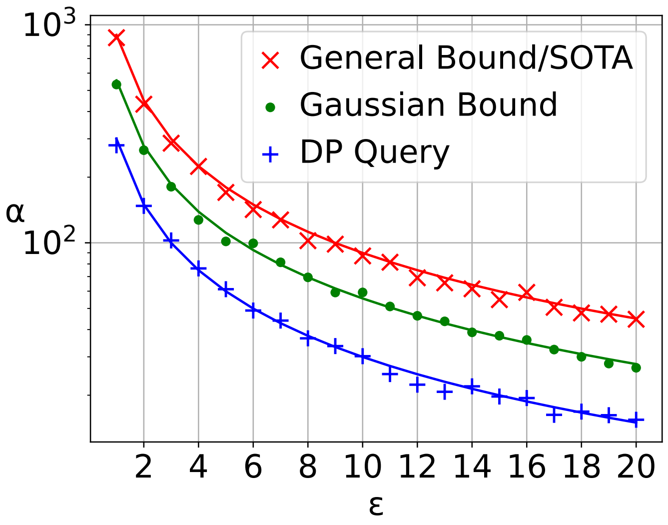


Figure: Galton, $n = 897$ $m = 3$

Key takeway:

Substantial utility gains compared to the general bound!

- More experiments with different real and synthetic datasets in our paper show similar results.

Markov Chain Correlation Model (Theoretical Results)

Main result (Informal)

- Let $\mathcal{M}$ be an ε-DP mechanism,
- input data sampled form Markov chain with transition matrix $P \in \mathbb{R}^{s \times s}$ and initial distribution $w \in \mathbb{R}^{s}$ with the following properties:

$$\text{(H1) For all } x, y \in \mathcal{S} \text{ we have } P_{x,y} > 0 \text{ and,} \qquad \text{(H2) } wP = w.$$

Then, $\mathcal{M}$ is an $(\varepsilon + 4 \ln \gamma)$-BDP mechanism where $\gamma = \frac{\max_{x,y \in \mathcal{S}} P_{xy}}{\min_{x,y \in \mathcal{S}} P_{xy}}$.

Markov Chain Correlation Model (Theoretical Results)

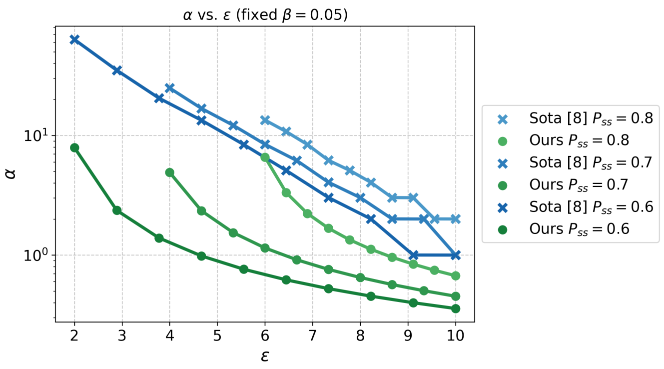
Main result (Informal)

- Let $\mathcal{M}$ be an ε-DP mechanism,
- input data sampled form Markov chain with transition matrix $P \in \mathbb{R}^{s \times s}$ and initial distribution $w \in \mathbb{R}^{s}$ with the following properties:

$$\text{(H1) For all } x, y \in \mathcal{S} \text{ we have } P_{x,y} > 0 \text{ and,} \qquad \text{(H2) } wP = w.$$

Then, $\mathcal{M}$ is an $(\varepsilon + 4 \ln \gamma)$-BDP mechanism where $\gamma = \frac{\max_{x,y \in \mathcal{S}} P_{xy}}{\min_{x,y \in \mathcal{S}} P_{xy}}$.

Previous mechanism	Ours
$P_{xy} > 0$	$P_{xy} > 0$
stationary	stationary
lazy	
binary	
symmetric	
$\varepsilon' > 0$	$\varepsilon' > 4 \ln(\gamma)$



Markov Chain Correlation Model (Impact on Real Databases)

- **Use-case:** Counting queries with Laplace mechanism. $\theta = f(D) + Z$ with $Z \sim \mathrm{Lap(b)}$.

Markov Chain Correlation Model (Impact on Real Databases)

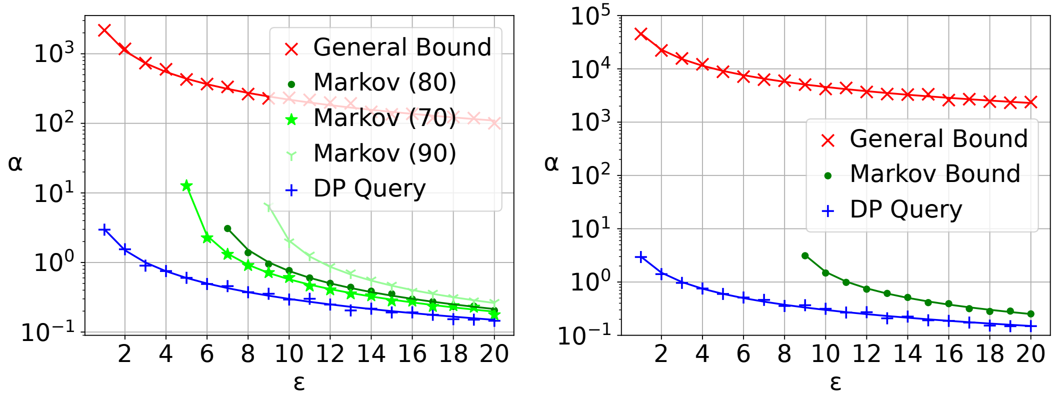
- **Use-case:** Counting queries with Laplace mechanism. $\theta = f(D) + Z$ with $Z \sim \mathrm{Lap}(\mathrm{b})$.
- **Strategy:** We calibrate b to obtain BDP using our theorem.

Markov Chain Correlation Model (Impact on Real Databases)

- **Use-case:** Counting queries with Laplace mechanism. $\theta = f(D) + Z$ with $Z \sim \mathrm{Lap}(\mathrm{b})$.
- **Strategy:** We calibrate b to obtain BDP using our theorem.
- **Utility metric:** We set $\beta = 0.05$ (i.e., 95% confidence) and measure $-(\alpha, \beta)$-accuracy, $\times$ upper bound of a $(1 - \beta)$ confidence interval for the absolute query error.

Markov Chain Correlation Model (Impact on Real Databases)

- **Use-case:** Counting queries with Laplace mechanism. $\theta = f(D) + Z$ with $Z \sim \mathrm{Lap}(b)$.
- **Strategy:** We calibrate b to obtain BDP using our theorem.
- **Utility metric:** We set $\beta = 0.05$ (i.e., 95% confidence) and measure $-(\alpha, \beta)$-accuracy, $\times$ upper bound of a $(1 - \beta)$ confidence interval for the absolute query error.



(g) Electricity, $n = 731$.
(h) Activity, $n = 17\,568$.

Key takeway:

- **Substantial utility gains** compared to the general bound!
- Markov bound independent of n $\Rightarrow$ **huge improvement for large datasets.**

Conclusion

- ✓ We provide a **feasible method** to generate a **BDP mechanism** by **recalibrating** existing DP methods, tailored to **Gaussian** and **Markov** models.

Conclusion

- ✓ We provide a **feasible method** to generate a **BDP mechanism** by **recalibrating** existing DP methods, tailored to **Gaussian** and **Markov** models.
- ✓ We offer **significantly better utility than prior results**.

Conclusion

- ✓ We provide a **feasible method** to generate a **BDP mechanism** by **recalibrating** existing DP methods, tailored to **Gaussian** and **Markov** models.
- ✓ We offer **significantly better utility than prior results**.

Key takeway:

BDP becomes usable when correlations are structured.

Conclusion

- ✓ We provide a **feasible method** to generate a **BDP mechanism** by **recalibrating** existing DP methods, tailored to **Gaussian** and **Markov** models.
- ✓ We offer **significantly better utility than prior results**.

Key takeway:

BDP becomes usable when correlations are structured.

Paper

Code

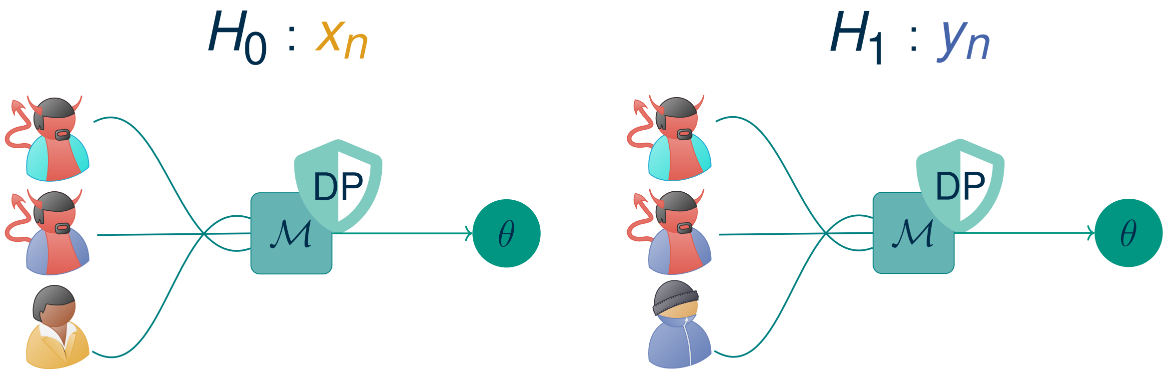
Future Work:

- Other distributions **?**
- Can we build methods from scratch instead or recycling **?**
- What if we calibrate directly to the attack advantage **?**

Backup Slides

Membership Inference Attack Knowing D_{-}

The attacker receives θ and aims to distinguish between:



D_{-} is known:

$H_0 = D_{x_n}$ Vs. $H_1 = D_{y_n}$

Type I error:
$\alpha = \Pr_{A \circ \mathcal{M}}(y_n \mid D_{x_n})$

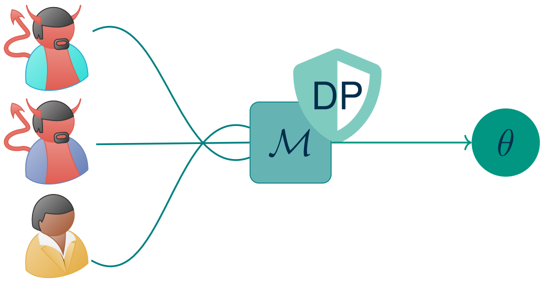
Type II error:
$\beta = 1 - \Pr_{A \circ \mathcal{M}}(y_n \mid D_{y_n})$

$A \circ \mathcal{M}$ is ε-DP $\implies$ $1 - \beta \leq e^{\varepsilon}\alpha$, $\alpha \leq e^{\varepsilon}(1 - \beta)$ $\implies$ $\beta \geq \max\{1 - e^{\varepsilon}\alpha, e^{\varepsilon}(1 - \alpha)\}$

Membership Inference Attack Knowing D_{-}

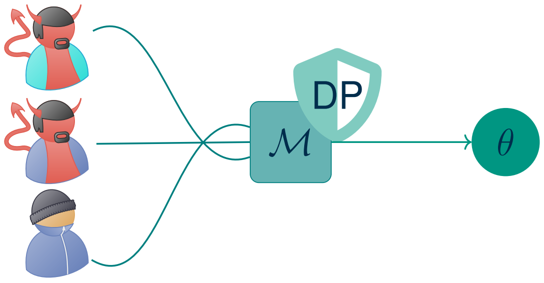
The attacker receives θ and aims to distinguish between:

$H_0 : x_n$

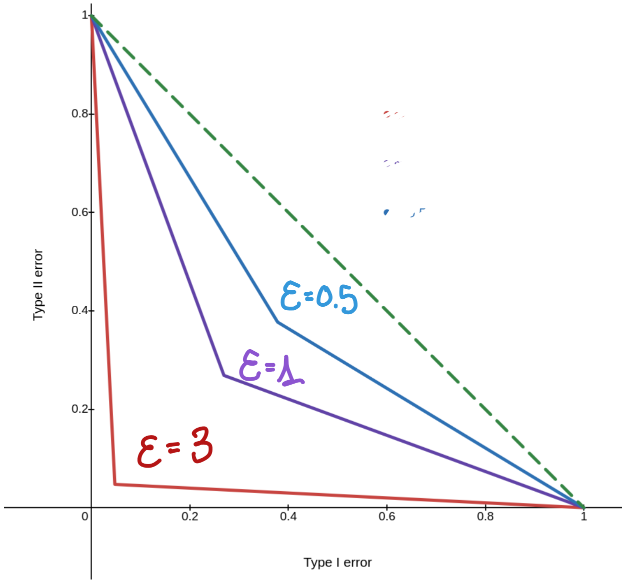


Type I error:
$\alpha = \Pr_{A \circ \mathcal{M}}(y_n \mid D_{x_n})$

$H_1 : y_n$



Type II error:
$\beta = 1 - \Pr_{A \circ \mathcal{M}}(y_n \mid D_{y_n})$

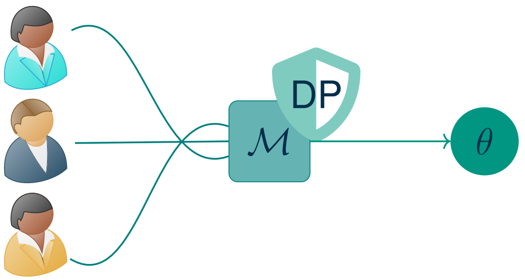


$A \circ \mathcal{M}$ is ε-DP $\Rightarrow$ $1 - \beta \leq e^{\varepsilon}\alpha$, $\alpha \leq e^{\varepsilon}(1 - \beta)$ $\Rightarrow$ $\beta \geq \max\{1 - e^{\varepsilon}\alpha, e^{\varepsilon}(1 - \alpha)\}$

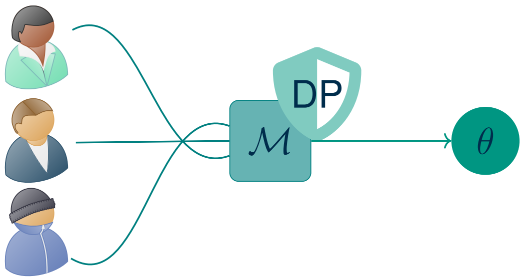
Membership Inference Attack With Dependencies

The attacker receives θ and aims to distinguish between:

$H_0 : x_n$



$H_1 : y_n$



D_- is **unknown**:

$$H_0 = \{D \colon x_n \in D\} \text{ Vs. } H_1 = \{D \colon y_n \in D\}$$

Type I error:

$$\alpha = \Pr_{A \circ \mathcal{M}}(y_n \mid x_n) = \sum_{D_-} \Pr_{A \circ \mathcal{M}}(y_n \mid D_{x_n}) \, \pi(D_- \mid x_n)$$

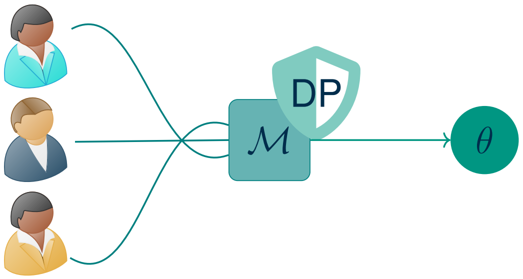
Type II error:

$$\beta = 1 - \Pr_{A \circ \mathcal{M}}(y_n \mid y_n) = 1 - \sum_{D_-} \Pr_{A \circ \mathcal{M}}(y_n \mid D_{y_n}) \, \pi(D_- \mid y_n)$$

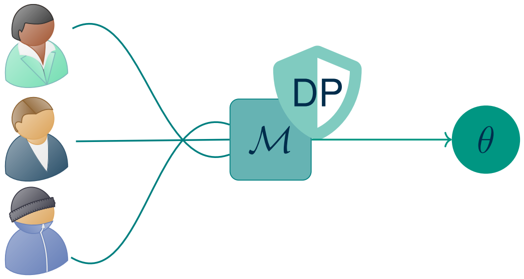
Membership Inference Attack With Dependencies

The attacker receives θ and aims to distinguish between:

$H_0 : x_n$



$H_1 : y_n$



D_- is **unknown**:

$H_0 = \{D \colon x_n \in D\}$ Vs. $H_1 = \{D \colon y_n \in D\}$

Type I error:

$$\alpha = \Pr_{A \circ \mathcal{M}}(y_n \mid x_n) = \sum_{D_-} \Pr_{A \circ \mathcal{M}}(y_n \mid D_{x_n})\, \pi(D_- \mid x_n)$$

Type II error:

$$\beta = 1 - \Pr_{A \circ \mathcal{M}}(y_n \mid y_n) = 1 - \sum_{D_-} \Pr_{A \circ \mathcal{M}}(y_n \mid D_{y_n})\, \pi(D_- \mid y_n)$$

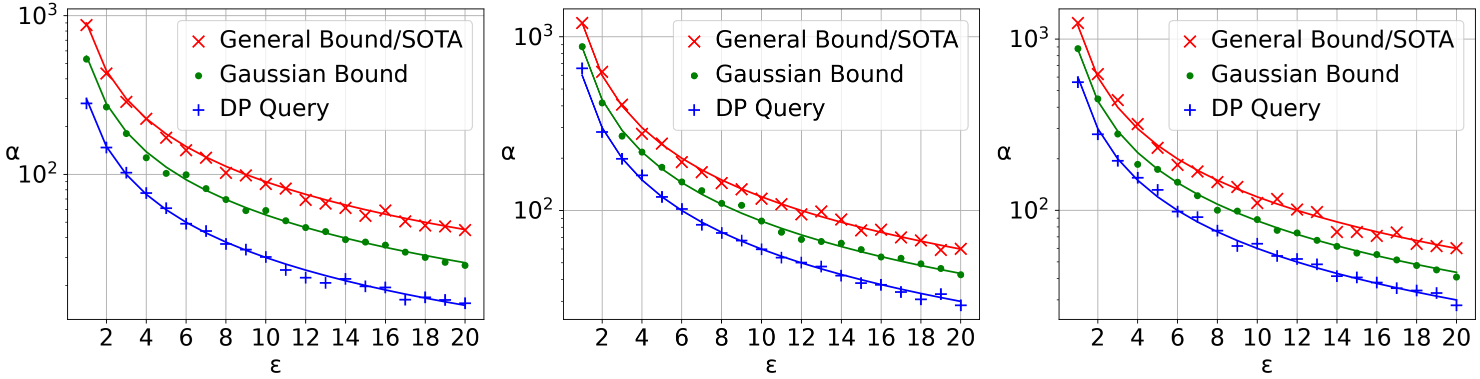
$A \circ \mathcal{M}$ is ε-DP $\implies$ $$1 - \beta \leq \sum_{D_-} e^{\varepsilon} \Pr_{A \circ \mathcal{M}}(y_n \mid D_{x_n}) \pi(D_- \mid y_n) = e^{\varepsilon} \sum_{D_-} \Pr_{A \circ \mathcal{M}}(y_n \mid D_{x_n}) \pi(D_- \mid y_n) \neq e^{\varepsilon} \alpha$$

Experiment Details

Database	n	m	Parameters	Sensitivity
Galton	897	3	$\rho = 0.275$	$\Delta q = 254cm$
FamilyIQ	868	2	$\rho = 0.4483$	$\Delta q = 120$
SyntheticIQ	20000	2	$\rho = 0.45$	$\Delta q = 120$
Activity	17568	n	$\gamma = 7.54$	$\Delta q = 1$
Activity Single Day	288	n	$\gamma = 7.54$	$\Delta q = 1$
Electricity	731	n	70 kWh, $\gamma = 3.29$ 80 kWh, $\gamma = 4.49$ 90 kWh, $\gamma = 8.43$	$\Delta q = 1$

Table: Data description. m is the max number of correlated records and n the total amount.

Multivariate Gaussian More Results



(i) Galton, $n = 897$ $m = 3$ (j) FamilyIQ, $n = 868$, $m = 2$. (k) SyntheticIQ, $n = 20000$, $m = 2$.

Figure: Gaussian data results. Lines show theoretical error at $\beta = 5\%$ and markers indicate empirical 95% upper bounds.