FCAI Finnish Center for Artificial Intelligence

Differentially Private Markov Chain Monte Carlo

Mikko Heikkilä *[1], Joonas Jälkö *[2], Onur Dikmen [3] and Antti Honkela [4]

* Equal contribution
[1] Helsinki Institute for Information Technology HIIT, Department of Mathematics and Statistics, University of Helsinki, Finland
[2] Helsinki Institute for Information Technology HIIT, Department of Computer Science, Aalto University, Finland
[3] Center for Applied Intelligent Systems Research (CAISR), Halmstad University, Sweden
[4] Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki, Finland

UNIVERSITY OF HELSINKI

Aalto University

HALMSTAD UNIVERSITY

Abstract

Recent developments in differentially private (DP) machine learning and DP Bayesian learning have enabled learning under strong privacy guarantees for the training data subjects. In this paper, we further extend the applicability of DP Bayesian learning by presenting the first general DP Markov chain Monte Carlo (MCMC) algorithm whose privacy-guarantees are not subject to unrealistic assumptions on Markov chain convergence and that is applicable to posterior inference in arbitrary models. Our algorithm is based on a decomposition of the Barker acceptance test that allows evaluating the Rényi DP privacy cost of the accept-reject choice.

Background : Differential privacy

(ϵ, δ)-differential privacy

A randomized algorithm $\mathcal{M} : \mathcal{X}^N \to \mathcal{I}$ satisfies (ϵ, δ) differential privacy [1], if for all adjacent datasets $\mathbf{x}, \mathbf{x}' \in \mathcal{X}^N$ and for all measurable $I \subset \mathcal{I}$ it holds that

$$\Pr(\mathcal{M}(\mathbf{x}) \in I) \le e^{\epsilon} \Pr(\mathcal{M}(\mathbf{x}') \in I) + \delta. \tag{1}$$

Rényi divergence

Rényi divergence [3] between two distributions P and Q defined over $\mathcal{I}$ is defined as

$$D_\alpha(P \,||\, Q) = \frac{1}{\alpha - 1} \log \mathbb{E}_P \left[\left(\frac{p(X)}{q(X)} \right)^{\alpha-1} \right]. \tag{2}$$

Rényi differential privacy (RDP)

A randomized algorithm $\mathcal{M} : \mathcal{X}^N \to \mathcal{I}$ is (α, ϵ)-RDP [2], if for all adjacent datasets $\mathbf{x}, \mathbf{x}'$ it holds that

$$D_\alpha(\mathcal{M}(\mathbf{x}) \,||\, \mathcal{M}(\mathbf{x}')) \le \epsilon \triangleq \epsilon(\alpha). \tag{3}$$

Background : Markov chain Monte Carlo

We want to draw samples from distribution $\pi(\theta)$ that can only be evaluated up to a normalizing constant. More specifically, we are interested in drawing samples from a posterior distribution

$$p(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)p(\theta) = \prod_i p(x_i|\theta)p(\theta), \tag{4}$$

where $p(\mathbf{x}|\theta)$ is the likelihood, $p(\theta)$ is the prior for parameters, and the last equality holds assuming exchangeability.

Metropolis-Hastings (M-H) [4] [5]

Accept proposed move θ' from distribution q with probability $\min\{\exp(\Delta), 1\}$

$$\Delta := \Delta(\theta', \theta) = \sum_{x_i \in \mathbf{x}} \log \frac{p(x_i|\theta')}{p(x_i|\theta)} + \log \frac{p(\theta')q(\theta|\theta')}{p(\theta)q(\theta'|\theta)}. \tag{5}$$

Background: Subsampled MCMC (Seita et al. [6])

- Instead of M-H acceptance test, use Barker acceptance test.
- Accept proposed move if $\Delta + V_{log} > 0$, where $V_{log} \sim \text{Logistic}(0, 1)$.
- Satisfies detailed balance.
- To ease the computational burden use only a random subset $S \subset \mathbf{x}$ of size b to evaluate acceptance.

$$\Delta^* := \Delta^*(\theta', \theta) = \frac{N}{b} \sum_{x_i \in S} \log \frac{p(x_i|\theta')}{p(x_i|\theta)} + \log \frac{p(\theta')q(\theta|\theta')}{p(\theta)q(\theta'|\theta)} \tag{6}$$

- As a sum of i.i.d. random variables Δ^* is approximately normal according to CLT with mean Δ.
- Decompose V_{log} as $V_{log} = V_{norm} + V_{cor}$, where V_{norm} has normal distribution and V_{cor} is a suitable correction.
- Approximate the full data test using only a minibatch as

$$\Delta^* + V_{cor} \simeq \Delta + V_{norm} + V_{cor} \simeq \Delta + V_{log} \tag{7}$$

Tempering

- When sample size is very large, we tend to overfit the posterior distribution (see e.g. [7]).
- To address this, we scale the log-likelihood ratios in (5) and (6) by factor τ.

DP MCMC

- Repurpose decomposition idea for guaranteeing privacy.
- Logistic r.v. V_{log} has variance $\pi^2/3$.
- Fix $0 < C < \pi^2/3$ and write

$$V_{log} \simeq \mathcal{N}(0, C) + V_{cor}^{(C)}. \tag{8}$$

- Thus testing if $\Delta + V_{log} > 0 \Leftrightarrow \Delta + \mathcal{N}(0, C) + V_{cor}^{(C)} > 0$.
  - Gaussian mechanism with variance C.
- Want to use as large C as possible to achieve tight privacy guarantees.
- Approximate the distribution of $V_{cor}^{(C)}$ using a Gaussian mixture model.

DP subsampled MCMC

- Again we decompose $V_{log} = \mathcal{N}(0, C) + V_{cor}^{(C)}$.
- Now according to CLT we have

$$\Delta^* = \Delta + \tilde{V}_{norm}, \tag{9}$$

where $\tilde{V}_{norm}$ is approximately normal with variance $\text{Var}(\tilde{V}_{norm}) = \sigma^2_{\Delta^*}$, which is the subsampling induced variance.
- Assuming $C > \sigma^2_{\Delta^*}$, we further decompose the normal variable $\mathcal{N}(0, C)$ as

$$\mathcal{N}(0, C) = \mathcal{N}(0, C - \sigma^2_{\Delta^*}) + \tilde{V}_{norm} \tag{10}$$

- Now testing if $\Delta^* + \mathcal{N}(0, C - \sigma^2_{\Delta^*}) + V_{cor}^{(C)} > 0$ approximates the full test.
  - Gaussian mechanism again, but the variance depends on the data.

Privacy analysis

- Need to bound the Rényi divergences between two Gaussians resulting from neighbouring datasets.
- Known analytical form, can be bounded with standard techniques under some assumptions.
- The assumptions can be forced and do not rely on the chain's convergence.

Analysis for DP MCMC (using full data)

- Denote the acceptance test using data $\mathbf{x}$ with $\mathcal{M}(\mathbf{x})$.
- Assuming either

$$|\log p(x_i \,|\, \theta') - \log p(x_i \,|\, \theta)| \le B \tag{11}$$

or

$$|\log p(x_i \,|\, \theta) - \log p(x_j \,|\, \theta)| \le B, \tag{12}$$

leads to $\epsilon(\alpha) = D_\alpha(\mathcal{M}(\mathbf{x})||\mathcal{M}(\mathbf{x}')) \le 2\alpha B^2/C$.
- After T iterations the RDP composition [2] implies that the algorithm satisfies $(\alpha, T2\alpha B^2/C)$-RDP.
- Can satisfy the condition (11) with sufficiently smooth likelihoods and a proposal distribution with a bounded domain.
- Can also force the condition (11) by clipping the log-likelihood ratios in (5).

Analysis for DP subsampled MCMC (using a minibatch)

- Assuming $\alpha < b/5$ and

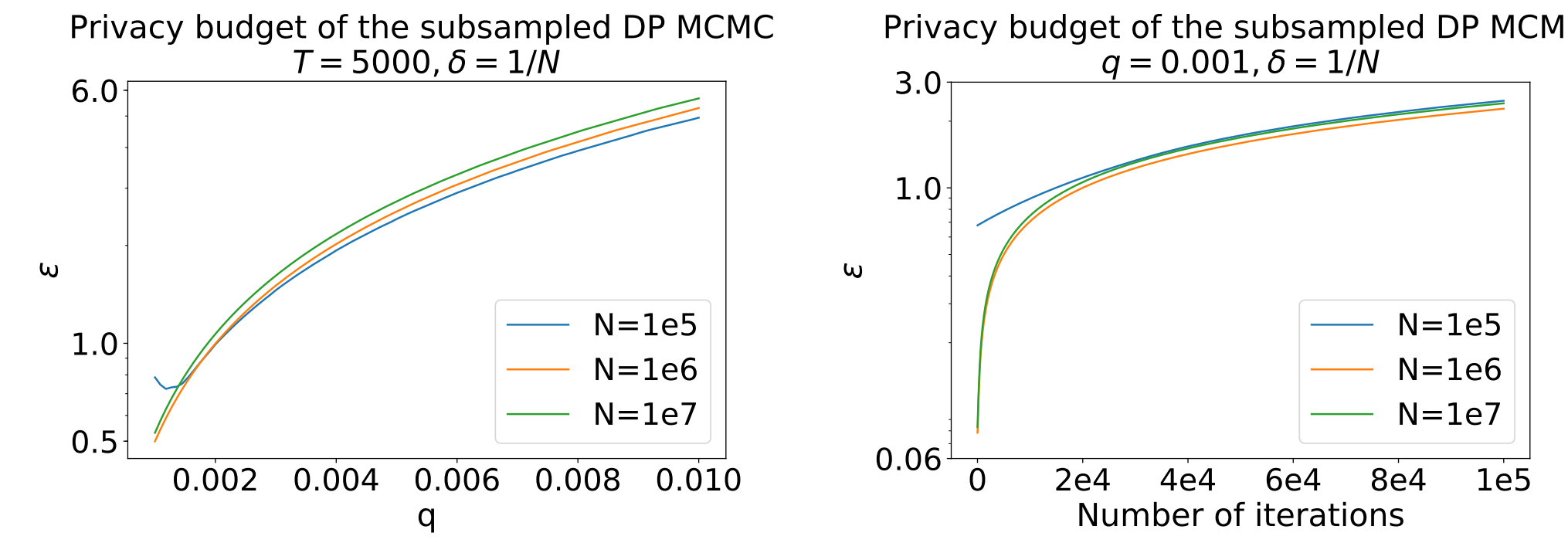
$$|\log p(x_i|\theta') - \log p(x_i|\theta)| \le \frac{\sqrt{b}}{N} \tag{13}$$

where b is the size of the minibatch S and N is the dataset size, releasing a sample from subsampled MCMC method satisfies $(\alpha, \epsilon(\alpha))$-RDP with

$$\epsilon(\alpha) = \frac{5}{2b} + \frac{1}{2(\alpha - 1)} \ln \frac{2b}{b - 5\alpha} + \frac{2\alpha}{b - 5\alpha}. \tag{14}$$

- Apply subsampling amplification [8] and composition [2] results to get the privacy cost for the full mechanism.
- Figures 1(a) and 1(b) illustrate how changing the parameters q and T affects the privacy cost.
- Can satisfy the condition (13) similarly as in the full data case.
- The bound in (13) gets tighter with increasing N.
  - We temper the log-likelihoods with $\tau = 1/(1 + (N - N_0)/N_0)$.
  - Then instead of condition (13) we require

$$|\log p(x_i|\theta') - \log p(x_i|\theta)| \le \frac{\sqrt{b}}{N_0}. \tag{15}$$



(a) Privacy vs. subsampling ratio

(b) Privacy vs. iterations

Figure 1: Parameter effects. Calculating total privacy budget for subsampled method for different dataset sizes: in Figure 1(a) as a function of subsampling ratio, and in Figure 1(b) as a function of number of iterations.

Example: Mixture of Gaussians

Description of data

We demonstrate our proposed method using a 2-dimensional Gaussian mixture model

$$\theta_j \sim \mathcal{N}(0, \sigma_j^2,), \quad j = 1, 2 \tag{16}$$

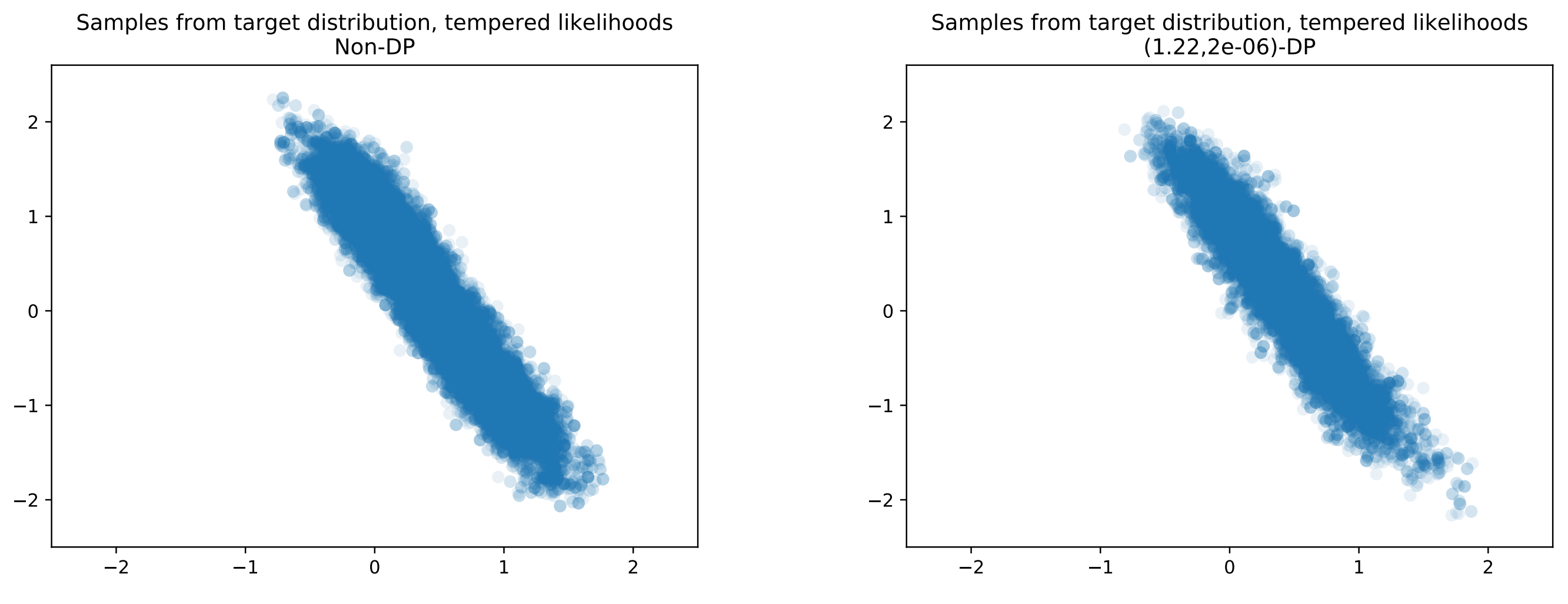
$$x_i \sim 0.5 \cdot \mathcal{N}(\theta_1, \sigma_x^2) + 0.5 \cdot \mathcal{N}(\theta_1 + \theta_2, \sigma_x^2), \tag{17}$$

where $\sigma_1^2 = 10, \sigma_2^2 = 1, \sigma_x^2 = 2$. For the observed data, we use fixed parameter values $\theta = (0, 1)$. We generate 10^6 samples from the model to use as training data. We use $b = 1000$ for the minibatches, and adjust the temperature of the chain s.t. $N_0 = 100$ in (15). This corresponds to the temperature used by Seita et al. [6] in their non-private test. The parameters were initialized using DPVI method [9] with a small privacy budget $(0.22, 10^{-6})$.

Notations cheat sheet

$\alpha, \epsilon = \epsilon(\alpha)$	Privacy parameters for RDP
$\mathbf{x}$	Data
N	Dataset size
θ	Model parameters
$\Delta = \Delta(\theta', \theta)$	Log-likelihood ratio
$\Delta^* = \Delta^*(\theta', \theta)$	Log-likelihood ratio for a minibatch
$b > 5\alpha$	Batch size for subsampled DP-MCMC
$C \in (0, \pi^2/3)$	Noise variance, in subsampled case we set $C = 2$
V_{log}	Standard logistic rv
V_{norm}	$\mathcal{N}(0, C)$ rv
$V_{cor}^{(C)}$	Rv s.t $V_{norm} + V_{cor} \simeq V_{log}$
$\tilde{V}_{norm}$	Approximately normal rv with variance $\sigma^2_{\Delta^*} < C$
B	Bound for the log-likelihood ratios (llr)
T	Number of MCMC draws
β	Parameter for tempering

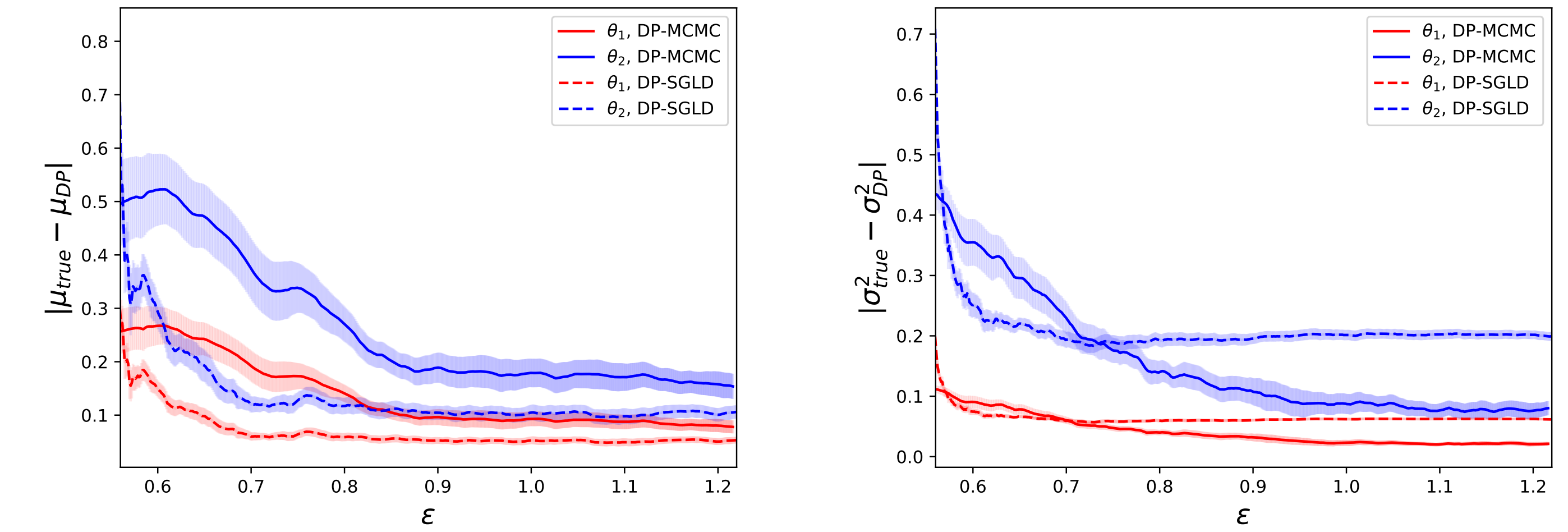
Results



(a) Samples without privacy

(b) Samples with DP

Figure 2: Results for the GMM experiment with tempered likelihoods: 2(a) shows 5000 samples from the chain without privacy and 2(b) with privacy. The results with strict privacy are very close to the non-private results.



(a) Posterior mean accuracy

(b) Posterior variance accuracy

Figure 3: Intermediate private posterior statistics from DP SGLD and DP MCMC compared against the baseline given by a non-private chain after 40000 iterations. Lines showing the mean error between 20 runs of the algorithm with errorbars illustrating the standard error of the mean between the runs. DP SGDL converges quickly towards the posterior mean, but does not properly capture the posterior variance.

Conclusions

- We present a new generic DP-MCMC method with strict, non-asymptotic privacy guarantees that hold independently of the chain's convergence.
- The proposed method allows for structurally new kind of assumptions to guarantee privacy through forcing bounds on the proposal instead of or in addition to the likelihood.

References

[1] Dwork, Cynthia and Frank McSherry and Kobbi Nissim and Adam Smith. Calibrating Noise to Sensitivity in Private Data Analysis. TCC 2006.
[2] Mironov, Ilya. Rényi differential privacy. Computer Security Foundations Symposium (CSF), 2017 IEEE 30th.
[3] Rényi, Alfréd. On Measures of Entropy and Information. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics 1961.
[4] Metropolis, Nicholas and Rosenbluth, Arianna W. and Rosenbluth, Marshall N. and Teller, Augusta H. and Teller, Edward. Equation of State Calculations by Fast Computing Machines. The Journal of Chemical Physics 1953.
[5] W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika 1970.
[6] Seita, Daniel and Pan, Xinlei and Chen, Haoyu and Canny, John F. An Efficient Minibatch Acceptance Test for Metropolis–Hastings. UAI 2017.
[7] Robust Bayesian inference via coarsening. Miller, Jeffrey W. and Dunson, David B. Journal of the American Statistical Association 2019.
[8] Wang, Yu-Xiang and Balle, Borja and Kasiviswanathan, Shiva Prasad. Subsampled Rényi Differential Privacy and Analytical Moments Accountant. AISTATS 2019.
[9] Jälkö, Joonas and Dikmen, Onur and Honkela, Antti. Differentially Private Variational Inference for Non-conjugate Models. UAI 2017.

Acknowledgements

The authors would like to thank Daniel Seita for sharing the original code for their paper. This work has been supported by the Academy of Finland [Finnish Center for Artificial Intelligence FCAI and grants 294238, 303815, 313124].