Appendix A Definitions, proofs, and related work

Here, we provide missing definitions of the KLD and the β-divergence.

Definition 2 (Kullback-Leibler divergence [53]). *The* KLD *between probability densities* $g(\cdot)$ *and* $f(\cdot)$ *is given by*

$$\text{KLD}(g||f) = \int g(x) \log \frac{g(x)}{f(x)} dx.$$

Definition 3 (β-divergence [9, 63]). *The β-divergence is defined as*

$$D_B^{(\beta)}(g||f) = \frac{1}{\beta(\beta-1)} \int g(x)^{\beta} dx + \frac{1}{\beta} \int f(x)^{\beta} dx - \frac{1}{\beta - 1} \int g(x) f(x)^{\beta-1} dx,$$

where $\beta \in \mathbb{R} \backslash \{0, 1\}$. $D_B^{(\beta)}$ *is a member of the Bregman-divergence family [16] with* $\psi(t) = \frac{1}{\beta(\beta-1)} t^{\beta}$. *When* $\beta \to 1$, $D_B^{(1)}(g(x)||f(x)) \to \text{KLD}(g(x)||f(x))$.

The β-divergence has often been referred to as the *density-power divergence* in the statistics literature [9] where it is often parameterised with $\beta_{DPD} = \beta - 1$.

Intuition for how βD-Bayes provides DP estimation is provided in Figure 5 which shows the divergence between the posterior before and after adding an observation y that is $|y - \mu|$ standard deviations away from the posterior mean μ when updating using a Gaussian distribution under KLD-Bayes and βD-Bayes. The influence of observations under KLD-Bayes is steadily increasing, making the posterior sensitive to extreme observations and therefore leaking their information. Under βD-Bayes, the influence initially increases before being maximised at a point depending on the value of β, before decreasing to 0. Therefore, each observation has bounded influence on the posterior, allowing for DP estimation.

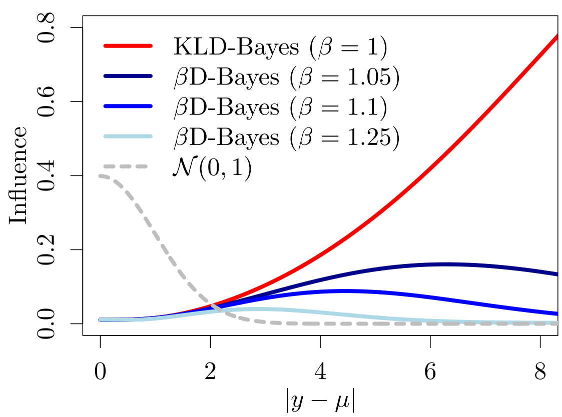

Figure 5: The influence of adding an observation y with $|y - \mu|$ on the posterior conditioned on a sample of 1000 points from a $\mathcal{N}(0, 1)$ when fitting a $\mathcal{N}(\mu, \sigma^2)$.

A.1 Bernstein-von Mises theorem for βD-Bayes

The general Bernstein-von Mises theorem for generalised posteriors [Theorem 4; 64] can be applied to the βD-Bayes posterior to show that

$$\int \left| \tilde{\pi}^{(\beta)}(\phi) - \mathcal{N}\left(\phi; 0, (H_0^{(\beta)})^{-1}\right) \right| d\phi \underset{n \to \infty}{\longrightarrow} 0 \tag{5}$$

where $\tilde{\pi}^{(\beta)}$ denotes the density of $\sqrt{n}(\tilde{\theta} - \hat{\theta}_n^{(\beta)})$ when $\tilde{\theta} \sim \pi^{(\beta)}(\cdot; D)$, $\mathcal{N}(x; \mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2, and

$$\hat{\theta}_n^{(\beta)} := \underset{\theta \in \Theta}{\arg\min} \sum_{i=1}^{n} \ell^{(\beta)}(D_i, f(\cdot; \theta)), \quad \theta_0^{(\beta)} := \underset{\theta \in \Theta}{\arg\min}\, \mathbb{E}_g \left[\ell^{(\beta)}(D, f(\cdot; \theta)) \right]$$

$$H_0^{(\beta)} := \left(\frac{\partial}{\partial \theta_i \partial \theta_j} \mathbb{E}_D \left[\ell^{(\beta)}(D, f(\cdot; \theta_0^{(\beta)})) \right] \right)_{i,j}.$$

That is to show that the βD-Bayes posterior converges to a Gaussian distribution centered around the βD minimising parameter $\theta_0^{(\beta)}$ in total variation distance.

A.2 Proofs

A.2.1 Proof of Lemma 1

Lemma 1 (Bounded sensitivity of the βD-Bayes loss)**.** *Under Condition 1 the sensitivity of the βD-Bayes-loss for any $\beta > 1$ is $\left|\ell^{(\beta)}(D, f(\cdot;\theta)) - \ell^{(\beta)}(D', f(\cdot;\theta))\right| \leq \frac{M^{\beta-1}}{\beta-1}$.*

Proof. By (4), for $\beta > 1$

$$\begin{aligned}\left|\ell^{(\beta)}(D, f(\cdot;\theta)) - \ell^{(\beta)}(D', f(\cdot;\theta))\right| &= \frac{1}{\beta-1}\left(f(D';\theta)^{\beta-1} - f(D;\theta)^{\beta-1}\right) \\ &\leq \max_D \frac{1}{\beta-1} f(D;\theta)^{\beta-1}) \\ &\leq \frac{M^{\beta-1}}{\beta-1}\end{aligned}$$

$\square$

A.2.2 Proof of Theorem 1

Theorem 1 (Differential privacy of the βD-Bayes posterior)**.** *Under Condition 1, a draw $\tilde{\theta}$ from the βD-Bayes posterior $\pi^{(\beta)}(\theta|D)$ in (3) is $(\frac{2M^{\beta-1}}{\beta-1}, 0)$-differentially private.*

Proof. Define $D = \{D_1, \ldots, D_n\}$, $D' = \{D'_1, \ldots, D'_n\}$ and let j be the index such that $D_j \neq D'_j$ with $D_j = D'_i$ for all $i \neq j$. Firstly, the normalising constant of the βD-Bayes posterior combining (3) with (4) is

$$P^{\ell}(D) := \int \pi(\theta) \exp\left(-w\sum_{i=1}^{n} \ell\{\theta, D_i\}\right) d\theta \tag{6}$$

Then,

$$\begin{aligned}\log\frac{\pi^{(\beta)}(\theta|D)}{\pi^{(\beta)}(\theta|D')} &= \sum_{i=1}^{n}\ell^{(\beta)}(D'_i, f(\cdot;\theta)) - \sum_{i=1}^{n}\ell^{(\beta)}(D_i, f(\cdot;\theta)) + \log\frac{P^{(\beta)}(D')}{P^{(\beta)}(D)} \\ &= \ell^{(\beta)}(D'_j; f(\cdot;\theta)) - \ell^{(\beta)}(D_j; f(\cdot;\theta)) + \log\frac{P^{(\beta)}(D')}{P^{(\beta)}(D)}\end{aligned}$$

where $P^{(\beta)}(D')$ is the normaliser of the general Bayesian posterior defined in (3).

Now, by Condition 1 and Lemma 1,

$$\ell^{(\beta)}(D'_j; f(\cdot;\theta)) - \ell^{(\beta)}(D_j; f(\cdot;\theta)) \leq \frac{M^{\beta-1}}{\beta-1},$$

and

$$\begin{aligned}P^{(\beta)}(D') &= \int \exp\left\{-\sum_{i=1}^{n}\ell^{(\beta)}(D'_i, f(\cdot;\theta))\right\}\pi(\theta)d\theta \\ &= \int \exp\left\{\ell^{(\beta)}(D_j, f(\cdot;\theta)) - \ell^{(\beta)}(D'_j, f(\cdot;\theta)) - \sum_{i=1}^{n}\ell^{(\beta)}(D_i, f(\cdot;\theta))\right\}\pi(\theta)d\theta \\ &= \exp\left\{\frac{M^{\beta-1}}{\beta-1}\right\}\int \exp\left\{-\sum_{i=1}^{n}\ell^{(\beta)}(D_i, f(\cdot;\theta))\right\}\pi(\theta)d\theta,\end{aligned}$$

which combined provides that

$$\log\frac{\pi(\theta|D)}{\pi(\theta|D')} \leq 2\frac{M^{\beta-1}}{\beta-1}.$$

$\square$

A.2.3 Proof of Theorem 2

Theorem 2 (Consistency of βD-Bayes sampling)**.** *Under the conditions of Theorem 4 of [64],*

1. *a posterior sample $\tilde{\theta} \sim \pi^{(\beta)}(\theta|D)$ is a consistent estimator of $\theta_0^{(\beta)}$.*
2. *if data $D_1, \ldots, D_n \sim g(\cdot)$ were generated such that there exists θ_0 with $g(D) = f(D; \theta_0)$, then $\tilde{\theta} \sim \pi^{(\beta)}(\theta|D)$ for all $1 \leq \beta \leq \infty$ is consistent for θ_0.*

Proof. For part 1), define $B_r(x_0) = \{x \in \mathbb{R}^p : |x - x_0| < r\}$. Theorem 4 of [64] applied to βD-Bayes posterior proves that

$$\int_{B_\varepsilon(\theta_0^{(\beta)})} \pi^{(\beta)}(\theta|D) d\theta \underset{n\to\infty}{\longrightarrow} 1$$

for all $\varepsilon > 0$. This is enough to show that for $\tilde{\theta} \sim \pi^{(\beta)}(\theta|D) \to \theta_0^{(\beta)}$ in probability.

For part 2), note that if $g(D) = f(D; \theta_0)$, then for all $1 \leq \beta \leq \infty$

$$\begin{aligned}\theta_0^{(\beta)} :&= \arg\min_{\theta\in\Theta} \mathbb{E}_g \left[\ell^{(\beta)}(D; f(\cdot; \theta))\right] \\ &= \arg\min_{\theta\in\Theta} D_B^{(\beta)}(g||f(\cdot; \theta)) \\ &= \theta_0.\end{aligned}$$

□

A.2.4 Proof of Proposition 1

Proposition 1 (Asymptotic efficiency)**.** *Under the conditions of Theorem 4 of [64], $\tilde{\theta} \sim \pi^{(\beta)}(\theta|x)$ is asymptotically distributed as $\sqrt{n}(\tilde{\theta} - \theta_0^{(\beta)}) \overset{\text{weakly}}{\longrightarrow} \mathcal{N}(0, (H_0^{(\beta)})^{-1} K_0^{(\beta)} (H_0^{(\beta)})^{-1} + (H_0^{(\beta)})^{-1})$, where $K_0^{(\beta)}$ and $H_0^{(\beta)}$ are defined in Appendix A.1.*

Proof. Let $\tilde{\theta} \sim \pi^{(\beta)}(\theta|D)$. By the Bernstein-von Mises theorem [64] applied to βD-Bayes in (5),

$$\sqrt{n}(\tilde{\theta} - \hat{\theta}_n^{(\beta)}) \to \mathcal{N}(0, (H_0^{(\beta)})^{-1}).$$

By the asymptotic normality of $\hat{\theta}_n^{(\beta)}$ [10], we have that

$$\sqrt{n}(\hat{\theta}_n^{(\beta)} - \theta_0^{(\beta)}) \to^D \mathcal{N}(0, (H_0^{(\beta)})^{-1} K_0^{(\beta)} (H_0^{(\beta)})^{-1})$$

for $K_0 := \left(\frac{\partial}{\partial \theta_i} \mathbb{E}_D \left[\ell^{(\beta)}(D; f(\cdot; \theta_0^{(\beta)}))\right] \frac{\partial}{\partial \theta_j} \mathbb{E}_D \left[\ell^{(\beta)}(D; f(\cdot; \theta_0^{(\beta)}))\right]\right)_{i,j}$. The result then comes from the asymptotic independence of $\tilde{\theta} - \hat{\theta}_n^{(\beta)}$ and $\hat{\theta}_n^{(\beta)}$ [see e.g. 80] □

A.2.5 Proof of Proposition 2

Proposition 2 (DP-MCMC methods for the βD-Bayes-Posterior)**.** *Under Condition 1, the penalty algorithm of [Algorithm 1; 82],* DP-HMC *of [Algorithm 1; 72] and* DP-*Fast MH of [Algorithm 2; 84] and under further Condition 2* DP-SGLD *of [Algorithm 1; 56] can be used to produce (ϵ, δ)-*DP *estimation from the βD-Bayes posterior with $\delta > 0$ without requiring the clipping of any gradients.*

Condition 2 (Boundedness of the model density/mass function gradient)**.** *The model density or mass function $f(\cdot; \theta)$ is such that there exists $0 < G^{(\beta)} < \infty$ such that $\left|\nabla_\theta f(D; \theta) \times f(D; \theta)^{\beta-2}\right| \leq G^{(\beta)}, \forall \theta \in \Theta$.*

Proof. Algorithm 1 of [82], Algorithm 1 of [72] and Algorithm 2 of [84] requires a posterior whose log-likelihood has bounded sensitivity. For βD-Bayes posterior, this requires βD-Bayes-loss has bounded sensitvity which is provided by Condition 1 and Lemma 1.

Algorithm 1 of [56] requires a posterior whose log-likelihood has bounded gradient. For βD-Bayes posterior, this requires βD-Bayes-loss to have bounded gradient:

$$\begin{aligned}|\nabla_\theta \ell^{(\beta)}(D;\theta)| &= \nabla_\theta f(D;\theta) \times f(D;\theta)^{\beta-2} - \int \nabla_\theta f(D;\theta) \times f(D;\theta)^{\beta-1} dD \\ &= \nabla_\theta f(D;\theta) \times f(D;\theta)^{\beta-2} - \int \nabla_\theta f(D;\theta) \times f(D;\theta)^{\beta-2} \times f(D;\theta) dD \\ &\leq \max\{G^{(\beta)}, G^{(\beta)} M\},\end{aligned}$$

assuming we can interchange integration and differentiation and as $\left|\nabla_\theta f(D;\theta) \times f(D;\theta)^{\beta-2}\right| \leq G^{(\beta)}$ by Condition 2 not requiring the clipping of any gradients. □

A.3 Related work

Here, we would like to extend our discussion of two important areas within the related work.

A.3.1 Differentially private logistic regression

Chaudhuri et al. [19] propose a regularised DP logistic regression, solving (1). (1) adds the regulariser to the average loss and as a result, the impact of the regulariser does not diminish as $n \to \infty$. Even though the scale of the Laplace noise decreases as n grows, Chaudhuri et al. [19] consistently estimate a parameter that is not the data generating parameter. Alternatively, one could choose a regulariser $\lambda' := \frac{\lambda}{n}$ whose influence decreases as n grows. This would allow for unbiased inference as $n \to \infty$ (assuming a Bayesian model with corresponding prior distribution), but the n cancels in the scale of the Laplace noise and therefore the perturbation scale does not decrease in n, and the estimator is inconsistent. Choosing instead $\lambda' := \frac{\lambda}{n^r}$ with $0 < r < 1$, would help in constructing unbiased and consistent estimators. In our experiments, we did not find this choice to help.

A.3.2 Differentially private Monte Carlo methods

Wang et al. [80] propose using Stochastic Gradient Langevin Dynamics [SGLD; 81] with a modified burn-in period and bounded step-size to provide DP sampling when the log-likelihood has bounded gradient. Li et al. [56] improve upon [80], taking advantage of the moments accountant [1] to allow for a larger step-size and faster mixing for non-convex target posteriors. Foulds et al. [28] extend their privatisation of sufficient statistics to a Gibbs sampling setting where the conditional posterior distribution for a Gibbs update is from the exponential family. Yıldırım and Ermiş [82] use the penalty algorithm which adds noise to the log of the Metropolis-Hastings acceptance probability. Heikkilä et al. [38] use Barker's acceptance test [8, 75] and provide RDP guarantees. Räisä et al. [72] derive DP-HMC also using the penalty algorithm. Zhang and Zhang [84] propose a random batch size implementation of Metropolis-Hasting for a general proposal distribution that takes advantage of the inherent randomness of Metropolis-Hasting and is asymptotically exact. Lastly, Awan and Rao [7] consider DP rejection sampling.

A.4 Attack optimality

Remark 1. *Let $p(\tilde\theta|D)$ be the density of the privacy mechanism—i.e the Laplace density for [19] or the posterior (i.e. Equations 2,3) for* OPS. *An attacker estimating $\mathcal{M}(\tilde\theta, D, D') = \frac{p(\tilde\theta|D')}{(p(\tilde\theta|D)+p(\tilde\theta|D'))}$ is Bayes optimal. For* OPS, *$\mathcal{M}(\tilde\theta, D, D') = \exp\{\ell(D_l'; f(\cdot;\tilde\theta)) - \ell(D_l;\tilde\theta)\} \int \exp\{\ell(D_l; f(\cdot;\theta)) - \ell(D_l'; f(\cdot;\theta))\}\pi(\theta|D)d\theta$ where D, D' s.t. $D \setminus D' = \{D_l\}$ and $D' \setminus D = \{D_l'\}$ (see Appendix A.4).*

The privacy attacks outlined in Section 4 require the calculation of

$$\begin{aligned}\mathcal{M}(\tilde\theta, D, D') := p(m=1;\tilde\theta, D, D') &= p(\tilde\theta|D')/(p(\tilde\theta|D)) + p(\tilde\theta|D')) \\ &= 1/(p(\tilde\theta|D)/p(\tilde\theta|D') + 1)\end{aligned}$$

by Bayes Theorem. For [19], it is

$$p(\tilde\theta|D) = \mathcal{L}\left(\hat\theta(D), \frac{2}{n\lambda\epsilon}\right),$$

where $\hat\theta(D)$ was defined in (1).

For the OPS methods, Minami et al. [65] and βD-Bayes, $p(\tilde{\theta}|D)$ is the posterior

$$p(\tilde{\theta}|D) = \pi^{(\ell)}(\tilde{\theta}|D) \propto \pi(\theta)\exp\{-\sum_{i=1}^{n} \ell(D_i;\theta)\}$$

where for [65] $\ell(D_i; f(\cdot;\theta)) = -w \log f(D_i;\theta)$, and for βD-Bayes $\ell(D_i; f(\cdot;\theta)) = \ell^{(\beta)}(D_i; f(\cdot;\theta))$ given in (4). Without loss of generality, index observations within D and D' such that $D \setminus D' = \{D_l\}$ and $D' \setminus D = \{D_l'\}$. Then,

$$\begin{aligned}
\frac{\tilde{\pi}(\tilde{\theta}|D)}{\tilde{\pi}(\tilde{\theta}|D')} &= \frac{\pi(\tilde{\theta})\exp\{-\sum_{i=1}^{n}\ell(D_i; f(\cdot;\tilde{\theta}))\}}{\int \pi(\theta)\exp\{-\sum_{i=1}^{n}\ell(D_i; f(\cdot;\theta))\}d\theta} \Big/ \frac{\pi(\tilde{\theta})\exp\{-\sum_{i=1}^{n}\ell(D_i'; f(\cdot;\tilde{\theta}))\}}{\int \pi(\theta)\exp\{-\sum_{i=1}^{n}\ell(D_i'; f(\cdot;\theta))\}d\theta} \\
&= \exp\{\ell(D_l'; f(\cdot;\tilde{\theta})) - \ell(D_l; f(\cdot;\tilde{\theta}))\} \int \frac{\pi(\theta)\exp\{-\sum_{i=1}^{n}\ell(D_i'; f(\cdot;\theta))\}}{\int \pi(\theta)\exp\{-\sum_{i=1}^{n}\ell(D_i; f(\cdot;\theta))\}d\theta} d\theta \\
&= \exp\{\ell(D_l'; f(\cdot;\tilde{\theta})) - \ell(D_l;\tilde{\theta})\} \\
&\qquad \int \exp\{\ell(D_l;\theta) - \ell(D_l'; f(\cdot;\theta))\} \frac{\pi(\theta)\exp\{-\sum_{i=1}^{n}\ell(D_i; f(\cdot;\theta))\}}{\int \pi(\theta)\exp\{-\sum_{i=1}^{n}\ell(D_i; f(\cdot;\theta))\}d\theta} d\theta \\
&= \exp\{\ell(D_l'; f(\cdot;\tilde{\theta})) - \ell(D_l; f(\cdot;\tilde{\theta}))\} \int \exp\{\ell(D_l; f(\cdot;\theta)) - \ell(D_l'; f(\cdot;\theta))\}\pi(\theta|D)d\theta \\
&\approx \exp\{\ell(D_l'; f(\cdot;\tilde{\theta})) - \ell(D_l; f(\cdot;\tilde{\theta}))\} \frac{1}{N}\sum_{j}^{N} \exp\{\ell(D_l; f(\cdot;\theta_j)) - \ell(D_l'; f(\cdot;\theta_j))\},
\end{aligned}$$

where $\{\theta_j\}_{j=1}^N \sim \pi(\theta|D)$. The adversary only needs to sample from the posterior based on dataset D to be able to estimate $\mathcal{M}(\tilde{\theta}, D, D')$ for all D' differing from D in only one index l.

Appendix B Additional experimental details and results

Additional experimental details Unless otherwise specified, we choose $d = 2$ in the simulated experiments. The MCMC methods are run for 1000 warm-up steps, and 100 iterations. DPSGD is run for $15 + \lfloor \epsilon \rfloor$ epochs, with clipping norm 1, batch size 100, and learning rate of 10^{-2}. All other implementation details can be found on `https://anonymous.4open.science/r/beta-bayes-ops-6626`.

Neural network classification Similarly to neural network regression, we can use βD-Bayes for neural network classification. As we see in Figure 6, βD-Bayes regularly outperforms DPSGD for $\epsilon > 0.2$ on simulated and real data, except on `abalone`.

Sensitivity in number of features Please refer to Figure 7 for the sensitivity of the private methods w.r.t. the number of features in the data set. We see that the RMSE of the data generating parameter θ (divided by the number of dimensions of θ) increases. The reason for this is two-fold: 1) The methods of [19] and [65] provide their privacy guarantees w.r.t. the number of features. While more noise has to be added for [19], the influence of the prior increases for [65] when the number of features increases for a fixed privacy budget. 2) A single sample from a posterior is of higher variance the higher-dimensional the posterior is, negatively influencing OPS methods such as [65] and βD-Bayes.

Membership inference attacks For $\epsilon \in \{0.2, 1, 2, 7, 10, 20\}$, we run 10,000 rounds of the attack presented in Section 4. In Figure 8, we use the approach presented by [44] to estimate a lower bound on ϵ given the false positive and negative rates of the attacks. Note that these lower bounds are unrealistic for $\epsilon < 1$. We see that, for any RMSE value, βD-Bayes achieves a lower practical bound on ϵ than [19], which gives exact privacy guarantees.

Compute While the final experimental results can be run within approximately two hours on a single Intel(R) Xeon(R) Gold 5118 CPU @ 2.30GHz core, the complete compute needed for the final results, debugging runs, and sweeps amounts to around 11 days.

Licenses The UCI data sets are licensed under Creative Commons Attribution 4.0 International license (CC BY 4.0).

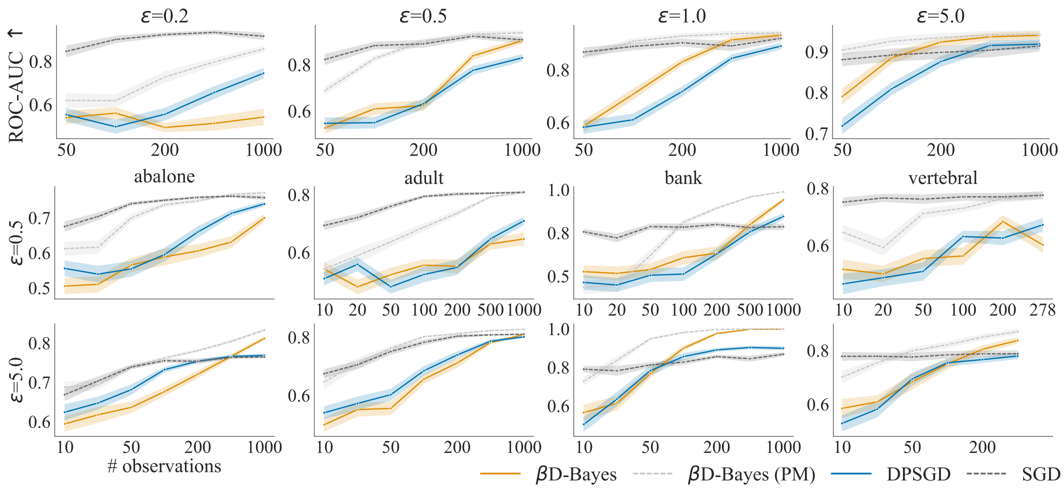

Figure 6: Test set predictive ROC-AUC of DP estimation for neural network classification as the number of observations n increases on simulated and UCI data.

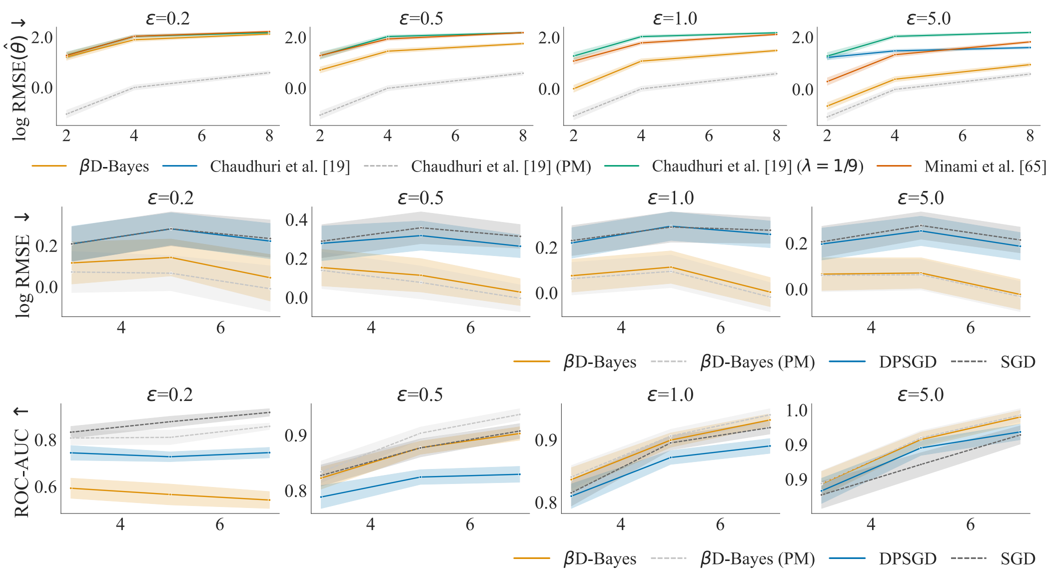

Figure 7: Parameter log RMSE of DP logistic regression (**first row**), test set predictive log RMSE of DP neural network regression (**second row**), and test set ROC-AUC of DP neural network classification (**third row**) as the number of features d increases on simulated data with $n = 1000$.

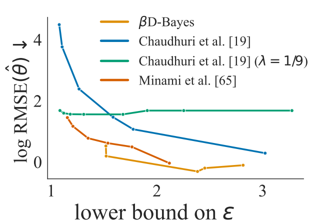

Figure 8: Lower bound on ϵ against log RMSE. Points correspond to values of ϵ.

References

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 308–318, 2016.

[2] Haleh Akrami, Wenhui Cui, Anand A Joshi, and Richard M. Leahy. Learning from imperfect training data using a robust loss function: application to brain image segmentation, 2022.

[3] Haleh Akrami, Anand A Joshi, Jian Li, Sergül Aydöre, and Richard M Leahy. A robust variational autoencoder using beta divergence. *Knowledge-Based Systems*, 238:107886, 2022.

[4] Matias Altamirano, François-Xavier Briol, and Jeremias Knoblauch. Robust and scalable Bayesian online changepoint detection. *arXiv preprint arXiv:2302.04759*, 2023.

[5] Marco Avella-Medina. The role of robust statistics in private data analysis. *Chance*, 33(4): 37–42, 2020.

[6] Marco Avella-Medina. Privacy-preserving parametric inference: a case for robust statistics. *Journal of the American Statistical Association*, 116(534):969–983, 2021.

[7] Jordan Awan and Vinayak Rao. Privacy-aware rejection sampling. *Journal of Machine Learning Research*, 24(74):1–32, 2023.

[8] Anthony Alfred Barker. Monte Carlo calculations of the radial distribution functions for a proton? electron plasma. *Australian Journal of Physics*, 18(2):119–134, 1965.

[9] Ayanendranath Basu, Ian R Harris, Nils L Hjort, and MC Jones. Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3):549–559, 1998.

[10] Ayanendranath Basu, Hiroyuki Shioya, and Chanseok Park. *Statistical inference: the minimum distance approach*. CRC Press, 2011.

[11] Robert H Berk et al. Limiting behavior of posterior distributions when the model is incorrect. *The Annals of Mathematical Statistics*, 37(1):51–58, 1966.

[12] Garrett Bernstein and Daniel R Sheldon. Differentially private Bayesian inference for exponential families. *Advances in Neural Information Processing Systems*, 31, 2018.

[13] Garrett Bernstein and Daniel R Sheldon. Differentially private Bayesian linear regression. *Advances in Neural Information Processing Systems*, 32, 2019.

[14] Michael Betancourt. A conceptual introduction to hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*, 2017.

[15] PG Bissiri, CC Holmes, and Stephen G Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2016.

[16] Lev M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217, 1967.

[17] Bob Carpenter, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 20, 2016.

[18] Kamalika Chaudhuri and Daniel Hsu. Convergence rates for differentially private statistical estimation. In *Proceedings of the... International Conference on Machine Learning. International Conference on Machine Learning*, volume 2012, page 1327. NIH Public Access, 2012.

[19] Kamalika Chaudhuri, Claire Monteleoni, and Anand D Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12(3), 2011.

[20] Wenhui Cui, Haleh Akrami, Anand A Joshi, and Richard M Leahy. Semi-supervised learning using robust loss. *arXiv preprint arXiv:2203.01524*, 2022.

[21] Soham De, Leonard Berrada, Jamie Hayes, Samuel L Smith, and Borja Balle. Unlocking high-accuracy differentially private image classification through scale. *arXiv preprint arXiv:2204.13650*, 2022.

[22] Christos Dimitrakakis, Blaine Nelson, Aikaterini Mitrokotsa, and Benjamin IP Rubinstein. Robust and private Bayesian inference. In *Algorithmic Learning Theory: 25th International Conference, ALT 2014, Bled, Slovenia, October 8-10, 2014. Proceedings 25*, pages 291–305. Springer, 2014.

[23] Christos Dimitrakakis, Blaine Nelson, Zuhe Zhang, Aikaterini Mitrokotsa, and Benjamin IP Rubinstein. Differential privacy for Bayesian inference through posterior sampling. *The Journal of Machine Learning Research*, 18(1):343–381, 2017.

[24] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017. URL http://archive.ics.uci.edu/ml.

[25] Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, 1987.

[26] Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 371–380, 2009.

[27] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

[28] James Foulds, Joseph Geumlek, Max Welling, and Kamalika Chaudhuri. On the theory and practice of privacy-preserving Bayesian data analysis. *arXiv preprint arXiv:1603.07294*, 2016.

[29] Hironori Fujisawa and Shinto Eguchi. Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis*, 99(9):2053–2081, 2008.

[30] Arun Ganesh and Kunal Talwar. Faster differentially private samplers via rényi divergence analysis of discretized Langevin MCMC. *Advances in Neural Information Processing Systems*, 33:7222–7233, 2020.

[31] Joseph Geumlek, Shuang Song, and Kamalika Chaudhuri. Renyi differential privacy mechanisms for posterior sampling. *Advances in Neural Information Processing Systems*, 30, 2017.

[32] Sahra Ghalebikesabi, Harry Wilde, Jack Jewson, Arnaud Doucet, Sebastian Vollmer, and Chris Holmes. Mitigating statistical bias within differentially private synthetic data. In *Uncertainty in Artificial Intelligence*, pages 696–705. PMLR, 2022.

[33] Abhik Ghosh and Ayanendranath Basu. Robust Bayes estimation using the density power divergence. *Annals of the Institute of Statistical Mathematics*, 68(2):413–437, 2016.

[34] Paolo Girardi, Luca Greco, Valentina Mameli, Monica Musio, Walter Racugno, Erlis Ruli, and Laura Ventura. Robust inference for non-linear regression models from the tsallis score: Application to coronavirus disease 2019 contagion in italy. *Stat*, 9(1):e309, 2020.

[35] Federica Giummolè, Valentina Mameli, Erlis Ruli, and Laura Ventura. Objective Bayesian inference with proper scoring rules. *Test*, 28(3):728–755, 2019.

[36] Frank R Hampel, Elvezio M Ronchetti, Peter J Rousseeuw, and Werner A Stahel. *Robust statistics: the approach based on influence functions*, volume 114. John Wiley & Sons, 2011.

[37] S Harini, K Nivedha, Selva Keerthana BG, R Gokul, BS Jayasree, et al. Data anomaly detection in wireless sensor networks using β-variational autoencoder. In *2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS)*, pages 631–636. IEEE, 2023.

[38] Mikko Heikkilä, Joonas Jälkö, Onur Dikmen, and Antti Honkela. Differentially private Markov chain Monte Carlo. *Advances in Neural Information Processing Systems*, 32, 2019.

[39] Matthew D Hoffman and Andrew Gelman. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

[40] Naoise Holohan, Stefano Braghin, Pól Mac Aonghusa, and Killian Levacher. Diffprivlib: the ibm differential privacy library. *arXiv preprint arXiv:1907.02444*, 2019.

[41] Antti Honkela, Mrinal Das, Arttu Nieminen, Onur Dikmen, and Samuel Kaski. Efficient differentially private learning improves drug sensitivity prediction. *Biology direct*, 13:1–12, 2018.

[42] Peter J Huber and EM Ronchetti. Robust statistics, series in probability and mathematical statistics, 1981.

[43] Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(Apr):695–709, 2005.

[44] Matthew Jagielski, Jonathan Ullman, and Alina Oprea. Auditing differentially private machine learning: How private is private sgd? *Advances in Neural Information Processing Systems*, 33: 22205–22216, 2020.

[45] Joonas Jälkö, Onur Dikmen, and Antti Honkela. Differentially private variational inference for non-conjugate models. *arXiv preprint arXiv:1610.08749*, 2016.

[46] Jack Jewson, Jim Smith, and Chris Holmes. Principles of Bayesian inference using general divergence criteria. *Entropy*, 20(6):442, 2018.

[47] Jack Jewson, Jim Q Smith, and Chris Holmes. On the stability of general Bayesian inference. *arXiv preprint arXiv:2301.13701*, 2023.

[48] Jeremias Knoblauch, Jack Jewson, and Theodoros Damoulas. Doubly robust Bayesian inference for non-stationary streaming data using β-divergences. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 64–75, 2018.

[49] Jeremias Knoblauch, Jack Jewson, and Theodoros Damoulas. An optimization-centric view on Bayes' rule: Reviewing and generalizing variational inference. *Journal of Machine Learning Research*, 23(132):1–109, 2022.

[50] Hyeyoung Ko, Suyeon Lee, Yoonseo Park, and Anna Choi. A survey of recommendation systems: recommendation models, techniques, and application fields. *Electronics*, 11(1):141, 2022.

[51] Michael R Kosorok. *Introduction to empirical processes and semiparametric inference.* Springer, 2008.

[52] Tejas Kulkarni, Joonas Jälkö, Antti Koskela, Samuel Kaski, and Antti Honkela. Differentially private Bayesian inference for generalized linear models. In *International Conference on Machine Learning*, pages 5838–5849. PMLR, 2021.

[53] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[54] Anthony Lee, Arnaud Doucet, and Krzysztof Łatuszyński. Perfect simulation using atomic regeneration with application to sequential Monte Carlo. *arXiv preprint arXiv:1407.5770*, 2014.

[55] Jing Lei. Differentially private m-estimators. *Advances in Neural Information Processing Systems*, 24, 2011.

[56] Bai Li, Changyou Chen, Hao Liu, and Lawrence Carin. On connecting stochastic gradient MCMC and differential privacy. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 557–566. PMLR, 2019.

[57] Mengchu Li, Thomas B Berrett, and Yi Yu. On robustness and local differential privacy. *arXiv preprint arXiv:2201.00751*, 2022.

[58] Samuel Livingstone, Michael Betancourt, Simon Byrne, and Mark Girolami. On the geometric ergodicity of hamiltonian Monte Carlo. 2019.

[59] Lauri Lode et al. Sub-sampled and differentially private hamiltonian Monte Carlo. 2019.

[60] Takuo Matsubara, Jeremias Knoblauch, François-Xavier Briol, Chris Oates, et al. Robust generalised Bayesian inference for intractable likelihoods. *arXiv preprint arXiv:2104.07359*, 2021.

[61] H Brendan McMahan, Galen Andrew, Ulfar Erlingsson, Steve Chien, Ilya Mironov, Nicolas Papernot, and Peter Kairouz. A general approach to adding differential privacy to iterative training procedures. *arXiv preprint arXiv:1812.06210*, 2018.

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

[63] Minami Mihoko and Shinto Eguchi. Robust blind source separation by beta divergence. *Neural computation*, 14(8):1859–1886, 2002.

[64] Jeffrey W Miller. Asymptotic normality, concentration, and coverage of generalized posteriors. *The Journal of Machine Learning Research*, 22(1):7598–7650, 2021.

[65] Kentaro Minami, HItomi Arai, Issei Sato, and Hiroshi Nakagawa. Differential privacy without sensitivity. *Advances in Neural Information Processing Systems*, 29, 2016.

[66] Darakhshan J Mir. *Differential privacy: an exploration of the privacy-utility landscape*. Rutgers The State University of New Jersey, School of Graduate Studies, 2013.

[67] Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*, pages 263–275. IEEE, 2017.

[68] Milad Nasr, Shuang Songi, Abhradeep Thakurta, Nicolas Papernot, and Nicholas Carlin. Adversary instantiation: Lower bounds for differentially private machine learning. In *2021 IEEE Symposium on security and privacy (SP)*, pages 866–882. IEEE, 2021.

[69] Radford M Neal. *Bayesian learning for neural networks*, volume 118. Springer Science & Business Media, 2012.

[70] Mijung Park, James Foulds, Kamalika Chaudhuri, and Max Welling. VariationalBayes in private settings (vips). *Journal of Artificial Intelligence Research*, 68:109–157, 2020.

[71] Melanie F Pradier, Niranjani Prasad, Paidamoyo Chapfuwa, Sahra Ghalebikesabi, Max Ilse, Steven Woodhouse, Rebecca Elyanow, Javier Zazo, Javier Gonzalez, Julia Greissl, et al. Airiva: A deep generative model of adaptive immune repertoires. *arXiv preprint arXiv:2304.13737*, 2023.

[72] Ossi Räisä, Antti Koskela, and Antti Honkela. Differentially private hamiltonian Monte Carlo. *arXiv preprint arXiv:2106.09376*, 2021.

[73] Alexandre Sablayrolles, Matthijs Douze, Cordelia Schmid, Yann Ollivier, and Hervé Jégou. White-box vs black-box: Bayes optimal strategies for membership inference. In *International Conference on Machine Learning*, pages 5558–5567. PMLR, 2019.

[74] Jeremy Seeman, Matthew Reimherr, and Aleksandra Slavković. Exact privacy guarantees for Markov chain implementations of the exponential mechanism with artificial atoms. *Advances in Neural Information Processing Systems*, 34:13125–13136, 2021.

[75] Daniel Seita, Xinlei Pan, Haoyu Chen, and John Canny. An efficient minibatch acceptance test for Metropolis-Hastings. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 5359–5363, 2018.

[76] Adam Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 813–822, 2011.

[77] Theresa Stadler, Bristena Oprisanu, and Carmela Troncoso. Synthetic data–anonymisation groundhog day. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 1451–1468, 2022.

[78] Shonosuke Sugasawa. Robust empirical Bayes small area estimation with density power divergence. *Biometrika*, 107(2):467–480, 2020.

[79] Stephen G Walker. Bayesian inference with misspecified models. *Journal of Statistical Planning and Inference*, 143(10):1621–1633, 2013.

[80] Yu-Xiang Wang, Stephen Fienberg, and Alex Smola. Privacy for free: Posterior sampling and stochastic gradient Monte Carlo. In *International Conference on Machine Learning*, pages 2493–2502. PMLR, 2015.

[81] Max Welling and Yee W Teh. Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 681–688, 2011.

[82] Sinan Yıldırım and Beyza Ermiş. Exact MCMC with differentially private moves: revisiting the penalty algorithm in a data privacy framework. *Statistics and Computing*, 29:947–963, 2019.

[83] Qiyiwen Zhang, Zhiqi Bu, Kan Chen, and Qi Long. Differentially private Bayesian neural networks on accuracy, privacy and reliability. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part IV*, pages 604–619. Springer, 2023.

[84] Wanrong Zhang and Ruqi Zhang. DP-Fast MH: Private, fast, and accurate Metropolis-Hastings for large-scale Bayesian inference. *arXiv preprint arXiv:2303.06171*, 2023.

[85] Zuhe Zhang, Benjamin Rubinstein, and Christos Dimitrakakis. On the differential privacy of Bayesian inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

[86] Shijie Zheng. *The differential privacy of Bayesian inference*. PhD thesis, Harvard University, 2015.