(ε, δ)-Differential Privacy of Gibbs Posteriors

Kentaro MINAMI
Department of Mathematical Informatics
Graduate School of Information Science and Technology
The University of Tokyo
kentaro_minami@mist.i.u-tokyo.ac.jp

Hiromi ARAI
Information Technology Center
The University of Tokyo
arai@dl.itc.u-tokyo.ac.jp

Issei SATO
Department of Complexity Science and Engineering
Graduate School of Frontier Sciences
The University of Tokyo
sato@k.u-tokyo.ac.jp

Abstract

The exponential mechanism is a general method to construct a randomized estimator that satisfies $(\varepsilon, 0)$-differential privacy. Recently, Wang et al. [12] showed that the Gibbs posterior, which is a data-dependent probability distribution that contains the Bayesian posterior, is essentially equivalent to the exponential mechanism under certain boundedness conditions on the loss function. While the exponential mechanism provides a way to build an $(\varepsilon, 0)$-differential private algorithm, it requires boundedness of the loss function, which is quite stringent for some learning problems. In this paper, we focus on (ε, δ)-differential privacy of Gibbs posteriors with convex and Lipschitz loss functions. Our result extends the classical exponential mechanism, allowing the loss functions to have an unbounded sensitivity.

1 Introduction

Differential privacy is a notion of privacy that provides a statistical measure of privacy protection for randomized statistics. In the field of privacy-preserving learning, constructing estimators that satisfy (ε, δ)-differential privacy is a fundamental problem. In recent years, differentially private algorithms for various statistical learning problems have been developed [4, 8, 1].

At an abstract level, the estimator construction procedure in statistical learning can be regarded as the following problem. Given a dataset $D_n = \{x_1, \dots, x_n\}$, a statistician chooses a *parameter* θ that minimizes a certain cost function $\mathcal{L}(\theta, D_n)$. A typical example of cost function is the empirical risk function, that is, a sum of *loss function* $\ell(\theta, x_i)$ evaluated at each sample point $x_i \in D_n$. For example, the maximum likelihood estimator (MLE) is given by the minimizer of empirical risk with loss function $\ell(\theta, x) = -\log p(x \mid \theta)$.

To achieve a differentially private estimator, one natural idea is to construct an algorithm based on a *posterior sampling*, namely drawing a sample from a certain data-dependent probability distribution. The exponential mechanism [10], which can be regarded as a posterior sampling, provides a general

Table 1: Regularity conditions for (ε, δ)-differential privacy of the Gibbs posterior. Instead of the boundedness of the loss function, our analysis in Theorem 7 requires its Lipschitz property and convexity. Unlike the classical exponential mechanism, our result explains "shrinkage effect" or "contraction effect", namely, the upper bound for β depends on the concavity of the prior π and the size of the dataset n.

	(ε, δ)	Loss function ℓ	Prior π	Shrinkage
Exponential mechanism [10]	$\delta = 0$	Bounded sensitivity	Arbitrary	No
Theorem 7	$\delta > 0$	Lipschitz and convex	Log-concave	Yes
Theorem 10	$\delta > 0$	Bounded, Lipschitz and strongly convex	Log-concave	Yes

method to construct a randomized estimator that satisfies $(\varepsilon, 0)$-differential privacy. The probability density of the output of the exponential mechanism is proportional to $\exp(-\beta\mathcal{L}(\theta, D_n))\pi(\theta)$, where $\pi(\theta)$ is an arbitrary prior density function, and $\beta > 0$ is a parameter that controls the degree of concentration. Obviously, the resulting distribution is highly concentrated around the minimizer $\theta^* \in \operatorname{argmin}_\theta \mathcal{L}(\theta, D_n)$. Note that most differential private algorithms involve a procedure to add some noise (e.g. the Laplace mechanism [6], objective perturbation [4, 8], and gradient perturbation [1]), while the posterior sampling explicitly designs the density of the output distribution.

We define the density of the *Gibbs posterior distribution* as

$$G_\beta(\theta \mid D_n) := \frac{\exp(-\beta \sum_{i=1}^n \ell(\theta, x_i))\pi(\theta)}{\int \exp(-\beta \sum_{i=1}^n \ell(\theta, x_i))\pi(\theta)\mathrm{d}\theta}. \tag{1}$$

The Gibbs posterior plays important roles in several learning problems, especially in PAC-Bayesian learning theory [3, 13]. In the context of differential privacy, Wang et al. [12] recently pointed out that the Bayesian (Gibbs) posterior, which is a special version of (1) with $\beta = 1$ and a specific loss function, satisfies $(\varepsilon, 0)$-differential privacy because it is equivalent to the exponential mechanism under a certain regularity condition.

In this paper, we study the (ε, δ)-differential privacy of the posterior sampling with $\delta > 0$. In particular, we consider the following statement.

Claim 1. Under a suitable condition on loss function ℓ and prior π, there exists an upper bound $B(\varepsilon, \delta) > 0$, and the Gibbs posterior $G_\beta(\theta \mid D_n)$ with $\beta \leq B(\varepsilon, \delta)$ satisfies (ε, δ)-differential privacy. The value of $B(\varepsilon, \delta)$ does not depend on the boundedness of the loss function.

We point out here the analyzes of $(\varepsilon, 0)$-differential privacy and (ε, δ)-differential privacy with $\delta > 0$ are very different. The largest difference arises in the regularity conditions they require. On one hand, the exponential mechanism essentially requires the boundedness of the loss function to satisfy $(\varepsilon, 0)$-differential privacy. On the other hand, the boundedness is not a necessary condition in (ε, δ)-differential privacy. In this paper, we give a new sufficient condition for (ε, δ)-differential privacy based on the convexity and the Lipschitz property. To our knowledge, the only other work on the (ε, δ)-differential privacy of the posterior sampling is one by Dimitrakakis et al. [5], which requires some modification of the definition of the neighborhood on the database. In contrast, the analysis in this paper does not require any modification of the original definition.

Our analysis widens the application ranges of the exponential mechanism in the following aspects (See also Table 1).

- (Removal of boundedness assumption) If the loss function is unbounded, which is usually the case when the parameter space is unbounded, the Gibbs posterior does not satisfy $(\varepsilon, 0)$-differential privacy in general. Still, in some cases we can build an (ε, δ)-differential private estimator.
- (Tighter evaluation of β) Even when the difference of the loss function is bounded, our analysis can yield a better scheme in determining the appropriate value of β for a given privacy level. Figure 1 shows an example of logistic loss.
- (Shrinkage and contraction effect) Intuitively speaking, the Gibbs posterior becomes robust against a small change of the dataset, if the prior π has a strong shrinkage effect (e.g. a

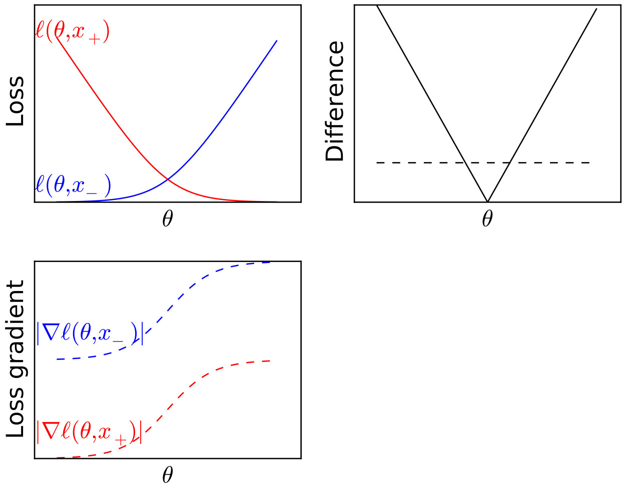

Figure 1: An example of a logistic loss function $\ell(\theta, x) := \log(1+\exp(-y\theta^\top z))$. Considering two points $x_\pm = (z, \pm 1)$, the difference of the loss $|\ell(\theta, x_+) - \ell(\theta, x_-)|$ increases proportionally to the size of the parameter space (solid lines). In this case, the value of the β in the exponential mechanism, which is inversely proportional to the maximum difference of the loss function, becomes very small. On the other hand, the difference of the gradient $|\nabla\ell(\theta, x_+) - \nabla\ell(\theta, x_-)|$ does not exceed twice of the Lipschitz constant (dashed lines). Hence, our analysis based on Lipschitz property does not be influenced by the size of the parameter space.

Gaussian prior with a small variance), or if the size of the dataset n tends to infinity. In our analysis, the upper bound of β depends on π and n, which explains such shrinkage and contraction effects.

2 Preliminary

2.1 Differential privacy

Differential privacy is a notion of privacy that provides a degree of privacy protection in a statistical sense. More precisely, differential privacy formalizes a closeness between output probability distributions that correspond to two adjacent datasets.

In this paper, we assume that a dataset $D = D_n = (x_1, \ldots, x_n)$ is a vector that consists of n points in abstract attribute space $\mathcal{X}$, where each entry $x_i \in \mathcal{X}$ represents information contributed by one individual. Two data sets D, D' are said to be adjacent if $d_H(D, D') = 1$, where d_H is the Hamming distance defined on the space of all possible datasets $\mathcal{X}^d$.

Let Θ be a measurable space. The set of all probability measures on Θ is denoted by $\mathcal{M}^1_+(\Theta)$. A randomized estimator is a map $\rho : \mathcal{X}^n \to \mathcal{M}^1_+(\Theta)$ from the space of datasets to the space of probability measures. We describe the definition of differential privacy in terms of random estimators.

Definition 2 (Differential privacy)**.** Let $\varepsilon > 0$ and $\delta \geq 0$ be given privacy parameters. We say that a random estimator $\rho : \mathcal{X}^n \to \mathcal{M}^1_+(\Theta)$ satisfies (ε, δ)-differential privacy, if for any adjacent datasets $D, D' \in \mathcal{X}^n$, an inequality

$$\rho_D(A) \leq e^\varepsilon \rho_{D'}(A) + \delta \tag{2}$$

holds for every measurable set $A \subset \Theta$. We say that ρ satisfies ε-differential privacy if it satisfies $(\varepsilon, 0)$-differential privacy.

2.1.1 The exponential mechanism

Here, we provide a brief review of the exponential mechanism [10]. For an arbitrary function $f : \Theta \times \mathcal{X}^n \to \mathbb{R}$, we define the sensitivity by

$$\Delta_f := \sup_{\substack{D,D' \in \mathcal{X}^n: \\ d_H(D,D')=1}} \sup_{\theta \in \Theta} |f(\theta, D) - f(\theta, D')|, \tag{3}$$

which is the largest possible difference of two "adjacent" functions $f(\cdot, D)$ and $f(\cdot, D')$ with respect to supremum norm. In short, the basic theorem of the exponential mechanism (Theorem 3) guarantees $(\varepsilon, 0)$-differential privacy of the Gibbs posterior.

Theorem 3 (McSherry and Talwar)**.** Suppose that the sensitivity of the function $\mathcal{L}(\theta, D_n)$ is finite. Let π be an arbitrary base measure on Θ. Take a positive number β so that $\beta \leq \varepsilon / 2\Delta_{\mathcal{L}}$. Then a probability distribution whose density with respect to π is proportional to $\exp(-\beta \mathcal{L}(\theta, D_n))$ satisfies $(\varepsilon, 0)$-differential privacy.

It is convenient to understand the exponential mechanism as a composition of two Lipschitz maps. We define a distance d_{DP} between two probability measures $\mu_1, \mu_2 \in \mathcal{M}^1_+(\Theta)$ by

$$d_{\mathrm{DP}}(\mu_1, \mu_2) := \sup_{A \subset \Theta} |\log \mu_1(A) - \log \mu_2(A)|, \tag{4}$$

where the supremum is taken over measurable sets. $d_{\mathrm{DP}}(\mu_1, \mu_2)$ is defined to be $+\infty$ if μ_1 and μ_2 are not absolutely continuous. Recall that a map between two metric spaces $f : (X, d_X) \to (Y, d_Y)$ is said to be L-Lipschitz, if $d_Y(f(x_1), f(x_2)) \leq L d_X(x_1, x_2)$ holds for all $x_1, x_2 \in X$. It is easy to check that the $(\varepsilon, 0)$-differential privacy of randomized estimator ρ is equivalent to the ε-Lipschitz property as a map between two metric spaces $\rho : \mathcal{X}^n \to \mathcal{M}^1_+(\Theta)$. We define a function space $\mathbb{R}^\Theta := \{f : \Theta \to \mathbb{R}\}$ equipped with supremum distance $d_\infty(f, g) := \sup_\theta |f(\theta) - g(\theta)|$. If the sensitivity $\Delta_{\mathcal{L}}$ is finite, a function-valued function $\mathcal{L} : D_n \mapsto \mathcal{L}(\cdot, D_n)$ is $\Delta_{\mathcal{L}}$-Lipschitz with respect to d_H and d_∞. We define a Gibbs map $G_\beta : \mathbb{R}^\Theta \to \mathcal{M}^1_+(\Theta)$ as follows: given a function f, $G_\beta(f)$ is a probability distribution whose density w.r.t. π is proportional to $\exp(-\beta f)$. We can check that the Gibbs map is 2β-Lipschitz. Eventually, Theorem 3 states that the exponential mechanism is $2\beta\Delta_{\mathcal{L}}$-Lipschitz, because it is a composition of two Lipschitz functions:

$$(\mathcal{X}^n, d_H) \xrightarrow{\mathcal{L}} (\mathbb{R}^\Theta, d_\infty) \xrightarrow{G_\beta} (\mathcal{M}^1_+(\Theta), d_{\mathrm{DP}}). \tag{5}$$

We now consider the particular case that the cost function is given as sum form $\mathcal{L}(\theta, D_n) = \sum_{i=1}^n \ell(\theta, x_i)$. Recently, Wang et al. [12] examined two typical cases in which $\Delta_{\mathcal{L}}$ is finite. The following statement slightly generalizes their result.

Theorem 4 (Wang, et al.)**.** (a) Suppose that the loss function ℓ is bounded by A, namely $|\ell(\theta, x)| \leq A$ holds for all $x \in \mathcal{X}$ and $\theta \in \Theta$. Then $\Delta_{\mathcal{L}} \leq 2A$, and the Gibbs posterior (1) satisfies $(4\beta A, 0)$-differential privacy.

(b) Suppose that for any fixed $\theta \in \Theta$, the difference $|\ell(\theta, x_1) - \ell(\theta, x_2)|$ is bounded by L for all $x_1, x_2 \in \mathcal{X}$. Then $\Delta_{\mathcal{L}} \leq L$, and the Gibbs posterior (1) satisfies $(2\beta L, 0)$-differential privacy.

The condition $\Delta_{\mathcal{L}} < \infty$ requires the boundedness of the loss function in some sense because it is a Lipschitz condition with respect to a degenerated distance d_H. In practice, statistical models of interest do not necessarily satisfy the boundedness condition of Theorem 4. Thus, in the next section, we will consider an alternative argument for (ε, δ)-differential privacy so that it does not require such boundedness conditions.

3 Main Results

In this section, we state our main results in the form of Claim 1.

There is a well-known sufficient condition for the (ε, δ)-differential privacy:

Theorem 5 (See for example Lemma 2 of [7])**.** Let $\varepsilon > 0$ and $\delta > 0$ be privacy parameters. Suppose that a randomized estimator $\rho : \mathcal{X}^n \to \mathcal{M}^1_+(\Theta)$ satisfies a tail-bound inequality of log-density ratio

$$\rho_D \left\{ \log \frac{\mathrm{d}\rho_D}{\mathrm{d}\rho_{D'}} \geq \varepsilon \right\} \leq \delta \tag{6}$$

for every adjacent pair of datasets D, D'. Then ρ satisfies (ε, δ)-differential privacy.

To control the tail behavior (6) of the log-density ratio function $\log \frac{\mathrm{d}\rho_D}{\mathrm{d}\rho_{D'}}$, we consider the concentration around its expectation. Roughly speaking, inequality (6) holds if there exists an increasing function $\alpha(t)$ that satisfies an inequality

$$\forall t > 0, \quad \rho_D \left\{ \log \frac{\mathrm{d}\rho_D}{\mathrm{d}\rho_{D'}} \geq D_{\mathrm{KL}}(\rho_D, \rho_{D'}) + t \right\} \leq \exp(-\alpha(t)), \tag{7}$$

where $\log \frac{\mathrm{d}G_{\beta,D}}{\mathrm{d}G_{\beta,D'}}$ is the log-density ratio function, and $D_{\mathrm{KL}}(\rho_D, \rho_{D'}) := \mathbb{E}_{\rho_D} \log \frac{\mathrm{d}\rho_D}{\mathrm{d}\rho_{D'}}$ is the Kullback-Leibler (KL) divergence. Suppose that the Gibbs posterior $G_{\beta,D}$, whose density $G(\theta \mid D)$ is defined by (1), satisfies an inequality (7) for a certain $\alpha(t) = \alpha(t, \beta)$. Then $G_{\beta,D}$ satisfies (6) if there exist $\beta, t > 0$ that satisfy the following two conditions.

1. KL-divergence bound: $D_{\mathrm{KL}}(G_{\beta,D}, G_{\beta,D'}) + t \leq \varepsilon$
2. Tail-probability bound: $\exp(-\alpha(t, \beta)) \leq \delta$

In Section 3.1, we consider a regularity condition that the loss function is Lipschitz and convex, in which the sensitivity is allowed to be unbounded. In Section 3.2, we consider another condition for bounded and strongly convex loss functions. This provides an alternative analysis to the exponential mechanism, and some nice properties (e.g. dependency on the sample size and prior) are obtained.

3.1 Convex and Lipschitz loss

Here, we examine the case in which the loss function ℓ is Lipschitz and convex, and the parameter space Θ is the entire Euclidean space $\mathbb{R}^d$. Due to the unboundedness of the domain, the sensitivity $\Delta_{\mathcal{L}}$ can be infinite, in which case the exponential mechanism cannot be applied.

Recall that a C^2-function f defined on a subset of $\mathbb{R}^d$ is said to be $m(>0)$-strongly convex, if the eigenvalues of the Hessian $\nabla^2 f$ are bounded by m from below.

Assumption 6. (i) $\Theta = \mathbb{R}^d$.

(ii) For any $x \in \mathcal{X}$, $\ell(\cdot, x)$ is twice differentiable, L-Lipschitz, and convex.

(iii) $-\log \pi(\cdot)$ is twice differentiable and m_π-strongly convex.

In Assumption 6, the loss function $\ell(\cdot, x)$ and the difference $|\ell(\cdot, x_1) - \ell(\cdot, x_2)|$ can be unbounded. Thus, the classical argument of the exponential mechanism in Section 2.1.1 cannot be applied. Nevertheless, our analysis shows that the Gibbs posterior satisfies (ε, δ)-differential privacy.

Theorem 7. Let $\beta \in (0, 1]$ be a fixed parameter, and $D, D' \in \mathcal{X}^n$ be an adjacent pair of datasets. Under Assumption 6, inequality

$$G_{\beta,D} \left\{ \log \frac{\mathrm{d}G_{\beta,D}}{\mathrm{d}G_{\beta,D'}} \geq \varepsilon \right\} \leq \exp\left(-\frac{m_\pi}{8L^2\beta^2} \left(\varepsilon - \frac{2L^2\beta^2}{m_\pi} \right)^2 \right) \tag{8}$$

holds for any $\varepsilon > \frac{2L^2\beta^2}{m_\pi}$.

Gibbs posterior $G_{\beta,D}$ satisfies (ε, δ)-differential privacy if $\beta > 0$ is taken so that the right-hand side of (8) is bounded by δ. It is elementary to check the following statement:

Corollary 8. Let $\varepsilon > 0$ and $0 < \delta < 1$ be privacy parameters.

(i) Gibbs posterior $G_{\beta,D}$ satisfies $(\varepsilon, e^{-(1+\varepsilon)/4})$-differential privacy for any $\beta \leq \sqrt{m_\pi \varepsilon / 2L^2}$.

(ii) Taking β so that it satisfies

$$\beta \leq \frac{\varepsilon}{2L} \sqrt{\frac{m_\pi}{1 + 2\log(1/\delta)}}, \tag{9}$$

Gibbs posterior $G_{\beta,D}$ satisfies (ε, δ)-differential privacy.

Note that the right-hand side of (8) depends on the strong concavity m_π. The strong concavity parameter corresponds to the precision (i.e. inverse variance) of the Gaussian, and a distribution with large m_π becomes spiky. Intuitively, if we use a prior that has a strong shrinkage effect, then the posterior becomes robust against a small change of the dataset, and consequently the differential privacy can be satisfied with little effort. This observation is justified in the following sense: the upper bound of β grows proportionally to $\sqrt{m_\pi}$. In contrast, the classic exponential mechanism does not have that kind of prior-dependency.

3.2 Strongly convex loss

Let $\tilde{\ell}$ be a strongly convex function defined on the entire Euclidean space $\mathbb{R}^d$. If ℓ is a restriction of $\tilde{\ell}$ to a compact L_2-ball, the Gibbs posterior can satisfy $(\varepsilon, 0)$-differential privacy with a certain privacy level $\varepsilon > 0$ because of the boundedness of ℓ. However, using the boundedness of $\nabla \ell$ rather than that of ℓ itself, we can give another guarantee for (ε, δ)-differential privacy.

Assumption 9. Suppose that a function $\tilde{\ell} : \mathbb{R}^d \times \mathcal{X} \to \mathbb{R}$ is a twice differentiable and m_ℓ-strongly convex with respect to its first argument. Let $\tilde{\pi}$ be a finite measure over $\mathbb{R}^d$ that $-\log \tilde{\pi}(\cdot)$ is twice differentiable and m_π-strongly convex. Let $\tilde{G}_{\beta,D}$ is a Gibbs posterior on $\mathbb{R}^d$ whose density with respect to the Lebesgue measure is proportional to $\exp(-\beta \sum_i \tilde{\ell}(\theta, x_i))\tilde{\pi}(\theta)$. Assume that the mean of $\tilde{G}_{\beta,D}$ is contained in a L_2-ball of radius κ:

$$\forall D \in \mathcal{X}^n, \quad \|\mathbb{E}_{\tilde{G}_{\beta,D}}[\theta]\|_2 \le \kappa. \tag{10}$$

Define a positive number $\alpha > 1$.

Assume that (Θ, ℓ, π) satisfies the following conditions.

(i) Θ is a compact L_2-ball centered at the origin, and its radius R_Θ satisfies $R_\Theta \le \kappa + \alpha\sqrt{d/m_\pi}$.

(ii) For any $x \in \mathcal{X}$, $\ell(\cdot, x)$ is twice differentiable, L-Lipschitz, and convex. In other words, $L := \sup_{x\in\mathcal{X}} \sup_{\theta\in\Theta} \|\nabla_\theta \ell(\theta, x)\|_2$ is bounded.

(iii) π is given by a restriction of $\tilde{\pi}$ to Θ.

The following statements are the counterparts of Theorem 7 and its corollary.

Theorem 10. Let $\beta \in (0, 1]$ be a fixed parameter, and $D, D' \in \mathcal{X}^n$ be an adjacent pair of datasets. Under Assumption 9, inequality

$$G_{\beta,D}\left\{\log \frac{\mathrm{d}G_{\beta,D}}{\mathrm{d}G_{\beta,D'}} \ge \varepsilon\right\} \le \exp\left(-\frac{nm_\ell\beta + m_\pi}{4C'\beta^2}\left(\varepsilon - \frac{C'\beta^2}{nm_\ell\beta + m_\pi}\right)^2\right) \tag{11}$$

holds for any $\varepsilon > \frac{C'\beta^2}{nm_\ell\beta + m_\pi}$. Here, we defined $C' := 2CL^2(1 + \log(\alpha^2/(\alpha^2 - 1)))$, where $C > 0$ is a universal constant that does not depend on any other quantities.

Corollary 11. Under Assumption 9, there exists an upper bound $B(\varepsilon, \delta) = B(\varepsilon, \delta, n, m_\ell, m_\pi, \alpha) > 0$, and $G_\beta(\theta \mid D_n)$ with $\beta \le B(\varepsilon, \delta)$ satisfies (ε, δ)-differential privacy.

Similar to Corollary 8, the upper bound on β depends on the prior. Moreover, the right-hand side of (11) decreases to 0 as the size of dataset n increases, which implies that (ε, δ)-differential privacy is satisfied almost for free if the size of the dataset is large.

3.3 Example: Logistic regression

In this section, we will show an application of Theorem 7 to the problem of linear binary classification. Let $\mathcal{Z} := \{z \in \mathbb{R}^d, \|z\|_2 \le R\}$ be a space of the input variables. The space of the observation is the set of input variables equipped with binary label $\mathcal{X} := \{x = (z, y) \in \mathcal{Z} \times \{-1, +1\}\}$. The problem is to determine a parameter $\theta = (a, b)$ of linear classifier $f_\theta(z) = \mathrm{sgn}(a^\top z + b)$.

Define a loss function ℓ_{LR} by

$$\ell_{\mathrm{LR}}(\theta, x) := \log(1 + \exp(-y(a^\top z + b))). \tag{12}$$

The ℓ_2-regularized logistic regression estimator is given by

$$\hat{\theta}_{\mathrm{LR}} = \operatorname*{argmin}_{\theta \in \mathbb{R}^{d+1}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \ell_{\mathrm{LR}}(\theta, x_i) + \frac{\lambda}{2} \|\theta\|_2^2 \right\}, \tag{13}$$

where $\lambda > 0$ is a regularization parameter. Corresponding Gibbs posterior has a density

$$G_\beta(\theta \mid D) \propto \prod_{i=1}^{n} \sigma(y_i(a^\top z_i + b))^\beta \phi_{d+1}(\theta \mid 0, (n\lambda)^{-1} I), \tag{14}$$

where $\sigma(u) = (1 + \exp(-u))^{-1}$ is a sigmoid function, and $\phi_{d+1}(\theta \mid \mu, \Sigma)$ is a density of $(d+1)$-dimensional Gaussian distribution.

It is easy to check that $\ell_{\mathrm{LR}(\cdot,x)}$ is R-Lipschitz and convex, and $-\log \phi_{d+1}(\cdot \mid 0, (n\lambda^{-1})I)$ is $(n\lambda)$-strongly convex. Hence, by Corollary 8, the Gibbs posterior satisfies (ε, δ)-differential privacy if

$$\beta \leq \frac{\varepsilon}{2R} \sqrt{\frac{n\lambda}{1 + 2\log(1/\delta)}}. \tag{15}$$

4 Proofs

In this section, we give a formal proof of Theorem 7 and a proof sketch of 10.

There is a vast literature on techniques to obtain a concentration inequality in (7) (see, for example, [2]). Logarithmic Sobolev inequality (LSI) is a useful tool for this purpose. We say that a probability measure μ over $\Theta \subset \mathbb{R}^d$ satisfies LSI with constant D_{LS} if inequality

$$\mathbb{E}_\mu[f^2 \log f^2] - \mathbb{E}_\mu[f^2] \log \mathbb{E}_\mu[f^2] \leq 2 D_{\mathrm{LS}} \mathbb{E}_\mu \|\nabla f\|_2^2 \tag{16}$$

holds for any integrable function f, provided the expectations in the expression are defined. It is known that [9, 2], if μ satisfies LSI, then every real-valued L-Lipschitz function F behaves in a sub-Gaussian manner:

$$\mu\{F \geq \mathbb{E}_\mu[F] + t\} \leq \exp\left(-\frac{t^2}{2L^2 D_{\mathrm{LS}}} \right). \tag{17}$$

In our analysis, we utilize the LSI technique for the following two reasons: (a) a sub-Gaussian tail bound of the log-density ratio is obtained from (17), and (b) an upper bound on the KL-divergence is directly obtained from LSI, which appears to be difficult to prove by any other argument.

Roughly speaking, LSI holds if the logarithm of the density is strongly concave. In particular, for a Gibbs measure on $\mathbb{R}^d$, the following fact is known.

Lemma 12 ([9])**.** Let $U : \mathbb{R}^d \to \mathbb{R}$ be a twice differential, m-strongly convex and integrable function. Let μ be a probability measure on $\mathbb{R}^d$ whose density is proportional to $\exp(-U)$. Then μ satisfies LSI (16) with constant $D_{\mathrm{LS}} = m^{-1}$.

Proof of Theorem 7. Since $U(\cdot) = \beta \sum_i \ell(\cdot, x_i) - \log \pi(\cdot)$ is m_π-strongly convex, Gibbs posterior $G_{\beta,D}$ satisfies LSI with constant m_π^{-1}.

Let $D, D' \in \mathcal{X}^n$ be a pair of adjacent datasets. Considering appropriate permutation of the elements, we can assume that $D = (x_1, \ldots, x_n)$ and $D' = (x'_1, \ldots, x'_n)$ differ in the first element, namely, $x_1 \neq x'_1$ and $x_i = x'_i$ $(i = 2, \ldots, n)$. By the assumption that $\ell(\cdot, x)$ is L-Lipschitz, we have

$$\left\| \nabla \log \frac{\mathrm{d}G_{\beta,D}}{\mathrm{d}G_{\beta,D'}} \right\|_2 = \beta \|\nabla(\ell(\theta, x_1) - \ell(\theta, x'_1))\|_2 \leq 2\beta L, \tag{18}$$

and log-density ratio $\log \frac{\mathrm{d}G_{\beta,D}}{\mathrm{d}G_{\beta,D'}}$ is $2\beta L$-Lipschitz. Then, by concentration inequality for Lipschitz function (17), we have

$$\forall t > 0, \quad G_{\beta,D} \left\{ \log \frac{\mathrm{d}G_{\beta,D}}{\mathrm{d}G_{\beta,D'}} \geq D_{\mathrm{KL}}(G_{\beta,D}, G_{\beta,D'}) + t \right\} \leq \exp\left(-\frac{m_\pi t^2}{8L^2\beta^2} \right) \tag{19}$$

We will show an upper bound of the KL-divergence. To simplify the notation, we will write $F := \frac{\mathrm{d}G_{\beta,D}}{\mathrm{d}G_{\beta,D'}}$. Noting that

$$\|\nabla\sqrt{F}\|_2^2 = \|\nabla \exp(2^{-1}\log F)\|_2^2 = \|\frac{\sqrt{F}}{2}\nabla \log F\|_2^2 \leq \frac{F}{4}\cdot(2\beta L)^2 \tag{20}$$

and

$$\begin{aligned} D_{\mathrm{KL}}(G_{\beta,D}, G_{\beta,D'}) &= \mathbb{E}_{G_{\beta,D}}[\log F] \\ &= \mathbb{E}_{G_{\beta,D'}}[F\log F] - \mathbb{E}_{G_{\beta,D'}}[F]\mathbb{E}_{G_{\beta,D'}}[\log F], \end{aligned} \tag{21}$$

we have, from LSI (16) with $f = \sqrt{F}$,

$$D_{\mathrm{KL}}(G_{\beta,D}, G_{\beta,D'}) \leq \frac{2}{m_\pi}\mathbb{E}_{G_{\beta,D'}}\|\nabla\sqrt{F}\|_2^2 \leq \frac{2L^2\beta^2}{m_\pi}\mathbb{E}_{G_{\beta,D'}}[F] = \frac{2L^2\beta^2}{m_\pi}. \tag{22}$$

Combining (19) and (22), we have

$$\begin{aligned} G_{\beta,D}\left\{\log\frac{\mathrm{d}G_{\beta,D}}{\mathrm{d}G_{\beta,D'}} \geq \varepsilon\right\} &\leq G_{\beta,D}\left\{\log\frac{\mathrm{d}G_{\beta,D}}{\mathrm{d}G_{\beta,D'}} \geq \varepsilon + D_{\mathrm{KL}}(G_{\beta,D}, G_{\beta,D'}) - \frac{2L^2\beta^2}{m_\pi}\right\} \\ &\leq \exp\left(-\frac{m_\pi}{8L^2\beta^2}\left(\varepsilon - \frac{2L^2\beta^2}{m_\pi}\right)^2\right) \end{aligned} \tag{23}$$

for any $\varepsilon > \frac{2L^2\beta^2}{m_\pi}$. □

Proof sketch for Theorem 10. The proof is almost the same as that of Theorem 7. It is sufficient to show that the set of Gibbs posteriors $\{G_{\beta,D},\ D \in \mathcal{X}^n\}$ simultaneously satisfies LSI with the same constant. Since the logarithm of the density is $m := (nm_\ell\beta + m_\pi)$-strongly convex, a probability measure $\tilde{G}_{\beta,D}$ satisfies LSI with constant m^{-1}. By the Poincaré inequality for $\tilde{G}_{\beta,D}$, the variance of $\|\theta\|_2$ is bounded by $d/m \leq d/m_\pi$. By the Chebyshev inequality, we can check that the mass of parameter space is lower-bounded as $\tilde{G}_{\beta,D}(\Theta) \geq p := 1 - \alpha^{-2}$. Then, by Corollary 3.9 of [11], $G_{\beta,D} := \tilde{G}_{\beta,D}|_\Theta$ satisfies LSI with constant $C(1 + \log p^{-1})m^{-1}$, where $C > 0$ is a universal numeric constant. □

5 Conclusion

In this paper, we have proved (ε, δ)-differential privacy of the Gibbs posterior under certain regularity conditions based on Lipschitz property and convexity. The proofs of the main theorems are based on the logarithmic Sobolev inequalities, which provides a useful tool to prove the concentration of measure inequalities.

Acknowledgments

We would like to thank Hiroshi Nakagawa and Junpei Komiyama for helpful discussions. We also thank the anonymous reviewers for their helpful comments.

References

[1] R. Bassily, A. Smith, and A. Thakurta. Differentially private empirical risk minimization: Efficient algorithms and tight error bounds. In *FOCS*, 2014.

[2] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[3] O. Catoni. *Pac-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning*. IMS, 2007.

[4] K. Chaudhuri, C. Monteleoni, and A.D. Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12:1069–1109, 2011.

[5] C. Dimitrakakis, B. Nelson, and B. Rubinstein. Robust and private Bayesian inference. In *Algorithmic Learning Theory*, 2014.

[6] C. Dwork. Differential privacy. In *ICALP*, pages 1–12, 2006.

[7] R. Hall, A. Rinaldo, and L. Wasserman. Differential privacy for functions and functional data. *Journal of Machine Learning Research*, 14:703–727, 2013.

[8] D. Kifer, A. Smith, and A. Thakurta. Private convex empirical risk minimization and high-dimensional regression. In *COLT*, 2012.

[9] M. Ledoux. *Concentration of Measure and Logarithmic Sobolev Inequalities*, volume 1709 of *Séminaire de Probabilités XXXIII Lecture Notes in Mathematics*. Springer, 1999.

[10] F. McSherry and K. Talwar. Mechanism design via differential privacy. In *FOCS*, 2007.

[11] E. Milman. Properties of isoperimetric, functional and Transport-Entropy inequalities via concentration. *Probability Theory and Related Fields*, 152:475–507, 2012.

[12] Y. Wang, S. Fienberg, and A. Smola. Privacy for free: Posterior sampling and stochastic gradient monte carlo. In *ICML*, 2015.

[13] T. Zhang. From ε-entropy to KL-entropy: Analysis of minimum information complexity density estimation. *The Annals of Statistics*, 34(5):2180–2210, 2006.