Gibbs事後分布の測度集中と差分プライバシ（研究紹介）

中川研究室 修士2年 南 賢太郎
指導教員： 中川 裕志 教授

2015年6月12日

概 要

差分プライバシとは，個人情報を含むデータから構成した推定量や学習器を外部に公開する場合のプライバシ保護基準として代表的なものである．制約条件として差分プライバシを満たしつつ，汎化誤差の意味でできるだけ良いランダム推定量を構成する問題を差分プライベート学習という．Gibbs 事後分布とはデータ依存の確率分布であり，ベイズの事後分布を逆温度パラメータによって一般化したものに相当する．Gibbs 事後分布は PAC-Bayes 学習と呼ばれる分野では基本的な役割をもつため，Gibbs 事後分布が差分プライバシ制約を満たす条件を調べることは，既存の統計的学習の枠組みと差分プライベート学習を結びつける意味で重要である．

(ε, δ)-差分プライバシという性質を調べる際には，尤度比の測度集中不等式を示すことが鍵となる．本発表では，損失関数と事前分布が凸性の仮定を満たす場合に，Gibbs 事後分布が差分プライバシを満たすための条件を測度集中の理論を利用して調べる．また，サンプルサイズおよび事前分布の縮小の強さと,「プライバシ保護のしやすさ」の関係について論じる．

1 はじめに

個人情報を含むデータから抽出した何らかの統計情報を第三者に公開する場合，その値を通して元のデータセットに含まれる特定個人の情報が漏洩する可能性がある．推定量や学習器といった統計情報は，一見すると元のデータ形式とはかけ離れており,「匿名化された」値となっていることが多い．しかし，他の公開情報との組み合わせによる名寄せ行為の危険性を考慮すると，特に機微性のある個人情報を保護するためには，任意の統計的手法によって特定個人のデータを逆推定することの困難性が保証されることが望ましい．差分プライバシ [?] は，そのような保証を定量的に与える尺度として代表的なものである．

差分プライバシは，公開する値を適当にランダム化し，データセットに含まれる1個人分のデータが変化したとき，それに伴う出力の変化が確率分布の

意味で小さいということを主張している．この保証により，直感的には，たとえ攻撃者がある特定の 1 人以外の全てのデータを所持していたとしても，出力を用いて残りの 1 人の値を有意に特定することが困難となる．差分プライバシの基本的なアイデアは，本来あるべき出力にノイズを加えることによって撹乱するというものであり，一般に，強い差分プライバシを保証しようとすればするほど推定や予測の精度は悪くなるというトレードオフの関係がある．そこで，差分プライバシ制約を満たした上で，いかに精度のよい学習を行うことができるかということは興味深い問題であり，近年盛んに研究されている [**?**, **?**, **?**, **?**].

差分プライバシを満たしつつ与えられた目的関数をなるべく小さくするようなサンプリングアルゴリズムとして，最も基本的なものは指数メカニズム [**?**] である．Wang, et al. [**?**] は，パラメトリック統計モデルにおけるベイズ事後分布

$$\frac{\prod_{i=1}^{n} p(x_i \mid \theta)\pi(\theta)}{\int \prod_{i=1}^{n} p(x_i \mid \theta)\pi(\theta)\mathrm{d}\theta} \tag{1}$$

から θ をサンプリングすることが指数メカニズムの特別な場合であることを指摘し，対数尤度 $\log p(x|\theta)$ が有界である場合には差分プライバシが常に成立することを示した．よって，この場合は差分プライバシとベイズ的な学習手法が同一の枠組みに収まり，プライバシ保護と学習精度のトレードオフなどの議論の見通しが良くなると考えられる．しかし一般に，対数尤度が有界となるような確率モデルは非常に限られており，[**?**] の主張はそのままの形で適用することができない．

一方，統計学や機械学習において，ある種の正則化の目的で推定量をランダム化するという手段が取られることがある．特に，PAC-Bayes 学習 [**?**, **?**] では，Gibbs 事後分布 (Gibbs posterior) または擬ベイズ事後分布 (pseudo-Bayesian posterior) と呼ばれるデータ依存の条件付き分布

$$\frac{\exp\left(-\beta \sum_{i=1}^{n} \ell(\theta, x_i)\right) \pi(\theta)}{\int \exp\left(-\beta \sum_{i=1}^{n} \ell(\theta, x_i)\right) \pi(\theta)\mathrm{d}\theta} \tag{2}$$

からサンプルしたランダム推定量が考察される．$\ell(\theta, x)$ は損失関数であり，パラメトリック統計モデルにおける推定の場合には $-\log p(x \mid \theta)$ がこれに相当する．このような推定量も，密度関数の式の形のみに着目する限りでは，指数メカニズムと似ている．よって，何らかの条件のもとで差分プライバシを満たすということが示せれば，プライバシ保護と PAC-Bayes 学習を同一の枠組みで議論することができ，アルゴリズム設計や理論解析の幅が広がることが期待できる．

本論文では，適当な条件のもとで Gibbs 事後分布 (**??**) からのサンプリングが (ε, δ)-差分プライバシを満たすことを示す．本論文の結果の特徴は次のとおりである：

- **（非有界損失への対応）** 損失関数がパラメータ θ に関して Lipschitz で

ある場合，有界性を用いずに差分プライバシを示すことができる．この意味で，損失関数が有界な場合の [?] の結果の拡張になっている．(定理??)

- **（事前分布の縮小効果とプライバシの関係）** 事前分布 π が縮小効果の強い（尖った）分布であるとき，事後分布は事前分布の影響を受けやすく，データの変化に関してロバストであることが期待される．したがって，そのような事前分布を使った場合の方が差分プライバシを満たしやすい．本論文の結果は，事前分布の強凸性パラメータに応じて逆温度パラメータ β の設定可能な範囲が広がることを示しており，この性質を反映した結果となっている．(定理??，定理??)

- **（サンプルサイズとプライバシの関係）** サンプルサイズ n が大きいほど，1 つのデータあたりの寄与が小さくなり，差分プライバシを満たしやすくなることが期待される．このような効果は，逆温度パラメータの上界という形で直接的に示される．(定理??)

本論文の構成は以下の通りである．第 2 節では，差分プライバシの定義と基本的な性質を与え，事後分布からのサンプリングに関する [?] の結果を紹介する．ここでは，先行研究でよく知られている結果を，関数の Lipschitz 性を通して簡潔に書くことを試みる．本論文における主結果は第 3 節で述べる．3.1 節では，(ε, δ)-差分プライバシの十分条件である対数尤度比の裾確率のバウンドについて説明する．3.2 節では擬ベイズ事後分布 (??) が差分プライバシを満たすことを主張した主定理を述べる．第 4 節では，第 3 節で得られた結果を利用して，いくつかの新しい差分プライベート学習のアルゴリズムを導出する．4.1 節では，ロジスティック回帰のベイズ的な対応物による差分プライベートな線形分類器について説明する．4.2 節では，正規分布の密度推定の例を紹介する．4.3 節では，マルコフ連鎖モンテカルロ法によって Gibbs 事後分布を近似する場合のプライバシ保証について説明する．第 5 節では結論と今後の課題を述べる．

2 差分プライバシ

2.1 定義と Lipschitz 性による特徴づけ

差分プライバシは，ユーザー 1 人分のデータでのみ異なるデータセット同士に対応する出力が，統計的な意味で互いに区別しにくい，ということを定量化したものである．本論文では一般性を失うことなく，[1]データセット D は n 人のユーザーの個人データ $x_i \in \mathcal{X}$ からなる列 $D = (x_1, \ldots, x_n)$ とし，

[1]文献によってはデータサイズの n 違いや，置換を考えた編集距離を利用することもある．現実的な意味を考えるとそちらが妥当な場合もあるが，本論文で主に扱う統計的学習の問題では，そのような違いは本質的に無視できる．

2 つのデータセット D, D' が隣接するとは $d_H(D, D') = 1$ であることとする．ここで d_H はデータセットの空間 $\mathcal{D} = \mathcal{X}^n$ で定義されたハミング距離 $d_H(D, D') = \sum_{i=1}^n 1_{\{x_i \neq x'_i\}}$ である．

データセット D を入力とし，ある（可測）空間 $(\Theta, \mathcal{T})$ の値を出力として公開することを考える．Θ 上の確率測度の全体を $\mathcal{M}(\Theta)$ で表す．$\rho : \mathcal{D} \to \mathcal{M}(\Theta)$ を考える．つまり ρ はデータセット D を入力として確率分布 ρ_D を出力する関数である．実際には，ρ_D に従う確率変数をサンプリングする行為が推定量の公開に相当する．

定義 1 (差分プライバシ [?])**.** $\varepsilon > 0, \delta \geq 0$ を与えられたプライバシ強度とする．$\rho : \mathcal{D} \to \mathcal{M}(\Theta)$ が (ε, δ)-差分プライバシを満たすとは，任意の $D, D' \in \mathcal{D}$, $d_H(D, D') = 1$ に対して，

$$\rho_D(A) \leq e^{\varepsilon} \rho_{D'}(A) + \delta, \quad \forall A \in \mathcal{T} \tag{3}$$

が成り立つことをいう．特に，$\delta = 0$ のとき ρ は ε-差分プライバシを満たすという．

$\delta = 0$ のとき，不等式 (**??**) は

$$d_{\mathrm{priv}}(\rho_D, \rho_{D'}) := \sup_{A \in \mathcal{T}} \left| \log \frac{\rho_D(A)}{\rho_{D'}(A)} \right| \leq \varepsilon \tag{4}$$

であることと等価である．ここで上の d_{priv} は $\mathcal{M}(\Theta)$ 上の「強い」距離[2]を定めることがわかるが，ρ が ε-差分プライバシを満たすとは，Lipschitz 性 $d_{\mathrm{priv}}(\rho_D, \rho_{D'}) \leq \varepsilon d_H(D, D')$ と同値であることがわかる．一般に，統計的な推定や検定の問題の難しさは，大雑把に言えば異なる確率分布同士の区別のしやすさで決まる．攻撃者の目的は，公開された統計量をもとに，ハミング距離 d_H の意味でデータセットを区別することである．差分プライバシは，データセット同士がその距離の意味で近いほど，統計的手法では区別できないということを述べている．

なお，ハミング距離は「縮退した」距離であり，Lipschitz 性が本質的にはデータセットの空間または損失関数の有界性を要請するため，特にデータ空間が非有界な場合の差分プライバシの扱いは難しい．Dimitrakakis, et al. [?] では，差分プライバシの定義をより一般の距離まで拡張することで，ハミング距離では扱いづらい正規分布モデルなどのプライバシ保証を示している．しかし，本論文では伝統的な定義通り，ハミング距離による隣接関係のみ考察する．

[2]ゼロ集合を除くなどして適切に定義すると，Kullback-Leibler ダイバージェンスや全変動よりも収束が強い距離が得られる．詳細は省略する．

2.2 指数メカニズム

データセット D を入力すると，θ の非負関数 $H(\theta, D)$ が決まるとする．H は情報 θ を公開した場合の損失関数（または負の効用関数）に相当し，ε-差分プライバシの制約を満たす範囲で，この値をなるべく小さくするような θ を公開することが望ましい．このとき，どのような確率分布から θ をサンプリングすればよいかが問題になるが，指数メカニズム (exponential mechanism)[**?**] がその一般的な指針を与える．本節ではこのアルゴリズムについて，Lipschitz 性に着目した新たな説明を与える．

関数に値を取る写像 $D \mapsto H(\cdot, D)$ が sup ノルムに関して Lipschitz であるとする．つまり，ある $\Delta_1 > 0$ が存在して

$$\| H(\cdot, D) - H(\cdot, D') \|_\infty \leq \Delta_1 d_H(D, D') \tag{5}$$

であったとする．差分プライバシの文脈では，Lipschitz 定数 Δ_1 はしばしば L_1-sensitivity と呼ばれる．(式 (**??**) は実質的には $d_H(D, D') = 1$ のところでしか情報を持たないことに注意する.)

次に，Θ 上の（非負）関数 f に対して，確率分布

$$\mathrm{d}G_{\beta,f}(\theta) = \frac{\exp(-\beta f(\theta))\pi(\theta)}{Z}\mathrm{d}\theta \tag{6}$$

を対応させる写像 $G_{\beta,\cdot}$ を考える．ここで，$\beta > 0$ は与えられた正数，π は与えられた基底測度で，Z は $G_{\beta,f}$ が確率密度になるようにするための正規化定数である．π の寄与を考えなければ，$G_{\beta,f}$ は，f の値が大きいところに集中した分布である．本論文では $G_{\beta,\cdot}$ を Gibb 写像と呼ぶことにする．Gibbs 写像は 2β-Lipschitz であることがわかる．すなわち，

$$d_{\mathrm{priv}}(G_{\beta,f}, G_{\beta,g}) \leq 2\beta \| f - g \|_\infty \tag{7}$$

である．一般に，C_1-Lipschitz 関数と C_2-Lipschitz 関数との合成は C_1C_2-Lipschitz であるから，$H(\cdot, D)$ と Gibbs 写像との合成は $2\Delta_1\beta$-差分プライバシを満たす．以上をまとめると次のようになる．

定理 2 (指数メカニズム [**?**])**.** データセット D に対して，

$$\exp\left(-\frac{\varepsilon}{2\Delta_1}H(\theta, D)\right)\pi(\theta) \tag{8}$$

に比例した密度をもつ確率分布から θ をサンプルすることを考える．この対応は ε-差分プライバシを満たす.

さて，$H(\theta, D) = -\sum_{i=1}^n \log p(x_i \mid \theta)$ として，Gibbs 事後分布

$$q_\beta(\theta \mid D) = \frac{\left(\prod_{i=1}^n p(x_i \mid \theta)\right)^\beta \pi(\theta)}{\int \left(\prod_{i=1}^n p(x_i \mid \theta)\right)^\beta \pi(\theta)\mathrm{d}\theta} \tag{9}$$

を考える．$\beta > 0$ は逆温度パラメータである．$q_\beta(\theta \mid D)$ は $\beta = 1$ のとき通常の意味でのベイズ事後分布，$\beta \to \infty$ のとき最尤推定量の上への点質量と一致することに注意する．また，PAC-Bayes 学習では収束レートの関係でしばしば $0 < \beta < 1$ の場合（高温度）に興味がある．$q_\beta(\theta \mid D)$ は形式的には指数メカニズムと同じ形をしている．実際にそうであるためには，H が Lipschitz であればよい．次は [**?**] を少し一般化したものである．証明は明らかである．

定理 3 (Wang, et al.(2015)[**?**], Theorem 4)**.** (i) 任意の $x \in \mathcal{X}$ に対して対数尤度関数が $\| \log p(x|\cdot) \|_\infty \leq B$ を満たすとする．このとき，$H(\cdot, D) = \sum_i \log p(x_i \mid \cdot)$ は $2B$-Lipschitz であり，$q_\beta(\theta \mid D)$ は $4\beta B$-差分プライバシを満たす．特に，$\beta = 1$ とすると，ベイズ事後分布 $p(\theta \mid D)$ は $4B$-差分プライバシを満たす．

(ii) $H(\cdot, D)$ が $\mathcal{D}$ 上のある距離 d に関して L-Lipschitz であり，さらに $\mathcal{D}$ の距離 d による直径が有界 $\sup_{D,D'} d(D, D') \leq 2R$ であるとする．このとき，$H(\cdot, D)$ は d_H に関して $2LR$-Lipschitz であるから，$q_\beta(\theta \mid D)$ は $4\beta LR$-差分プライバシを満たす．

定理**??**は，ベイズ的な枠組みと差分プライバシを結びつける意味で非常に重要であると考えられるが，実際には (i)(ii) の仮定を満たす統計モデルは多くない．例えば，正規分布モデルでは対数尤度関数は θ について有界でも，x について Lipschitz でもない．

では，他の対数尤度関数（より一般には，損失関数）に対して，指数メカニズムが成り立たない場合にも，何らかの意味で差分プライバシを示すことはできないだろうか．やや不正確な議論ではあるが，直感的には次のようなことが考えられる．$\beta = 1$ とすると，ベイズ事後分布は，Bernstein-von Mises の定理 [**?**] より，$n \to \infty$ の極限では正規分布に近い挙動をする．一方，平均または分散が少しでも異なる 2 つの正規分布に対して，距離 d_{priv} は $+\infty$ となる．したがって，もし十分大きな有限の n において事後分布が正規分布で近似されるならば，それらは $\varepsilon(< \infty)$-差分プライバシを満たさない．一方，そのような場合でも $\delta > 0$ を適当にとれば，(ε, δ)-差分プライバシを満たすようにできる．そこで，次節では，Gibbs 事後分布が (ε, δ)-差分プライバシを満たす条件を調べることにする．

3 Gibbs 事後分布による差分プライバシ

3.1 (ε, δ)-差分プライバシの十分条件

$\delta > 0$ とする．(ε, δ)-差分プライバシを満たすための十分条件として，次の定理がよく知られている．証明は [**?**], Lemma 2 を参照されたい．

定理 4. $\rho : \mathcal{D} \to \mathcal{M}(\Theta)$ が (ε, δ)-差分プライバシを満たすための十分条件は，任意の $D, D' \in \mathcal{D}$, $d_H(D, D') = 1$ に対して，

$$\rho_D \left\{ \log \frac{\mathrm{d}\rho_D}{\mathrm{d}\rho_{D'}} \geq \varepsilon \right\} \leq \delta \tag{10}$$

が成り立つことである．ここで $\mathrm{d}\rho_D / \mathrm{d}\rho_{D'}$ は密度関数の比である．

与えられた D, D' に対して，不等式 (**??**) が成立するためには何が必要だろうか．(**??**) は対数尤度比 $\log \frac{\mathrm{d}\rho_D}{\mathrm{d}\rho_{D'}}$ の裾確率に関する不等式である．この関数の平均値は

$$\mathbb{E}_{\rho_D} \left[\log \frac{\mathrm{d}\rho_D}{\mathrm{d}\rho_{D'}} \right] = D_{\mathrm{KL}}(\rho_D, \rho_{D'}) \tag{11}$$

であり，Kullback-Leibler(KL) ダイバージェンスに一致する．もし平均値 $D_{\mathrm{KL}}(\rho_D, \rho_{D'})$ が十分小さく ($D_{\mathrm{KL}}(\rho_D, \rho_{D'}) \approx \varepsilon$)，さらに $\log \frac{\mathrm{d}\rho_D}{\mathrm{d}\rho_{D'}}$ が平均値のまわりに十分集中していれば，つまり，$t > 0$ に関して減少する関数 $\psi(t)$ が存在して

$$\rho_1 \left\{ \log \frac{\mathrm{d}\rho_D}{\mathrm{d}\rho_{D'}} \geq D_{\mathrm{KL}}(\rho_D, \rho_{D'}) + t \right\} \leq \psi(t) \tag{12}$$

のような不等式を示すことができれば，それを通して裾確率 (**??**) を評価することができると考えられる．

3.2 Gibbs 事後分布のプライバシ

パラメータ空間 Θ を $\mathbb{R}^d$ またはその部分集合とし，Gibbs 分布

$$\mathrm{d}G_{\beta,H}(\theta) = \frac{\exp(-\beta H(\theta))\pi(\theta)}{Z} \mathrm{d}\theta \tag{13}$$

を考える．$H(\theta)$，$\pi(\theta)$ はそれぞれ適当な滑らかさをもつエネルギー関数と密度関数であり，$\beta \in (0, 1]$ とする．$Z = \int_\Theta \exp(-\beta H(\theta))\pi(\theta)\mathrm{d}\theta$ は正規化定数である．本節の目的は，$H(\cdot, D) = \sum_{x \in D} \ell(\cdot, x)$ であるときに，Gibbs 事後分布 $G_{\beta,H(\cdot,D)}$ が差分プライバシを満たす条件を調べることである．すなわち，$G_\beta : D \mapsto G_{\beta,H(\cdot,D)}$ に対して不等式 (**??**) を考える．

本論文では，裾確率の評価は測度集中の一般論を利用して行うが，そこで損失関数の凸性や Lipschitz 性が重要な役割をもつ．2 回連続微分可能な関数 $f : \mathbb{R}^d \to \mathbb{R}$ が $m(> 0)$-強凸であるとは，Hessian$\nabla^2 \ell = (\frac{\partial^2}{\partial\theta_i \partial\theta_j})_{1 \leq i,j \leq n}$ について $\nabla^2 f \succeq mI$ が成り立つことをいう．また，本論文では便宜上，凸関数であるが任意の $m > 0$ に対して m-強凸とならないとき，またそのときに限り 0-強凸であるということとする．

本論文では，次の仮定を満たすモデルについて考える．仮定**??**は基本的な仮定であり，全ての場合に共通して成り立つとする．仮定**??**および仮定**??**はどちらか一方が成り立つとする．

仮定 5 (基本的な仮定). パラメータ空間 $\Theta \subset \mathbb{R}^d$, 損失関数 $\ell : \Theta \times \mathcal{X} \to \mathbb{R}_{\geq 0}$, 事前分布 $\pi \in \mathcal{M}(\Theta)$ に関して，次のことが成り立つとする．

(i) Θ は凸集合である．

(ii) $\ell(\cdot, x)$ は 2 回連続微分可能であり，任意の $x \in \mathcal{X}$ について m_ℓ-強凸である．また，$-\log \pi(\theta)$ は 2 回連続微分可能であり，m_π-強凸である．

仮定 6 (非強凸，Lipschitz 損失)． (i) $m_\ell = 0$ かつ $m_\pi > 0$ である．

(ii) $x \in \mathcal{X}$ を任意に固定したとき，$\ell(\cdot, x)$ は Θ 上で L-Lipschitz である．

仮定 7 (強凸損失，有界パラメータ空間)． (i) Θ は原点を中心としたコンパクトな球である．球の半径については後の定理で定める．

(ii) $m_\ell > 0$ かつ $m_\pi > 0$ である．ただし，損失関数 $\ell(\cdot, x)$ は $\mathbb{R}^d$ 全体で定義されている m_ℓ-強凸関数を Θ 上に制限したものであるとする．また，π は $\mathbb{R}^d$ に台をもち $-\log \pi$ が m_π-強凸であるような測度 $\bar{\pi}$ を Θ 上に制限した有限測度であるとする．

(iii) $\mathcal{X}$ はある距離 d_X について有界な集合であり，$R_X = \sup_{x,x'} d_X(x, x') < \infty$ であるとする．(ii) において，$\mathbb{R}^d$ 全体に損失関数を拡張し，Lebesgue 測度に関する密度が $\exp(-\beta \sum_i \ell(\theta, x_i) - \log \bar{\pi}(\theta)$ に比例するようにとった確率測度を $\bar{G}_{\beta,D}$ とおく．このとき，ある有界な非負関数 $\kappa(n, m_\ell, m_\pi, R_X)$ が存在して，任意のデータセット $D \in \mathcal{X}^n$ に対する平均 $\mathbb{E}_{\bar{G}_{\beta,D}}[\theta]$ が

$$\| \mathbb{E}_{\bar{G}_{\beta,D}}[\theta] \|_2 \leq \kappa(n, m_\ell, m_\pi, R_X) \tag{14}$$

を満たす．

(iv) $\sup_{x \in X} \sup_{\theta \in \Theta} \| \nabla \ell(\theta, x) \|_2 = L < \infty$ である．特に，任意の $x \in \mathcal{X}$ を固定したとき，$\ell(\cdot, x)$ は L-Lipschitz である．

以上の仮定に関するいくつかの注意点を述べる．仮定**??**について，本論文で考えるモデルは損失関数 ℓ および $-\log \pi$ がともになめらかな凸関数であるものに限る．直感的には，$m_\ell > 0$ は損失関数が 2 次関数よりも早く増大するということを意味する．よって，Θ が非有界（例えば $\Theta = \mathbb{R}^d$）であるとき，$m_\ell > 0$ であることと ℓ が Lipschitz であることは一般に両立しない．そこで，考えているモデルを，$m_\ell = 0$ であって ℓ が Lipschitz である場合（仮定**??**），$m_\ell > 0$ であって Θ が有界である場合（仮定**??**）の 2 通りに分類して議論する．仮定**??**の条件 (iii) は複雑に見えるが，本質的には $\mathcal{X}$ が有界であって，事後平均がハミング距離に関して離れすぎないための十分条件を与えたものである．一方，$m_\pi > 0$ は常に仮定しておく．こちらは，ある程度縮小効果の強い事前分布を使うことによって，β を小さくしても事後分布の分散が大きくなりすぎないようにするためである．

次の定理??は本論文における主結果であり，対数尤度比関数の裾確率の上界を与える．定理??および系??の証明は，概略のみ付録に示す．

定理 8 (Lipschitz 損失の場合)．D_1, D_2 は $\mathcal{X}^n$ に属する 2 つのデータセットであり，$d_H(D_1, D_2) = 1$ であるとする．仮定??および仮定??が成り立つとする．このとき，次の (i), (ii) が成り立つ．

(i) $\beta \in (0, 1]$ を固定する．このとき，任意の $t > 0$ に対して

$$G_{\beta,D_1}\left\{\log\frac{\mathrm{d}G_{\beta,D_1}}{\mathrm{d}G_{\beta,D_2}} > D_{\mathrm{KL}}(G_{\beta,D_1}, G_{\beta,D_2}) + t\right\} \leq \exp\left(-\frac{m_\pi t^2}{8L^2\beta^2}\right) \tag{15}$$

が成立する．

(ii) 任意の $\beta \in (0, 1]$ に対して，KL ダイバージェンスは上界

$$D_{\mathrm{KL}}(G_{\beta,D_1}, G_{\beta,D_2}) \leq \frac{2L^2\beta^2}{m_\pi} \tag{16}$$

をもつ．また，$\varepsilon > \frac{2L^2\beta^2}{m_\pi}$ のとき

$$G_{\beta,D_1}\left\{\log\frac{\mathrm{d}G_{\beta,D_1}}{\mathrm{d}G_{\beta,D_2}} > \varepsilon\right\} \leq \exp\left(-\frac{m_\pi}{8L^2\beta^2}\left(\varepsilon - \frac{2L^2\beta^2}{m_\pi}\right)^2\right) \tag{17}$$

が成り立つ．

定理 9 (強凸損失の場合)．D_1, D_2 は $\mathcal{X}^n$ に属する 2 つのデータセットであり，$d_H(D_1, D_2) = 1$ であるとする．仮定??および仮定??が成り立つとする．$\alpha > 1$ を固定し，定義域の半径 R_Θ を

$$R_\Theta \geq \kappa(n, m_\ell, m_\pi, R_X) + \alpha\sqrt{\frac{d}{m_\pi}} \tag{18}$$

ととる．(R_Θ のとり方に依存して L も変化することに注意する．) このとき，次の (i), (ii) が成り立つ．

(i) $\beta \in (0, 1]$ を固定する．このとき，任意の $t > 0$ に対して

$$G_{\beta,D_1}\left\{\log\frac{\mathrm{d}G_{\beta,D_1}}{\mathrm{d}G_{\beta,D_2}} > D_{\mathrm{KL}}(G_{\beta,D_1}, G_{\beta,D_2}) + t\right\} \leq \exp\left(-\frac{(nm_\ell\beta + m_\pi)t^2}{8L^2C(1 + \log(\alpha^2/(\alpha^2-1)))\beta^2}\right) \tag{19}$$

が成立する．ただし，$C > 0$ は普遍的な定数である．

(ii) 任意の $\beta \in (0, 1]$ に対して，KL ダイバージェンスは上界

$$\begin{aligned} D_{\mathrm{KL}}(G_{\beta,D_1}, G_{\beta,D_2}) \leq \quad & \frac{2L^2C(1 + \log(\alpha^2/(\alpha^2-1)))\beta^2}{nm_\ell\beta + m_\pi} \\ =: \quad & D_+(\beta, n, m_\ell, m_\pi, \alpha, L) = D_+ \end{aligned} \tag{20}$$

をもつ．また，$\varepsilon > D_+$ のとき

$$G_{\beta,D_1}\left\{\log\frac{\mathrm{d}G_{\beta,D_1}}{\mathrm{d}G_{\beta,D_2}} > \varepsilon\right\} \leq \exp\left(-\frac{1}{4D_+}\left(\varepsilon - D_+\right)^2\right) \tag{21}$$

が成り立つ．

定理**??**と定理**??**を応用して，Gibbs 事後分布が満たす (ε,δ)-差分プライバシに関する結果が得られる．

系 10. 仮定**??**と仮定**??**が成り立つとする．

(i) $\varepsilon > 0$ とする．

$$\beta < \sqrt{\frac{m_\pi \varepsilon}{2L^2}} \tag{22}$$

であるような β に対して，Gibbs 事後分布 $G_{\beta,D}$ は $(\varepsilon, e^{-(1+\varepsilon)/4})$-差分プライバシを満たす．

(ii) $\varepsilon > 0$, $1 > \delta > 0$ とする．

$$\beta \leq \frac{\varepsilon}{2L}\sqrt{\frac{m_\pi}{1 + 2\log(1/\delta)}} \tag{23}$$

であるような β に対して，Gibbs 事後分布 $G_{\beta,D}$ は (ε,δ)-差分プライバシを満たす．

系 11. 仮定**??**と仮定**??**が成り立つとし，さらに定理**??**のように R_Θ を定める．$\varepsilon > 0, 1 > \delta > 0$ を与えたプライバシパラメータとする．このとき，ある定数 $B = B(\varepsilon,\delta,n,m_\ell,m_\pi,\alpha)$ が存在して，$\beta < B$ のとき，Gibbs 事後分布 $G_{\beta,D}$ は (ε,δ)-差分プライバシを満たす．

系**??**について，$m_\ell = 0$ かつ $m_\pi > 0$ の場合には，損失関数 ℓ が有界でない場合も含まれる．よって，定理**??**–(i) が直接的に適用できないモデルにも，こちらが適用できる場合がある．その他に定理**??**と異なるところは，強凸性パラメータ m_π によって β の設定可能な範囲が変化することである．これは，不自然な結果というわけではない．例えば，$\pi(\theta)$ として正規事前分布を考えると，m_π を大きくすることは分散が小さい尖った事前分布を用いることに相当し，データセット D の変化に対して事後分布がよりロバストになるためと考えられる．

一方，$m_\ell > 0$ の場合は，さきに述べた注意点のように，暗黙に Θ の有界性を仮定することになる．このとき，ℓ 自体も有界になり，定理**??**より Gibbs 事後分布からのサンプリングが ε-差分プライバシを満たすようにできる．そのため，系**??**のような結果は，一見必要ないように感じられる．しかし，具体的なモデルで $B(\varepsilon,\delta,n,m_\ell,m_\pi,\alpha)$ を計算することで，事前分布 π やサンプルサイズ n が β の上界に与える影響を調べることができる．例えば，仮定

??–(iii) における κ を n に依存しないようにとれる場合は，定理**??**における上界 (**??**) を計算することにより，B は n に関する単調増加関数となることが確認できる（4.1 節および 4.2 節の例も参照）．したがって，サンプルサイズ n が大きいほど，β を大きく設定できるため，差分プライバシが満たしやすいということになる．このようなサンプルサイズと差分プライバシとの関係は，サブサンプリングに関する性質（[**?**] などを参照）が知られているが，本論文の結果ではサブサンプリングを行うことなく β の上界という形で直接的に与えられる．

4　差分プライベート学習への応用

4.1　ロジスティック回帰

$\mathcal{Z} = \{z \in \mathbb{R}^d, \| z \|_2 \leq R\}$ とし，データは説明変数 $z \in \mathcal{Z}$ と二値ラベル $y \in \{-1, 1\}$ の組 $x = (z, y)$ であるとする．データセット $D = \{x_1, \ldots, x_n\}$ が与えられたとき，線形分類器 $f_\theta(z) = \mathrm{sgn}(a^\top z + b)$ のパラメータ $\theta = (a, b)$, $a \in \mathbb{R}^d$, $b \in \mathbb{R}$ を学習したい．

θ を決定する方法のひとつとして，ℓ_2-正則化ロジスティック回帰を考える．すなわち，損失関数を

$$\ell_{\mathrm{LR}}(\theta, x) = \log(1 + \exp(-y(a^\top x + b))) \tag{24}$$

で定義し，

$$\hat{\theta}_{\mathrm{LR}} = \operatorname*{argmin}_\theta \left\{ \frac{1}{n} \sum_{i=1}^n \ell_{\mathrm{LR}}(\theta, x_i) + \frac{\lambda}{2} \| \theta \|_2 \right\} \tag{25}$$

によって識別面ベクトル $\hat{\theta}_{\mathrm{LR}}$ を決定する．損失関数 (**??**) は定義域上で非有界であることに注意する．ロジスティック回帰は決定的な推定量であるが，対応する Gibbs 事後分布として

$$G_{\beta,D}(\theta) \propto \exp\left(-\beta \sum_{i=1}^n \ell_{\mathrm{LR}}(\theta, x_i) \right) \phi_{d+1}(\theta \mid 0, \frac{1}{n\lambda} I) \tag{26}$$

を考えることができる．ただし，$\phi_k(\cdot \mid \mu, \Sigma)$ は平均 μ，共分散行列 Σ の k-次元正規分布の密度関数である．

微分の計算により，ℓ_{LR} は R-Lipschitz であることがわかる．また，$-\log \phi_{d+1}(\theta \mid 0, \frac{1}{n\lambda} I)$ は $n\lambda$-強凸である．よって，系**??**より，

$$\beta < \frac{\varepsilon}{2R} \sqrt{\frac{n\lambda}{1 + 2\log(1/\delta)}} \tag{27}$$

とすれば Gibbs 事後分布 (**??**) は (ε, δ)-差分プライバシを満たす．(**??**) より，正則化パラメータ λ（事前分布の縮小の強さ）が大きいほど，またサンプル

サイズ n が大きいほど，β の上界は大きくとれるため，差分プライバシ保証を容易に満たすことができることがわかる．これは直感とも合うが，指数メカニズムに基づく定理**??**からは得られなかった知見である．

4.2 正規分布の制約モデルによる密度推定

定理**??**の応用として，分散が既知の正規分布における平均パラメータの推定において，共役事前分布を用いた場合を考える．ただし，正規分布のように負の対数尤度が強凸である場合，定理**??**の適用のためにはパラメータ空間を有界なところに制限した「制約モデル」を考える必要がある．

簡単のため，1 次元の場合の例を示す．データの空間 X は既知の閉区間 $X = [-a, a]$，$(a > 0)$ であるとする．データセット D が与えられたとき，データを生成する確率分布を正規分布モデル $\{\phi(\cdot \mid \theta, \sigma^2),\, \theta \in \Theta\}$ によって推定すること考える．ここで，$\Theta = [-b, b]$ はパラメータの制約区間であり，$b > 0$ はあとで定義する．

分散既知の正規分布モデルにおいて，平均に関する共役事前分布は正規分布である．そこで，上記の密度推定の問題において，θ の事前分布として，正規分布 $N(0, \tau^2)$ を区間 Θ に制限した測度を用いることにする．

通常の $\mathbb{R}$ 全体での推定を考えたとき，Gibbs 事後分布は正規分布 $N(\theta_{\beta,D}, \sigma^2_{\beta,D})$ である．ただし

$$\theta_{\beta,D} = \frac{\frac{\beta}{\sigma^2}|\frac{1}{n}\sum_{i=1}^n x_i|}{\frac{\beta}{\sigma^2} + \frac{1}{n\tau^2}} \leq \frac{\frac{\beta}{\sigma^2}a}{\frac{\beta}{\sigma^2} + \frac{1}{n\tau^2}} \leq a, \tag{28}$$

かつ

$$\sigma^2_{\beta,D} = \frac{1}{\frac{n\beta}{\sigma^2} + \frac{1}{\tau^2}} \leq \tau^2 \tag{29}$$

である．したがって，制約領域の半径を $b = a + \alpha\tau$ と定めることにより，定理**??**を適用できる．

$-\partial_\theta \log \phi(x \mid \theta, \sigma^2) = (\theta - x)/\sigma^2$ に注意すると，$L = \sup_x \sup_\theta |-\partial_\theta \log \phi(x \mid \theta, \sigma^2)| \leq (a+b)/\sigma^2$ である．これにより，定理**??**の裾確率の上界 (**??**) が計算できる．特に，系**??**の定数 B が計算できる．簡単のため，$\sigma^2 = 1, \tau^2 = 1$ とすると，(**??**) はある $C' > 0$ を用いて

$$D_+ = \frac{C'\beta^2}{n\beta + 1} \tag{30}$$

と書ける．

$$u_{\varepsilon,\delta} = \varepsilon + 2\log(1/\delta) - \sqrt{(\varepsilon + 2\log(1/\delta))^2 - \varepsilon^2} \tag{31}$$

とおくと，初等的な計算により，

$$\beta \leq B = \frac{nu_{\varepsilon,\delta} + \sqrt{n^2u^2_{\varepsilon,\delta} - 4u_{\varepsilon,\delta}C'}}{2C'} \tag{32}$$

とすれば，Gibbs 事後分布は (ε, δ)-差分プライバシを満たすことがわかる．上界 (**??**) は n に関する増加関数となっていることに注意する．

また，この例の応用として，1 次元のリッジ線形回帰に対応する Gibbs 事後分布からのサンプルが差分プライバシを満たすこともわかる．

4.3 MCMC による近似とプライバシ

損失関数および事前分布が与えられていたとしても，Gibbs 事後分布を解析的に計算することが困難な場合がある．特に，パラメータ空間が高次元の場合，Gibbs 事後分布からのサンプルを得るためには近似手法としてマルコフ連鎖モンテカルロ法 (MCMC) を用いることが一般的である．しかし，MCMC は対象となる分布に漸近的に収束するサンプリング手法であるため，有限回の反復によって得られるサンプルの分布は，真に求めたい事後分布とは一般には異なる．そのため，Gibbs 事後分布で成立していた差分プライバシ保証が破綻する可能性がある．

一方，全変動距離の意味で Gibbs 事後分布を一様に近似することができれば，差分プライバシ保証を保存することが可能である．次の定理は，ε-差分プライバシに関して知られている結果 [**?**, **?**] を (ε, δ)-差分プライバシに関して述べ直したものである．

定理 12. $\rho : \mathcal{D} \to \mathcal{M}(\Theta)$ は (ε, δ)-差分プライバシを満たすアルゴリズムとする．各データセット D に対して，全変動距離 d_{TV} の意味で $\rho(D)$ を γ-近似する，すなわち

$$d_{\mathrm{TV}}(\rho(D), \rho'(D)) \leq \gamma \tag{33}$$

を満たすアルゴリズム ρ' が存在したとする．このとき，ρ' は $(\varepsilon, \delta + (\mathrm{e}^{\varepsilon} + 1)\gamma)$-差分プライバシを満たす．

これにより，MCMC の第 k ステップにおける分布と定常分布の全変動距離が定数で抑えられるような有限の k（ミキシング時刻）を知ることができれば，本論文の定理との組み合わせで差分プライバシを満たすアルゴリズムが構成できる．ミキシング時刻に関する研究は古くから存在する [**?**] にも関わらず，厳密な定数も含めて収束レートが計算できる状況は依然として限られている．例えば，Stochastic gradient Langevin dynamics (SGLD)[**?**, **?**] など，機械学習の分野で頻繁に用いられる手法の厳密な収束レートを議論することは現状では難しい．Langevin dynamics の収束レートに関する研究としては [**?**] がある．

5 結論

本論文では，対数尤度比の測度集中不等式を証明することにより，Gibbs 事後分布からのサンプルが (ε, δ)-差分プライバシを満たす条件を解析した．特

に，対数尤度比の裾確率のバウンド (**??**)(**??**) によって，サンプルサイズ n および事前分布の縮小の強さ m_π がプライバシ保護強度に与える影響についての示唆が得られた．さらに，定理**??**，定理**??**の直接的な応用として，差分プライバシを満たすいくつかの新しいアルゴリズムが得られた．

A　定理の証明

A.1　測度集中に関する理論

定理**??**および定理**??**で利用する測度集中の理論について簡単に説明する．$(X, \mathcal{B}, \mu)$ を確率空間とする．ただし X は距離空間の構造をもつとし，$\mathcal{B}$ は Borel σ-field とする．また，X 上の実数値関数 f に対して，距離構造に基いて一般化された勾配 ∇f が定義されているとする．以下では，具体的に $X \subset \mathbb{R}^d$ で ∇f は通常の意味での微分である場合を扱う．

非負の実数値関数 f に対して，μ に関するエントロピーを

$$\mathrm{Ent}_\mu(f) = \mathbb{E}_\mu[f \log f] - \mathbb{E}_\mu[f] \log \mathbb{E}_\mu[f] \tag{34}$$

によって定義する．（右辺の積分が存在する場合に定義される．）また，f のエネルギーを

$$\mathcal{E}(f) = \mathbb{E}_\mu \parallel \nabla f \parallel_2^2 \tag{35}$$

によって定義する．ある定数 $D_{\mathrm{LS}} > 0$ が存在して，エントロピーおよびエネルギーが定義される任意の f に対して

$$\mathrm{Ent}_\mu(f^2) \leq 2 D_{\mathrm{LS}} \mathcal{E}(f) \tag{36}$$

が成立するとき，確率測度 μ は対数 Sobolev 不等式を満たすという．対数 Sobolev 不等式が成り立つとき，任意の L-Lipschitz 関数 $F : X \to \mathbb{R}$ の測度集中不等式

$$\mu\{F \geq \mathbb{E}_\mu[F] + t\} \leq \exp\left(\frac{-t^2}{2L^2 D_{\mathrm{LS}}}\right), \quad \forall t > 0 \tag{37}$$

が成り立つことがよく知られている [**?**, **?**].

対数 Sobolev 不等式の例を挙げる．$X = \mathbb{R}^d$ とし，$U : \mathbb{R}^d \to \mathbb{R}$ を C^2-級で e^{-U} が Lebesgue 可積分である関数とする．このとき，ポテンシャル関数 U をもつ Gibbs 分布 G_U の密度関数を

$$\mathrm{d}G_U(x) = Z^{-1} \mathrm{e}^{-U(x)} \mathrm{d}x \tag{38}$$

によって定義する．ここで，ポテンシャル関数 U が m-強凸である場合には，対数 Sobolev 不等式 (**??**) が定数 $D_{\mathrm{LS}} = m^{-1}$ として成立することが知られて

いる [**?**]．例えば，$U = \| x \|_2^2 / 2$ とすると (**??**) は標準正規分布と一致するが，この場合は $D_{\mathrm{LS}} = 1$ となり，Gross の対数 Sobolev 不等式と呼ばれる．

$\mathbb{R}^d$ 上の確率測度 μ が定数 D_{LS} で対数 Sobolev 不等式を満たすとき，部分集合 $X \subset \mathbb{R}^d$ 上に μ を制限した測度 $\mu|_X$ もやはり対数 Sobolev 不等式を満たすかどうか知りたい場合がある．この性質は対数 Sobolev 不等式の安定性と呼ばれ，次の結果が知られている．

定理 13 ([**?**], Corollary 3.9)**.** $\mathbb{R}^d$ 上の Gibbs 分布 (**??**) が定数 D_{LS} で対数 Sobolev 不等式を満たすとする．$X \subset \mathbb{R}^d$ は凸集合であって，ポテンシャル関数 U の X への制限 $U|_X$ は凸関数であり，

$$G_U(X) = p > 0 \tag{39}$$

が成り立つとする．このとき，G_U を X 上に制限して規格化した測度 $G'_U = G_U|_X / G_U(X)$ は，定数 $C(1 + \log(1/p))D_{\mathrm{LS}}$ で対数 Sobolev 不等式を満たす．ここで，$C > 0$ は普遍的な定数である．

本論文では，定理**??**に現れる定数 C は既知のものとして扱う．詳細については [**?**] を参照されたい．

式 (**??**) において，エントロピーの代わりに分散を評価した式

$$\mathrm{Var}_\mu(f) \leq D_{\mathrm{Poin}} \mathcal{E}(f) \tag{40}$$

を Poincaré 不等式という．Poincaré 不等式は対数 Sobolev 不等式より弱い主張であることが知られている．実際，確率測度 μ が定数 D_{LS} で対数 Sobolev 不等式を満たすとき，μ は定数 $D_{\mathrm{Poin}} = D_{\mathrm{LS}}/2$ で Poincaré 不等式を満たす [**?**].

A.2 定理??の証明

仮定より $\Theta = \mathbb{R}^d$ である．Gibbs 事後分布 $G_{\beta,D}$ は，Gibbs 分布 (**??**) において $U(\theta) = \beta \sum_{i=1}^n \ell(\theta, x_i) - \log \pi(\theta)$ としたものであることに注意する．仮定**??**–(i) より，U は m_π-強凸関数であるから，対数 Sobolev 不等式 (**??**) を定数 m_π^{-1} で満たす．

$D_1, D_2 \in \mathcal{X}^n$ を $d_H(D_1, D_2) = 1$ である 2 つのデータセットとする．一般性を失うことなく，D_1 と D_2 は第 1 成分でのみ異なる，すなわち $D_1 = (x_1, x_2, \ldots, x_n)$，$D_2 = (x'_1, x_2, \ldots, x_n)$ としてよい．このとき，仮定**??**–(ii) より

$$\left\| \nabla \log \frac{\mathrm{d}G_{\beta,D_1}}{\mathrm{d}G_{\beta,D_2}} \right\|_2 = \beta \| \nabla(\ell(\theta, x_1) - \ell(\theta, x'_1)) \|_2 \leq 2\beta L \tag{41}$$

であるから，対数尤度比は $2\beta L$-Lipschitz である．よって，対数 Sobolev 不等式が成り立つ空間での測度集中の式 (**??**) より，(**??**) を得る．

次に，KL ダイバージェンスの上界 (**??**) を求める．対数 Sobolev 不等式 (**??**) が成り立つとき，L-Lipschitz 関数 f に対して

$$\mathrm{Ent}_\mu[\mathrm{e}^f] \leq \frac{D_{\mathrm{LS}}}{2}\mathbb{E}_\mu\left[\|\nabla f\|_2^2 \mathrm{e}^f\right] \leq \frac{D_{\mathrm{LS}}L^2}{2}\mathbb{E}_\mu[\mathrm{e}^f] \tag{42}$$

が成り立つことに注意する．(**??**) に $e^f = \mathrm{d}G_{\beta,D_1}/\mathrm{d}G_{\beta,D_2}$ を代入することにより (**??**) が得られる．

A.3　定理??の証明（概略）

仮定**??**–(ii) より，Gibbs 事後分布を Gibbs 分布とみなしたときのポテンシャルは，$\mathbb{R}^d$ 上に $nm_\ell\beta + m_\pi$-強凸となるように拡張される．このポテンシャルをもつ $\mathbb{R}^d$ 上の Gibbs 分布を $\bar{G}_{\beta,D}$ とおく．$\bar{G}_{\beta,D}(\Theta) \geq p$ であるとき，定理**??**より，Gibbs 事後分布 $G_{\beta,D}$ は定数 $C(1+\log(1/p))/(nm_\ell\beta + m_\pi)$ で対数 Sobolev 不等式を満たす．

そこで，任意の D に対して，$\bar{G}_{\beta,D}(\Theta)$ が一様な下界をもつように Θ の半径を定めることを考える．今，仮定**??**–(iii) より，$\bar{G}_{\beta,D}(\Theta)$ の平均は上界 κ をもつ．また，Poincaré 不等式 (**??**) より，分散は $d/(nm_\ell\beta + m_\pi) \leq d/m_\pi$ で上から抑えられる．よって，定理のように R_Θ を定めれば，Chebyshev の不等式から，任意の D に対して $\bar{G}_{\beta,D}(\Theta) \geq \quad 1-\alpha^{-2}$ が成り立つ．ゆえに，Gibbs 事後分布 $G_{\beta,D}$ は定数

$$\frac{C(1+\log(\alpha^2/(\alpha^2-1)))}{nm_\ell\beta + m_\pi} \tag{43}$$

で対数 Sobolev 不等式を満たす．あとは，定理**??**と同様の議論によって，測度集中不等式 (**??**) を得る．

A.4　定理??の証明

μ_i, μ_i', $(i=1,2)$ を同じ測度空間で定義された確率測度とし，$d_{\mathrm{TV}}(\mu_i,\mu_i') \geq \gamma$, $(i=1,2)$，および任意の可測集合 A に対し

$$\mu_1'(A) \leq \mathrm{e}^\varepsilon \mu_2'(A) + \delta \tag{44}$$

が成り立つとする．このとき，

$$\begin{aligned}\mu_1(A) &\leq \mu_1'(A) + \gamma \leq \mathrm{e}^\varepsilon\mu_2'(A) + \delta + \gamma \\ &\leq \mathrm{e}^\varepsilon\mu_2(A) + (\mathrm{e}^\varepsilon + 1)\gamma + \delta\end{aligned} \tag{45}$$

が成り立つことから主張を得る．