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．