On the Differential Privacy of Bayesian Inference

Zuhe Zhang[†] and **Benjamin I. P. Rubinstein**[⋆]
[†]School of Mathematics and Statistics,
[⋆]Department of Computing and Information Systems,
The University of Melbourne, Australia
zhang.zuhe@gmail.com, brubinstein@unimelb.edu.au

Christos Dimitrakakis[‡§]
[‡]Univ-Lille-3, France
[§]Chalmers University of Technology, Sweden
christos.dimitrakakis@gmail.com

Abstract

We study how to communicate findings of Bayesian inference to third parties, while preserving the strong guarantee of differential privacy. Our main contributions are four different algorithms for private Bayesian inference on probabilistic graphical models. These include two mechanisms for adding noise to the Bayesian updates, either directly to the posterior parameters, or to their Fourier transform so as to preserve update consistency. We also utilise a recently introduced posterior sampling mechanism, for which we prove bounds for the specific but general case of discrete Bayesian networks; and we introduce a maximum-a-posteriori private mechanism. Our analysis includes utility and privacy bounds, with a novel focus on the influence of graph structure on privacy. Worked examples and experiments with Bayesian naïve Bayes and Bayesian linear regression illustrate the application of our mechanisms.

Introduction

We consider the problem faced by a statistician $\mathscr{B}$ who analyses data and communicates her findings to a third party $\mathscr{A}$. While $\mathscr{B}$ wants to learn as much as possible *from* the data, she doesn't want $\mathscr{A}$ to learn *about* any individual datum. This is for example the case where $\mathscr{A}$ is an insurance agency, the data are medical records, and $\mathscr{B}$ wants to convey the efficacy of drugs to the agency, without revealing the specific illnesses of individuals in the population. Such requirements of *privacy* are of growing interest in the learning (Chaudhuri and Hsu 2012; Duchi, Jordan, and Wainwright 2013), theoretical computer science (Dwork and Smith 2009; McSherry and Talwar 2007) and databases communities (Barak et al. 2007; Zhang et al. 2014) due to the impact on individual privacy by real-world data analytics.

In our setting, we assume that $\mathscr{B}$ is using *Bayesian inference* to draw conclusions from observations of a system of random variables by updating a prior distribution on parameters (*i.e.*, *latent* variables) to a posterior. Our goal is to release an approximation to the posterior that preserves privacy. We adopt the formalism of *differential privacy* to characterise how easy it is for $\mathscr{A}$ to discover facts about the *individual* data from the *aggregate* posterior. Releasing the posterior permits external parties to make further inferences at will. For example, a third-party pharmaceutical might use the released posterior as a prior on the efficacy of drugs, and update it with their own patient data. Or they could form a predictive posterior for classification or regression, all while preserving differential privacy of the original data.

Our focus in this paper is Bayesian inference in *probabilistic graphical models* (PGMs), which are popular as a tool for modelling conditional independence assumptions. Similar to the effect on statistical and computational efficiency of non-private inference, a central tenet of this paper is that independence structure should impact privacy. *Our mechanisms and theoretical bounds are the first to establish such a link between PGM graph structure and privacy.*

Main Contributions. We develop the first mechanisms for Bayesian inference on the flexible PGM framework (*cf.* Table 1). We propose two posterior perturbation mechanisms for networks with likelihood functions from exponential families and conjugate priors, that add Laplace noise (Dwork et al. 2006) to posterior parameters (or their Fourier coefficients) to preserve privacy. The latter achieves stealth through consistent posterior updates. For general Bayesian networks, posteriors may be non-parametric. In this case, we explore a mechanism (Dimitrakakis et al. 2014) which samples from the posterior to answer queries—no additional noise is injected. We complement our study with a maximum *a posteriori* estimator that leverages the exponential mechanism (McSherry and Talwar 2007). Our utility and privacy bounds connect privacy and graph/dependency structure, and are complemented by illustrative experiments with Bayesian naïve Bayes and linear regression.

Related Work. Many individual learning algorithms have been adapted to maintain differential privacy, including regularised logistic regression (Chaudhuri and Monteleoni 2008), the SVM (Rubinstein et al. 2012; Chaudhuri, Monteleoni, and Sarwate 2011), PCA (Chaudhuri, Sarwate, and Sinha 2012), the functional mechanism (Zhang et al. 2012) and trees (Jagannathan, Pillaipakkamnatt, and Wright 2009).

Probabilistic graphical models have been used to preserve privacy. Zhang et al. (2014) learned a graphical model from data, in order to generate *surrogate data* for release; while Williams and McSherry (2010) fit a model to the response of private mechanisms to clean up output and improve accuracy. Xiao and Xiong (2012) similarly used Bayesian credible intervals to increase the utility of query responses.

Little attention has been paid to private inference in the Bayesian setting. We seek to adapt Bayesian inference to preserve differential privacy when releasing posteriors. Dimitrakakis et al. (2014; 2015) introduce a differentially-private mechanism for Bayesian inference based on posterior sampling—a mechanism on which we build—while Zheng (2015) considers further refinements. Wang, Fienberg, and Smola (2015) explore Monte Carlo approaches to Bayesian inference using the same mechanism, while Mir (2012) was the first to establish differential privacy of the Gibbs estimator (McSherry and Talwar 2007) by minimising risk bounds.

This paper is the first to develop mechanisms for differential privacy under the general framework of Bayesian inference on multiple, dependent r.v.'s. Our mechanisms consider graph structure and include a purely Bayesian approach that only places conditions on the prior. We show how the (stochastic) Lipschitz assumptions of Dimitrakakis et al. (2014) lift to graphs of r.v.'s, and bound KL-divergence when releasing an empirical posterior based on a modified prior. While Chaudhuri, Monteleoni, and Sarwate (2011) achieve privacy in regularised Empirical Risk Minimisation through objective randomisation, we do so through conditions on priors. We develop an alternate approach that uses the additive-noise mechanism of Dwork et al. (2006) to perturb posterior parameterisations; and we apply techniques due to Barak et al. (2007), who released marginal tables that maintain consistency in addition to privacy, by adding Laplace noise in the Fourier domain. Our motivation is novel: we wish to guarantee privacy against omniscient attackers and stealth against unsuspecting third parties.

Problem Setting

Consider a Bayesian statistician $\mathscr{B}$ estimating the parameters $\boldsymbol{\theta}$ of some family of distributions $\mathcal{F}_\Theta = \{ p_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \}$ on a system of r.v.'s $\boldsymbol{X} = \{ X_i : i \in \mathcal{I} \}$, where $\mathcal{I}$ is an index set, with observations denoted $x_i \in \mathcal{X}_i$, where $\mathcal{X}_i$ is the sample space of X_i. $\mathscr{B}$ has a prior distribution[1] ξ on Θ reflecting her prior belief, which she updates on an observation $\boldsymbol{x}$ to obtain posterior

$$\xi(B \mid \boldsymbol{x}) = \frac{\int_B p_{\boldsymbol{\theta}}\left(\boldsymbol{x}\right) \mathrm{d}\xi\left(\boldsymbol{\theta}\right)}{\phi(\boldsymbol{x})}, \ \forall B \in \mathfrak{S}_\Theta$$

where $\phi(\boldsymbol{x}) \triangleq \int_\Theta p_{\boldsymbol{\theta}}\left(\boldsymbol{x}\right) d\xi\left(\boldsymbol{\theta}\right)$. Posterior updates are iterated over an i.i.d. dataset $D \in \mathcal{D} = (\prod_i \mathcal{X}_i)^n$ to $\xi(\cdot \mid D)$.

$\mathscr{B}$'s goal is to communicate her posterior distribution to a third party $\mathscr{A}$, while limiting the information revealed about the original data. From the point of view of the data provider, $\mathscr{B}$ is a trusted party.[2] However, she may still inadvertently reveal information. We assume that $\mathscr{A}$ is computationally unbounded, and has knowledge of the prior ξ and the family $\mathcal{F}_\Theta$. To guarantee that $\mathscr{A}$ can gain little additional information about D from their communication, $\mathscr{B}$ uses Bayesian inference to learn from the data, and a differentially-private posterior to ensure disclosure to $\mathscr{A}$ is carefully controlled.

Probabilistic Graphical Models

Our main results focus on PGMs which model conditional independence assumptions with joint factorisation

$$p_{\boldsymbol{\theta}}(\boldsymbol{x}) = \prod_{i \in \mathcal{I}} p_{\boldsymbol{\theta}}\left(x_i \mid x_{\pi_i}\right), \quad x_{\pi_i} = \{ x_j : j \in \pi_i \} \ \ ,$$

where π_i are the parents of the i-th variable in a Bayesian network—a directed acyclic graph with r.v.'s as nodes.

Example 1. *For concreteness, we illustrate some of our mechanisms on systems of Bernoulli r.v.'s $X_i \in \{0, 1\}$. In that case, we represent the conditional distribution of X_i given its parents as Bernoulli with parameters $\theta_{i,j} \in [0, 1]$:*

$$(X_i \mid X_{\pi_i} = j) \sim \textit{Bernoulli}(\theta_{i,j}) \ \ .$$

The choice of conjugate prior $\xi(\boldsymbol{\theta}) = \prod_{i,j} \xi_{i,j}(\theta_{i,j})$ has Beta marginals with parameters $\alpha_{i,j}, \beta_{i,j}$, so that:

$$(\theta_{i,j} \mid \alpha_{i,j} = \alpha, \beta_{i,j} = \beta) \sim \textit{Beta}(\alpha, \beta) \ \ .$$

Given observation $\boldsymbol{x}$, the updated posterior Beta parameters are $\alpha_{i,j} := \alpha_{i,j} + x_i$ and $\beta_{i,j} := \beta_{i,j} + (1 - x_i)$ if $x_{\pi_i} = j$.

Differential Privacy

$\mathscr{B}$ communicates to $\mathscr{A}$ by releasing information about the posterior distribution, via randomised mechanism M that maps dataset $D \in \mathcal{D}$ to a response in set $\mathcal{Y}$. Dwork et al. (2006) characterise when such a mechanism is private:

Definition 1 (Differential Privacy)**.** *A randomised mechanism $M : \mathcal{D} \to \mathcal{Y}$ is (ϵ, δ)-DP if for any* neighbouring *$D, \tilde{D} \in \mathcal{D}$, and measurable $B \subseteq \mathcal{Y}$:*

$$\mathbb{P}[M(D) \in B] \le e^\epsilon \mathbb{P}[M(\tilde{D}) \in B] + \delta,$$

where $D = (\boldsymbol{x}^i)_{i=1}^n$, $\tilde{D} = (\tilde{\boldsymbol{x}}^i)_{i=1}^n$ are neighbouring if $\boldsymbol{x}^i \neq \tilde{\boldsymbol{x}}^i$ for at most one i.

This definition requires that neighbouring datasets induce similar response distributions. Consequently, it is impossible for $\mathscr{A}$ to identify the true dataset from bounded mechanism query responses. Differential privacy assumes no bounds on adversarial computation or auxiliary knowledge.

Privacy by Posterior Perturbation

One approach to differential privacy is to use additive Laplace noise (Dwork et al. 2006). Previous work has focused on the addition of noise directly to the outputs of a non-private mechanism. We are the first to apply Laplace noise to the posterior parameter updates.

Laplace Mechanism on Posterior Updates

Under the setting of Example 1, we can add Laplace noise to the posterior parameters. Algorithm 1 releases perturbed parameter updates for the Beta posteriors, calculated simply by *counting*. It then adds zero-mean Laplace-distributed noise

[1] Precisely, a probability measure on a σ-algebra $(\Theta_i, \mathfrak{S}_{\Theta_i})$.

[2] Cryptographic tools for untrusted $\mathscr{B}$ do not prevent information leakage to $\mathscr{A}$ *cf. e.g.*, (Pagnin et al. 2014).

	DBN only	**Privacy**	**Utility type**	**Utility bound**
Laplace	✓	$(\epsilon, 0)$	closeness of posterior	$\mathcal{O}\left(mn\ln n\right)\left[1-\exp\left(-\frac{n\epsilon}{2\lvert\mathcal{I}\rvert}\right)\right]+\sqrt{-\mathcal{O}\left(mn\ln n\right)\ln\delta}$
Fourier	✓	$(\epsilon, 0)$	close posterior params	$\frac{4\lvert\mathcal{N}_{\mathcal{I}}\rvert}{\epsilon}\left(2^{\lvert\pi_i\rvert}\log\frac{\lvert\mathcal{N}_{\mathcal{I}}\rvert}{\delta}+t\lvert\mathcal{N}_{\mathcal{I}}\rvert\right)$
Sampler	✗	$(2L, 0)$ if Lipschitz; or $(0, \sqrt{M/2})$ stochastic Lipschitz	expected utility functional wrt posterior	$\mathcal{O}\left(\eta+\sqrt{\ln(1/\delta)/N}\right)$ (Dimitrakakis et al. 2015)
MAP	✗	$(\epsilon, 0)$	closeness of MAP	$\mathbb{P}(S_{2t}^c)\leq\exp(-\epsilon t)/\xi\left(S_t\right)$

Table 1: Summary of the privacy/utility guarantees for this paper's mechanisms. See below for parameter definitions.

Algorithm 1 Laplace Mechanism on Posterior Updates

```
Input data $D$; graph $\mathcal{I}, \{\pi_i \mid i \in \mathcal{I}\}$; parameter $\epsilon > 0$
calculate posterior updates: $\Delta\alpha_{i,j}, \Delta\beta_{i,j}$ for all
    $i \in \mathcal{I}, j \in \{0,1\}^{|\pi_i|}$
perturb updates: $\Delta\alpha'_{i,j} \triangleq \Delta\alpha_{i,j} + \mathrm{Lap}\left(\frac{2|\mathcal{I}|}{\epsilon}\right)$,
    $\Delta\beta'_{i,j} \triangleq \Delta\beta_{i,j} + \mathrm{Lap}\left(\frac{2|\mathcal{I}|}{\epsilon}\right)$.
truncate: $Z^{(1)}_{i,j} \triangleq \mathbf{1}_{[0,n]}(\Delta\alpha'_{i,j}), Z^{(2)}_{i,j} \triangleq \mathbf{1}_{[0,n]}(\Delta\beta'_{i,j})$
output $\boldsymbol{Z}_{i,j} = (Z^{(1)}_{i,j}, Z^{(2)}_{i,j})$
```

to the updates $\Delta\boldsymbol{\omega} = (\cdots, \Delta\alpha_{i,j}, \Delta\beta_{i,j}, \cdots)$. This is the final dependence on D. Finally, the perturbed updates $\Delta\boldsymbol{\omega}'$ are truncated at zero to rule out invalid Beta parameters and are upper truncated at n. This yields an upper bound on the raw updates and facilitates an application of McDiarmid's bounded-differences inequality (*cf.* full report Zhang, Rubinstein, and Dimitrakakis 2015) in our utility analysis. Note that this truncation only improves utility (relative to the utility pre-truncation), and does not affect privacy.

Privacy. To establish differential privacy of our mechanism, we must calculate a Lipschitz condition for the vector $\boldsymbol{\Delta\omega}$ called *global sensitivity* (Dwork et al. 2006).

Lemma 1. *For any* neighbouring *datasets* $D, \tilde{D}$*, the corresponding updates* $\Delta\boldsymbol{\omega}, \Delta\tilde{\boldsymbol{\omega}}$ *satisfy* $\left\|\Delta\boldsymbol{\omega} - \Delta\tilde{\boldsymbol{\omega}}\right\|_1 \leq 2|\mathcal{I}|$.

Proof. By changing the observations of one datum, at most two counts associated with each X_i can change by 1. □

Corollary 1. *Algorithm 1 preserves* ϵ*-differential privacy.*

Proof. Based on Lemma 1, the intermediate $\Delta\boldsymbol{\omega}'$ preserve ϵ-differential privacy (Dwork et al. 2006). Since truncation depends only on $\Delta\boldsymbol{\omega}'$, the $\boldsymbol{Z}$ preserves the same privacy. □

Utility on Updates. Before bounding the effect on the posterior of the Laplace mechanism, we demonstrate a utility bound on the posterior update counts.

Proposition 1. *With probability at least* $1-\delta$*, for* $\delta \in (0,1)$*, the update counts computed by Algorithm 1 are close to the non-private counts*

$$\|\Delta\boldsymbol{\omega} - \Delta\boldsymbol{\omega}'\|_\infty \leq \frac{2|\mathcal{I}|}{\epsilon}\ln\left(\frac{2m}{\delta}\right) \quad ,$$

where $m = \sum_{i \in I} 2^{|\pi_i|}$.

This bound states that w.h.p., none of the updates can be perturbed beyond $O(|\mathcal{I}|^2/\epsilon)$. This implies the same bound on the deviation between $\Delta\boldsymbol{\omega}$ and the revealed truncated $\boldsymbol{Z}$.

Utility on Posterior. We derive our main utility bounds for Algorithm 1 in terms of posteriors, proved in the full report (Zhang, Rubinstein, and Dimitrakakis 2015). We abuse notation, and use ξ to refer to the prior density; its meaning will be apparent from context. Given priors $\xi_{i,j}(\theta_{i,j}) = \mathrm{Beta}\left(\alpha_{i,j}, \beta_{i,j}\right)$, the posteriors on n observations are

$$\xi_{i,j}(\theta_{i,j}|D) = \mathrm{Beta}(\alpha_{i,j} + \Delta\alpha_{i,j}, \beta_{i,j} + \Delta\beta_{i,j}) \enspace .$$

The privacy-preserving posterior parametrised by the output of Algorithm 1 is

$$\xi'_{i,j}(\theta_{i,j}|D) = \mathrm{Beta}\left(\alpha_{i,j} + Z^{(1)}_{i,j}, \beta_{i,j} + Z^{(2)}_{i,j}\right) \enspace .$$

It is natural to measure utility by the KL-divergence between the joint product posteriors $\xi(\boldsymbol{\theta}|D)$ and $\xi'(\boldsymbol{\theta}|D)$, which is the sum of the component-wise divergences, with each having known closed form. In our analysis, the divergence is a random quantity, expressible as the sum $\sum_{i,j}^{m} f_{i,j}(\boldsymbol{Z}_{i,j})$, where the randomness is due to the added noise. We demonstrate this r.v. is not too big, w.h.p.

Theorem 1. *Let* $m = \sum_{i \in \mathcal{I}} 2^{|\pi_i|}$*. Assume that* $Z_{i,j}$ *are independent and* f *is a mapping from* $\mathcal{Z}^m$ *to* $\mathbb{R}$*:* $f(\cdots, \boldsymbol{z}_{i,j}, \cdots) \triangleq \sum_{i,j} f_{i,j}(\boldsymbol{z}_{i,j})$*. Given* $\delta > 0$*, we have*

$$\mathbb{P}\left[f(\boldsymbol{Z}) \geq \mathbb{E}(f(\boldsymbol{Z})) + \left(-\frac{1}{2}\sum_{i,j} c_{i,j}\ln\delta\right)^{\frac{1}{2}}\right] \leq \delta$$

where $c_{i,j} \leq (2n+1)\ln[(\alpha_{i,j}+n+1)+(\beta_{i,j}+n+1))$ *and* $\mathbb{E}(f_{i,j}(\boldsymbol{Z}_{i,j})] \leq n\ln((\alpha_{i,j} + \Delta\alpha_{i,j})(\beta_{i,j} + \Delta\beta_{i,j})) = \mathcal{U}$*. Moreover, when* $n \geq b = \frac{2|\mathcal{I}|}{\epsilon}$*, the bound for expectation can be refined as the following*

$$\ln[(\alpha_{i,j} + n + 1)(\beta_{i,j} + n + 1)]\left(\frac{n}{2}\exp\left(-\frac{n\epsilon}{2|\mathcal{I}|}\right)\right) \enspace .$$

The loss of utility measured by KL-divergence is no more than

$$\mathcal{O}\left(mn\ln n\right)\left[1 - \exp\left(-\frac{n\epsilon}{2|\mathcal{I}|}\right)\right] + \sqrt{-\mathcal{O}\left(mn\ln n\right)\ln\delta}$$

with probability at least $1-\delta$.

Note that m depends on the structure of the network: bounds are better for networks with an underlying graph having smaller average in-degree.

Laplace Mechanism in the Fourier Domain

Algorithm 1 follows *Kerckhoffs's Principle* (Kerckhoffs 1883) of "no security through obscurity": differential privacy defends against a mechanism-aware attacker. However *additional stealth* may be required in certain circumstances. An oblivious observer will be tipped off to our privacy-preserving activities by our independent perturbations, which are likely inconsistent with one-another (*e.g.*, noisy counts for X_1, X_2 and X_2, X_3 will say different things about X_2). To achieve differential privacy and stealth, we turn to Barak et al. (2007)'s study of consistent marginal contingency table release. This section presents a particularly natural application to Bayesian posterior updates.

Denote by $h \in \mathbb{R}^{\{0,1\}^{|\mathcal{I}|}}$ the *contingency table* over r.v.'s $\mathcal{I}$ induced by D: *i.e.*, for each combination of variables $j \in \{0,1\}^{|\mathcal{I}|}$, component or *cell* h_j is a non-negative count of the observations in D with characteristic j. Geometrically h is a real-valued function over the $|\mathcal{I}|$-dimensional Boolean hypercube. Then the parameter delta's of our first mechanism correspond to cells of $(|\pi_i| + 1)$-way marginal contingency tables $\mathrm{C}^{\overline{\pi}_i}(h)$ where vector $\overline{\pi}_i \triangleq \pi_i + e_i$ and the projection/marginalisation operator is defined as

$$\left(\mathrm{C}^j(h)\right)_\gamma \quad \triangleq \quad \sum_{\eta:\langle \eta, j\rangle = \gamma} h_\eta \quad . \tag{1}$$

We wish to release these statistics as before, however we will not represent them under their Euclidean coordinates but instead in the Fourier basis $\{f^j : j \in \{0,1\}^{|\mathcal{I}|}\}$ where

$$f^j_\gamma \quad \triangleq \quad (-1)^{\langle \gamma, j\rangle} 2^{-|\mathcal{I}|/2} \quad .$$

Due to this basis structure and linearity of the projection operator, any marginal contingency table must lie in the span of few projections of Fourier basis vectors (Barak et al. 2007):

Theorem 2. *For any table $h \in \mathbb{R}^{\{0,1\}^{|\mathcal{I}|}}$ and set of variables $j \in \{0,1\}^{|\mathcal{I}|}$, the marginal table on j satisfies $\mathrm{C}^j(h) = \sum_{\gamma \preceq j} \langle f^\gamma, h\rangle \, \mathrm{C}^j(f^\gamma)$.*

This states that marginal j lies in the span of only those (projected) basis vectors f^γ with γ contained in j. The number of values needed to update X_i is then $2^{|\pi_i|+1}$, potentially far less than suggested by (1). To release updates for two r.v.'s $i, j \in \mathcal{I}$ there may well be significant overlap $\langle \overline{\pi}_i, \overline{\pi}_j\rangle$; we need to release once, coefficients $\langle f^\gamma, h\rangle$ for γ in the downward closure of variable neighbourhoods:

$$\mathcal{N}_\mathcal{I} \quad \triangleq \quad \bigcup_{i\in\mathcal{I}} \bigcup_{j \preceq \overline{\pi}_i} j \quad .$$

Privacy. By (Barak et al. 2007, Theorem 6) we can apply Laplace additive noise to release these Fourier coefficients.

Corollary 2. *For any $\epsilon > 0$, releasing for each $\gamma \in \mathcal{N}_\mathcal{I}$ the Fourier coefficient $\langle f^\gamma, h\rangle + \mathrm{Lap}\left(2|\mathcal{N}_\mathcal{I}|\epsilon^{-1}2^{-|\mathcal{I}|/2}\right)$ (and Algorithm 2) preserves ϵ-differential privacy.*

Remark 1. *Since $|\mathcal{N}_\mathcal{I}| \leq |\mathcal{I}|2^{1+\max_{i\in\mathcal{I}} \mathrm{indeg}(i)}$, at worst we have noise scale $|\mathcal{I}|2^{2+\max_i \mathrm{indeg}(i) - |\mathcal{I}|/2}/\epsilon$. This compares favourably with Algorithm 1's noise scale provided no r.v. is child to more than half the graph. Moreover the denser the graph—the more overlap between nodes' parents and the less conditional independence assumed—the greater the reduction in scale. This is intuitively appealing.*

Consistency. What is gained by passing to the Fourier domain, is that the perturbed marginal tables of Corollary 2 are consistent: anything in the span of projected Fourier basis vectors corresponds to some valid contingency table on $\mathcal{I}$ with (possibly negative) real-valued cells (Barak et al. 2007).

Algorithm 2 Laplace Mechanism in the Fourier Domain

```
Input data $D$; graph $\mathcal{I}, \{\pi_i \mid i \in \mathcal{I}\}$; prior parameters $\boldsymbol{\alpha}, \boldsymbol{\beta} \succeq \mathbf{0}$; parameters $t, \epsilon > 0$
define contingency table $h \in \mathbb{R}^{\{0,1\}^{|\mathcal{I}|}}$ on $D$
define downward closure $\mathcal{N}_\mathcal{I} = \bigcup_{i\in\mathcal{I}} \bigcup_{j\preceq\overline{\pi}_i} j$
for $\gamma \in \mathcal{N}_\mathcal{I}$ do
    Fourier coefficient $z_\gamma = \langle f^\gamma, h\rangle + \mathrm{Lap}\left(\frac{2|\mathcal{N}_\mathcal{I}|}{\epsilon 2^{|\mathcal{I}|/2}}\right)$
end for
increment first coefficient $z_\mathbf{0} \leftarrow z_\mathbf{0} + \frac{4t|\mathcal{N}_\mathcal{I}|^2}{\epsilon 2^{|\mathcal{I}|/2}}$
for $i \in \mathcal{I}$ do
    project marginal for $X_i$ as $h^i = \sum_{\gamma\preceq\overline{\pi}_i} z_\gamma \mathrm{C}^{\overline{\pi}_i}(f^\gamma)$
    for $j \preceq \pi_i$ do
        output posterior param $\left(\alpha_{ij} + h^i_{e_i+j}, \beta_{ij} + h^i_j\right)$
    end for
end for
```

Non-negativity. So far we have described the first stage of Algorithm 2. The remainder yields *stealth* by guaranteeing releases that are non-negative w.h.p. We adapt an idea of Barak et al. (2007) to increase the coefficient of Fourier basis vector $f^\mathbf{0}$, affecting a small increment to each cell of the contingency table. While there is an exact minimal amount that would guarantee non-negativity, it is data dependent. Thus our efficient $\mathcal{O}(|\mathcal{N}_\mathcal{I}|)$-time approach is randomised.

Corollary 3. *For $t > 0$, adding $4t|\mathcal{N}_\mathcal{I}|^2\epsilon^{-1}2^{-k/2}$ to $f^\mathbf{0}$'s coefficient induces a non-negative table w.p. $\geq 1 - \exp(-t)$.*

Parameter t trades off between the probability of non-negativity and the resulting (minor) loss to utility. In the rare event of negativity, re-running Algorithm 2 affords another chance of stealth at the cost of privacy budget ϵ. We could alternatively truncate to achieve validity, sacrificing stealth but not privacy.

Utility. Analogous to Proposition 1, each perturbed marginal is close to its unperturbed version w.h.p.

Theorem 3. *For each $i \in \mathcal{I}$ and $\delta \in (0,1)$, the perturbed tables in Algorithm 2 satisfy with probability at least $1 - \delta$:*

$$\left\|\mathrm{C}^{\overline{\pi}_i}(h) - h^i\right\|_1 \quad \leq \quad \frac{4|\mathcal{N}_\mathcal{I}|}{\epsilon}\left(2^{|\pi_i|}\log\frac{|\mathcal{N}_\mathcal{I}|}{\delta} + t|\mathcal{N}_\mathcal{I}|\right) \quad .$$

Note that the scaling of this bound is reasonable since the table h^i involves $2^{|\pi_i|+1}$ cells.

Privacy by Posterior Sampling

For general Bayesian networks, $\mathscr{B}$ can release samples from the posterior (Dimitrakakis et al. 2014) instead of perturbed samples of the posterior's parametrisation. We now develop a calculus of building up (stochastic) Lipschitz properties of systems of r.v.'s that are locally (stochastic) Lipschitz. Given smoothness of the entire network, differential privacy and utility of posterior sampling follow.

(Stochastic) Lipschitz Smoothness of Networks

The distribution family $\{p_\theta : \theta \in \Theta\}$ on outcome space $\mathcal{S}$, equipped with pseudo metric[3] ρ, is *Lipschitz continuous* if

Assumption 1 (Lipschitz Continuity)**.** *Let $d(\cdot, \cdot)$ be a metric on $\mathbb{R}$. There exists $L > 0$ such that, for any $\theta \in \Theta$:*

$$d(p_\theta(x), p_\theta(y)) \leq L\rho(x, y), \forall x, y \in \mathcal{S}.$$

We fix the distance function d to be the absolute log-ratio (*cf.* differential privacy). Consider a general Bayesian network. The following lemma shows that the individual Lipschitz continuity of the conditional likelihood at every $i \in \mathcal{I}$ implies the global Lipschitz continuity of the network.

Lemma 2. *If there exists $\boldsymbol{L} = (L_1, \cdots, L_{|\mathcal{I}|}) \geq \mathbf{0}$ such that $\forall i \in \mathcal{I}$, $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X} = \prod_{i=1}^{|\mathcal{I}|} \mathcal{X}_i$ we have $d(p_{\boldsymbol{\theta}}(x_i|x_{\pi_i}), p_{\boldsymbol{\theta}}(y_i|y_{\pi_i})) \leq L_i\rho_i(x_i, y_i)$, then $d(p_{\boldsymbol{\theta}}(\boldsymbol{x}), p_{\boldsymbol{\theta}}(\boldsymbol{y})) \leq \|\boldsymbol{L}\|_\infty \boldsymbol{\rho}(\boldsymbol{x}, \boldsymbol{y})$ where $\boldsymbol{\rho}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{|I|} \rho_i(x_i, y_i)$.*

Note that while Lipschitz continuity holds uniformly for some families *e.g.*, the exponential distribution, this is not so for many useful distributions such as the Bernoulli. In such cases a relaxed assumption requires that the prior be concentrated on smooth regions.

Assumption 2 (Stochastic Lipschitz Continuity)**.** *Let the set of L-Lipschitz θ be*

$$\Theta_L \triangleq \left\{\theta \in \Theta : \sup_{x,y \in \mathcal{S}} \{d(p_\theta(x), p_\theta(y)) - L\rho(x, y)\} \leq 0\right\}$$

Then there exists constants $c, L_0 > 0$ such that, $\forall L \geq L_0$: $\xi(\Theta_L) \geq 1 - e^{-cL}$.

Lemma 3. *For the conditional likelihood at each node $i \in \mathcal{I}$, define the set $\Theta_{i,L}$ of parameters for which Lipschitz continuity holds with Lipschitz constant L. If $\exists \mathbf{c} = (c_1, \cdots, c_{|\mathcal{I}|})$ such that $\forall i, L \geq L_0$, $\xi(\Theta_{i,L}) \geq 1 - e^{-c_i L}$, then $\xi(\Theta_L) \geq 1 - e^{-c' L}$, where $c' = \min_{i \in \mathcal{I}} c_i - \ln|\mathcal{I}|/L_0$ when $|\mathcal{I}| \leq e^{L_0 \min_{i \in \mathcal{I}} c_i}$.*

Therefore, (Dimitrakakis et al. 2015, Theorem 7) asserts differential privacy of the Bayesian network's posterior.

Theorem 4. *Differential privacy is satisfied using the log-ratio distance, for all $B \in \mathfrak{S}_\Theta$ and $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{\mathcal{X}}$:*

1. *Under the conditions in Lemma 2:*

$$\xi(B \mid \boldsymbol{x}) \leq \exp\{2L\boldsymbol{\rho}(\boldsymbol{x}, \boldsymbol{y})\}\xi(B \mid \boldsymbol{y})$$

i.e., *the posterior ξ is $(2\|\boldsymbol{L}\|_\infty, 0)$-differentially private under pseudo-metric $\boldsymbol{\rho}(\boldsymbol{x}, \boldsymbol{y})$.*

2. *Under the conditions in Lemma 3, if $\boldsymbol{\rho}(\boldsymbol{x}, \boldsymbol{y}) \leq (1 - \delta)c$ uniformly for all $\boldsymbol{x}, \boldsymbol{y}$ for some $\delta \in (0, 1)$:*

$$|\xi(B \mid \boldsymbol{x}) - \xi(B \mid \boldsymbol{y})| \leq \sqrt{\frac{M}{2} \cdot \max\{\boldsymbol{\rho}(\boldsymbol{x}, \boldsymbol{y}), 1\}}$$

where $M = \Big(\frac{\kappa}{c} + L_0(\frac{1}{1-e^{-\omega}} + 1) + \ln C + \ln\left(e^{-L_0\delta c}(e^{-\omega(1-\delta)} - e^{-\omega})^{-1} + e^{L_0(1-\delta)c}\right)\Big) C$; constants $\kappa = 4.91081$ and $\omega = 1.25643$; $C = \prod_i^{|I|} C_i$; and

$$C_i = \sup_{\boldsymbol{x} \in \boldsymbol{\mathcal{X}}} \frac{p_{\theta^\star_{i,\mathrm{MLE}}}(x_i \mid x_{\pi_i})}{\int_{\Theta_i} p_{\theta_i}(x_i \mid x_{\pi_i}) d\xi(\theta_i)} \ ,$$

the ratio between the maximum and marginal likelihoods of each likelihood function. Note that $M = \mathcal{O}\left(\left(\frac{1}{c} + \ln C + L_0\right) C\right)$ i.e., *the posterior ξ is $\left(0, \sqrt{\frac{M}{2}}\right)$-differentially private under pseudo-metric $\sqrt{\boldsymbol{\rho}}$ for $\boldsymbol{\rho}(\boldsymbol{x}, \boldsymbol{y}) \geq 1$.*

MAP by the Exponential Mechanism

As an application of the posterior sampler, we now turn to releasing MAP point estimates via the exponential mechanism (McSherry and Talwar 2007), which samples responses from a likelihood exponential in some score function. By selecting a utility function that is maximised by a target non-private mechanism, the exponential mechanism can be used to privately approximate that target with high utility. It is natural then to select as our utility u the posterior likelihood $\xi(\cdot|D)$. This u is maximised by the MAP estimate.

Algorithm 3 Mechanism for MAP Point Estimates

```
1: Input data $D$; prior $\xi(\cdot)$; appropriate smoothness parameters $c, L, M > 0$; parameters distance $r > 0$, privacy $\epsilon > 0$
2: calculate posterior $\xi(\theta|D)$
3: set $\Delta = \begin{cases} \sqrt{Lr}\ , & \text{if Lipschitz continuous} \\ \sqrt{0.5M}\ , & \text{if stochastic Lipschitz} \end{cases}$
4: output $\hat{\theta}$ sampled $\propto \exp\left(\frac{\epsilon\xi(\theta|D)}{2\Delta}\right)\xi(\theta)$
```

Formally, Algorithm 3, under the assumptions of Theorem 4, outputs response θ with probability proportional to $\exp(\epsilon u(D, \theta)/2\Delta)$ times a base measure $\mu(\theta)$. Here Δ is a Lipschitz coefficient for u with sup-norm on responses and pseudo-metric ρ on datasets as in the previous section. Providing the base measure is non-trivial in general, but for discrete finite outcome spaces can be uniform (McSherry and Talwar 2007). For our mechanism to be broadly applicable, we can safely take $\mu(\theta)$ as $\xi(\theta)$.[4]

Corollary 4. *Algorithm 3 preserves ϵ-differential privacy wrt pseudo-metric ρ up to distance $r > 0$.*

[3] Meaning that $\rho(x, y) = 0$ does not necessarily imply $x = y$.

[4] In particular the base measure guarantees we have a proper density function: if $u(D, \theta)$ is bounded by M, then we have normalising constant $\int_\theta \exp(\epsilon u(D, \theta))\mu(\theta)d\theta \leq \exp(M\epsilon) < \infty$.

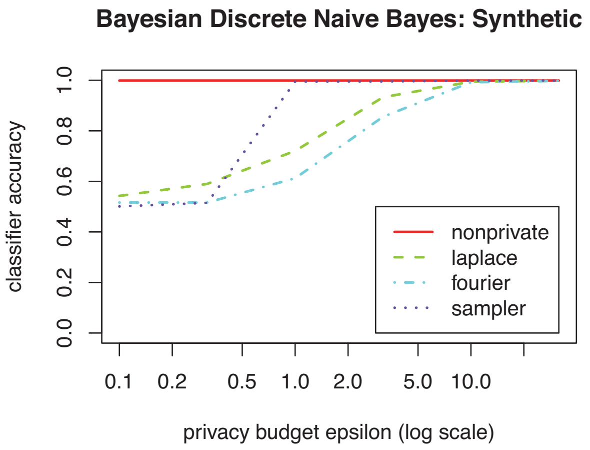

Figure 1: Effect on Bayesian naïve Bayes predictive-posterior accuracy of varying the privacy level.

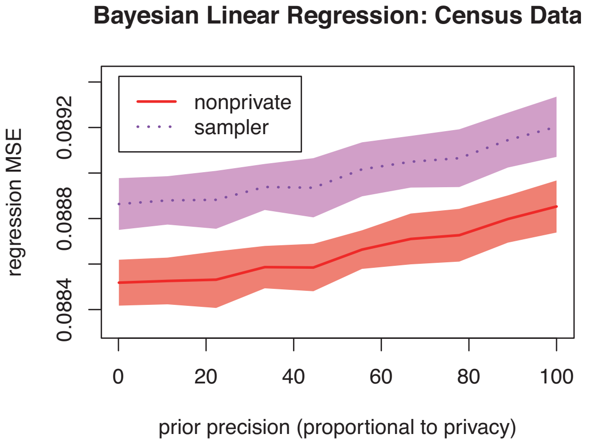

Figure 2: Effect on linear regression of varying prior concentration. Bands indicate standard error over repeats.

Proof. The sensitivity of the posterior score function corresponds to the computed Δ (Dimitrakakis et al. 2015, Theorem 6) under either Lipschitz assumptions. The result then follows from (McSherry and Talwar 2007, Theorem 6). □

Utility for Algorithm 3 follows from (McSherry and Talwar 2007), and states that the posterior likelihood of responses is likely to be close to that of the MAP.

Lemma 4. *Let* $\theta^\star = \max_\theta \xi\left(\theta|D\right)$ *with maximizer the MAP estimate, and let* $S_t = \{\theta \in \Theta : \xi\left(\theta|D\right) > \theta^\star - t\}$ *for* $t > 0$. *Then* $\mathbb{P}(S^c_{2t}) \le \exp(-\epsilon t)/\xi\left(S_t\right)$.

Experiments

Having proposed a number of mechanisms for approximating exact Bayesian inference in the general framework of probabilistic graphical models, we now demonstrate our approaches on two simple, well-known PGMs: the (generative) naïve Bayes classifier, and (discriminative) linear regression. This section, with derivations in the full report (Zhang, Rubinstein, and Dimitrakakis 2015), illustrates how our approaches are applied, and supports our extensive theoretical results with experimental observation. We focus on the trade-off between privacy and utility (accuracy and MSE respectively), which involves the (private) posterior via a predictive posterior distribution in both case studies.

Bayesian Discrete Naïve Bayes

An illustrative example for our mechanisms is a Bayesian naïve Bayes model on Bernoulli class Y and attribute variables X_i, with full conjugate Beta priors. This PGM directly specialises the running Example 1. We synthesised data generated from a naïve Bayes model, with 16 features and 1000 examples. Of these we trained our mechanisms on only 50 examples, with uniform Beta priors. We formed predictive posteriors for $Y|\boldsymbol{X}$ from which we thresholded at 0.5 to make classification predictions on the remaining, unseen test data so as to evaluate classification accuracy. The results are reported in Figure 1, where average performance is taken over 100 repeats to account for randomness in train/test split, and randomised mechanisms.

The small size of this data represents a challenge in our setting, since privacy is more difficult to preserve under smaller samples (Dwork et al. 2006). As expected, privacy incurs a sacrifice to accuracy for all private mechanisms.

For both Laplace mechanisms that perturb posterior updates, note that the d Boolean attributes and class label (being sole parent to each) yields nodes $|\mathcal{I}| = d + 1$ and downward closure size $|\mathcal{N}_\mathcal{I}| = 2d + 2$. Following our generic mechanisms, the noise added to sufficient statistics is independent on training set size, and is similar in scale. t was set for the Fourier approach, so that stealth was achieved 90% of the time—those times that contributed to the plot. Due to the small increments to cell counts for Fourier, necessary to achieve its *additional stealth property*, we expect a *small decrease to utility which is borne out in Figure 1.*

For the posterior sampler mechanism, while we can apply Assumption 2 to a Bernoulli-Beta pair to obtain a generalised form of (ϵ, δ)-differential privacy, we wish to compare with our ϵ-differentially-private mechanisms and so choose a route which satisfies Assumption 1 as detailed in the full report (Zhang, Rubinstein, and Dimitrakakis 2015). We trim the posterior before sampling, so that probabilities are lower-bounded. Figure 1 demonstrates that for small ϵ, the minimal probability at which to trim is relatively large resulting in a poor approximate posterior. But past a certain threshold, *the posterior sampler eventually outperforms the other private mechanisms.*

Bayesian Linear Regression

We next explore a system of continuous r.v.'s in Bayesian linear regression, for which our posterior sampler is most appropriate. We model label Y as i.i.d. Gaussian with known-variance and mean a linear function of features, and the linear weights endowed with multivariate Gaussian prior with zero mean and spherical covariance. To satisfy Assumption 1 we conservatively truncate the Gaussian prior (*cf.* the full report Zhang, Rubinstein, and Dimitrakakis 2015), and sample from the resulting truncated posterior; form a predictive posterior; then compute mean squared error. To evaluate our approach we used the U.S. census records dataset from the *Integrated Public Use Microdata Series* (Minnesota Population Center 2009) with 370k records and 14 demographic features. To predict *Annual Income*, we train on 10% data with the remainder for testing. Figure 2 displays MSE under varying prior precision b (inverse of covariance) and weights with bounded norm $10/\sqrt{b}$ (chosen conservatively). As expected, more concentrated prior (larger b) leads to worse MSE for both mechanisms, as stronger priors reduce data influence. Compared with linear regression, private regression suffers only slightly worse MSE. At the same time the posterior sampler enjoys increasing privacy (that is proportional to the bounded norm as given in the full report).

Conclusions

We have presented a suite of mechanisms for differentially-private inference in graphical models, in a Bayesian framework. The first two perturb posterior parameters to achieve privacy. This can be achieved either by performing perturbations in the original parameter domain, or in the frequency domain via a Fourier transform. Our third mechanism relies on the choice of a prior, in combination with posterior sampling. We complement our mechanisms for releasing the posterior, with private MAP point estimators. Throughout we have proved utility and privacy bounds for our mechanisms, which in most cases depend on the *graph structure of the Bayesian network: naturally, conditional independence affects privacy.* We support our new mechanisms and analysis with applications to two concrete models, with experiments exploring the privacy-utility trade-off.

Acknowledgements. This work was partially supported by the Swiss National Foundation grant "Swiss Sense Synergy" CRSII2-154458, and by the Australian Research Council (DE160100584).

References

Barak, B.; Chaudhuri, K.; Dwork, C.; Kale, S.; McSherry, F.; and Talwar, K. 2007. Privacy, accuracy, and consistency too: A holistic solution to contingency table release. In *PODS '07*, 273–282.

Chaudhuri, K., and Hsu, D. 2012. Convergence rates for differentially private statistical estimation. In *ICML'12*.

Chaudhuri, K., and Monteleoni, C. 2008. Privacy-preserving logistic regression. In *NIPS'08*, 289–296.

Chaudhuri, K.; Monteleoni, C.; and Sarwate, A. D. 2011. Differentially private empirical risk minimization. *Journal of Machine Learning Research* 12(Mar):1069–1109.

Chaudhuri, K.; Sarwate, A.; and Sinha, K. 2012. Near-optimal differentially private principal components. In *NIPS'12*, 989–997.

Dimitrakakis, C.; Nelson, B.; Mitrokotsa, A.; and Rubinstein, B. 2014. Robust and private Bayesian inference. In *ALT'14*.

Dimitrakakis, C.; Nelson, B.; Zhang, Z.; Mitrokotsa, A.; and Rubinstein, B. 2015. Differential privacy in a Bayesian setting through posterior sampling. Technical Report 1306.1066, arXiv.

Duchi, J. C.; Jordan, M. I.; and Wainwright, M. J. 2013. Local privacy and statistical minimax rates. Technical Report 1302.3203, arXiv.

Dwork, C., and Smith, A. 2009. Differential privacy for statistics: What we know and what we want to learn. *Journal of Privacy and Confidentiality* 1(2):135–154.

Dwork, C.; McSherry, F.; Nissim, K.; and Smith, A. 2006. Calibrating noise to sensitivity in private data analysis. In *TCC'06*, 265–284.

Jagannathan, G.; Pillaipakkamnatt, K.; and Wright, R. 2009. A practical differentially private random decision tree classifier. In *ICDM'09*, 114–121.

Kerckhoffs, A. 1883. La cryptographie militaire. *Journal des sciences militaires* IX:5–83 January; 161–191, February.

McSherry, F., and Talwar, K. 2007. Mechanism design via differential privacy. In *FOCS'07*, 94–103.

Minnesota Population Center. 2009. Integrated public use microdata series - international: Version 5.0. 2009. https://international.ipums.org accessed 2015-08-30.

Mir, D. 2012. Differentially-private learning and information theory. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, 206–210. ACM.

Pagnin, E.; Dimitrakakis, C.; Abidin, A.; and Mitrokotsa, A. 2014. On the leakage of information in biometric authentication. In *Indocrypt 2014*.

Rubinstein, B. I. P.; Bartlett, P. L.; Huang, L.; ; and Taft, N. 2012. Learning in a large function space: Privacy-preserving mechanisms for SVM learning. *Journal of Privacy and Confidentiality* 4(1).

Wang, Y.-X.; Fienberg, S.; and Smola, A. 2015. Privacy for free: Posterior sampling and stochastic gradient monte carlo. In Blei, D., and Bach, F., eds., *ICML'15*, 2493–2502.

Williams, O., and McSherry, F. 2010. Probabilistic inference and differential privacy. In *NIPS'10*, 2451–2459.

Xiao, Y., and Xiong, L. 2012. Bayesian inference under differential privacy. arXiv preprint arXiv:1203.0617.

Zhang, J.; Zhang, Z.; Xiao, X.; Yang, Y.; and Winslett, M. 2012. Functional mechanism: regression analysis under differential privacy. *Proc. VLDB Endowment* 5(11):1364–1375.

Zhang, J.; Cormode, G.; Procopiuc, C. M.; Srivastava, D.; and Xiao, X. 2014. Privbayes: Private data release via bayesian networks. In *SIGMOD'14*, 1423–1434.

Zhang, Z.; Rubinstein, B.; and Dimitrakakis, C. 2015. On the differential privacy of Bayesian inference. Technical Report https://hal.inria.fr/hal-01234215, HAL.

Zheng, S. 2015. The differential privacy of Bayesian inference. Bachelor's thesis, Harvard College.