Differential Privacy and Private Bayesian Inference*

Christos Dimitrakakis[1], Blaine Nelson[2]**,
Aikaterini Mitrokotsa[1], and Benjamin I. P. Rubinstein[3]

[1] Chalmers University of Technology, Sweden
[2] University of Potsdam, Germany
[3] The University of Melbourne, Australia

We consider a Bayesian statistician ($\mathcal{B}$) communicating with an untrusted third party ($\mathcal{A}$). $\mathcal{B}$ wants to convey useful answers to the queries of $\mathcal{A}$, but without revealing private information. For example, we may want to give statistics about how many people suffer from a disease, but without revealing whether a particular person has it. This requires us to strike a good balance between utility and privacy. In this extended abstract, we summarise our results on the inherent privacy and robustness properties of Bayesian inference [1]. We formalise and answer the question of whether $\mathcal{B}$ can select a prior distribution so that a computationally unbounded $\mathcal{A}$ cannot obtain private information from queries. Our setting is as follows:

(i) $\mathcal{B}$ selects a model family ($\mathcal{F}_\Theta$) and a prior (ξ).
(ii) $\mathcal{A}$ is allowed to see $\mathcal{F}_\Theta$ and ξ and is computationally unbounded.
(iii) $\mathcal{B}$ observes data x and calculates the posterior $\xi(\theta|x)$ but does not reveal it. Instead, $\mathcal{B}$ responds to queries at times $t = 1, \ldots$ as follows.
(iv) $\mathcal{A}$ sends a query q_t to $\mathcal{B}$.
(v) $\mathcal{B}$ responds $q_t(\theta_t)$ where θ_t is drawn from the posterior: $\theta_t \sim \xi(\theta|x)$.

We show that by choosing $\mathcal{F}_\Theta$ or ξ appropriately, the resulting posterior-sampling mechanism satisfies generalised differential privacy and indistinguishability properties. The intuition is that robustness and privacy are linked via smoothness. Learning algorithms that are smooth mappings—their output (*eg.* a spam filter) varies little with perturbations to input (*e.g.* similar training corpora)—are robust: outliers have reduced influence, and adversaries cannot easily discover private information. Consequently, robustness and privacy may be simultaneously achieved and perhaps are deeply linked.

Our results [1] show that mild assumptions are sufficient to obtain a differentially-private mechanism in the Bayesian setting. As a first step, we generalise the definition of differential privacy [2] to arbitrary dataset spaces $\mathcal{S}$. To do so, we introduce the notion of differential privacy under a pseudo-metric ρ on the space of all datasets.

* This work was partially supported by the Marie Curie Project ESDeMUU grant No: 237816 and the FP7 STREP project BEAT, grant No: 284989
** Blaine Nelson is now at Google, Mountain View

Definition 1 ((ϵ, δ)-differential privacy under ρ.). *A conditional distribution $P(\cdot \mid x)$ on $(\Theta, \mathfrak{S}_\Theta)$ is* (ϵ, δ)-differentially private under *a pseudo-metric $\rho : \mathcal{S} \times \mathcal{S} \to \mathbb{R}_+$ if, for all $B \in \mathfrak{S}_\Theta$ and for any $x \in \mathcal{S}$, then $P(B \mid x) \leqslant e^{\epsilon\rho(x,y)} P(B \mid y) + \delta\rho(x, y)\ \forall y$.*

Our first assumption is that the $\mathcal{F}_\Theta$ is smooth with respect to some metric d:

Assumption 1 (Lipschitz continuity) *Let $d(a, b) \triangleq |\ln a/b|$. There exists $L > 0$ such that, for any $\theta \in \Theta$: $d(p_\theta(x), p_\theta(y)) \leqslant L\rho(x, y), \qquad \forall x, y \in \mathcal{S}$.*

As it can be hard for this assumption to hold uniformly over Θ, we relax it by only requiring that $\mathcal{B}$'s *prior* probability ξ is concentrated in the smoothest members of the family:

Assumption 2 (Stochastic Lipschitz continuity) *Let Θ_L be the set of L-Lipschitz parameters. Then $\exists c > 0$ such that, $\forall L \geqslant 0$: $\xi(\Theta_L) \geqslant 1 - \exp(-cL)$.*

One consequence of either of those assumption is that the posterior is robust, in the sense that small dataset changes result in small changes in the posterior:

Theorem 1. *If ξ is a prior on Θ and $\xi(\cdot \mid x)$ and $\xi(\cdot \mid y)$ are the respective posterior distributions for datasets $x, y \in \mathcal{S}$, then the posterior KL-divergence satisfies: $D\left(\xi(\cdot \mid x) \parallel \xi(\cdot \mid y)\right) \leqslant O(\rho(x, y))$, with linear terms depending on L, c.*

Consequently, one way to answer queries would be to use samples from the poster distribution. In fact, we show that such posterior-sampling mechanisms are differentially private:

Theorem 2. *Under Assumption 1, the posterior is $(2L, 0)$-differentially private under ρ. Under Assumption 2, the posterior ξ is $\left(0, \sqrt{\frac{\kappa}{2c}}\right)$-differentially private under $\sqrt{\rho}$.*

As the adversary performs more queries, he obtains more information about the true dataset. Finally, we bound the effort required by an adversary to be ϵ-close to the true dataset:

Theorem 3. *The adversary can distinguish between data x, y with probability $1 - \delta$ if $\rho(x, y) \geqslant O(\frac{\ln 1/\delta}{n})$, with a linear dependency on L or c.*

We have shown that both the privacy and robustness properties of Bayesian inference are inherently linked through the choice of prior distribution. Such prior distributions exist for example in well known conjugate families. There is also a natural *posterior sampling* mechanism through which differential privacy and dataset indistinguishability can be achieved.

References

[1] Dimitrakakis, C., Nelson, B., Mitrokotsa, A., Rubinstein, B.: Robust and private bayesian inference. In: Proc. of ALT. (2014)

[2] Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Proc. of TCC. (2006) 265–284