Notes on the Definition of Differential Privacy

Salil Vadhan

February 9, 2013

Extracted from scribe notes taken by Alan Deckelbaum, Emily Shen, and Thomas Steinke during MIT 6.889 "New Developments in Cryptography," Spring 2011.

1 Defining Differential Privacy

We have the following two desiderata in our definition of differential privacy:

1. Strong notion of privacy
2. Utility- "noisy" answers are still useful.

We now have the following definition.

Definition 1. *$M : \mathcal{X}^n \times \mathcal{Q} \to \mathcal{Y}$ is **differentially private** iff for all $q \in \mathcal{Q}$ and for all $x, x' \in \mathcal{X}$ differing only on a single row, the distributions $M(x, q)$ and $M(x', q)$ are "similar."*

In the above definition, $\mathcal{X}$ is the space from which the rows come, $\mathcal{Q}$ is the space of questions, and $\mathcal{Y}$ is the output space. The definition captures the idea that no individual's data has a significant influence on the output. Note that the "similarity" of distributions depends only on the coin tosses of the mechanism, and we need not worry about what supplemental information is available to the user.

We say that a mechanism is **ϵ-differentially private** if the distributions $M(x, q)$ and $M(x', q)$ have "distance" at most ϵ. We now must choose what to use for the notion of "distance" of two distributions.

1.1 Statistical Difference

A first attempt for our definition is to use statistical difference, where the statistical difference between distributions A and B is defined to be

$$\max_{T \subset Y} |Pr[A \in T] - Pr[B \in T]|.$$

Unfortunately, this is a bad choice for differential privacy. Consider the following two cases:

- Case 1: $\epsilon \leq \frac{1}{10n}$. In this case, we can use a hybrid argument by changing one entry at a time to show that, with probability at least 90%, we gain no useful information from our query. In particular, the statistical difference between $M(x, q)$ and $M(0^n, q)$ is less than 0.1.
- Case 2: $\epsilon \geq \frac{1}{10n}$. Consider a mechanism which releases a random row from the database with probability $\frac{1}{10}$. This mechanism satisfies the statistical difference constraint, but intuitively shouldn't be allowed. (Alternatively, we could look at a mechanism which releases the first row with probability $\frac{1}{10n}$.)

1.2 Actual Choice of Distance Between Distributions

The condition we choose to use for distance between probability distributions A and B being at most ϵ is, for all $T \subset \mathcal{Y}$,

$$Pr[A \in T] \leq e^{\epsilon} Pr[B \in T].$$

We will think of ϵ being in the range $\frac{1}{n} \leq \epsilon \leq 1$. (We still cannot think of ϵ being arbitrarily small, due to critiques similar to case 1 above.) We note that $e^{\epsilon} \approx 1 + \epsilon$, and that the above constraint implies that the statistical difference between distributions is $O(\epsilon)$. This is in fact a stronger condition than merely bounding statistical difference. In particular, the bad mechanisms from case 2 above are eliminated. This definition is more useful for rare events (events which have probability less than ϵ) than the statistical difference definition.

2 Max-Divergence

We define the max-divergence of A and B by

$$D_{\infty}(A \| B) = max_{T \subset \mathcal{Y}} \ln \left(\frac{Pr[A \in T]}{Pr[B \in T]} \right) = max_{y \in \mathcal{Y}} \ln \left(\frac{Pr[A = y]}{Pr[B = y]} \right).$$

A mechanism M is ϵ-differentially private iff for all $q \in \mathcal{Q}$, for all $x, x' \in \mathcal{X}^n$ differing on 1 row,

$$D_{\infty}(M(x, q) \| M(x', q)) \leq \epsilon.$$

Compare this with the KL Divergence

$$D(A \| B) = \mathbb{E}_{y \in A} \left[\ln \left(\frac{Pr[A = y]}{Pr[B = y]} \right) \right].$$

Max-divergence is a worst-case analog of KL divergence, similar to the way that min-entropy is a worst-case analog of Shannon entropy.

3 Simulation-Based Definition

As in cryptography, we can consider a simulation-based definition.

Definition 2. *A polynomial-time mechanism M is* ϵ-simulation-differentially private *if there exists a polynomial-time simulator S s.t. $\forall q \in \mathcal{Q}, x \in \mathcal{X}^n, i \in [n]$,*

$$D_{\infty}(M(q, x) \| S(q, x_{-i})) \leq \epsilon$$

and

$$D_{\infty}(S(q, x_{-i}) \| M(q, x)) \leq \epsilon,$$

where x_{-i} denotes x without the ith row.

This says that the amount one learns from M(x) about any row in x is within ϵ of what one can learn about it from the rest of the database. Note that whatever can be learned about an individual from the other rows is not protected.)

Claim 1. *1. If M is ϵ-differentially private, then M is ϵ-simulation-differentially private.*

2. If M is ϵ-simulation-differentially private, then M is 2ϵ-differentially private.

4 A Bayesian Definition

In this section, we will see a Bayesian definition of privacy that is robust to the choice of metric on distributions that is used, yet is equivalent to the standard formulation of differential privacy in terms of max-divergence. This provides a justification for why the multiplicative measure is the "right" one to use in the definition of differential privacy.

Roughly, the next proposition states that the output of a differentially private mechanism does not significantly change an adversary's beliefs about an individual. Note that the output of a (useful) mechanism will give the adversary information about the global distribution from which the database was generated. Thus we fix all but one row of the database, and only consider the adversary's beliefs about the remaining individual (under the implicit assumption that the adversary knows the rest of the database).

Proposition 1. *If M is an ε-differentially private, then, for every $x \in \mathcal{X}^n$, $i \in [n]$, distribution X_i on $\mathcal{X}$ (the adversary's prior on x_i), and output y of $M(x)$,*

$$D_\infty(X_i||(X_i|M(X_i, x_{-i}) = y)) \leq \varepsilon \quad \wedge \quad D_\infty((X_i|M(X_i, x_{-i}) = y)||X_i) \leq \varepsilon,$$

where $X_i|M(X_i, x_{-i}) = y$ denotes the distribution of X_i conditioned on the output of M on X_i and the rest of the database x_{-i} being y.

Proof: By Bayes' theorem

$$\begin{aligned} \Pr[X_i = x_i'|M(X_i, x_{-i}) = y] &= \frac{\Pr[M(x_i', x_{-i}) = y]}{\Pr[M(X_i, x_{-i}) = y]} \Pr[X_i = x_i'] \\ &\in e^{\pm\varepsilon} \Pr[X_i = x_i'] \quad \text{(by } \varepsilon\text{-differential privacy).} \end{aligned} \tag{1}$$

So the prior and posterior only differ by a multiplicative factor in the range $[e^{-\varepsilon}, e^{+\varepsilon}]$, which gives the result. □

Proposition 2 is a converse to Proposition 1. Moreover, we see a different metric being used—statistical distance.

Proposition 2. *Let $M : \mathcal{X}^n \to \mathcal{Y}$ be a randomized mechanism. Suppose that, for every $x \in \mathcal{X}^n$, $i \in [n]$, distributions X_i on $\mathcal{X}$, and $y \in \mathcal{Y}$,*

$$\Delta(X_i, (X_i|M(X_i, x_{-i}) = y)) \leq \varepsilon.$$

Then M is $O(\varepsilon)$-differentially private.

Proof: Let $x, x' \in \mathcal{X}^n$ differ on one row $i \in [n]$; let $x = (x_i, x_{-i})$ and $x' = (x_i', x_{-i})$. Define a distribution X_i by

$$\Pr[X_i = x_i] = \Pr[X_i = x_i'] = 1/2.$$

Choose γ such that

$$\Pr[X_i = x_i|M(X_i, x_{-i}) = y] = (1 + \gamma)/2 \quad \wedge \quad \Pr[X_i = x_i'|M(X_i, x_{-i}) = y] = (1 - \gamma)/2.$$

Then X_i and $X_i|M(X_i, x_{-i}) = y$ are Bernoulli random variables, whence

$$\gamma/2 = \Delta(X_i, (X_i|M(X_i, x_{-i}) = y)) \leq \varepsilon.$$

By Bayes' theorem,

$$\begin{aligned} \frac{\Pr[M(x') = y]}{\Pr[M(x) = y]} &= \frac{\Pr[M(x_i', x_{-i}) = y]}{\Pr[M(X_i, x_{-i}) = y]} \Big/ \frac{\Pr[M(x_i, x_{-i}) = y]}{\Pr[M(X_i, x_{-i}) = y]} \\ &= \frac{\Pr[X_i = x_i'|M(X_i, x_{-i}) = y]}{\Pr[X_i = x_i']} \Big/ \frac{\Pr[X_i = x_i|M(X_i, x_{-i}) = y]}{\Pr[X_i = x_i]} \quad \text{(cf. (1))} \\ &= \frac{1 + \gamma}{1 - \gamma} = e^{O(\varepsilon)}. \end{aligned}$$

□

Remarks:

- Note that the only fact about statistical distance used in the proof of Proposition 2 is that if

$$\Delta\left(\text{Bernoulli}\left(\frac{1}{2}\right), \text{Bernoulli}\left(\frac{1+\gamma}{2}\right)\right)$$

  is small (say at most ϵ), then γ is small (specifically, at most 2ϵ in the proof of above). Replacing statistical distance with any other metric that satisfies this property will imply the standard definition of differential privacy (with the max-divergence measure of distance).

- Propositions 1 and 2 can be generalized to (ε, δ)-differential privacy. For negligible δ, the statements hold with $1 -$ negligible probability over the randomness of M.

- We allow an arbitrary prior. So differential privacy is resilient to arbitrary side information. So an individual's data is safe, given that it is localized to one row in the database. There is no guarantee if the data is spread over the whole database.