Verifying Snapping Mechanism

November 5, 2019

1 Formalization

Definition 1 ($\mathsf{Snap}(\mu, a) : \mathsf{Distr}(U) \to A \to \mathsf{Distr}(B)$)
The ideal Snapping mechanism $\mathsf{Snap}(\mu, a)$ is defined as:

$$u \xleftarrow{\$} \mu; y = \frac{\ln(|u|)}{\epsilon}; s\mathsf{sign}(u); z = s * y; x = f(a); w = x + z; w' = \lfloor w \rceil_\Lambda; r = \mathsf{clamp}_B(w')$$

where f is the query function over input $a \in A$, ϵ is the privacy budget and S sampled from $\{-1, +1\}$ with $\mathsf{Bernoulli}(0.5)$.

Definition 2
Let $\epsilon \leq 0$. The ϵ-DP divergence $\Delta_\epsilon(\mu_1, \mu_2)$ between two sub-distributions $\mu_1 \in \mathsf{Distr}(U)$, $\mu_2 \in \mathsf{Distr}(U)$is defined as:

$$\underset{E \in U}{sup}\Big(\Pr_{x \leftarrow \mu_1}[x \in E] - \exp(\epsilon) \Pr_{x \leftarrow \mu_2}[x \in \cdot E]]\Big)$$

Definition 3 **(ϵ - dilation)**
.

Let $\epsilon \geq 0$. The ϵ-dilation $D_\epsilon(\mu_1, \mu_2)$ between two sub-distributions $\mu_1 \in \mathsf{Distr}(U)$, $\mu_2 \in \mathsf{Distr}(U)$is defined as:

$$\underset{E \in U}{sup}\Big(\Pr_{x \leftarrow \mu_1}[x \in E] - \exp(\epsilon) \Pr_{x \leftarrow \mu_2}[x \in \exp(-\epsilon) \cdot E]]\Big)$$

Proposition 1 **((ϵ, δ)-differential privacy)**
For every pair of sub-distributions $\mu_1 \in \mathsf{Distr}(U)$, $\mu_2 \in \mathsf{Distr}(U)$, s.t.

$$D_\epsilon(\mu_1, \mu_2) \leq \delta,$$

The snapping mechanism $\mathsf{Snap}(\mu, a) : \mathsf{Distr}(U) \to A \to \mathsf{Distr}(B)$ is (ϵ, δ) - differentially private w.r.t. an adjacency relation Φ for every two adjacent inputs a, a' and μ_1, μ_2

Proof. Followed directly by unfolding the Snap mechanism.

$$\begin{aligned}
\Pr_{x \leftarrow \mathsf{Snap}(\mu_1, a)}[x = e] &= \Pr_{u \leftarrow \mu_1}[\lfloor f(a) + \tfrac{S \cdot \log(u)}{\epsilon} \rceil_\Lambda = e] \\
&= \Pr_{u \leftarrow \mu_1}[u \in [\tfrac{\exp((e - \frac{\Lambda}{2} - f(a))\epsilon)}{S}, \tfrac{\exp((e + \frac{\Lambda}{2} - f(a))\epsilon)}{S})] \\
&\leq \exp(\epsilon) \Pr_{u \leftarrow \mu_2}[u \in \exp(-\epsilon)[\tfrac{\exp((e - \frac{\Lambda}{2} - f(a))\epsilon)}{S}, \tfrac{\exp((e + \frac{\Lambda}{2} - f(a))\epsilon)}{S})] \\
&= \exp(\epsilon) \Pr_{u \leftarrow \mu_2}[\lfloor f(a') + \tfrac{S \cdot \log(u)}{\epsilon} \rceil_\Lambda = e] \\
&= \exp(\epsilon) \Pr_{x \leftarrow \mathsf{Snap}(\mu_2, a')}[x = e]
\end{aligned}$$

□

$$\frac{}{u_1 \xleftarrow{\$} \mu \sim_{\epsilon,0} u_2 \xleftarrow{\$} \mu : T \Rightarrow e^{-\epsilon} u_2 \leq u_1 \leq e^{\epsilon} u_2} \text{ AxUNIF}$$

$$\frac{}{y_1 = \dfrac{\ln(|u_1|)}{\epsilon} \sim_{0,0} y_2 = \dfrac{\ln(|u_2|)}{\epsilon} : e^{-\epsilon} u_2 \leq u_1 \leq e^{\epsilon} u_2 \Rightarrow y_2 - 1 \leq y_1 \leq 1 + y_2}$$

$$\frac{}{s_1 = \mathsf{sign}(u_1) \sim_{0,0} s_2 = \mathsf{sign}(u_2) : e^{-\epsilon} u_2 \leq u_1 \leq e^{\epsilon} u_2 \Rightarrow s_1 = s_2}$$

$$\frac{}{z_1 = s_1 * y_1 \sim_{0,0} z_2 = s_2 * y_2 : s_1 = s_2 \wedge y_2 - 1 \leq y_1 \leq 1 + y_2 \Rightarrow |z_1 - z_2| \leq 1}$$

$$\frac{}{x_1 = f(a_1) \sim_{0,0} x_2 = f(a_2) : a_1 = a_2 + 1 \Rightarrow x_1 = x_2 + 1}$$

$$\frac{}{w_1 = x_1 + z_1 \sim_{0,0} w_2 = x_2 + z_2 : x_1 = x_2 + 1 \wedge |z_1 - z_2| \leq 1 \wedge -2 \leq k \leq 0 \Rightarrow w_1 + k = w_2}$$

$$\frac{}{w_1' = \lfloor w_1 \rfloor_\Lambda \sim_{0,0} w_2' = \lfloor w_2 \rfloor_\Lambda : w_1 + k = w_2 \wedge -2 \leq k \leq 0 \Rightarrow w_1' + k = w_2'}$$

$$\frac{}{r_1 = \mathsf{clamp}_B(w_1') \sim_{0,0} r_2 = \mathsf{clamp}_B(w_2') : w_1' + k = w_2' \wedge -2 \leq k \leq 0 \Rightarrow r_1 + k = r_2}$$

Figure 1: Coupling Derivation of two Snap mechanisms: $\mathsf{Snap}(\mu_1, a_1)$, $\mathsf{Snap}(\mu_2, a_2)$

Definition 4 **((ϵ, δ) - lifting [1])**
Two sub-distributions $\mu_1 \in \mathsf{Distr}(U_1)$, $\mu_2 \in \mathsf{Distr}(U_2)$are related by the (ϵ, δ) - dilation lifting of $\Psi \subseteq U_1 \times U_2$, written $\mu_1 \Psi^{\#(\epsilon,\delta)} \mu_2$, if there exist two witness sub-distributions $\mu_L \in \mathsf{Distr}(U_1 \times U_2)$ and $\mu_R \in \mathsf{Distr}(U_1, U_2)$ s.t.:

1. $\pi_1(\mu_L) = \mu_1$ and $\pi_2(\mu_R) = \mu_2$;
2. $\mathsf{supp}(\mu_L) \subseteq \Psi$ and $\mathsf{supp}(\mu_R) \subseteq \Psi$; and
3. $\Delta_\epsilon(\mu_L, \mu_R) \leq \delta$.

Theorem 2
Let $\mu_1 \in \mathsf{Distr}(\mathbb{R})$, $\mu_2 \in \mathsf{Distr}(\mathbb{R})$ are defined:

$$\mu_1(x) = \mathsf{unif}(x)$$

$$\mu_2(y) = \mathsf{unif}(y)$$

where unif is uniform distribution over $(-1,1)$ whoes pdf. is defined as:

$$\mathsf{pdf}_{\mathsf{unif}}(x) = \begin{cases} \frac{1}{2} & x \in (-1,1) \\ 0 & o.w. \end{cases}.$$

Then, $\mu_1 \Psi^{\#(\epsilon,0)} \mu_2$, where

$$\Psi = \{(x,y) \in \mathbb{R} \times \mathbb{R} | x \cdot e^{-\epsilon} \le y \le x \cdot e^{\epsilon}\}$$

Proof. Existing $\mu_L, \mu_R \in \mathsf{Distr}(\mathbb{R} \times \mathbb{R})$:

$$\mu_L(x,y) = \begin{cases} \mathsf{unif}(x) & x \cdot e^{-\epsilon} = y \land x \in (-1,1) \\ 0 & o.w. \end{cases} \mu_R(x,y) = \begin{cases} \mathsf{unif}(y) & x \cdot e^{-\epsilon} = y \land y \in (-1,1) \\ 0 & o.w. \end{cases}.$$

Their pdf. are defined:

$$\mathsf{pdf}_{\mu_L}(x,y) = \begin{cases} \mathsf{pdf}_{\mathsf{unif}}(x) & x \cdot e^{-\epsilon} = y \land x \in (-1,1) \\ 0 & o.w. \end{cases}$$

$$\mathsf{pdf}_{\mu_R}(x,y) = \begin{cases} \mathsf{pdf}_{\mathsf{unif}}(y) & x \cdot e^{-\epsilon} = y \land y \in (-1,1) \\ 0 & o.w. \end{cases}.$$

- $\mathsf{supp}(\mu_L) \in \Psi \land \mathsf{supp}(\mu_R) \in \Psi$
  - $\mathsf{supp}(\mu_L) \subseteq \Psi$
    By definition of the pdf of μ_L, we have: $\Pr_{(x,y) \xleftarrow{\$} \mu_L}[(x,y) \notin \Psi] = 0.$
    Then we can derive $\mathsf{supp}(\mu_L) \in \Psi$
  - $\mathsf{supp}(\mu_R) \subseteq \Psi$
    By definition of the pdf of μ_R, we have: $\Pr_{(x,y) \xleftarrow{\$} \mu_R}[(x,y) \notin \Psi] = 0.$
    Then we can derive $\mathsf{supp}(\mu_L) \in \Psi$
- $\pi_1(\mu_L) = \mu_1 \land \pi_2(\mu_R) = \mu_2$
  - $\pi_1(\mu_L) = \mu_1$
    By definition of the π_1 and pdf of μ_L, we have $\forall x \in \mathbb{R}$:

$$\mathsf{pdf}_{\pi_1(\mu_L)}(x) = \begin{cases} \int_y \mathsf{pdf}_{\mathsf{unif}}(x) & (x,y) \in \Psi \land x \in (-1,1) \\ 0 & o.w. \end{cases} = \begin{cases} \mathsf{pdf}_{\mathsf{unif}}(x) & x \in (-1,1) \\ 0 & o.w. \end{cases} = \mathsf{pdf}_{\mu_1}(x)$$

  - $\pi_1(\mu_R) = \mu_2$
    Equivalent to show$\mathsf{pdf}_{\pi_2(\mu_R)} = \mathsf{pdf}_{\mu_2}$.
    By definition of the π_2 and pdf of μ_R, we have $\forall y \in \mathbb{R}$:

$$\mathsf{pdf}_{\pi_2(\mu_R)}(y) = \begin{cases} \int_x \mathsf{pdf}_{\mathsf{unif}}(y) & (x,y) \in \Psi \land y \in (-1,1) \\ 0 & o.w. \end{cases} = \begin{cases} \mathsf{pdf}_{\mathsf{unif}}(y) & y \in (-1,1) \\ 0 & o.w. \end{cases} = \mathsf{pdf}_{\mu_2}(y)$$

- $\Delta_\epsilon(\mu_L, \mu_R) \leq 0$

  By definition of $\epsilon-$DP divergence, we have:

$$\begin{aligned} \Delta_\epsilon(\mu_L, \mu_R) &= \mathsf{Sup}_S\Big(\Pr_{(x,y) \xleftarrow{\$} \mu_L} [(x, y) \in S] - e^\epsilon \Pr_{(x,y) \xleftarrow{\$} \mu_R} [(x, y) \in S] \Big) \\ &= \mathsf{Sup}_S\Big(\int_{(x,y) \in S} \mathsf{pdf}_{\mu_L}(x, y) - e^\epsilon \int_{(x,y) \in S} \mathsf{pdf}_{\mu_R}(x, y) \Big) \end{aligned}$$

case $S \subseteq \{(x, y) | x \in (-1, 1) \wedge x \cdot e^{-\epsilon} = y\}$:

$$\begin{aligned} \Delta_\epsilon(\mu_L, \mu_R) &= \int_{(x,y) \in S} \mathsf{pdf}_{\mathsf{unif}}(x) - e^\epsilon \int_{(x,y) \in S} \mathsf{pdf}_{\mathsf{unif}}(y) \\ &= \int_{(x,y) \in S} \mathsf{pdf}_{\mathsf{unif}}(x) - e^\epsilon \int_{(x,y) \in S} \mathsf{pdf}_{\mathsf{unif}}(x * e^{-\epsilon}) \\ &= \int_{(x,y) \in S} \mathsf{pdf}_{\mathsf{unif}}(x) - e^\epsilon * e^{-\epsilon} \int_{(x,y) \in S} \mathsf{pdf}_{\mathsf{unif}}(x) \\ &= 0 \end{aligned}$$

case $S \subseteq \{(x, y) | x \in [1, e^\epsilon) \cup (-e^\epsilon, -1] \wedge x \cdot e^{-\epsilon} = y\}$:

$$\begin{aligned} \Delta_\epsilon(\mu_L, \mu_R) &= 0 - e^\epsilon \int_{(x,y) \in S} \mathsf{pdf}_{\mathsf{unif}}(y) \\ &< 0 \end{aligned}$$

case o.w.

$$\Delta_\epsilon(\mu_L, \mu_R) \quad = 0 - 0 = 0$$

□

References

[1] Gilles Barthe, Marco Gaboardi, Benjamin Grégoire, Justin Hsu, and Pierre-Yves Strub. Proving differential privacy via probabilistic couplings. In *LICS 2016*.