

Verifying Snapping Mechanism

November 5, 2019

1 Formalization

Definition 1 ($\text{Snap}(\mu, a) : \text{Distr}(U) \rightarrow A \rightarrow \text{Distr}(B)$)

The ideal Snapping mechanism $\text{Snap}(\mu, a)$ is defined as:

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

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

Definition 2

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

$$\sup_{E \in \mathcal{U}} \left(\Pr_{x \leftarrow \mu_1} [x \in E] - \exp(\epsilon) \Pr_{x \leftarrow \mu_2} [x \in \cdot E] \right)$$

Definition 3 (ϵ - dilation)

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

$$\sup_{E \in \mathcal{U}} \left(\Pr_{x \leftarrow \mu_1} [x \in E] - \exp(\epsilon) \Pr_{x \leftarrow \mu_2} [x \in \exp(-\epsilon) \cdot E] \right)$$

Proposition 1 ((ϵ, δ) -differential privacy)

For every pair of sub-distributions $\mu_1 \in \text{Distr}(U)$, $\mu_2 \in \text{Distr}(U)$, s.t.

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

The snapping mechanism $\text{Snap}(\mu, a) : \text{Distr}(U) \rightarrow A \rightarrow \text{Distr}(B)$ 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.

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

□

Definition 4 ((ϵ, δ) - lifting [1])

Two sub-distributions $\mu_1 \in \text{Distr}(U_1)$, $\mu_2 \in \text{Distr}(U_2)$ 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 $\mu_L \in \text{Distr}(U_1 \times U_2)$ and $\mu_R \in \text{Distr}(U_1, U_2)$ s.t.:

$$\begin{array}{c}
\frac{}{u_1 \stackrel{s}{\leftarrow} \mu \sim_{\epsilon,0} u_2 \stackrel{s}{\leftarrow} \mu : T \Rightarrow e^{-\epsilon} u_2 \leq u_1 \leq e^{\epsilon} u_2} \text{AxUNIF} \\
\\
\frac{}{y_1 = \frac{\ln(|u_1|)}{\epsilon} \sim_{0,0} y_2 = \frac{\ln(|u_2|)}{\epsilon} : e^{-\epsilon} u_2 \leq u_1 \leq e^{\epsilon} u_2 \Rightarrow y_2 - 1 \leq y_1 \leq 1 + y_2} \\
\\
\frac{}{s_1 = \text{sign}(u_1) \sim_{0,0} s_2 = \text{sign}(u_2) : e^{-\epsilon} u_2 \leq u_1 \leq e^{\epsilon} u_2 \Rightarrow s_1 = s_2} \\
\\
\frac{}{z_1 = s_1 * y_1 \sim_{0,0} z_2 = s_2 * y_2 : s_1 = s_2 \wedge y_2 - 1 \leq y_1 \leq 1 + y_2 \Rightarrow |z_1 - z_2| \leq 1} \\
\\
\frac{}{x_1 = f(a_1) \sim_{0,0} x_2 = f(a_2) : a_1 = a_2 + 1 \Rightarrow x_1 = x_2 + 1} \\
\\
\frac{}{w_1 = x_1 + z_1 \sim_{0,0} w_2 = x_2 + z_2 : x_1 = x_2 + 1 \wedge |z_1 - z_2| \leq 1 \wedge -2 \leq k \leq 0 \Rightarrow w_1 + k = w_2} \\
\\
\frac{}{w'_1 = \lfloor w_1 \rfloor_{\Lambda} \sim_{0,0} w'_2 = \lfloor w_2 \rfloor_{\Lambda} : w_1 + k = w_2 \wedge -2 \leq k \leq 0 \Rightarrow w'_1 + k = w'_2} \\
\\
\frac{}{r_1 = \text{clamp}_B(w'_1) \sim_{0,0} r_2 = \text{clamp}_B(w'_2) : w'_1 + k = w'_2 \wedge -2 \leq k \leq 0 \Rightarrow r_1 + k = r_2}
\end{array}$$

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

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

Theorem 2

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

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

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

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

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

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

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

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

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

Their pdf. are defined:

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

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

- $\text{supp}(\mu_L) \in \Psi \wedge \text{supp}(\mu_R) \in \Psi$

- $\text{supp}(\mu_L) \subseteq \Psi$

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 $\text{supp}(\mu_L) \in \Psi$

- $\text{supp}(\mu_R) \subseteq \Psi$

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 $\text{supp}(\mu_L) \in \Psi$

- $\pi_1(\mu_L) = \mu_1 \wedge \pi_2(\mu_R) = \mu_2$

- $\pi_1(\mu_L) = \mu_1$

By definition of the π_1 and pdf of μ_L , we have $\forall x \in \mathbb{R}$:

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

- $\pi_1(\mu_R) = \mu_2$

Equivalent to show $\text{pdf}_{\pi_2(\mu_R)} = \text{pdf}_{\mu_2}$.

By definition of the π_2 and pdf of μ_R , we have $\forall y \in \mathbb{R}$:

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

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

By definition of ϵ -DP divergence, we have:

$$\begin{aligned} \Delta_\epsilon(\mu_L, \mu_R) &= \sup_S \left(\Pr_{(x,y) \xleftarrow{\$} \mu_L} [(x, y) \in S] - e^\epsilon \Pr_{(x,y) \xleftarrow{\$} \mu_R} [(x, y) \in S] \right) \\ &= \sup_S \left(\int_{(x,y) \in S} \text{pdf}_{\mu_L}(x, y) - e^\epsilon \int_{(x,y) \in S} \text{pdf}_{\mu_R}(x, y) \right) \end{aligned}$$

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

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

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

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

case o.w.

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

□

References

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