Sécurité Appliquée
Protection de la vie privée-PVP. TD4

Jean-François COUCHOT
couchot[arobase]femto-st[point]fr

12 décembre 2021

1 Confidentialité différentielle locale

On donne la définition suivante :

DÉFINITION 1.1 (ϵ-CONFIDENTIALITÉ DIFFÉRENTIELLE LOCALE). Soit $\mathcal{X}$ un ensemble des valeurs possibles en entrée et $\epsilon \in \mathbb{R}^+$ un budget de fuite. $\mathcal{M}$ est un algorithme probabiliste non déterministe qui respecte la ϵ-confidentialité différentielle locale si

$$\begin{array}{ll} \forall x_1, x_2 \in \mathcal{X} & (x_1 \text{ et } x_2 \text{ deux données d'entrée}) \\ \forall r \text{ t.q. } r \in \mathcal{M}(\mathcal{X}), & (\text{pour tte image de l'algo.}) \\ \Pr[\mathcal{M}(x_1) = r] \leq e^{\epsilon} \Pr[\mathcal{M}(x_2) = r] & \end{array}$$

2 Application à l'algorithme de réponse randomisée

On considère la méthode de sondage [1] qui permet de répondre de manière "anonyme" à une question sur des données sensibles. On la particularise à la question suivante :

« Qui trouve ce cours ennuyeux [2] ? »

— Peu oseront dire OUI, même si c'est la vérité.

— Vous lancez une pièce 2 fois sans me montrer les résultat t_1 et t_2 :

  — Si c'est $t_1 = P$: si $t_2 = P$, vous annoncez NON, sinon OUI.

  — Si c'est $t_1 = F$: vous relancez la pièce et vous dites la vérité pour tout t_2.

Exercice 2.1 (Code implantant le nettoyage de la réponse). *Donner le code (3 lignes tout au plus) de la fonction* `reponse_randomisee(rep_o)` *qui prend en paramètre une réponse originale booléenne et qui retourne une réponse du même type selon l'algorithme donné ci-dessus.*

Exercice 2.2 (Arbre de porbabilité et estimation des réponses).

1. *Construire l'arbre de probabilité associé à ce mécanisme.*
2. *On suppose que N personnes ont participé à ce sondage et que f est le nombre de personnes qui auraient voulu répondre* OUI *à la question posée. Quel est le nombre o de personnes qui devraient finalement fournir la réponse* OUI *à ce sondage après application de l'algorithme ci-dessus ?*
3. *En déduire une estimation $\hat{f}$ du nombre de personnes qui auraient voulu répondre* OUI.

Exercice 2.3 (Code implantant l'estimateur). *Donner le code (3 lignes tout au plus) de la fonction* `estimation_nbre_oui_initiaux(reps)` *qui prend en paramètre la série de réponses nettoyées et retourne une estimation du nombre de* OUI *qui étaient présents initialement.*

1. Warner, S. L. (1965). Randomized response : A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 60(309), 63-69.
2. http ://www.lix.polytechnique.fr/čatuscia/teaching/MPRI/19-20/lecture3.pdf

Exercice 2.4 (Valeur de epsilon).

1. *Pour $x \in \{True, False\}$ et $r \in \{True, False\}$, évaluer les quatre probabilités* $\Pr[\mathcal{M}(x) = r]$.
2. *Montrer alors que*
$$\forall x_1, x_2 \in \mathcal{X}, \forall r \in \mathcal{M}(\mathcal{X}), \frac{\Pr[\mathcal{M}(x_1) = r]}{\Pr[\mathcal{M}(x_2) = r]} \leq 3.$$
3. *En déduire que* $\epsilon = \ln(3)$.
4. *Trouver une formule qui donne la variance de $\hat{f}$. En discuter.*