Sécurité Appliquée PVP. TD4
Mécanismes pour la confidentialités différentielle locale

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

3 janvier 2023

1 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.

Exercice 1.1 (Variance de l'estimateur).

1. *On rappelle que la variance est définie comme la moyenne des carrés des écarts à la moyenne. Montrer que pour une VAR* X, $\mathrm{Var}[X] = E[X^2] - E[X]^2$.
2. *On considère un utilisateur dont la réponse originale à la question est* OUI. *Soit* X *la VAR qui vaut 1 si sa réponse randomisée est* OUI *et 0 sinon. Évaluer* $E[X]$, $E[X^2]$. *En déduire que* $\mathrm{Var}[X] = \delta_1(1 - \delta_1)$ *avec* $\delta_1 = 3/4$.
3. *De manière similaire, on considère un utilisateur dont la réponse originale à la question est* NON. *Soit* Y *la VAR qui vaut 1 si sa réponse randomisée est* OUI *et 0 sinon. Montrer que* $\mathrm{Var}[Y] = \delta_0(1-\delta_0)$ *avec* $\delta_0 = 1/4$.
4. *Soit* R *la VAR qui compte le nombre de réponses randomisées égales à* OUI. *Montrer que* $\mathrm{Var}[R] = f.\mathrm{Var}[X] + (N - f).\mathrm{Var}[Y]$.
5. *Montrer alors que* $\mathrm{Var}[\hat{f}] = 4 \times \mathrm{Var}[R] = \frac{3N}{4}$. *Discuter de cette variance.*

2 Comparaison théorique entre $\mathrm{Var}_{\mathbf{Stair}}$ et $\mathrm{Var}_{\mathbf{Lb}}$

Exercice 2.1 (Etude du signe de la différence $\mathbf{Var_{Lb} - Var_{Stair}}$).

On rappelle que $\mathrm{Var}_{Lb} = 2(\frac{\Delta}{\epsilon})^2$ *et que* $\mathrm{Var}_{Stair} = \Delta^2 \times \dfrac{2^{-2/3}e^{-2\epsilon/3}(1+e^{-\epsilon})^{2/3} + e^{-\epsilon}}{(1-e^{-\epsilon})^2}$

1. *Pourquoi s'intéresse-t-on à ce signe ? Quelles sont les conséquences d'un signe toujours positif ?*
2. *Travailler avec ces deux fonctions de* ϵ *n'est pas aisé. A lieu de cela, on comparera un autre indicateur de dispersion qu'est l'écart absolu moyen (EBM). On a* $EBM_{Lb} = \frac{\Delta}{\epsilon}$ *et* $EBM_{Stair} = \Delta\frac{e^{\epsilon/2}}{e^{\epsilon}-1}$.

   Montrer qu'étudier le signe de $EBM_{Lb} - EBM_{Stair}$ *revient à étudier le signe de* $e^x - 1 - x.e^{x/2}$ *pour* $x \geq 0$.
3. *Etudier la dérivée de cette fonction. Conclure quant à cet indicateur de dispersion.*

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.

3 Etude de GRR

On reprend dans cette partie l'algorithme GRR (transparent 12 du cours 4).

Exercice 3.1. *On considère un ensemble de N personnes, chacune ayant une réponse personnelle dans $\{v_1, \ldots, v_k\}$. Soit f_i le nombre de fois où la valeur v_i apparaît dans l'ensemble initial de réponses personnelles.*

Soit r_i le nombre de fois où la valeur v_i apparaît dans l'ensemble obtenu après application du mécanisme $\mathcal{M}_{GRR}$ à chaque réponse individuelle.

1. *Montrer que $\hat{f}_i$ défini par*
$$\hat{f}_i = \frac{k-1+e^\epsilon}{e^\epsilon - 1} r_i - \frac{N}{e^\epsilon - 1} \tag{1}$$
est un estimateur de f_i.

2. *Montrer que la variance de $\hat{f}_i$, notée $\mathrm{Var}[\hat{f}_i]$ est définie par*

$$\begin{aligned}
\mathrm{Var}[\hat{f}_i] &= \left(\frac{k-1+e^\epsilon}{e^\epsilon - 1}\right)^2 Var[\hat{r_i}] & (2)\\
\mathrm{Var}[\hat{r_i}] &= f_i \times \delta_1(1-\delta_1) + (N - f_i) \times \delta_0(1-\delta_0) & (3)
\end{aligned}$$

avec $\delta_1 = \dfrac{e^\epsilon}{k-1+e^\epsilon}$ *et* $\delta_0 = \dfrac{1}{k-1+e^\epsilon}$