Sécurité Appliquée
Protection de la vie privée-PVP
Confidentialité différentielle

Jean-François Couchot

Université de Franche-Comté, UFR-ST

Plan

Confidentialité différentielle : formalisation

Confidentialité différentielle : mise en œuvre

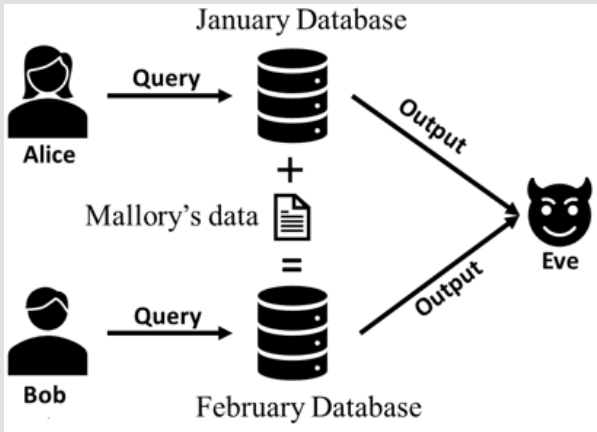
Propriétés générales de la ϵ-DP

Plan

Plan

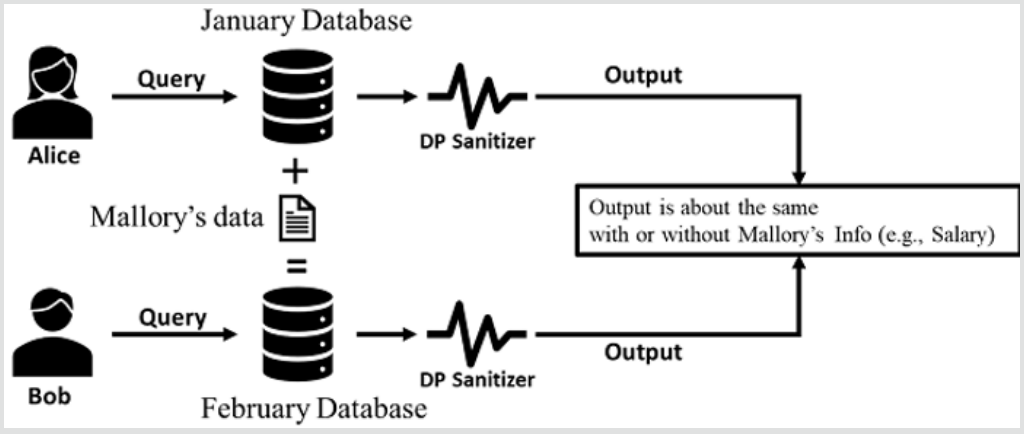
Ex. de requêtes sur des bases voisines [1]

Sans Confidentialité Différentielle



- Requête mensuelle : (nb empl., salaire moyen)
- Res. :
  $\{\text{jan.} : (100, \$55000), \text{fev.} : (101, \$56000)\}$
- Connaiss. suppl. : 0 sortie + Mallory en fev.
- $\rightsquigarrow$ salaire de Mallory : \$156000

Avec Confidentialité Différentielle



- Même requêtes, mêmes connaiss. suppl.
- Res. nettoyés :
  $\{\text{jan} : (102, \$55551),$
  $\text{fev} : (97, \$55975)\}$
- Salaire de Mallory ?

1. Privacy-Preserving Machine Learning. Manning Early Access Program Publications, 2021.

Idées clefs

Intuition pour 2 bases voisines D_1 et D_2

- Résultats (aggrégés, statistiques,. . .) proches
- $\Leftrightarrow$ *Probabilités* sur $\mathcal{M}(D_1)$ et $\mathcal{M}(D_2)$ égales (à ϵ près)

Pourquoi une confidentialité différentielle ?

- Les données privées : affectent peu les résultats
- $\rightsquigarrow$ Difficile de distinguer si une personne particulière *participe ou non*
- $\rightsquigarrow$ Propriétaire des données : moins inquiet-e de partager ses données

Plan

Formalisation de la DP [2]

Définition (ϵ-confidentialité différentielle (DP))

Soit $\epsilon \in \mathbb{R}^+$. L'algorithme probabiliste non déterministe $\mathcal{M}$ respecte la ϵ-confidentialité différentielle si

$$\begin{array}{ll} \forall D_1, D_2 \in \mathbb{N}^{|\mathcal{X}|} \text{ t.q. } \|D_1 - D_2\|_1 = 1, & (D_1 \text{ et } D_2 \text{ voisines}) \\ \forall R \text{ t.q. } R \subseteq \mathcal{M}(\mathbb{N}^{|\mathcal{X}|}), & (\text{pour tte image de l'algo.}) \\ \Pr[\mathcal{M}(D_1) \in R] \leq e^{\epsilon} \Pr[\mathcal{M}(D_2) \in R] & (\text{si } \epsilon \text{ petit, } e^{\epsilon} \approx 1 + \epsilon) \end{array}$$

Budget de fuite $\epsilon \in \mathbb{R}^+$: déviation permise, fuite autorisée

- $\Pr[\mathcal{M}(D_1) \in R] \leq e^{\epsilon} \Pr[\mathcal{M}(D_2) \in R]$: résultats approximativement (mais pas nécessairement) égaux avec/sans les informations d'1 utilisateur
- $\epsilon = 0$: aucune déviation permise (sorties toutes égales avec/sans les informations d'1 utilisateur), données parfaitement protégées, mais inutiles
- ϵ petit : petite déviation permise, grande protection, utilité moindre
- $\epsilon \in [0.001; 1]$ pour des tâches statistiques (moyennes, est^{on} de fréquence)
- $\epsilon \in [0.1; 20]$ pour de l'apprentissage machine

2. Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006, March). Calibrating noise to sensitivity in private data analysis. In Theory of cryptography conference (pp. 265-284). Springer, Berlin, Heidelberg.

Plan

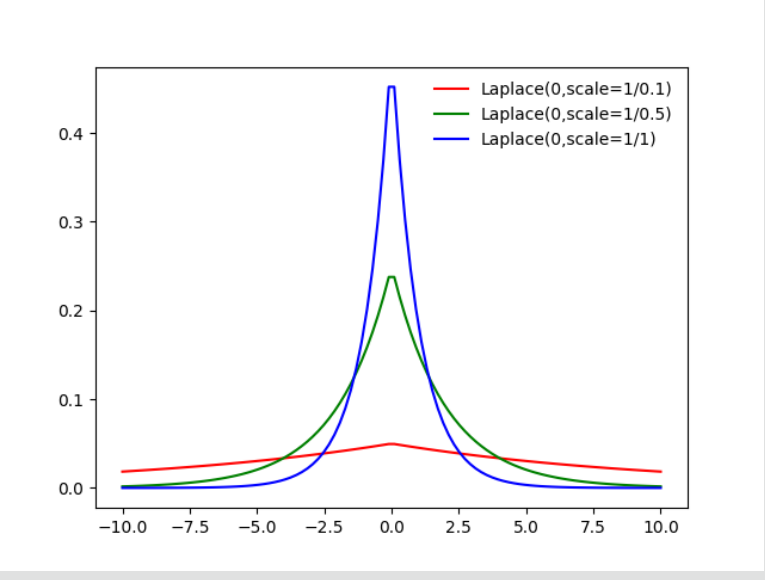
Plan

Requête 1 : nombre d'employés

Objectifs, données, idée

- Publier un nombre d'employé avec un mécanisme ϵ-DP
- jan. :100, fev. :101. . . .
- Ajouter un bruit centré en 0 dépendant d'ϵ :

MEO : bruit laplacien centré en 0, $\mathcal{M}_L(D) = Q(D) + v$, $v \sim Lap(0, \epsilon^{-1})$

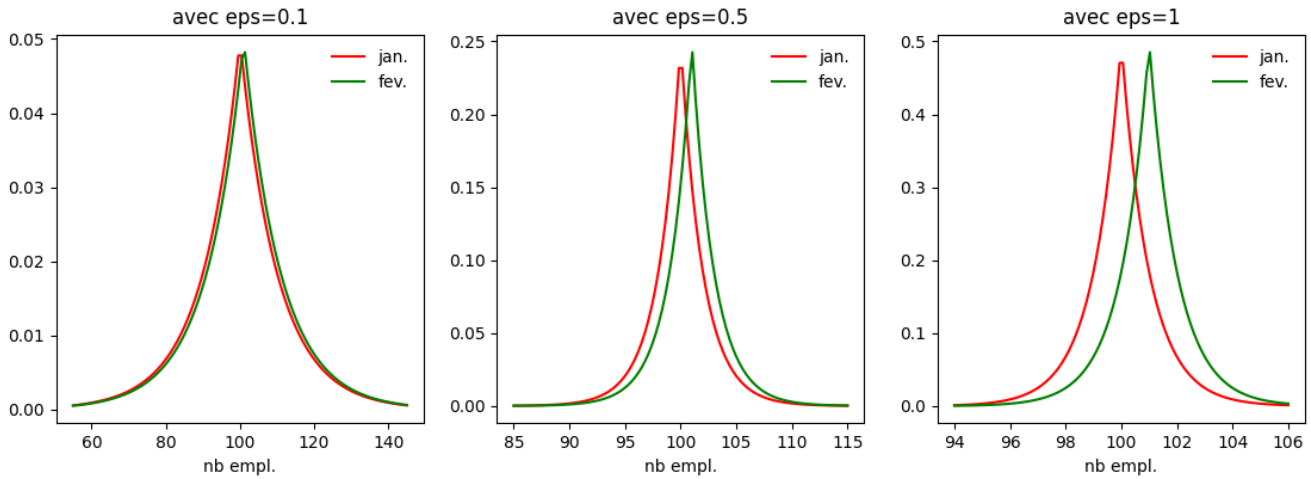


Requête 1 : nombre d'employés

Objectifs, données, idée

- Publier un nombre d'employé avec un mécanisme ϵ-DP
- jan. :100, fev. :101. . . .
- Ajouter un bruit centré en 0 dépendant d'ϵ :

MEO : bruit laplacien centré en 0, $\mathcal{M}_L(D) = Q(D) + v$, $v \sim Lap(0, \epsilon^{-1})$



Requête 2 : salaire moyen

Objectifs, données, idée

- Publier le salaire moyen avec un mécanisme ϵ-DP
- jan. :\$55000, fev. :\$56000 ...

MEO

- $\mathcal{M}_L(D) = Q(D) + v$, $v \sim Lap(0, \epsilon^{-1})$?
- Toutes les valeurs de janvier vont appartenir à $[54900, 55100]$
- Toutes les valeurs de fevrier vont appartenir à $[55900, 56100]$
- Le bruit doit déprendre de la sensibilité Δ_Q de la requête Q

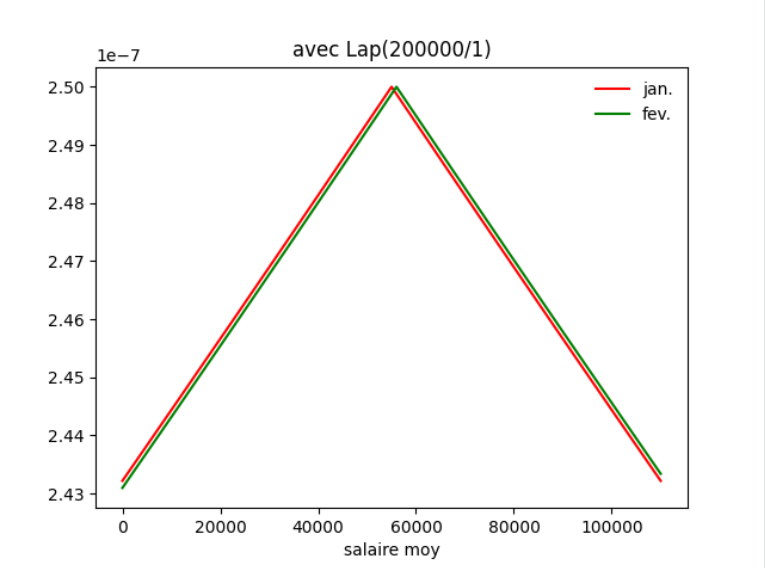
Requête 2 : salaire moyen

Objectifs, données, idée

- Publier le salaire moyen avec un mécanisme ϵ-DP
- jan. :\$55000, fev. :\$56000 ...

MEO

- $\mathcal{M}_L(D) =$
  $Q(D) + v,\ v \sim Lap(0, \frac{\Delta_Q}{\epsilon})$?



Plan

Mécanisme de Laplace, définition

Sensibilité de la requête $Q : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^k$

$$\Delta_Q = \max_{D_1, D_2 \text{ t.q. } \|D_1 - D_2\|_1 = 1} \|Q(D_1) - Q(D_2)\|_1$$

- $Q(D)$ = nb pers. aux yeux bleus $\rightsquigarrow \Delta_Q = 1$
- $Q(D)$ = (nb pers. aux yeux bleus, nb pers. aux yeux noirs) $\rightsquigarrow \Delta_Q = 1$
- $Q(D)$ = (nb pers. aux yeux bleus, nb pers. de plus d'1m60) $\rightsquigarrow \Delta_Q = 2$

Mécanisme laplacien (ϵ, δ)-DP pour $Q : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}$

$$\mathcal{M}_L(D) = Q(D) + v \text{ t.q. } v \sim Lap(\frac{\Delta_Q}{\epsilon})$$

- chaque donnée : perturbée indépendamment selon Laplace (sensibilité, ϵ)
- bruit : croît avec Δ_Q, décroît avec ϵ

Mécanisme de Laplace, preuve

Densité de probabilité de la loi de Laplace

- Une variable aléatoire possède une distribution $Laplace(0, b)$ si sa densité de probabilité est $f_{Lap(b)}(z) = \frac{e^{\frac{-|z|}{b}}}{2b}$
- Rappel : pour a et b dans $\mathbb{R}$ on a $\Pr[v_1 \leq v \leq v_2] = \int_{v_1}^{v_2} f_{Lap(b)}(z)\,\mathrm{d}z$

Preuve

- Montrons que $\Pr[\mathcal{M}(D_1) \in]-\infty, r]] \leq e^{\epsilon} \Pr[\mathcal{M}(D_2) \in]-\infty, r]]$
- $\Pr[\mathcal{M}_L(D_1) \leq r] = \Pr[Q(D_1) + v \leq r] = \Pr[v \leq r - Q(D_1)]$
- Pour $r - Q(D_1)$ négatif (autre cas similaire) :

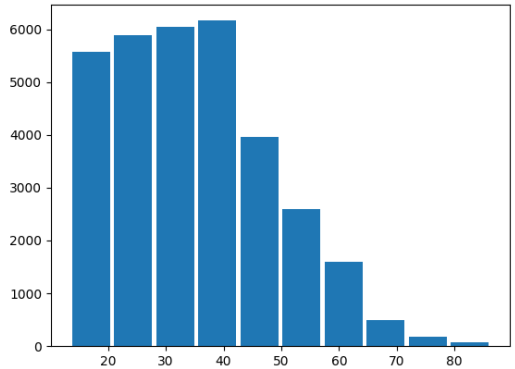
$$\Pr[\mathcal{M}_L(D_1) \leq r] = \int_{-\infty}^{r-Q(D_1)} \frac{e^{\frac{-z}{b}}}{2b}\,dz = \left[-\frac{e^{\frac{-z}{b}}}{2}\right]_{-\infty}^{r-Q(D_1)} = -\frac{e^{-\frac{r-Q(D_1)}{b}}}{2}$$

- $$\frac{\Pr[\mathcal{M}_L(D_1) \leq r]}{\Pr[\mathcal{M}_L(D_2) \leq r]} = e^{\frac{(r-Q(D_2))-(r-Q(D_1))}{b}} = e^{\frac{\epsilon(Q(D_1)-Q(D_2))}{\Delta_Q}} \leq e^{\epsilon}$$

Application aux histogrammes

Sans confidentialité différentielle

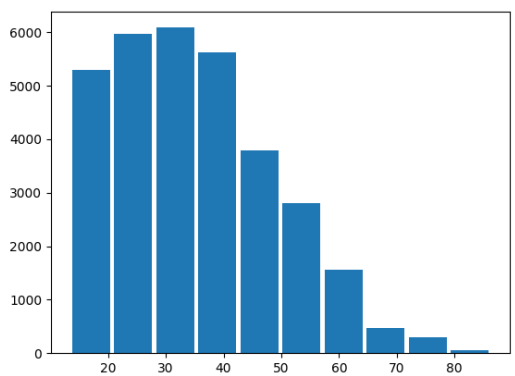
```
import numpy as np
import matplotlib.pyplot as plt
ages_adult=np.loadtxt("adult.data",usecols=0, delimiter=", ")
hist, bins=np.histogram(ages_adult)
plt.bar(bins[:-1], hist, width=(bins[1]-bins[0]) * 0.9)
plt.show()
```



Avec confidentialité différentielle

- Sensibilité $= 1$: ajouter/retrancher une personne ne change que de 1 l'effectif d'une classe

```
...
hist, bins=np.histogram(ages_adult)
eps=0.01
hist= [v+np.random.laplace(0,1/eps) for v in hist]
# et si valeur négative ?
```



Plan

Mécanisme exponentiel : intro

Ajouter un bruit numérique à une requête $Q : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}$: toujours sensé ?

- à réponse textuelle :« couleur d'yeux la plus fréquente dans la classe ? » $\rightsquigarrow$ $\mathcal{R} = \{\text{rouge}, \text{bleu}, \text{vert}, \text{marron}, \text{noisette}\}$
- à réponse détruite par le bruit [3] : « Soit une offre abondante de citrouilles et 4 acheteurs : A, F, I, K, où A, F, I offrent chacun 1,00\$ et K offre 3,01\$. Quel est le prix optimal ? A 3,01\$, le revenu est de 3,01\$, à 3,00\$ et à 1,00\$, le revenu est de 3,00\$, mais à 3,02\$, le revenu est nul ! » $\rightsquigarrow$ $R = \{1.00\$, 3.00\$, 3.01\$, \ldots\}$

Une fonction d'utilité $u : \mathbb{N}^{|\mathcal{X}|} \times R \to \mathbb{R}$

Intuition : pour $D_1 \in \mathbb{N}^{|\mathcal{X}|}$, et $r \in R$, $u(D_1, r)$ sera élevé si r est "important" pour D_1

Sensibilité de la fonction d'utilité u

$$\Delta_u = \max_{r \in R, D_1, D_2 \text{ t.q. } \|D_1 - D_2\|_1 = 1} \|u(D_1, r) - u(D_2, r)\|_1$$

3. Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3-4), 211-407.

Mécanisme exponentiel : def et exemple

Définition pour D, $u : \mathbb{N}^{|\mathcal{X}|} \times R \to \mathbb{R}$

$\mathcal{M}_E(D) = r$ avec une probabilité proportionnelle à $\exp(\frac{\epsilon u(D,r)}{2\Delta_u})$

Exemple : nationalité la plus commune dans D ?[4], $\epsilon = 2$

- $R = \{\text{chinoise}, \text{indienne}, \text{américaine}, \text{grecque}\}$, $D = (6, 5, 3, 2)$
- $u(D, r) =$ « nombre d'individus dans D de nationalité r » $\rightsquigarrow \Delta_u = 1$

r	chinoise	indienne	américaine	grecque
$\exp(\frac{\epsilon u(D,r)}{2\Delta_u})$	$e^6 \approx 403$	$e^5 \approx 148$	$e^3 \approx 20$	$e^2 \approx 7$
$\Pr[\mathcal{M}_E(D) = r]$	0.70	0.26	0.03	0.01

Remarques

- preuve d'$(\epsilon, 0)$-DP : sera vue en TD
- garantie élevée d'utilité de $\mathcal{M}_E$: résultats avec note d'utilité faible écartés exponentiellement, rapidement

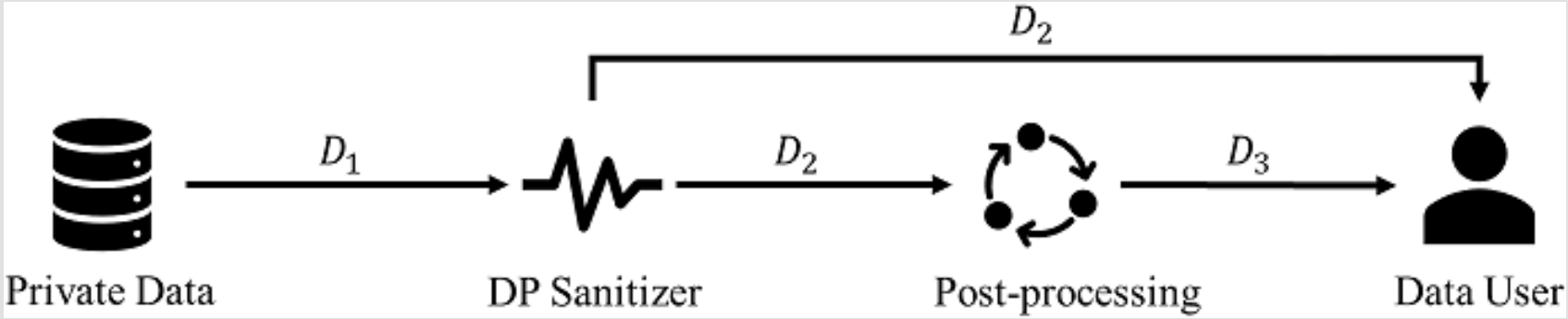
4. Benkhelif, T. (2018). Publication de données individuelles respectueuse de la vie privée : une démarche fondée sur le co-clustering (Doctoral dissertation, Université de Nantes).

Plan

Plan

Robustesse au post-traitement, idées

Intuition pour une base D_1 [5]



Interprétations

- *Post – traitement* s'il est vu comme un algorithme ultérieur (suppressions des valeurs insensées p.ex.) : peu importe, seul l'algo de DP est à considérer avec soin
- *Post – traitement* vue comme une attaque d'un-e adversaire : elle/il peut incorporer autant d'informations auxiliaires qu'elle/il veut ; la garantie de confidentialité reste valable quoi qu'elle/il fasse

5. Privacy-Preserving Machine Learning. Manning Early Access Program Publications, 2021.

Robustesse au post-traitement, formalisation

Théorème pour $\mathcal{M}$ ϵ-DP

Pour toute fonction $f : \mathcal{M}(\mathbb{N}^{|\mathcal{X}|}) \rightarrow f(\mathcal{M}(\mathbb{N}^{|\mathcal{X}|}), f(\mathcal{M})$ est ϵ-DP.

Preuve

$\forall (D_1, D_2)$ t.q. $\|D_1 - D_2\|_1 = 1$,
$\forall S$ t.q. $S \subseteq f(\mathcal{M}(\mathbb{N}^{|\mathcal{X}|})$,

$$\begin{aligned}
\Pr[f(\mathcal{M}(D_1)) \in S] &= \Pr[\mathcal{M}(D_1)) \in f^{-1}(S)] \\
&\leq e^{\epsilon} \Pr[\mathcal{M}(D_2)) \in f^{-1}(S)] \\
&\leq e^{\epsilon} \Pr[f(\mathcal{M}(D_2))) \in S]
\end{aligned}$$

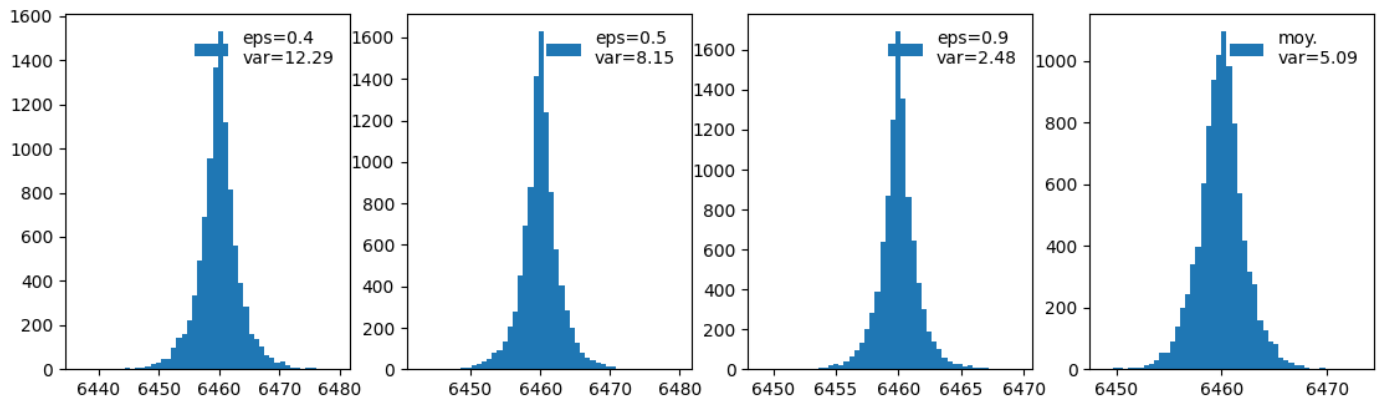
Plan

Composition séquentielle, intuitions

Séquences de fuites

- Usuel de requêter itérativ^{nt} la même base (effectif en jan. ? en fev. ? p.ex.)
- Chaque requête $\equiv$ une fuite de données : possibilité de trouver une valeur proche de la réalité en moyennant les réponses bruitées
- Valeur de la fuite totale ϵ pour la séquence de fuites ϵ_1, ϵ_2 ?

Approche expérimentale sur le nombre de personnes de plus de 50 ans



Compositions formalisation

Définition (Compositions de $\mathcal{M}_1 : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}_1$ et $\mathcal{M}_2 : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}_2$)

La composition $\mathcal{M}_{1,2} : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}_1 \times \mathcal{R}_2$ de $\mathcal{M}_1$ et $\mathcal{M}_2$ est définie par $\mathcal{M}_{1,2}(D) = (\mathcal{M}_1(D), \mathcal{M}_2(D))$.

Théorème (Compositions de $\mathcal{M}_1$ ϵ_1-DP et $\mathcal{M}_2$ ϵ_2-DP)

- *Si $\mathcal{M}_1$ et $\mathcal{M}_2$ opèrent sur des ensembles disjoints (composition paralelle) : $\mathcal{M}_{1,2}$ est* $\max(\epsilon_1, \epsilon_2)$*-DP*
- *Si $\mathcal{M}_1$ et $\mathcal{M}_2$ opèrent sur des ensembles non disjoints (composition séquentielle) : $\mathcal{M}_{1,2}$ est $\epsilon_1 + \epsilon_2$-DP*

Exemple de combinaison avec $\epsilon = 0.5 + 0.5$

Données et requêtes

Sx	Zip Code	Age	Disease	Salary
M	400071	35	bronchitis	10k
M	400182	37	pneumonia	11k
M	400095	39	stomach cancer	12k
F	440672	54	gastritis	12k
F	440123	58	Flu	15k
M	440893	54	bronchitis	16k
M	400022	41	gastric ulcer	16k
M	400135	46	gastritis	17k
F	400182	44	stomach cancer	18k

- Q_1 : « nb de pers. avc pb gastrique »
- Q_2 : « age moyen des patients »
- D', à distance 1 : obtenue en supprimant/ajoutant 1 ligne
- $\Delta_{Q_1} = 1$
- $\Delta_{Q_2} = \frac{U-L}{n+1}$, U la borne max, l la borne min se démontre $\rightsquigarrow$ $\Delta_{Q_2} \approx \frac{100-20}{10} = 8$

Combinaison

- $Q_1(D) + Lap(0, 1/0.5) = 5 - 2.25 = 2.75 \rightsquigarrow 3$ grace à une approximation (post traitement)
- $Q_2(D) + Lap(0, 8/0.5) = 45.33 + 8.92$
- publication de $(3, 54.25)$ pour $\epsilon = 1$

Questions ouvertes

- Les mécanismes précédents lorsqu'ils sont utilisés pour du comptage ne doivent renvoyer que des valeurs positives.... quid des valeurs négatives ?
- Les mécanismes précédents visent à répondre à une requête. Adaptés à de l'apprentissage machine ?
- Quel mécanisme choisir lorsque deux sont possibles ?