Sécurité Appliquée
Protection de la vie privée-PVP
Confidentialité différentielle

Jean-François COUCHOT
Université de Franche-Comté, UFR-ST

Plan

Confidentialité différentielle : formalisation
- Motivation
- Bases de données voisines
- Propriété sur l'algorithme de réponse anonymisée

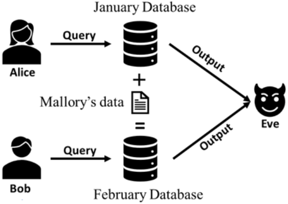
Confidentialité différentielle : mise en œuvre

Propriétés générales de la ϵ-DP

Plan

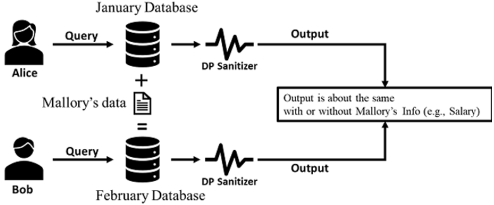
Ex. de requêtes sur des bases voisines [1]

Sans Confidentialité Différentielle



- Requête mensuelle : (nb empl., salaire moyen)
- Res. : $\{$jan. : $(100, \$55000)$, 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 : $\{$jan : $(102, \$55551)$, fev : $(97, \$55975)\}$
- Salaire de Mallory ?

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

Idées clefs

Intuition pour 2 bases D_1 et D_2 voisines l'une de l'autre

- Résultats (aggrégés, statistiques,...) proches
- $\Leftrightarrow$ *Probabilités* sur $\mathcal{M}(D_1)$ et $\mathcal{M}(D_2)$ égales (à ϵ près)
- 2 bases D_1 et D_2 voisines l'une de l'autre : reste à le formaliser

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

D'une base à son histogramme

Exemples d'univers $\mathcal{X}$ des valeurs possibles

- toutes les tailles de vêtements : $\mathcal{X} = \{XXS, \ldots, XXL\}$,
- tous les salaires annuels : $\mathcal{X} = \mathbb{N}$,
- toutes les coordonnées géographique : $\mathcal{X} = [-90, 90] \times [-180, 180]$ ...

Représentation sous forme d'histogramme

- Base D : multi-ensemble d'éléments de $\mathcal{X}$, représentée par un histogramme : $D \in \mathbb{N}^{|\mathcal{X}|}$
  - Exemple d'une base $D = [5'0, 5'1, 5'2, 5'1, 5'4]$ de tailles (pieds) sur $\mathcal{X} = \{5'0, 5'1, 5'2, 5'1, 5'3, 5'4\}$
  - Connaissant $\mathcal{X}$, on peut définir D comme un histogramme $D = (1, 2, 1, 0, 1)$

Distance entre bases de données

Norme 1 pour un vecteur $x = (x_1, \ldots, x_k)$ de $\mathbb{R}^k$

La norme 1 de x est :

$$|x\|_1 = |x_1| + \ldots + |x_k|$$

Norme 1 entre deux bases D_1 et D_2

- Exprimer D_1 et D_2 comme deux vecteurs
- Norme-1 du vecteur de différences entre les deux $\|D_1 - D_2\|_1$
- Exemple avec $D_1 = [5'0, 5'1, 5'2, 5'1, 5'4]$ et $D_2 = [5'0, 5'1, 5'2, 5'0, 5'3]$
  - $D_1 = (1, 2, 1, 0, 1) \rightsquigarrow \|D_1\|_1 = |1| + |2| + |1| + |0| + |1| = 5$
  - $D_2 = (2, 1, 1, 1, 0) \rightsquigarrow \|D_2\|_1 = |2| + |1| + |1| + |1| = 5$
  - $\|D_1 - D_2\|_1 = \|(-1, 1, 0, -1, 1)\| = |-1| + |1| + |-1| + |1| = 4$

Bases de données D_1 et D_2 voisines

Définition pour deux bases de données D_1 et D_2

D_1 et D_2 sont voisines si la distance entre elles vaut 1.

$D_1 = (1, 2, 1, 0, 1), D_2 = (2, 1, 1, 1, 0), D_3 = (1, 2, 2, 0, 1), D_4 = (0, 2, 1, 1, 0)$

- $\|D_1 - D_2\|_1 = 4 \rightsquigarrow D_1$ et D_2 non voisines : même nombre de données mais diffèrent sur la donnée d'au moins 2 persones
- $\|D_1 - D_3\|_1 = 1 \rightsquigarrow D_1$ et D_3 voisines : diffèrent sur la donnée d'une seule personne, une donnée en $+$ dans D_3
- $\|D_1 - D_4\|_1 = 3 \rightsquigarrow D_1$ et D_4 non voisines : diffèrent sur la donnée d'au moins 2 personnes, une donnée en - dans D_4

Plan

Confidentialité différentielle : formalisation

- Motivation
- Bases de données voisines
- Propriété sur l'algorithme de réponse anonymisée

Confidentialité différentielle : mise en œuvre

Propriétés générales de la ϵ-DP

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 égaux (mais pas nécessairement) avec/sans la donnée d'1 personne
- $\epsilon = 0$: aucune déviation permise (sorties toutes égales avec/sans la donnée d'1 personne), 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$^{\text{on}}$ 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

Confidentialité différentielle : formalisation

Confidentialité différentielle : mise en œuvre

- Intuitions
- Requêtes numériques : mécanisme de Laplace
- Requêtes non numériques : mécanisme exponentiel

Propriétés générales de la ϵ-DP

Plan

Confidentialité différentielle : formalisation

Confidentialité différentielle : mise en œuvre

- Intuitions
- Requêtes numériques : mécanisme de Laplace
- Requêtes non numériques : mécanisme exponentiel

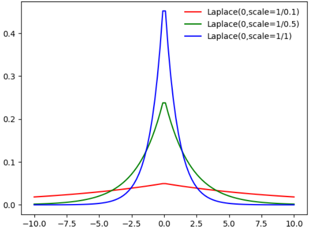
Propriétés générales de la ϵ-DP

Requête Q_1 : nombre d'employés dans la base

Objectifs, données, idée

- Publier un nombre d'employés avec un mécanisme ϵ-DP
- $Q_1(D_{\text{jan}}) = 100$, $Q_1(D_{\text{fev}}) = 101 \dots$
- Ajouter un bruit centré en 0 dépendant d'ϵ :

MEO : bruit laplacien centré en 0, $\mathcal{M}_L(D) = Q_1(D) + v$, $v \sim Lap(0, \epsilon^{-1})$

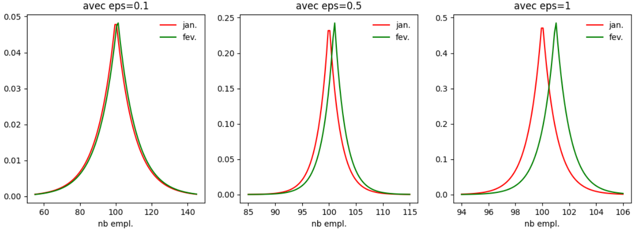


Requête Q_1 : nombre d'employés dans la base

Objectifs, données, idée

- Publier un nombre d'employés avec un mécanisme ϵ-DP
- $Q_1(D_{\text{jan}}) = 100,\ Q_1(D_{\text{fev}}) = 101. \ldots$
- Ajouter un bruit centré en 0 dépendant d'ϵ :

MEO : bruit laplacien centré en 0, $\mathcal{M}_L(D) = Q_1(D) + v,\ v \sim Lap(0, \epsilon^{-1})$



Requête Q_2 : salaire moyen

Objectifs, données, idée

- Publier le salaire moyen avec un mécanisme ϵ-DP
- $Q_2(D_{\text{jan}}) = \$55000$, $Q_2(D_{\text{fev}}) = \$56000 \ldots$

MEO

- $\mathcal{M}_L(D) = Q_2(D) + v$, $v \sim Lap(0, \epsilon^{-1})$?
- 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

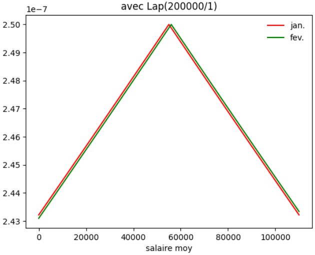
Requête Q_2 : salaire moyen

Objectifs, données, idée

- Publier le salaire moyen avec un mécanisme ϵ-DP
- $Q_2(D_{\text{jan}}) = \$55000$, $Q_2(D_{\text{fev}}) = \$56000 \ldots$

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 $Q : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}$

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

- Q= “nb pers. aux yeux bleus” $\rightsquigarrow \Delta_Q = 1$
- Q= “nb pers. aux yeux bleus, nb pers. aux yeux noirs” $\rightsquigarrow \Delta_Q = 1$
- Q= “nb pers. aux yeux bleus, nb pers. de plus d’1m60” $\rightsquigarrow \Delta_Q = 2$

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

$$\mathcal{M}_L(D) = Q(D) + v \text{ t.q. } v \sim Lap(0, \frac{\Delta_Q}{\epsilon})$$

- 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 $f_{Lap(b)}(z) = \frac{e^{\frac{-|z|}{b}}}{2b}$
- Ici $b = \frac{\Delta}{\epsilon}$

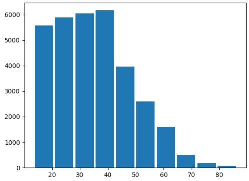
Preuve : montrons que $\forall r \in R \,.\, \frac{\Pr[\mathcal{M}(D_1)=r]}{\Pr[\mathcal{M}(D_2)=r]} \leq e^{\epsilon}$

- $\frac{\Pr[\mathcal{M}(D_1)=r]}{\Pr[\mathcal{M}(D_2)=r]} = \frac{\Pr[Q(D_1)+v=r]}{\Pr[Q(D_2)+v=r]} = \frac{\Pr[v=r-Q(D_1)]}{\Pr[v=r-Q(D_2)]} = (*) = \frac{e^{\frac{|r-Q(D_1)|}{b}}}{2b} \cdot \frac{2b}{e^{\frac{|r-Q(D_2)|}{b}}}$
- $= e^{\frac{|r-Q(D_1)|-|r-Q(D_2)|}{b}} \leq (**) \leq e^{\frac{|Q(D_1)-Q(D_2)|}{b}} \leq e^{\epsilon}$
- (*) fonct° f de densité d'une var. aléatoire réelle X et $\mathrm{d}t \in \mathbb{R}^+$ infinit petit
  - $\Pr[t_0 < X < t_0 + \mathrm{d}t] = f(t_0)\,\mathrm{d}t. \rightsquigarrow \frac{\Pr[t_0<X<t_0+\mathrm{d}t]}{\Pr[t_1<X<t_1+\mathrm{d}t]} = \frac{f(t_0)}{f(t_1)}$
- (**) $|r-Q(D_1)| = |r-Q(D_2)+Q(D_2)-Q(D_1)| \leq |r-Q(D_2)|+|Q(D_2)-Q(D_1)|$

Application aux histogrammes

Sans confidentialité différentielle

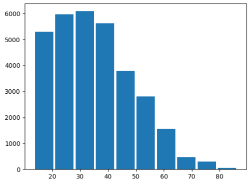
```python
import numpy as np
import matplotlib.pyplot as plt
ages_adult=np.loadtxt("adult.data",usecols=0, delimiter=", ")
hist, bins=np.histogram(ages_adult)
plt.bar(bins[:-1], hist, width=(bins[1]-bins[0])*0.9)
plt.show()
```



Avec confidentialité différentielle

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

```python
...
hist, bins=np.histogram(ages_adult)
eps=0.01
hist= [max(0,q+np.random.laplace(0,1/eps))
       for q in hist] # pourquoi max...
```



Plan

Confidentialité différentielle : formalisation

Confidentialité différentielle : mise en œuvre

- Intuitions
- Requêtes numériques : mécanisme de Laplace
- Requêtes non numériques : mécanisme exponentiel

Propriétés générales de la ϵ-DP

Mécanisme exponentiel : intro

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

- à réponse textuelle :« couleur d'yeux la plus fréquente dans la classe ? » $\rightsquigarrow$ $R = \{\text{rouge}, \text{bleu}, \text{vert}, \text{marron}, \text{noisette}\}$
- 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 et K offre \$3.01. Quel est le prix optimal ? A \$3.01, le revenu est de \$3.01, à \$3 et à \$1, le revenu est de \$3, mais à \$3.02, le revenu est nul ! » $\rightsquigarrow$ $R = \{\$1, \$3, \$3.01\}$

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 : definition et exemple

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

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

- Remarque : valeurs avec utilité faible écartées exponentiel[t] rapidement
- Remarque : preuve d'ϵ-DP vue en TD

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

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).

Mécanisme exponentiel : definition code

Code brut pour le mécanisme exponentiel

```python
import numpy as np
import pandas as pd

def u(D, r):
    return D.value_counts()[r]

def exp_prob(D,R,u_func, sensitivity, eps):
    freq = [np.exp(eps*u_func(D,r)/(2*sensitivity)) for r in R]
    return freq/sum(freq)

def exponential(D, R, prob):
    return np.random.choice(R, 1, p=prob)[0]

df = pd.Series(["chinoise","chinoise","chinoise","chinoise","chinoise",
    "chinoise","indienne","indienne","indienne","indienne","indienne",
    "américaine","américaine","américaine","grecque","grecque"])
vc = df.value_counts()
prob = exp_prob(df,vc.index,u,1,2)
print(exponential(df, vc.index, prob))
```

Plan

Confidentialité différentielle : formalisation

Confidentialité différentielle : mise en œuvre

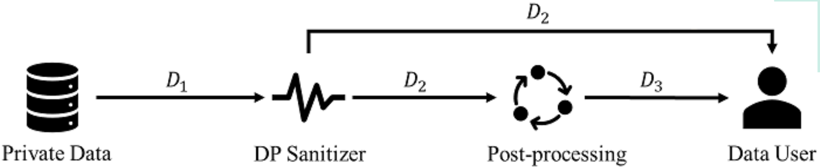
Propriétés générales de la ϵ-DP

- Post-traitements de mécanismes ϵ-DP
- Compositions de mécanismes ϵ-DP

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* vu 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

$$\begin{aligned}
&\forall (D_1, D_2) \text{ t.q. } \|D_1 - D_2\|_1 = 1, \\
&\forall S \text{ t.q. } S \subseteq f(\mathcal{M}(\mathbb{N}^{|\mathcal{X}|}), \\
&\Pr[f(\mathcal{M}(D_1)) \in S] && = \Pr[\mathcal{M}(D_1)) \in f^{-1}(S)] \\
& && \leq e^{\epsilon} \Pr[\mathcal{M}(D_2)) \in f^{-1}(S)] \\
& && \leq e^{\epsilon} \Pr[f(\mathcal{M}(D_2))) \in S]
\end{aligned}$$

Plan

Confidentialité différentielle : formalisation

Confidentialité différentielle : mise en œuvre

Propriétés générales de la ϵ-DP

Post-traitements de mécanismes ϵ-DP

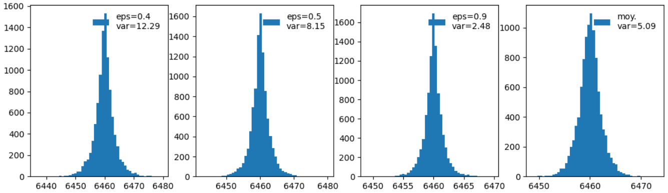
Compositions de mécanismes ϵ-DP

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 ?