Quelques avancées en analyses de données multidimensionnelles respectueuses de la vie privée.

Jean-François Couchot[1], *Héber* Hwang-Arcolezi[2]

[1]Université de Franche-Comté, FEMTO-ST

[2]École Polytechnique, Inria Saclay, LIX

Plan

Plan

Données multidim. et respect de la VP

A propos des données multidimensionnelles

- Ensemble de tuples portant sur d attributs $\{A_1, \ldots, A_d\}$ discrets
- Relation/Table sur des données au sens SGBD

Objectif

- Construire des histogrammes pour ces attributs en respectant la vie privée.

Objectifs réduits, intérêt limité ?

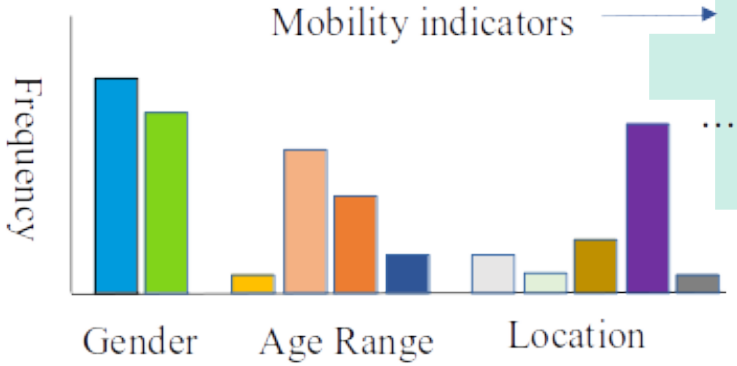
- Voir transparent suivant

Analyses multidime^{elles} respectueuses : exemples

Mobilité humaine par Orange/Flux Vision

Flux Vision
Mobility models

orange



GAFAM

RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response

Úlfar Erlingsson
Google, Inc.
ulfar@google.com

Vasyl Pihur
Google, Inc.
vpihur@google.com

Aleksandra Korolova
University of Southern California
korolova@usc.edu

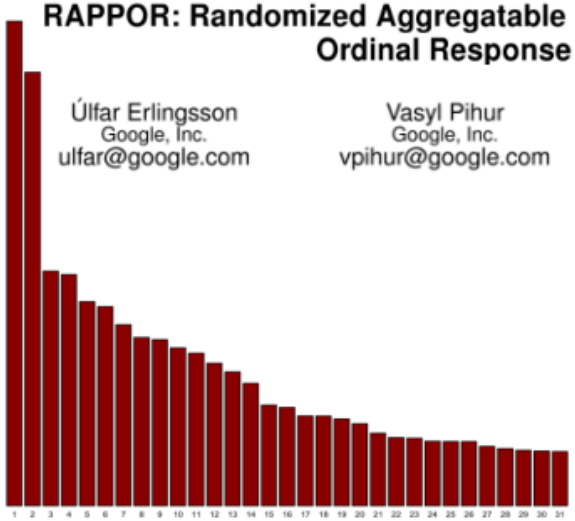
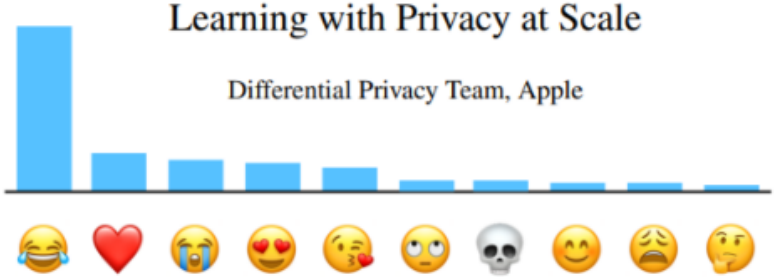


Figure 6: Relative frequencies of the top 31 unexpected Chrome homepage domains found by analyzing ∼14 million RAPPOR reports, excluding expected domains (the homepage "google.com", etc.).

Learning with Privacy at Scale

Differential Privacy Team, Apple



The Count Mean Sketch technique allows Apple to determine the most popular emoji to help design better ways to find and use our favorite emoji. The top emoji for US English speakers contained some surprising favorites.

Collecting Telemetry Data Privately

Bolin Ding, Janardhan Kulkarni, Sergey Yekhanin
Microsoft Research
{bolind, jakul, yekhanin}@microsoft.com

Windows Insiders in Windows 10 Fall Creators Update to protect users' privacy while collecting application usage statistics.

Respecter la vie privée

Pourquoi ?

- Dans la déclaration universelle des droits de l'homme
- Contexte d'utilisation des données personnelles formalisé dans le RGPD

Comment ?

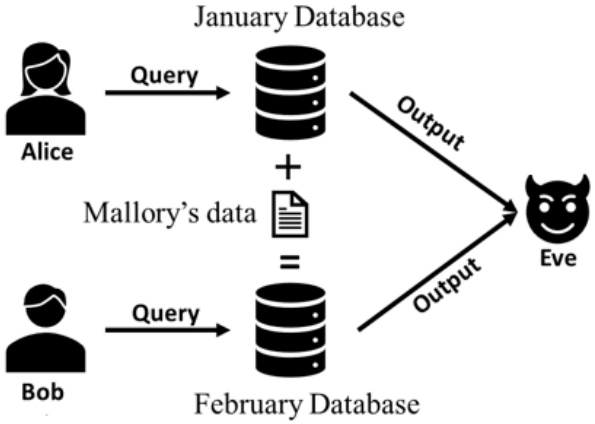
- Besoin de méthodes dont la fuite de données est mesurable
  - Optimiser une utilité p.r. à cette fuite

Plan

Plan

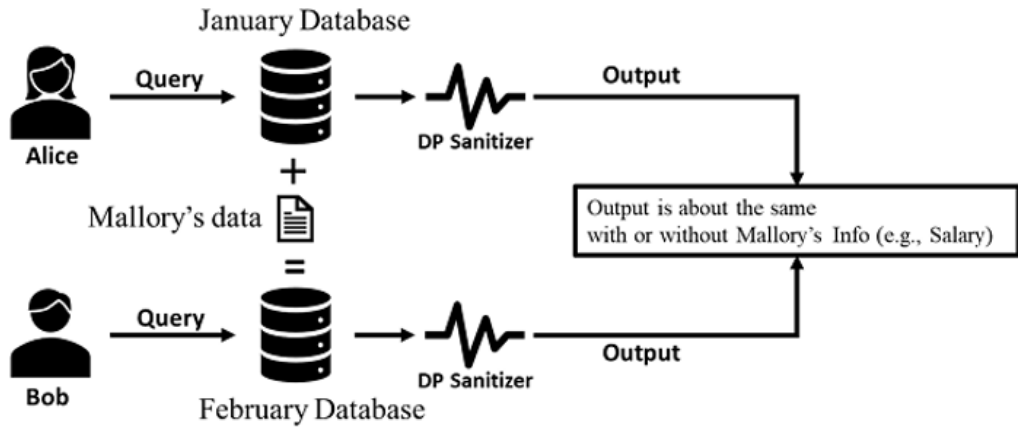
Ex. de requêtes sur des bases voisines [1]

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 D_1 et D_2 voisines l'une de l'autre

- Résultats (agrégés, 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 : souhait qu'elles 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

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

colab lister tous les fichiers d'un répertoire drive

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. . . grand : tout dépend de la fuite qu'on s'autorise

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.

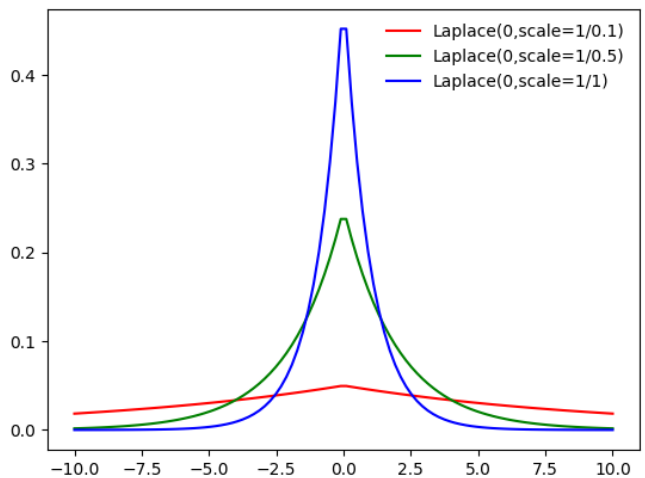
Plan

Requête Q_1 : nombre d'employé·e·s dans la base

Objectifs, données, idée

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

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

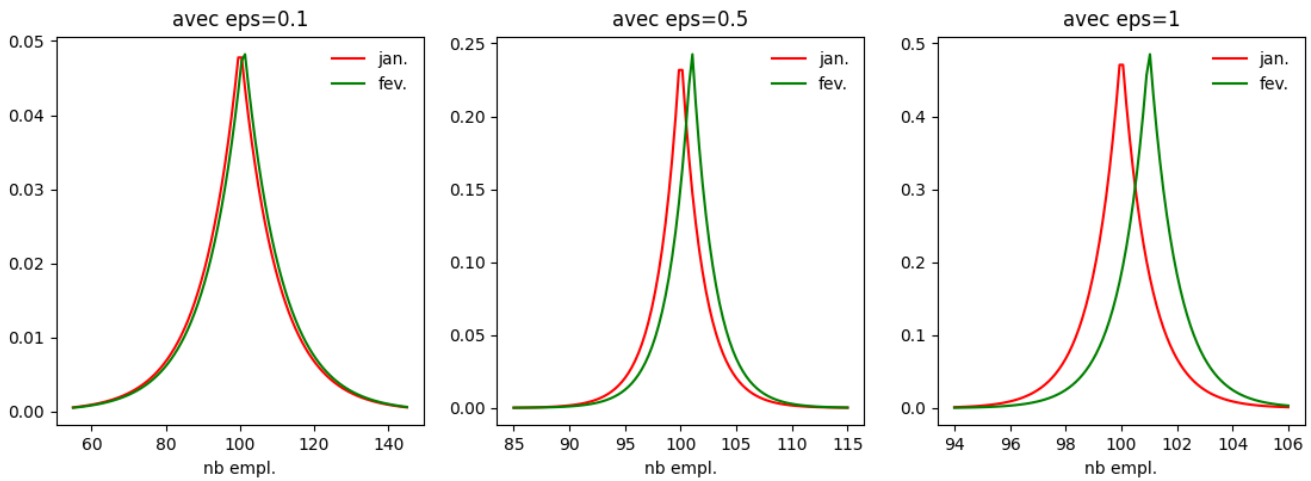


Requête Q_1 : nombre d'employé·e·s dans la base

Objectifs, données, idée

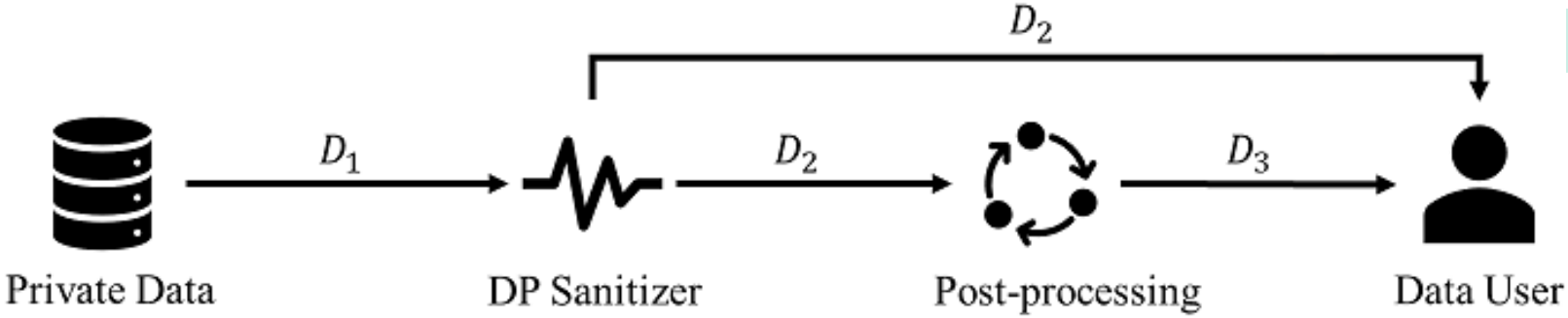
- Publier un nombre d'employé·e·s avec un mécanisme ϵ-DP
- $Q_1(D_{\text{jan}}) = 100$, $Q_1(D_{\text{fev}}) = 101$. . . .
- Ajouter un bruit centré en 0 dépendant d'ϵ :

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



Robustesse au post-traitement

Intuition pour une base D_1 [3]



Interprétations

- *Post – traitement* s'il est vu comme un algorithme ultérieur (suppressions des valeurs insensées p.ex.) : 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

Théorème (Post-traitement d'un mécanisme ϵ-DP)

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

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

Composition séquentielle

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 ?

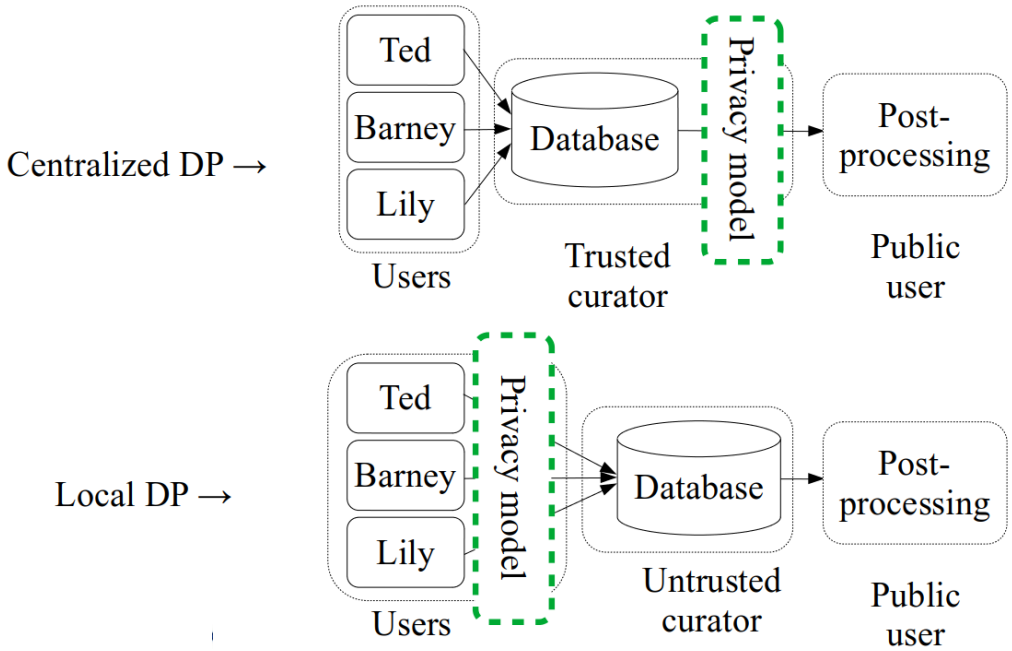
Théorème (Compositions séquentielle de $\mathcal{M}_1$ ϵ_1-DP et $\mathcal{M}_2$ ϵ_2-DP)

Si $\mathcal{M}_1$ et $\mathcal{M}_2$ opèrent sur des ensembles non disjoints, $\mathcal{M}_{1,2}$ est $\epsilon_1 + \epsilon_2$-DP

Plan

Motivations

En images



DP vs. LDP

- Confiance nécessaire envers le SGBD.
- Bruit optimal par requête.
- Bruit individuel pour tous les post-traitements (ML p.ex).
- Confiance superflue envers le SGBD.

Définition [4] et propriétés

Définition de ϵ-confidentialité différentielle locale (ϵ-LDP)

- $\mathcal{X}$: ensemble des valeurs possibles en entrée
- $\epsilon \in \mathbb{R}^+$: budget de fuite
- $\mathcal{M}$: algorithme probabiliste non déterministe qui respecte la ϵ-confidentialité différentielle locale si

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

Propriétés similaires à celles de la DP

- Robustesse au post-traitement
- Combinaison de deux mécanismes ϵ_1-LDP et ϵ_2-LDP : est $\epsilon_1 + \epsilon_2$-LDP

4. Duchi, J. C., Jordan, M. I., & Wainwright, M. J. (2013, October). Local privacy and statistical minimax rates. In 2013 IEEE 54th Annual Symposium on Foundations of Computer Science (pp. 429-438). IEEE.

Motivation ; données embarrassante à nettoyer [6]

Table avec 1 seul attribut binaire : Q_1="avez-vous triché au moins une fois ?"

- Embarras : tentation pour un·e étudiant·e de ne pas répondre honnêtement.

Bruiter selon Warner [5]

- Chaque étudiant·e lance 2 fois une pièce de monnaie $\{\text{Pile}, \text{Face}\}$ sans montrer les 2 résultat successifs t_1 et t_2.
- Ajout de la question Q_2 : « Est-ce que t_2 est égal à Pile ? ».
  - Si t_1 vaut Pile, l'étudiant·e répond honnêtement à la question Q_1.
  - Sinon ($t_1 = \text{Face}$), l'étudiant·e répond honnêtement à la question Q_2.

Analyse de l'extension

- Réponse partiellement aléatoire : on ne sait pas si une réponse OUI d'un·e étudiant·e provient d'une tricherie ou d'un Pile au second tirage.
- Honnêteté de l'étudiant·e renforcée : c'est lui·elle qui modifie ses données.

5. Warner, S. L. (1965). Randomized response : A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 60(309), 63-69.

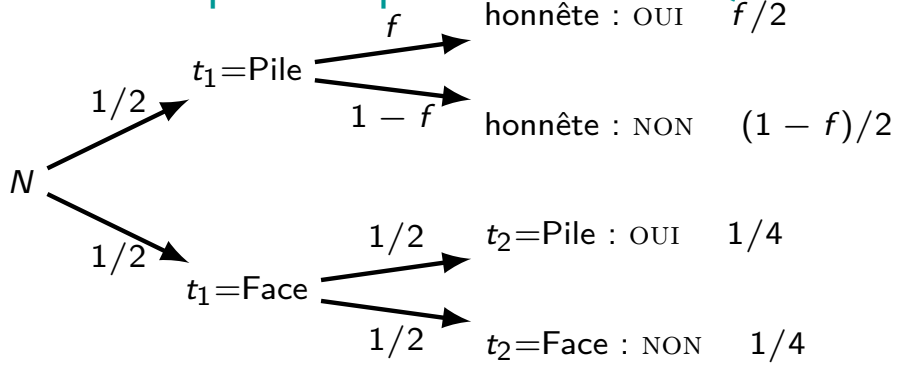
6. https://fr.coursera.org/lecture/stanford-statistics/warners-randomized-response-model-ck65a

Motivation : estimation du pourcentage de tricheurs

Point clef

- Un OUI individuel : on ne connaît pas exactement son origine.
- Après calcul du pourcentage global des OUI : on doit pouvoir estimer le pourcentage des étudiants ayant triché au moins une fois.

Estimateur pour N personnes et $f < N$ tricheurs



- Fréquence observée de OUI : $r \approx 1/4 + f/2$
- Estimation $\hat{f}$ du nbre. original de OUI : $\hat{f} = 2r - 1/2$

x \ y	OUI	NON
OUI	3/4	1/4
NON	1/4	3/4

- $\frac{\Pr[\mathcal{M}(x_1)=y]}{\Pr[\mathcal{M}(x_2)=y]} \leq \frac{\Pr[\mathcal{M}(\text{OUI})=\text{OUI}]}{\Pr[\mathcal{M}(\text{OUI})=\text{OUI}]} \leq 3$
- $\rightsquigarrow$ Mécanisme ln(3)-LDP

LDP sur des données continues : encore Laplace

Intervalle continu d'amplitude Δ : mécanisme Laplacien borné $\mathcal{M}_{Lb}$

- $\mathcal{M}_{Lb}(x) = x + v$ t.q. $v \sim Lap(\frac{\Delta}{\epsilon})$
- si $x + v$ pas dans l'intervalle, réappliquer $\mathcal{M}_{Lb}$

Et pour les attributs catégoriels ?

- Voir section suivante.

Plan

Plan

$\mathcal{M}_{GRR}$: Réponse Randomisée Généralisée [7]

Définition (Mécanisme de réponses randomisées généralisées)

- Domaine : $\{v_1, \ldots, v_k\}$
- $\Pr[\mathcal{M}_{GRR}(x) = v_i] = \begin{cases} p = \dfrac{e^\epsilon}{k - 1 + e^\epsilon} & \text{pour } v_i = x \\ q = \dfrac{1}{k - 1 + e^\epsilon} & \text{sinon} \end{cases}$

Estimation de la fréquence d'apparition de v_i

- N personnes, f_i (resp. r_i) la fréquence initiale de v_i (resp. après application de $\mathcal{M}_{GRR}$ à chaque réponse indiv.)
- Estimateur non biaisé de f_i, $\hat{f}_i = \dfrac{r_i - q}{p - q}$

$$\mathrm{Var}[\hat{f}_i] = \frac{q(1-q)}{N(p-q)^2} + \frac{f_i(1-p-q)}{N(p-q)}$$

7. Kairouz, P., Bonawitz, K., & Ramage, D. (2016). Discrete distribution estimation under local privacy. arXiv preprint arXiv :1602.07387.

$\mathcal{M}_{SUE}$: Encod. Unaire Sym. (RAPPOR basique) [8]

Définition (Mécanisme d'encodage unaire symétrique)

- Domaine : $\{v_1, \ldots, v_k\}$, v_i encodée en $[0, \ldots, 0, 1, 0, \ldots, 0]$
- $\Pr[\mathcal{M}_{SUE}(v) = 1] = \begin{cases} p = \dfrac{e^{\epsilon/2}}{e^{\epsilon/2}+1} & \text{pour } v = 1 \\ q = \dfrac{1}{e^{\epsilon/2}+1} & \text{pour } v = 0 \end{cases}$

Estimation de la fréquence d'apparition de v_i

- N personnes, f_i (resp. r_i) la fréquence initiale de v_i (resp. après application de $\mathcal{M}_{GRR}$ à chaque réponse indiv.)
- Estimateur non biaisé de f_i, $\hat{f}_i = \dfrac{r_i - q}{p - q}$

$$\mathrm{Var}[\hat{f}_i] = \frac{q(1-q)}{N(p-q)^2} + \frac{f_i(1-p-q)}{N(p-q)}$$

8. Erlingsson, Ú., Pihur, V., & Korolova, A. (2014, November). Rappor : Randomized aggregatable privacy-preserving ordinal response. In Proceedings of the 2014 ACM SIGSAC conference on computer and communications security (pp. 1054-1067).

Autres mécanismes avec même estimateur/variance

Références vers ces mécanismes

- $\mathcal{M}_{OUE}$ [9] : $p = \frac{1}{2}$ et $q = \frac{1}{e^{\epsilon}+1}$ choisis pour minimiser la variance
- $\mathcal{M}_{BLH}$ [10] : hash sur g bits, puis $\mathcal{M}_{GRR}$
- ...

Choisir le mécanismes minimisant la dispersion ?

- $\mathrm{Var}[\hat{f}_i] = \frac{q(1-q)}{N(p-q)^2} + \frac{f_i(1-p-q)}{N(p-q)}$ p et q fonctions des mécanismes.
- Remarque (discutable) : variance en $1/N \rightsquigarrow$ intérêt lorsque N est petit
- Hypothèse (discutable) de l'état de l'art : $f_i = 0 \rightsquigarrow \mathrm{Var}*$ (ord. à l'origne)

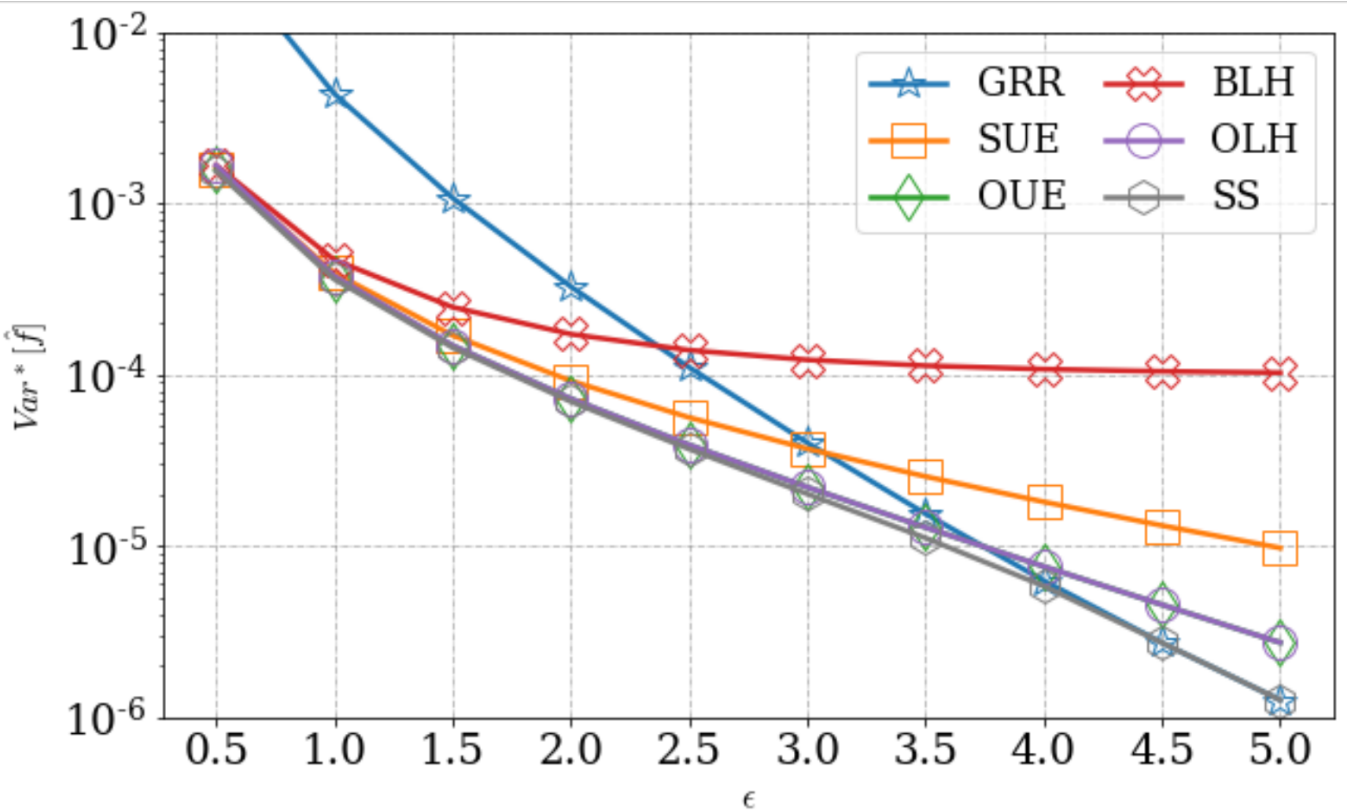
---

9. Wang, T., Blocki, J., Li, N., & Jha, S. (2017). Locally differentially private protocols for frequency estimation. In 26th USENIX Security Symposium (USENIX Security 17) (pp. 729-745).

10. Bassily, R., & Smith, A. (2015, June). Local, private, efficient protocols for succinct histograms. In Proceedings of the forty-seventh annual ACM symposium on Theory of computing (pp. 127-135).

Choisir un mécanisme pour un attribut

Var$*$ théoriques avec $N = 10^5$, $k = 128$



Méthode (discutable)

- Donnés : ϵ et k,
- Choix du mécanisme minimisant Var$*$

Plan

Diviser le budget entre attributs [11]

Assainissment par de d attributs avec un budget de ϵ/d

- Pour chaque attribut : $\mathrm{Var} * (\epsilon/d) \geq \mathrm{Var} * (\epsilon)$
- Méthode communément adoptée pour chaque enregistrement
  1. Sélection aléatoire d'un attribut $j \in \{1, \ldots, d\}$
  2. A l'agrégateur, envoi du résultat $(j, \mathcal{M}(v_j, \epsilon))$

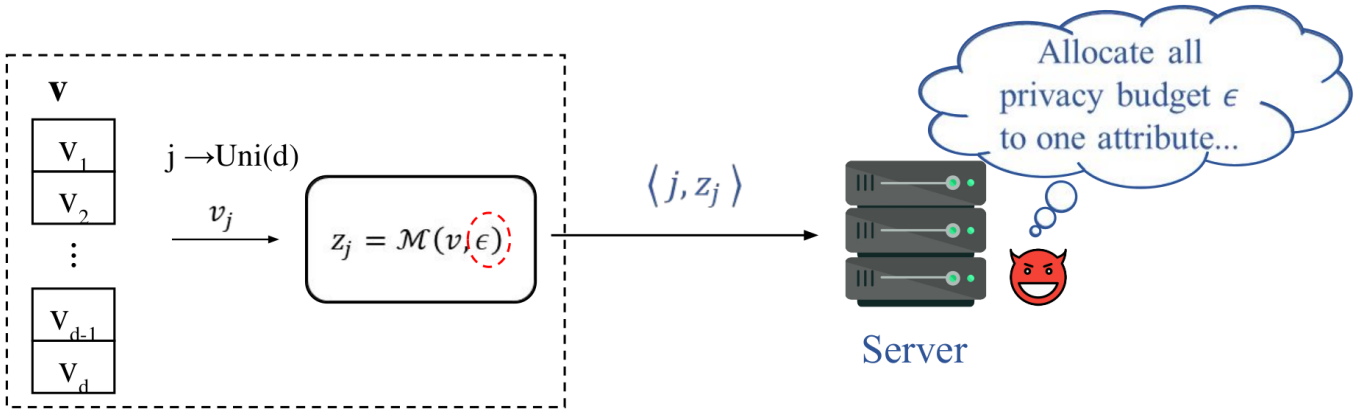
Un résultat problématique ?

- Résultats limités à $\mathrm{Var}*$ ne tenant pas compte de f_i
- Attributs avec différents niveaux de sensibilité $\rightsquigarrow$ équité de la divulgation d'un seul attribut ?

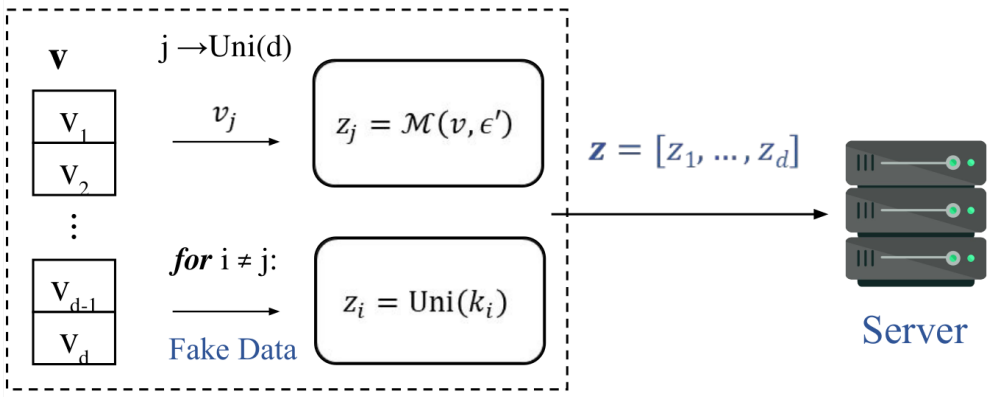
11. Wang, N., Xiao, X., Yang, Y., Zhao, J., Hui, S. C., Shin, H., ... & Yu, G. (2019, April). Collecting and analyzing multidimensional data with local differential privacy. In 2019 IEEE 35th International Conference on Data Engineering (ICDE) (pp. 638-649) ?

Choix inéquitable de l'attribut partagé : une solution

Uniquement choix aléatoire de l'attribut



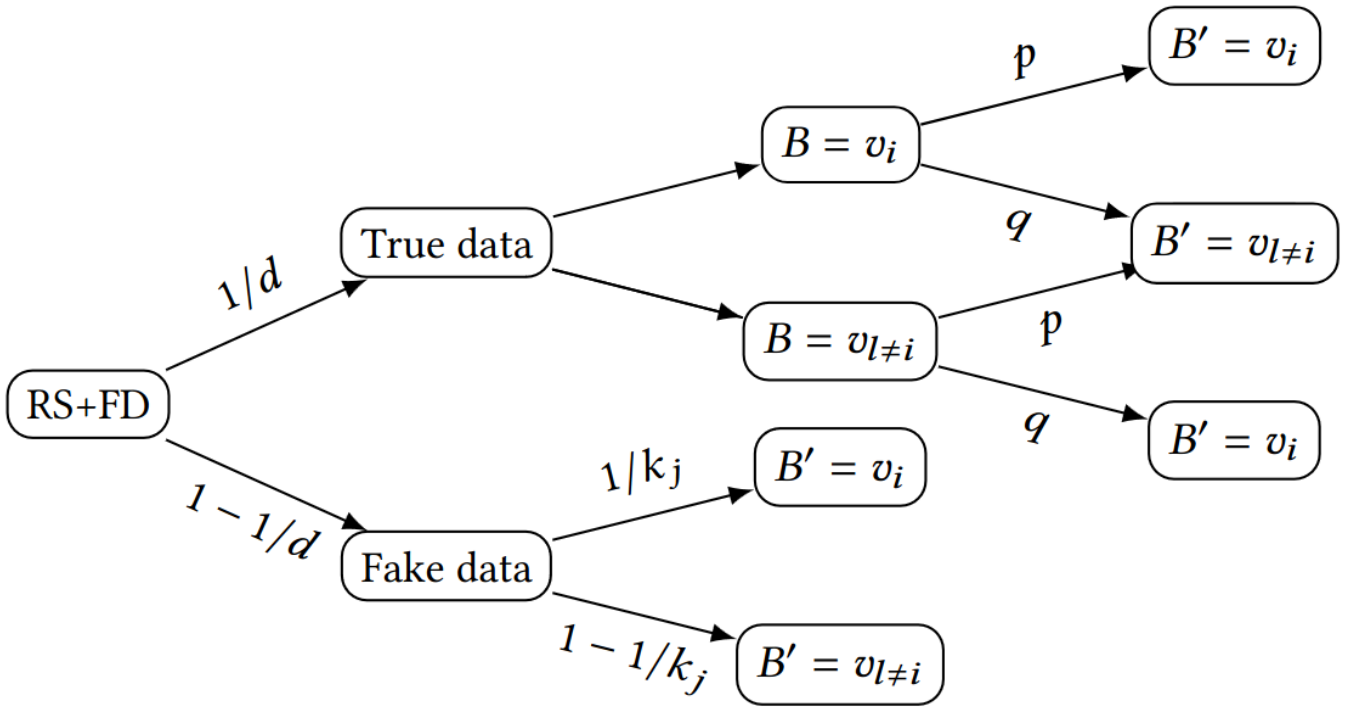
1 attribut choisi aléatoirement, les autres fictifs [12]



12. Arcolezi, H. H., Couchot, J. F., Al Bouna, B., & Xiao, X. (2021, October). Random sampling plus fake data : Multidimensional frequency estimates with local differential privacy. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management (pp. 47-57).

Mise en RS+FD de $\mathcal{M}_{GRR}$

Arbre de probabilités



Estimateur et variance

- Estimateur non biaisé de f_i, $\hat{f}_i = \dfrac{r_i d k_j - (q k_j + d - 1)}{(p-q)k_j}$

$$\mathrm{Var}[\hat{f}_i] = \frac{d^2}{N(p-q)^2}\left(f_i.\delta_1(1-\delta_1) + \frac{1-f_i}{N}\delta_0(1-\delta_0)\right)$$

$$\delta_1 = \frac{pk_j + d - 1}{dk_j} \text{ et } \delta_0 = \frac{qk_j + d - 1}{dk_j}$$

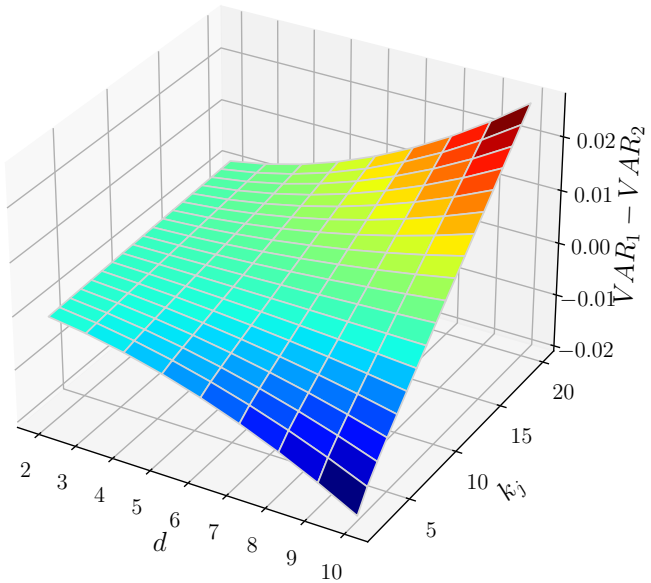
Un mécanisme adaptatif ADP

Connus

- Paramètre p, q de chaque macanisme pour vérifier ϵ-LDP
- Une estimation du nombre de participants N
- d attribut et pour chacun : son nombre de valeurs k
- Pour chaque estimateur : la variance approximative de l'estimateur $\mathrm{Var}*[\hat{f}]$

Pour chaque attribut

- Choix du mécanisme minimisant $\mathrm{Var}*[\hat{f}]$
- Partage de ce choix avec l'agrégateur
- Exemple avec VAR_1 : $RS + FD[GRR]$, $VAR_2 = RS + FD[OUE]$, $N = 10,000$ $\epsilon = \ln(3)$

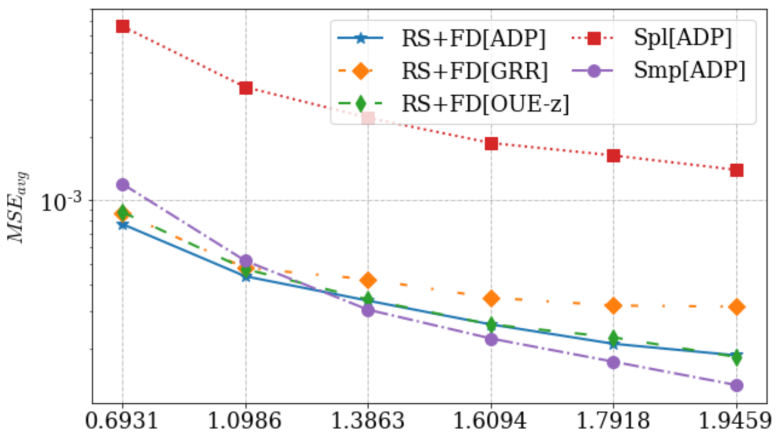


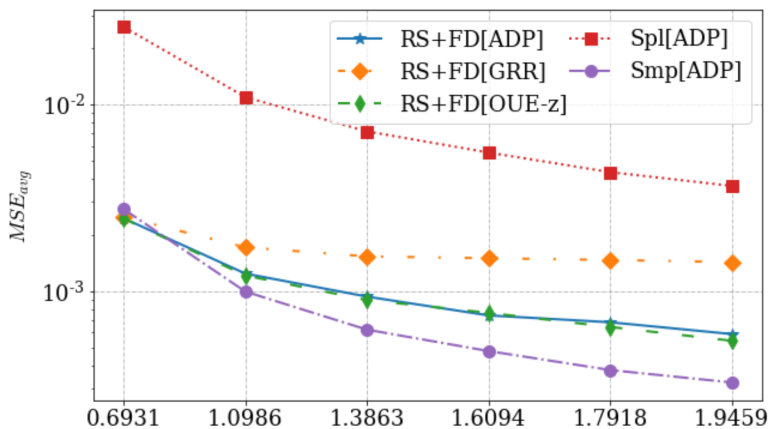
Quelques expérimentations

Caractéristiques

- Données d'UCI : adults, nursery
- $\epsilon \in \{\ln(2), \ln(3), \ldots \ln(7)\}$
- Erreur quadratique moyenne sur 100 essais à chaque fois

Quelques courbes imparfaites (pas de dispersion de l'erreur)





Plan

Retrouver l'attribut non fictif ?

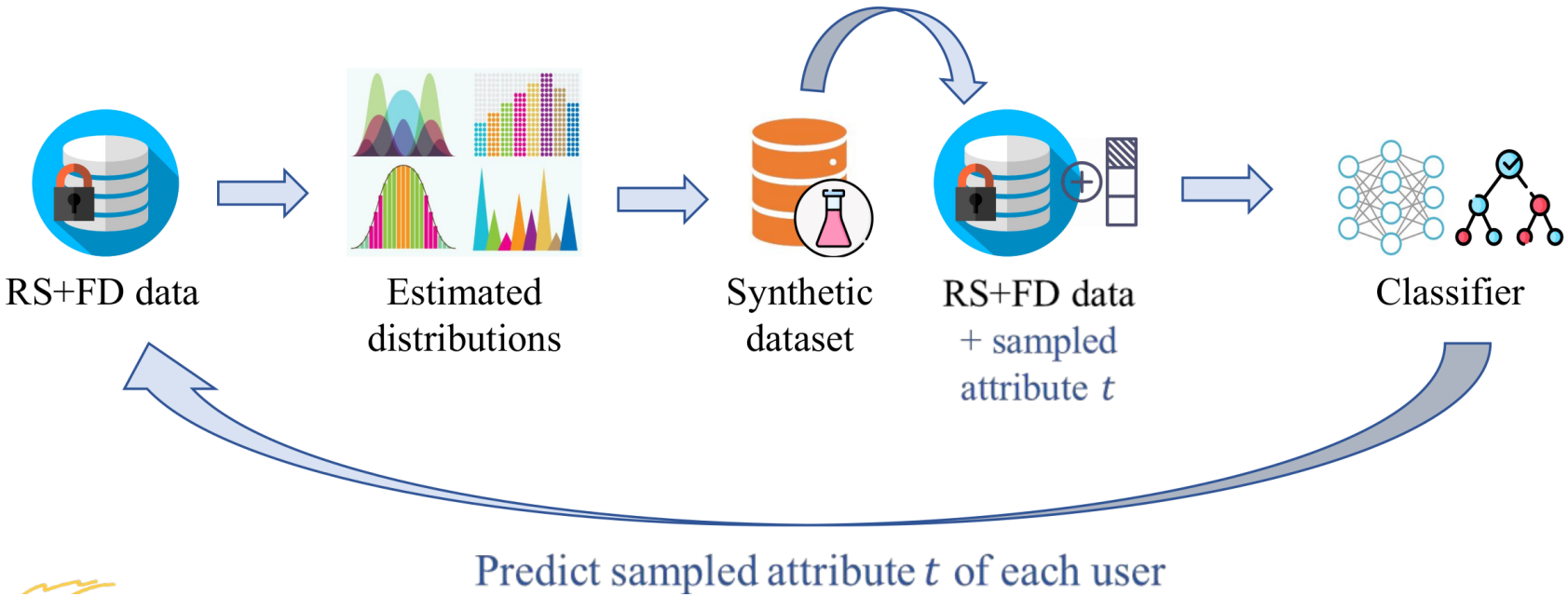
Motivation

- Rappel : valeur de chaque attribut fictif choisie selon une distrib. unif.
- Choix de l'attribut non fictif : inféré par ML ?

Attaque avec connaissances minimales (ϵ)

Connaissances et démarche de l'attaquant

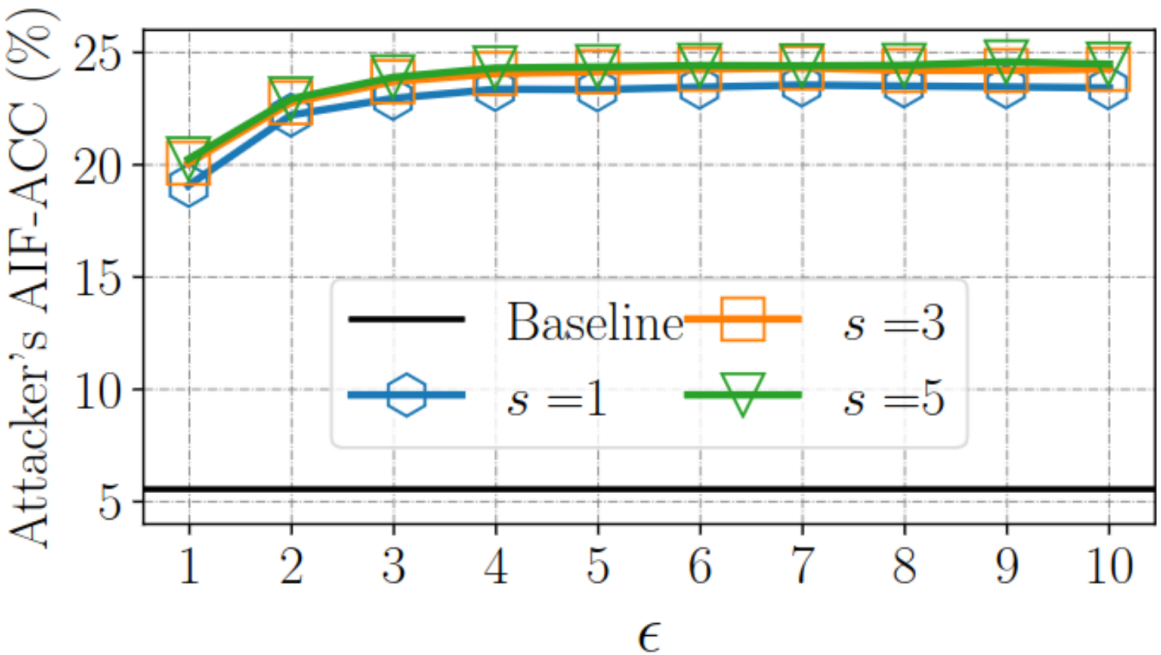
1. Inféré : cardinalité de chaque attribut $\rightsquigarrow$
   - 1.1 Mécanisme LDP pour celui-ci et estimateur associé (théorie)
   - 1.2 Distribution originale des valeurs de chaque attribut
2. Génération possible d'un dataset synthétique
3. Rejouer RS + FD sur ce dataset et mémoriser l'attribut non fictif
4. Classifieur XGBoost : état de l'art pour les données tabulaires



Résultats de l'attaque de RS+FD

Contexte expérimental

- GRR utilisé partout
- Jeu de données : ACSEmployement [13]
- Cardinalité des profiles synthétiques : $1N$, $2N$ ... $5N$
- Nombre d'expérimentations pour chaque mesure : 20

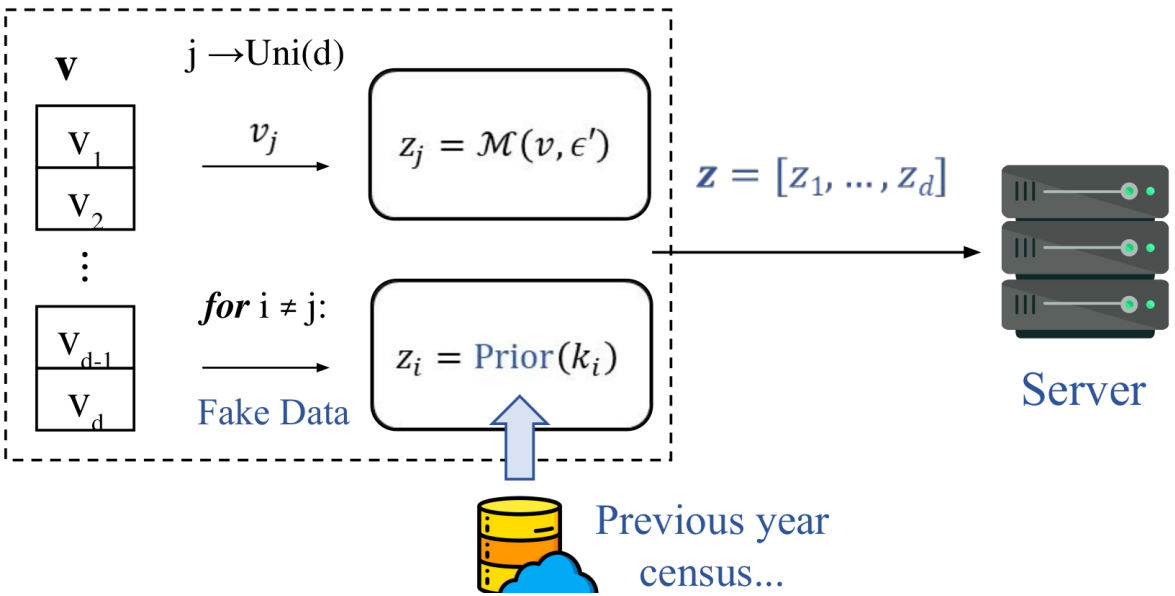


13. F. Ding, M. Hardt, J. Miller, and L. Schmidt. 2021. Retiring adult : New datasets for fair machine learning. Advances in Neural Information Processing Systems 34 (2021)

Une contremesure parfois possible

Post-processing de la LDP

- Autorise l'utilisation a posteriori de données assainies sans consommer le ϵ
- Exploitation de ces connaissances pour générer des données fictives réalistes

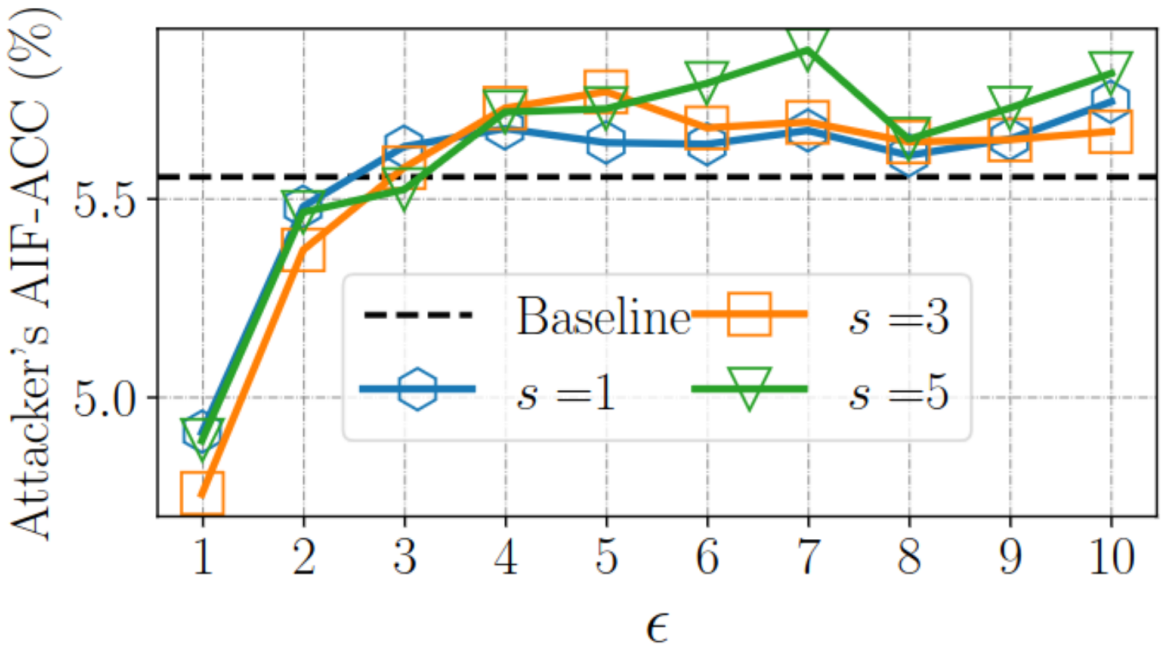


raisses alimentaire

Résultats de l'attaque de RS+ Realistic FD

Contexte expérimental (copie du transparent $n-2$

- GRR utilisé partout
- Jeu de données : ACSEmployement
- Cardinalité des profiles synthétiques : $1N$, $2N$ ...$5N$
- Nombre d'expérimentations pour chaque mesure : 20



Plan

Résumé, perspectives et références

Analyses de données multidimensionnelles respectueuses de la vie privée

- Avancées en LDP, particulièrement concernant l'attribut partagé
- A nouveau : utilité (variance réduite) vs robustesse (précision de réidentification)

Perspectives

- Points soulevés ($\mathrm{Var}*$, gestion des N petits) ...

Pour aller plus loin

- H.H. Arcolezi, S. Gambs, J.-F. Couchot, C. Palamidessi: On the Risks of Collecting Multidimensional Data Under Local Differential Privacy. Proc. VLDB Endow. 16(5): 1126-1139 (2023)
- H.H. Arcolezi, J.-F. Couchot, B. al Bouna, X. Xiao: Random Sampling Plus Fake Data: Multidimensional Frequency Estimates With Local Differential Privacy. CIKM 2021: 47-57

- Site Github d'Héber Hwang Arcolezi : codes de tous les mécanismes, toutes les expérimentations