Rapport de stage de confidentialité différentielle

Clémence Audibert

Juillet-août 2024
Stage encadré par Aurélien Garivier

1 Introduction

Les avancées récentes de l'IA ont fait ressortir l'importance des enjeux de confidentialité des données. Mal les maîtriser a pu salir l'image de certaines entreprises. Ainsi, en 2006, Netflix livrait des ensembles de données "anonymisés" lors d'un concours ouvert aux chercheurs en *machine learning*. En faisant correspondre ces données avec celles d'une autre base, Arvind Narayanan, alors doctorant, retrouva des informations sensibles liées aux utilisateurs de Netflix, provoquant un scandale.

Depuis, plusieurs chercheurs ont travaillé sur le sujet. Cynthia Dwork a posé dans [3] les bases d'un type particulier de confidentialité, la confidentialité différentielle. Selon ses règles, le résultat d'un calcul sur des données doit être bruité, pour être insensible au changement de l'une de ces données.

Après avoir rappelé les définitions de base du domaine, nous présenterons dans un premier temps des simulations numériques [1] de mécanismes de calcul d'espérance et de quantiles, permettant de mieux saisir les enjeux de base de la confidentialité différentielle. Nous nous inspirerons alors du travail de Clément Lalanne ([6]) portant sur les compromis entre précision et confidentialité dans l'apprentissage statistique. Il y fournit un résumé clair des algorithmes existants en calcul de quantiles et de leurs avantages comparés. Puis nous étudierons la convergence de mécanismes exponentiels d'estimation de moyenne avec discrétisation des réponses vers un mécanisme de Laplace, avant d'en donner une petite généralisation. Enfin, nous nous pencherons sur des mécanismes d'*inverse sensitivity*, estimant le paramètre d'une loi de Bernoulli. Nous analyserons leur optimalité quant à une borne inférieure de précision, donnée par le travail très récent d'Asi et Duchi ([2]).

1. Toutes les simulations ont été réalisées en Python sur *Visual Studio Code*.

2 Définitions de base

La plupart des définitions suivantes sont données par Dwork et Roth dans [3].

Un **mécanisme** est un algorithme prenant en entrée
— un ensemble de données X
— l'univers de toutes les données possibles
— des bits aléatoires
— et facultativement, un ensemble de *queries* (questions portant sur X).
Il retourne une chaîne de caractères qu'on espère pouvoir décoder pour
— répondre de manière relativement précise aux *queries*, s'il y en a ;
— trouver des réponse à de futures *queries* dans le cas contraire.
Notations :
— Quand il n'y a pas d'ambiguïté, on note simplement $M(X)$ le résultat du mécanisme M sur l'ensemble de données X.

— Pour X, X' deux ensembles de données, on note $d_{ham}(X, X')$ la distance de Hamming entre X et X'.

— On note $X \sim X'$ quand X et X' sont voisins, c'est-à-dire quand :
  — il existe une permutation σ des indices telle que $d_{ham}(\sigma(X), X') \leq 1$: relation de voisinage de ***substitution*** ;
  — $X = X'$ ou X' peut être obtenu en ajoutant ou supprimant un seul élément de X : relation d'***addition / replacement***.

— On note Δf la *sensitivity* d'une fonction f (par exemple une *query*) :

$$\Delta f = \max_{X \sim X'} |f(X) - f(X')|.$$

Un mécanisme M à valeurs dans $\mathcal{T}$ est dit (ϵ, δ)-*differentially private* lorsque pour tout $\mathcal{S} \subset \mathcal{T}$, pour tous $X \sim X'$,

$$\mathbb{P}[M(X) \in \mathcal{S}] \leq e^{\epsilon}\mathbb{P}[M(X') \in \mathcal{S}] + \delta.$$

Un mécanisme $(\epsilon, 0)$-*differentially private* sera simplement dit ϵ-*differentially private*.

Donnons trois types de mécanismes très classiques :
— Le **mécanisme de Laplace**.
  Définition préliminaire : La **distribution de Laplace** (centrée en 0) **d'échelle** α est la distribution de densité de probabilité

$$g_{\mathcal{L}}(x|\alpha) = \frac{1}{2\alpha} e^{-|x|/\alpha}.$$

  On note $\mathcal{L}(\alpha)$ la loi associée. Sa variance est de $2\alpha^2$.

Soit $\alpha > 0$. Le mécanisme de Laplace estimant f est défini par $M : X \mapsto f(X) + \mathcal{L}(\alpha)$. Comme l'énonce Lalanne dans [6], dès lors que $\alpha \geq \frac{\Delta f}{\epsilon}$, ce mécanisme est ϵ-D.P.

— Soit un ensemble fini $\mathcal{O}$ et une fonction u prenant en entrée un ensemble de données X et un élément o de $\mathcal{O}$. Le **mécanisme exponentiel** $M(X, u, \mathcal{O})$ sélectionne et retourne un élément o avec une probabilité proportionnelle à $e^{\frac{\epsilon u(X,o)}{2\Delta u}}$.
Remarque : La *sensitivity* d'une telle u (associée à un mécanisme exponentiel, et nommée *utility function*) se définit par

$$\Delta u := \max_{o \in \mathcal{O}} \max_{X \sim X'} |u(X', o) - u(X, o)|.$$

D'après le théorème 3.10 de [3], un tel mécanisme est ϵ-*differentially private*.

— Soit $\alpha > 0$. Le **mécanisme gaussien** estimant f est défini par $M : X \mapsto f(X) + \mathcal{N}(0, \alpha)$. Comme l'énonce le théorème 3.22 de [3], pour $c^2 > 2\ln(1.25/\delta)$, ce mécanisme est ϵ-D.P dès lors que $\alpha \geq \left(\frac{c\Delta f}{\epsilon}\right)^2$.

3 Simulations : mécanisme de calcul d'espérance

Dans cette section, on se donne des variables aléatoires $X_1, ..., X_n \overset{i.i.d.}{\sim} \mathbb{P}$, qui représentent nos données. On note θ l'espérance des X_i.
On pose $\hat{\theta}_n := \frac{1}{n}\sum_{i=1}^n X_i$ l'estimateur non *differentially private* de la moyenne des X_i, et $\tilde{\theta}_n := \frac{1}{n}\sum_{i=1}^n X_i + B$ son estimateur D.P., où B est une VA centrée représentant le bruit. Différentes options s'offrent concernant la loi de B.

3.1 Avec un bruit de distribution de Laplace

La première solution est d'utiliser le mécanisme de Laplace, défini dans [3] au paragraphe 3.3.
Supposons $B \sim \alpha\mathcal{L}(1)$. D'après le Fact 2.3.3. de [6], si $\alpha \geq \frac{\Delta f}{\epsilon}$, alors le mécanisme de Laplace est ϵ-*differentially private*. Pour avoir le meilleur compromis confidentialité / précision, nous prenons donc $\alpha = \frac{\Delta f}{\epsilon}$.

Une simulation des estimations obtenues avec un bruit de Laplace est donnée Figure 1.
Jouons ensuite sur les paramètres pour s'assurer que tout est cohérent : en divisant b par 10, le support du pic est lui aussi divisé par 10, et son altitude est décuplée. C'est normal car (en configuration de remplacement) $\alpha = \frac{b-a}{n\epsilon}$, donc ce paramètre d'échelle a aussi été divisé par 10.
Augmenter ϵ fait aussi augmenter α donc le support du bruit augmente.

On réalise ensuite la même simulation avec des X_i gaussiennes, de loi $\mathcal{N}(\frac{a+b}{2}, \frac{a+b}{8})$. Les résultats sont similaires.

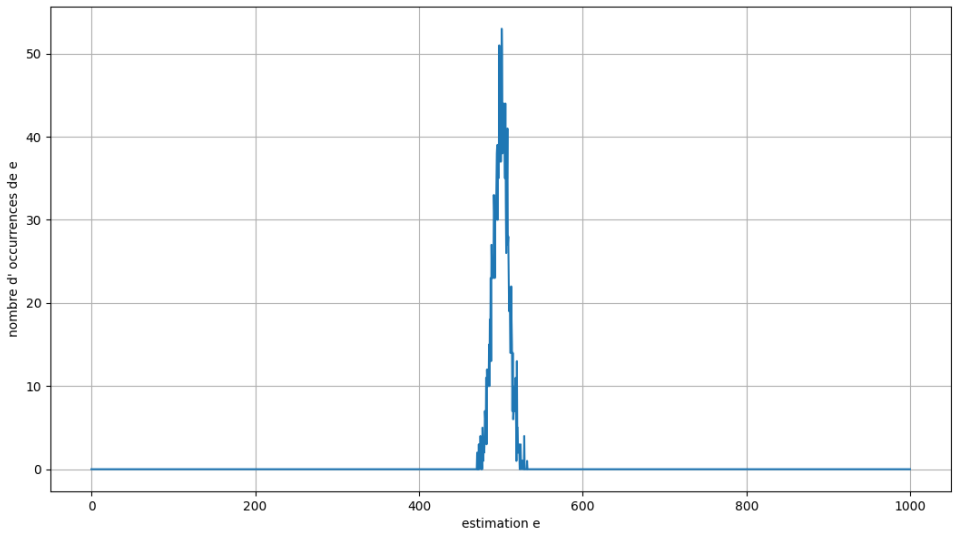

FIGURE 1 – Estimations obtenues pour 1000 VA uniformes sur $[0, 1000]$, avec un bruit de distribution de Laplace, une précision de 0.5, 2000 simulations et $\epsilon = 0.8$.

3.2 Avec un bruit gaussien

Demandons-nous ensuite si on ne pourrait pas avoir un bruit B gaussien. Nous refaisons donc la simulation de 2.1 avec un bruit gaussien, en prenant le c (cf. [3], théorème 3.22) presque minimal (5 au lieu d'un peu plus de 4.8) pour qu'il n'y ait pas d'erreur, et $\delta = \frac{1}{100n}$. Nous obtenons la figure 3. Excepté son pic plus fin et plus haut, elle est semblable à son équivalente avec bruit de Laplace.

Si l'on diminue ensuite ϵ, on obtient un pic beaucoup plus brouillé et élargi que pour la même simulation en Laplace (cf. figure 2) ; c'est-à-dire qu'avec un bruit gaussien, augmenter la confidentialité nuit plus à la précision. C'est pourquoi le bruit de Laplace est davantage utilisé dans cette situation.

3.3 Calcul des risques quadratiques

Posons $\hat{\theta}_n = \overline{X}_n$, la moyenne non bruitée des données.

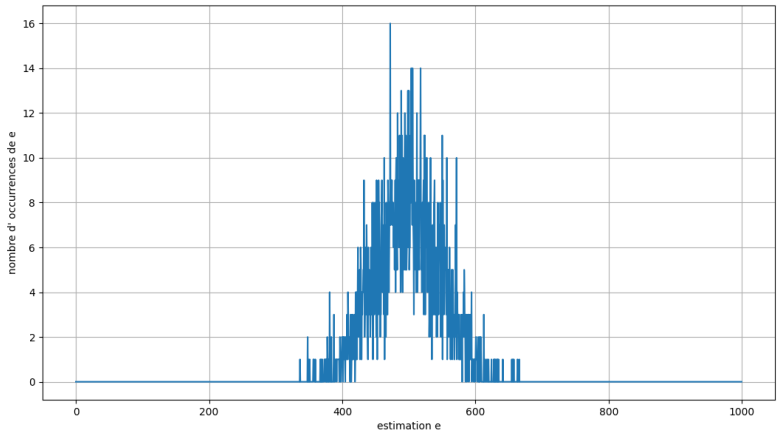

FIGURE 2 – Estimations obtenues pour 1000 VA uniformes sur $[0, 1000]$, avec un bruit gaussien, une précision de 0.5, 2000 simulations, $\epsilon = 0.1$ et $\delta = 0.00001$.

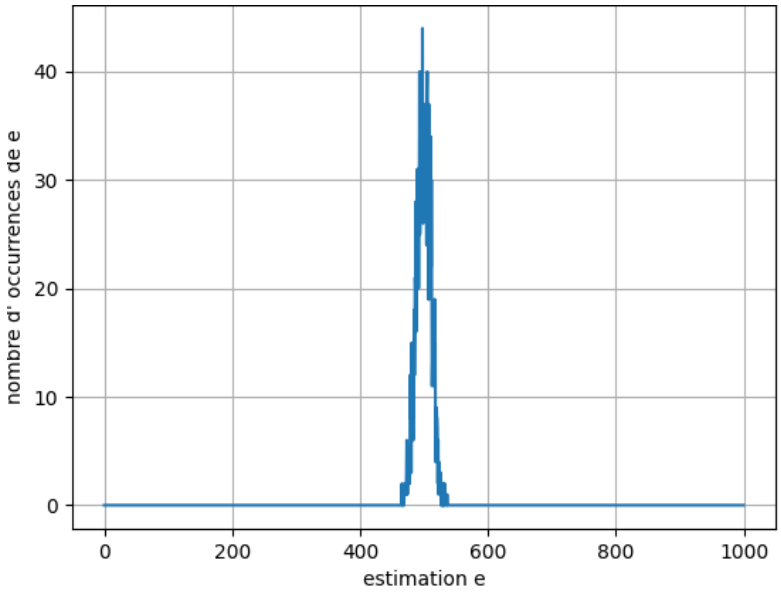

FIGURE 3 – Estimations obtenues pour 1000 VA uniformes sur $[0, 1000]$, avec un bruit gaussien, une précision de 0.5, 2000 simulations, $\epsilon = 0.8$ et $\delta = 0.00001$.

3.3.1 Risque quadratique sans bruit

$$\begin{aligned}\mathbb{E}[(\hat{\theta}_n - \theta)^2] &= \mathbb{E}[\frac{\sum_{i=1}^n X_i}{n} - \theta]^2 + Var[\frac{\sum_{i=1}^n X_i}{n} - \theta] \\ &= 0^2 + \frac{1}{n^2}\sum_{i=1}^n Var[X_i] \text{ (en supposant les } X_i \text{ indépendantes)} \\ &= \frac{Var[X_1]}{n}\end{aligned}$$

3.3.2 Risque quadratique avec bruit de Laplace

Posons $\tilde{\theta}_n = \hat{\theta}_n + \alpha\mathcal{L}(1)$.

$$\begin{aligned}\mathbb{E}[(\tilde{\theta}_n - \theta)^2] &= \mathbb{E}[\frac{\sum_{i=1}^n X_i}{n} + \alpha\mathcal{L}(1) - \theta]^2 + Var[\frac{\sum_{i=1}^n X_i}{n} + \alpha\mathcal{L}(1) - \theta] \\ &= 0^2 + \frac{Var[X_1]}{n} + 2\alpha^2\end{aligned}$$

(en supposant les X_i indépendantes du bruit).

3.3.3 Risque quadratique avec un bruit gaussien

Posons $\tilde{\theta}_n = \hat{\theta}_n + \mathcal{N}(0, \frac{c^2(b-a)^2}{n^2\epsilon^2})$.

$$\begin{aligned}\mathbb{E}[(\tilde{\theta}_n - \theta)^2] &= \mathbb{E}[\frac{\sum_{i=1}^n X_i}{n} + \mathcal{N}(0, \frac{c^2(b-a)^2}{n^2\epsilon^2}) - \theta]^2 + Var[\frac{\sum_{i=1}^n X_i}{n} + \mathcal{N}(0, \frac{c^2(b-a)^2}{n^2\epsilon^2}) - \theta] \\ &= 0^2 + \frac{Var[X_1]}{n} + \frac{c^2(b-a)^2}{n^2\epsilon^2} \\ &= \mathbb{E}[(\tilde{\theta}_{n,Lap} - \theta)^2] + \frac{(c^2-2)(b-a)^2}{n^2\epsilon^2} \\ &\geq \mathbb{E}[(\tilde{\theta}_{n,Lap} - \theta)^2], \text{ ce qui confirme l'utilisation préférentielle du mécanisme de Laplace ici.}\end{aligned}$$

4 Mécanisme de calcul de quantiles

D'après [6], l'estimation d'un seul quantile se faisait au départ par mécanisme de Laplace, mais la forte *sensitivity* de cette *query* rendait le résultat imprécis. On utilise donc actuellement un mécanisme exponentiel, qu'on nommera "QExp" dans la suite.

Pour estimer m quantiles en une seule fois, on a d'abord pensé à utiliser m fois QExp de manière indépendante, mais cette méthode était très sous-optimale. Gillenwater et al. (cf. [8]) ont alors développé le mécanisme "JointExp", où le tirage aléatoire des quantiles s'appuie sur une *utility function*

adaptée au calcul joint de plusieurs quantiles. Enfin, en 2022, Kaplan et al. ont proposé "RecExp", mécanisme utilisant récursivement QExp et d'une précision battant tous les records à ce jour.

4.1 Algorithme QExp

Nous étudions dans cette sous-section l'algorithme 2 fourni par [8].

4.1.1 Présentation

Cet algorithme prend en entrée une liste de données $(X_1, ..., X_n)$, un ordre de quantile p, un degré de confidentialité ϵ et un *bounding parameter* Λ. L'algorithme commence à trier les X_i dans l'ordre croissant et à les borner par 0 et Λ avant de leur ajouter ces deux valeurs. Puis il crée une liste $(Y_0, ..., Y_n)$ telle que pour tout $i \in [\![0, n]\!]$,

$$Y_i = (X_{i+1} - X_i)e^{-\epsilon|i-pn|}.$$

Intuitivement, Y_i est l'écart entre X_i et X_{i+1} diminué d'un facteur d'autant plus grand que i est éloigné de l'indice du quantile recherché.

L'algorithme normalise ensuite Y, puis tire un Y_{i_0} avant de retourner un élément aléatoire de $[X_{i_0}, X_{i_0+1}]$ (ainsi, plus i est proche du quantile recherché, plus la probabilité de retourner un élément de $[X_i, X_{i+1}]$ est élevée).

On remarque que plus ϵ diminue, plus l'écart entre X_i et X_{i+1} augmente donc les probabilités seront *a priori* mieux réparties (car $t \mapsto e^{-\epsilon t}$ variera moins, la valeur absolue de sa dérivée diminuant alors).

4.1.2 Simulations

On trace les résultats d'un nombre N de simulations sur une liste de données aléatoire. Comme on s'y attend et comme on le remarque, la précision des estimations diminue avec ϵ (cf. figures 4, 5 et 6).

On écrit ensuite une fonction de même but, mais prenant en entrée le vecteur X déjà existant et calculant automatiquement le plus petit ϵ tel que la probabilité d'une erreur plus grande que $(b-a)/10$ (largeur d'une colonne de l'histogramme) soit inférieure à 0.01. (On a ici noté $a \approx minX$ et $b \approx maxX$). Dans cette fonction, on utilise la proposition démontrée par Kaplan et al. en 2022 :

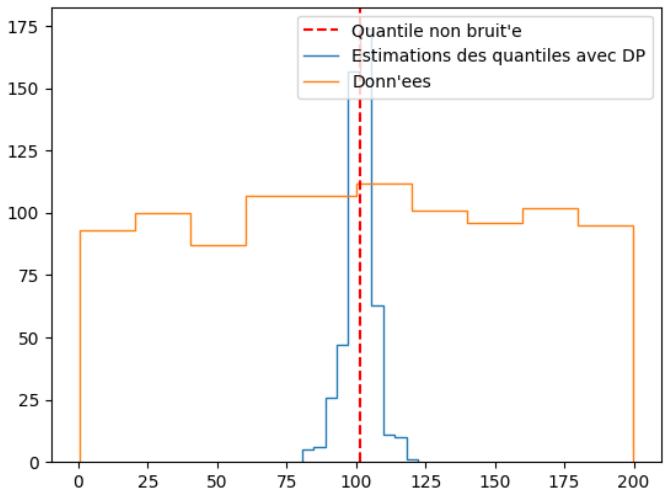

FIGURE 4 – Estimations de la médiane obtenues avec QExp pour 1000 VA uniformes sur $[0, 200]$, avec 500 simulations et $\epsilon = 0.05$.

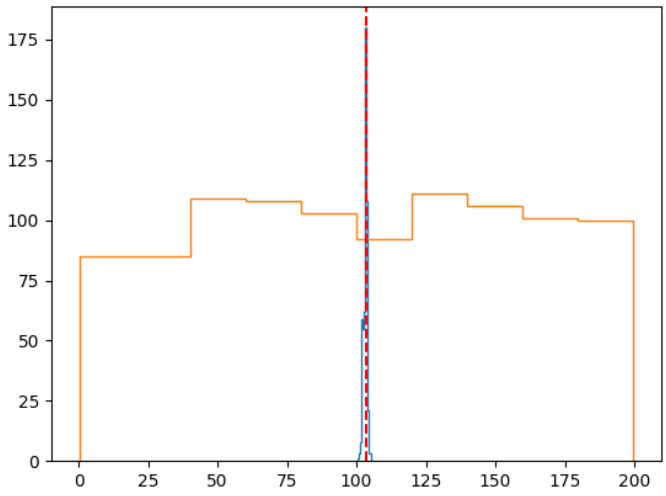

FIGURE 5 – Estimations de la médiane obtenues avec QExp pour 1000 VA uniformes sur $[0, 200]$, avec 500 simulations et $\epsilon = 0.8$. On observe une nette amélioration de la précision par rapport à la figure précédente.

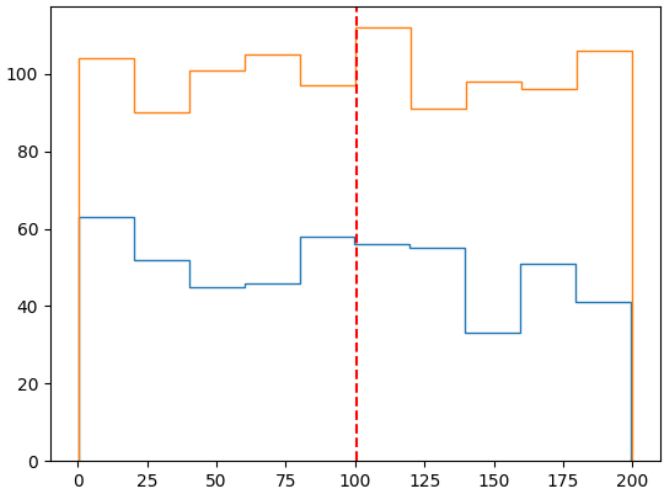

FIGURE 6 – Estimations de la médiane obtenues avec QExp pour 1000 VA uniformes sur $[0, 200]$, avec 500 simulations et $\epsilon = e^{-32}$. On a perdu toute précision : le bruit est devenu trop fort.

Proposition : Soit $X \in (a, b)^n$ et $q \in [0, 1]$ un ordre de quantile. Soit o le vrai quantile d'ordre q. Avec probabilité $1 - \beta$, QExp retourne un v tel que

$$n(o, v) \leq 2\frac{log\psi - log\beta}{\epsilon},$$

où $\psi = \frac{b-a}{min_{k \in [\![1, n+1]\!]}\{x_k - x_{k-1}\}}$.

La figure 7 montre un graphe fourni par cette fonction. On observe un très bon compromis précision / confidentialité.

4.1.3 Comparaison de plusieurs algorithmes de calcul d'un quantile

Cf. figures 8, 10 et 9.

4.2 Algorithme JointExp

Cet algorithme est expliqué en détail dans [4]. Le mécanisme exponentiel employé utilise une astuce polynomiale abaissant la complexité de $O(n^m)$ à $O(mnlog(n) + m^2n)$.

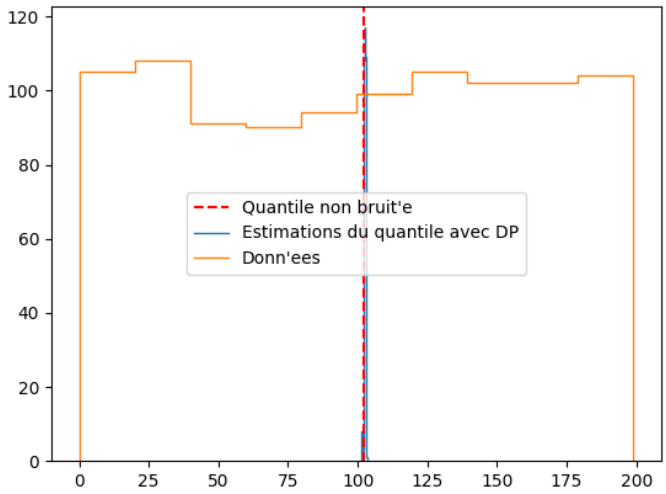

FIGURE 7 – Estimations obtenues de la médiane par la fonction calculant le meilleur compromis pour ϵ (ici, environ 0,99). On a fourni en entrée 1000 VA uniformes sur $[0, 200]$ et utilisé 500 simulations.

4.3 Algorithme RecExp

Cet algorithme fonctionne récursivement en utilisant à chaque appel l'algorithme QExp. Pour tout appel, il commence par considérer l'ordre $p_{\hat{m}}$ de quantile d'indice $\hat{m} \approx m/2$;

- il appelle d'abord QExp sur les données initiales et $p_{\hat{m}}$, il obtient alors v ;
- puis il effectue un appel récursif avec les données $X_i < v$ et les ordres p_i pour $i < \hat{m}$;
- il agit symétriquement pour les données $X_i > v$ et les ordres d'indice $i > \hat{m}$;
- enfin, il retourne la concaténation de ces trois résultats.

Le degré de *privacy* en entrée de chaque appel à QExp vaut $\frac{\epsilon}{log_2(m)+1}$, où ϵ est fourni en entrée de RecExp. Or le nombre d'appels est de l'ordre de $log_2(m)$. Par théorème de composition (cf. Théorème 3.16 de [3] : un mécanisme égal à un k-uplet de mécanismes ϵ_j-D.P. est $\sum_{j=1}^{k} \epsilon_j$-D.P.), on obtient bien un algorithme ϵ-*differentially private*.

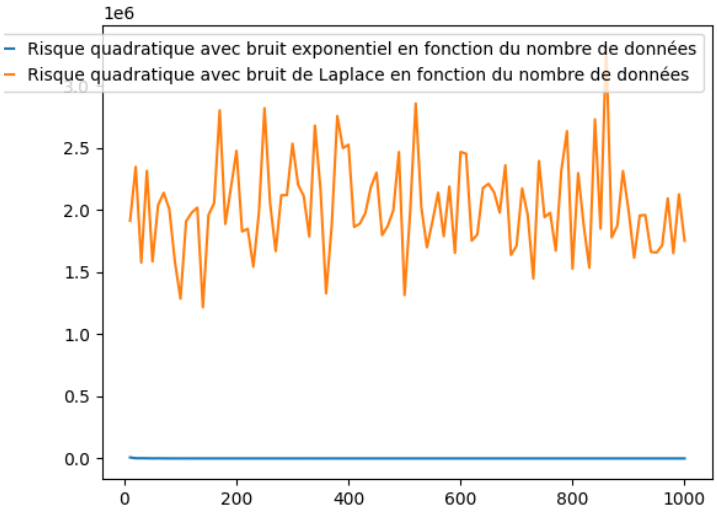

FIGURE 8 – Risques quadratiques d'algorithmes de calcul d'un seul quantile, de bruits respectivement gaussien et de Laplace, en fonction de n, le nombre de données. On a généré, pour chaque valeur de n, n valeurs de V.A. uniformes sur $[0, 200)$, et utilisé pour chaque valeur de n 150 simulations. On a considéré le premier quartile et pris $\epsilon = 0.2$. Remarque : Les valeurs de la courbe correspondant à QExp sont de l'ordre de 10^2.

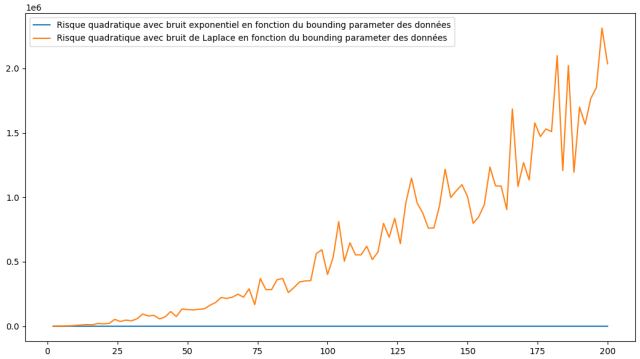

FIGURE 9 – Risques quadratiques d'algorithmes de calcul d'un seul quantile, de bruits respectivement gaussien et de Laplace, en fonction de Λ, *bounding parameter* des données. On a généré, pour chaque valeur de Λ, 1000 valeurs de V.A. uniformes sur $[0, \Lambda)$, et utilisé pour chaque valeur de Λ 150 simulations. On a considéré le premier quartile et pris $\epsilon = 0.2$. Remarque : Les valeurs de la courbe correspondant à QExp sont de l'ordre de 90.

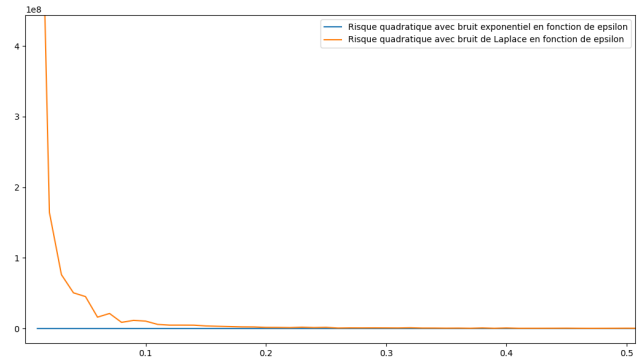

FIGURE 10 – Risques quadratiques d'algorithmes de calcul d'un seul quantile en fonction de ϵ. On a généré, pour chaque valeur de ϵ, 10 valeurs de V.A. uniformes sur $[0, 200)$, et utilisé pour chaque valeur de ϵ 150 simulations. On a considéré le premier quartile.

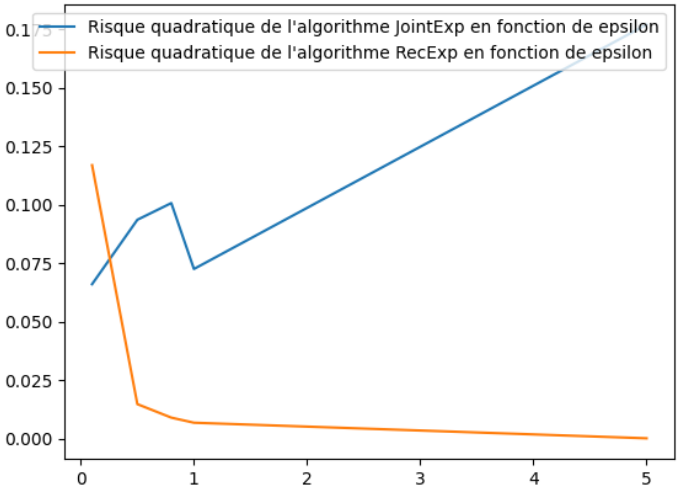

FIGURE 11 – Risques quadratiques de JointExp et RecExp en fonction de ϵ. Pour chaque valeur de ϵ, on a généré 100 valeurs de V.A. uniformes sur $[0, 1)$ et utilisé 50 simulations. Les quantiles utilisés ici sont les trois quartiles.

4.4 Comparaison des algorithmes de calcul de plusieurs quantiles

4.4.1 Comparaison des risques quadratiques

Voir les figures 11 et 12.

Remarque : La 12 semble contredire les figures p. 8 de [7], mais ce n'est pas le cas. Certes la courbe correspondant à RecExp est située au-dessus de celle d'IndExp, mais il s'agit ici du risque quadratique $\mathbb{E}\left[\|\text{estimateurs - quantiles théoriques}\|_2^2\right]$. L'article [7] étudie quant à lui $\mathbb{E}\left[\|\text{estimateurs - quantiles théoriques}\|_\infty\right]$ et pour X suivant une loi Bêta. En traçant la même chose (cf. 13), on obtient bien une courbe RecExp en-dessous de la courbe IndExp.

4.4.2 Comparaison des temps d'exécution

Les tracés des figures 14, 15, 16 et 17 sont cohérents avec la figure 8 p. 8 de [5].

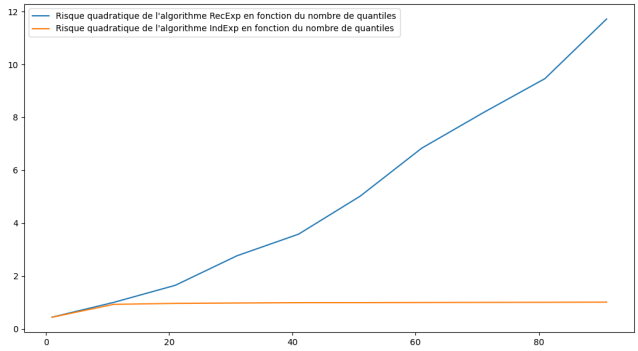

FIGURE 12 – Risques quadratiques de IndExp et RecExp en fonction de m. On a généré (une fois pour toutes) 100 valeurs de V.A. uniformes sur $[0, 1)$, et utilisé pour chaque valeur de m 50 simulations. m va ici de 10 en 10, et on considère à chaque fois les $m - 1$ quantiles d'ordre m.

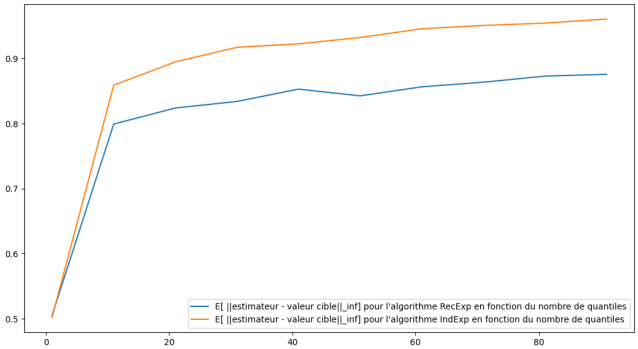

FIGURE 13 – Risques $\mathbb{E}\Big[\|\text{estimateurs - quantiles théoriques}\|_\infty\Big]$ de IndExp et RecExp en fonction de m. On a généré (une fois pour toutes) 100 valeurs de V.A. de loi Bêta(2,2) sur $[0, 1)$, et utilisé pour chaque valeur de m 50 simulations. m va ici de 10 en 10, et on considère à chaque fois les $m - 1$ quantiles d'ordre m.

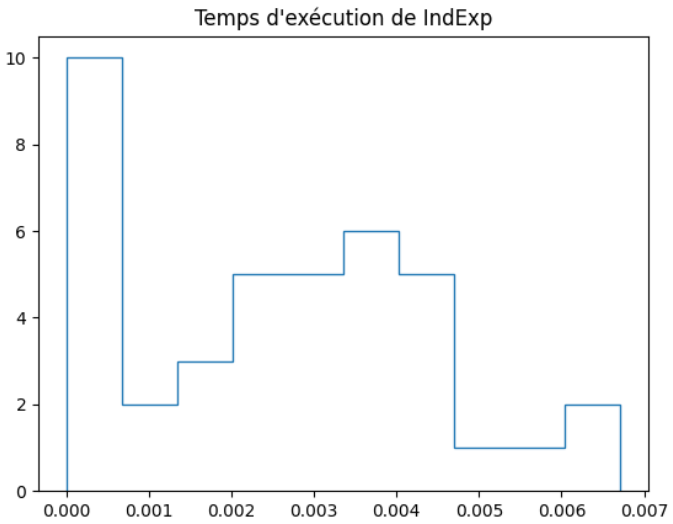

FIGURE 14 – Histogramme des temps d'exécution obtenus lors de 40 simulations de IndExp(X, $p = [0.2, 0.4, 0.6, 0.8]$, $\epsilon = 1$, $b = 1$), où X est constitué de 100 V.A. uniformes entre 0.01 et 0.99.

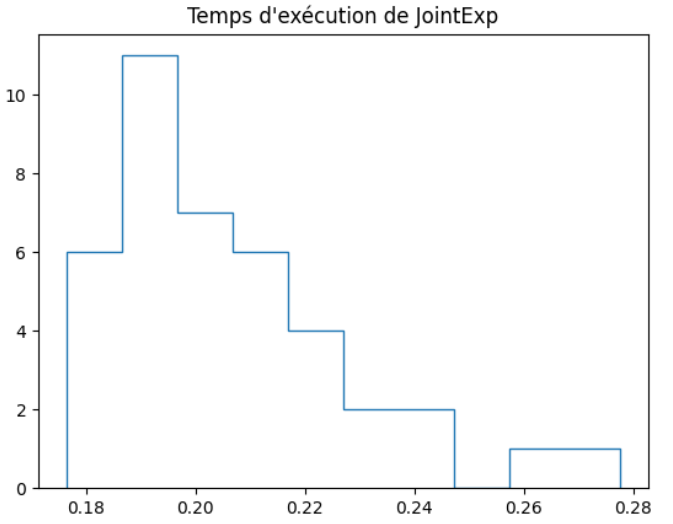

FIGURE 15 – Histogramme des temps d'exécution obtenus lors de 40 simulations de JointExp(X, $p = [0.2, 0.4, 0.6, 0.8]$, $\epsilon = 1$, $a = 0$, $b = 1$), où X est constitué de 100 V.A. uniformes entre 0.01 et 0.99.

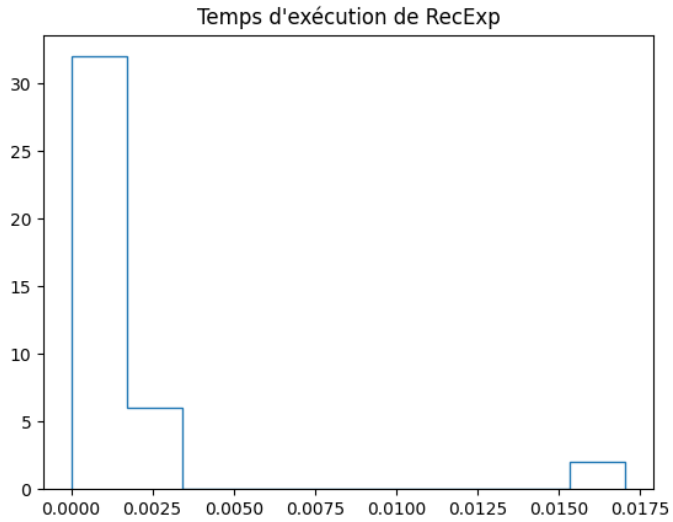

FIGURE 16 – Histogramme des temps d'exécution obtenus lors de 40 simulations de RecExp(X, $p = [0.2, 0.4, 0.6, 0.8]$, $\epsilon = 1$, $a = 0$, $b = 1$), où X est constitué de 100 V.A. uniformes entre 0.01 et 0.99.

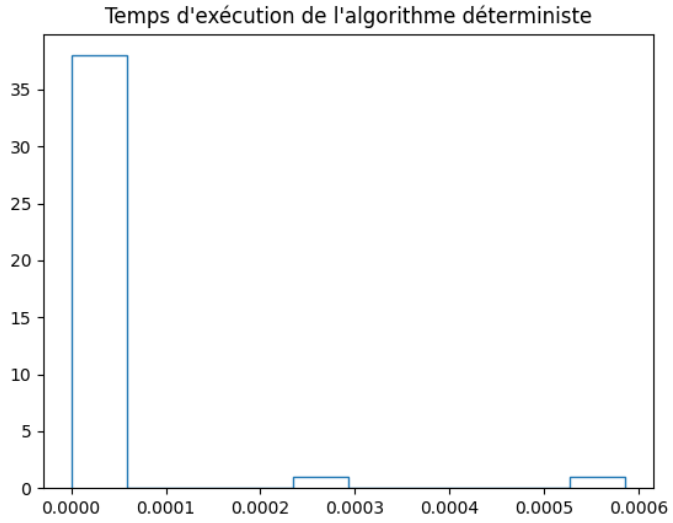

FIGURE 17 – Histogramme des temps d'exécution obtenus lors de 40 simulations d'un algorithme non D.P. de calcul de quantiles, avec les paramètres $p = [0.2, 0.4, 0.6, 0.8]$ et X constitué de 100 V.A. uniformes entre 0.01 et 0.99.

5 Convergence de mécanismes

5.1 Estimation de moyenne par mécanisme exponentiel avec discrétisation de l'ensemble des réponses

Considérons X, ensemble de n V.A. uniformes sur $[0,1]$, dont on veut calculer la moyenne par mécanisme exponentiel. Ce mécanisme étant à valeurs dans un ensemble fini [2], nous allons discrétiser $[0,1]$ en N sous-intervalles puis faire tendre N vers l'infini, et regarder s'il existe un mécanisme limite.

Prenons la fonction d'utilité $u : (X, o) \mapsto -|o - \frac{1}{2}|$. Ainsi, le mécanisme retournera o avec probabilité $p(o) \propto e^{\frac{\epsilon u(X,o)}{2\Delta u}}$.Or :

$$\begin{aligned}
\Delta u &= \sup_{X \sim X' o \in O} \sup \underbrace{||o - \mathbb{E}[X]| - |o - \mathbb{E}[X']||}_{\leq |\mathbb{E}[X] - \mathbb{E}[X']|} \\
&= \sup_{X \sim X'} |\mathbb{E}[X] - \mathbb{E}[X']| \\
&= \begin{cases} \frac{1}{n} & \text{pour une } \textit{neighboring relation} \text{ de substitution} \\ 1 & \text{pour une } \textit{relation} \text{ d'}\textit{addition / replacement} \end{cases}
\end{aligned}$$

En traçant les estimations obtenues en fonction de N pour un même ensemble de données X, on obtient la figure 18. L'algorithme semble converger assez rapidement vers 0.5, et cette vitesse augmente (sans surprise) avec ϵ : en comparant la Figure 19, où $\epsilon = 3$, à la figure 20 ($\epsilon = 0.1$), on voit une nette différence.

5.1.1 Preuve de la convergence en loi

En traçant les probabilités liées à ce mécanisme exponentiel (Figure 22), et en les comparant à un histogramme tracé selon la loi de Laplace (Figure 23), on observe que les courbes de notre mécanisme se rapprochent de plus en plus de cette distribution de Laplace. Montrons en fait la convergence en loi du mécanisme vers un mécanisme de Laplace d'un degré ϵ' que nous exprimerons en fonction de ϵ.

On note Y_N la réponse de l'algorithme pour N intervalles de discrétisation, et Y la V.A. de loi $\frac{1}{2} + \frac{1}{n\epsilon'}\mathcal{L}(1)$ (pour un certain ϵ'), bornée à $[0,1]$ (et donc renormalisée de la manière adéquate). Montrons qu'il existe une valeur de ϵ' tel que $Y_N \xrightarrow{\mathcal{L}} Y$.
Utilisons pour cela le théorème de Lévy (nous donnons les détails des calculs en appendice).

2. Cf. [3] p. 38 : sur un grand ensemble, ce mécanisme, produisant une distribution complexe dessus, serait difficilement implémentable.

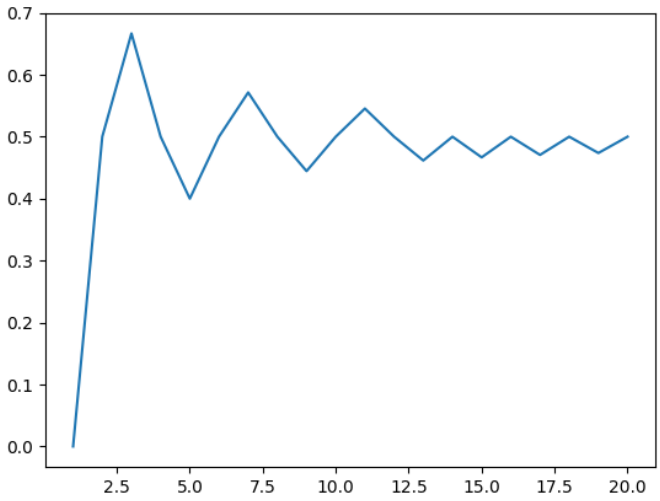

FIGURE 18 – Estimations obtenues par mécanisme exponentiel de la moyenne de 1000 VA uniformes sur $[0, 1]$, en fonction de N, nombre d'intervalles de discrétisation de l'ensemble des réponses. On a pris ici $\epsilon = 0.8$.

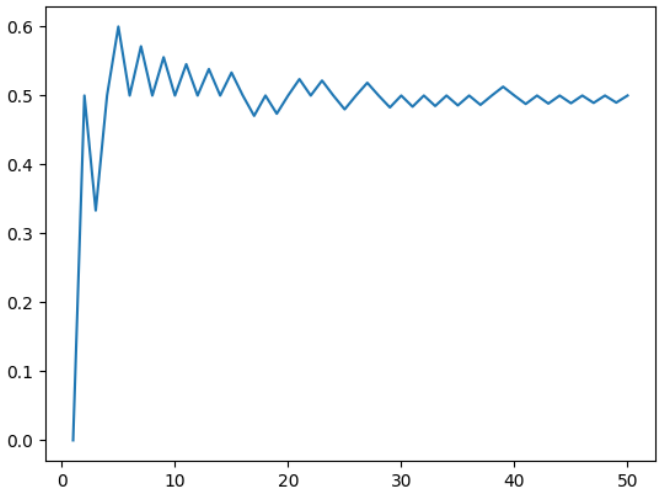

FIGURE 19 – Estimations obtenues par mécanisme exponentiel de la moyenne de 1000 VA uniformes sur $[0, 1]$, en fonction de N, nombre d'intervalles de discrétisation de l'ensemble des réponses. On a pris ici $\epsilon = 3$.

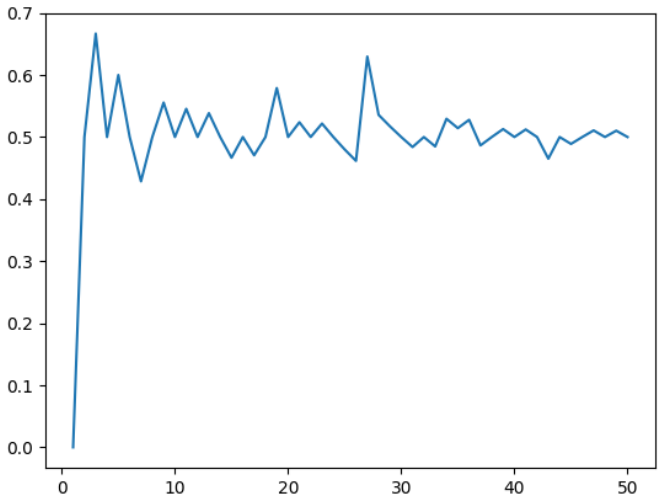

FIGURE 20 – Estimations obtenues par mécanisme exponentiel de la moyenne de 1000 VA uniformes sur $[0,1]$, en fonction de N, nombre d'intervalles de discrétisation de l'ensemble des réponses. On a pris ici $\epsilon = 0.1$.

$$\Phi_{Y_N}(t) = \frac{1}{\mathcal{CN}(N)} \sum_{k=0}^{N-1} e^{it\frac{k}{N}} e^{-\frac{n\epsilon}{2}\left|\frac{k}{N}-\frac{1}{2}\right|},$$

où $\mathcal{CN}(N)$ est le coefficient de normalisation de la loi de probabilité de Y_N. Etudions séparément les cas N pair et impair.

— Après calculs (cf. appendice), on montre :

$$\Phi_{Y_{2L+1}}(t) \underset{L\to+\infty}{\to} \frac{n\epsilon}{4(1-e^{-n\epsilon/4})}\left[\frac{e^{it/2}-e^{-n\epsilon/4}}{it+\frac{n\epsilon}{2}} + \frac{e^{it-\frac{n\epsilon}{4}}-e^{it/2}}{it-\frac{n\epsilon}{2}}\right]$$

— De même pour $\Phi_{Y_{2L}}(t)$: après calculs, on retrouve

$$\Phi_{Y_{2L}}(t) \underset{L\to+\infty}{\to} \frac{n\epsilon}{4(1-e^{-n\epsilon/4})}\left[\frac{e^{it/2}-e^{-n\epsilon/4}}{it+\frac{n\epsilon}{2}} + \frac{e^{it-\frac{n\epsilon}{4}}-e^{it/2}}{it-\frac{n\epsilon}{2}}\right]$$

— Ensuite, on montre :

$$\Phi_Y(t) = \frac{n\epsilon'}{2(1-e^{-n\epsilon'/2})}\left[\frac{e^{it/2}-e^{-n\epsilon'/2}}{it+n\epsilon'} + \frac{e^{it-\frac{n\epsilon'}{2}}-e^{it/2}}{it-n\epsilon'}\right]$$

Ainsi, pour $\epsilon' = \frac{\epsilon}{2}$:

$$\Phi_Y(t) = \lim_{N\to+\infty} \Phi_{Y_N}(t)$$

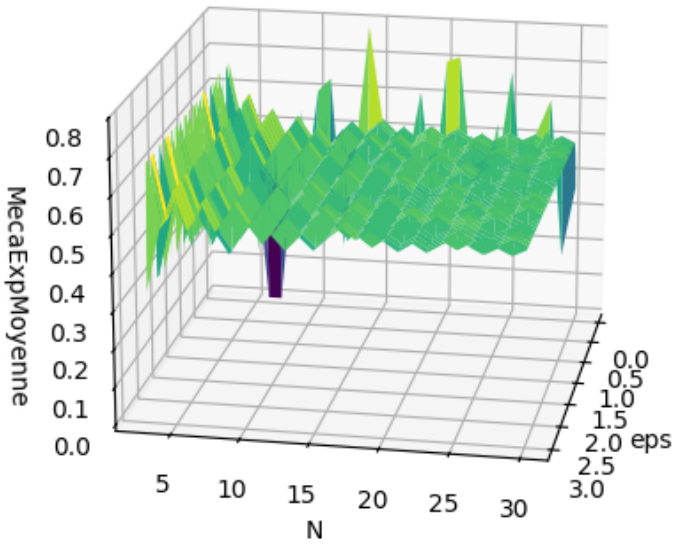

FIGURE 21 – Estimations obtenues par mécanisme exponentiel de la moyenne de 1000 VA uniformes sur $[0, 1]$, en fonction de N, nombre d'intervalles de discrétisation, et de ϵ. On observe bien une convergence plus rapide pour les grands ϵ.

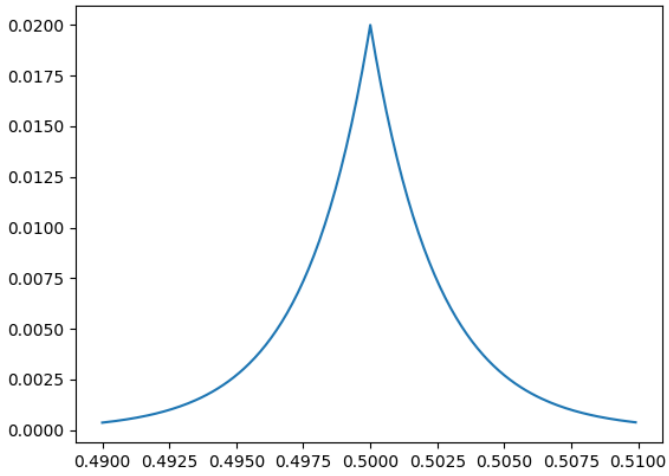

FIGURE 22 – Probabilités (discrètes, malgré l'impression donnée par la figure) pour notre mécanisme exponentiel, avec $n = 1000$ VA uniformes sur $[0, 1]$, et $\epsilon = 0.8$.

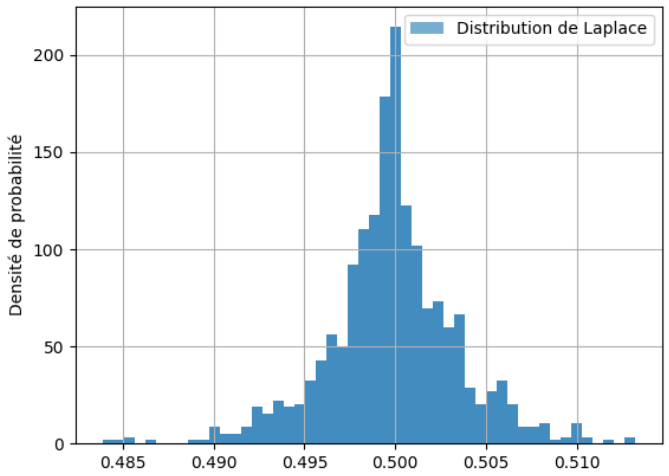

FIGURE 23 – Distribution de probabilité de $\mathcal{L}(\frac{1}{2}, \frac{1}{n\epsilon'})$, avec $n = 1000$ et $\epsilon' = 0.4$. (Il s'agit ici d'un histogramme, la distribution de probabilité serait bien entendu normalisée.)

5.1.2 Non-convergence en probabilité

Remarquons qu'il n'y a en revanche pas convergence en probabilité. Les détails sont donnés en appendice.

5.2 Généralisation

Soit $\mathcal{M}$ un mécanisme de degré de *privacy* ϵ', de densité de probabilité f continue et à valeurs dans $[0,1]$. Montrons que le mécanisme exponentiel de domaine de réponse $\{\frac{k}{N} | k \in [\![0, N-1]\!]\}$ et d'*utility function* vérifiant $u(X,y) = \ln(f(X,y))$ tend vers $\mathcal{M}$ quand N tend vers $+\infty$.
Posons à cette fin Y à valeurs dans $[0,1]$ de densité f, et, pour tout N, Y_N également à valeurs dans $[0,1]$ et vérifiant $\mathbb{P}(Y_N = \frac{k}{N}) \propto e^{\frac{\epsilon \ln(f(X,\frac{k}{N}))}{2\Delta u}}$.

Montrons que $\boxed{Y_N \xrightarrow{\mathcal{L}} Y \text{ pour } \epsilon' = \frac{\epsilon}{2}}$.

— Commençons par calculer Δu.

$$\Delta u = \sup_{X \sim X'} \sup_{k \in [\![0,N-1]\!]} | \underbrace{\ln(\frac{f(X,\frac{k}{N})}{f(X',\frac{k}{N})})}_{privacy\ loss \text{ de Y}} | = \epsilon'.$$

— De plus :

$$\Phi_Y(t) = \int_0^1 e^{ity} f(X,y) dy.$$

— Et

$$\Phi_{Y_N}(t) = \sum_{k=0}^{N-1} e^{\frac{itk}{N} + \frac{\epsilon}{2\Delta u} \ln(f(X,\frac{k}{N}))} \frac{1}{\mathcal{CN}(N)},$$

où $\mathcal{CN}(N)$ est le coefficient de normalisation de la loi de probabilité de Y_N. Il vérifie :

$$\mathcal{CN}(N) = \sum_{k=0}^{N-1} e^{\ln(f(X,\frac{k}{N}))} = \sum_{k=0}^{N-1} f(X,\frac{k}{N}).$$

Ainsi

$$\Phi_{Y_N}(t) = \frac{\sum_{k=0}^{N-1} \frac{1}{N} e^{\frac{itk}{N}} f(X,\frac{k}{N})}{\sum_{k=0}^{N-1} \frac{1}{N} f(X,\frac{k}{N})}.$$

— Or on a pris f continue. Donc en utilisant les sommes de Riemann, on obtient :

$$\begin{cases} \sum_{k=0}^{N-1} \frac{1}{N} e^{\frac{itk}{N}} f(X,\frac{k}{N}) \underset{N \to +\infty}{\longrightarrow} \int_0^1 e^{ity} f(X,y) dy \\ \sum_{k=0}^{N-1} \frac{1}{N} f(X,\frac{k}{N}) \underset{N \to +\infty}{\longrightarrow} \int_0^1 f(X,y) dy = 1. \end{cases}$$

Ainsi,

$$\Phi_{Y_N}(t) \underset{N \to +\infty}{\longrightarrow} \Phi_Y(t)$$

D'où, par le théorème de Lévy :

$$Y_N \xrightarrow{\mathcal{L}} Y$$

pour $\epsilon' = \frac{\epsilon}{2}$.

6 Bornes inférieures de précision

6.1 Optimalité du mécanisme d'*inverse sensitivity*

Dans toute la suite, nous nous appuyons sur l'article [2] d'Asi et Duchi. Les démonstrations seront données en appendice. Analysons la borne inférieure qu'Asi et Duchi proposent dans le théorème 1 ; pour ce faire, nous nous pencherons sur le cas de n-uplets ($n \geq 1$) $X^{(n)} = (X_1, ..., X_n)$ de variables suivant une loi de Bernoulli. Un n-uplet $X^{(n)}$ sera parfois noté plus simplement X, et son paramètre sera pris dans $]0, 1[$. $X^{(n)}$ et $X'^{(n)}$ n'auront pas forcément le même paramètre de Bernoulli [3]. Les mécanismes utilisés chercheront à estimer la moyenne $f(X^n)$ d'un ensemble de données (ce qui donnera un ordre d'idées de leur paramètre de Bernoulli) ; cette moyenne (des X_i) sera donc notre *query*. Un mécanisme principal étudié ici sera celui d'*inverse senstivity*.

Définition : Soit un ensemble de données X. On appelle ***inverse sensitivity*** de cible $\frac{k}{n}$ la quantité $|nf(X) - k|$. Intuitivement, cela correspond au nombre minimal de données à changer dans X pour atteindre une moyenne de $\frac{k}{n}$.

Définition : On appelle **mécanisme d'*inverse sensitivity*** le mécanisme exponentiel $M_{inv,n}$ de fonction d'utilité égale à l'opposé de l'*inverse sensitivity*.

Remarques : 1. Quand il n'y a pas d'ambiguïté, ce mécanisme sera plus simplement noté M_{inv}.

2. Pour un degré de *privacy* ϵ, la loi de probabilité de M_{inv} vérifie

$$\mathbb{P}\Big(M_{inv}(X) = \frac{k}{n}\Big) = \frac{e^{-\epsilon|nf(X)-k|/2}}{\sum_{j=0}^{n} e^{-\epsilon|nf(X)-j|/2}}.$$

En effet, la *sensitivity* de sa fonction d'utilité vaut

$$\max_{j\in[\![0,n]\!]} \max_{X\sim X'} \underbrace{\big||nf(X) - j| - |nf(X') - j|\big|}_{\leq |nf(X)-nf(X')|} = 1.$$

Après calcul, on trouve que le coefficient de normalisation est égal à

$$\mathcal{CN}(X) = \frac{e^{-n\epsilon f(X)/2} - e^{\epsilon/2} - 1 + e^{n\epsilon(f(X)-1)/2}}{1 - e^{\epsilon/2}}.$$

3. Ceci impliquera que le mécanisme limite (pour $n \to +\infty$) ne sera jamais *D.P.*. Le raisonnement sera alors bien cohérent, puisque ce mécanisme limite sera un mécanisme déterministe.

Définition : Un mécanisme M à valeurs dans $\mathcal{T}$ et approximant une *query* g sera dit ***unbiased*** si pour tout ensemble de données Z et pour tout $t \in \mathcal{T}$,

$$\mathbb{E}[|M(Z) - g(Z)|] \leq \mathbb{E}[|M(Z) - t|].$$

$\mathbb{E}[|M(Z) - g(Z)|]$ sera appelé l'***expected loss*** en Z.

Proposition : M_{inv} est *unbiased*.

Nous abordons maintenant l'adaptation à notre cas du Théorème 1 de [2].

Théorème : Pour tout mécanisme ϵ-D.P. et *unbiased* M d'estimation de f, pour tout X de taille n,

$$\mathbb{E}[|M(X) - f(X)|] \geq \frac{1}{2n(e^{2\epsilon} + 1)}$$

Remarquons que M_{inv} est ϵ-*differentially private*. La *differential privacy* vient en effet de la remarque en partie 2, puisque M_{inv} est exponentiel. On a de plus vu ci-dessus que M_{inv} était bien *unbiased*.
M_{inv} vérifie donc le théorème précédent.

Définition : Un mécanisme M sera dit **optimal** s'il vérifie le théorème précédent et si son *expected loss* est en $O(\frac{1}{n})$.
Ainsi, un autre mécanisme aura une erreur nécessairement supérieure.

Proposition :

$$\mathbb{E}[|M_{inv}(X^{(n)}) - f(X^{(n)})|] \underset{n \longrightarrow +\infty}{\sim} \frac{1}{2n(e^{2\epsilon} + 1)} \times \frac{4(e^{2\epsilon} + 1)}{1 - e^{-\epsilon}}$$

Sur la simulation dont le résultat est donné figure 24, on voit que, quand $n \longrightarrow +\infty$, l'*expected loss* se rapproche à la fois de l'équivalent donné et de la borne inférieure du théorème, sans jamais la minorer.
M_{inv} a une précision de l'ordre de $\frac{1}{n}$.

Par le théorème précédent, M_{inv} **est donc optimal.**

6.2 Optimalité du mécanisme limite

Faisons tendre n vers $+\infty$. Nous considérons une suite d'ensembles de données $(X^{(n)})_n$, où toutes les données $X_i^{(n)}$ ont le même paramètre de Bernoulli. De la même manière qu'en partie 5, lorsque $n \longrightarrow +\infty$, notre mécanisme exponentiel admet un mécanisme limite M_l. Cela se voit bien intuitivement : la distribution de probabilité de $M_{inv,n}$ suit une double exponentielle centrée en $f(X)$, et ce pic s'affine et croît de plus en plus. On verra qu'il tend vers un Dirac

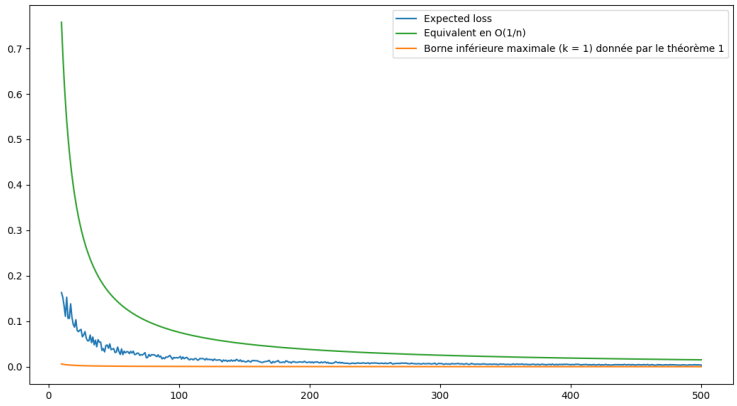

FIGURE 24 – Borne inférieure donnée par le théorème et équivalent donné par la proposition, avec l'*expected loss* de M_{inv} en fonction de n. On a pris ici un paramètre de Bernoulli égal à 0.5, $\epsilon = 1$ et utilisé 100 échantillons pour chaque n lors du calcul de l'*expected loss*.

en $f(X^\infty)$ [4]. $M_{inv,n}$ étant optimal pour tout n, M_l l'est aussi (à ϵ-*differential privacy* près, comme on le verra).
Détaillons cela plus formellement (nous démontrons les résultats en appendice). Soit $Y_n := M_{inv,n}(X^{(n)})$ et X^∞ la limite d'une suite $X^{(n)}$ d'ensembles de données de taille n.

Proposition : $M_{inv,n}$ tend quand n tend vers $+\infty$ vers le mécanisme constant M_l centré en $f(X^\infty)$.

Etudions maintenant l'optimalité de M_l.
En tant que limite de mécanismes optimaux, M_l est *a priori* optimal. On peut d'abord le vérifier facilement.

— Il n'existe plus d'ϵ pour lequel ce mécanisme limite est ϵ-*differentially private*. En effet, si l'on prend X^∞ limite d'une suite $X^{(n)}$ où les X_i suivent une loi $\mathcal{B}\left(\frac{1}{4}\right)$, et X'^∞ limite analogue avec des $X'_i \sim \mathcal{B}\left(\frac{3}{4}\right)$, on aura, par loi des grands nombres :

$$\mathbb{P}\big(M_l(X^\infty) = \frac{1}{4}\big) = 1; \mathbb{P}\big(M_l(X'^\infty) = \frac{1}{4}\big) = 0.$$

Il n'existe donc aucun ϵ tel que $\mathbb{P}\big(M_l(X^\infty) = \frac{1}{4}\big) \leq e^\epsilon \mathbb{P}\big(M_l(X'^\infty) = \frac{1}{4}\big)$. (C'est concrètement cohérent, puisque M_l est déterministe.)

4. On note ainsi la limite de la moyenne de $X^{(n)}$. Il s'agit *a priori* du paramètre de la loi de Bernoulli suivie par les X_i.

— En revanche, on vérifie facilement que M_l est *unbiased*.
— Lorsque $n \to +\infty$, la borne inférieure du théorème en 6.1 revient à

$$"\mathbb{E}[|M(X) - f(X)|] \geq 0",$$

ce qui est trivialement vérifié par M_l.
A ϵ-*differential privacy* près, M_l est donc bien optimal.

Remarque : Cette situation est analogue à celle étudiée en 5, où nous avions une suite de mécanismes exponentiels qui convergeaient vers un mécanisme de Laplace. En effet, un Dirac correspond à un Laplace de variance nulle.

7 Conclusion

La *differential privacy* assure ainsi des sorties d'algorithmes (moyennes, quantiles, etc.) confidentielles en y ajoutant du bruit (Laplace ou gaussien par exemple) ou en déterminant l'*output* grâce à une loi de probabilité (exponentielle, classiquement). Certains chercheurs ont de plus travaillé sur les meilleurs compromis entre confidentialité et précision (cf. thèse de Clément Lalanne, [6]). Travailler sur des ensembles finis de sortie de mécanisme est souvent plus pratique. Comme nous l'avons montré en 5 (dans le cas particulier d'estimation de moyenne de variables uniformes sur [0, 1]), travailler ainsi permet bien d'approcher un mécanisme ayant un ensemble infini de réponses, en conservant de plus la *differential privacy*.
Enfin, pour juger de l'efficacité d'un mécanisme, on le compare à une borne théorique de précision, comme celle donnée dans [2] et que nous avons étudiée en 6. Dans notre cas particulier (estimation du paramètre d'une loi de Bernoulli), les mécanismes d'*inverse sensitivity* respectaient cette borne de façon optimale. De plus, augmenter le nombre de données n'influençait pas cette propriété (même si, asymptotiquement, on perdait la *differential privacy*).

Toutes ces propriétés rendent les mécanismes *differentially private* maniables et sûrs, sans trop s'affranchir de précision toutefois.

A Appendice

A.1 Preuves des résultats de convergence de mécanismes (section 5)

A.1.1 Preuve de la convergence des mécanismes exponentiels de calcul de moyenne (cf. 5.1.1)

Démonstration. — Pour calculer $\Phi_{Y_{2L+1}}(t)$, déterminons d'abord le coefficient de normalisation de la loi de probabilité de Y_{2L+1}.

$$\mathcal{CN}(2L+1) = \sum_{k=0}^{L} e^{\frac{n\epsilon}{2(2L+1)}k} \times e^{-n\epsilon/4} + e^{n\epsilon/4} \sum_{k=L+1}^{2L} e^{-\frac{n\epsilon}{2(2L+1)}k}$$

Posons $r := e^{-n\epsilon/4}$ et $q := e^{\frac{n\epsilon}{2(2L+1)}}$.

$$\begin{aligned}
\mathcal{CN}(2L+1) &= r\sum_{k=0}^{L} q^k + \frac{1}{r}\sum_{k'=L+1}^{2L}\frac{1}{q^{k'}} \\
&= r\sum_{k=0}^{L} q^k + \frac{1}{r}\frac{1}{q^{2L}}\sum_{k=0}^{L-1} q^k \\
&= \sum_{k=0}^{L-1} q^k\left(r + \frac{1}{rq^{2L}}\right) + rq^L \\
&= \frac{1-q^L}{1-q}\frac{r^2q^{2L+1}+q}{rq^{2L+1}} + rq^L \text{ or } r^2q^{2L+1} = 1 \\
&= \frac{1-q^L}{1-q}\frac{1+q}{1/r} + rq^L \\
&= r\frac{1+q-q^L-q^{L+1}+q^L-q^{L+1}}{1-q} \\
&= r\frac{1+q-2q^{L+1}}{1-q}
\end{aligned}$$

Ainsi,

$$\mathcal{CN}(2L+1) = e^{-n\epsilon/4}\frac{1+e^{\frac{n\epsilon}{2(2L+1)}} - 2e^{\frac{n\epsilon(L+1)}{2(2L+1)}}}{1-e^{\frac{n\epsilon}{2(2L+1)}}}$$

Donc

$$\mathcal{CN}(2L+1) \sim \frac{4(1-e^{-n\epsilon/4})(2L+1)}{n\epsilon}$$

— Avec ceci, nous pouvons calculer $\Phi_{Y_{2L+1}}(t)$:

$$\begin{aligned}\Phi_{Y_{2L+1}}(t) &= \frac{1}{\mathcal{CN}(2L+1)} \sum_{k=0}^{2L} e^{it\frac{k}{2L+1}} e^{-\frac{n\epsilon}{2}\left|\frac{k}{2L+1}-\frac{1}{2}\right|} \\ &= \frac{1}{\mathcal{CN}(2L+1)} \Bigg[\sum_{k=0}^{L} e^{\frac{k}{2L+1}\left(it+\frac{n\epsilon}{2}\right)} e^{-n\epsilon/4} \\ &+ e^{n\epsilon/4} \sum_{k=L+1}^{2L} e^{\frac{k}{2L+1}\left(it-\frac{n\epsilon}{2}\right)} \Bigg] \\ &= \frac{1}{\mathcal{CN}(2L+1)} \Bigg[\underbrace{e^{-n\epsilon/4} \frac{1-e^{\frac{2it+n\epsilon}{2(2L+1)}(L+1)}}{1-e^{\frac{2it+n\epsilon}{2(2L+1)}}}}_{:=(1)} \\ &+ \underbrace{e^{n\epsilon/4} \frac{e^{\frac{2it-n\epsilon}{2(2L+1)}(L+1)} - e^{it-\frac{n\epsilon}{2}}}{1-e^{\frac{2it-n\epsilon}{2(2L+1)}}}}_{:=(2)} \Bigg]\end{aligned}$$

Or

$$(1) \sim e^{-n\epsilon/4} \frac{1-e^{\frac{2it+n\epsilon}{4}}}{-\frac{2it+n\epsilon}{2(2L+1)}} = \frac{e^{it/2}-e^{-n\epsilon/4}}{2it+n\epsilon} \times 2(2L+1)$$

Donc

$$\frac{(1)}{\mathcal{CN}(2L+1)} \sim e^{-n\epsilon/4} \frac{1-e^{\frac{2it+n\epsilon}{4}}}{-\frac{2it+n\epsilon}{2(2L+1)}} = \frac{e^{it/2}-e^{-n\epsilon/4}}{2it+n\epsilon} \times \frac{2n\epsilon}{4(1-e^{-n\epsilon/4})}$$

De plus

$$(2) \sim \frac{e^{it-\frac{n\epsilon}{4}}-e^{it/2}}{2it-n\epsilon} \times 2(2L+1)$$

Donc

$$\frac{(2)}{\mathcal{CN}(2L+1)} \sim \frac{e^{it-\frac{n\epsilon}{4}}-e^{it/2}}{2it-n\epsilon} \times \frac{2n\epsilon}{4(1-e^{-n\epsilon/4})}$$

Ainsi :

$$\Phi_{Y_{2L+1}}(t) \underset{L\to+\infty}{\to} \frac{n\epsilon}{4(1-e^{-n\epsilon/4})} \left[\frac{e^{it/2}-e^{-n\epsilon/4}}{it+\frac{n\epsilon}{2}} + \frac{e^{it-\frac{n\epsilon}{4}}-e^{it/2}}{it-\frac{n\epsilon}{2}}\right]$$

— De même pour $\Phi_{Y_{2L}}(t)$: on calcule

$$\begin{aligned}\mathcal{CN}(2L) &= e^{-n\epsilon/4}\sum_{k=0}^{L-1} e^{\frac{n\epsilon}{4L}k} + e^{n\epsilon/4}\sum_{k=L}^{2L-1} e^{\frac{-n\epsilon}{4L}k} \\ &= \frac{-1+e^{-n\epsilon/4}}{1-e^{\frac{n\epsilon}{4L}}} + \frac{1-e^{-n\epsilon/4}}{1-e^{\frac{-n\epsilon}{4L}}} \\ &= (1-e^{-n\epsilon/4})\frac{1-e^{\frac{n\epsilon}{2L}}}{-\left(1-e^{\frac{n\epsilon}{4L}}\right)^2} \\ &= (1-e^{-n\epsilon/4})[1+\frac{2e^{\frac{n\epsilon}{4L}}}{(1-e^{\frac{n\epsilon}{4L}})^2}] \\ &\sim \frac{8L(1-e^{-n\epsilon/4})}{n\epsilon}\end{aligned}$$

Puis

$$\begin{aligned}\Phi_{Y_{2L}}(t) &= \frac{1}{\mathcal{CN}(2L)}[\sum_{k=0}^{L-1} e^{it\frac{k}{2L}-\frac{n\epsilon}{2}(\frac{1}{2}-\frac{k}{2L})} + \sum_{k=L}^{2L-1} e^{it\frac{k}{2L}-\frac{n\epsilon}{2}(\frac{k}{2L}-\frac{1}{2})}] \\ &= \frac{1}{\mathcal{CN}(2L)}[e^{-n\epsilon/4}\sum_{k=0}^{L-1} e^{\frac{2it+n\epsilon}{4L}k} + e^{n\epsilon/4}\sum_{k=L}^{2L-1} e^{\frac{2it-n\epsilon}{4L}k}] \\ &= \frac{1}{\mathcal{CN}(2L)}[e^{-n\epsilon/4}\frac{1-e^{\frac{2it+n\epsilon}{4}}}{1-e^{\frac{2it+n\epsilon}{4L}}} + e^{n\epsilon/4}e^{\frac{2it-n\epsilon}{4}}\frac{1-e^{\frac{2it-n\epsilon}{4}}}{1-e^{\frac{2it-n\epsilon}{4L}}}]\end{aligned}$$

Finalement, on retrouve :

$$\Phi_{Y_{2L}}(t) \underset{L\to+\infty}{\to} \frac{n\epsilon}{4(1-e^{-n\epsilon/4})}[\frac{e^{it/2}-e^{-n\epsilon/4}}{it+\frac{n\epsilon}{2}} + \frac{e^{it-\frac{n\epsilon}{4}}-e^{it/2}}{it-\frac{n\epsilon}{2}}]$$

— Déterminons ensuite $\Phi_Y(t)$. Calculons d'abord la densité f_Y de Y (de support $[0,1]$ et donc renormalisée) :

$$f_Y(y) = \frac{n\epsilon'}{2}\frac{e^{-n\epsilon'\left|y-\frac{1}{2}\right|}}{\int_0^1 \frac{n\epsilon'}{2}e^{-n\epsilon'\left|y-\frac{1}{2}\right|}dy}\mathbb{1}_{[0,1]}(y)$$

or

$$\begin{aligned}\int_0^1 \frac{n\epsilon'}{2}e^{-n\epsilon'\left|y-\frac{1}{2}\right|}dy &= \frac{n\epsilon'}{2}\int_0^{1/2} e^{-n\epsilon'\left(\frac{1}{2}-y\right)}dy + \frac{n\epsilon'}{2}\int_{1/2}^1 e^{-n\epsilon'\left(y-\frac{1}{2}\right)}dy \\ &= \frac{1-e^{-n\epsilon'/2}}{2} + \frac{1-e^{-n\epsilon'/2}}{2} \\ &= 1-e^{-n\epsilon'/2}\end{aligned}$$

Donc

$$f_Y(y) = \frac{n\epsilon'}{2} \frac{e^{-n\epsilon'\left|y-\frac{1}{2}\right|}}{1-e^{-n\epsilon'/2}} \mathbb{1}_{[0,1]}(y)$$

D'où

$$\begin{aligned}
\Phi_Y(t) &= \int_0^1 \frac{n\epsilon'}{2(1-e^{-n\epsilon'/2})} e^{ity - n\epsilon'\left|y-\frac{1}{2}\right|} dy \\
&= \frac{n\epsilon'}{2(1-e^{-n\epsilon'/2})} \left[\int_0^{1/2} e^{-n\epsilon'/2} e^{(it+n\epsilon')y} dy + \int_{1/2}^1 e^{n\epsilon'/2} e^{(it-n\epsilon')y} dy \right] \\
&= \frac{n\epsilon'}{2(1-e^{-n\epsilon'/2})} \left[\frac{e^{it/2} - e^{-n\epsilon'/2}}{it+n\epsilon'} + \frac{e^{it-\frac{n\epsilon'}{2}} - e^{it/2}}{it-n\epsilon'} \right]
\end{aligned}$$

□

A.1.2 Preuve de la non-convergence en probabilité (cf. 5.1.2)

Démonstration. Il y a convergence en probabilité ssi $\forall \eta > 0, \mathbb{P}(|Y_N - Y| \geq \eta) \underset{N\to+\infty}{\longrightarrow} 0$.

Soit donc $\eta > 0$. Pour tout N,

$$\mathbb{P}(|Y_N - Y| \geq \eta) = \mathbb{P}(Y_N - Y \geq \eta) + \mathbb{P}(Y - Y_N \geq \eta)$$

Or

$$\mathbb{P}(Y_N - Y \geq \eta) = \sum_{k=0}^{N-1} \mathbb{P}(Y \leq \frac{k}{N} - \eta) e^{-\frac{n\epsilon}{2}|\frac{1}{2} - \frac{k}{N}|} \frac{1}{\mathcal{CN}(N)} \text{ (probabilités totales)}$$

Après calcul, on trouve finalement :

$$\mathbb{P}(Y_N - Y \geq \eta) \underset{\substack{N\to+\infty \\ N impair}}{\longrightarrow} \frac{1}{8} + \frac{e^{-n\epsilon(1+2\eta)/4} + e^{-n\epsilon\eta/2}}{4} > 0$$

Or $\mathbb{P}(Y - Y_N \geq \eta) \geq 0$, donc $\mathbb{P}(|Y_N - Y| \geq \eta)$ ne tend pas vers 0 pour N (impair) tendant vers $+\infty$.
Il n'y a donc pas convergence en probabilité. □

A.2 Preuve des résultats concernant les mécanismes d'*inverse sensitivity* et les bornes inférieures de précision

A.2.1 Preuves des résultats d'optimalité du mécanisme d'*inverse sensitivity*

Démonstration. (Première proposition en 6.1) D'après la remarque suivant la Définition 3.1 de [1], M_{inv} vérifie :

$$\forall X, X', \forall \rho > 0, \mathbb{P}(|M_{inv}(X) - f(X)| \leq \rho) \geq \mathbb{P}(|M_{inv}(X) - f(X')| \leq \rho) \quad (1)$$

Soit $k \in [\![0, n]\!]$, montrons $\mathbb{E}[|M_{inv}(X) - f(X)|] \leq \mathbb{E}[|M_{inv}(X) - \frac{k}{n}|]$.
Fixons pour cela X' tel que $f(X') = \frac{k}{n}$.

$$\begin{aligned}
\mathbb{E}[|M_{inv}(X) - f(X)|] &= \sum_{j=0}^{n} \frac{j}{n} \mathbb{P}(|M_{inv}(X) - f(X)| = \frac{j}{n}) \\
&= \frac{1}{n} \sum_{j=1}^{n} j \Big[\mathbb{P}\big(|M_{inv}(X) - f(X)| \leq \frac{j}{n}\big) - \mathbb{P}\big(|M_{inv}(X) - f(X)| \leq \frac{j-1}{n}\big) \Big] \\
&= \frac{1}{n} \Big[n \underbrace{\mathbb{P}(|M_{inv}(X) - f(X)| \leq \frac{n}{n})}_{=1} - \sum_{j=0}^{n-1} \mathbb{P}(|M_{inv}(X) - f(X)| \leq \frac{j}{n}) \Big] \text{ (télescopage)} \\
&= 1 - \frac{1}{n} \sum_{j=0}^{n-1} \underbrace{\mathbb{P}\Big(|M_{inv}(X) - f(X)| \leq \frac{j}{n}\Big)}_{\geq \mathbb{P}\left(|M_{inv}(X) - f(X')| \leq \frac{j}{n}\right) \text{ par (1)}} \\
&\leq 1 - \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{P}\Big(|M_{inv}(X) - \frac{k}{n}| \leq \frac{j}{n}\Big) \text{ car } f(X') = \frac{k}{n} \\
&= \mathbb{E}\Big[\Big|M_{inv}(X) - \frac{k}{n}\Big|\Big]
\end{aligned}$$

La dernière ligne se montre de la même manière qu'on a prouvé $\mathbb{E}[|M_{inv}(X) - f(X)|] = 1 - \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{P}\Big(|M_{inv}(X) - f(X)| \leq \frac{j}{n}\Big)$. □

Démonstration. (Théorème en 6.1)
Nous nous inspirons dans ce raisonnement de la démonstration du théorème 1 de [2].

— Montrons que pour tous $X \sim X'$,

$$\mathbb{E}[|M(X) - f(X)|] \leq e^{\epsilon} \mathbb{E}[|M(X') - f(X')|].$$

Pour tout $k \in [\![0, n]\!]$,

$$\begin{aligned}
\mathbb{E}\Big[\Big|M(X) - \frac{k}{n}\Big|\Big] &= \sum_{j=0}^{n} \mathbb{P}\big(M(X) = \frac{j}{n}\big) \frac{|k-j|}{n} \\
&\leq \sum_{j=0}^{n} e^{\epsilon} \mathbb{P}\big(M(X') = \frac{j}{n}\big) \frac{|k-j|}{n} \text{ par } \epsilon\text{-D.P., puisque } X \sim X' \\
&= e^{\epsilon} \mathbb{E}\Big[\Big|M(X') - \frac{k}{n}\Big|\Big].
\end{aligned}$$

Donc

$$\begin{aligned}
\frac{\mathbb{E}[|M(X) - f(X)|]}{\mathbb{E}[|M(X') - f(X')|]} &= \frac{\mathbb{E}[|M(X) - f(X)|]}{\mathbb{E}[|M(X) - f(X')|]} \frac{\mathbb{E}[|M(X) - f(X')|]}{\mathbb{E}[|M(X') - f(X')|]} \\
&\leq e^{\epsilon} \qquad (1)
\end{aligned}$$

par l'inégalité précédente et car M est *unbiased*.

— Soit dans la suite α tel que $\mathbb{E}[|M(X) - f(X)|] \leq \alpha$. Par l'inégalité de Markov,

$$\mathbb{P}\Big(|M(X) - f(X)| \geq \frac{1}{2n}\Big) \leq 2n\alpha.$$

Fixons X' tel que $X \sim X'$ et $|f(X) - f(X')| = \frac{1}{n}$ (par exemple $X' = (1 - X_1, X_2, X_3, ..., X_n)$).

— Nous nous intéressons maintenant à

$$\mathbb{P}\Big(|M(X) - f(X')| \geq \frac{1}{2n}\Big) :$$

$$\begin{aligned}\mathbb{P}\Big(|M(X) - f(X')| \geq \frac{1}{2n}\Big) &\leq e^{\epsilon}\mathbb{P}\Big(|M(X') - f(X')| \geq \frac{1}{2n}\Big) \text{ par } \epsilon\text{-D.P.} \\ &\leq 2ne^{2\epsilon}\alpha \text{ par l'inégalité de Markov et par (1).}\end{aligned}$$

— or $|f(X)-f(X')| = \frac{1}{n}$, donc $\{|M(X)-f(X)| < \frac{1}{2n}\}$ et $\{|M(X)-f(X')| < \frac{1}{2n}\}$ sont disjoints. Ainsi :

$$\begin{aligned}1 &\geq \mathbb{P}\Big(|M(X) - f(X)| < \frac{1}{2n}\Big) + \mathbb{P}\Big(|M(X) - f(X')| < \frac{1}{2n}\Big) \\ &\geq 1 - 2n\alpha + 1 - 2ne^{2\epsilon}\alpha\end{aligned}$$

d'après les deux points précédents. Ce qui suffit à conclure.

□

Démonstration. (Seconde proposition en 6.1)

$$\begin{aligned}\mathcal{CN}(X^{(n)})\mathbb{E}[|M_{inv}(X^{(n)}) - f(X^{(n)})|] &= \sum_{j=0}^{n} \frac{1}{n}|j - nf(X^{(n)})|e^{-\epsilon|nf(X^{(n)})-j|/2} \\ &= \sum_{j=0}^{nf(X^{(n)})} \frac{1}{n}(nf(X^{(n)}) - j)e^{-\epsilon(nf(X^{(n)})-j)/2} + \sum_{j=nf(X^{(n)})+1}^{n} \frac{1}{n}(j - nf(X^{(n)}) \\ &= \sum_{l=0}^{nf(X^{(n)})} \frac{1}{n}le^{-\epsilon l/2} \\ &+ \sum_{l=1}^{n(1-f(X^{(n)}))} \frac{1}{n}le^{-\epsilon l/2} \text{ (changement de variables } [l = nf(X^{(n)}) - j]) \\ &= \frac{1}{n}\frac{1}{(1 - e^{-\epsilon/2})^2}\Big[\Big(1 - (nf(X^{(n)}) + 2)e^{-\epsilon(nf(X^{(n)})+1)/2} + (nf(X^{(n)}) + 1)\Big) \\ &+ \Big(1 - (n - nf(X^{(n)}) + 2)e^{-\epsilon(n-nf(X^{(n)})+1)/2} + (n - nf(X^{(n)}) + 1)e^{-\epsilon(n}\end{aligned}$$

Or, ce dernier crochet tend vers $\frac{2}{(1-e^{-\epsilon/2})^2}$ quand n tend vers $+\infty$.
De plus,

$$\mathcal{CN}(X^{(n)}) = \frac{e^{-n\epsilon f(X^{(n)})/2} - e^{\epsilon/2} - 1 + e^{n\epsilon(f(X^{(n)})-1)/2}}{1 - e^{\epsilon/2}},$$

donc

$$\frac{1}{\mathcal{CN}(X^{(n)})} \underset{n\to+\infty}{\longrightarrow} \frac{e^{\epsilon/2}-1}{1+e^{\epsilon/2}}.$$

Ainsi,

$$\begin{aligned}\mathbb{E}[|M_{inv}(X^{(n)}) - f(X^{(n)})|] \times 2n(e^{2\epsilon}+1) &\underset{n\to+\infty}{\longrightarrow} \frac{2}{(1-e^{-\epsilon/2})^2} \times \frac{e^{\epsilon/2}-1}{1+e^{\epsilon/2}} \times 2(e^{2\epsilon}+1) \\ &= \frac{4(e^{2\epsilon}+1)}{1-e^{-\epsilon}}\end{aligned}$$

□

A.2.2 Preuves des résultats concernant l'optimalité du mécanisme limite (partie 6.2)

Démonstration. (Proposition en 6.2)
Utilisons le théorème de Lévy. Soit $t \in \mathbb{R}$.

$$\begin{aligned}\Phi_{Y_n}(t) &= \sum_{k=0}^{n} \frac{e^{it\frac{k}{n} - \epsilon|nf(X^{(n)})-k|/2}}{\mathcal{CN}(X^{(n)}} \\ &= \sum_{k=0}^{nf(X^{(n)})} [e^{\frac{it}{n}+\frac{\epsilon}{2}}]^k \ \frac{e^{-\epsilon nf(X^{(n)})/2}}{\mathcal{CN}(X^{(n)})} + \sum_{k=nf(X^{(n)})+1}^{n} [e^{\frac{it}{n}-\frac{\epsilon}{2}}]^k \ \frac{e^{\epsilon nf(X^{(n)})/2}}{\mathcal{CN}(X^{(n)})} \\ &= \frac{1}{\mathcal{CN}(X^{(n)})} \Bigg[\underbrace{e^{-\epsilon nf(X^{(n)})/2} \frac{1 - \left[e^{\frac{it}{n}+\frac{\epsilon}{2}}\right]^{nf(X^{(n)})+1}}{1 - e^{\frac{it}{n}+\frac{\epsilon}{2}}}}_{=:(1)} \\ &\quad + \underbrace{e^{\epsilon nf(X^{(n)})/2} \times \frac{1 - \left[e^{\frac{it}{n}-\frac{\epsilon}{2}}\right]^{n(1-f(X^{(n)}))}}{1 - e^{\frac{it}{n}-\frac{\epsilon}{2}}} \times \left[e^{\frac{it}{n}-\frac{\epsilon}{2}}\right]^{nf(X^{(n)})+1}}_{=:(2)} \Bigg],\end{aligned}$$

cette dernière ligne s'obtenant par somme de termes consécutifs d'une suite géométrique. Or

—

$$(1) = \frac{e^{-\epsilon n f(X^{(n)})/2} - e^{itf(X^{(n)}) + \frac{it}{n} + \frac{\epsilon}{2}}}{1 - e^{\frac{it}{n} + \frac{\epsilon}{2}}} \underset{n \to +\infty}{\longrightarrow} \frac{e^{itf(X^\infty) + \frac{\epsilon}{2}}}{e^{\epsilon/2} - 1}$$

car $f(X^{(n)}) \underset{n \to +\infty}{\longrightarrow} f(X^\infty)$ (par définition de $f(X^\infty)$)

—

$$(2) = \frac{e^{itf(X^{(n)}) + \frac{it}{n} - \frac{\epsilon}{2}} - e^{\frac{it}{n} - \frac{\epsilon}{2}(n+1+nf(X^{(n)})) + it}}{1 - e^{\frac{it}{n} - \frac{\epsilon}{2}}} \underset{n \to +\infty}{\longrightarrow} \frac{e^{itf(X^\infty) - \frac{\epsilon}{2}}}{1 - e^{-\epsilon/2}}$$

—

$$\mathcal{CN}(X^{(n)}) = \frac{e^{-n\epsilon f(X^{(n)})/2} - e^{\epsilon/2} - 1 + e^{n\epsilon(f(X^{(n)})-1)/2}}{1 - e^{\epsilon/2}} \underset{n \to +\infty}{\longrightarrow} \frac{e^{\epsilon/2} + 1}{e^{\epsilon/2} - 1}$$

Ainsi,

$$\Phi_{Y_n}(t) \underset{n \to +\infty}{\longrightarrow} \frac{e^{\epsilon/2} - 1}{e^{\epsilon/2} + 1} \left[\frac{e^{itf(X^\infty) + \frac{\epsilon}{2}}}{e^{\epsilon/2} - 1} + \frac{e^{itf(X^\infty) - \frac{\epsilon}{2}}}{1 - e^{-\epsilon/2}} \right] = e^{itf(X^\infty)} = \Phi_Y(t),$$

où Y est une V.A. de loi Dirac en $f(X^\infty)$. □

Références

[1] Hilal Asi et John C. Duchi. « Instance-optimality in differential privacy via approximate inverse sensitivity mechanisms ». In : *34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.* (2020).

[2] Hilal Asi et John C. Duchi. « Near Instance-Optimality in Differential Privacy ». In : *https://arxiv.org/abs/2005.10630* (2020).

[3] Cynthia Dwork et Aaron Roth. *The Algorithmic Foundations of Differential Privacy.* T. 9. 3-4. Foundations et Trends ®in Theoretical Computer Science, 2014, p. 1-277. doi : 10.1561/0400000042.

[4] Jennifer Gillenwater, Matthew Joseph et Alex Kulesza. « Differentially private quantiles ». In : *Proceedings of the 38th International Conference on Machine Learning* (2021).

[5] H. Kaplan, S. Schnapp et U. Stemmer. « Differentially private approximate quantiles ». In : *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA, volume 162 of Proceedings of Machine Learning Research, 10751–10761.* (2022).

[6] Clément Lalanne. « On the tradeoffs of statistical learning with privacy ». In : (Machine Learning [cs.LG]. ENS Lyon, 2023. English. NNT : 2023ENSL0068. tel-04431813).

[7] Clément Lalanne, Aurélien Garivier et Rémi Gribonval. « Private Statistical Estimation of Many Quantiles ». In : *Proceedings of the 40th International Conference on Machine Learning* (2023).

[8] Adam Smith. « Privacy-preserving Statistical Estimation with Optimal Convergence Rates ». In : *STOC '11: Proceedings of the forty-third annual ACM symposium on Theory of computing* (2011).