Rapport du Semestre de Stage à l'Etranger (S8)

Foundation of Private and Fair Statistics

Amina MANSEUR - 2A

Stage effectué du 10 février 2023 au 7 juillet 2023

Tuteur école : Mme Mitra FOULADIRAD
Maître de stage : M. Di WANG

Établissement d'enseignement : Ecole Centrale Méditerranée
Organisation d'accueil : King Abdullah University of Science and Technology (KAUST)

Résumé

Résumé version française

Le Semestre de Stage à l'Étranger (SSE) à l'École Centrale Méditerranée constitue une opportunité exceptionnelle pour les étudiants de découvrir la recherche dans un contexte international. Pour mon SSE, j'ai choisi de rejoindre le laboratoire PRADA de King Abdullah University of Science and Technology (KAUST) en Arabie Saoudite. Ce laboratoire se concentre sur des aspects avancés de l'apprentissage automatique, de la confidentialité des données, de l'apprentissage robuste et de l'intelligence artificielle explicative. Mon stage de cinq mois a été axé sur deux projets théoriques liés à la confidentialité différentielle et à l'apprentissage robuste. Le premier projet portait sur les mécanismes de véracité permettant de garantir la précision des estimateurs tout en préservant la confidentialité des données. Le second projet se penchait sur les bandits multi-bras dans le contexte de l'apprentissage robuste. Mon expérience de stage m'a permis d'appliquer mes connaissances théoriques en algorithmique, théorie des jeux, probabilités et statistiques. J'ai pu développer des compétences méthodologiques et acquérir une compréhension approfondie des enjeux de la recherche en apprentissage automatique. En travaillant aux côtés de chercheurs et d'étudiants internationaux, j'ai également pu enrichir ma vision du monde et développer mon sens de l'adaptabilité. En résumé, ce stage a comblé mes attentes en termes de développement de compétences, de collaboration internationale et d'enrichissement personnel. Mon rapport de stage vise à rendre compte en détail de mon travail au sein du laboratoire PRADA et de l'impact de cette expérience sur ma formation.

English version abstract

The Semester Abroad Internship at École Centrale Méditerranée provides an exceptional opportunity for students to explore research within an international context. For this internship, I chose to join the PRADA laboratory at the King Abdullah University of Science and Technology (KAUST) in Saudi Arabia. This laboratory focuses on advanced aspects of machine learning, data privacy, robust learning, and explainable artificial intelligence. My five-month internship was centered on two theoretical projects related to differential privacy and robust learning. The first project revolved around veracity mechanisms to ensure estimator accuracy while preserving data privacy. The second project focused on multi-armed bandits in the context of robust learning. My internship experience enabled me to apply my theoretical knowledge in algorithms, probability, statistics, and game theory. I developed methodological skills and gained an in-depth understanding of the challenges in machine learning research. Working with international researchers and students has enriched my global perspective and strengthened my ability to adapt. In summary, this internship has met my expectations in terms of skill development, international collaboration, and personal enrichment. My internship report aims to provide a detailed account of my work within the PRADA laboratory and the impact of this experience on my education.

Mots-clés : Stage, recherche, expérience internationale, probabilités, statistiques, informatique, algorithmique, machine learning, théorie des jeux, données sensibles, confidentialité, confidentialité différentielle, bandits multi-bras, mécanisme de véracité

Remerciements

Je tiens tout d'abord à remercier sincèrement ma tutrice de stage, Madame Mitra Fouladirad, ainsi que tous les professeurs et étudiants qui m'ont apporté leur aide tout au long de ma recherche de stage. Leurs précieux conseils, recommandations et partages d'expériences ont été pour moi une véritable source d'aide et d'inspiration.

Je tiens également à exprimer ma profonde gratitude au professeur Di Wang, responsable du laboratoire PRADA, ainsi qu'aux étudiants et professeurs qui le composent. Leur accueil chaleureux et leur confiance ont joué un rôle essentiel dans le bon déroulement de cette expérience. Leurs connaissances et leur expertise ont été une source d'apprentissage inestimable et j'ai été honorée de collaborer avec eux.

Enfin, je tiens à remercier chaleureusement toutes les personnes que j'ai eu le privilège de rencontrer au cours de ce séjour, que ce soit lors de simples discussions ou au cours de moments de partage et de convivialité. Elles ont grandement contribué à faire de cette expérience un moment inoubliable et empreint de sens.

Table des matières

1 Introduction

Le Semestre de Stage à l'Étranger (SSE) de deuxième année à l'École Centrale Méditerranée offre aux étudiants l'opportunité de découvrir la recherche au sein de laboratoires universitaires ou de services de R&D d'entreprises à l'étranger. Cette expérience, essentielle pour les futurs ingénieurs menés à évoluer dans un contexte international et/ou interculturel, leur permet également d'acquérir de nouvelles compétences dans leur domaine d'études.

Pour ma part, j'ai décidé de saisir cette opportunité et d'effectuer un stage de recherche de cinq mois au sein du laboratoire PRADA (Provable Responsible AI and Data Analytics) de l'université KAUST, en Arabie Saoudite. Les projets proposés au sein de ce laboratoire ont particulièrement attiré mon attention car ils mettent en application des notions que j'avais déjà abordées lors de mon cursus en école d'ingénieurs. Cependant, je souhaitais approfondir ces connaissances et les appliquer spécifiquement dans les domaines de l'apprentissage automatique et de l'intelligence artificielle.

De plus, le laboratoire se focalise sur des aspects cruciaux de ces deux domaines, notamment la préservation de la confidentialité, l'apprentissage robuste, l'intelligence artificielle explicative et le désapprentissage. Ces thématiques étaient donc parfaitement alignées sur mes intérêts et sur ce que je recherchais.

Au cours de mon SSE, j'ai participé à deux projets théoriques centrés sur la notion de confidentialité différentielle, essentielle dans un contexte où la protection des données est primordiale et explorant des aspects avancés de l'algorithmique, de la théorie des jeux, des probabilités et des statistiques. Le premier projet explorait les mécanismes de véracité qui permettent de garantir la précision des estimateurs malgré les préoccupations des agents concernant leur vie privée. Le second se penchait sur les bandits multibras dans le cadre de l'apprentissage robuste, visant à renforcer la fiabilité des systèmes face à des perturbations.

Le stage a été une occasion de mettre en pratique mes connaissances acquises à l'École Centrale Méditerranée, en développant des compétences méthodologiques et en comprenant les enjeux de la recherche dans un domaine en constante évolution

Outre l'aspect scientifique, ce SSE m'a également offert l'opportunité unique de m'immerger dans une nouvelle culture et d'élargir mes horizons personnels. Vivre et travailler au sein d'un environnement international et interculturel m'a permis de développer mon sens de l'adaptabilité et d'enrichir ma vision du monde.

Mes attentes pour ce SSE ont été largement comblées, permettant l'approfondissement des connaissances en apprentissage automatique, l'exploration de questions cruciales liées à ce domaine, et le développement de compétences en recherche et développement au sein d'un environnement international.

En résumé, ce stage au sein du laboratoire PRADA de KAUST a apporté des avantages considérables en termes d'acquisition de compétences, de collaboration internationale, et d'enrichissement personnel et culturel. Ce rapport vise à rendre compte clairement et en profondeur du travail accompli et de l'apport scientifique de cette expérience.

2 Présentation du laboratoire

2.1 Présentation générale

J'ai effectué mon semestre de stage à l'étranger au sein du laboratoire PRADA [PRA] de l'université KAUST [KAU].

Le laboratoire PRADA, anciennement connu sous le nom de PART Lab (Privacy Awareness, Responsibility and Trustworthy Lab), a été fondé en janvier 2021 par le Professeur Di Wang au sein du Département d'Informatique de KAUST. Ce changement de nom a eu lieu durant mon séjour à KAUST et reflète l'évolution et l'élargissement des domaines d'étude du laboratoire, qui ne se limitent plus uniquement à la préservation de la vie privée dans l'apprentissage automatique, mais englobent désormais des aspects plus larges de l'intelligence artificielle responsable. Situé au sein de la Division des Sciences Informatiques, Électriques, Mathématiques et de Génie de KAUST, le laboratoire PRADA réunit une équipe diversifiée de plus de 20 professeurs, chercheurs invités, étudiants en doctorat, en master et stagiaires, et bénéficie ainsi d'une expertise diverse et variée.

Le laboratoire travaille en collaboration avec KAUST, le Royaume d'Arabie saoudite ainsi qu'avec d'autres laboratoires à travers le monde. Son objectif principal est de relever les défis fondamentaux de l'intelligence artificielle et de l'apprentissage automatique fiables, afin de contribuer aux avancées et de répondre aux besoins mondiaux.

Le laboratoire PRADA se distingue par son engagement dans des sujets de pointe tels que la préservation de la vie privée dans l'apprentissage automatique, l'apprentissage robuste, l'intelligence artificielle explicative et le désapprentissage automatique. Il joue par ailleurs un rôle actif dans la communauté scientifique en publiant régulièrement des articles de recherche et en participant à des conférences et des collaborations internationales.

2.2 Les differents groupes et domaines de recherche du laboratoire

Le laboratoire PRADA se divise en différents groupes de recherche, chaque groupe étant spécialisé dans un domaine spécifique.

Il y a tout d'abord, le groupe "Privacy-preserving Machine Learning". Ce groupe se consacre à la préservation de la confidentialité des données dans le domaine de l'apprentissage automatique. Leurs recherches se concentrent sur des mécanismes tels que la confidentialité différentielle et l'apprentissage fédéré, et visent à garantir la protection des données tout en maintenant la performance des modèles. Leur objectif est de trouver des solutions innovantes pour exploiter les données sensibles de manière sécurisée et responsable.

Puis, il y a le groupe "Machine Unlearning" qui se concentre sur le développement d'algorithmes et de systèmes de désapprentissage automatique. Leur objectif est de concevoir des méthodes permettant de supprimer sélectivement les données apprises ou les informations

FIG. 1 – Les membres du laboratoire PRADA.

spécifiées des modèles d'apprentissage automatique. Cette approche vise à prévenir la génération de contenu nuisible et à se conformer aux réglementations en matière de confidentialité et de protection des données.

Le groupe "Faithful and Interpretable AI", quant à lui, se penche sur le développement de méthodes d'intelligence artificielle explicables et fidèles. Leur objectif est de fournir des explications précises et interprétables des prédictions effectuées par les modèles d'apprentissage automatique. En rendant les décisions des modèles compréhensibles, ils renforcent la confiance et facilitent la prise de décisions éclairées basées sur ces modèles.

Le groupe "Optoelectronic Acceleration for Machine Learning Algorithms" explore les possibilités d'accélérer les algorithmes d'apprentissage automatique en utilisant des processeurs optoélectroniques. Ces processeurs offrent des avantages tels qu'une consommation d'énergie réduite et une capacité de calcul rapide, ce qui les rend prometteurs pour les futurs systèmes d'intelligence artificielle. Les chercheurs de ce groupe explorent les applications potentielles de cette technologie afin d'améliorer les performances des algorithmes d'apprentissage automatique.

Enfin, le groupe "Artificial Intelligence and Optical Systems" se concentre sur l'amélioration des systèmes optiques utilisés dans les spectromètres miniaturisés. Leurs travaux portent sur les avancées en nano photonique, en algorithmes d'intelligence artificielle et en ingénierie des systèmes optiques pour améliorer la reconnaissance des couleurs et les performances des systèmes de spectroscopie miniaturisés.

Cette diversité contribue à l'expertise variée du laboratoire et donne ainsi la possibilité d'explorer les différentes dimensions de l'apprentissage automatique, de l'intelligence artificielle et de la science des données.

2.3 Equipe de recherche

L'équipe de recherche de ce laboratoire est un groupe multidisciplinaire qui se caractérise par la variété de ses membres. À sa tête se trouve le professeur assistant Di WANG, un chercheur titulaire d'un doctorat en informatique et qui a acquis son expertise à l'Université d'État de New York (SUNY) à Buffalo. Son parcours académique comprend également un master en mathématiques de l'Université de Western Ontario et une licence en mathématiques et mathématiques appliquées de l'Université de Shandong en Chine.

Outre le directeur, l'équipe compte des chercheurs invités et des post-doctorants, dont certains sont associés à d'autres universités internationales telles que Huazhong Agricultural University et Nanjing Tech University. Leurs domaines de recherche sont tout aussi variés que leurs origines académiques, couvrant des sujets allant des réseaux sociaux au transport intelligent, en passant par l'apprentissage fédéré, le traitement du langage naturel, la confidentialité différentielle, et bien d'autres.

Le laboratoire est également composé d'étudiants en doctorat, qui apportent une perspective dynamique à l'équipe. Ils proviennent d'horizons académiques divers et possèdent des diplômes de master et de licence dans des domaines variés. Leurs recherches s'étendent de l'expliquabilité de l'IA à la confidentialité des données, en passant par la sécurité de l'apprentissage automatique et l'IA pour les sciences.

Les étudiants en master participent également activement aux projets de recherche ; ils ont suivi des formations variées et travaillent sur des sujets tels que l'exploitation de données spatio-temporelles, la sécurité de l'apprentissage automatique et d'autres domaines de l'IA.

En plus des membres permanents, des étudiants en échange universitaire apportent des perspectives nouvelles à l'équipe.

Cette diversité de profils académiques favorise une approche collaborative et innovante au sein de l'équipe.

2.4 Organisation au sein du laboratoire

En ce qui concerne l'organisation du travail au sein du laboratoire, plusieurs points méritent d'être soulignés.

Premièrement, le directeur organise des réunions de groupe une à deux fois par semestre, au cours desquelles les projets de recherche liés aux thématiques du laboratoire ainsi que les projets personnels de chaque membre sont partagés et discutés. Ces réunions permettent avant tout de présenter les travaux réalisés et de partager les résultats avec les autres membres du laboratoire. Les échanges d'idées, les retours constructifs et les propositions de collaboration avec d'autres laboratoires renforcent l'esprit d'équipe et créent un environnement stimulant,

jouant ainsi un rôle essentiel dans la dynamique du groupe et l'avancement des projets. Ces réunions de groupe sont également l'occasion de définir de nouvelles stratégies et axes de recherche en réponse aux avancées récentes, afin de rester compétitifs et de proposer des solutions novatrices et adaptées aux nouveaux défis. Elles permettent à chacun d'avoir une vue d'ensemble sur les projets en cours au sein du laboratoire, d'échanger des idées et de contribuer avec son expertise. Par ailleurs, ces réunions contribuent à renforcer la cohésion du groupe.

En dehors du travail scientifique, des activités de cohésion sont également organisées, telles que des sorties plongées avec tuba pour explorer la nature environnante, ou encore des repas partagés. Nous avons par exemple partagé un iftar (repas de coupure du jeûne) pendant le mois de Ramadan lors de la visite d'une chercheuse française qui voulait établir des collaborations avec le laboratoire PRADA. Ces moments de convivialité favorisent les liens entre les membres du laboratoire, et renforcent ainsi la collaboration et l'harmonie au sein de l'équipe.

FIG. 2 – Repas de cohésion de groupe.

Deuxièmement, il est attendu que chaque étudiant participe au moins à un séminaire par semaine, durant lequel il doit présenter deux ou trois articles récents et pertinents liés au groupe de recherche dans lequel il travaille. Ces séminaires sont organisés de manière hebdomadaire pour les différents groupes du laboratoire, offrant ainsi une occasion de partager de nouvelles connaissances pouvant être utiles au sein du groupe.

Pour ma part, cette expérience m'a permis d'apprendre à cibler les éléments importants d'un article lors de sa lecture, afin de savoir où trouver les informations pertinentes et nécessaires lors d'un projet de recherche. Participer à ces séminaires m'a été très bénéfique, car cela m'a permis d'approfondir ma compréhension des sujets abordés dans notre groupe de recherche et de rester informée des dernières avancées dans le domaine du Machine Learning.

Pour finir, toutes les deux semaines, tous les étudiants membres du laboratoire bénéficient d'entretiens individuels avec le directeur du laboratoire afin de discuter de l'avancement de leur projet. Ces rencontres privilégiées offrent l'opportunité de faire le point sur les progrès réalisés, mais aussi d'aborder les éventuels obstacles rencontrés et de recevoir des conseils per-

sonnalisés. Ces réunions permettent également de bénéficier de l'expertise et de l'expérience du directeur, qui peut fournir des aides précieuses pour orienter les recherches dans la bonne direction. Elles favorisent ainsi une approche personnalisée et un suivi régulier, garantissant que chaque étudiant puisse bénéficier d'un soutien adapté et développer ses compétences de recherche de manière optimale.

2.5 Collaborations et partenariats

L'université favorise activement les échanges, les collaborations et les partenariats avec d'autres laboratoires de recherche internationaux. Ainsi, plusieurs événements sont organisés pour promouvoir et faciliter ces interactions.

Par exemple, l'université organise régulièrement des ateliers et invitent pour cette occasion des chercheurs du monde entier. Ceci permet aux étudiants et aux enseignants-chercheurs de KAUST de rencontrer et d'échanger avec des professeurs et des chercheurs spécialisés dans leur domaine. Ces ateliers offrent également l'opportunité d'assister à des mini-cours et à des présentations sur de nombreux sujets. Durant mon séjour, j'ai eu l'occasion d'assister à plusieurs de ces ateliers comme le workshop intitulé "Stochastic Numerics and Statistical Learning" pour lequel une douzaine de professeurs de plus de 7 pays différents ont répondu présents. Ce workshop avait été organisé par un laboratoire de KAUST en collaboration avec l'université.

Un autre exemple est le "Rising Stars in AI Symposium 2023," auquel j'ai pu également participer. Cet événement s'adressait essentiellement aux jeunes chercheurs, doctorants, post-doctorants et professeurs en début de carrière, qui ont récemment contribué à la recherche en intelligence artificielle. Cet événement a favorisé les discussions et le partage des idées de recherche passionnantes. Les orateurs sélectionnés pour participer à cet événement ont eu l'opportunité de présenter leurs travaux acceptés lors de conférences prestigieuses sur l'IA telles que NeurIPS, ICML, ICLR, CVPR, ECCV, ICCV, EMNLP, ACL, et bien d'autres. Durant cet événement, j'ai eu l'occasion d'assiter à une présentation donnée par Gautam Kamath de l'Université de Waterloo, spécialiste reconnu pour ses travaux sur la confidentialité différentielle, et dont j'avais suivi les cours sur Youtube afin d'en apprendre davantage sur ce sujet essentiel pour mon stage de recherche.

De plus, des collaborations internes sont encouragées au sein de chaque laboratoire avec la possibilité d'inviter des professeurs à KAUST, pour présenter leurs travaux et explorer des opportunités de partenariats. Par exemple, nous avons eu l'honneur de recevoir Mme Laure Berti-Équille, la directrice de recherche à l'IRD (Institut de recherche pour le développement). Son intervention a mis en lumière le rôle crucial de l'apprentissage automatique dans la résolution de défis majeurs liés aux objectifs de développement durable des Nations unies.

Ces différents événements favorisent la recherche interdisciplinaire et la collaboration internationale, et contribue ainsi à relever les défis actuels et futurs.

2.6 Impact et applications

Les travaux de recherche du laboratoire PRADA ont un impact significatif dans divers domaines et ouvrent la voie à des applications concrètes et à des solutions concernant des défis majeurs. Avant l'intervention de Laure Berti-Équille, j'étais déjà consciente de l'importance de ce type de recherche, mais la présentation de son travail a permis de mettre en lumière certains domaines d'application clés en lien avec les enjeux mondiaux actuels.

L'une de ces applications de ce type de recherche réside dans la capacité à répondre aux crises climatiques, aux catastrophes naturelles et aux pandémies. Certaines techniques, comme celles développées au sein du laboratoire PRADA, permettent de modéliser et d'anticiper la propagation de ces événements, aidant ainsi à prendre des mesures de prévention et à atténuer leurs effets dévastateurs. De plus, la recherche en analyse de données et en apprentissage automatique est cruciale pour la prévision de phénomènes tels que la pauvreté et la perte de biodiversité à différentes échelles géographiques et temporelles. Ces prévisions intègrent diverses sources de données et exigent des modèles sophistiqués qui combinent des aspects de la physique avec des approches de machine learning avancées.

L'impact positif de ces recherches sur la société est certain, et la présentation de Laure Berti-Équille n'a fait que renforcer cette conviction. De plus, il est probable que d'autres applications prometteuses existent, et il est enthousiasmant de penser que la recherche au sein du laboratoire PRADA contribue à la résolution de défis contemporains importants.

2.7 Ma place au sein du laboratoire

Durant mon stage, j'ai eu la chance de faire partie du programme VSRP (Visiting Student Research Program) de l'université KAUST et d'intégrer, en tant que stagiaire, le groupe “Privacy Preserving Machine Learning" du laboratoire PRADA.

Ma première mission était de me familiariser avec les notions clés sur lesquelles ce groupe travaillait, telles que la préservation de la vie privée dans l'apprentissage automatique, les mécanismes d'acquisition de données privées, les mécanismes de véracité et les bandits multibras.

J'ai pu mener à bien cette mission notamment grâce à la lecture et à la présentation d'articles lors des séminaires, ainsi qu'au suivi de cours en ligne sur la notion de ‘Differential Privacy”. Après une période d'immersion et de compréhension approfondie des concepts, j'ai pu contribuer activement à deux projets en collaboration avec deux étudiants du groupe, dans le but de rédiger deux articles scientifiques.

Le premier projet sur lequel j'ai travaillé est intitulé "Sparse Truthful Linear Regression". J'ai collaboré avec Liyang, un étudiant en master qui se consacrait à l'étude des mécanismes de véracité dans le cadre de sa thèse de master. Ce premier projet visait à développer une méthode de régression linéaire parcimonieuse garantissant l'honnêteté des agents et la précision des résultats recherchés. J'ai participé à la conception de l'algorithme ainsi qu'à la recherche des résultats d'efficacité relatifs à celui-ci en terme de : protection de la vie privée,

véracité des données, précision des estimations, rationalité individuelle et gestion du budget total alloué aux individus.

Le second projet auquel j'ai contribué, intitulé "Alpha Fraction Private and Robust Bandits", a été réalisé en étroite collaboration avec Yulian, spécialisée dans les bandits multibras en lien avec la préservation de la vie privée. Notre objectif était de concevoir des algorithmes de bandits respectueux de la vie privée, et robustes, en contaminant une fraction alpha des données disponibles. Ce projet relevait d'une importance particulière, car nous devions terminer nos travaux avant le mois d'août afin de soumettre nos résultats pour une conférence spécifique. La pression était donc élevée, car nous devions être efficaces et rigoureux dans notre travail afin de présenter nos contributions avant d'autres équipes travaillant sur des problématiques similaires... Finalement, Yulian a soumis l'article auprsè de l'"International Conference on Artificial Intellogence and Statistics" au début du mois d'octobre 2023.

Dans l'ensemble, ma mission au sein du laboratoire PRADA était de me familiariser avec les concepts clés, de contribuer aux projets du groupe "Privacy Preserving Machine Learning" et d'essayer de répondre aux enjeux liés à chaque projet. J'ai eu l'opportunité de travailler aux côtés d'étudiants talentueux, d'apporter des idées et de participer activement au développement d'algorithmes respectueux de la vie privée. Mon expérience a été enrichissante et m'a permis de comprendre l'importance de la recherche dans ces domaines.

Par ailleurs, cette expérience m'a permis d'en apprendre davantage sur le fonctionnement d'un laboratoire de recherche, en termes de recherche et de collecte des données, de rédaction d'articles reprenant des résultats de la recherche, de la présentation à des conférences, de révision et de publication dans des revues scientifiques.

Ce que j'ai également apprécié lors de mon stage est l'entraide qui régne au sein du laboratoire et de l'université dans son ensemble. J'ai constaté une collaboration fructueuse entre les différents laboratoires et chercheurs, qui se soutenaient mutuellement dans leurs expériences et projets de recherche. Par exemple, j'ai pu apporter mon expertise en codage Python à Hadeel, une étudiante en astrophysique qui en avait besoin pour ses expériences, tandis qu'un autre étudiant du laboratoire PRADA avait besoin de l'aide d'un ou d'une hispanophone afin de mener à bien ses simulations, ce qui a été rendu possible grâce à mon amie Maria.

3 Présentation de la mission

Comme mentionné précédemment, mon stage m'a permis de travailler sur deux projets distincts, chacun ayant ses propres problématiques, méthodes et déroulement de mission. Dans cette partie, je vais présenter en détail les enjeux spécifiques de chaque projet, les méthodes que j'ai employées pour les aborder, le déroulement des missions, les solutions apportées pour résoudre les problèmes et enfin, les conclusions tirées de chacun des projets.

3.0.1 Enjeux liés aux deux projets

De nos jours, nous observons une demande croissante en termes de données, y compris de données sensibles et ce pour diverses raisons.

En effet, la collecte de données représente une réelle opportunité pour la recherche, l'innovation ou encore la résolution de problèmes complexes. Les données sensibles, notamment celles liées à la santé, jouent un rôle considérable dans la personnalisation des soins, l'avancement de la recherche médicale et scientifique, mais aussi dans la personnalisation des services en ligne par exemple. Ces données permettent de réelles avancées en apprentissage automatique et intelligence artificielle et sont sources de progrès dans divers domaines, de la médecine à la gestion de chaîne d'approvisionnement en passant par la recommandation de services. Toutefois, cette utilisation croissante des données sensibles soulève des questions en ce qui concerne la protection de la vie privée des individus. Leur utilisation nécessite donc une gestion responsable des informations afin de permettre une utilisation bénéfique tout en préservant la confidentialité et la sécurité des individus.

Plusieurs techniques qui semblent préserver la vie privée comme l'anonymisation ou la pseudonimisation des bases de données, peuvent être victimes de ce que l'on appelle des attaques, permettant la récupération de l'identité des individus.

Voici un exemple de cas dans lequel l'anonymisation n'a pas été utile pour préserver la vie privée. En 1997, l'État du Massachusetts a publié un vaste ensemble de données sur l'assurance maladie, soigneusement dépourvu d'identifiants individuels. À cette époque, une doctorante du nom de L. Sweeney a décidé de mettre à l'épreuve cette anonymisation. Elle a choisi de suivre les traces du gouverneur de l'État, W. Weld. Sachant que Weld vivait à Cambridge, MA, une ville de 54 000 habitants avec seulement 7 codes postaux, elle a acheté les listes électorales complètes de Cambridge. En croisant les deux ensembles de données, elle a découvert que seuls 6 individus partageaient la date de naissance du gouverneur, que 3 d'entre eux étaient des hommes, et que parmi ces 3, seul le gouverneur vivait dans son propre code postal. Elle a alors envoyé les dossiers de santé du gouverneur à son bureau, démontrant ainsi les limites de l'anonymisation des données. Elle a également montré, par une étude ultérieure menée en 2000, que 53 % de tous les Américains pouvaient être identifiés de manière unique en utilisant uniquement leur code postal, leur date de naissance et leur sexe.

De ce besoin de préservation de la vie privée naît une nouvelle notion, celle de confidentialité différentielle, introduite par Cynthia Dwork, dans un article écrit en 2006 [Dwo+06]. La confidentialité différentielle permet de répondre à deux défis importants dans l'exploitation et l'analyse de données : la préservation de la vie privée des pourvoyeurs de données et l'analyse précise de ces données. Elle permet de garantir que l'ajout ou la suppression d'une donnée n'influence pas de manière significative le résultat de l'analyse. Ainsi, il est impossible de déterminer avec certitude la contribution spécifique d'un individu aux résultats globaux. En voici une définition mathématique.

Soit $\epsilon \geq 0$. Un mécanisme M est ϵ-confidentiellement privé lorsque, pour toutes bases de données X et X' voisines, et pour tout ensemble mesurable S :

$$P(M(X) \in S) \leq e^{\epsilon} P(M(X') \in S) \tag{1}$$

où e est la base du logarithme naturel.

Remarque 1. Deux bases de données X et X' sont dites voisines lorsqu'elles diffèrent d'une seule entrée.

Cette définition peut être interprétée ainsi : quelles que soient les sorties du mécanisme, la probabilité d'observer cette sortie si la contribution d'un utilisateur était modifiée ne changerait pas d'un certain ratio, qui est quantifié par ϵ.

Les deux projets sur lesquels j'ai travaillé reposent sur cette notion de confidentialité différentielle dans le cadre de mon travail dans le groupe "Privacy-preserving Machine Learning".

3.0.2 Ressources à disposition

Durant mon premier mois de stage, je n'étais affectée à aucun projet, ma mission était de me familiariser avec les notions en lien avec les domaines de recherche du groupe "Privacy-preserving Machine Learning". Cela comprenait la lecture d'articles dont ceux que je devais présenter lors des séminaires hebdomadaires ; la lecture de quelques chapitres importants du livre "High-Dimensional Probability", [Ver19] pour m'introduire aux probabilités et statistiques en haute dimension ; ainsi que le visionnage et le suivi d'un cours sur la confidentialité différentielle donné en ligne par Gautam [Kam].

Voici la liste des articles présentés lors des séminaires hebdomadaires :

— A Statistical Framework for Differential Privacy, [WZ09]

— Sample and Threshold Differential Privacy : Histograms and applications, [BC22]

— Learning to Infer Structures of Network Games, [Ros+22]

— Tight Accounting in the Shuffle Model of Differential Privacy, [KHH22]

— Differentially Private Bayesian Linear Regression, [AYB23]

— Sparse Mixed Linear Regression with Guarantees : Taming an Intractable Problem with Invex Relaxation, [BH22]

— Fast Sparse Classification for Generalized Linear and Additive Models, [Liu+22]

— Differentially Private Regression with Unbounded Covariates, [Mil+22]

— Flexible Accuracy for Differential Privacy, [Ban+21]

J'ai également été sensibilisée sur les conférences et les journaux scientifiques fiables grâce auxquels je pouvais trouver des articles intéressants.

1. Dans le domaine du Machine Learning et des Statistiques :
   — ICML : "International Conference on Machine Learning"
   — AISTATS : "International Conference on Artificial Intelligence and Statistics"
   — NeurIPS : "Neural Information Processing Systems"
2. Dans le domaine de la Théorie des Jeux :
   — EC Conference : "Conference on Economics and Computation"
   — WINE : "Conference on Web and Internet Economics"
3. Dans le domaine des Statistiques :
   — Annals of Statistics
   — JASA : "Journal of the American Statistical Association"
   — JRSSB : "Journal of the Royal Statistical Society Series B"
   — Bernoulli Journal

J'avais accès à toutes ces conférences et ces journaux ainsi que d'autres de manière libre et gratuite.

De plus, dans le cadre de mon apprentissage des domaines mentionnés précédemment, ces lectures et présentations d'articles m'ont également doté de compétences essentielles pour l'identification rapide des aspects cruciaux et novateurs d'un article. Cette compétence est particulièrement utile pour évaluer si un article mérite une étude plus approfondie dans le contexte de mes travaux de recherche. Les éléments clés à repérer dans un article afin de déterminer sa pertinence incluent son thème, sa motivation, les méthodes employées, les résultats théoriques obtenus, ainsi que la comparaison avec les travaux antérieurs. Cette aptitude à identifier rapidement ces éléments est un atout précieux pour la recherche bibliographique lors de travaux de recherche.

3.1 Projet 1 : Régression linéaire véridique et parcimonieuse en haute dimension

3.1.1 Problématique et exploration des résultats existantes

Dans le cadre de ce premier projet, le professeur Di Wang a proposé que l'on s'intéresse à de nouveaux défis auxquels les analystes de données peuvent être confrontés de nos jours.

Tout d'abord, l'utilisation croissante des jeux de données de grande dimension dans des domaines tels que la génomique, la finance et les sciences sociales pose des défis quant à la modélisation et l'analyse de ces données. En effet, la grande dimensionnalité rend l'utilisation de modèles classiques inadaptée et nécessite donc l'exploration et l'exploitation de nouvelles méthodes.

D'autre part, le professeur Di Wang, nous a proposé de travailler sur le modèle de régression linéaires qui est l'un des outils les plus couramment utilisés pour modéliser la relation entre une variable cible et des variables explicatives ou descriptives, la variable cible étant celle que nous cherchons à connaître via une relation linéaire avec les variables explicatives. Cependant, malgré sa popularité, le modèle de régression linéaire n'est pas adapté aux ensembles de données de grande dimension ; le problème nécessite donc des adaptations ou des méthodes spécifiques permettant de relever ce défi.

Par ailleurs, deux autres notions importantes sont utilisées dans ce projet : la confidentialité différentielle introduite en 1 et les mécanismes de véracité. Les mécanismes de véracité ("truthful mechanisms") reposent sur l'idée que les pourvoyeurs de données peuvent être animés par un intérêt personnel lors du report de données et qu'il leur est possible de ne pas dire la vérité, les données pouvant être sensibles dans de nombreux cas. Ainsi, le but du mécanisme est d'encourager les participants à reporter leurs données de manière fiable et honnête. En plus de garantir des estimateurs précis et la préservation de la vie privée, ces mécanismes doivent intégrer un système de compensation pour les individus basé sur leur fonction de coût de confidentialité. Cette fonction capture le niveau de préoccupation d'un individu concernant les violations de la vie privée lorsqu'il reporte ses données de manière véridique à l'analyste. De plus, la compensation doit également tenir compte de l'alignement des données rapportées par un individu avec le modèle construit à partir des données de ses pairs. Ces approches visent également à minimiser le budget total alloué lorsque le nombre d'individus augmente.

Dans notre projet, nous nous appuyons sur des travaux existants afin de relever le défi de l'analyse de données de grande dimension dans le cadre de la régression linéaire dans le cas où les individus peuvent être soucieux de leur vie privée et où on cherche donc à la préserver. Nous avons travailler à partir de deux articles, "Truthful Linear Regression" [CIL18] et "Truthful Generalized Linear Models" [QLW22] qui reprennent l'idée du mécanisme de véracité dans le cadre de la confidentialité différentielle pour le cas des modèles de régressions linéaires et linéaires généralisés, respectivement. Cependant, nous nous distinguons de ces travaux en nous concentrant sur la haute dimensionnalité. Nous cherchons à obtenir des résultats qui garantissent la protection de la vie privée, la véracité des données, la précision des estimations, la rationalité individuelle et la gestion du budget, en nous appuyant sur les idées développées dans ces deux articles.

3.1.2 Méthodes et déroulement de la mission

Après un mois de recherche et de lecture d'articles, puis suite mon à affectation au projet "Sparse High Dimensional Truthful Linear Regression", ou "Régression linéaire véridique et parcimonieuse en haute dimension" en français. J'ai entamé mon travail sur ce projet en me replongeant dans la littérature scientifique spécialisée. J'ai passé du temps à explorer des articles pertinents qui abordent les concepts clés, tels que la régression linéaire, les mécanismes de véracité, et la confidentialité différentielle. Mon objectif était de bien maîtriser ces notions

essentielles et de commencer à réfléchir à la manière dont nous pourrions les intégrer dans notre projet, tout en tenant compte des différences par rapport aux travaux précédents.

Pendant cette phase de lecture et de recherche bibliographique, je gardais en tête le problème considéré dans ce projet qui est le suivant :

Si l'on considère une base de données $X = (X_i, y_i)_{i=1}^n \in \mathcal{X}^n$ regroupant les variables descriptives $(X_i)_{i=1}^n \in (\mathbb{R}^d)^n$ et les variables cibles $(y_i)_{i=1}^n \in \mathbb{R}^n$, le but de ce projet est de concevoir un algorithme permettant de trouver un estimateur privé et précis de ω^* vérifiant :

$$\forall i \in \{1, ..., n\}, y_i = \langle X_i, \omega^* \rangle + \sigma_i = X_i^T \omega^* + \sigma_i \tag{2}$$

dans le cas où $d >> n$, avec d le nombre de caractéristiques de chaque individu et σ_i une variable de bruit de moyenne nulle.
En intégrant l'idée du mécanisme de véracité, l'algorithme devra également renvoyer le paiement que l'on devra allouer à chaque individu de sorte à minimiser le budget total et à adapter chaque paiement en fonction de l'utilité de l'individu et la véracité de ses données.

Je vais en premier lieu introduire des notions importantes tirées d'articles scientifiques que j'ai pu lire durant cette phase de compréhension du projet et que nous avons repris dans notre projet.

(a) Régression linéaire en faible dimension ($d << n$) [XLS]

Dans le cas d'une régression linéaire classique, le but est d'estimer le paramètre ω^* dans l'équation (2), ce qui revient à résoudre le problème de minimisation suivant :

$$\min_{\omega^*} \frac{1}{2} \sum_{i=1}^{n} \left(X_i^T \omega^* - y_i \right)^2 \tag{3}$$

Il s'agit ici de la régression OLS (ou des moindres carrés ordinaires). Cette méthode correspond à la minimisation de la somme des écarts quadratiques entre les valeurs observées et les valeurs prédites. Cette minimisation conduit à l'estimateur $\tilde{\omega}$ de ω^* suivant :

$$\tilde{\omega} = \left(X^T X \right)^{-1} X^T y \tag{4}$$

(b) Mécanisme de véracité [QLW22],[CIL18]

On peut avoir besoin d'utiliser un mécansime de véracité dans le cas où l'on considère des données sensibles et que les agents sont soucieux de leur vie privée. Ils peuvent donc être tentés de mentir quant à leurs données afin de préserver leur confidentialité. Dans ce cas, le mécanisme de véracité doit renvoyer, en plus de l'estimation du paramètre considéré, une liste de paiements pour chaque individu relatifs à la véracité de leurs données.

A chaque individu i est associé une fonction de coût de la vie privée : $f_i(c_i, \epsilon)$ où c_i est un paramètre de coût de confidentialité qui mesure le coût qu'il encourt lorsque ses données sont utilisées dans un mécanisme (ϵ, δ)-JDP (voir la définition 6). En outre, si l'individu i reçoit un paiement π_i, son utilité sera : $u_i = \pi_i - f_i(c_i, \epsilon)$, soit la différence entre le paiement alloué par le mécanisme et le coût de divulguer ses données.

Parmi les résultats que nous souhaitons que le mécanisme vérifie, il y a la rationnalité individuelle qui correspond au fait pour l'utilité de chaque individu doit être positive. De plus, nous souhaitons que le mécanisme guarantisse un budget asymptotique faible, c'est-à-dire que le bugdet doit tendre vers 0 lorsque le nombre d'individu tend vers l'infini. Ceci reflète le fait que plus il y a de données et moins les données d'un individu spécifique sont importantes pour établir le modèle.

Dans notre article, nous considérons la règle de paiement Brier redimensionné, utilisée dans les travaux précédents [QLW22], [CIL18], à savoir :

$$B_{a_1,a_2}(p,q) = a_1 - a_2(p - 2pq + q^2) \tag{5}$$

où $a_1, a_2 > 0$ sont des paramètres à déterminer, q est la prédiction de la réponse de l'agent i en fonction des valeurs reportées, et p est la prédiction de la réponse de l'agent i en fonction de son vecteur de caractéristiques et des reports de ses pairs.
Cette une fonction est strictement concave de q, maximisée lorsque $q = p$, c'est-à-dire lorsque la prédiction de la réponse de l'agent i en fonction de ses informations est alignée avec celle donnée par les informations de ses pairs.

Par ailleurs, on utilise deux autres notions dans le cadre des mécanismes de véracité.

Définition 1 (Stratégie de seuil)**.** *La stratégie de seuil σ_τ est définie comme suit :*

$$\hat{y}_i = \sigma_\tau(x_i, y_i, c_i) = \begin{cases} y_i, & si \quad c_i \leq \tau, \\ \text{valeur arbitraire dans } \mathcal{Y}, & si \quad c_i > \tau. \end{cases}$$

Définition 2 (Seuil $\tau_{\alpha,\beta}$)**.** *Fixons une fonction de densité de probabilité $p(c)$ du paramètre de coût de confidentialité, et définissons*

$$\tau^1_{\alpha,\beta} = \inf\{\tau > 0 : P_{(c_1,\cdots,c_n)\sim p^n}(\#\{i : c_i \leq \tau\} \geq (1-\alpha)n) \geq 1-\beta\},$$
$$\tau^2_{\alpha} = \inf\{\tau > 0 : \inf_{D_i} P_{c_j \sim p(c|D_i)}(c_j \leq \tau) \geq 1-\alpha\}.$$

Définissons $\tau_{\alpha,\beta}$ comme le plus grand de ces deux seuils : $\tau_{\alpha,\beta} = \max\{\tau^1_{\alpha,\beta}, \tau^2_{\alpha}\}$.

Remarque 2.
- $\tau^1_{\alpha,\beta}$ est un seuil tel que, avec une probabilité d'au moins $1-\beta$, au moins une fraction de $1-\alpha$ des agents ont un coefficient de coût $c_i \leq \tau_{\alpha,\beta}$,
- τ^2_{α} est un seuil tel que, conditionnellement à leur propre ensemble de données D_i, chaque agent i croit qu'avec une probabilité de $1-\alpha$, tout autre agent j a un coefficient de coût $c_j \leq \tau_{\alpha,\beta}$.

(c) Confidentialité différentielle [Vad16], [DR14],[Dwo+17] Plusieurs définitions de la confidentialité différentielle permettent d'envisager cette notion selon différentes situations.

Celle établie en (1) correspond au cas de la confidentialité différentielle pure et permet de dire que si un mécanisme est ϵ-différentiellement privé, alors on ne peut pas savoir si les

données d'un individu spécifique sont incluses ou non dans l'ensemble de données original en observant la sortie du mécanisme.

Dans l'article "Truthful Generalized Linear Models" [QLW22], la notion de confidentialité différentielle est présentée afin de rendre compte du fait que l'estimateur privé $\tilde{\omega}$ calculé par le mécanisme est un résultat observable publiquement, alors que chaque paiement π_i ne peut être observé que par l'agent i.

Définition 3 (La confidentialité différentielle conjointe)**.** *Considérons un mécanisme aléatoire $M : D^n \to \Theta \times \Pi^n$ avec des ensembles de réponses arbitraires (Θ, Π^n). Pour chaque $i \in [n]$, définissons $M(.)_{-i} = (\theta, \pi^{-i}) \in \Theta \times \Pi^{n-1}$ comme la portion de la sortie du mécanisme qui est observable par les observateurs extérieurs et les agents $j \neq i$. Alors, le mécanisme M est (ϵ, δ)-différentiellement privé conjointement (JDP) si, pour chaque agent i, chaque ensemble de données $D \in D^n$ et chaque $D'_i, D_i \in D$, nous avons :*

$$\forall S \subseteq \Theta \times \Pi^{n-1}, \mathbb{P}(M(D_i, D_{-i})^{-i} \in S | (D_i, D_{-i})) \leq e^{\epsilon} \mathbb{P}(M(D'_i, D_{-i})^{-i} \in S | (D'_i, D_{-i})) + \delta. \tag{6}$$

où δ représente la probabilité qu'un désastre arrive.

(d) Mécanismes d'ajout de bruit pour respecter la confidentialité différentielle [Vad16], [DR14],[Dwo+17]

Pour obtenir un estimateur privé, il va falloir ajouter du bruit à celui trouvé grâce à un premier algorithme d'estimation.
Deux mécanismes principaux intègrent un facteur de perturbation, qui permet à l'estimateur ainsi formulé de maintenir la confidentialité différentielle.

Définition 4 (Mécanisme de Laplace)**.** *Soit $f : \mathcal{X}^n \to \mathbb{R}^k$ une fonction définie de telle sorte que $\Delta^{(f)}$ existe, avec $\Delta^{(f)} = \max_{X \sim X'} \|f(X) - f(X')\|_1$. $\Delta^{(f)}$ est appelée la l_1-sensibilité de la fonction f.*

On définit $M : \mathcal{X}^n \to \mathbb{R}^k$, un mécanisme de Laplace, comme suit : $M(X) = f(X) + (Y_1, \ldots, Y_k)$ avec $(Y_i)_i$ une suite de variables aléatoires indépendantes identiquement distribuées de loi $\text{Lap}\left(\frac{\Delta}{\epsilon}\right)$, Δ étant la l_1-sensibilité de f et $\epsilon > 0$. $(Y_1, \ldots, Y_k)$ est appelée un bruit laplacien.

M est un mécanisme ϵ-différentiellement privé.

Définition 5 (Mécanisme gaussian)**.** *Soit $f : \mathcal{X}^n \to \mathbb{R}^k$. On considère $\Delta_2^{(f)}$ la l_2-sensibilité de la fonction f, avec $\Delta^{(f)} = \max_{X \sim X'} \|f(X) - f(X')\|_2$.*

On dit que $M : \mathcal{X}^n \to \mathbb{R}^k$ est un mécanisme gaussien si : $M(X) = f(X) + (Y_1, \ldots, Y_k)$ avec $(Y_i)_i$ une suite de variables aléatoires indépendantes identiquement distribuées de loi $\mathcal{N}\left(0, \frac{\Delta_2^2}{\epsilon^2} \log\left(\frac{1}{\delta}\right)\right)$, avec $\epsilon > 0$ et $\delta > 0$. $(Y_1, \ldots, Y_k)$ est appelée un bruit gaussien.

M est un mécanisme (ϵ, δ)-différentiellement privé.

Plusieurs réunions avec le professeur Di Wang et mon collègue Liyang ont été organisées pour discuter des hypothèses d'étude, de la forme du mécanisme et des problèmes rencontrés.

Un des premiers enjeux était de trouver des méthodes permettant de traiter la haute dimensionalité des données. Mon collègue Liyang, qui était déjà expérimenté dans le domaine (puisque ce projet était dans la continuité de son projet de thèse sur lequel il a travaillé jusqu'à mi-mai), m'expliqua trois méthodes que l'on allait utiliser : la troncature, la projection et la notion de sparsité.

Hypothèses 1. *Voici les hypothèses que nous avons émises concernant les différents paramètres du problème, suite à nos différentes réflexions et réunions avec le professeur :*

- ω^* *est s-clairsemé, c'est-à-dire que* $\|\omega^*\|_0 \leq s$*, soit qu'il possède au plus s composantes non nulles, avec* $0 < s \ll d$*.*
- *La norme 1 du vecteur* ω^* *est bornée par une constante que l'on prendra égale à 1 :* $\|\omega^*\|_1 \leq 1$*.*
- *Pour tout agent i, la fonction coût* f_i *vérifie :* $f_i(c_i, \epsilon) \leq c_i\epsilon^2$*.*
- *La norme 1 du vecteur caractéristique de chaque agent i est bornée par une constante :* $\|X_i\|_1 \leq O(1)$*.*
- *ainsi que les hypothèses 3, 4 et 5 de l'article "Truthful Generalized Linear Models", [QLW22].*

En s'inspirant des mécanismes des deux articles existants (présentés précedemment) et du mécanisme que Liyang avait utilisé dans son précédent projet, on a pu imaginé un algorithme utilisant les notions de troncature, de bruit gaussien, de confidentialité différentielle jointe et d'estimateur OLS. Il est introduit dans la prochaine section. Toutefois l'algorithme d'estimation du paramètre repose sur les principales idées suivantes :

- Prétraitement des données pour gérer la haute dimensionnalité
- Calcul de quantités essentielles dans l'estimation OLS à partir des données prétraitées : $XX^T = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$ et $XY = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$
- Intégration d'un bruit gaussien indépendant à chacune des quantités précédentes

Par la suite, ma mission était de réfléchir aux performances du mécanisme en terme de préservation de la vie privée, de rationnalité individuelle, de budget et de véracité ; Liyang s'occupait du résultat sur la précision.

Pour trouver les résultats voulus et surtout les méthodes pour y parvenir, j'ai lu en profondeur les preuves des résultats des articles [QLW22] et [CIL18]. J'ai ensuite cherché des résultats similaires en prenant en compte les différences avec les hypothèses et les paramètres bien spécifiques de notre projet.

Au début de ce projet, je travaillais plutôt seule car mon collègue Liyang travaillait essentiellement sur le projet lié à sa thèse jusqu'à mi-mai. Cependant, lui et le professeur étaient toujours disponibles pour répondre à mes interrogations et m'aider à avancer en cas de difficulté.

Par la suite, lors de la réunion de groupe qui s'est déroulée mi-avril, j'ai demandé à pouvoir travailler sur un nouveau projet, en lien avec les bandits multi-bras. En effet, je commençais

à ressentir une certaine monotonie dans le cadre de mon travail sur ce projet exclusivement, et je souhaitais diversifier mes tâches et m'investir dans un tout nouveau projet. À partir de ce moment, j'ai commencé à m'investir de manière simultanée dans les deux projets.

En ce qui concerne ce premier projet, j'ai terminé de rédiger les preuves des théorèmes, j'ai également participé à la rédaction l'introduction et la partie sur les travaux apparentés, pendant que mon collègue s'occupait de développer la partie explicative du mécanisme et le théorème concernant la précision de l'estimateur.

3.1.3 Réponse apportée au problème posé et résultats théoriques

Les algorithmes décrivant le mécanisme établi par Liyang et moi-même sont les suivants :

Algorithm 1 (ε, δ)-DP Algorithm for Sparse Linear Regression

1: **Input :** Private data $\{(x_i, y_i)\}_{i=1}^n \in \left(\mathbb{R}^d \times \mathbb{R}\right)^n$. Predefined parameters τ_y, λ_n.
2: **for** each user $i \in [n]$ **do**
3: Release $x_i x_i^T$ to the server.
4: Clip $\tilde{y}_i := \text{sgn}(y_i) \min\{|y_i|, \tau_y\}$. Release $x_i \tilde{y}_i$ to the server.
5: **end for**
6: The server aggregates $\hat{\Sigma}_{XX} = \frac{1}{n}\sum_{i=1}^n x_i x_i^T$ and $\hat{\Sigma}_{X\widetilde{Y}} = \frac{1}{n}\sum_{i=1}^n \widehat{x_i \tilde{y}_i}$.
7: Add noise $\dot{\Sigma}_{XX} = \hat{\Sigma}_{XX} + N_1$, where $N_1 \in \mathbb{R}^{d\times d}$ is a symmetric matrix and each entry of the upper triangular matrix is sampled from $\mathcal{N}(0, \frac{32r^4 \log\frac{2.5}{\delta}}{n^2\varepsilon^2})$. Add noise $\dot{\Sigma}_{X\widetilde{Y}} = \hat{\Sigma}_{X\widetilde{Y}} + N_2$, where the vector $N_2 \in \mathbb{R}^d$ is sampled from $\mathcal{N}(0, \frac{32r^2\tau_y^2 \log\frac{2.5}{\delta}}{n^2\varepsilon^2} I_d)$.
8: Apply hard-thresholding to $\dot{\Sigma}_{XX}$ and obtain $\ddot{\Sigma}_{XX}$ where each entry is defined as $\ddot{\sigma}_{xx^T,ij} = \dot{\sigma}_{xx^T,ij} \cdot I[|\dot{\sigma}_{xx^T,ij}| > \mathit{Thres}]$, where $\mathit{Thres} = \gamma\sqrt{\frac{\log d}{n}} + \frac{4r^2\sqrt{2\ln 1.25/\delta}\sqrt{\log d}}{n\varepsilon}$ and γ is some constant depending on σ_{N_1}.
9: The server outputs $\hat{\theta}^P(D) = S_{\lambda_n}([\ddot{\Sigma}_{XX}]^{-1}\dot{\Sigma}_{X\widetilde{Y}})$.

Cet algorithme permet de calculer un estimateur privé du paramètre sous-jacent au modèle de régression linéaire en prenant en compte la haute dimensionalité des données et en utilisant donc des méthodes adaptées comme la troncature et la projection. L'ajout de bruit par le biais du mécanisme gaussien permet de conserver la confidentialité différentielle de l'estimateur.

Ce deuxième algorithme renvoie les paiements de chaque individus calculés selon la règle de Brier 5. On reprend l'idée utilisée dans les travaux précédents afin que le mécanisme incite les individus à reporter de manière véridique leurs données.

Voici les résultats satisfaits par le mécanisme ainsi défini, en terme de confidentialité, de véracité, de rationalité individuelle, de budget et de précision.

Théorème 1 (Confidentialité)**.** *La sortie de l'Algorithme 2 satisfait la* $(2\varepsilon, 3\delta)$*-JDP.*

Théorème 2 (Véracité)**.** *Soient* $\varepsilon > 0$ *un paramètre de confidentialité,* $1-\alpha$ *un objectif de participation et* β *un paramètre de confiance souhaité. Alors, avec une probabilité d'au moins* $1 - \beta - O(n^{-\Omega(1)}) - nd^{\frac{9}{2}}$*, la stratégie de seuil symétrique* $\sigma_{\tau_{\alpha,\beta}}$ *est un équilibre de Nash*

Algorithm 2 General Framework for Truthful and Private High Dimensional Linear Regression

1: Ask all agents to report their data $\hat{D}_1, \cdots, \hat{D}_n$;
2: Randomly partition agents into two groups, with respective data pairs $\hat{D}^0, \hat{D}^1$
3: For each dataset $\hat{D}, \hat{D}^0$ and $\hat{D}^1$, compute private estimators $\hat{\theta}^P(\hat{D})$, $\hat{\theta}^P(\hat{D}^b)$ for $b = 0, 1$ according to Algorithm 1
4: Compute estimators $\bar{\theta}^P(\hat{D}) = \Pi_{\tau_\theta}(\hat{\theta}^P(\hat{D}))$ and $\bar{\theta}^P(\hat{D}^b) = \Pi_{\tau_\theta}(\hat{\theta}^P(\hat{D}^b))$ for $b = 0, 1$
5: Set parameters a_1, a_2, and compute payments to each agent i : if agent i's is in group $1 - b$, then he will receive payment

$$\pi_i = B_{a_1,a_2}\left(\langle x_i, \bar{\theta}^P(\hat{D}^b)\rangle, \langle x_i, E_{\theta\sim p(\theta|\hat{D}_i)}[\theta]\rangle\right).$$

bayésien approximatif η dans le mécanisme de régression linéaire privée 2 avec

$$\eta = O\left(\frac{a_2 n\alpha^2 r^6 k \log d \log\frac{1}{\delta}}{\varepsilon^2} + \tau_{\alpha,\beta}\varepsilon^2\right).$$

Théorème 3 (Rationalité Individuelle)**.** *Avec une probabilité d'au moins $1-\beta-O(n^{-\Omega(1)})-nd^{-\frac{9}{2}}$, le mécanisme est individuellement rationnel pour tous les agents avec des coefficients de coût $c_i \leq \tau_{\alpha,\beta}$ tant que*

$$a_1 \geq a_2(r\tau_\theta + 3r^2\tau_\theta^2) + \tau_{\alpha,\beta}4(1+\delta)\varepsilon^2$$

où a_1, a_2 sont introduits dans la règle de paiement 5.

Remarque 3. Notre mécanisme peut ne pas être individuellement rationnel pour tous les agents. Cela est dû au fait que nous supposons que les coûts liés à la confidentialité suivent une distribution particulière. Cette supposition est fondée sur l'idée qu'un petit groupe de personnes pourraient avoir des coûts de confidentialité très élevés. En d'autres termes, il y a des individus qui ne voudront peut-être pas partager leurs informations, même si on leur offre une compensation.

Théorème 4 (Budget)**.** *Avec une probabilité d'au moins $1-\beta-O(n^{-\Omega(1)})-nd^{-\frac{9}{2}}$, le budget total attendu $\mathcal{B} = \sum_{i=1}^n \mathbb{E}(\pi_i)$ requis par l'analyste pour exécuter le mécanisme sous la stratégie d'équilibre de seuil $\sigma_{\tau_{\alpha,\beta}}$ satisfait*

$$\mathcal{B} \leq n(a_1 + a_2(r\tau_\theta + r^2\tau_\theta^2)).$$

Remarque 4. Avec un réglage approprié du paramètre α, il est raisonnable de s'attendre à ce que le budget global attendu diminue vers zéro à mesure que la taille de l'échantillon augmente. Cela est dû au fait que si la taille de l'échantillon est infinie, chaque agent a une contribution minime, l'analyste n'a donc pas besoin de payer pour inciter les agents.

Le résultat ayant été développé par Liyang est surtout celui de la précision de l'estimateur, que voici :

Théorème 5 (Précision)**.** *Soit $\varepsilon > 0$ un paramètre de confidentialité, $1-\alpha$ un objectif de participation et β un paramètre de confiance souhaité dans la définition 2. Alors, avec la stratégie de seuil symétrique $\sigma_{\tau_{\alpha,\beta}}$, la sortie $\bar{\theta}^P(\hat{D})$ de l'algorithme 2 satisfait que, avec une probabilité d'au moins $1-\beta-O(n^{-\Omega(1)})$,*

$$\mathbb{E}[\|\bar{\theta}^P(\hat{D}) - \theta^*\|_2^2] \leq O\left(kn\alpha^2 \frac{r^4 \log d \log \frac{1}{\delta}}{\varepsilon^2}\right)$$

3.1.4 Conclusions et implication personnelle

Durant cette première partie de mon stage, j'ai plongé dans l'univers complexe de la régression linéaire en haute dimension, de la confidentialité différentielle et des mécanismes de véracité. Mon objectif était de participer à la conception d'un mécanisme innovant qui permettrait de résoudre le problème de la régression linéaire dans des situations où la dimension des données dépasse largement la taille de l'échantillon ; ceci tout en garantissant la confidentialité et la véracité des participants.

Pour atteindre cet objectif, j'ai étudié en profondeur les différentes notions clées, en m'appuyant sur des articles de recherche pertinents. J'ai collaboré avec mon collègue Liyang et le professeur Di Wang pour formuler des hypothèses spécifiques et développer un mécanisme qui répondrait à ces défis. Ensemble, nous avons conçu un algorithme qui intègre des techniques de troncature, de bruit gaussien, et de confidentialité différentielle conjointe pour obtenir un estimateur privé précis.

Mon implication personnelle dans ce projet a été significative. J'ai participé à la réflexion sur la conception du mécanisme, contribué à la rédaction de l'introduction et de la partie sur les travaux apparentés, participé à la recherche des performances du mécanisme en termes de vie privée, de rationnalité individuelle, de budget, et de véracité, et rédigé les preuves des théorèmes associés. Mon apprentissage a été enrichissant, car j'ai acquis une compréhension plus profonde de ces différents concepts et j'ai eu la possibilité de contribuer activement à la recherche en participant au développement d'un mécanisme original.

Dans la suite de ce rapport, je vais présenter une deuxième partie de mon stage, qui m'a permis de travailler sur un autre projet passionnant. Ce nouveau projet m'a apporté une perspective différente et m'a permis d'élargir mes compétences tout en continuant à contribuer au domaine de l'apprentissage automatique.

3.2 Projet 2 : Bandits robustes et privés avec alpha-contamination

3.2.1 Problématique et exploration des résultats existantes

Dans le cadre de ce deuxième projet, j'ai eu l'occasion de collaborer avec une étudiante en doctorat Yulian, qui travaille essentiellement sur les problèmes de bandits multi-bras dans le cadre de la préservation de la confidentialité en Machine Learning.

Elle m'a invité à prendre part à un projet déjà commencé, intitulé : "Bandits robustes et privés avec alpha-contamination".

En premier lieu, ce projet fait intervenir le concept de bandits multi-bras, qui peut être considéré comme un élement essentiel en intelligence artificielle et surtout dans la prise de décision séquentielle. Ils permettent de gérer des situations dans lesquelles nous devons choisir parmi plusieurs options sans connaître à l'avance les récompenses qui y sont associées ; le but étant de maximiser les récompenses cumulatives totales. Ils trouvent des applications dans la recommandation de produits, l'optimisation des campagnes publicitaires en ligne, la personnalisation des soins de santé et à la recommandation de contenu.

Ces algorithmes permettent d'équilibrer l'exploration (essayer de nouvelles actions pour apprendre) et l'exploitation (choisir les actions jugées les plus prometteuses), en vue de maximiser la récompense totale. Les bandits multi-bras sont ainsi un outil puissant pour améliorer l'efficacité de la prise de décision dans divers domaines de notre société moderne.

Les bandits multi-bras robustes étendent le concept des bandits multi-bras classiques pour prendre en compte l'incertitude qui se présente lorsque les données sont bruitées ou incertaines, mais aussi lorsque les récompenses potentielles des actions peuvent évoluer avec le temps en réponse à des facteurs inconnus.
Pour modéliser le fait que certaines données peuvent être erronées ou corrompues et produire des algorithmes robustes à ce type d'incertitudes, on utilise le modèle de "forte contamination" ou "strong contamination". C'est un modèle de contamination robuste utilisé pour étudier la confidentialité différentielle dans le contexte des bandits. Il permet de prendre en compte des erreurs introduites de manière stratégique par un adversaire qui a une connaissance des données non corrompues. Cela signifie que l'adversaire peut introduire des erreurs de manière sélective.

Tous ces concepts sont étudiés en relation avec la confidentialité différentielle. En effet, le but de cet article est de développer des algorithmes de bandits multi-bras robustes et privés, tout en montrant comment la confidentialité et la robustesse affectent les performances du mécanisme.

Tout au long de ce projet, j'ai essentiellement travaillé à partir de deux articles : "On Private and Robust Bandits", [Wu+23] et "Distributed Differential Privacy in Multi Armed-Bandits", [CZ22].

3.2.2 Méthodes et déroulement de la mission

Comme introduit précédemment, j'ai eu l'opportunité de travailler sur un projet lié aux bandits multi-bras en collaboration avec le professeur Di Wang et ma collègue Yulian. Après avoir exprimé mon intérêt pour ce domaine de recherche, Yulian m'a intégrée à un de ses projets, qui portait sur la confidentialité différentielle dans le contexte des bandits multi-bras robustes. À ce stade, elle avait déjà commencé à travailler sur ce projet d'article et obtenu des résultats pour deux des trois modèles de confidentialité différentielle considérés, à savoir la confidentialité différentielle centrale (CDP) et la confidentialité différentielle locale (LDP). Mon rôle consistait à étendre ces résultats au troisième modèle, celui de confidentialité différentielle mélangée ou "shuffle differential privacy" (SDP).

J'ai abordé ce projet en m'appuyant sur des connaissances préalables acquises grâce à mes lectures d'articles sur les bandits multi-bras et sur la SDP. J'ai rapidement réalisé que, bien que je maîtrisais déjà les concepts de base, la complexité du nouveau problème reposait sur le fait qu'il fallait relier divers éléments issus de plusieurs articles. Notamment, j'ai consulté les articles recommandés par Yulian et j'ai poursuivi mes propres recherches pour mieux comprendre les notions nécessaires à ce projet. Lors de cette phase de préparation, j'ai notamment eu recours à la lecture de l'article "Best Arm Identification for Contaminated Bandits", [ABM19], pour approfondir la notion de "strong contamination" qui s'avérait cruciale pour la compréhension du projet.

Je vais introduire quelques notions clés tirées d'articles scientifiques qu'il m'a fallut bien comprendre et que nous avons utilisé pour notre article.

(a) <u>Les modèles de confidentialité différentielle</u>

Définition 6 (Confidentialité différentielle centrale (CDP), [Dwo+06])**.** *Un algorithme aléatoire $\mathcal{A}$ est (ϵ, δ)-différentiellement privé (confidentialité différentielle centrale) si pour tous les ensembles de données voisins $D \sim D'$ et pour tous les événements S dans l'espace de sortie de $\mathcal{A}$, nous avons*

$$\Pr(\mathcal{A}(D) \in S) \leq e^{\epsilon} \Pr\left(\mathcal{A}\left(D'\right) \in S\right) + \delta,$$

lorsque $\delta = 0$ et que $\mathcal{A}$ est ϵ-différentiellement privé (confidentialité différentielle centrale pure).

Définition 7 (Confidentialité différentielle locale (LDP), [**duch**])**.** *Un algorithme $\mathcal{M} : \mathcal{X} \to \mathcal{Y}$ est ϵ-localement différentiellement privé (LDP) si pour toute paire de valeurs d'entrée $x, x' \in \mathcal{X}$, et tout sous-ensemble mesurable $\mathcal{O} \subset \mathcal{Y}$, il est vrai que $\mathbb{P}\left[\mathcal{M}(x) \in \mathcal{O}\right] \leq e^{\epsilon} \cdot \mathbb{P}\left[\mathcal{M}(x') \in \mathcal{O}\right]$.*

Par rapport à la CDP, la confidentialité différentielle locale (LDP) offre une meilleure garantie de confidentialité, car elle protège la vie privée de l'utilisateur au niveau des données individuelles avant toute agrégation de la part de l'analyseur.

Voici un schéma repris de l'article [Fal+23] permet d'illustrer les notions de confidentialité différentielle décrites précédemment.

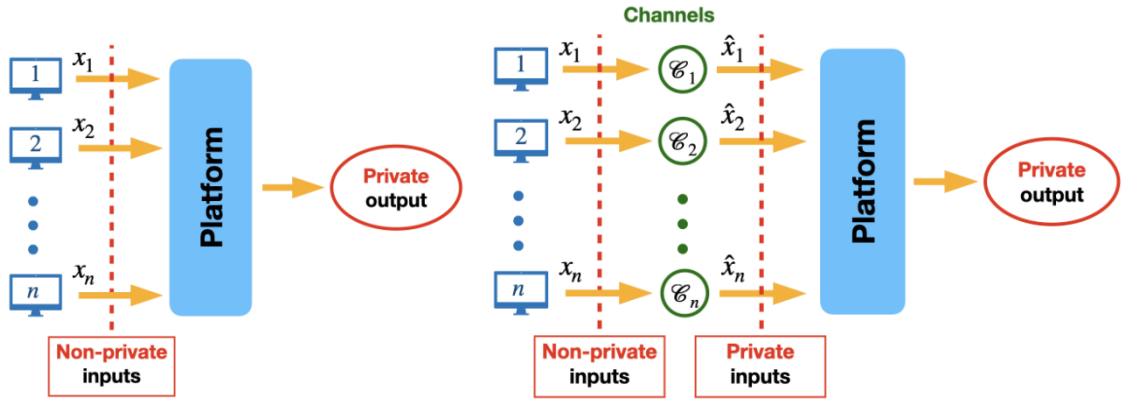

FIG. 3 – (a) le cadre central et (b) le cadre local.

Dans le cadre central, les individus ont confiance à l'agrégateur/analyseur et lui partagent leurs données non privatisées. L'analyseur ajoute ensuite du bruit à l'estimateur trouvé afin de respecter la confidentialité différentielle à la sortie.

Au contraire, dans le cadre local, les utilisateurs privatisent leurs données avant de les partager avec l'analyseur, qui estime ensuite le paramètre souhaité à partir des données bruitées.

Le modèle de "shuffle differential privacy" reprend ces deux idées. [Bal+19]

Définition 8 (Confidentialité différentielle mélangée (SDP), [CZ22])**.** *Un protocole distribué* $\mathcal{P} = (\mathcal{R}, \mathcal{S}, \mathcal{A})$ *se compose de trois parties :*

- *un randomiseur (local)* $\mathcal{R}$ *du côté de chaque utilisateur*
- *un protocole de mélange intermédiaire* $\mathcal{S}$
- *un analyseur* $\mathcal{A}$ *au niveau du serveur central*

Dans ce schéma, chaque utilisateur applique initialement le randomiseur $\mathcal{R}$ *localement à ses données brutes* D_i*. Les données randomisées sont ensuite transmises à un protocole de calcul aléatoire* $\mathcal{S}$*. Ce protocole intermédiaire génère des entrées pour le serveur central, qui, à l'aide d'un analyseur* $\mathcal{A}$*, calcule la sortie en utilisant les messages reçus de* $\mathcal{S}$*.*
Un protocole $\mathcal{P} = (\mathcal{R}, \mathcal{S}, \mathcal{A})$ *est considéré comme satisfaisant à la* ϵ*-SDP si le mécanisme de composition* $\mathcal{M}_{\mathcal{P}} = \mathcal{S} \circ \mathcal{R}$ *répond à la définition 6.*

Dans ce modèle, les individus partagent leurs données avec une première plateforme sécurisée, qui mélange aléatoirement les données, perturbant ainsi l'ordre d'origine. Ensuite, ces données mélangées sont transmises à une deuxième plateforme qui ajoute du bruit à chacune des données de manière contrôlée.

Enfin, l'analyseur traite la base de données mélangée pour obtenir l'estimateur du paramètre souhaité. Ce modèle permet l'utilisation d'un bruit moins important que dans le modèle de LDP, grâce au mélange aléatoire des données introduisant un élément d'aléatoire supplémentaire.

L'ordre des étapes peut varier, mais l'essentiel est que le mélange et l'ajout de bruit sont effectués avant que les données ne soient transmises à l'analyseur, qui est considéré comme non sécurisé.

Pour en revenir au déroulement de la mission, ma première partie de travail consistait à examiner les résultats déjà obtenus pour les modèles de CDP et LDP et d'analyser les hypothèses sous-jacentes. Pour se faire j'ai également pris l'initiative d'intégrer les mécanismes utilisés dans le cas de ces deux modèles à l'article, car certains algorithmes n'y étaient pas. Cette démarche m'a permis de mieux comprendre la base sur laquelle je devais construire le mécanisme pour le modèle de SDP. Cette première approche avait pour but de me permettre de comprendre la notion de confidentialité différentielle dans le cadre des bandits multi-bras avec des données contaminées.

Au cours de ce projet, j'ai principalement travaillé en autonomie pour concevoir le mécanisme du modèle de SDP. Ma collègue Yulian était disponible pour répondre à mes interrogations, mais j'ai privilégié une approche autonome. Cette démarche m'a permis de me sentir plus impliquée et de mieux assimiler les diverses notions de l'article sur lequel nous travaillions. J'ai cherché à comprendre en profondeur chaque section et résultat de l'article, car ayant commencé à travailler sur ce projet en cours de route, j'avais à cœur de bien comprendre chaque aspect, afin de ne pas me sentir freinée par mon arrivée tardive sur le projet.

Mon exploration a commencé par la notion de confidentialité différentielle dans le contexte des bandits multi-bras. J'ai ensuite plongé dans la littérature pour découvrir l'origine des résultats relatifs aux modèles CDP et LDP, en m'efforçant de comprendre les mécanismes employés. J'ai examiné les articles précédents traitant de ces mêmes notions, cherchant à saisir les subtilités de leurs méthodes et comment les résultats avaient été obtenus, notamment en ce qui concerne la taille des lots et les bornes supérieures du regret (dont je parlerais dans la suite).

Voici d'autres notions importantes afin de comprendre le projet et les résultats recherchés.

(b) Bandits multi-bras sous contamination

Définition 9 (Bandits multi-bras [CZ22])**.** *Dans un problème stochastique de bandits à bras multiples (MAB), il y a un apprenant qui interagit avec un environnement de manière séquentielle sur T tours. L'apprenant est confronté à un ensemble de K bras indépendants $\{1, \dots, K\}$. À chaque tour $t \in [T]$, l'apprenant choisit un bras $a_t \in [K]$ à tirer, puis obtient une récompense r_t tirée de manière i.i.d. à partir d'une distribution de probabilité fixe mais inconnue à queues lourdes P_{a_t} associée au bras choisi et satisfaisant (9). Désignons par μ_a la moyenne de chaque distribution P_a pour $a \in [K]$, et par $\mu^* = \max_{a \in [K]} \mu_a$ le maximum. On définit $\Delta_a \triangleq \mu^* - \mu_a$ comme l'écart de récompense moyen pour le bras a. L'apprenant cherche à maximiser sa récompense cumulative attendue au fil du temps,* i.e., *à minimiser le regret, défini comme :*

$$\mathcal{R}_T \triangleq T\mu^* - \mathbb{E}\left[\sum_{t=1}^{T} r_t\right], \tag{7}$$

où l'espérance est prise par rapport à tous les éléments aléatoires.

Définition 10 (Modèle de contamination forte [Dia18]). *Étant donné un paramètre* $0 < \alpha < 1$ *et une distribution de valeurs intermédiaires* P*, un algorithme reçoit des échantillons de* P *et renvoie le même nombre d'échantillons avec une contamination de* α*-fraction comme suit : l'algorithme spécifie un nombre entier d'échantillons* n*, et* n *échantillons sont tirés indépendamment de la distribution* P*. Un adversaire est ensuite autorisé à inspecter ces échantillons, à en retirer jusqu'à* αn *et à les remplacer par des points arbitraires. L'ensemble modifié de* n *points est alors renvoyé à l'algorithme.*

Pour reprendre le modèle de bandits multi-bras et y ajouter le modèle de contamination forte α introduite précédemment, nous autorisons un adversaire à corrompre les récompenses r_t de n'importe quelle manière tant que pour chaque temps t il n'y a pas plus d'une fraction α de récompenses contaminées pour n'importe quelle action. En d'autres termes, si nous désignons par $N_a(t)$ le nombre de fois où l'action choisie a a été jouée au cours du tour t, nous contraignons l'adversaire à ce que :

$$\forall a \in [K], \forall t \in [T], \sum_{t=1}^{N_a(t)} \mathbb{I}\{r_a(t) \neq x_a(t)\} \leq \alpha N_a(t) \tag{8}$$

L'apprenant observe les récompenses contaminées par l'adversaire et doit prendre en compte cette contamination éventuelle dans le problème de maximisation de sa récompenses cumulative attendue.

Après les premières recherches de mon côté, Yulian et moi avons discuté des hypothèses d'étude.

(c) Hypothèses d'étude

Soit $\mathcal{P}$ une classe de distributions sur un espace d'échantillonnage Ω et soit $\mathcal{G}$ l'ensemble de toutes les distributions sur Ω. L'article étudie les distributions de récompenses soutenues sur la ligne réelle, c'est-à-dire $\Omega = \mathbb{R}$ avec des moments d'ordre k finis et bornés par une constante que l'on prend égale à 1.

$$\mathcal{P} = \mathcal{P}_k = \{P := \mathbb{E}_{X \sim P}\left[|X|^k\right] \leq 1\}, \tag{9}$$

où $k \geq 2$. On parle de moments bruts finis d'ordre k.

Nous considérons de plus le modèle de contamination α : étant donné n échantillons $i.i.d$, $\{X_i'\}_{i=1}^n$ de vraies distributions P, au plus αn d'entre eux sont contaminés par une distribution arbitraire. En d'autres termes, les échantillons après contamination sont $\{X_i\}_{i=1}^n$ où pour au moins $(1-\alpha)n$ choix de i, $X_i' = X_i$. Le moment d'ordre k des échantillons est fini ($k \geq 2$), et est limité par une certaine constante (disons 1) et la vraie moyenne $\mathbb{E}_{X' \sim P} = \mu \in [-1, 1]$.

(d) Les mécanismes et les théorèmes

Pour mieux comprendre les mécanismes sur lesquels Yulian avait travaillé jusqu'alors, je me suis plongée dans les deux articles [Wu+23] et [CZ22]. En effet, lorsqu'elle me donna accès au projet, il y avait seulement les résultats, mais sans les explications détaillées ni les

Algorithm 3 Mean estimation under CDP for the finite raw moment case

Input : A collection of data $\{x_i\}_{i=1}^n$, truncation parameter M, additional parameters $\Phi = \{\varepsilon\}$

for $i = 1, 2, \ldots, n$ **do**

Truncate data $\bar{x}_i = x_i \cdot \mathbb{I}_{\{|x_i| \leq M\}}$

end for

Return private estimate $\widetilde{\mu} = \frac{\sum_{i=1}^n \bar{x}_i}{n} + \text{Lap}\left(\frac{2M}{n\varepsilon}\right)$

méthodes utilisées pour les obtenir. Mon objectif était de saisir les concepts et les résultats découverts, en cherchant à comprendre le "comment" derrière les résultats, notamment les preuves et les algorithmes sous-jacents.

Des deux articles cités précédemment, j'ai tiré les deux algorithmes suivants d'estimation de la moyenne sous CDP et LDP dans le cas des moments bruts finis d'ordre k.

Algorithm 4 Mean estimation under LDP for the finite raw moment case

Input : A collection of data $\{x_i\}_{i=1}^n$, truncation parameter M, additional parameters $\Phi = \{\varepsilon\}$

for $i = 1, 2, \ldots, n$ **do**

Truncate data $\bar{x}_i = x_i \cdot \mathbb{I}_{\{|x_i| \leq M\}}$

Perturb the truncated data : $\widetilde{x}_i = \bar{x}_i + \text{Lap}(\frac{2M}{\epsilon})$

end for

Return private estimate $\widetilde{\mu} = \frac{1}{n}\sum_{i=1}^n \widetilde{x}_i$

Ces deux algorithmes permettent d'estimer la moyenne des données collectées en respectant les deux modèles de confidentialité considérés, dans le cas des moments bruts finis d'ordre k.

On peut remarquer que les mécanismes ajoutent du bruit Laplacien, celui-ci étant ajouté à l'estimateur de la moyenne dans le cas du modèle de CDP et à chaque donnée dans le cas du modèle de LDP, afin de respecter les définitions respectives des deux types de confidentialité. Ces deux mécanismes satisfont respectivement les définitions de ϵ-CDP et ϵ-LDP.

En ce qui concerne le paramètre de troncature M, celui-ci a été explicité de manière à minimiser la différence entre l'estimateur et la vraie valeur de la moyenne.

Théorème 6 (Estimation Privée et Robuste de la Moyenne sous la CDP)**.** *Pour les données x_i de moments bruts finis d'ordre k et $\epsilon \in]0, 1]$, nous concevons une estimation privée et robuste de la moyenne sous la confidentialité différentielle centrale en utilisant le mécanisme Laplacien dans l'algorithme 3 : $\tilde{\mu} = \frac{\sum_{i\in[n]} \bar{x}_i}{n} + Lap(\frac{2M}{n\epsilon})$. Ensuite, avec une probabilité d'au moins $1 - 2\delta$*

$$|\tilde{\mu} - \mu| \leq \sqrt{\frac{2\log(2/\delta)}{n}} + \frac{1}{M^{k-1}} + 2\alpha M + \frac{4M\log(2/\delta)}{\epsilon n} \tag{10}$$

Nous choisissons $M = \min((\frac{n\epsilon}{4\log(1/\delta)})^{1/k}, (2\alpha)^{-1/k})$ et nous obtenons que l'estimation privée

et robuste de la moyenne sous LDP est bornée par

$$|\tilde{\mu} - \mu| \leq \sqrt{\frac{2\log(2/\delta)}{n}} + 2\left(\frac{4\log(1/\delta)}{n\epsilon}\right)^{1-1/k} + 2(2\alpha)^{1-1/k}. \tag{11}$$

Théorème 7 (Estimation Privée et Robuste de la Moyenne sous la LDP)**.** *Pour les données x_i de moments bruts finis d'ordre k et $\epsilon \in (0, 1]$, nous concevons une estimation privée et robuste de la moyenne sous la confidentialité différentielle locale en utilisant le mécanisme Laplacien dans l'algorithme 4. Ensuite, avec une probabilité d'au moins $1 - 2\delta$,*

$$|\tilde{\mu} - \mu| \leq \sqrt{\frac{2\log(2/\delta)}{n}} + \frac{4M\log(2/\delta)}{3n} + \frac{1}{M^{k-1}} + 2\alpha M + \frac{4M\sqrt{\log(2/\delta)}}{\epsilon\sqrt{n}} \tag{12}$$

Nous choisissons :

$$M = \min((3n/4\log(1/\delta))^{1/k}, \left(\frac{\sqrt{n}\epsilon}{4\sqrt{\log(1/\delta)}}\right)^{1/k}, (2\alpha)^{-1/k}),$$

et nous obtenons que l'estimation privée et robuste de la moyenne sous LDP est bornée par

$$\sqrt{\frac{2\log(2/\delta)}{n}} + 2(2\alpha)^{1-\frac{1}{k}} + 2\left(\frac{4\sqrt{\log(2/\delta)}}{\epsilon\sqrt{n}}\right)^{1-\frac{1}{k}} + 2\left(\frac{4\log(2/\delta)}{3n}\right)^{1-\frac{1}{k}} \tag{13}$$

De plus, des résultats introduisant la borne supérieure du regret dans les deux modèles ont été trouvé par Yulian avant mon affectation au projet. Toutefois, le mécanisme permettant d'étudier le problème de bandits multi-bras n'était pas bien détaillé et j'ai du y réflechir de mon côté.

En analysant le travail précédent, j'ai pu mieux comprendre ma mission principale. Celle-ci était d'établir deux algorithmes, un permettant d'estimer la moyenne des données collectées de manière à respecter la confidentalité dans le cas du modèle de SDP ; et un permettant d'éliminer les bras non viables dans le cas du problème de bandits multi-bras. Deux résultats étaient à trouver également, le premier concernant le paramètre de troncature utilisé dans l'algorithme d'estimation de la moyenne et le second étant la borne supérieure du regret.

Je me suis inspirée des mécanismes et résultats précédents ainsi que de leurs preuves pour trouver le même genre de résultats dans le cas du modèle de SDP, utilisant des résultats sur les variables aléatoire Polya ou encore l'inégalité de Hoeffding. Par ailleurs, j'ai lu et tiré inspiration des deux articles [Wu+23] et [CZ22].

Cependant, je n'ai pas été épargnée par des défis au cours de ce projet, notamment dans la recherche d'algorithmes. Dans la conception du mécanisme, je n'ai pas compris comment formuler le problème dès le début. J'avais envie d'essayer de comprendre le problème toute seule et d'en discuter ensuite avec Yulian en lui présentant mes premières idées. Au début, j'ai eu une certaine confusion quant à l'objectif de l'algorithme. Mon interprétation initiale était centrée sur le calcul de la moyenne des récompenses des bras viables. Je croyais que l'algorithme consistait à choisir des bras viables à l'aide d'un algorithme de sélection, puis à estimer la moyenne en effectuant un certain nombre de tirages de ces bras. Cependant, ce

que l'on voulait c'était un algorithme qui renvoyait le bras optimal établi après exploration des différentes actions, et élimination des bras non viables/prometteurs à chaque tour. J'avais toutefois compris les principales idées, les principaux éléments à intégrer à l'algorithme et trouver les articles dont on allait s'inspirer pour notre travail.

Au fil de mon analyse, j'ai également saisi comment l'algorithme parvient à éliminer les bras non prometteurs. Il utilise l'algorithme de calcul de la moyenne pour chaque bras viable, ce qui permet d'estimer la qualité de chaque bras en termes de récompenses potentielles. Ensuite, en se basant sur ces estimations, l'algorithme identifie les bras qui ont des estimations de moyenne significativement plus faibles que la meilleure estimation et les élimine progressivement de l'ensemble de bras viables. Concrètement, lorsque la différence entre l'estimation de la moyenne du bras le plus prometteur (pour le moment) et celle d'un autre bras excède un certain seuil, ces bras moins prometteurs sont retirés de l'ensemble de bras viables. Ce processus d'élimination séquentielle se poursuit à chaque tour jusqu'à ce qu'un nombre prédéfini de tirages soit atteint, et l'algorithme prend ainsi des décisions tout en garantissant la confidentialité des données sensibles.

3.2.3 Réponse apportée au problème posé et résultats

Voici les algorithmes décrit précédemment, le premier étant celui qui permet de calculer la moyenne d'un ensemble de n données de moments bruts finis d'ordre k sous le modèle de SDP.

Suite à la conception de l'algorithme d'estimation de la moyenne, j'ai cherché à estimer la valeur du paramètre de troncature M permettant de minimiser l'erreur entre l'estimateur et la vraie valeur de la moyenne.

Théorème 8 (Estimation de la moyenne sous SDP)**.** *Pour les données x_i de k-ième moment fini, $\epsilon \in]0,1]$ et $\delta \in [0,1]$, nous considérons un niveau de précision $g = \lceil \varepsilon\sqrt{n} \rceil$, un niveau d'exactitude $\tau = \lceil \frac{g}{\varepsilon} \log(2/p) \rceil$ et un modulo $m = ng + \tau + 1$. Nous supposons que le bruit η_i généré par le randomiseur $\mathcal{R}$ dans l'algorithme 5 est tiré comme suit : $\eta_i = \gamma_i^+ - \gamma_i^-$ où $\gamma_i^+, \gamma_i^- \sim^{i.i.d.} \text{Polya}(1/n, e^{-\varepsilon/g})$. Avec une probabilité d'au moins $1 - 3\delta$, nous avons :*

$$|\tilde{\mu} - \mu| \leq \sqrt{\frac{2\log(2/\delta)}{n}} + \frac{1}{M^{k-1}} + 2\alpha M + \frac{4M\log(1/\delta)}{n\epsilon} \tag{14}$$

Nous choisissons $M = \min((2\alpha)^{-1/k}, \left(\frac{n\varepsilon}{4\log(1/\delta)}\right)^{1/k}$, et nous obtenons :

$$\mathcal{T} \leq \sqrt{\frac{2\log(2/\delta)}{n}} + 2\left(\frac{n\varepsilon}{4\log(1/\delta)}\right)^{1/k-1} + 2(2\alpha)^{1-1/k} \tag{15}$$

Pour ce qui est de l'algorithme d'élimination privée et robuste des bras non prometteurs, décrit dans la partie précédente, il peut être utilisé pour les trois modèles de confidentialité différentielle, la seule différence étant dans l'estimation de la moyenne.

Pour trouver la borne supérieure du regret, j'ai utilisé des résultats et des étapes issus des articles [CZ22] et [Wu+23].

Algorithm 5 Mean Estimation under SDP

Input : Data $\{x_i\}_{i=1}^n$

Parameter : precision $g \in \mathbb{N}$, modulo $m \in \mathbb{N}$, batch size $n \in \mathbb{N}$, privacy level ε, accuracy level $\tau \in \mathbb{R}$

Output : The sum estimator z

Local Randomizer $\mathcal{R}$
(Input : x_i, Output : y_i)
Truncate data $\bar{x}_i = x_i \mathbb{I}_{\{|x_i| \leq M\}}$
Let $\tilde{x}_i = \frac{\bar{x}_i}{2M} + \frac{1}{2}$
Encode x_i as $\hat{x}_i = \lfloor \tilde{x}_i g \rfloor + \text{Ber}(\tilde{x}_i g - \lfloor \tilde{x}_i g \rfloor)$
Generate discrete noise η_i depending on n,ε and g
Add noise and modulo clip $y_i = (\hat{x}_i + \eta_i) \bmod m$
Secure Aggregation $\mathcal{S}$
(Input : $y_1, \ldots, y_n$, Output : $\hat{y}$)
Securely compute $\hat{y} = (\sum_{i=1}^n y_i) \bmod m$
Analyzer $\mathcal{A}$
(Input : $\hat{y}$, Output : $\tilde{\mu}$)
if $\hat{y} > ng + \tau$ **then**
    Set $\tilde{\mu} = \frac{2M}{n} \cdot (\frac{\hat{y}-m}{g} - \frac{1}{2})$
else
    set $\tilde{\mu} = \frac{2M}{n} \cdot (\hat{y}/g - 1/2)$
end if

Algorithm 6 Private and Robust Arm Elimination under CDP/LDP/SDP

Input : Privacy parameter ϵ, robust parameter α and total rounds T

Initialize : $N^t = 0, \tilde{\mu}_a^0 = 0, t = 0$ for every $a \in [K]$
Let $V = [K]$ denote the set of viable arms
repeat
    $t = t + 1$
    $N^t = 2^t, \hat{\mu}_a^t = 0$ for every $a \in V$ (use the idea of "forgetfulness")
    for arm $a \in V$ **do**
        pulls arm a for N^t rounds and observes contaminated rewards $\{r_{a,i}^t\}_{i=1}^{N^t}$
        private and robust mean estimation $\tilde{\mu}_a^t$ under CDP/LDP/SDP by Algorithm 3/4/5
    end for
    Calculate confidence radius β^t
    Calculate $\tilde{\mu}_{\max}^t = \max_{a \in V} \tilde{\mu}_a^t$
    Remove all arm a from V such that : $\tilde{\mu}_{\max}^t - \tilde{\mu}_a^t > 2\beta^t$
until Total number of pulls reaches T, exit

Théorème 9 (Regret Bound under SDP)**.** *La borne supérieure du regret est :*

$$\mathcal{R}_T \leq O\left(\sqrt{KT\log T} + \left(\frac{K\log T}{\varepsilon}\right)^{\frac{k-1}{k}} T^{\frac{1}{k}} + T\alpha^{1-\frac{1}{k}}\right).$$

3.2.4 Conclusions et implication personnelle

Ce projet sur les "Bandits robustes et privés avec alpha-contamination" a été une expérience intellectuelle très enrichissante. En collaborant avec Yulian, j'ai eu l'opportunité de découvrir un domaine de recherche complexe et toujours en évolution, qui réunit des éléments liés à la convergence de la confidentialité des données, de l'apprentissage automatique, et de la prise de décision séquentielle. Ma mission dans ce projet consistait à étendre les résultats préexistants des modèles de confidentialité différentielle centrale (CDP) et locale (LDP) au modèle de confidentialité différentielle mélangée (SDP).

Au fil de cette mission, j'ai développé une compréhension approfondie d'importants concepts de la confidentialité différentielle et des bandits multi-bras. J'ai également exploré les subtilités associées à la gestion de données contaminées de manière stratégique. Naviguer à travers la littérature académique et découvrir des travaux pertinents m'a permis d'acquérir de nouvelles compétences en recherche et en résolution de problèmes complexes.

De plus, ce projet a été un terrain favorable pour développer mes compétences en communication scientifique. Les échanges réguliers avec Yulian et la rédaction de ce rapport ont renforcé ma capacité à expliquer des concepts complexes de manière claire et précise (du moins je l'espère).

Finalement, ce projet a également été une occasion d'apprendre à rechercher des travaux connexes et à mener une analyse bibliographique approfondie. J'ai entamé ce projet à un certain moment de mon stage, et bien qu'il ait fallu intégrer un projet déjà bien avancé, cela a représenté un défi enrichissant.

4 Etude et développement du sujet théorique : les bandits multi-bras

4.1 Introduction

Dans de nombreux aspects de notre vie quotidienne et professionnelle, nous sommes confrontés à des prises de décisions séquentielles dans lesquelles nous devons maximiser nos gains ou notre satisfaction en présence d'incertitudes.

Par exemple, imaginez-vous dîner pour la première fois dans le nouveau restaurant de votre ville. Ce restaurant est d'ores-et-déjà connu pour son menu appétissant et varié, mais vous avez un budget limité et un appétit restreint. Vous êtes maintenant confrontés au défi de trouver les plats qui vous apporteront la plus grande satisfaction gustative, sans avoir la possibilité de tous les essayer. C'est ici que le problème des bandits multi-bras intervient. Dans ce contexte, chaque plat du menu est considéré comme un "bras" ou une action que nous pouvons choisir. Cependant, nous ne connaissons pas à l'avance les résultats ou les récompenses associées à chaque bras. Notre objectif est de maximiser notre satisfaction globale en prenant des décisions séquentielles, tout en tenant compte de l'incertitude entourant chaque choix.
Pour résoudre ce défi, nous devons trouver un équilibre entre l'exploration de nouvelles options et l'exploitation des choix qui semblent être les plus prometteurs. Au début, nous pouvons choisir les plats au hasard pour explorer et collecter des informations sur leur récompense potentielle. À mesure que nous acquérons de l'expérience, nous utilisons ces connaissances pour exploiter les plats qui semblent nous procurer le plus de satisfaction jusqu'à présent. Cet exemple met en évidence les notions et concepts clés du problème des bandits multi-bras, une branche de l'apprentissage par renforcement en plein essor.

Au cours de cette leçon, nous explorerons les concepts clés, les approches probabilistiques et statiques ainsi que les stratégies d'apprentissage utilisées pour résoudre le problème des bandits multi-bras. De plus, nous découvrirons certains algorithmes développés pour implémenter ces stratégies afin de maximiser les gains dans des environnements incertains.

Par ailleurs, nous utiliserons les quelques références suivantes : [LS20], [Ten+21], [Sli19], [NT19], [CZ22] ,[ABM19]

4.1.1 Motivations et domaines d'application

La notion des bandits multi-bras est apparue dans les années 1930, principalement grâce à un article de Thompson, et depuis, son intérêt ne cesse de croître au sein de la communauté scientifique. En effet, avec le développement de l'intelligence artificielle et l'augmentation des données manipulées, les chercheurs en apprentissage automatique et en statistiques s'intéressent de plus en plus à ce cadre de travail qui est d'ores et déjà utilisé dans de nombreux domaines d'application.

Pour citer quelques exemples, considérons tout d'abord le domaine de la publicité en ligne, dans lequel les annonceurs doivent sélectionner les publicités les plus efficaces pour maximiser le taux de clics. Le rôle du bandit multi-bras est d'explorer différentes stratégies publicitaires puis d'exploiter celles qui ont montré les meilleures performances.

Un autre domaine d'application est la recommandation de contenus personnalisés sur les sites de commerce en ligne, sur les réseaux sociaux ou sur les plateformes de streaming. Ces plateformes utilisent des algorithmes de bandits multi-bras afin de proposer des recommandations pertinentes aux utilisateurs en explorant différentes options et en adaptant les recommandations en fonction des préférences et des retours de ces mêmes utilisateurs.

Les bandits multi-bras sont également connus et utilisés dans le domaine de la santé. Ces algorithmes peuvent être utilisés pour déterminer les traitements les plus efficaces pour les patients, ceci afin de rendre plus performants les essais cliniques.

Pour citer un dernier exemple, les bandits multi-bras sont aussi utilisés dans le cadre de gestion de portefeuilles financiers. En effet, les décideurs veulent placer leurs ressources sur différents actifs afin de maximiser le rendement tout en limitant le risque. Les algorithmes utilisés permettent d'évaluer les différentes stratégies possibles et d'adapter la répartition des ressources en fonction des performances observées grâce à la phase d'exploration. Ces exemples permettent d'illustrer l'utilité des bandits multi-bras dans une grande variété de domaines d'application. De manière plus générale, chaque situation dans laquelle des décisions sont prises de manière séquentielle dans un environnement incertain peut être illustrée par le problème de bandits multi-bras.

4.1.2 Le langage des bandits multi-bras

Dans cette partie, nous allons introduire de manière plus formelle le problème de bandits multi-bras ainsi que les notations que nous allons utiliser dans la suite de la leçon [Sli19], [LS20]

Un problème de bandits est un jeu séquentiel dans lequel un joueur est confronté à un environnement composé de K actions, appelées bras. Une distribution de probabilité P_i est associée à chaque action $i \in \{1, \dots, K\}$. A chaque instant $t \in \{1, \dots, T\}$, le joueur doit choisir un bras A_t auquel est associé une récompense X_t, générée de manière stochastique à partir de la distribution de probabilité associée à A_t. Le nombre total de tours T est appelé l'horizon de jeu.

L'objectif du joueur est de maximiser sa récompense cumulée totale, c'est-à-dire, de maximiser la somme des récompenses obtenues au fil du temps, et ce, en prenant des décisions réfléchies et adaptées. Toutefois, le joueur ne connaît pas à l'avance les récompenses associées à chaque bras et doit donc apprendre en interagissant avec l'environnement. Pour guider ses décisions, il va utiliser des politiques, qui sont des stratégies de sélection de bras basées sur l'historique $H_{t-1} = (A_1, X_1, \dots, A_{t-1}, X_{t-1})$.

L'enjeu principal dans ce problème est donc l'exploration et l'exploitation. L'exploitation consiste à tester différentes actions pour découvrir les bras qui pourraient offrir des

récompenses élevées. L'exploitation consiste, quant à elle, à choisir les bras qui ont donné de bonnes récompenses dans le passé afin de maximiser les gains. L'équilibre entre exploration et exploitation est crucial pour obtenir de bonnes performances dans le problème de bandits multi-bras.

Une mesure courante pour évaluer les performances du joueur est le regret R_T. Le regret représente la différence entre la récompense cumulée maximale qui pourrait être obtenue en choisissant le bras optimal $a^* = \text{argmax}_{i\in[K]}\,\mu_i$ à chaque tour et la récompense cumulée réellement obtenue par le joueur :

$$R_T = T.\max_{i\in[K]}\mu_i - \mathbb{E}\left[\sum_{t=1}^{T} X_t\right]$$

où $\forall i \in [K], \mu_i = \mathbb{E}_{X\sim P_i}(X)$
Dans le contexte mentionné précédemment, l'objectif du joueur est de maximiser sa récompense totale cumulée, ce qui revient à minimiser le regret.

En résumé, les problèmes de bandits multi-bras sont des scénarios d'apprentissage séquentiel où le joueur doit explorer et exploiter les actions afin de maximiser sa récompense totale cumulée et donc de minimiser le regret associé aux décisions qu'il a prises.

4.2 Modélisation probabiliste des bandits multi-bras

Dans le cadre des bandits multi-bras, il est important de comprendre certaines notions clés de probabilité. Parmi celles-ci, on trouve les processus stochastiques, les chaînes de Markov et les martingales, ainsi que les notions qui y sont liées comme l'espérance conditionnelle ou encore les filtrations et les temps d'arrêt. De plus, les variables aléatoires sous-gaussiennes et certaines inégalités de concentration sont des outils très utiles pour l'analyse et l'optimisation de stratégies de sélection de bras dans les bandits. En développant certains de ces concepts, nous serons en mesure de définir le problème du bandit stochastique dans un premier temps, puis d'explorer efficacement les différentes approches et algorithmes utilisés pour résoudre ce type de problèmes.

4.2.1 Notions utiles

Pour étudier les problèmes de bandits multi-bras, il est utile d'avoir des connaissances préalables en probabilité. Dans la suite de la leçon, nous supposons que le lecteur possède des connaissances concernant les notions suivantes :

— les espaces de probabilité (tribus, mesures de probabilité, variables aléatoires...),
— les probabilités conditionnelles et l'espérance conditionnelle,
— les différents types de convergence de suite de variables aléatoires,
— les processus stochastiques,
— les martingales et temps d'arrêt,
— les chaînes de Markov.

4.2.2 Inégalités de concentration

Nous allons maintenant présenter quelques inégalités de concentration couramment utilisées. Pour des raisons de concision, les démonstrations de ces inégalités ne seront pas toutes détaillées. Par ailleurs, on ne rappelle pas toutes les inégalités de concentration de base comme celles de Markov, de Bienaymé-Tchebytchev ou encore le théorème Central Limite.

Théorème 10 (Inégalité de Hoeffding, [Ver18]). *Soit $X_1, \ldots, X_n$ une suite de variables aléatoires réelles indépendantes. On suppose que pour tout $i \in \{1, ..., n\}$, $X_i \in [m_i, M_i]$, avec $m_i < M_i$. Alors, pour tout $t > 0$, nous avons :*

$$\mathbb{P}\left[\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i)) \geq t\right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(M_i - m_i)^2}\right)$$

Théorème 11 (Inégalité de Berstein,[Ver18]). *Soit $X_1, X_2, \cdots, X_n$ une suite de variables aléatoires centrées bornées, telles que pour tout i, $|X_i| \leq M$ et $\mathbb{E}[X_i] = 0$. De plus, on note $\sigma^2 = \sum_{i=1}^{n}\mathbb{E}[X_i^2]$ la variance de la somme. Alors, pour tout $t > 0$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq t\right) \leq 2e^{-\frac{t^2/2}{\sigma^2 + Mt/3}}$$

Définition 11 (Variable aléatoire sous-gaussienne, [Ver18]). *Une variable aléatoire réelle X de moyenne nulle est dite σ-sous-gaussienne si sa fonction génératrice des moments satisfait*

$$\mathbb{E}[\exp(tX)] \leq \exp(\frac{\sigma^2 t^2}{2})$$

pour tout $t > 0$. On le note $X \sim \mathrm{subG}(\sigma^2)$.

Définition 12 (Vecteur aléatoire sous-gaussien, [Ver18]). *Un vecteur aléatoire $X \in \mathbb{R}^d$ de moyenne nulle est dit sous-gaussien avec une variance σ^2 si $\langle X, u\rangle$ est sous-gaussienne avec une variance σ^2 pour tout vecteur unitaire $u \in \mathbb{R}^d$. On le note $X \sim \mathrm{subG}_d(\sigma^2)$.*

Remarque 5. Pour une variable aléatoire X non centrée, nous abusons de la notation en disant que X est σ-sous-gaussienne si le bruit $X - \mathbb{E}[X]$ est σ-sous-gaussien.

Propriétés des variables aléatoires sous-gaussiennes [LS20]
La classe des variables aléatoires sous-gaussiennes contient les variables aléatoires bornées et les variables aléatoires gaussiennes, et elle présente des propriétés de concentration fortes.

Théorème 12. *Si X est σ-sous-gaussienne, alors pour tout $\varepsilon > 0$,*

$$\mathbb{P}(X \geq \varepsilon) \leq \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$$

Preuve. A travers cette preuve, nous allons présenter la méthode de Cramer-Chernoff.
Soit $\lambda > 0$ une constante a déterminer par la suite. Alors,

$$\begin{aligned}\mathbb{P}(X \geq \varepsilon) &= \mathbb{P}(\exp(\lambda X) \geq \exp(\lambda\varepsilon)) \\ &\leq \mathbb{E}[\exp(\lambda X)]\exp(-\lambda\varepsilon) \\ &\leq \exp\left(\frac{^2\sigma^2}{2} - \lambda\varepsilon\right)\end{aligned}$$

L'égalité de la première ligne correspond la méthode de Cramer-Chernoff. La deuxième ligne utilise l'inégalité de Markov et la dernière utilise la définition d'une variable aléatoire sous-gaussienne.
Pour terminer cette preuve, il suffit de choisir $\lambda = \varepsilon/\sigma^2$

Remarque 6. La méthode de Cramer-Chernoff repose sur l'utilisation des fonctions génératrices des moments des variables aléatoires. Elle utilise les propriétés de ces fonctions pour obtenir des inégalités exponentielles qui permettent de majorer les probabilités d'événements rares.

Si $X \sim \text{subG}(\sigma^2)$, alors pour tout $t > 0$, on a $\mathbb{P}(|X| > t) \leq 2\exp(-\frac{t^2}{2\sigma^2})$.

Lemme 1. *Considérons deux variables aléatoires indépendantes, X et Y, respectivement σ et σ' sous-gaussiennes, alors nous avons :*

1. $\mathbb{E}[X] = 0$ *et* $\mathbb{V}[X] \leq \sigma^2$
2. *Pour tout réel c, cX est $|c|\sigma$-sous-gaussienne*
3. $X + Y$ *est* $\sqrt{\sigma^2 + \sigma'^2}$*-sous-gaussienne*

On considère $(X_i)_i$ des variables aléatoires telles que $X_i - \mu$ sont des variables aléatoires indépendantes et σ-sous-gaussiennes. Alors, pour tout $\varepsilon > 0$, nous avons :

$$\mathbb{P}(\hat{\mu} \geq \mu + \varepsilon) \leq \exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right) \text{ et } \mathbb{P}(\hat{\mu} \leq \mu + \varepsilon) \leq \exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right)$$

où $\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} X_t$
Preuve. En appliquant le lemme précédent $\hat{\mu} - \mu = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)$ est $\frac{\sigma}{\sqrt{n}}$-sous-gaussien. On trouve ensuite le résultat annoncé en appliquant le théorème 12.

4.2.3 Application des probabilités aux bandits multi-bras : présentation du bandit stochastique

(a) Définition et hypothèses de base, [LS20]

Un **bandit stochastique** est un ensemble de distributions $\nu = (P_a : a \in \mathcal{A})$, où $\mathcal{A}$ est l'ensemble des actions disponibles. Le joueur et l'environnement jouent de manière séquentielle sur T tours. A chaque tour $t \in \{1, \ldots, T\}$, le joueur choisit une action $A_t \in \mathcal{A}$. L'environnement échantillonne ensuite une récompense $X_t \in \mathbb{R}$ à partir de la distribution P_{A_t} et la révèle au joueur. Les interactions entre le joueur et l'environnement induisent une mesure de probabilité sur la suite de résultats $A_1, X_1, \ldots, A_n, X_n$. Cette suite de résultats doit satisfaire les propriétés suivantes :

1. La distribution conditionnelle des récompenses X_t sachant $A_1, X_1, \ldots, A_{t-1}, X_{t-1}, A_t$ est P_{A_t}, c'est-à-dire que :

$$\forall x \in \mathbb{R}, \mathbb{P}(X_t = x | A_1, X_1, \ldots, A_{t-1}, X_{t-1}, A_t) = P_{A_t}(x)$$

2. La loi conditionelle de l'action A_t sachant $A_1, X_1, \ldots, A_{t-1}, X_{t-1}$ est $\pi_t(.|A_1, X_1, \ldots, A_{t-1}, X_{t-1})$:

$$\forall a \in \mathcal{A}, \mathbb{P}(A_t = a|A_1, X_1, \ldots, A_{t-1}, X_{t-1}) = \pi_t(a|A_1, X_1, \ldots, A_{t-1}, X_{t-1})$$

Cette condition assure le fait que le joueur ne peut pas utiliser les observations futures dans ces décisions actuelles du fait qu'il ne connaît pas les distributions de probabilité liées à chaque bras.

(b) L'objectif d'apprentissage

L'objectif du joueur est de maximiser la récompense totale $S_T = \sum_{t=1}^{T} X_t$ qui est une quantité aléatoire dépendant des actions du joueur et des récompenses générées par l'environnement.

Remarque 7. Il ne s'agit pas d'un problème d'optimisation pour plusieurs raisons ; notamment le fait que la récompense totale cumulée est une variable aléatoire, mais aussi du fait que le joueur ne connaît pas les distributions déterminant les récompenses pour chaque bras, ni le nombre de tours T qui doivent être joués. L'horizon T est parfois considéré comme étant connu, mais cela peut varier en fonction des situations.

Notation : Dans tout le reste de la leçon, on notera $[n]$ pour désigner l'ensemble des entiers $\{1, ..., n\}$, pour tout $n \in \mathbb{N}^*$.

Comme discuté dans l'introduction 4.1.2, le joueur est confronté à un compromis entre exploration et exploitation. En effet, il ne possède pas d'information sur l'ensemble des distributions $\nu = (P_a : a \in \mathcal{A})$ relatives aux actions possibles, mais souhaite maximiser sa récompense totale cumulée S_T : il va donc devoir explorer le jeu afin de trouver les bras qui lui permettent d'obtenir les meilleures récompenses. Le compromis réside dans le fait que si à un instant $t \in [T]$, il ne choisit que les actions qui ont les meilleures récompenses jusqu'à présent (exploitation), il va peut-être manquer des actions qui pourraient en offrir des meilleures à long terme. D'autre part, s'il tire des bras de manière aléatoire à chaque tour (exploration), il risque de jouer de manière inefficace et de ne pas maximiser ses récompenses à court terme.

Il est donc important de trouver un équilibre entre l'exploration et l'exploitation pour maximiser les récompenses à long terme. Des stratégies d'apprentissage par renforcement peuvent aider à atteindre cet équilibre en permettant une exploration initiale pour comprendre les préférences de l'environnement, suivie d'une exploitation pour maximiser les récompenses à long terme. Nous introduirons certaines de ces stratégies dans la partie 4.3.

(c) Classe d'environnement et type de bandits

Même dans le cas où l'horizon T est connu par le joueur, l'ensemble des distributions $\nu = (P_a)_{a \in \mathcal{A}}$ associées à chaque bras reste inconnu pour le joueur qui souhaite maximiser sa récompense totale cumulée sur T tours. Habituellement, le joueur est supposé avoir des informations partielles sur ν dont il sait appartenir à l'**environnement de classe $\mathcal{E}$**.

On distingue deux types de problèmes de bandits en fonction de la manière dont les récompenses sont générées : les bandits structurés et les bandits non structurés.

Bandits non structurés. Un bandit est dit non structuré ou indépendant si les bras ou actions disponibles sont considérés comme des expériences indépendantes les unes des autres. Chaque bras génère une récompense de manière aléatoires et ces récompenses ne sont pas liées ou influencées par les récompenses des autres bras.

Bandits structurés. Au contraire, dans un problème de bandits structurés, appelé aussi bandits dépendants ou séquentiels, les récompenses associées à chaque bras sont interdépendantes. Cela signifie que la récompense obtenue en tirant un bras $a \in \mathcal{A}$ peut donner des informations sur les distributions d'un ou de plusieurs autres bras $b \neq a$. Dans ce cas, le joueur doit se servir de ces informations pour ses décisions futures, et ce, afin d'améliorer ses récompenses et maximiser ses gains potentiels.

(d) La notion de regret

Afin d'évaluer la performance du joueur, nous allons présenter la notion de regret cumulé introduite dans la partie 4.1.2. Dans cette partie, nous avons défini le regret comme étant l'écart entre la récompense obtenue par le joueur et celle qu'il aurait obtenue s'il avait suivi la politique optimale.

Soit $\nu = (P_a : a \in \mathcal{A})$ un bandit stochastique, l'espérance mathématique de la récompense associée à l'action $a \in \mathcal{A}$ est définie par :

$$\mu_a(\nu) = \int_{-\infty}^{+\infty} x dP_a(x)$$

Notons $\mu^*(\nu) = \max_{a \in \mathcal{A}} \mu_a(\nu)$ la plus grande moyenne sur tous les bras.
Dans toute la suite, on suppose que $\mu_a(\nu)$ existe pour toute action $a \in \mathcal{A}$ et que $\arg\max_{a \in \mathcal{A}} \mu_a(\nu)$ est non vide.

Le regret associé à la politique π sur le problème de bandit ν correspond à la différence entre la récompense optimale espérée et la récompense totale obtenue en suivant la politique π sur un certain nombre de tours de jeu T.

$$R_T(\pi, \nu) = T\mu^*(\nu) - \mathbb{E}\left[\sum_{t=1}^{T} X_t\right] \tag{16}$$

Le regret mesure l'écart entre les gains que l'on aurait pu espérer en adoptant une stratégie optimale et les gains effectivement obtenus avec la politique π choisie. L'objectif du joueur est donc de minimiser le regret, ce qui équivaut à maximiser l'espérance de la récompense totale S_T.

Lemme 2 (Décomposition du regret, , [LS20])**.** *Soient π une politique et ν un environnement de bandit stochastique où l'ensemble des actions $\mathcal{A}$ est fini ou dénombrable et où l'horizon est $T \in \mathbb{N}$. Le regret R_T de la politique π sur le problème de bandit ν se décompose comme suit :*

$$R_T = \sum_{a \in \mathcal{A}} \Delta_a \mathbb{E}[N_a(T)] \tag{17}$$

où $N_a(t) = \sum_{s=1}^{t} \mathbb{I}\{A_s = a\}$ est le nombre de fois que l'action a est choisi par le joueur après t tours et $\Delta_a(\nu) = \mu^*(\nu) - \mu_a(\nu)$ est appelé l'écart de sous-optimalité ou l'écart d'action de a.

Remarque 8. Ce lemme décompose le regret en termes de pertes dues à l'utilisation de chaque bras. Il nous dit que pour garder le regret petit, l'apprenant doit essayer de minimiser l'utilisation de bras dont l'écart sous-optimal est élevé. Notons que l'écart de sous-optimalité est nul pour le bras optimal.

4.3 Algorithmes d'apprentissage et bandits multi-bras

Dans cette partie, nous allons introduire deux types d'algorithmes fondamentaux ainsi que des outils d'analyse pour le problème de bandits stochastiques non structurés avec un nombre fini d'actions.

Pour plus de clarté et de simplicité, nous considérons dans le reste de la leçon, un problème de bandit multi-bras $\nu = (P_i)_{i=1}^k$ dans lequel la récompense associée à chaque action est générée indépendemment à partir d'une distribution de probabilité sous-Gaussienne de paramètre $\sigma = 1 : \forall t \in [T], X_t \sim \text{SubG}(1)$. Par cette simplification, on fait abstraction du fait que l'algorithme dépend de σ et que les distributions de probabilité des récompenses P_i associées à chaque action i peuvent suivre des lois sous-Gaussiennes de paramètres différents. Un autre cas également courant est de considérer le modèle de Bernoulli qui permet de modéliser la probabilité de succès ou d'échec de chaque action. Ce qui diffère entre le cas sous-Gaussien et les autres cas possibles sont les inégalités de concentration utilisées pour l'analyse.

Il existe plusieurs méthodes et approches pour résoudre le problème du bandit multi-bras, nous présenterons certaines d'entre elles dans cette partie. Tout d'abord, la méthode la plus élémentaire et intuitive est la méthode ETC (explore-then-commit) qui consiste à explorer l'environnement en tirant chaque bras un même nombre de fois puis à exploiter les résultats obtenus pour les tours restants, [LS20]. La méthode epsilon-greedy est quelque plus élaborée ; elle consiste à choisir de manière aléatoire un bras avec une probabilité ε (exploration) et à choisir le bras avec la meilleure récompense estimée jusqu'à présent avec une probabilité 1-ε (exploitation)[Sli19]. La méthode UCB (Upper Confidence Bound) quant à elle, utilise une borne supérieure de confiance pour chaque bras, qui estime la récompense potentielle avec une certaine confiance. L'agent choisit alors le bras avec la plus grande borne supérieure de confiance, [Sli19].

L'étude de ces méthodes permet, entre autre, une compréhension approfondie du compromis entre l'exploration et l'exploitation introduit dans la partie 4.1.2 et peut être étandue à d'autres problèmes.

4.3.1 Méthodes simples : Exploration uniforme

(a) Algorithme "Explore-Then-Commit" (ETC)

Cet algorithme consiste à explorer initialement toutes les actions possibles pendant une période d'exploration définie, puis à sélectionner l'action avec la récompense moyenne la plus élevée et à s'y tenir pendant le reste de l'expérience, durant la phase d'exploitation.

Ici, on considère un problème de bandit avec k actions possibles qui se joue sur T tours. On note $m \in \mathbb{N}$ le nombre de fois que l'algorithme va explorer chaque bras $i \in [k]$. L'algorithme va donc en tout explorer mk fois l'ensemble $\mathcal{A}$ avant de choisir l'action à jouer pour le reste des tours. On note $\hat{\mu}_i(t)$ la récompense moyenne générée par l'exploitation du bras i après t tours :

$$\hat{\mu}_i(t) = \frac{1}{N_i(t)} \sum_{s=1}^{t} \mathbb{I}\{A_s = i\} X_s \tag{18}$$

où $N_i(t) = \sum_{s=1}^{t} \mathbb{I}\{A_s = i\}$ est le nombre de fois que l'action i a été choisie par le joueur après t tours.

Nous allons à présent implémenter cet algorithme en reprenant l'idée de [LS20].

Algorithm 7 Explore-then-commit.

Input : Paramètre d'exploration m

for $t \in [T]$ **do**
  if $t \leq mk$ **then**
    Choisir l'action $A_t = (t \bmod k) + 1$
  else
    Choisir l'action $A_t = \text{argmax}_i \, \hat{\mu}_i(mk)$
  end if
end for

Remarque 9. Durant les mk premiers tours, le fait de choisir l'action $A_t = (t \bmod k) + 1$ permet de jouer chaque action m fois.

Théorème 13 (Borne supérieure du regret)**.** *On considère le problème de bandits introduit précédemment, à savoir, un bandit stochastique non structuré fini dont les récompenses sont générées indépendamment à partir d'une distribution* 1*-sous-gaussienne. Lorsque l'algorithme (7) est appliqué à ce problème et que* $1 \leq m \leq T/k$*, il vient alors :*

$$R_T \leq m \sum_{i=1}^{k} \Delta_i + (T - mk) \sum_{i=1}^{k} \Delta_i \exp\left(-\frac{m\Delta_i^2}{4}\right)$$

avec $\forall i \in [k], \Delta_i = \mu^* - \mu_i$ *l'écart de sous-optimalité du bras* i*.*

Remarque 10. **Mais comment trouver la valeur optimale du paramètre** m **pour atteindre un bon équilibre entre l'exploration et l'exploitation ?**

Le théorème précédent illustre le compromis entre ces deux notions. En effet, si m est grand, alors la politique explore pendant trop longtemps : le premier terme sera important et donc la borne supérieure du regret sera importante aussi. D'autre part, si m est trop petit, alors la probabilité que l'algorithme se trompe en choisissant le mauvais bras augmentera et le deuxième terme deviendra important. La question est comment choisir m.

Supposons que $k = 2$ et que le premier bras est optimal, de sorte que $\Delta_1 = 0$, et posons $\Delta = \Delta_2$. Alors la borne dans le théorème 13 se simplifie comme suit

$$R_T \leq m\Delta + (T - 2m)\Delta \exp\left(-\frac{m\Delta^2}{4}\right) \leq m\Delta + T\Delta \exp\left(-\frac{m\Delta^2}{4}\right) \tag{19}$$

Pour T assez grand, la quantité du côté droit de l'équation 19 est minimisée (à une possible erreur d'arrondi près) par :

$$m = \max\left\{1, \left\lceil \frac{4}{\Delta^2} \log\left(\frac{T\Delta^2}{4}\right)\right\rceil\right\}$$

Pour ce choix et pour tout T, le regret est borné par :

$$R_T \leq \min\left\{T\Delta, \Delta + \frac{4}{\Delta}\left(1 + \max\left\{0, \log\left(\frac{T\Delta^2}{4}\right)\right\}\right)\right\}$$

Remarque 11. L'algorithme ETC est simple à implémenter et peut être efficace dans certaines situations, en particulier lorsque le nombre d'actions est relativement petit. Cependant, il peut ne pas être optimal dans tous les cas et peut entraîner une exploration excessive ou insuffisante selon les paramètres choisis.

(b) Amélioration : algorithme "ε-greedy"

L'algorithme ε-Greedy est également un algorithme simple et couramment utilisé pour résoudre le problème des bandits multi-bras. L'idée de cet algorithme est de choisir l'action avec la récompense moyenne maximale. c'est-à-dire l'action la plus rentable, avec une probabilité élevée $(1-\varepsilon)$ à chaque étape, mais aussi d'explorer d'autres actions avec une faible probabilité (ε).

Concrètement, l'algorithme ε-Greedy alterne entre l'exploration, qui consiste à choisir une action aléatoire, et l'exploitation, qui consiste à choisir l'action la plus rentable jusqu'à présent, et ce, en ajustant le paramètre epsilon pour contrôler la quantité d'exploration.

Algorithm 8 Epsilon-greedy avec probabilité d'exploration ε

Input : Probabilité d'exploration ε

for $t = 1, 2, \ldots$ **do**
  $x \leftarrow$ nombre aléatoire entre 0 et 1 réparti uniformément
  if $x < \varepsilon$ **then**
    Explorer : choisir un bras de manière aléatoire et uniforme.
  else
    Exploiter : choisir le bras avec la récompense moyenne la plus élevée.
  end if
end for

L'algorithme ε-Greedy est simple à implémenter et peut être efficace dans de nombreux cas. Cependant, il peut nécessiter un réglage minutieux du paramètre epsilon pour trouver un bon équilibre entre l'exploration et l'exploitation, afin d'éviter une exploration excessive ou insuffisante.

4.3.2 Méthodes avancées : Exploration adaptative

L'exploration adaptative, abordée dans le cadre des bandits multi-bras, consiste à trouver un équilibre entre l'exploration de nouvelles options et l'exploitation des choix prometteurs. Cette stratégie dynamique permet à l'algorithme d'évoluer en fonction des nouvelles données collectées, favorisant l'exploration des bras incertains au début, puis l'exploitation des bras les plus prometteurs. [Sli19]

Dans cette partie, on introduit une méthode plus performante que les deux précédentes en termes de bornes supérieures du regret : la méthode UCB (Upper Confidence Bound).

(a) Une première approche optimiste

Soit $(X_t)_{t=1}^T$ une suite de variables aléatoires indépendantes 1-sous-Gaussiennes de moyenne μ. On note $\hat{\mu} = \frac{1}{T}\sum_{i=1}^T X_t$. D'après le lemme 1, la variable aléatoire $\hat{\mu} - \mu = \frac{1}{T}\sum_{i=1}^T (X_t - \mu)$ est $\frac{1}{\sqrt{T}}$ - sous-Gaussienne. En appliquant le théorème 12 à la variable aléatoire précédemment définie, alors pour tout $\delta \in [0,1]$, l'inégalité suivante est vérifiée avec une probabilité $1-\delta$:

$$\mu \leq \hat{\mu} + \sqrt{\frac{2\log(1/\delta)}{n}} \tag{20}$$

De maniere symétrique, nous avons aussi avec une probabilité $1-\delta$:

$$\mu \geq \hat{\mu} - \sqrt{\frac{2\log(1/\delta)}{n}} \tag{21}$$

A partir de ces deux équations, on définit une borne supérieure de confiance (UCB), *Upper Confidence Bound*, et une borne inférieure de confiance (LCB), *Lower Confidence Bound*, pour chaque bras $i \in [k]$, temps $t \in [T]$ et $\delta \in [0,1]$:

$$UCB_i(t-1,\delta) = \hat{\mu}_i(t-1) + r_i(t-1,\delta)$$
$$LCB_i(t-1,\delta) = \hat{\mu}_i(t-1) - r_i(t-1,\delta)$$

où $r_i(t-1,\delta) = \sqrt{\frac{2\log(1/\delta)}{T_i(t-1)}}$ est appelé le bonus d'exploration ou le rayon de confiance, et $\hat{\mu}_i(t)$ est défini par l'équation 18.

(b) Algorithme de sélection de bras

Le premier algorithme que nous pouvons implémenter grâce à cette méthode est celui de sélection de bras. Pour cela, nous n'allons avoir besoin que de l'UCB.

Théorème 14 (Borne supérieure du regret)**.** *Pour tout horizon de jeu T, si $\delta = 1/T^2$, alors le regret obtenu en appliquant l'algorithme 9 a un problème de bandit stochastique 1-sous-Gaussien à k bras, alors :*

$$R_T \leq 3\sum_{i=1}^T \Delta_i + \sum_{i:\Delta_i>0} \frac{16\log(T)}{\Delta_i} \tag{22}$$

Algorithm 9 Sélection de bras UCB(δ)

Input k et δ

for $t \in 1, \ldots, T$ **do**

Choisir l'action $A_t = \text{argmax}_i \, UCB_i(t-1, \delta)$

Observer la recompense X_t et mettre à jour la borne supérieure de confiance UCB

end for

(c) Algorithme d'élimination des bras

Selon la même approche que précédemment, nous définissons l'UCB et la LCB avec un rayon de confiance défini de cette façon : $r_i(t) = \sqrt{\frac{2 \log T}{N_i(t)}}$ où T est l'horizon du jeu.

Algorithm 10 Eliminations successives UCB

Initialisation Activer tous les bras et noter $\mathcal{A}$ l'ensemble des bras actifs.

while $\mathcal{A}$ est non vide **do**

Jouer tous les bras actifs une fois

Désactiver tous les bras $a \in \mathcal{A}$ tels que, si t est le tour en cours, $\exists a_0 \in \mathcal{A}$ tel que $UCB_t(a) < LCB_t(a_0)$

end while

Tout comme l'algorithme 9, l'algorithme 10 permet de trouver les bras les plus prometteurs. En effet, il utilise les indices UCB et LCB introduits précédemment pour éliminer progressivement les bras moins performants jusqu'à ne conserver que les bras les plus susceptibles de donner les meilleures récompenses.

Théorème 15 (Borne supérieure du regret)**.** *L'algorithme d'éliminations successives satisfait :*

$$\mathbb{E}[R_T] \leq O(\log T). \sum_{i \in [k] : \mu_i < \mu^*} \left[\frac{1}{\mu^* - \mu_i} \right]$$

4.4 Simulation

Dans cette partie, nous allons étudier les performances des trois algorithmes introduits précédemment à l'aide d'un exemple de bandits multi-bras. Pour ce faire, nous allons nous intéresser aux nombres de tirages de chaque bras et aux gainx moyens par tour, pour les algorithmes ETC, Epsilon-greedy et UCB.

Tout au long de ces simulations, nous allons nous inspirer des travaux [Lou],[Mil] et considérer un problème de bandit à deux bras arm_1 et arm_2. La récompense liée à chaque bras suit une loi de Bernoulli de paramètres p_1 et p_2, respectivement.

4.4.1 Explore-Then-Commit

Nous considérons tout d'abord, l'algorithme "Explore-then-commit" dans le cadre du problème de bandits présenté précédemment.

Voici différentes simulations obtenues grâce au code Python fournit en annexe 1, en faisant varier les valeurs des probabilités de succès liées aux récompenses de chaque bras.

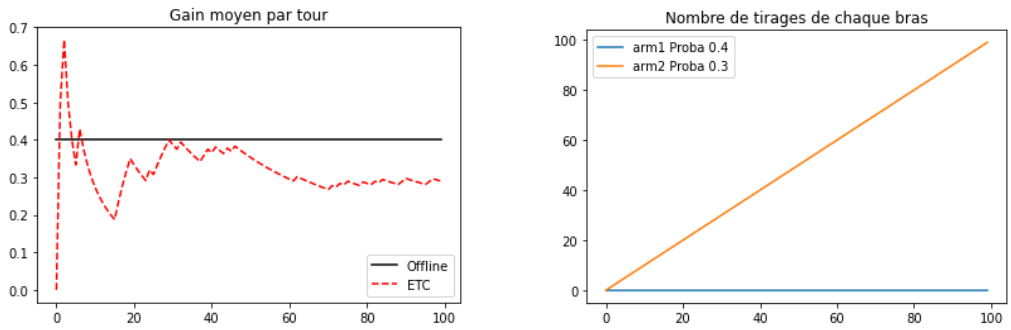

FIG. 4 – Gain moyen par tour et nombre de tirages de chaque bras pour $p_1 = 0.4$, $p_2 = 0.3$

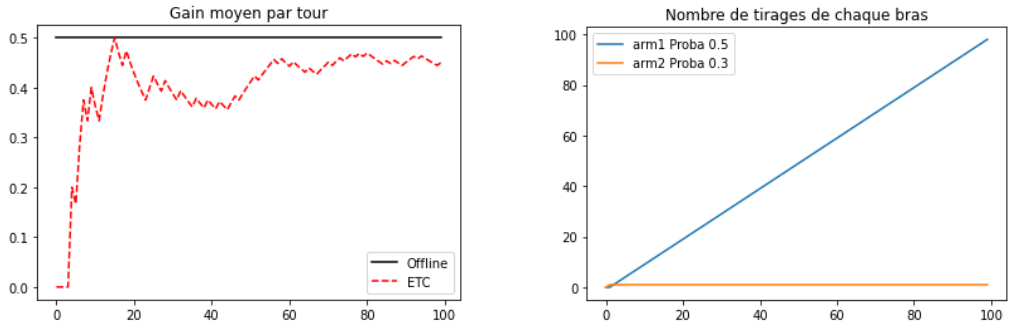

FIG. 5 – Gain moyen par tour et nombre de tirages de chaque bras pour $p_1 = 0.5$, $p_2 = 0.3$

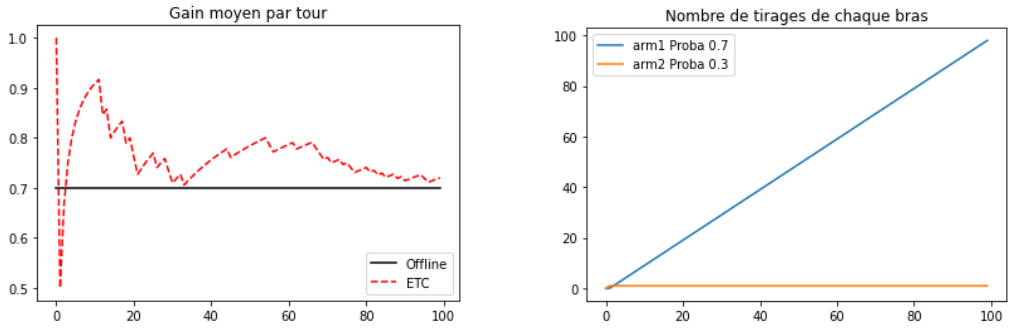

FIG. 6 – Gain moyen par tour et nombre de tirages de chaque bras pour $p_1 = 0.7$, $p_2 = 0.3$

Dans cette série de simulations d'Explore-Then-Commit (ETC), nous examinons comment l'algorithme réagit à différentes configurations de probabilités de succès des bras.

Remarque 12. Ici, on a choisi un nombre d'itérations égal à 100 et un nombre de tirages d'exploration égal à 1.

Ainsi, lorsque les probabilités sont proches, telles que $p_1 = 0.4$ et $p_2 = 0.3$ dans 4, l'algorithme peut initialement choisir le bras moins prometteur, car chaque bras n'a été tiré qu'une seule fois au début et l'algorithme exploite celui qui a donné une récompense maximale durant le reste des itérations. Ceci se traduit par une oscillation du gain moyen autour de $p_2 = 0.3$. En revanche, lorsque les probabilités sont plus éloignées, par exemple $p_1 \in \{0.5, 0.7\}$ et $p_2 = 0.3$, figures 5 et 6, l'algorithme identifie rapidement le bras optimal, conduisant à une stabilisation du gain moyen autour de $p_1 = 0.7$.

4.4.2 Epsilon-greedy

Nous considérons ensuite, l'algorithme "Epsilon-greedy" dans le cadre du même problème de bandits que précédemment.

Voici différentes simulations obtenues grâce au code Python fourni en annexe 2. Ici, on fait varier les probabilités de succès liées aux récompenses de chaque bras, ainsi que la valeur du paramètre d'exploration ϵ.

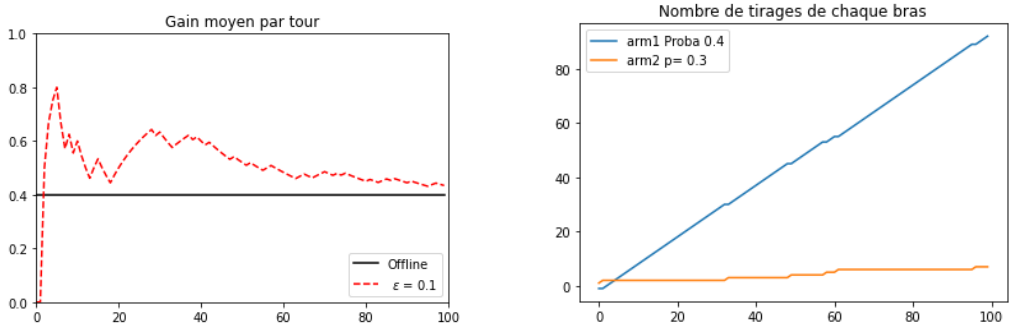

FIG. 7 – Résultats pour $p_1 = 0.4$, $p_2 = 0.3$ et $\varepsilon = 0.1$

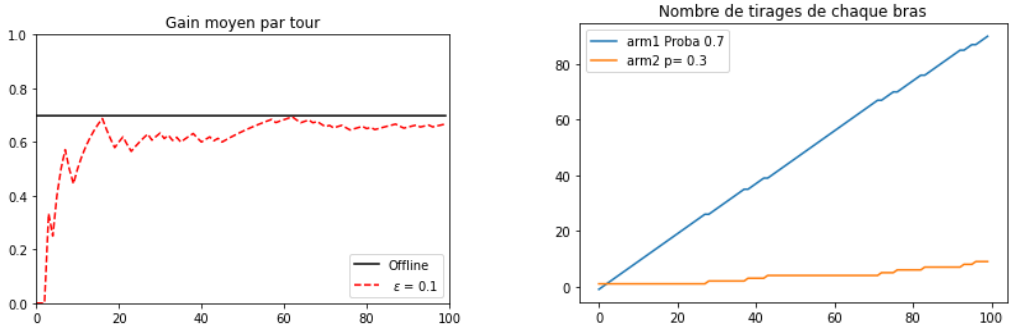

FIG. 8 – Résultats pour $p_1 = 0.7$, $p_2 = 0.3$ et $\varepsilon = 0.1$

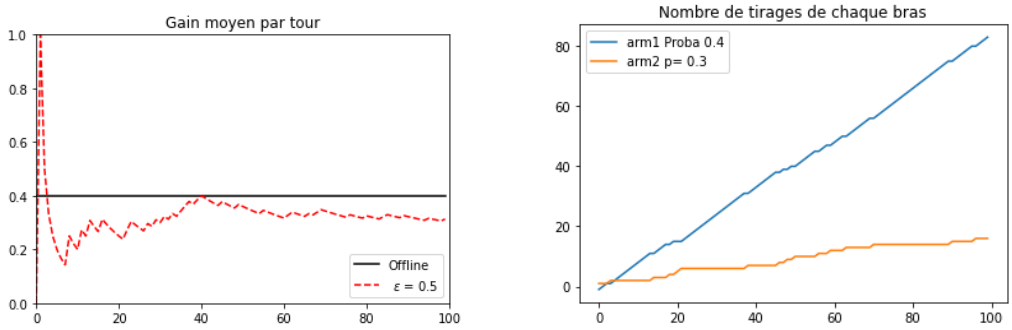

FIG. 9 – Résultats pour $p_1 = 0.4$, $p_2 = 0.3$ et $\varepsilon = 0.5$

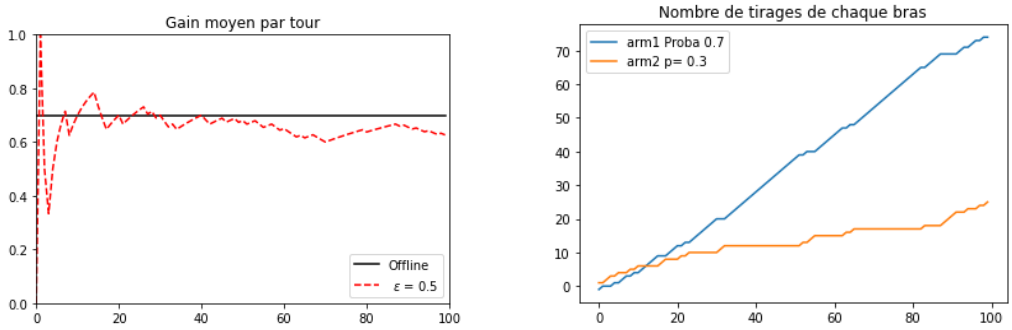

FIG. 10 – Résultats pour $p_1 = 0.7$, $p_2 = 0.3$ et $\varepsilon = 0.5$

Dans ces simulations d'Epsilon-greedy, nous explorons l'influence de ϵ sur le comportement de l'algorithme. Lorsque $\epsilon = 0.1$, l'algorithme exploite davantage les résultats passés pour déterminer le bras optimal et explore beaucoup moins les autres bras. Par exemple, avec $p_1 = 0.4$ et $p_2 = 0.3$, figure 7, l'algorithme commence par exploiter le bras moins prometteur, puis ajuste sa stratégie.

A l'inverse, dans les cas où la probabilité d'exploration est plus forte, figures 7 et 8, avec $\varepsilon = 0.5$, l'algorithme privilégie davantage l'exploration, comme on peut le voir sur les figures liées au nombre de tirages de chaque bras.

4.4.3 UCB

Nous considérons à présent, l'algorithme "UCB" pour le même problème de bandits.

Voici différentes simulations obtenues grâce au code Python fourni en annexe 3. Ici, on fait varier les valeurs des probabilités de succès liées aux récompenses de chaque bras.

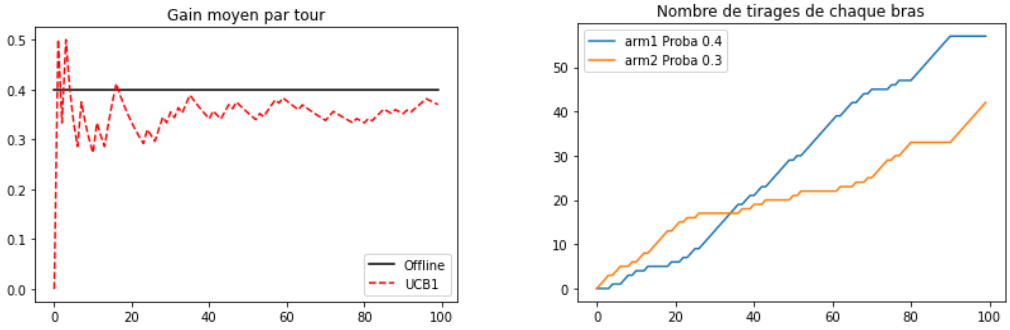

FIG. 11 – Gain moyen par tour et nombre de tirages de chaque bras pour $p_1 = 0.4$, $p_2 = 0.3$

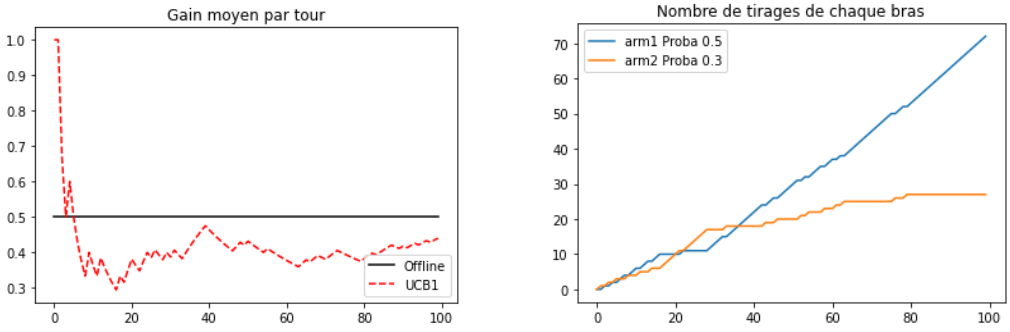

FIG. 12 – Gain moyen par tour et nombre de tirages de chaque bras pour $p_1 = 0.5$, $p_2 = 0.3$

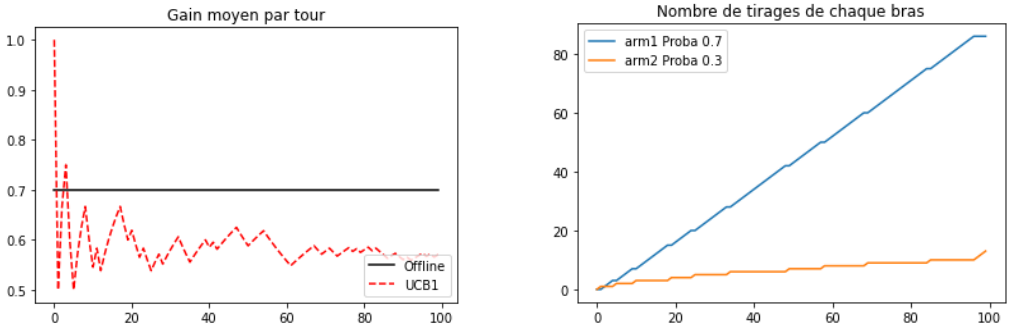

FIG. 13 – Gain moyen par tour et nombre de tirages de chaque bras pour $p_1 = 0.7$, $p_2 = 0.3$

Les simulations avec l'algorithme UCB mettent en lumière une exploration plus importante, surtout lorsque les probabilités des bras sont proches et qu'il est donc plus difficile d'identifier le bras le plus prometteur. Ceci se traduit par des gains moyens par tour oscillant autour de l'espérance la plus élevée.

4.4.4 Comparaison des performances

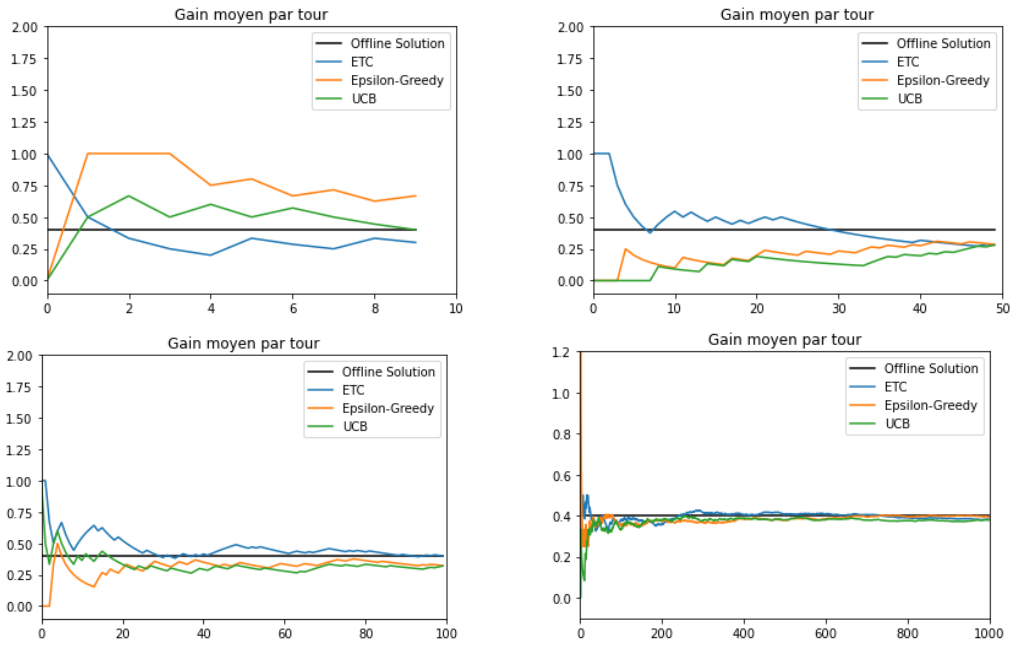

FIG. 14 – Comparaison des gains moyens par tour pour différents nombres d'itérations

Grâce à cette simulation, nous pouvons comparer le gain moyen par tour des différents algorithmes dans le cas où $p_1 = 0.4$ et $p_2 = 0.3$. Nous avons fait varier le nombre d'itérations afin de voir que le gain converge vers la valeur $p_1 = 0.4$ et que par conséquent chaque algorithme va savoir trouver un bon compromis entre exploration et exploitation, mais qu'il va nécessiter un nombre d'itérations plus ou moins important en fonction des algorithmes.

4.4.5 Conclusions tirées des simulations

En observant les résultats des simulations, on peut faire des observations plus générales sur les performances des trois algorithmes.

ETC semble performant lorsque les probabilités des bras sont significativement différentes. Il converge rapidement vers le bras optimal et stabilise le gain moyen autour de la probabilité la plus élevée.

Avec un ϵ faible, Epsilon-greedy peut se rapprocher de ETC, car il exploite davantage les informations passées. Toutefois, un ϵ plus élevé favorise l'exploration. Dans ces cas, la performance est plus robuste face à des probabilités de bras similaires, mais le gain moyen par tour peut être légèrement inférieur en général.

UCB démontre une exploration plus importante, en particulier lorsque les probabilités des bras sont proches. Cela peut conduire à une meilleure adaptation aux situations où les différences entre les bras sont plus subtiles, bien que cela puisse entraîner une exploration excessive dans certaines circonstances.

En résumé, ces simulations offrent un aperçu des réponses des algorithmes face à des scénarios de bandits multi-bras, et mettent en évidence la manière dont ils gèrent l'équilibre

entre exploration et exploitation. Le choix de l'algorithme va dépendre des caractéristiques spécifiques du problème et des ajustements fins des paramètres seraient nécessaires pour des conclusions plus définitives.

4.5 Conclusion

Pour conclure cette leçon, nous avons exploré un domaine passionnant concernant la prise de décisions séquentielles dans un contexte d'incertitude. À travers des exemples, nous avons introduit le défi existant entre exploration et exploitation, qu'il est essentiel de comprendre en vue de maximiser les gains dans les problèmes de bandits multi-bras.

L'étendue des domaines d'application, de la publicité en ligne à la santé, souligne l'impact croissant et universel des bandits multi-bras. Notre leçon a introduit un langage commun, défini des mesures de performance et abordé des concepts probabilistes fondamentaux.

Des méthodes simples, telles que "Explore-Then-Commit" (ETC) et "ε-Greedy", ont été présentées, suivies d'une approche plus sophistiquée avec "Upper Confidence Bound" (UCB). Des simulations ont permis d'étudier leurs performances.

En résumé, cette leçon offre une introduction aux bandits multi-bras, permettant l'accès à la compréhension de leurs principaux mécanismes et de leurs diverses applications. Elle constitue une porte d'entrée à des explorations plus approfondies en matière d'apprentissage par renforcement.

5 Apports personnels de cette expérience internationale

Mon expérience de stage à l'étranger a été une source inestimable d'apprentissage, tant sur le plan professionnel que personnel. Pendant cette période, j'ai eu l'opportunité d'acquérir un éventail de compétences professionnelles, liées au domaine de la recherche en particulier, mais pas seulement. J'ai par exemple perfectionné ma maîtrise du logiciel Overleaf, un outil essentiel pour la rédaction d'articles ou de rapports scientifiques entre autres, tout en développant mes compétences en lecture et compréhension d'articles académiques. La gestion de mon emploi du temps, laissée à ma propre initiative, m'a incité à me discipliner davantage. Par ailleurs, l'absence d'horaires fixes m'a permis d'être plus responsable et organisée.

Au cours de ce stage, j'ai eu l'opportunité d'explorer les défis et les enjeux de la recherche en machine learning. J'ai rapidement compris que rester informé sur les dernières avancées en matière d'algorithmes et de résultats théoriques était essentiel pour demeurer compétitif. Par ailleurs, le travail collaboratif avec des chercheurs du monde entier m'a permis de comprendre que chaque contribution individuelle peut avoir un impact significatif sur le travail collectif et que les échanges et le partage des connaissances sont importants. Cette expérience m'a sensibilisée à l'importance de la réactivité aux nouveautés. Mon immersion dans un environnement de travail multiculturel a été une expérience d'adaptabilité enrichissante qui m'a également permis de m'ouvrir à de nouveaux horizons. Travailler au sein d'un laboratoire de recherche était très différent de ce que j'avais connu jusqu'alors, cela comprenait des méthodes de travail assez spécifiques. Cette expérience m'a donc appris à m'adapter rapidement et à m'ouvrir à de nouvelles approches. Par ailleurs, l'interaction avec des collègues de différentes cultures a renforcé mon sens de l'ouverture d'esprit et de la compréhension interculturelle.

Sur un plan plus personnel, ce stage m'a permis de mieux me connaître en tant que professionnelle. J'ai réalisé que je suis plus productive lorsque je m'impose un certain nombre d'heures de travail par jour. Par ailleurs, j'ai découvert que le métier de chercheur était un métier assez solitaire, je me suis rendue compte que je pouvais travailler seule mais que j'avais besoin d'interactions avec les membres du laboratoire pour maintenir ma motivation. La possibilité de changer de lieu de travail régulièrement pour éviter la monotonie s'est avérée bénéfique pour moi.

En ce qui concerne mon avenir professionnel, ce stage m'a ouvert de nouvelles perspectives. J'ai reçu une proposition de doctorat, ce qui m'a amenée à réfléchir sur mon intérêt pour la recherche. Le fait de travailler sur des applications concrètes et utiles à plus courte échelle est quelque chose d'essentiel pour moi afin de voir plus rapidement la réalisation de mon travail et de me sentir utile. Toutefois, bien que mon stage ait été principalement théorique, il ne m'a pas découragée de poursuivre une carrière en recherche.

Je suis désormais ouverte à l'idée d'explorer des domaines plus appliqués, tout en continuant à travailler dans le domaine des statistiques.

Par ailleurs, cette expérience a renforcé ma conviction de l'importance de vivre et travailler dans un environnement qui valorise la diversité et l'inclusion et où les différences sont perçues comme une force plutôt qu'un obstacle ou une faiblesse.

Mon stage à l'étranger m'a offert l'opportunité de découvrir l'Arabie Saoudite d'une manière qui a dépassé les stéréotypes et les préjugés préexistants. J'ai été profondément touchée par la générosité et l'hospitalité du peuple saoudien. J'ai notamment été généreusement conviée par une de mes amies saoudiennes, à partager des moments en famille, lors de fêtes ou de voyages, me considérant comme un membre à part entière de leur cercle familial.

Enfin, j'ai également eu l'occasion de rencontrer des personnes provenant des quatre coins du monde (Vietnam, Chine, Kazakhstan, Brésil, Portugal, Angleterre, Etats-Unis, Tadjikistan, Equateur, Tunisie, Egypte, France...), ce qui a renforcé ma conviction que malgré nos origines culturelles différentes, nous partageons des valeurs et des aspirations communes. En outre, cette expérience m'a appris à communiquer efficacement dans un environnement multiculturel et à vivre et travailler harmonieusement dans la diversité.

6 Conclusion

En conclusion, ce semestre de stage à l'étranger au sein du laboratoire PRADA de KAUST a été une expérience riche en apprentissage, tant sur le plan professionnel que personnel. J'ai relevé avec enthousiasme des défis intellectuels stimulants centrés sur la confidentialité différentielle et l'apprentissage robuste, élargissant ainsi ma compréhension de l'apprentissage automatique.

Au fil de ces cinq mois, j'ai développé mes compétences dans le domaine de la recherche, découvrant ainsi ce secteur d'étude de manière approfondie.

Cette expérience m'a également permis d'approfondir ma connaissance de moi en tant que professionnelle. L'adaptabilité à un travail de chercheur relativement solitaire a été une découverte enrichissante, soulignant ma capacité à m'adapter rapidement et à persévérer face à des problématiques complexes.

Je suis reconnaissante d'avoir eu l'opportunité de collaborer dans un environnement culturellement diversifié, aux côtés de personnes brillantes dans leur domaine de travail ; une expérience qui a contribué à élargir mes horizons et à enrichir mes perspectives professionnelles.

Références

[AYB23] Baris Alparslan, Sinan Yildirim et Ilker Birbil. « Differentially Private Distributed Bayesian Linear Regression with MCMC ». In : *Proceedings of the 40th International Conference on Machine Learning*. 2023, p. 627-641. URL : https://proceedings.mlr.press/v202/alparslan23a.html.

[ABM19] Jason Altschuler, Victor-Emmanuel Brunel et Alan Malek. « Best Arm Identification for Contaminated Bandits ». In : *arXiv preprint* (2019). arXiv :1802.09514 [cs, math, stat]. DOI : 0.48550/arXiv.1802.09514.

[Bal+19] Borja Balle et al. « The Privacy Blanket of the Shuffle Model ». In : *arXiv preprint arXiv :1903.02837* (2019).

[Ban+21] Aman Bansal et al. « Flexible Accuracy for Differential Privacy ». In : *arXiv* (2021). DOI : 10.48550/arXiv.2110.09580. arXiv : 2110.09580 [cs].

[BH22] Adarsh Barik et Jean Honorio. « Sparse Mixed Linear Regression with Guarantees : Taming an Intractable Problem with Invex Relaxation ». In : *Proceedings of the 39th International Conference on Machine Learning*. 2022, p. 1627-1646. URL : https://proceedings.mlr.press/v162/barik22a.html.

[BC22] Akash Bharadwaj et Graham Cormode. « Sample and Threshold Differential Privacy : Histograms and Applications ». In : *arXiv* (2022). DOI : 10.48550/arXiv.2112.05693. arXiv : 2112.05693 [cs].

[CZ22] Sayak Ray Chowdhury et Xingyu Zhou. « Distributed Differential Privacy in Multi-Armed Bandits ». In : *arXiv preprint* (2022). arXiv :2206.05772 [cs].

[CIL18] Rachel Cummings, Stratis Ioannidis et Katrina Ligett. « Truthful Linear Regression ». In : *ArXiv preprint arXiv :1809.04356* (sept. 2018).

[Dia18] Ilias Diakonikolas. « Algorithmic high-dimensional robust statistics ». In : *Webpage http://www.iliasdiakonikolas.org/simons-tutorial-robust.html* (2018).

[DR14] Cynthia Dwork et Aaron Roth. « The Algorithmic Foundations of Differential Privacy ». In : *Foundations and Trends in Theoretical Computer Science* 9.3–4 (2014), p. 211-407. DOI : 10.1561/0400000042.

[Dwo+06] Cynthia Dwork et al. « Calibrating noise to sensitivity in private data analysis ». In : *Theory of cryptography conference*. Springer. 2006, p. 265-284.

[Dwo+17] Cynthia Dwork et al. « Calibrating Noise to Sensitivity in Private Data Analysis ». In : *Journal of Privacy and Confidentiality* (2017). Vol. 7, numéro 3, pp. 17-51. DOI : 10.29012/jpc.v7i3.405.

[Fal+23] Alireza Fallah et al. « Optimal and Differentially Private Data Acquisition : Central and Local Mechanisms ». In : *arXiv preprint* (2023). arXiv :2201.03968 [cs, math, stat]. DOI : 0.48550/arXiv.2201.03968.

[Hsu+14] Justin Hsu et al. « Private matchings and allocations ». In : *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*. 2014, p. 21-30.

[Kam] Gautam Kamath. *A Course In Differential Privacy*. [en ligne]. URL : https://www.youtube.com/playlist?list=PLmd_zeMNzSvRRNpoEWkVo6QY_6rR3SHjp.

[KAU] KAUST. *King Abdullah University of Science and Technology*. Consulté le 16 septembre 2023. URL : https://www.kaust.edu.sa/en/.

[KHH22] Antti Koskela, Mikko A. Heikkilä et Antti Honkela. « Tight Accounting in the Shuffle Model of Differential Privacy ». In : *arXiv* (2022). DOI : 10.48550/arXiv.2106.00477. arXiv : 2106.00477 [cs, stat].

[LS20] Tor Lattimore et Csaba Szepesvari. *Bandit Algorithms*. Cambridge University Press, 2020.

[Liu+22] Jiachang Liu et al. « Fast Sparse Classification for Generalized Linear and Additive Models ». In : *arXiv* (2022). DOI : 10.48550/arXiv.2202.11389. arXiv : 2202.11389 [cs, stat].

[Lou] J. Louedec. *demoBandits*. Disponible à l'adresse : https://www.math.univ-toulouse.fr/~jlouedec/demoBandits.html.

[McS09] Frank D McSherry. « Privacy integrated queries : an extensible platform for privacy-preserving data analysis ». In : *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*. 2009, p. 19-30.

[Mil+22] Jason Milionis et al. « Differentially Private Regression with Unbounded Covariates ». In : *arXiv* (2022). DOI : 10.48550/arXiv.2202.11199. arXiv : 2202.11199 [cs, stat].

[Mil] Tim Miller. *Multi-armed bandits — Introduction to Reinforcement Learning*. Disponible à l'adresse : https://gibberblot.github.io/rl-notes/single-agent/multi-armed-bandits.html.

[NT19] Laura Niss et Ambuj Tewar. « What You See May Not Be What You Get : UCB Bandit Algorithms Robust to -Contamination ». In : *arXiv preprint* (2019). eprint : 1910.05625. URL : https://arxiv.org/abs/1910.05625.

[PRA] PRADA LAB-Provable Responsible AI and Data Analytics (PRADA) Lab. *About PRADA LAB-Provable Responsible AI and Data Analytics (PRADA) Lab*. Consulté le 16 septembre 2023. URL : http://www.pradalab.org/?about_8/.

[QLW22] Yuan Qiu, Jinyan Liu et Di Wang. *Truthful Generalized Linear Models*. arXiv :2209.07815 [cs]. Sept. 2022. DOI : 10.48550/arXiv.2209.07815. URL : http://arxiv.org/abs/2209.07815 (visité le 12/12/2022).

[Ros+22] Emanuele Rossi et al. « Learning to Infer Structures of Network Games ». In : *Proceedings of the 39th International Conference on Machine Learning*. 2022, p. 18809-18827. URL : https://proceedings.mlr.press/v162/rossi22a.html.

[Sli19] Aleksandrs Slivkins. *Introduction to Multi-Armed Bandits*. T. 12. 1. Now Publishers Inc., 2019, p. 1-305.

[Ten+21] Jay Tenenbaum et al. « Differentially Private Multi-Armed Bandits in the Shuffle Model ». In : *arXiv preprint arXiv :2106.02900v1* (2021).

[Vad16] Salil Vadhan. « The Complexity of Differential Privacy ». In : *Center for Research on Computation & Society* (août 2016). URL : http://seas.harvard.edu/~salil.

[Ver18] Roman Vershynin. *High-Dimensional Probability*. Cambridge University Press, 2018, p. 1-36.

[Ver19] Roman Vershynin. *High-Dimensional Probability : An Introduction with Applications in Data Science*. 2019 Prose Award for Mathematics. Cambridge University Press, 2019.

[WZ09] Larry Wasserman et Shuheng Zhou. « A Statistical Framework for Differential Privacy ». In : *arXiv* (2009). DOI : 10.48550/arXiv.0811.2501. arXiv : 0811.2501 [math, stat].

[Wu+23] Yulian Wu et al. « On Private and Robust Bandits ». In : *arXiv preprint* (2023). arXiv :2302.02526 [cs, stat].

[XLS] XLSTAT. *Régression linéaire - méthode des moindres carrés (OLS)*. Consulté le 30 octobre 2023. URL : https://www.xlstat.com/fr/solutions/fonctionnalites/regression-lineaire-moindres-carres.

Glossaire

A | **C** | **D** | **E** | **I** | **K** | **N** | **P** | **R** | **T**

A

ACL L'Association for Computational Linguistics est la principale société importante en traitement automatique des langues. Elle organise chaque année la conférence scientifique la plus prestigieuse du domaine.

Apprentissage robuste L'apprentissage robuste se réfère à la capacité d'un modèle d'apprentissage automatique à maintenir ses performances et sa précision même en présence de perturbations, de bruit ou de variations dans les données d'entrée.

Apprentissage profond topologique L'apprentissage profond topologique se réfère à une approche d'apprentissage automatique intégrant des concepts de topologie, une branche mathématique étudiant les propriétés préservées par les transformations continues. Cette approche vise à saisir les caractéristiques topologiques complexes des données.

C

Confidentialité différentielle La confidentialité différentielle est un principe de protection de la vie privée en statistiques qui garantit que l'inclusion ou l'exclusion d'une seule observation n'affecte pas de manière significative les résultats d'une analyse, préservant ainsi l'anonymat des individus dans les données..

CVPR La Conference on Computer Vision and Pattern Recognition est une conférence scientifique annuelle de l'IEEE dédiée à la reconnaissance des formes et à la vision par ordinateur.

D

Désapprentissage Le désapprentissage se rapporte à la capacité d'un modèle à "oublier" ou à ajuster ses connaissances antérieures en fonction de nouvelles données.

Données sensibles Les données sensibles englobent des informations personnelles révélant l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses, l'appartenance syndicale, ainsi que des données génétiques, biométriques, de santé, et des détails sur la vie sexuelle ou l'orientation sexuelle d'une personne.

E

ECCV L'European Conference on Computer Vision est une conférence en vision par ordinateur organisée par la communauté scientifique et qui se déroule tous les deux ans. Elle alterne avec ICCV et est considérée comme aussi prestigieuse que CVPR ou ICCV.

EMNLP Empirical Methods in Natural Language Processing est une conférence de premier plan dans le domaine du traitement du langage naturel et de l'intelligence artificielle.

I

ICCV L'International Conference on Computer Vision est une conférence en vision par ordinateur organisée par l'IEEE et qui a lieu une fois tous les deux ans, en alternance avec ECCV.

ICLR L'International Conference on Learning Representations est une conférence annuelle sur l'apprentissage automatique.

ICML L'International Machine Learning Society est une organisation à but non lucratif dont l'objectif est d'encourager la recherche sur l'apprentissage automatique et dont la principale activité est la présentation d'une conférence annuelle, la conférence internationale sur l'apprentissage automatique.

IEEE La conférence internationale IEEE sur la robotique et l'automatisation est un événement académique annuel qui couvre les avancées en robotique.

Informatique responsable et fiable L'informatique responsable et fiable se réfère à une approche éthique et transparente du développement et de l'utilisation des technologies informatiques. Cette approche intègre la considération des implications sociales, éthiques et environnementales tout au long de l'exploitation et du cycle de vie des systèmes informatiques.

Intelligence artificielle explicative et le désapprentissage L'intelligence artificielle explicative concerne le développement de modèles d'IA capables d'expliquer de manière compréhensible et transparente leur processus de prise de décision.

Intelligence artificielle L'intelligence artificielle englobe les technologies et les systèmes informatiques capables d'accomplir des tâches qui nécessitent généralement l'intelligence humaine, telles que la résolution de problèmes complexes, la compréhension du langage naturel, la reconnaissance de formes et la prise de décision.

K

KAUST King Abdullah University of Science and Technology est une université internationale de recherche fondée en 2006 en périphérie de Djeddah, Arabie Saoudite.

N

NeurIPS La conférence Neural Information Processing Systems est un événement annuel majeur dans le domaine de l'intelligence artificielle et des neurosciences computationnelles. Elle se tient chaque année en décembre, réunissant des chercheurs éminents et offrant une plateforme cruciale pour les avancées scientifiques.

P

Préservation de la confidentialité La préservation de la confidentialité comprend l'ensemble des techniques et mécanismes visant à garantir que les informations sensibles ou personnelles dans les données ne sont ni compromises ni révélées lors de leur traitement ou de leur analyse.

R

Rationalité individuelle La rationalité individuelle fait référence au fait qu'un individu va prendre des décisions fondées sur la logique et la recherche du meilleur résultat possible en fonction de ses préférences et de ses objectifs personnels. Cela se caractérise par le principe selon lequel l'utilité d'un individu doit toujours être positive

.

T

Théorie de l'apprentissage automatique La théorie de l'apprentissage automatique englobe les principes fondamentaux qui guident le fonctionnement des algorithmes d'apprentissage automatique. Elle explore les concepts mathématiques et statistiques qui permettent aux machines d'apprendre à partir de données et de généraliser ces connaissances à de nouvelles situations.

A Annexe

Afin de maintenir la simplicité, les démonstrations des théorèmes sont présentées dans la langue originale des articles, qui est l'anglais.

A.1 Preuves des théorèmes de la partie 3.1

A.1.1 Lemmes utiles pour la garantie de la confidentialité

Lemme 3 (Post-processing). *Let $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ be an ε-differential private mechanism. Consider $F : \mathcal{Y} \to \mathcal{Z}$ as an arbitrary randomized mapping. Then the mechanism $F \circ \mathcal{M}$ is ε-differential private.*

Lemme 4 (Billboard lemma [Hsu+14]). *Let $\mathcal{M} : \mathcal{D}^n \to \mathcal{O}$ be an (ε, δ)-differential private mechanism. Consider a set of n functions $\pi_i : \mathcal{D} \times \mathcal{O} \to \mathcal{R}$, for $i \in [n]$. Then the mechanism $\mathcal{M}' : \mathcal{D}^n \to \mathcal{O} \times \mathcal{R}^n$ that computes $r = \mathcal{M}(D)$ and outputs $\mathcal{M}'(D) = (r, \pi_1(D_1, r), \cdots, \pi_n(D_n, r))$, where D_i is the agent i's data, is (ε, δ)-joint differential private.*

Théorème 16 (Parallel Composition Theorem). *Let $O_1, O_2, ..., O_k$ be n independent operations that satisfy $(\varepsilon_i, \delta$-DP for each $i \in [k]$, then the parallel composition of these operations guarantee $(\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_k, \delta)$-DP.*

Théorème 17 (Sequential Composition Theorem(Theorem 4 in [McS09])). *If $O_1, O_2, ..., O_k$ are sequential operations that satisfy ε-individual differential privacy with a failure probability of δ, then the sequential composition of these operations guarantees $(\varepsilon_1 + \varepsilon_2 + ... + \varepsilon_k, k \cdot \delta)$-individual differential privacy, where ε_i represents the ε of operation O_i, and k is the total number of operations performed.*

Preuve du théorème 1. By Gaussian mechanism and the post-processing property, it is easily to see that releasing $\ddot{\Sigma}_{XX}$ satisfies $(\frac{\varepsilon}{2}, \frac{\delta}{2})$-DP, releasing $\dot{\Sigma}_{X\widetilde{Y}}$ satisfies $(\frac{\varepsilon}{2}, \frac{\delta}{2})$-DP. Thus, the output of Algorithm 1 is (ε, δ)-DP.

Next, we show that the output of the mechanism satisfies joint differential privacy using Billboard Lemma (Lemma 4). The estimators $\hat{\theta}^P(\hat{D}^0)$ and $\hat{\theta}^P(\hat{D}^1)$ are computed in the same way as $\hat{\theta}^P(\hat{D})$, so $\hat{\theta}^P(\hat{D}^0)$ and $\hat{\theta}^P(\hat{D}^1)$ each satisfy (ε, δ)-JDP. Since $\hat{\theta}^P(\hat{D}^0)$ and $\hat{\theta}^P(\hat{D}^1)$ are computed on disjoint subsets of the data, then by the Parallel Composition Theorem, together they satisfy $(\varepsilon, 2\delta)$-JDP. By the Sequential Composition Theorem (Lemma 17), the estimators $(\hat{\theta}^P(\hat{D}), \hat{\theta}^P(\hat{D}^0), \hat{\theta}^P(\hat{D}^1))$ together satisfy $(2\varepsilon, 3\delta)$-JDP.Finally, using the post-processing property and Billboard Lemma 4, the output $(\bar{\theta}^P(\hat{D}), \bar{\theta}^P(\hat{D}^0), \bar{\theta}^P(\hat{D}^1), \{\pi_i(D_i, \bar{\theta}^P(\hat{D}^b))\}_{i=1}^n)$ of Algorithm 2 satisfies $(2\varepsilon, 3\delta)$-JDP. □

Preuve du théorème 2. Suppose all agents other than i are following strategy $\sigma_{\tau_{\alpha,\beta}}$. Let agent i be in group $1 - b, b \in \{0, 1\}$. We will show that $\sigma_{\tau_{\alpha,\beta}}$ achieves η-Bayesian Nash equilibrium by bounding agent i's incentive to deviate. Assume that $c_i \leq \tau_{\alpha,\beta}$, otherwise there is nothing to show because agent i would be allowed to submit an arbitrary report

under $\sigma_{\tau_{\alpha,\beta}}$. For ease of notation, we write σ for $\sigma_{\tau_{\alpha,\beta}}$ for the remainder of the proof. We first compute the maximum expected amount (based on their belief) that agent i can increase their payment by misreporting to the analyst, i.e.

$$\begin{aligned} & E\left[\pi_i(\hat{D}_i, \sigma(D^b, c^b))|D_i, c_i\right] - E[\pi_i(D_i, \sigma(D^b, c^b))|D_i, c_i\Big] \\ = & E\left[B_{a_1,a_2}\left(\langle x_i, \bar{\theta}^P(\hat{D}^b)\rangle), \langle x_i, E_{\theta\sim p(\theta|\hat{D}_i)}[\theta]\rangle\right)|D_i, c_i\right] \\ & - E\left[B_{a_1,a_2}\left(\langle x_i, \bar{\theta}^P(\hat{D}^b)\rangle, \langle x_i, E_{\theta\sim p(\theta|D_i)}[\theta]\rangle\right)|D_i, c_i\right]. \end{aligned} \tag{23}$$

Note that $B_{a_1,a_2}(p,q) = a_1 - a_2(p - 2pq + q^2)$ is linear with respect to p, and is a strictly concave function of q maximized at $q = p$. Thus, 23 is upper bounded by the following with probability $1 - C_1 n^{-\Omega(1)}$

$$\begin{aligned} & B_{a_1,a_2}\left[E\left[\langle x_i, \bar{\theta}^P(\hat{D}^b)\rangle|D_i, c_i\right], E\left[\langle x_i, \hat{\theta}^P(\hat{D}^b)\rangle|D_i, c_i\right]\right] \\ & - B_{a_1,a_2}\left[E[\langle x_i, \bar{\theta}^P(\hat{D}^b)\rangle|D_i, c_i], \langle x_i, E_{\theta\sim p(\theta|D_i)}[\theta]\rangle\right] \\ = & a_2\left(E[\langle x_i, \bar{\theta}^P(\hat{D}^b)\rangle|D_i, c_i] - \langle x_i, E_{\theta\sim p(\theta|D_i)}[\theta]\rangle\right)^2 \\ = & a_2\left(E[\langle x_i, \bar{\theta}^P(\hat{D}^b)\rangle - \langle x_i, E_{\theta\sim p(\theta|D_i)}[\theta]\rangle|D_i, c_i]\right)^2 \\ \leq & a_2\left(E[x_i^T(\bar{\theta}^P(\hat{D}^b) - E_{\theta\sim p(\theta|D_i)}[\theta])|D_i, c_i]\right)^2 \\ \leq & a_2\|x_i\|_2^2\|E[\bar{\theta}^P(\hat{D}^b) - E_{\theta\sim p(\theta|D_i)}[\theta]|D_i, c_i]\|_2^2 \\ \leq & r^2 a_2\|E[\bar{\theta}^P(\hat{D}^b) - E_{\theta\sim p(\theta|D_i)}[\theta]|D_i, c_i]\|_2^2. \end{aligned}$$

We continue by bounding the term $\|E[\bar{\theta}^P(\hat{D}^b) - E_{\theta\sim p(\theta|D_i)}[\theta]|D_i, c_i]\|_2$. By Lemma **??**

$$\begin{aligned} & \|E[\bar{\theta}^P(\hat{D}^b) - E_{\theta\sim p(\theta|D_i)}[\theta]|D_i, c_i]\|_2 \\ & \leq \|E[\bar{\theta}^P(\hat{D}^b) - \bar{\theta}^P(D^b)|D_i, c_i]\|_2 + \|E[\bar{\theta}^P(D^b)|D_i, c_i] - E_{\theta\sim p(\theta|D_i)}[\theta]|D_i, c_i]\|_2 \\ & \leq \|E[\hat{\theta}^P(\hat{D}^b) - \hat{\theta}^P(D^b)|D_i, c_i]\|_2 + \|E[\bar{\theta}^P(D^b)|D_i, c_i] - E_{\theta\sim p(\theta|D_i)}[\theta]|D_i, c_i]\|_2 \\ & \leq \|\hat{\theta}^P(\hat{D}^b) - \hat{\theta}^P(D^b)\|_2 + \|E[\bar{\theta}^P(D^b)|D_i] - E_{\theta\sim p(\theta|D_i)}[\theta]\|_2 \end{aligned}$$

For the first term of **??**, since agent i believes that with at least probability $1-\beta$, at most αn agents will misreport their datasets under threshold strategy $\sigma_{\tau_{\alpha,\beta}}$, datasets D^b and $\hat{D}^b$ differ only on at most αn agents' datasets. By Lemma **??** and Lemma **??**, we set the constraint bound $\lambda_n = O\left(\frac{r^2\sqrt{\log d \log\frac{1}{\delta}}}{\sqrt{n}\varepsilon}\right)$, when n is sufficient large such that $n \geq \Omega(\frac{s^2 r^4 \log d \log\frac{1}{\delta}}{\varepsilon^2 \kappa_\infty})$, with probability at least $1 - \beta - O(\alpha n d^{-\Omega(1)})$ we have that

$$\mathbb{E}\|\hat{\theta}^P(\hat{D}) - \hat{\theta}^P(D)\|_2 \leq 16\sqrt{k}\lambda_n \tag{24}$$

For the second term of (**??**) :

$$\begin{aligned} E[\bar{\theta}^P(D^b)|D_i] - E_{\theta\sim p(\theta|D_i)}[\theta] &= E_{D^b\sim p(D^b|D_i)}[\bar{\theta}^P(D^b)] - E_{\theta\sim p(\theta|D_i)}[\theta] \\ &= E_{\theta\sim p(\theta|D_i)}[E_{D^b\sim p(D^b|\theta)}[\bar{\theta}^P(D^b)]|\theta] - E_{\theta\sim p(\theta|D_i)}[\theta] \\ &= E_{\theta\sim p(\theta|D_i)}[E_{D^b\sim p(D^b|\theta)}[\bar{\theta}^P(D^b) - \theta]|\theta]. \end{aligned}$$

Since

$$p(D^b|\theta) = p(X^b, y^b|\theta) = p(y^b|X^b, \theta)p(X^b|\theta) = p(y^b|X^b, \theta)p(X^b),$$

we have

$$E_{D^b\sim p(D^b|\theta)}[\hat{\theta}(D^b) - \theta] = E_{X^b}[E_{y^b}[\bar{\theta}^P(X^b, y^b) - \theta]|X^b, \theta].$$

Since we have the prior knowledge that $\|\theta^*\|_2 \leq \tau_\theta$. Thus, for the posterior distribution $\theta \sim p(\theta|\hat{D}_i)$ it will also have $\|\theta\|_2 \leq \tau_\theta$. By Jensen's inequality, Theorem 5 and Lemma **??**, we have

$$\begin{aligned}\|E[\bar{\theta}^P(D^b)|D_i] - E_{\theta\sim p(\theta|D_i)}[\theta]\|_2 &\leq E_{\theta\sim p(\theta|D_i),X^b}[E_{y^b}[\|\bar{\theta}^P(X^b, y^b) - \theta\|_2|X^b, \theta]] \\ &\leq E_{\theta\sim p(\theta|D_i),X^b}[E_{y^b}[\|\hat{\theta}^P(X^b, y^b) - \theta\|_2|X^b, \theta]] \\ &\leq O\left(\frac{r^2\sqrt{k\log d\log\frac{1}{\delta}}}{\sqrt{n}\varepsilon}\right)\end{aligned}$$

In addition to an increased payment, agent i may also experience decreased privacy costs from misreporting. By Assumption **??**, this decrease in privacy costs is bounded above by $c_i 4\varepsilon^2$. Since we have assumed $c_i \leq \tau_{\alpha,\beta}$, the decrease in privacy costs for agent i is bounded above by $\tau_{\alpha,\beta}4\varepsilon^2$. Hence, agent i's total incentive to deviate is bounded above by

$$\eta = O\left(\frac{a_2 n\alpha^2 r^6 k\log d\log\frac{1}{\delta}}{\varepsilon^2} + \tau_{\alpha,\beta}\varepsilon^2\right).$$

□

Preuve du théorème 3. Let agent i have privacy cost $c_i \leq \tau_{\alpha,\beta}$ and consider agent i's utility from participating in the mechanism. Suppose agent i is in group $1-b$, then their expected utility is

$$\begin{aligned}E[u_i] &= E\left[B_{a_1,a_2}\left(\langle x_i, \bar{\theta}^P(\hat{D}^b)\rangle, \langle x_i, E_{\theta\sim p(\theta|\hat{D}_i)}[\theta]\rangle\right)|D_i, c_i\right] - f_i(c_i, \varepsilon) \\ &\geq B_{a_1,a_2}\left(E(\langle x_i, \bar{\theta}^P(\hat{D}^b)\rangle)|D_i, c_i, \langle x_i, E_{\theta\sim p(\theta|\hat{D}_i)}[\theta]\rangle\right) - \tau_{\alpha,\beta}4(1+\delta)\varepsilon^2.\end{aligned}$$

Note that

$$B_{a_1,a_2}(p, q) = a_1 - a_2(p - 2pq + q^2) \geq a_1 - a_2(|p| + 2|p||q| + |q|^2),$$

Since both $|\langle x_i, \bar{\theta}^P(\hat{D}^b)\rangle|$ and $|\langle x_i, E_{\theta\sim p(\theta|\hat{D}_i)}[\theta]\rangle|$ are bounded by $\|x_i\|_2\|\hat{\theta}(\hat{D}^b)\|_2 \leq r\tau_\theta$, thus by (**??**) and (25) agent i's expected utility is non-negative as long as

$$a_1 \geq a_2(r\tau_\theta + 3r^2\tau_\theta^2) + \tau_{\alpha,\beta}4(1+\delta)\varepsilon^2.$$

□

Preuve du théorème 4. Note that

$$B_{a_1,a_2}(p,q) \leq B_{a_1,a_2}(p,p) = a_1 - a_2(p-p^2) \leq a_1 + a_2(|p| + |p|^2),$$

thus

$$\begin{aligned}\mathcal{B} = \sum_{i=1}^{n} E[\pi_i] &= \sum_{i=1}^{n} E[B_{a_1,a_2}\left(\langle x_i, \bar{\theta}^P(\hat{D}^b)\rangle, \langle x_i, E_{\theta\sim p(\theta|\hat{D}_i)}[\theta]\rangle\right)|D_i, c_i] \\ &\leq n(a_1 + a_2(r\tau_\theta + r^2\tau_\theta^2)).\end{aligned}$$

□

A.2 Preuves des théorèmes de la partie 3.2

Les résultats trouvés dans cette partie sont basé sur un premier théorème, qui donne une estimation robuste et non-privée de la moyenne de données.

Théorème 18 (Estimation robuste et non-privée de la moyenne)**.** *Soient n échantillons i.i.d. x_i où $i = 1, \ldots, n$, échantillonnés à partir d'un modèle de contamination forte et la distribution des inliers est une distribution à queue lourde P avec un moment brut fini d'ordre k et $\alpha \in (0,1)$. Soit $\bar{x}_i$ la variable aléatoire tronquée, c'est-à-dire $\bar{x}_i = x_i I_{(|x_i|\leq M)}$. Alors, avec une probabilité d'au moins $1-\delta$,*

$$\left|\frac{1}{n}\sum_{i\in[n]} \bar{x}_i - \mu\right| \leq \sqrt{\frac{2\log(2/\delta)}{n}} + \frac{4M\log(2/\delta)}{3n} + \frac{1}{M^{k-1}} + 2\alpha M. \tag{25}$$

Nous choisissons $M = \min\left(\left(\frac{3n}{4\log(1/\delta)}\right)^{1/k}, (2\alpha)^{-1/k}\right)$, alors nous avons

$$\left|\frac{1}{n}\sum_{i\in[n]} \bar{x}_i - \mu\right| \leq \sqrt{\frac{2\log(2/\delta)}{n}} + 2\left(\frac{4\log(1/\delta)}{3n}\right)^{1-1/k} + 2(2\alpha)^{1-1/k}. \tag{26}$$

Preuve du théorème 18. Under the α-fraction contamination model, we can decompose $\mathcal{T} := \left|\frac{1}{n}\sum_{i=1}^{n} \bar{x}_i - \mu\right|$ term into

$$\left|\frac{1}{n}\sum_{i\in G} \bar{x}_i + \frac{1}{n}\sum_{i\in B} \bar{x}_i - \mu\right|, \tag{27}$$

where the set G and B represent the indices that are not contaminated (Good) and contaminated (Bad), respectively and $|B| \leq \alpha n$. Note that those samples in G may not be *i.i.d.* We have :

$$\mathcal{T} \leq \underbrace{\left|\frac{1}{n}\sum_{i\in G} \bar{x}_i - \mu\right|}_{1} + \underbrace{\left|\frac{1}{n}\sum_{i\in B} \bar{x}_i\right|}_{2}. \tag{28}$$

For $_2$, by truncation, we know

$$\mathcal{T}_2 \leq \frac{1}{n}\sum_{i\in B}|\bar{x}_i| = \frac{1}{n}\sum_{i\in B}|x_i|I_{(|x_i|\leq M)} \leq \alpha M.$$

For $_1$, we have :

$$\left|\frac{1}{n}\sum_{i\in G}\bar{x}_i - \mu\right| = \left|\frac{1}{n}\sum_{i\in G\cup B}\bar{x}_i - \mu - \frac{1}{n}\sum_{i\in B}\bar{x}_i\right| \tag{29}$$

$$\leq \left|\frac{1}{n}\sum_{i\in G\cup B}\bar{x}_i - \mu\right| + \left|\frac{1}{n}\sum_{i\in B}\bar{x}_i\right| \tag{30}$$

$$\leq \underbrace{\left|\frac{1}{n}\sum_{x\in\chi_n} xI_{(|x|\leq M)} - \mu\right|}_{3} + \alpha M. \tag{31}$$

where the last inequality is based on the upper bound result of $_2$. Now, $_3$ is the standard term (i.e., sum of n *i.i.d* samples from the true distribution P with truncation), which can be upper bounded as follows

$$_3 = \left|\frac{1}{n}\sum_{x\in\chi_n} xI_{(|x|\leq M)} - \mu\right| \tag{32}$$

$$= \left|\frac{1}{n}\sum_{x\in\chi_n} xI_{(|x|\leq M)} - \mathbb{E}_{x\sim P}[xI_{(|x|\leq M)}] + \mathbb{E}_{x\sim P}[xI_{(|x|\leq M)}] - \mu\right| \tag{33}$$

$$\leq \left|\frac{1}{n}\sum_{x\in\chi_n} xI_{(|x|\leq M)} - \mathbb{E}_{x\sim P}[xI_{(|x|\leq M)}]\right| + \left|\mathbb{E}_{x\sim P}[xI_{(|x|\leq M)}] - \mathbb{E}_{x\sim P}[x]\right| \tag{34}$$

$$\overset{(a)}{\leq} \sqrt{\frac{2\log(2/\delta)}{n}} + \frac{4M\log(2/\delta)}{3n} + (\mathbb{E}_{x\sim P}[|X_i|^k])^{\frac{1}{k}}(\mathbb{P}_{X_i\sim P_k}(|X_i| > M))^{\frac{k-1}{k}} \quad \text{w.p.} \quad 1-\delta \tag{35}$$

$$\overset{(b)}{\leq} \sqrt{\frac{2\log(2/\delta)}{n}} + \frac{4M\log(2/\delta)}{3n} + \frac{1}{M^{k-1}} \tag{36}$$

where the inequality (a) is based on Berstein's inequality in Lemma **??** and Hölder's Inequality in Lemma **??** and (b) is from Markov's inequality in Lemma **??**.

Putting them together yields with probability at least $1-\delta$

$$\leq \sqrt{\frac{2\log(2/\delta)}{n}} + \frac{4M\log(2/\delta)}{3n} + \frac{1}{M^{k-1}} + 2\alpha M. \tag{37}$$

□

Preuve du théorème 8.

$$|\tilde{\mu} - \mu| \leq \left|\frac{1}{n}\sum_{i=1}^{n}\bar{x}_i - \mu\right| + \left|\frac{1}{n}\sum_{i=1}^{n}\bar{x}_i - \tilde{\mu}\right|$$

Based on inequality (4) of [CZ22] which is the concentration of private noise introduced by shuffle protocol in Algorithm 5, we have with probability at least $1-2\delta$

$$\left|\frac{1}{n}\sum_{i=1}^{n}\bar{x}_i-\tilde{\mu}\right|\leq\frac{M\sqrt{2\log(2/\delta)}}{n\epsilon}+\frac{M\log(2/\delta)}{n\epsilon}$$

Thus, based on Theorem 18, we have with probability at least $1-3\delta$

$$|\tilde{\mu}-\mu|\leq\sqrt{\frac{2\log(2/\delta)}{n}}+\frac{1}{M^{k-1}}+2\alpha M+\frac{4M\log(1/\delta)}{n\epsilon}$$

where δ is small enough. □

Proof of Regret bound under CDP. For each $a\in V$, we have with probability at least $1-\frac{\delta}{2|V|t^2}$

$$\left|\tilde{\mu}_a^t-\mu_a\right|\leq\beta^t.$$

Note that here $\tilde{\mu}_a^t$ is the empirical mean in epoch t, and the batch size in epoch t is 2^t. Then the upper bound of empirical mean in the algorithm 6 is a little different from the upper bound for mean estimation for offline data due to the assumption of (8).

This is for the bad data in epoch t,

$$\left|\frac{1}{N^t}\sum_{i\in B}\bar{x}_i\right|\leq\frac{1}{N^t}\sum_{i\in B_t}|\bar{x}_i|\leq\frac{M|B_t|}{N^t}\leq\frac{M\sum_{i=1}^{t}|B_t|}{N^t}\leq\frac{M\alpha\sum_{i=1}^{t}N^t}{N^t}\leq\frac{2M\alpha N^t}{N^t}\leq 2M\alpha$$

we have with probability at least $1-2\delta$

$$|\tilde{\mu}_a^t-\mu_a|\leq\sqrt{\frac{2\log(2/\delta)}{N^t}}+\frac{1}{M^{k-1}}+3\alpha M+\frac{4M\log(2/\delta)}{\epsilon N^t}\tag{38}$$

We choose $M=\min((\frac{N^t\epsilon}{4\log(1/\delta)})^{1/k},(3\alpha)^{-1/k})$, we have

$$\beta^t=\sqrt{\frac{2\log(2/\delta)}{N^t}}+2\left(\frac{4\log(1/\delta)}{N^t\epsilon}\right)^{1-1/k}+2(3\alpha)^{1-1/k}.\tag{39}$$

We denote by $\mathcal{E}_t$ the event where for all $a\in V$ it holds that $|\hat{\mu}_a^t-\mu_a|\leq\beta^t$, and denote $\mathcal{E}=\cup\mathcal{E}_t$. By taking union-bound, we have

$$\mathbb{P}(\mathcal{E}_t)\geq 1-\frac{\delta}{2t^2},$$

and

$$\mathbb{P}(\mathcal{E})\geq 1-\frac{\delta}{2}\left(\sum_t t^{-2}\right)\geq 1-\delta.$$

In the following, we condition on the good event $\mathcal{E}$. We first show that the optimal arm a^* will always be active. We show this by contradiction. Suppose at the end of some batch b, a^* will be eliminated, i.e., $\mathrm{UCB}_{a^*}(t)<\mathrm{LCB}_{a'}(t)$ for some a'. This implies that under good event $\mathcal{E}$

$$\mu_{a^*}\leq\tilde{\mu}_{a^*}^t+\beta^t<\tilde{\mu}_{a'}^t-\beta^t\leq\mu_{a'}$$

which contradicts the fact that a^* is the optimal arm.

Then, we show that at the end of batch t, all arms such that $\Delta_a > 4\beta^t$ will be eliminated. To show this, we have that under good event $\mathcal{E}$:

$$\tilde{\mu}_a^t + \beta^t \leq \mu_a^t + 2\beta^t < \mu_{a^*}^t - 4\beta^t + 2\beta^t \leq \tilde{\mu}_{a^*}^t - \beta^t,$$

which implies that arm a will be eliminated by the rule. Thus, for each sub-optimal arm a, let $\tilde{t}_a$ be the last batch that arm a is not eliminated. By the above result, we have

$$\Delta_a \leq 4\beta^{\tilde{t}_a} \leq O\left(\sqrt{\frac{\log(|V|t^2/\delta)}{N^{\tilde{t}_a}}} + \left(\frac{\log(|V|t^2/\delta)}{\epsilon N^{\tilde{t}_a}}\right)^{1-1/k} + \alpha^{1-1/k}\right). \tag{40}$$

We divide the arms $a \in [K]$ into two groups : $\mathcal{G}_1 = \left\{a \in [K] : 2\alpha^{1-\frac{1}{k}} \leq \Delta_a\right\}$ and $\mathcal{G}_2 = \left\{a \in [K] : 2\alpha^{1-\frac{1}{k}} \geq \Delta_a\right\}$.

Group 1 : Now, for all arm $a \in \mathcal{G}_1$, we have

$$\Delta_a \leq O\left(\sqrt{\frac{\log(|V|t^2/\delta)}{N^{\tilde{t}_a}}} + \left(\frac{\log(|V|t^2/\delta)}{\varepsilon N^{\tilde{t}_a}}\right)^{1-1/k}\right).$$

Hence, we have for some absolute constants c_1 and c_2 :

$$N^{\tilde{t}_a} \leq \max\left\{\frac{c_1 \log\left(|V|t^2/\delta\right)}{\Delta_a^2}, \frac{c_2 \log\left(|V|t^2/\delta\right)}{\varepsilon \Delta_a^{\frac{k}{k-1}}}\right\}.$$

Since $|V| \leq K$ and $2^t = N^t \leq T$ for any batch t. Thus,

$$N^{\tilde{t}_a} \leq \max\left\{\frac{c_1 \log\left(K \log^2 T/\delta\right)}{\Delta_a^2}, \frac{c_2 \log\left(K \log^2 T/\delta\right)}{\varepsilon \Delta_a^{\frac{k}{k-1}}}\right\}.$$

Since the batch size doubles, we have $N_a(T) \leq 4N^{\tilde{t}_a}$ for each sub-optimal arm a. Therefore, for all arm $a \in \mathcal{G}_1$,

$$\mathcal{R}_T = \sum_{a \in \mathcal{G}_1} N_a(T)\Delta_a \leq 4N^{\tilde{t}_a}\Delta_a$$

Let η be a number in $(0, 1)$. For all arms $a \in \mathcal{G}_1$ with $\Delta_a \leq \eta$, the regret incurred by pulling these arms is upper bounded by $T\eta$. For any arm $a \in \mathcal{G}_1$ with $\Delta_a > \eta$, choose $\delta = \frac{1}{T}$ and assume $T \geq K$, then the expected regret incurred by pulling arm a is upper bounded by :

$$\begin{aligned}
&\mathbb{E}\left[\sum_{a \in \mathcal{G}_1, \Delta_a > \eta} \Delta_a N_a(T)\right] \\
&\leq \mathbb{P}(\overline{\mathcal{E}}) \cdot T + O\left(\sum_{a \in \mathcal{G}_1, \Delta_a > \eta}\left\{\frac{\log T}{\Delta_a} + \frac{\log T}{\varepsilon \Delta_a^{\frac{1}{k-1}}}\right\}\right) \\
&\leq O\left(\frac{K \log T}{\eta} + \frac{K \log T}{\varepsilon \eta^{\frac{1}{k-1}}}\right).
\end{aligned}$$

Thus the regret from group 1 is at most :

$$T\eta + O\left(\frac{K\log T}{\eta} + \frac{K\log T}{\varepsilon\eta^{\frac{1}{k-1}}}\right).$$

Taking $\eta = \max\left\{\sqrt{\frac{K\log T}{T}}, \left(\frac{K\log T}{T\varepsilon}\right)^{\frac{k-1}{k}}\right\}$, the regret from group 1 is at most :

$$O\left(\sqrt{KT\log T} + \left(\frac{K\log T}{\varepsilon}\right)^{\frac{k-1}{k}} T^{\frac{1}{k}}\right).$$

Group 2 : For all other arms $a \in \mathcal{G}_2$, we have the total regret is at most $O(T\Delta_a) = O\left(T\alpha^{1-\frac{1}{k}}\right)$.
Combine the two groups, choose $\delta = \frac{1}{T}$ and assume $T \geq K$, we have the that the expected regret satisfies :

$$\mathcal{R}_T \leq O\left(\sqrt{KT\log T} + \left(\frac{K\log T}{\varepsilon}\right)^{\frac{k-1}{k}} T^{\frac{1}{k}} + T\alpha^{1-\frac{1}{k}}\right).$$

□

Proof of Regret bound under LDP. Similar to the analysis of CDP case and based on the concentration inequality of truncated mean estimation under LDP in Theorem 7, we can get

$$\Delta_a \leq 4\beta^{\tilde{t}_a} \leq O\left(\sqrt{\frac{\log(|V|t^2/\delta)}{N^{\tilde{t}_a}}} + \left(\frac{\sqrt{\log(|V|t^2/\delta)}}{\epsilon\sqrt{N^{\tilde{t}_a}}}\right)^{1-1/k} + \left(\frac{\log(|V|t^2/\delta)}{N^{\tilde{t}_a}}\right)^{1-1/k} + \alpha^{1-1/k}\right). \tag{41}$$

We divide the arms $a \in [K]$ into two groups : $\mathcal{G}_1 = \left\{a \in [K] : 2\alpha^{1-\frac{1}{k}} \leq \Delta_a\right\}$ and $\mathcal{G}_2 = \left\{a \in [K] : 2\alpha^{1-\frac{1}{k}} \geq \Delta_a\right\}$.
Group 1 : Now, for all arm $a \in \mathcal{G}_1$, we have

$$\Delta_a \leq O\left(\sqrt{\frac{\log(|V|t^2/\delta)}{N^{\tilde{t}_a}}} + \left(\frac{\sqrt{\log(|V|t^2/\delta)}}{\varepsilon\sqrt{N^{\tilde{t}_a}}}\right)^{1-1/k} + \left(\frac{\log(|V|t^2/\delta)}{N^{\tilde{t}_a}}\right)^{1-1/k}\right).$$

Hence, we have for some absolute constants c_1 , c_2 and c_3 :

$$N^{\tilde{t}_a} \leq \max\left\{\frac{c_1\log\left(|V|t^2/\delta\right)}{\Delta_a^2}, \frac{c_2\log\left(|V|t^2/\delta\right)}{\varepsilon^2\Delta_a^{\frac{2k}{k-1}}}, \frac{c_3\log\left(|V|t^2/\delta\right)}{\Delta_a^{\frac{k}{k-1}}}\right\}.$$

Since $|V| \leq K$ and $2^t = N^t \leq T$ for any batch t. Thus,

$$N^{\tilde{t}_a} \leq \max\left\{\frac{c_1\log\left(K\log^2 T/\delta\right)}{\Delta_a^2}, \frac{c_2\log\left(K\log^2 T/\delta\right)}{\varepsilon^2\Delta_a^{\frac{2k}{k-1}}}, \frac{c_3\log\left(K\log^2 T/\delta\right)}{\Delta_a^{\frac{k}{k-1}}}\right\}.$$

Since the batch size doubles, we have $N_a(T) \leq 4N^{\tilde{t}_a}$ for each sub-optimal arm a. Therefore, for all arm $a \in \mathcal{G}_1$,

$$\mathcal{R}_T = \sum_{a \in \mathcal{G}_1} N_a(T)\Delta_a \leq 4N^{\tilde{t}_a}\Delta_a$$

Let η be a number in $(0, 1)$. For all arms $a \in \mathcal{G}_1$ with $\Delta_a \leq \eta$, the regret incurred by pulling these arms is upper bounded by $T\eta$. For any arm $a \in \mathcal{G}_1$ with $\Delta_a > \eta$, choose $\delta = \frac{1}{T}$ and assume $T \geq K$, then the expected regret incurred by pulling arm a is upper bounded by :

$$\begin{aligned}
&\mathbb{E}\left[\sum_{a \in \mathcal{G}_1, \Delta_a > \eta} \Delta_a N_a(T)\right] \\
&\leq \mathbb{P}(\overline{\mathcal{E}}) \cdot T + O\left(\sum_{a \in \mathcal{G}_1, \Delta_a > \eta} \left\{\frac{\log T}{\Delta_a} + \frac{\log T}{\varepsilon^2 \Delta_a^{\frac{k+1}{k-1}}} + \frac{\log T}{\Delta_a^{\frac{1}{k-1}}}\right\}\right) \\
&\leq O\left(\frac{K \log T}{\eta} + \frac{K \log T}{\varepsilon^2 \eta^{\frac{k+1}{k-1}}} + \frac{K \log T}{\eta^{\frac{1}{k-1}}}\right).
\end{aligned}$$

Thus the regret from group 1 is at most :

$$T\eta + O\left(\frac{K \log T}{\eta} + \frac{K \log T}{\varepsilon^2 \eta^{\frac{k+1}{k-1}}} + + \frac{K \log T}{\eta^{\frac{1}{k-1}}}\right).$$

Taking $\eta = \max\left\{\sqrt{\frac{K \log T}{T}}, \left(\frac{K \log T}{T\varepsilon^2}\right)^{\frac{k-1}{2k}}, \left(\frac{K \log T}{T}\right)^{\frac{k-1}{k}}\right\}$, the regret from group 1 is at most :

$$O\left(\sqrt{KT \log T} + \left(\frac{K \log T}{\varepsilon^2}\right)^{\frac{k-1}{2k}} T^{\frac{k+1}{2k}} + (K \log T)^{\frac{k-1}{k}} T^{\frac{1}{k}}\right).$$

Group 2 : For all other arms $a \in \mathcal{G}_2$, we have the total regret is at most $O(T\Delta_a) = O\left(T\alpha^{1-\frac{1}{k}}\right)$.
Combine the two groups, choose $\delta = \frac{1}{T}$ and assume $T \geq K$, we have the that the expected regret satisfies :

$$\mathcal{R}_T \leq O\left(\sqrt{KT \log T} + \left(\frac{K \log T}{\varepsilon^2}\right)^{\frac{k-1}{2k}} T^{\frac{k+1}{2k}} + T\alpha^{1-\frac{1}{k}} + (K \log T)^{\frac{k-1}{k}} T^{\frac{1}{k}}\right).$$

□

Proof of Regret bound under SDP 9. Similar to the proof of regret under the CDP model and based on Theorem 7, we have for SDP model, the regret upper bound is

$$\mathcal{R}_T \leq O\left(\sqrt{KT \log T} + \left(\frac{K \log T}{\varepsilon}\right)^{\frac{k-1}{k}} T^{\frac{1}{k}} + T\alpha^{1-\frac{1}{k}}\right).$$

□

A.3 Annexe relative à la leçon scientifique

Preuve du lemme 2. Considérons les mêmes paramètres que dans l'énoncé du lemme. Pour t fixé, nous avons $\sum_{a\in\mathcal{A}} \mathbb{I}\{A_t = a\} = 1$. Ainsi nous pouvons écrire : $S_T = \sum_{t=1}^{T} X_t = \sum_{t}^{T}\sum_{a\in\mathcal{A}} X_t \mathbb{I}\{A_t = a\}$ et donc d'après la définition du regret et par linéarité de l'espérance, nous avons :

$$R_T = T\mu^* - \mathbb{E}[S_n] = \sum_{t=1}^{T} \mathbb{E}\left(\mu^* - X_t\right) = \sum_{a\in\mathcal{A}}\sum_{t=1}^{T} \mathbb{E}\left[(\mu^* - X_t)\mathbb{I}\{A_t = a\}\right] = \sum_{a\in\mathcal{A}}\sum_{t=1}^{T} \mathbb{E}\left(\mathbb{E}[(\mu^* - X_t)\mathbb{I}\{A_t = a\}]|A_t\right) \tag{42}$$

Comme la récompense au tour t sachant que le joueur a tiré l'action A_t, par propriété de l'espérance conditionelle nous avons :

$$\begin{aligned}
\mathbb{E}[(\mu^* - X_t)\mathbb{I}\{A_t = a\}|A_t] &= \mathbb{I}\{A_t = a\}\mathbb{E}[\mu^* - X_t|A_t] \\
&= \mathbb{I}\{A_t = a\}(\mu^* - \mu_{A_t}) \\
&= \mathbb{I}\{A_t = a\}(\mu^* - \mu_a) \\
&= \mathbb{I}\{A_t = a\}\Delta_a
\end{aligned}$$

En reprenant l'équation 42, nous avons :

$$\begin{aligned}
R_T &= \sum_{a\in\mathcal{A}}\sum_{t=1}^{T} \mathbb{E}\left[(\mu^* - X_t)\mathbb{I}\{A_t = a\}\right] = \sum_{a\in\mathcal{A}}\sum_{t=1}^{T} \mathbb{E}\left[\mathbb{I}\{A_t = a\}\Delta_a\right] \\
&= \sum_{a\in\mathcal{A}} \Delta_a \mathbb{E}\left(\sum_{t=1}^{T} \mathbb{I}\{A_t = a\}\right) = \sum_{a\in\mathcal{A}} \Delta_a \mathbb{E}\left[N_a(t)\right]
\end{aligned}$$

Preuve du théorème 13. En reprenant l'idée de [LS20] et sans perte de généralité, on assume que le bras premier bras et le bras optimal, c'est-à-dire que $\mu_1 = \mu^* = \max_i \mu_i$. D'après le lemme de décomposition du regret 2, nous avons :

$$R_T = \sum_{i=1}^{T} \Delta_i \mathbb{E}\left[N_i(t)\right]$$

Commencons par trouver une borne supérieure à $\mathbb{E}\left[N_i(t)\right]$:

$$\mathbb{E}\left[N_i(t)\right] = \mathbb{E}\left[\sum_{s=1}^{T} \mathbb{I}\{A_s = i\}\right] = \mathbb{E}\left[\sum_{s=1}^{mk} \mathbb{I}\{A_s = i\}\right] + \mathbb{E}\left[\sum_{s=mk+1}^{T} \mathbb{I}\{A_s = i\}\right]$$

Les mk premiers tours sont déterministes, l'algorithme choisit exactement m fois chaque action, donc l'action i est également choisie m fois. Pour les $T - mk$ tours restants l'action qui est jouée est celle ayant maximisé la récompense moyenne durant l'exploration. On a donc :

$$\begin{aligned}
\mathbb{E}\left[T_i(n)\right] &= m + (T - mk)\mathbb{P}\left(A_{mk+1} = i\right) \\
&\leq m + (T - mk)\mathbb{P}\left(\hat{\mu}_i(mk) \geq \max_{j\neq i} \hat{\mu}_j(mk)\right)
\end{aligned}$$

De plus, on a :

$$\begin{aligned}\mathbb{P}\left(\hat{\mu}_i(mk) \geq \max_{j\neq i} \hat{\mu}_j(mk)\right) &\leq \mathbb{P}\left(\hat{\mu}_i(mk) \geq \hat{\mu}_1(mk)\right) \\ &= \mathbb{P}\left(\hat{\mu}_i(mk) - \mu_i - (\hat{\mu}_1(mk) - \mu_1) \geq \Delta_i\right)\end{aligned}$$

D'après la défintion d'une variable aléatoire sous-gaussienne et la définition des $(\hat{\mu}_j)_j$, nous pouvons affirmer que $\hat{\mu}_i(mk) - \mu_i - (\hat{\mu}_1(mk) - \mu_1)$ est $\sqrt{2/m}$-sous-gaussienne,
Donc d'après le corollaire 4.2.2,

$$\mathbb{P}\left(\hat{\mu}_i(mk) - \mu_i - \hat{\mu}_1(mk) + \mu_1 \geq \Delta_i\right) \leq \exp\left(-\frac{m\Delta_i^2}{4}\right)$$

En reprenant toutes ces inégalités dans l'équation de départ on retrouve le résultat du théorème.

Listing 1 – Code Python de la simulation Explore-then-Commit

```
##Explore-then-commit
import matplotlib.pyplot as plt
import random
import numpy

#Definition de la classe ETC
class ETC():
    def __init__(self, counts = [], values = [], n_arms = 0):
        self.n_arms = n_arms
        self.counts = [0 for col in range(n_arms)]
        self.values = [0.0 for col in range(n_arms)]

    def select_arm(self):
        """ Selectionne le bras avec la valeur de l'estimateur la plus
            haute"""
        for arm in range(self.n_arms):
            if self.counts[arm] == 0:
                return arm
        value_max = max(self.values)
        return self.values.index(value_max)

    def update(self, chosen_arm, reward):
        self.counts[chosen_arm] += 1
        n = self.counts[chosen_arm]
        value = self.values[chosen_arm]
        new_value = ((n - 1) /float(n)) * value + (reward / float(n))
        self.values[chosen_arm] =  new_value


##Partie simulation
#Simulation avec deux bras de loi de Bernoulli
```

```
31 def bernoulli_ETC(iteration, proba_arm_1, proba_arm_2):
32     print("UCB1 : cas 2 bras, suivant une loi de Bernoulli, arm1 : %s,
          arm2 : %s" % (proba_arm_1, proba_arm_2))
33     vector_arms_chosen = []
34     reward_vect = []
35     reward_cum = 0
36     algo = ETC([],[], 2)
37     i = 0
38     while i < iteration :
39         chosen_arm = algo.select_arm()
40         vector_arms_chosen.append(chosen_arm)
41         if chosen_arm == 0 :                          #si le bras 1 est
              choisi
42             if random.random() < proba_arm_1 :
43                 reward = 1
44             else :
45                 reward = 0
46         else :                                         #si le bras 2 est
              choisi
47             if random.random() < proba_arm_2 :
48                 reward = 1
49             else :
50                 reward = 0
51         algo.update(chosen_arm, reward)
52         reward_cum = reward_cum + reward
53         i += 1
54         reward_vect.append(reward_cum / float(i))
55     return reward_vect, algo.values, algo.counts, vector_arms_chosen
56
57 #Parametres
58 nb_iteration = 100
59 arm_1_probability = 0.7
60 arm_2_probability = 0.3
61
62 #Simulation
63 rewards_vect, values, count, arms_chosen = bernoulli_ETC(nb_iteration,
       arm_1_probability, arm_2_probability)
64 offline_solution = [max(arm_1_probability, arm_2_probability) for x in
       range(nb_iteration)]
65
66 #Affichage nombre de tirages de chaque bras
67 cum_sum_arm_2 = numpy.cumsum(arms_chosen) #arm_chosen = 1 si arm2
     selected, else arm_chosen = 0
68 cum_sum_arm_1 = [x - cum_sum_arm_2[x] for x in range(nb_iteration)]
69 plt.plot(range(nb_iteration),cum_sum_arm_1, range(nb_iteration),
     cum_sum_arm_2);
70 plt.title("Nombre de tirages de chaque bras");
```

```
plt.legend(["arm1 Proba %s" % arm_1_probability, "arm2 Proba %s" %
    arm_2_probability], loc="upper left");
plt.figure();

#Affichage gain moyen par tour
plt.plot(range(nb_iteration), offline_solution, 'k-', range(
    nb_iteration), rewards_vect, 'r--');
plt.title("Gain moyen par tour");
plt.legend(["Offline","ETC"], loc="lower right");
plt.figure();
```

Listing 2 – Code Python de la simulation Epsilon-Greedy

```
##Epsilon-greedy
import matplotlib.pyplot as plt
import random
import numpy

#Definition de la classe Epsilon-greedy
class EpsilonGreedy():
    def __init__(self, epsilon=0.05, counts=[], values=[], n_arms = 0)
        :
        """Recupere les parametres specifies lors de la creation d un
            objet de cette classe"""
        self.epsilon = epsilon
        self.n_arms = n_arms
        self.counts = [0 for col in range(n_arms)]       #nombre de
            fois que chaque bras a   t   choisi
        self.values = [0.0 for col in range(n_arms)]     #valeur des
            r compenses pour chaque bras tir s

    def select_arm(self):
        """S lectionne le bras avec la valeur la plus haute si random
            .radom() > Epsilon, choisit un bras au hasard sinon"""
        if sum(self.values) == 0:
            return(random.choice([i for i in range(self.n_arms)]))
        else :
            if random.random()>self.epsilon:
                return(self.values.index(max(self.values)))       #
                    renvoie le bras le plus prometteur avec proba 1-
                    epsilon
            else:
                return(random.choice([i for i in range(self.n_arms)]))

    def update(self, chosen_arm, reward):
        """Met a jour la liste self.values en fonction des recompenses
            obtenues"""
```

```
27          self.counts[chosen_arm]+=1
28          new_value = ((self.counts[chosen_arm]-1)/float(self.counts[
                chosen_arm])) * self.values[chosen_arm] + (reward /float(
                self.counts[chosen_arm]))
29          self.values[chosen_arm] = new_value
30
31  ##Partie simulation
32  #Simulation avec deux bras de loi de Bernoulli
33
34  def simulation_arm_bernoulli(proba):
35      """Cette fonction modelise la loi des differents bras"""
36      if random.random() < proba :
37          return(1)
38      else:
39          return(0)
40
41  def bernoulli_epsilongreedy(iteration, epsilon,  proba_arm_1,
        proba_arm_2):
42      print("Epsilon-greedy : cas de 2 bras suivant une loi de Bernoulli
            , arm1 : %s, arm2 : %s" % (proba_arm_1, proba_arm_2))
43
44      algo = EpsilonGreedy(epsilon,[],[], 2)
45      i = 0
46      vector_arms_chosen = []
47      reward_vect = []
48      reward_cum = 0
49      while i < iteration :
50          chosen_arm = algo.select_arm()
51          vector_arms_chosen.append(chosen_arm)
52          if chosen_arm == 0 :
53              reward =  simulation_arm_bernoulli(proba_arm_1)
54          else :
55              reward = simulation_arm_bernoulli(proba_arm_2)
56          algo.update(chosen_arm, reward)
57          reward_cum = reward_cum + reward
58          if i == 0:
59              reward_vect.append(0)
60          else:
61              reward_vect.append(reward_cum / float(i))
62
63          i += 1
64      return reward_vect, algo.values, algo.counts, vector_arms_chosen
65
66  #Parametres
67  nb_iteration = 100
68  arm_1_probability = 0.7
69  arm_2_probability = 0.3
```

```
epsilon = 0.5

#Simulation
rewards_vect, values, count, arms_chosen = bernoulli_epsilongreedy(
    nb_iteration, epsilon, arm_1_probability, arm_2_probability)
offline_solution = [max(arm_1_probability, arm_2_probability) for x in
     range(nb_iteration)]

#Affichage du nombre de tirages de chaque bras
cum_sum_arm_2 = numpy.cumsum(arms_chosen)
cum_sum_arm_1 = [ x - cum_sum_arm_2[x] for x in range(nb_iteration)]
plt.plot(range(nb_iteration),cum_sum_arm_1, range(nb_iteration),
    cum_sum_arm_2);
plt.title("Nombre de tirages de chaque bras");
plt.legend(["arm1 Proba %s" % arm_1_probability, "arm2 p= %s" %
    arm_2_probability], loc="upper left");
plt.figure();

#Affichage du gain moyen par tour
plt.plot(range(nb_iteration), offline_solution, 'k-', range(
    nb_iteration), rewards_vect, 'r--');
plt.axis([0,nb_iteration,0,1]);
plt.title("Gain moyen par tour");
plt.legend(["Offline",r" $\epsilon$ = %s" % epsilon], loc="lower right
    ");
plt.figure();
```

Listing 3 – Code Python de la simulation UCB

```
##UCB1
import matplotlib.pyplot as plt
import math
import random
import numpy

#Definition de la classe UCB1
class UCB1():
    def __init__(self, counts = [], values = [], n_arms = 0):
        self.n_arms = n_arms
        self.counts = [0 for col in range(n_arms)]
        self.values = [0.0 for col in range(n_arms)]

    def select_arm(self):
        """ Selectionne le bras avec la valeur de l'estimateur la plus
            haute"""
        for arm in range(self.n_arms):
            if self.counts[arm] == 0:
```

```
18                  return arm                          #on explore chaque
                        bras une premi re fois
19
20          ucb_values = [0.0 for arm in range(self.n_arms)]
21          total_counts = sum(self.counts)
22          for arm in range(self.n_arms):
23              bonus = math.sqrt((2 * math.log(total_counts)) / float(
                    self.counts[arm]))
24              ucb_values[arm] = self.values[arm] + bonus
25
26          value_max = max(ucb_values)
27          return ucb_values.index(value_max)
28
29      def update(self, chosen_arm, reward):
30          self.counts[chosen_arm] += 1
31          n = self.counts[chosen_arm]
32          value = self.values[chosen_arm]
33          new_value = ((n - 1) /float(n)) * value + (reward / float(n))
34          self.values[chosen_arm] =  new_value
35
36  ##Partie simulation
37  #Simulation avec deux bras de loi de Bernoulli
38  def bernoulli_UCB1(iteration, proba_arm_1, proba_arm_2):
39      print("UCB1 : cas 2 bras, suivant une loi de bernoulli, arm1 : %s,
            arm2 : %s" % (proba_arm_1, proba_arm_2))
40      vector_arms_chosen = []
41      reward_vect = []
42      reward_cum = 0
43      algo = UCB1([],[], 2)
44      i = 0
45      while i < iteration :
46          chosen_arm = algo.select_arm()
47          vector_arms_chosen.append(chosen_arm)
48          if chosen_arm == 0 :                          #si le bras 1 est
                choisi
49              if random.random() < proba_arm_1 :
50                  reward = 1
51              else :
52                  reward = 0
53          else :                                         #si le bras 2 est
                choisi
54              if random.random() < proba_arm_2 :
55                  reward = 1
56              else :
57                  reward = 0
58          algo.update(chosen_arm, reward)
59          reward_cum = reward_cum + reward
```

```
60          i += 1
61          reward_vect.append(reward_cum / float(i))
62      return reward_vect, algo.values, algo.counts, vector_arms_chosen
63
64  #Parametres
65  nb_iteration = 100
66  arm_1_probability = 0.7
67  arm_2_probability = 0.3
68
69  #Simulation
70  rewards_vect, values, count, arms_chosen = bernoulli_UCB1(nb_iteration
        , arm_1_probability, arm_2_probability)
71
72  offline_solution = [max(arm_1_probability, arm_2_probability) for x in
         range(nb_iteration)]
73
74  cum_sum_arm_2 = numpy.cumsum(arms_chosen) #arm_chosen = 1 if arm2
        selected, else arm_chosen = 0
75  cum_sum_arm_1 = [x - cum_sum_arm_2[x] for x in range(nb_iteration)]
76  plt.plot(range(nb_iteration),cum_sum_arm_1, range(nb_iteration),
        cum_sum_arm_2);
77  plt.title("Nombre de tirages de chaque bras");
78  plt.legend(["arm1 Proba %s" % arm_1_probability, "arm2 Proba %s" %
        arm_2_probability], loc="upper left");
79  plt.figure();
80
81  #Affichage
82  plt.plot(range(nb_iteration), offline_solution, 'k-', range(
        nb_iteration), rewards_vect, 'r--');
83  plt.title("Gain moyen par tour");
84  plt.legend(["Offline","UCB1"], loc="lower right");
```

Listing 4 – Code Python de la comparaison des performances

```
1
2   #Parametres
3   nb_iteration = 1000
4   arm_1_probability = 0.4
5   arm_2_probability = 0.3
6   epsilon = 0.1
7
8   offline_solution = [max(arm_1_probability, arm_2_probability) for x in
         range(nb_iteration)]
9
10  #SimulationETC
11  rewards_vect_ETC, values_ETC, count_ETC, arms_chosen_ETC =
        bernoulli_ETC(nb_iteration, arm_1_probability, arm_2_probability)
```

```
12
13 #Simulation EPSILON-GREEDY
14 rewards_vect_EPS, values_EPS, count_EPS, arms_chosen_EPS =
       bernoulli_epsilongreedy(nb_iteration, epsilon, arm_1_probability,
       arm_2_probability)
15
16 #Simulation UCB
17 rewards_vect_UCB, values_UCB, count_UCB, arms_chosen_UCB =
       bernoulli_UCB1(nb_iteration, arm_1_probability, arm_2_probability)
18
19 #Creation de la figure
20 fig, ax = plt.subplots()
21
22 #Affichage de la meilleure solution en moyenne
23 ax.plot(range(nb_iteration), offline_solution, 'k-', label='Offline␣
       Solution')
24
25 # Affichage des r sultats pour ETC
26 ax.plot(range(nb_iteration), rewards_vect_ETC, label='ETC')
27
28 # Affichage des resultats pour Epsilon-Greedy
29 ax.plot(range(nb_iteration), rewards_vect_EPS, label='Epsilon-Greedy')
30
31 #Affichage des resultats pour UCB
32 ax.plot(range(nb_iteration), rewards_vect_UCB, label='UCB')
33
34 # Definition des limites des axes
35 ax.axis([0, nb_iteration, 0.25, 0.45])
36 ax.set_ylim(-0.1, 1.2)
37
38 # Ajout du titre et des legendes
39 ax.set_title("Gain␣moyen␣par␣tour")
40 ax.legend()
41
42 # Affichage du graphique
43 plt.show()
```