

Répondez dans un rapport PDF. Remettez une archive contenant le rapport PDF ainsi que le code Python associé aux différentes questions. Votre rapport doit indiquer clairement quel fichier Python doit être exécuté pour répliquer les résultats présentés dans votre rapport.

Remise : Mercredi 7 octobre à 17h00

### 1. Algorithmes de base

- (a) ( $\frac{1}{2}$  point) Par des expériences empiriques, trouvez une instance de bandits dans la classe des two-armed bandits à distribution Bernoulli pour laquelle un algorithme Explore-Then-Commit (ETC) de phase exploratoire fixe à  $m = 10$  échoue (obtient un pseudo-regret cumulatif linéaire dans le temps) avec une probabilité d'au moins 20%. Présentez des résultats supportant votre réponse et expliquez pourquoi ça se produit.
- (b) ( $\frac{1}{2}$  point) Comment pouvez-vous rendre  $\epsilon$ -greedy optimiste à l'aide de l'initialisation des estimateurs empiriques  $\hat{\mu}_k$ ? Par des expériences sur des bandits à distribution Normale, montrez le potentiel de cette variante en la comparant avec  $\epsilon$ -greedy standard. Justifiez vos choix de design expérimental et discutez des résultats obtenus.

### 2. Algorithme kl-UCB

- (a) (1 point) Donnez le pseudo-code de l'algorithme kl-UCB pour des bandits à distribution Bernoulli et implémentez cet algorithme. Discutez des détails d'implémentation pertinents.
- (b) (1 point) Réalisez des expériences pour comparer la performance de cet algorithme avec un kl-UCB utilisant la divergence de Kullback-Leibler de distribution Gaussienne sur des bandits à distribution Bernoulli. Justifiez vos choix de design expérimental et discutez des résultats obtenus.

### 3. Algorithme Thompson Sampling (TS)

- (a) (1 point) Donnez le pseudo-code d'un algorithme TS pour des bandits à distribution Poisson,  $X \sim \text{Pois}(\lambda)$ , et implémentez cet algorithme.
- (b) (2 points) Réalisez des expériences et présentez des résultats montrant que votre algorithme fonctionne. Justifiez vos choix de design expérimental et discutez des résultats obtenus. Discutez notamment de l'impact des priors dans ce modèle.

### 4. Bandits contextuels et structurés

- (a) (1 point) Implémentez l'algorithme Kernel Thompson Sampling (Algorithme 1 dans [1]) et donnez le pseudo-code de la procédure pour l'appliquer à un problème de bandits contextuels.
- (b) (2 points) Réalisez des expériences et présentez des résultats montrant que votre algorithme fonctionne. Pour ce faire, considérez un two-armed bandit contextuel de bruit Gaussien et considérez l'espace de contextes  $\mathcal{S} = [0, 1]$  avec une arrivée des contextes suivant une distribution uniforme. Justifiez vos choix de design expérimental et discutez des résultats obtenus.

## References

- [1] Durand, A., Maillard, O. A., & Pineau, J. (2018). *Streaming kernel regression with provably adaptive mean, variance, and regularization*. The Journal of Machine Learning Research, 19(1), 650-683. <https://www.jmlr.org/papers/volume19/17-404/17-404.pdf>