Université de Bordeaux **Mathématiques**
Préparation à l'agrégation **Année 2019/2020**

TP : Martingales 2 - Le Bandit à Deux Bras

On considère une machine à sous à deux leviers A et B.

Pour le levier $L \in \{A, B\}$, le gain est de 1€ avec probabilité θ^L, et 0€ avec probabilité $1 - \theta^L$.

On suppose que $0 < \theta^A, \theta^B < 1$ avec $\theta^A \neq \theta^B$.

On introduit deux suites indépendantes $(E_n^A), (E_n^B)$ de v.a. i.i.d. de Bernoulli de paramètres θ^A et θ^B respectivement. On introduira encore une suite annexe (U_n) de v.a. i.i.d. uniformes sur $[0, 1]$ et qui est aussi indépendante des suites (E_n^A), (E_n^B).

À l'étape n, le joueur actionne alors le levier L_n et obtient le gain $X_n = E_n^{L_n}$.

On travaillera avec la filtration

$$\mathcal{F}_n = \sigma(U_0, U_1, X_1, \ldots, U_n, X_n) .$$

À l'étape n il choisit le levier $L_n \in \{A, B\}$ au vu des gains antérieurs $X_1, \cdots, X_{n-1}$, en rajoutant éventuellement de l'aléa (avec les variables uniformes). Autrement dit, on suppose que (L_n) est une suite prévisible pour la filtration $(\mathcal{F}_n)$.

Après l'étape n, le gain moyen s'écrit

$$G_n = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

Le joueur va chercher à optimiser son gain moyen en adoptant une stratégie pour le choix des leviers. Pour $n \geq 1$, on note N_n^L le nombre de fois où le joueur a choisi le levier L jusqu'à l'étape n. On pose aussi

$$M_n = \sum_{k=1}^{n} X_k - \theta^A N_n^A - \theta^B N_n^B .$$

Préliminaires

1. Quelle est la meilleure stratégie si le joueur connaît θ^A et θ^B ?
Vers quoi converge alors le gain moyen ?

2. Montrer que M_n est une martingale de carré intégrable et de processus croissant

$$\langle M \rangle_n = \theta^A(1 - \theta^A) N_n^A + \theta^B(1 - \theta^B) N_n^B .$$

3. Montrer que $M_n = o(n)$ p.s. En déduire que presque sûrement on a

$$\min(\theta^A, \theta^B) \leq \liminf G_n \leq \limsup G_n \leq \max(\theta^A, \theta^B).$$

On dira d'une stratégie qu'elle est **bonne** si presque sûrement

$$G_n \longrightarrow \max(\theta^A, \theta^B) .$$

On va estimer les probabilités inconnues θ^A et θ^B par

$$\widehat{\theta_n^L} = \begin{cases} \dfrac{1}{N_n^L} \displaystyle\sum_{k=1}^{n} \mathbf{1}_{\{L_k = L, X_k = 1\}} & \text{si } N_n^L \geq 1 \\ 0 \text{ sinon} \end{cases} .$$

Stratégie naïve

Une stratégie naturelle consiste à choisir, pour tout $n \geq 0$,

$$L_{n+1} = \begin{cases} A & \text{si } \widehat{\theta_n^A} > \widehat{\theta_n^B} \text{ ou } (\widehat{\theta_n^A} = \widehat{\theta_n^B} \text{ et } U_n < \frac{1}{2}) \\ B & \text{si } \widehat{\theta_n^A} < \widehat{\theta_n^B} \text{ ou } (\widehat{\theta_n^A} = \widehat{\theta_n^B} \text{ et } U_n \geqslant \frac{1}{2}) \end{cases} .$$

4. Implémenter la stratégie naturelle. Tracer $n \mapsto G_n$, $n \mapsto \widehat{\theta_n^A}$, et $n \mapsto \widehat{\theta_n^B}$.

5. Vérifier numériquement que cette stratégie n'est pas bonne.
Estimer numériquement la probabilité d'échec.

6. Montrer que cette stratégie n'est pas bonne. Quel est le problème ?

Une bonne stratégie

7. Trouver une martingale M_n^L de carré intégrable, qui vérifie $\widehat{\theta_n^L} - \theta^L = \dfrac{M_n^L}{N_n^L}$.

Vérifier que son processus croissant est $\langle M^L \rangle_n = \theta^L(1-\theta^L)N_n^L$.

8. En déduire que si $N_n^L \longrightarrow +\infty$ p.s., alors $\widehat{\theta_n^L} \longrightarrow \theta^L$ p.s. .

Pour s'assurer que N_n^A et N_n^B tendent vers l'infini, on fixe deux ensembles infinis d'indice Γ^A, Γ^B disjoints pour lesquels on va jouer respectivement A et B. On pose donc $\Gamma = \Gamma^A \cup \Gamma^B$ puis

$$L_{n+1} = \begin{cases} A & \text{si } n \in \Gamma^A \text{ ou } (n \notin \Gamma \text{ et } \widehat{\theta_n^A} \geqslant \widehat{\theta_n^B}) \\ B & \text{si } n \in \Gamma^B \text{ ou } (n \notin \Gamma \text{ et } \widehat{\theta_n^A} < \widehat{\theta_n^B}) \end{cases} .$$

On choisira Γ de telle sorte que sa densité tende vers 0 :

$$\frac{1}{n}|\Gamma \cup \{1, \ldots, n\}| \xrightarrow[n\to\infty]{} 0 .$$

9. Montrer qu'avec cette nouvelle stratégie, $G_n \longrightarrow \max(\theta^A, \theta^B)$ p.s.

10. Implémenter cette nouvelle stratégie. Représenter G_n ainsi que les estimateurs $\widehat{\theta_n^A}$ et $\widehat{\theta_n^B}$.

11. Supposons $|\Gamma \cup \{1, \ldots, n\}| = o(\sqrt{n})$. En écrivant

$$\sqrt{n}(G_n - \theta^A) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}(E_k^A - \theta^A) - \frac{1}{\sqrt{n}}\sum_{k=1}^{n}(E_k^A - E_k^B)\mathbf{1}_{U_k=B} ,$$

montrer qu'on a la convergence en loi

$$\sqrt{n}(G_n - \max(\theta^A, \theta^B)) \longrightarrow \mathcal{N}(0, \sigma^2) ,$$

où $\sigma^2 = \max(\theta^A, \theta^B)(1 - \max(\theta^A, \theta^B))$.

12. Illustrer cette convergence en loi.