ENS Cachan 2017-2018 **Préparation à l'Agrégation**
Arthur Leclaire **Option Probabilités et Statistiques**

Martingales - TP
Bandit à deux bras

Rappels et compléments de cours [1, 6]

Soit (M_n) une martingale de carré intégrable adaptée à une filtration $(\mathcal{F}_n)$.

Définition 1. *Le processus croissant $(\langle M\rangle_n)$ associé à (M_n) est défini par $\langle M\rangle_0 = 0$ et*

$$\forall n \in \mathbb{N}, \quad \langle M\rangle_{n+1} - \langle M\rangle_n = \mathbb{E}[(M_{n+1} - M_n)^2 | \mathcal{F}_n].$$

Le processus croissant est intégrable, positif et croissant, donc converge vers $\langle M\rangle_\infty$.

Théorème 2 (Loi des grand nombres pour les martingales)**.**

1. *Sur $\{\langle M\rangle_\infty < \infty\}$, M_n converge p.s. vers M_∞ de carré intégrable.*
2. *Supposons que $\langle M\rangle_\infty = \infty$ presque sûrement. Alors $\dfrac{M_n}{\langle M\rangle_n} \longrightarrow 0$ p.s.*

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

Pour le levier L, 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 variables i.i.d. de Bernoulli de paramètres θ^A et θ^B respectivement. À l'étape n le joueur choisit le levier $U_n \in \{A, B\}$ au vu des gains antérieurs $X_1, \cdots, X_{n-1}$. Il l'actionne et obtient le gain $X_n = E_n^{U_n}$.

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 adéquate pour le choix des leviers.

Pour $n \geqslant 1$, 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 $M_n = o(n)$. En déduire que presque sûrement on a

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

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

$$\widehat{\theta_n^L} = \frac{1}{N_n^L} \sum_{k=1}^{n} \mathbb{1}_{\{U_k=L, X_k=1\}} \quad \text{si } N_n^L \geqslant 1, \quad \text{et 0 sinon .}$$

Stratégie naïve

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

$$U_{n+1} = \begin{cases} A & \text{si } \widehat{\theta_n^A} > \widehat{\theta_n^B} \\ B & \text{si } \widehat{\theta_n^A} < \widehat{\theta_n^B} \\ A \text{ ou } B \text{ avec probabilité } \frac{1}{2} & \text{si } \widehat{\theta_n^A} = \widehat{\theta_n^B} \end{cases} .$$

4. Implémenter la stratégie naturelle. Tracer $n \mapsto G_n$, $n \mapsto \widehat{\theta_n^A}$, et $n \mapsto \widehat{\theta_n^B}$.
   Vérifier numériquement que cette stratégie n'est pas bonne.
   Estimer numériquement la probabilité d'échec.
5. Montrer que cette stratégie n'est pas bonne.
   Quel est le problème ? Calculer la probabilité d'échec.

Une bonne stratégie

6. 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 $\langle M^L \rangle_n = \theta^L(1-\theta^L)N_n^L$.
7. En déduire que si $N_n^L \to +\infty$ p.s., alors $\widehat{\theta_n^L} \to \theta^L$ p.s. .

8. En déduire une bonne stratégie et l'implémenter.
9. Représenter G_n ainsi que les estimateurs $\widehat{\theta_n^A}$ et $\widehat{\theta_n^B}$.
10. Affiner la stratégie pour avoir la convergence en loi

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

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

11. Illustrer cette convergence en loi.

Loi des grands nombres

Dans cette partie on va faire la preuve du Théorème 2 ci-dessus. Le point 1 a déjà été vu en cours. On se concentre donc sur le point 2. On pose

$$Y_0 = 0 \ , \quad Y_n = M_n - M_{n-1} \ , \quad V_n = \langle M \rangle_n \ .$$

Notons $\tau = \inf\{n \geqslant 0 \mid V_n > 1\}$.

12. Montrer que $\tau < \infty$ p.s. et que pour $t \geqslant 1$, $\{\tau \leqslant t\} \in \mathcal{F}_{t-1}$.
13. Pour $t \geqslant 1$, on pose $Y'_t = \dfrac{Y_t}{V_t} \ \mathbb{1}_{\tau \leqslant t}$ et $M'_t = Y'_1 + \ldots + Y'_t$.
    Montrer que (M'_t) est une martingale.
14. Montrer que le processus croissant V' associé à la martingale M' est donné par

$$V'_n = \sum_{t=1}^{n} \frac{V_t - V_{t-1}}{V_t^2} \ \mathbb{1}_{\tau \leqslant t} \ .$$

15. Montrer que sur $\{\tau < \infty\}$ on a pour tout n, $V'_n \leqslant \dfrac{2}{V_\tau}$.
    En déduire que $M'_n = \displaystyle\sum_{t=1}^{n} \frac{Y_t}{V_t} \mathbb{1}_{\tau \leqslant t}$ converge presque sûrement.
16. Montrer que presque sûrement,

$$\frac{1}{V_n} \sum_{t=\tau}^{n} Y_t \xrightarrow[n \to \infty]{} 0$$

et en déduire $M_n = o(V_n)$ presque sûrement. On pourra utiliser le lemme de Kronecker : si $\sum u_n$ converge et si $b_n > 0$ tend vers $+\infty$, alors

$$\frac{1}{b_n} \sum_{k=1}^{n} b_k u_k \longrightarrow 0 \ .$$

Optimisation stochastique, Stratégie de Narendra [2, 4]

La stratégie de Narendra est la suivante : à l'étape $n+1$, on choisit le levier A avec probabilité p_{n+1}, où p_{n+1} est calculé récursivement à l'aide des expériences antérieures. La règle choisie est la suivante

$$p_{n+1} = \begin{cases} p_n + \gamma_n(1-p_n) & \text{si } U_n = A \text{ et } X_n = 1 \\ p_n - \gamma_n p_n & \text{si } U_n = B \text{ et } X_n = 1 \\ p_n & \text{si } X_n = 0 \end{cases} .$$

On choisit $\gamma_n = (\frac{C}{n+C})^\alpha$ pour $\alpha \in (0,1]$ et $C > 0$.

17. Vérifier que $p_{n+1} = p_n + \gamma_n f(U_n, X_n, p_n)$ où

$$f(u,x,p) = \big((1-p)\mathbb{1}_{u=A} - p\mathbb{1}_{u=B}\big)\mathbb{1}_{x=1} .$$

18. Montrer que

$$\forall n \in \mathbb{N}, \quad \mathbb{E}\big[f(U_n, X_n, p_n)|p_n\big] = (\theta^A - \theta^B)p_n(1-p_n) .$$

19. En déduire une écriture de (p_n) sous la forme

$$p_{n+1} = p_n + \gamma_n(h(p_n) + \epsilon_{n+1})$$

où ϵ_{n+1} est intégrable et $\mathbb{E}(\epsilon_{n+1}|p_n) = 0$.

20. Implémenter la stratégie de Narendra.
21. Vérifier numériquement que c'est une bonne stratégie lorsque $\alpha = 1$ et $c \leqslant \frac{1}{\theta^B}$.
22. Vérifier numériquement que ce n'est pas une bonne stratégie lorsque $\alpha < 1$ ou lorsque $\alpha = 1$ et $c > \frac{1}{\theta^B}$. (On prendra garde à lancer l'expérience plusieurs fois.)

Références

[1] Bercu et Chafaï. *Modélisation Stochastique et Simulation*. Sciences Sup, 2007.

[2] Duflo. *Algorithmes Stochastiques*. Springer, 1997.

[3] Kaufmann. *Modèles de bandit : une histoire bayésienne et fréquentiste*. Matapli (Bulletin de la Société de Mathématiques Appliquées et Industrielles), 2016.

[4] Lamberton, Pagès et Tarrès. *When can the two-armed bandit algorithm be trusted ?* The Annals of Applied Probability, 2004.

[5] Rivoirard et Stoltz. *Statistique en Action*. Vuibert, 2009.

[6] Rivoirard et Stoltz. *Machines à sous*. Document en ligne http://www.math.ens.fr/statenaction/PDF/MachinesSous-Principal.pdf .