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

Martingales - TP
Bandit à deux bras - Corrigé

Le présent document a été fortement inspiré du corrigé "Machines à sous" annexé au livre "Statistique en action" de Rivoirard et Stoltz. En complément on recommande la lecture de cette annexe, qui est disponible sur Internet.

1. La meilleure stratégie est de jouer toujours avec le levier A si $\theta^A > \theta^B$ et toujours avec le levier B sinon. On a alors $G_n \to \max(\theta^A, \theta^B)$ p.s. par la loi des grands nombres.

2. Soit $\mathcal{F}_n$ la filtration $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Par définition de la notion de stratégie (non randomisée), la variable aléatoire U_{n+1} est $\mathcal{F}_n$-mesurable. Donc M_n est bien $\mathcal{F}_n$ mesurable, et elle est de carré intégrable car bornée :

$$|M_n| \leqslant \sum_{k=1}^{n} |X_k| + N_n^A + N_n^B \leqslant 2n \ .$$

De plus,

$$\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n - \theta^A \mathbb{1}_{U_{n+1}=A} - \theta^B \mathbb{1}_{U_{n+1}=B} + \mathbb{E}[X_{n+1}|\mathcal{F}_n] \ ,$$

car $X_1, \cdots, X_n, U_1, \cdots, U_{n+1}$ sont mesurables par rapport à $\mathcal{F}_n$. Sachant $U_{n+1} = L$, X_{n+1} est une variable de Bernoulli de paramètre θ^L. Par conséquent,

$$\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] = \theta^A \mathbb{1}_{U_{n+1}=A} + \theta^B \mathbb{1}_{U_{n+1}=B} \ .$$

On a donc bien $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$ donc M_n est une martingale de carré intégrable.

Calculons le processus croissant de M_n. On a

$$(M_{n+1} - M_n)^2 = \left(X_{n+1} - \theta^A \mathbb{1}_{U_{n+1}=A} - \theta^B \mathbb{1}_{U_{n+1}=B}\right)^2 .$$

Par suite,

$$\begin{aligned} \mathbb{E}[(M_{n+1} - M_n)^2|\mathcal{F}_n] &= \mathbb{E}\big[(X_{n+1} - \theta^A \mathbb{1}_{U_{n+1}=A} - \theta^B \mathbb{1}_{U_{n+1}=B})^2\big] \\ &= \mathbb{E}\big[(X_{n+1} - \theta^A)^2|\mathcal{F}_n\big]\mathbb{1}_{U_{n+1}=A} + \mathbb{E}\big[(X_{n+1} - \theta^B)^2|\mathcal{F}_n\big]\mathbb{1}_{U_{n+1}=B} \\ &= \theta^A(1-\theta^A)\mathbb{1}_{U_{n+1}=A} + \theta^B(1-\theta^B)\mathbb{1}_{U_{n+1}=B} \\ &\Big(= \theta^{U_{n+1}}(1-\theta^{U_{n+1}})\Big) \ . \end{aligned}$$

Noter que ce terme est exactement la variance conditionnelle de l'accroissement $M_{n+1} - M_n$ (ici la loi conditionnelle de $M_{n+1} - M_n$ conditionnellement à $\mathcal{F}_n$ est bien une Bernoulli de paramètre $\theta^{U_{n+1}}$).

Donc en sommant

$$\langle M\rangle_n = \sum_{k=1}^{n} \mathbb{E}\left[(M_k - M_{k-1})^2 | \mathcal{F}_{k-1}\right] = \theta^A(1-\theta^A)N_n^A + \theta^B(1-\theta^B)N_n^B \ .$$

3. On a

$$\min(\theta^A(1-\theta^B), \theta^B(1-\theta^B)) \cdot n \ \leqslant \ \langle M\rangle_n \ \leqslant \ N_n^A + N_n^B = n \ ,$$

d'où $\langle M\rangle_n \to +\infty$ p.s. et donc $\langle M\rangle_\infty = \infty$.

D'après la loi des grands nombres pour les martingales, on a

$$\frac{M_n}{\langle M\rangle_n} \to 0 \quad \text{p.s.} \ .$$

Ainsi,

$$\frac{M_n}{n} = \frac{M_n}{\langle M\rangle_n}\frac{\langle M\rangle_n}{n} \to 0 \text{ p.s.} \ .$$

Donc $M_n = o(n)$ presque sûrement.

On a

$$G_n = \frac{M_n}{n} + \theta^A \cdot \frac{N_n^A}{n} + \theta^B \cdot \frac{N_n^B}{n} \ ,$$

Comme $N_n^A + N_n^B = n$, on en déduit

$$\frac{M_n}{n} + \min(\theta^A, \theta^B) \leqslant G_n \leqslant \frac{M_n}{n} + \max(\theta^A, \theta^B) \ .$$

Comme $\frac{M_n}{n} \to 0$, il s'ensuit

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

4. Voir corrigé du code.

5. Supposons $\theta^A > \theta^B$. Les quatre situations possibles à l'issue de l'étape 1 sont

(a) $U_1 = A$ et $X_1 = 1$, avec probabilité $\frac{1}{2}\theta^A \to$ Réussite.

(b) $U_1 = A$ et $X_1 = 0$, avec probabilité $\frac{1}{2}(1-\theta^A) \to$ Les estimateurs sont inchangés.

(c) $U_1 = B$ et $X_1 = 1$, avec probabilité $\frac{1}{2}\theta^B \to$ Échec.

(d) $U_1 = B$ et $X_1 = 0$, avec probabilité $\frac{1}{2}(1-\theta^B) \to$ Les estimateurs sont inchangés.

Dans le cas (c), pour tout $n \geqslant 1$, $U_n = B$ car $\widehat{\theta_n^B} > 0 = \widehat{\theta_n^A}$.

Tant que $X_i = 0$ (quel que soit U_i), on a $\widehat{\theta_n^A} = \widehat{\theta_n^B} = 0$.

Dès que $X_p = 1$, on a $\widehat{\theta_p^A} > 0$ ou $\widehat{\theta_p^B} > 0$ selon le résultat du p-ième tirage. Puis pour tout $n \geqslant p$, $\widehat{\theta_n^A} > 0$ alors que $\widehat{\theta_n^B} = 0$ (ou réciproquement).

La probabilité d'échec s'exprime donc par

$$\begin{aligned}
\mathbb{P}(\text{échec}) &= \mathbb{P}\Big(\exists p \in \mathbb{N}^*, X_1 = \cdots = X_{p-1} = 0, U_p = B \text{ et } X_p = 1\Big) \\
&= \sum_{p \geqslant 1} \mathbb{P}(X_1 = \cdots = X_{p-1} = 0, U_p = B \text{ et } X_p = 1) \\
&= \sum_{p \geqslant 1} \mathbb{P}(U_p = B \text{ et } X_p = 1 | X_1 = \cdots = X_{p-1} = 0) \mathbb{P}(X_1 = \cdots = X_{p-1} = 0) \\
&= \frac{\theta^B}{2} \sum_{p \geqslant 1} \mathbb{P}(X_1 = \cdots = X_{p-1} = 0) \\
&= \frac{\theta^B}{2} \sum_{p \geqslant 1} \left(\frac{1-\theta^A}{2} + \frac{1-\theta^B}{2} \right)^{p-1} \\
&= \frac{\theta^B}{\theta^A + \theta^B} \ .
\end{aligned}$$

Le problème vient du fait qu'on ne joue pas avec chaque levier une infinité de fois.

6. Posons

$$M_n^L = \sum_{k=1}^n X_k \mathbb{1}_{\{U_k = L\}} - \theta^L N_n^L \ .$$

On a bien $\widehat{\theta_n^L} - \theta^L = \frac{M_n^L}{N_n^L}$. M_n^L est $\mathcal{F}_n$-mesurable et de carré intégrable car

$$|M_n^L| \leqslant \sum_{k=1}^n |X_k| + \theta^L N_n^L \leqslant n + 1 \times N_n^L \leqslant 2n \ .$$

De plus,

$$\begin{aligned}
\mathbb{E}[M_{n+1}^L | \mathcal{F}_n] &= M_n^L + \mathbb{1}_{U_{n+1} = L} \mathbb{E}\left[X_{n+1} | \mathcal{F}_n\right] - \theta^L \mathbb{1}_{U_{n+1} = L} \\
&= M_n^L + \theta^L \mathbb{1}_{U_{n+1} = L} - \theta^L \mathbb{1}_{U_{n+1} = L} = M_n^L \ ,
\end{aligned}$$

car $X_1, \ldots, X_n, U_1, \ldots, U_{n+1}$ sont $\mathcal{F}_n$-mesurables. M_n^L est donc une martingale.

Et on peut calculer son processus croissant car

$$\begin{aligned}
\mathbb{E}\left[(M_{n+1}^L - M_n^L)^2 | \mathcal{F}_n\right] &= \mathbb{E}\left[(X_{n+1} \mathbb{1}_{U_{n+1} = L} - \theta^L \mathbb{1}_{U_{n+1} = L})^2 | \mathcal{F}_n\right] \\
&= \mathbb{E}\left[(X_{n+1} - \theta^L)^2 | \mathcal{F}_n\right] \mathbb{1}_{U_{n+1} = L} \\
&= \theta^L (1 - \theta^L) \mathbb{1}_{U_{n+1} = L} \ .
\end{aligned}$$

Donc en sommant

$$\langle M^L \rangle_n = \theta^L (1 - \theta^L) N_n^L.$$

7. Comme $N_n^L \to +\infty$ p.s., on a

$$\langle M^L \rangle_n = \theta^L (1 - \theta^L) N_n^L \to +\infty \quad \text{p.s.} \ ,$$

et donc $\langle M\rangle_\infty = \infty$ p.s. D'après la loi des grands nombres pour les martingales,

$$\frac{M_n^L}{\langle M^L\rangle_n} \to 0 \quad \text{p.s. .}$$

Or

$$\frac{M_n^L}{\langle M^L\rangle_n} = \frac{M_n^L}{\theta^L(1-\theta^L)N_n^L} = \frac{\widehat{\theta_n^L} - \theta^L}{\theta^L(1-\theta^L)}.$$

On a donc $\widehat{\theta_n^L} \to \theta^L$ p.s.

8. On va modifier la stratégie précédente de manière à s'assurer que l'on joue une infinité de fois chaque levier. Pour cela, on fixe deux ensembles infinis d'indices Γ^A, Γ^B disjoints pour lesquels on va jouer respectivement A et B. On pose donc $\Gamma = \Gamma^A \cup \Gamma^B$ puis

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

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

$$\frac{1}{n}\,\mathrm{Card}(\Gamma \cap \{1,\ldots,n\}) \xrightarrow[n\to\infty]{} 0 \ .$$

On va montrer qu'alors la stratégie définie ci-dessus est optimale.

À partir de maintenant, on supposera par exemple $\theta^A > \theta^B$. Par la question précédente, on a presque sûrement

$$\widehat{\theta_n^A} \to \theta^A \quad \text{et} \quad \widehat{\theta_n^B} \to \theta^B \ .$$

En particulier, p.s. ω, il existe un rang $n_0(\omega)$ tel que

$$n \geqslant n_0(\omega) \quad \Longrightarrow \quad \widehat{\theta_n^A} \geqslant \widehat{\theta_n^B} \quad ,$$

et donc pour $n \geqslant n_0(\omega)$ et $n \notin \Gamma$, on a $U_{n+1} = A$.

Écrivons alors

$$G_n - \theta^A = \frac{N_n^A}{n}(\widehat{\theta_n^A} - \theta^A) + \frac{N_n^B}{n}(\widehat{\theta_n^B} - \theta^A) \ .$$

Pour $n \geqslant n_0(\omega)$, on a

$$n \geqslant N_n^A \geqslant n - n_0(\omega) - \mathrm{Card}(\Gamma^B \cap \{n_0(\omega),\ldots,n\}) \ ,$$

et

$$N_n^B \leqslant \mathrm{Card}(\Gamma^B \cap \{1,\ldots,n\}) + n_0(\omega) \ . \tag{1}$$

Puisque $\mathrm{Card}(\Gamma^L \cap \{1,\ldots,n\}) = o(n)$, on en déduit que presque sûrement,

$$\frac{N_n^A}{n} \longrightarrow 1 \quad \text{et} \quad \frac{N_n^B}{n} \longrightarrow 0 \quad \text{donc} \quad G_n \longrightarrow \theta^A = \max(\theta^A,\theta^B) \quad \text{p.s. .}$$

9. Voir la correction du code.

10. On va supposer maintenant que

$$\mathrm{Card}(\Gamma \cap \{1, \ldots, n\}) = o(\sqrt{n}).$$

En considérant uniquement les excitations données par le bras A, le théorème central-limite usuel donne

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} (E_k^A - \theta^A) \xrightarrow{\text{loi}} \mathcal{N}\big(0, \theta^A(1-\theta^A)\big).$$

Or

$$\begin{aligned}
\sqrt{n}(G_n - \theta^A) &= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - \theta^A) \\
&= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \Big(E_k^A - (E_k^A - E_k^B) \mathbb{1}_{U_k = B} - \theta^A \Big) \\
&= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (E_k^A - \theta^A) - \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (E_k^A - E_k^B) \mathbb{1}_{U_k = B}.
\end{aligned}$$

Remarquons qu'en valeur absolue, le deuxième terme est $\leqslant \frac{1}{\sqrt{n}} \mathrm{Card}(\Gamma^B \cap \{1, \ldots, n\})$ et donc tend vers zéro p.s. grâce aux choix de Γ^L. Le lemme de Slutsky permet de conclure que

$$\sqrt{n}(G_n - \theta^A) \xrightarrow{\text{loi}} \mathcal{N}\big(0, \theta^A(1-\theta^A)\big).$$

11. Voir la correction du code.

12. On a $\tau < \infty$ p.s. car $V_n = \langle M \rangle_n \to +\infty$ p.s. De plus

$$\langle \tau \leqslant t \rangle = \bigcup_{n \leqslant t} \{V_n > 1\} \in \mathcal{F}_{t-1}$$

car (V_n) est prévisible.

13. Comme on a $V_t > 1$ sur l'événement $\{\tau \leqslant t\}$, on a par définition de Y_t'

$$\mathbb{E}[|Y_t'|^2] \leqslant \mathbb{E}[|Y_t|^2]$$

donc Y_t' est de carré intégrable. De plus, M_t' est $\mathcal{F}_t$ mesurable et comme V_n est prévisible, on a

$$\mathbb{E}[M_t' - M_{t-1}' | \mathcal{F}_{t-1}] = \mathbb{E}[Y_t' | \mathcal{F}_{t-1}] = \mathbb{E}\left[\frac{Y_t}{V_t} \mathbb{1}_{\tau \leqslant t} \right] = \frac{\mathbb{1}_{\tau \leqslant t}}{V_t} \mathbb{E}[Y_t | \mathcal{F}_{t-1}] = 0 \ .$$

Ainsi (M_t') est une martingale de carré intégrable.

14. De même, comme (V_n) est prévisible, on a

$$\mathbb{E}\big[(M'_t - M'_{t-1})^2|\mathcal{F}_{t-1}\big] = \frac{\mathbb{1}_{\tau\leqslant t}}{V_t^2}\mathbb{E}\big[Y_t^2|\mathcal{F}_{t-1}\big] = \frac{\mathbb{1}_{\tau\leqslant t}}{V_t^2}\mathbb{E}\big[(M_t - M_{t-1})^2|\mathcal{F}_{t-1}\big] = \frac{\mathbb{1}_{\tau\leqslant t}}{V_t^2}(V_t - V_{t-1})$$

donc

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

15. Sur $\{\tau < \infty\}$,

$$\begin{aligned} V'_n &= \frac{1}{V_\tau} + \sum_{t=\tau+1}^{n} \frac{V_t - V_{t-1}}{V_t^2} = \frac{1}{V_\tau} + \sum_{t=\tau+1}^{n} \int_{V_{t-1}}^{V_t} \frac{ds}{V_t^2} \\ &\leqslant \frac{1}{V_\tau} + \sum_{t=\tau+1}^{n} \int_{V_{t-1}}^{V_t} \frac{ds}{s^2} \\ &\leqslant \frac{1}{V_\tau} + \int_{V_\tau}^{+\infty} \frac{ds}{s^2} = \frac{2}{V_\tau} \ . \end{aligned}$$

Ainsi, (V'_n) est p.s. bornée et donc $\langle M' \rangle_\infty < \infty$ p.s. donc la première partie de la loi des grands nombres pour les martingales donne que $M'_n \to M'_\infty$ p.s.

16. On utilise le lemme de Kronecker avec $b_n = V_{n+\tau-1}$, $u_k = \frac{Y_{k+\tau-1}}{V_{k+\tau-1}}$ de sorte que le résultat de la question précédente s'exprime par $\sum u_k$ converge et on obtient donc

$$\frac{1}{V_n}\sum_{t=\tau}^{n} Y_t = \frac{1}{b_{n-\tau+1}} \sum_{k=1}^{n-\tau+1} b_k u_k \longrightarrow 0 \quad \text{p.s.}$$

Comme $\tau < \infty$ p.s. et que $V_n \to \infty$ p.s., cela donne aussi

$$\frac{1}{V_n}\sum_{t=1}^{n} Y_t \longrightarrow 0 \quad \text{p.s.} \qquad \text{i.e.} \qquad \frac{M_n}{V_n} \longrightarrow 0 \quad \text{p.s.}$$

17. C'est une simple réécriture de la règle de mise à jour de p_n.

NB : on notera bien ici que la stratégie de Narendra n'est pas une stratégie déterministe au sens où le choix U_{n+1} du levier à l'étape $n+1$ n'est pas une fonction mesurable des résultats précédents. Autrement dit U_{n+1} n'est même pas mesurable par rapport à $\sigma(U_1, X_1, \ldots, U_n, X_n)$. En toute rigueur, on introduirait une suite annexe Z_n (indépendante des U_k et E_k^L) de variables aléatoires uniformes sur $[0,1]$, et on pose

$$U_{n+1} = \begin{cases} A & \text{si} \quad Z_n < p_{n+1} \\ B & \text{sinon} \end{cases}$$

De la sorte, U_{n+1} est mesurable par rapport à $\sigma(U_1, X_1, Z_1, \ldots, U_n, X_n, Z_n)$.

18. On notera bien aussi que p_n est aussi une variable aléatoire, prévisible par rapport à la filtration précédente. Vu comment U_n et X_n sont tirés sachant p_n, on obtient

$$\begin{aligned}\mathbb{E}[f(U_n, X_n, p_n)|p_n] &= (1-p_n)\mathbb{E}[\mathbb{1}_{U_n=A, X_n=1}|p_n] - p_n\mathbb{E}[\mathbb{1}_{U_n=B, X_=1}|p_n] \\ &= (1-p_n)p_n\theta^a - p_n(1-p_n)\theta^B = (\theta^A - \theta^B)p_n(1-p_n)\ .\end{aligned}$$

19. Il suffit de poser

$$\begin{aligned}h(p_n) &= (\theta^A - \theta^B)p_n(1-p_n) \\ \epsilon_{n+1} &= f(U_n, X_n, p_n) - (\theta^A - \theta^B)p_n(1-p_n)\ .\end{aligned}$$

20. **21.** **22.** Voir la correction du code. La preuve de convergence de cette stratégie

est donnée dans l'article de Lamberton, Pagès et Tarrès, *When can the two-armed bandit be trusted* (Then Annals of Applied Probability, 2004). On y retrouvera notamment la condition $c\theta_B > 1$.

Remarque : L'intérêt de l'écriture trouvée dans la dernière question est que l'on va étudier le comportemetn de la suite (p_n) en s'inspirant de l'étude de l'équation différentielle

$$x'(t) = h(x(t))$$

i.e.

$$x'(t) = (\theta^A - \theta^B)x(t)(1-x(t))\ .$$

On pourra voir notamment que cette équation admet 0 et 1 pour équilibres, et que quand $\theta^A > \theta^B$, l'équilibre 0 est répulsif, alors que l'équilibre 1 est attractif (d'où l'intuition que p_n convergera vers 1, c'est-à-dire l'on activera de plus en plus le bon bras). Pour plus de détails, voir l'article de Lamberton et al.