Forum scientifique et pédagogique

Forum mathématique

Envoyez vos contributions de préférence à Claude MORIN (claude.morin@ac-limoges.fr) par courrier électronique ou par courrier postal au secrétariat.

Probabilités associées aux runs

par Arnaud BÉGYN, Bruno JAFFUEL et Édouard LUCAS

Contexte et notations

On s'intéresse ici aux runs (séries, séquences, salves, jets, courses $\cdots$) dans un lancer de pile ou face. On fixe un réel $p \in]0,1[$.

On se donne un espace de probabilité $(\Omega, \mathscr{F}, \mathrm{P})$ sur lequel est définie une suite $(\mathrm{X}_n)_{n\in\mathbb{N}^*}$ de variables aléatoires mutuellement indépendantes et suivant toutes la même loi de Bernoulli $\mathscr{B}(p)$ de paramètre p. Pour simplifier les notations, on supposera que $\Omega = \{0,1\}^{\mathbb{N}^*}$ et que, pour tout $u \in \Omega$ et tout $n \in \mathbb{N}^*$, $\mathrm{X}_n(u) = u_n$. Ainsi, pour $u \in \Omega$, la suite $(\mathrm{X}_n(u))_{n\in\mathbb{N}^*}$ est égale à u.

On note P_1 la probabilité conditionnellement à $\left\{\mathrm{X}_1 = 1\right\}$ et P_0 la probabilité conditionnellement à $\left\{\mathrm{X}_1 = 0\right\}$. Sous ces probabilités, X_1 est déterministe mais les X_n pour $n \geqslant 2$ sont toujours des $\mathscr{B}(p)$ indépendantes.

- Pour $k \in \mathbb{N}^*$, X_k correspond au $k-$ième tirage.
- Pour $u \in \Omega$, on appelle run toute sous-suite constante de u qui ne peut être prolongée en une sous-suite constante.

  Pour $n \in \mathbb{N}^*$, on définit L_n la variable aléatoire qui donne la longueur du n-ième run s'il existe, et qui vaut par exemple 0 s'il n'y en a pas.
  Par exemple, pour $u = 1,1,1,0,0,1,1,0,0,0,0,0,1,\ldots$, les quatre premiers runs sont $(1,1,1)$, $(0,0)$, $(1,1)$ et $(0,0,0,0,0)$. Ainsi $\mathrm{L}_1(u) = 3$, $\mathrm{L}_2(u) = 2$, $\mathrm{L}_3(u) = 2$, $\mathrm{L}_4(u) = 5$.
- Enfin, pour $n \in \mathbb{N}^*$, on définit N_n la variable aléatoire qui compte le nombre de runs au bout de n tirages. Pour $u \in \Omega$, $\mathrm{N}_n(u)$ est le nombre de runs de la sous-suite $(u_k)_{1\leqslant k\leqslant n}$.

  On peut remarquer que $\mathrm{N}_n = 1 + \displaystyle\sum_{k=1}^{n-1} |\mathrm{X}_{k+1} - \mathrm{X}_k|$.

  En reprenant, l'exemple précédent on a : $\mathrm{N}_3(u) = 1$; $\mathrm{N}_4(u) = 2$ et $\mathrm{N}_6(u) = 3$.

Objectifs

Nous cherchons à établir les résultats suivants.

Lemme 1. Presque sûrement, il y a une infinité de runs. En dehors d'un ensemble négligeable, les L_n sont donc à valeurs dans $\mathbb{N}^*$ et la suite $(N_n)_{n\in\mathbb{N}^*}$ diverge vers $+\infty$.

Cela équivaut à dire que l'ensemble $\bigcup_{m=1}^{+\infty}\bigcap_{j=m}^{+\infty}\{X_j = X_m\}$ des suites stationnaires (les seules qui posent problème) est négligeable pour la probabilité P (et donc aussi pour P_1 et P_0).

Résultat 1. Pour n et $k \in \mathbb{N}^*$, on a

$$P(L_n = k) = \begin{cases} p(1-p)^k + (1-p)p^k & \text{si } n \text{ impair,} \\ p^2(1-p)^{k-1} + (1-p)^2 p^{k-1} & \text{si } n \text{ pair.} \end{cases}$$

Résultat 2. Si $p = \dfrac{1}{2}$ alors, pour $n \in \mathbb{N}^*$, $N_n - 1$ suit la binomiale $\mathscr{B}(n-1, \frac{1}{2})$.
Cela équivaut à dire que la fonction génératrice G_{N_n} de N_n est donnée par

$$\forall s \in \mathbb{R},\ G_{N_n}(s) = s\left(\frac{1+s}{2}\right)^{n-1}.$$

Lemme 2. Soit n et $k \in \mathbb{N}^*$ et $\alpha \in \{0,1\}$. On a l'égalité *intuitive*

$$P_\alpha(L_n \geqslant k+1) = \begin{cases} P(X_1 = \alpha)\cdot P_\alpha(L_n \geqslant k) & \text{si } n \text{ est impair ;} \\ P(X_1 = 1-\alpha)\cdot P_\alpha(L_n \geqslant k) & \text{si } n \text{ est pair.} \end{cases}$$

Les preuves

Démonstration du lemme 1.

Soit $m \in \mathbb{N}^*$. On a $\bigcap_{j=m}^{+\infty}\{X_j = X_m\} = \left(\bigcap_{j=m}^{+\infty}\{X_j = 1\}\right)\bigcup\left(\bigcap_{j=m}^{+\infty}\{X_j = 0\}\right)$.

Soit $i \in \mathbb{N}$; $\bigcap_{j=m}^{+\infty}\{X_j = 1\} \subset \bigcap_{j=m}^{m+i}\{X_j = 1\}$, donc $0 \leqslant P\left(\bigcap_{j=m}^{+\infty}\{X_j = 1\}\right) \leqslant p^{i+1}$. En passant à la limite, comme $p \in]0,1[$, on obtient $P\left(\bigcap_{j=m}^{+\infty}\{X_j = 1\}\right) = 0$.

De même, $P\left(\bigcap_{j=m}^{+\infty}\{X_j = 0\}\right) = 0$, car $1-p \in]0,1[$.

Comme $\bigcap_{j=m}^{+\infty}\{X_j = X_m\} = \left(\bigcap_{j=m}^{+\infty}\{X_j = 1\}\right)\bigcup\left(\bigcap_{j=m}^{+\infty}\{X_j = 0\}\right)$, on a $P\left(\bigcap_{j=m}^{+\infty}\{X_j = X_m\}\right) = 0$, d'où $P\left(\bigcup_{m=1}^{+\infty}\bigcap_{j=m}^{+\infty}\{X_j = X_m\}\right) = 0$ (réunion dénombrable d'ensembles négligeables), *quod erat demonstrandum.*

Démonstration du lemme 2.

Si n est impair [respectivement pair] la valeur du n-ième run est égale à la valeur α de X_1 [respectivement $1-X_1$].
Sans perte de généralité, nous supposerons que n est impair et que $\alpha = 1$. Il s'agit de prouver que $P_1(L_n \geqslant k+1) = p \cdot P_1(L_n \geqslant k)$ avec l'hypothèse n impair.

On considère la variable aléatoire $G_n = \sum_{i=1}^{n-1} L_i$ qui donne le nombre de tirages qui précèdent le n-ième run de u. Alors $\{L_n \geqslant k+1\} = \bigcup_{m\in\mathbb{N}} \left(\{G_n = m\} \cap \{L_n \geqslant k+1\}\right)$.
On remarque qu'il s'agit d'une réunion dénombrable disjointe. Par conséquent,

$$P_1(L_n \geqslant k+1) = \sum_{m=0}^{+\infty} P_1(G_n = m, L_n \geqslant k+1).$$

Soit $m \in \mathbb{N}$; $\{G_n = m\} \cap \{L_n \geqslant k+1\} = \{G_n = m\} \cap \{L_n \geqslant k\} \cap \{X_{m+k+1} = X_1\}$ car n est impair.
Or $\{G_n = m\} = \left\{ \sum_{j=0}^{m} |X_{j+1} - X_j| = n, |X_{m+1} - X_m| = 1 \right\}$, avec la convention $X_0 = 1 - X_1$, et $\{G_n = m\} \cap \{L_n \geqslant k\} = \left\{ \sum_{j=0}^{m} |X_{j+1} - X_j| = n, |X_{m+1} - X_m| = 1, \sum_{j=m+1}^{m+k-1} |X_{j+1} - X_j| = 0 \right\}$. De plus, $\{G_n = m\} \cap \{L_n \geqslant k\}$ ne dépend que de $(X_j)_{1\leqslant j\leqslant m+k}$ et est donc indépendant de X_{m+k+1}.
On en déduit que

$$P_1(G_n = m, L_n \geqslant k+1) = P(X_{m+k+1} = 1) \cdot P_1(G_n = m, L_n \geqslant k) = p \cdot P_1(G_n = m, L_n \geqslant k).$$

Après mise en facteur, on obtient $P_1(L_n \geqslant k+1) = p \cdot \sum_{m=0}^{+\infty} P_1(G_n = m, L_n \geqslant k)$, ce qui donne le résultat souhaité :

$$P_1(L_n \geqslant k+1) = p \cdot P_1(L_n \geqslant k).$$

Démonstration du résultat 1.

Plaçons nous d'abord dans le cas *n impair* et $k \in \mathbb{N}^*$.
De $P_1(L_n \geqslant 1) = 1$ et de la relation de récurrence $P_1(L_n \geqslant k+1) = p \cdot P_1(L_n \geqslant k)$, on tire $P_1(L_n \geqslant k) = p^{k-1}$ (suite géométrique). Donc

$$P_1(L_n = k) = P_1(L_n \geqslant k) - P_1(L_n \geqslant k+1) = (1-p)p^{k-1}.$$

De manière analogue, $P_0(L_n = k) = p(1-p)^{k-1}$, donc

$$P(L_n = k) = pP_1(L_n = k) + (1-p)P_0(L_n = k) = (1-p)p^k + p(1-p)^k.$$

Dans le cas *n pair*, de façon analogue, $P_1(L_n \geqslant 1) = 1$ et $P_1(L_n \geqslant k+1) = (1-p) \cdot P_1(L_n \geqslant k)$.
On en déduit $P_1(L_n \geqslant k) = (1-p)^{k-1}$ et ainsi $P_1(L_n = k) = p(1-p)^{k-1}$.
De manière analogue, $P_0(L_n = k) = (1-p)p^{k-1}$. D'où

$$P(L_n = k) = p^2(1-p)^{k-1} + (1-p)^2 p^{k-1},$$

quod erat demonstrandum.

Démonstration du résultat 2.

On rappelle qu'ici $p = \frac{1}{2}$.

Méthode 1 (d'après Arnaud Bégyn).

Soit $k, n \in \mathbb{N}^*$. Soit $s \in \mathbb{R}$.

Comme $P(N_1 = 1) = 1$, $G_{N_1}(s) = \sum_{k=0}^{\infty} P(N_1 = k)s^k = s$.

D'autre part, on dispose de la relation de récurrence

$$P(N_{n+1} = k) = \frac{1}{2}P(N_n = k) + \frac{1}{2}P(N_n = k-1).$$

Pour l'obtenir, on utilise le fait que $\{N_{n+1} = k\}$ est égal à la réunion disjointe de $\{N_n = k\} \cap \{X_n = 0\} \cap \{X_{n+1} = 0\}$, $\{N_n = k\} \cap \{X_n = 1\} \cap \{X_{n+1} = 1\}$, $\{N_n = k-1\} \cap \{X_n = 0\} \cap \{X_{n+1} = 1\}$ et $\{N_n = k-1\} \cap \{X_n = 1\} \cap \{X_{n+1} = 0\}$. Comme les variables aléatoires $X_1, \cdots, X_{n+1}$ sont mutuellement indépendantes et que N_n ne dépend que de $X_1, \cdots, X_n$, alors, pour tous a, b et c, on a $P(N_n = a, X_n = b, X_{n+1} = c) = P(N_n = a, X_n = b) \cdot P(X_{n+1} = c)$ (d'après la propriété du programme : « Si $X_1, X_2, \ldots, X_n$ sont des variables aléatoires mutuellement indépendantes alors, pour tout m compris entre 1 et $n-1$, et toutes fonctions f et g, les variables $f(X_1, \ldots, X_m)$ et $g(X_{m+1}, \ldots, X_n)$ sont indépendantes. »). De plus, $P(X_{n+1} = 1) = P(X_{n+1} = 0) = \frac{1}{2}$. On obtient

$$P(N_{n+1} = k) = \frac{1}{2}\sum_{i=0}^{1}\Big(P\big(\{N_n = k\} \cap \{X_n = i\}\big) + P\big(\{N_n = k-1\} \cap \{X_n = i\}\big)\Big).$$

Comme Ω est la réunion disjointe de $\{X_n = 0\}$ et de $\{X_n = 1\}$, par regroupements convenables, on conclut que la relation de récurrence est vérifiée.

Ensuite, compte tenu des relations $P(N_n = n+1) = 0$ et $P(N_n = 0) = 0$,

$$\begin{aligned} G_{N_{n+1}}(s) &= \sum_{k=1}^{n+1} P(N_{n+1} = k)s^k \\ &= \frac{1}{2}\sum_{k=1}^{n+1} P(N_n = k)s^k + \frac{1}{2}\sum_{k=1}^{n+1} P(N_n = k-1))s^k = \frac{1}{2}G_{N_n}(s) + \frac{s}{2}G_{N_n}(s). \end{aligned}$$

On obtient ainsi la relation de récurrence $G_{N_{n+1}}(s) = \dfrac{1+s}{2}G_{N_n}(s)$. Le résultat attendu s'en déduit alors par récurrence immédiate.

Méthode 2.

Soit n et $k \in \mathbb{N}^*$. On considère l'application ϕ de Ω vers $\mathscr{P}([\![2, n]\!])$ qui à u associe l'ensemble d'entiers $\{i \in [\![2, n]\!] \mid u_i \neq u_{i-1}\}$. Ainsi, pour $k \in \mathbb{N}^*$, $\{N_n = k\} = \{u \in \Omega \mid \mathrm{card}\,\phi(u) = k-1\}$, donc $\{N_n = k\}$ est la réunion disjointe des $\{u \in \Omega \mid \phi(u) = A\}$ lorsque A parcourt l'ensemble des parties de $[\![2, n]\!]$ ayant exactement $k-1$ éléments ; il y en a $\binom{n-1}{k-1}$.

Maintenant, pour une partie A de $[\![2, n]\!]$ à exactement $k-1$ éléments, on peut écrire :

$$P(\phi = A, X_1 = \alpha_1) = P(X_1 = \alpha_1, X_2 = \alpha_2, \ldots, X_n = \alpha_n),$$

où $\alpha_1, \alpha_2, \ldots, \alpha_n$ ne dépendent que du choix de $\alpha_1 \in \{0, 1\}$ et de celui de A. Donc, pour tout $\alpha_1 \in \{0, 1\}$,

$$P(\phi = A, X_1 = \alpha_1) = p^n = \frac{1}{2^n}.$$

Ainsi, on obtient $P(\phi = A) = P(\phi = A, X_1 = 0) + P(\phi = A, X_1 = 1) = \dfrac{1}{2^{n-1}}$. Ceci permet de conclure que $P(N_n = k) = \binom{n-1}{k-1}\dfrac{1}{2^{n-1}}$. Finalement,

$$G_{N_n}(s) = \sum_{k=1}^{n} \binom{n-1}{k-1} \frac{s^k}{2^{n-1}} = s\Big(\frac{1+s}{2}\Big)^{n-1}.$$

Méthode 3 (d'après Bruno Jaffuel).

Soit $n \in \mathbb{N}$ tel que $n \geqslant 2$.

Soit $k \in [\![2, n]\!]$, on pose $Y_k = |X_k - X_{k-1}|$; Y_k est une variable aléatoire à valeurs dans $\{0, 1\}$.

Or $P(Y_k = 1) = P(X_k = 0, X_{k-1} = 1) + P(X_k = 1, X_{k-1} = 0) = \dfrac{2}{4} = \dfrac{1}{2}$; ainsi, Y_k suit une loi de Bernoulli de paramètre $\dfrac{1}{2}$.

Comme $N_n = 1 + \sum_{k=2}^{n} Y_k$, pour établir que $N_n - 1 \hookrightarrow \mathscr{B}(n-1, 1/2)$, il reste à montrer que les Y_k sont mutuellement indépendantes.

On prouve en fait l'indépendance de $X_1, Y_2, \ldots Y_n$. Soit $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) \in \{0, 1\}^n$. On va montrer $P(X_1 = \varepsilon_1, Y_2 = \varepsilon_2, \ldots, Y_n = \varepsilon_n) = P(X_1 = \varepsilon_1) \prod_{j=2}^{n} P(Y_j = \varepsilon_j)$, c'est-à-dire

$$P(X_1 = \varepsilon_1, Y_2 = \varepsilon_2, \ldots, Y_n = \varepsilon_n) = \frac{1}{2^n}.$$

On considère l'application ψ de $\{0, 1\}^2$ dans $\{0, 1\}$ définie par

$$\psi(u, v) = u(1 - v) + v(1 - u)\ ;$$

ainsi, $\psi(0, 0) = 0, \psi(1, 0) = 1, \psi(1, 1) = 0, \psi(0, 1) = 1$.

On définit alors la fonction $f : \{0, 1\}^n \longrightarrow \{0, 1\}^n$ par

$$f(\lambda_1, \lambda_2, \ldots, \lambda_{n-1}, \lambda_n) = (\mu_1, \mu_2, \ldots, \mu_{n-1}, \mu_n),$$

où $\mu_1 = \lambda_1$ et, pour $i \in [\![1, n-1]\!]$, $\mu_{i+1} = \psi(\mu_i, \lambda_{i+1})$; f est bijective (il suffit d'établir l'injectivité, qui utilise une récurrence).

Si on note $(\theta_1, \ldots, \theta_n) = f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$, cette construction donne :

$$P(X_1 = \varepsilon_1, Y_2 = \varepsilon_2, \ldots, Y_n = \varepsilon_n) = P(X_1 = \theta_1, X_2 = \theta_2, \ldots, X_n = \theta_n).$$

« C'est étudié pour ! », comme aurait dit Fernand Raynaud.

On peut conclure dans l'allégresse.