Une application des séries à la marche aléatoire de Bernoulli

Aimé Lachal

22 mai 2008

Avertissement : ce problème qui illustre la théorie des séries (séries numériques, séries de fonctions, séries entières) nécessite quelques connaissances élémentaires en probabilités qui sont rappelées dans l'appendice en fin de document. On résout une équation de convolution discrète (équation (2)) à l'aide de certaines séries entières (fonctions génératrices).

1 Présentation du problème

Considérons une personne se déplaçant au hasard sur une droite (représentée par $\mathbb{Z}$), effectuant à chaque instant (à valeurs dans $\mathbb{N}$) soit un pas en avant avec une probabilité $p \in]0,1[$, soit un pas en arrière avec la probabilité complémentaire $q = 1 - p$. À l'instant 0, elle est située à l'origine.

Problème : cette personne a-t-elle une chance de repasser par sa position de départ ? Si oui, avec quelle probabilité et à quel instant ? Pour répondre à ces questions, nous déterminons :

- la loi de probabilité du premier instant de retour au point de départ ;
- le temps moyen de retour au point de départ.

2 Modélisation du problème

Notons à l'instant $n \in \mathbb{N}^*$: $X_n = +1$ si le marcheur avance d'un pas, $X_n = -1$ s'il recule d'un pas. Pour chaque n, la quantité X_n est une variable aléatoire de Bernoulli de paramètre p à valeurs dans $\{-1, +1\}$: $\mathbb{P}\{X_n = +1\} = p$ et $\mathbb{P}\{X_n = -1\} = q$. On suppose les variables X_n, $n \in \mathbb{N}^*$, indépendantes, c'est-à-dire le choix du prochain pas ne dépend pas des précédents. Posons $S_0 = 0$ (la position initiale est l'origine) et $S_n = X_1 + \cdots + X_n$, $n \in \mathbb{N}^*$. La variable aléatoire S_n représente la position du marcheur au bout de n pas.

Les variables aléatoires de Bernoulli standards étant à valeurs dans $\{0, 1\}$, il est naturel d'introduire $Y_n = (X_n + 1)/2$ et alors Y_n est une variable aléatoire de Bernoulli de paramètre p : $\mathbb{P}\{Y_n = 1\} = p$ et $\mathbb{P}\{Y_n = 0\} = q$. Puis $T_n = Y_1 + \cdots + Y_n = (S_n + n)/2$ est une variable aléatoire binomiale de paramètres (n, p). La variable T_n est à valeurs dans $\{0, 1, \ldots, n\}$ et S_n à valeurs dans $\{-n, -n+2, \ldots, n-2, n\}$. Ces considérations fournissent la loi de probabilité de T_n puis celle de S_n :

$$\forall k \in \{0, 1, \ldots, n\},\ \mathbb{P}\{S_n = 2k - n\} = \mathbb{P}\{T_n = k\} = \binom{n}{k} p^k q^{n-k},$$

soit encore

$$\mathbb{P}\{S_n = j\} = \binom{n}{(n+j)/2} p^{(n+j)/2} q^{(n-j)/2} \tag{1}$$

pour j de même parité que n tel que $|j| \leqslant n$.

Introduisons le premier instant de retour à l'origine lorsqu'il existe :

$$\tau_0 = \min\{n \geqslant 1 : S_n = 0\}.$$

L'égalité $\tau_0 = n$ pour un $n \in \mathbb{N}^*$ signifie que $S_0 = 0$, $S_1, S_2, \dots, S_{n-1} \neq 0$ et $S_n = 0$; le marcheur repasse par l'origine pour la première fois à l'instant n. Si $\forall n \in \mathbb{N}^*,\ S_n \neq 0$, on pose $\tau_0 = +\infty$; dans ce cas, le marcheur ne repasse jamais par l'origine.

On cherche la loi de probabilité de τ_0 *i.e.* la famille de probabilités $\mathbb{P}\{\tau_0 = n\}$, $n \in \mathbb{N}^* \cup \{+\infty\}$.

3 Résolution du problème

3.1 Une équation de convolution discrète

L'idée consiste à regarder une marche qui passe par l'origine au temps n, donc pour laquelle le premier retour en l'origine a eu lieu à un instant inférieur ou égal à n (c'est-à-dire $\tau_0 \leqslant n$), disons k, $k \in \{1, \dots, n\}$. Puis on découpe cette marche en deux sous-marches indépendantes, l'une entre les temps 1 et k, l'autre entre les temps k et n. La deuxième sous-marche démarre donc de l'origine et se reproduit à l'identique (en loi) d'une marche entre les instants 0 et $n-k$ (voir Fig. 1).

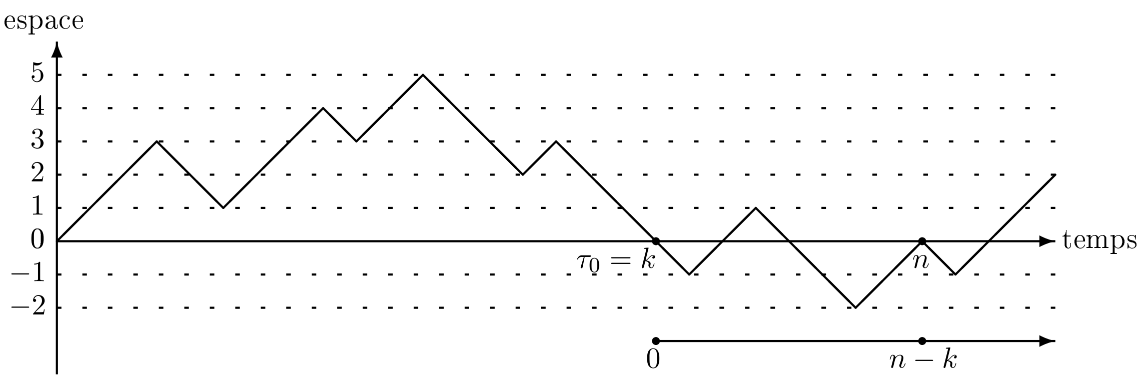

FIGURE 1 – Marche aléatoire, loi horaire

Cette discussion conduit à la relation

$$\mathbb{P}\{S_n = 0\} = \mathbb{P}\{\tau_0 \leqslant n, S_n = 0\} = \sum_{k=1}^{n} \mathbb{P}\{\tau_0 = k, S_n = 0\} = \sum_{k=1}^{n} \mathbb{P}\{\tau_0 = k\}\mathbb{P}\{S_{n-k} = 0\}.$$

Posons $u_n = \mathbb{P}\{S_n = 0\}$ et $v_n = \mathbb{P}\{\tau_0 = n\}$ pour tout $n \geqslant 1$. Il est naturel de poser également $u_0 = 1$ et $v_0 = 0$ puisque $S_0 = 0$ et $\tau_0 \geqslant 1$. Nous avons obtenu l'« équation de convolution discrète »

$$\forall n \in \mathbb{N}^*,\ u_n = \sum_{k=0}^{n} u_k v_{n-k} \qquad \text{avec } u_0 = 1. \tag{2}$$

3.2 Fonctions génératrices

Pour résoudre l'équation de convolution (2), on fait appel aux fonctions génératrices des suites $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$:

$$G(z) = \sum_{n=0}^{+\infty} u_n z^n \text{ et } H(z) = \sum_{n=0}^{+\infty} v_n z^n.$$

Notons que $H(z)$ représente l'espérance $\mathbb{E}\left(z^{\tau_0}\mathbb{1}_{\{\tau_0<+\infty\}}\right)$ qui n'est autre que la fonction génératrice de la variable aléatoire τ_0 assujettie à rester finie.

D'une part, on a d'après (1)

$$G(z)=\sum_{n=0}^{+\infty}\binom{n}{n/2}(pqz^2)^{n/2}=\sum_{n=0}^{+\infty}\binom{2n}{n}(pqz^2)^n.$$

Remarquons que $\binom{2n}{n}=\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\cdots\left(-\frac{1}{2}-(n-1)\right)\frac{(-4)^n}{n!}$. En utilisant le développement en série de Taylor

$$(1+\zeta)^\alpha=\sum_{n=0}^{+\infty}\alpha(\alpha-1)\cdots(\alpha-n+1)\,\frac{\zeta^n}{n!}$$

qui est une série entière de rayon de convergence 1, on obtient, avec les choix $\alpha=-1/2$ et $\zeta=-4pqz^2$:

$$G(z)=\frac{1}{\sqrt{1-4pqz^2}}.$$

Le rayon de convergence de cette série entière est $1/(2\sqrt{pq})$; ce rayon est supérieur ou égal à 1 comme on peut le voir à l'aide de l'inégalité élémentaire $p(1-p)\leqslant 1/4$.

D'autre part, comme $\forall n\in\mathbb{N},\ v_n\in[0,1]$ et

$$\sum_{n=0}^{+\infty}v_n=\sum_{n=1}^{+\infty}v_n=\mathbb{P}\left(\bigcup_{n=1}^{+\infty}\{\tau_0=n\}\right)=\mathbb{P}\{\tau_0<+\infty\},$$

on voit que la série entière $H(z)$ converge absolument en $z=1$, donc son rayon de convergence est au moins 1 et $H(1)=\mathbb{P}\{\tau_0<+\infty\}$.

L'équation de convolution (2) conduit alors, en utilisant le théorème sur les séries-produits, à l'équation fonctionnelle $G(z)H(z)=G(z)-1$, $|z|<1$, de laquelle on tire $H(z)=1-\frac{1}{G(z)}$, soit

$$H(z)=1-\sqrt{1-4pqz^2},\quad |z|<1. \tag{3}$$

La convergence de la série à termes positifs $(\sum v_n)$ montre que la série de fonctions $(\sum(x\mapsto v_nx^n))$ est normalement convergente sur l'intervalle $[-1,1]$ et donc

$$\lim_{\substack{x\to1^-\\x\in\mathbb{R}}}\sum_{n=1}^{+\infty}v_nx^n=\sum_{n=1}^{+\infty}v_n.$$

En conséquence, (3) est valable aussi pour $z=1$:

$$H(1)=\lim_{\substack{x\to1^-\\x\in\mathbb{R}}}H(x)=1-\sqrt{1-4pq}. \tag{4}$$

3.3 Inversion de la fonction génératrice H et loi de τ_0

En développant en série de Taylor la quantité $\sqrt{1-4pqz^2}$ apparaissant dans (3), il vient

$$H(z)=\sum_{n=1}^{+\infty}\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\cdots\left(\frac{1}{2}-(n-1)\right)\frac{(-4pqz^2)^n}{n!}=\sum_{n=1}^{+\infty}\frac{\binom{2n}{n}}{2n-1}(pq)^nz^{2n}.$$

Par identification avec l'expression initiale de $H(z)$, on obtient finalement

$$v_{2n}=\frac{\binom{2n}{n}}{2n-1}(pq)^n\ \text{et}\ v_{2n+1}=0.$$

La loi de probabilité de la variable aléatoire τ_0 est ainsi donnée par

$$\boxed{\mathbb{P}\{\tau_0 = n\} = \begin{cases} \frac{\binom{n}{n/2}}{n-1}\,(pq)^{n/2} & \text{si } n \text{ est pair non nul,} \\ 0 & \text{si } n \text{ est impair.} \end{cases}}$$

Avec (4), on a $H(1) = \mathbb{P}\{\tau_0 < +\infty\} = 1 - \sqrt{1-4pq}$. En remplaçant q par $1-p$, on trouve

$$\mathbb{P}\{\tau_0 = +\infty\} = \sqrt{1-4pq} = \sqrt{4p^2 - 4p + 1} = |2p-1|,$$

soit

$$\boxed{\mathbb{P}\{\tau_0 = +\infty\} = |2p-1|.}$$

On a en particulier

$$\mathbb{P}\{\tau_0 = +\infty\} \begin{cases} = 0 & \text{si } p = 1/2, \\ > 0 & \text{si } p \neq 1/2. \end{cases}$$

Ce résultat signifie que pour une marche équilibrée ($p = 1/2$), le marcheur repassera presque sûrement (*i.e.* avec probabilité 1) par sa position initiale : on parle de marche *récurrente*, alors que dans le cas d'une marche désiquilibrée ($p \neq 1/2$), il y a une dérive vers $+\infty$ si $p > 1/2$, vers $-\infty$ si $p < 1/2$ et le retour à l'origine pourrait se produire, mais ce avec une probabilité inférieure à 1. Dans ce cas il y a une probabilité non nulle de ne jamais retourner à l'origine : on parle de marche *transitoire* ou *transiente*.

3.4 Espérance de τ_0

Plaçons-nous à présent dans le cas $p = 1/2$. On a $H(z) = 1 - \sqrt{1-z^2}$ pour $|z| < 1$. À l'intérieur du disque de convergence, *i.e.* pour $|z| < 1$, on peut dériver la série H terme à terme :

$$H'(z) = \frac{z}{\sqrt{1-z^2}} = \sum_{n=1}^{+\infty} n v_n z^{n-1}.$$

Examinons la limite de $H'(x)$ lorsque $x \to 1^-, x \in \mathbb{R}$. Si la série numérique $(\sum n v_n)$ était convergente, la série de fonctions $(\sum (x \mapsto n v_n x^{n-1}))$ serait normalement convergente sur l'intervalle $[-1,1]$ et on aurait

$$\lim_{\substack{x \to 1^- \\ x \in \mathbb{R}}} \sum_{n=1}^{+\infty} n v_n x^{n-1} = \sum_{n=1}^{+\infty} n v_n.$$

Or

$$\forall x \in \,]-1,1[, \; \sum_{n=1}^{+\infty} n v_n x^{n-1} = H'(x) = \frac{x}{\sqrt{1-x^2}}$$

et donc

$$\lim_{x \to 1^-} \sum_{n=1}^{+\infty} n v_n x^{n-1} = \lim_{x \to 1^-} \frac{x}{\sqrt{1-x^2}} = +\infty$$

d'où une contradiction. La série $(\sum n v_n)$ est donc divergente et

$$\sum_{n=1}^{+\infty} n v_n = +\infty.$$

Notant que cette somme représente l'espérance de la variable aléatoire τ_0, on obtient

$$\boxed{\mathbb{E}(\tau_0) = +\infty.}$$

Ce résultat signifie que, bien que le marcheur repasse presque sûrement par sa position initiale, le temps moyen de retour est infini.

Appendice : quelques rappels de probabilités

A.1 Loi de Bernoulli et loi binomiale

Une variable aléatoire X prenant les deux valeurs 0 et 1 avec les probabilités respectives $1-p$ et p suit la loi de Bernoulli de paramètre p.

Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes de Bernoulli de paramètre commun p. La variable aléatoire $S = X_1 + \cdots + X_n$, représente le nombre de 1 obtenus parmi les $X_1, \dots, X_n$; elle suit la loi binomiale de paramètres (n, p). Cette loi de probabilité est donnée par

$$\mathbb{P}\{S = k\} = \binom{n}{k} p^k q^{n-k}, \ k \in \{0, 1, \dots, n\}.$$

A.2 Formule des probabilités totales

Si l'on a une partition de l'univers $\Omega = \cup_{n=1}^{+\infty} B_n$ (les B_n étant des parties non vides de Ω deux à deux disjointes), alors pour tout événement A :

$$\mathbb{P}(A) = \sum_{n=1}^{+\infty} \mathbb{P}(A \cap B_n).$$

En particulier, si X est une variable aléatoire à valeurs dans $\mathbb{N}^*$, les événements $\{X = n\}$ forment une partition de l'univers et l'on a

$$\mathbb{P}(A) = \sum_{n=1}^{+\infty} \mathbb{P}(A \cap \{X = n\}).$$

A.3 Espérance d'une variable aléatoire discrète

Si X est une variable aléatoire à valeurs dans $\mathbb{N}^*$ de loi de probabilité la famille $(p_n)_{n \in \mathbb{N}^*}$ ($p_n = \mathbb{P}\{X = n\}$), l'espérance de X est définie selon

$$\mathbb{E}(X) = \sum_{n=1}^{+\infty} n p_n.$$

Si cette série est divergente, on a $\mathbb{E}(X) = +\infty$.

A.4 Fonction génératrice d'une variable aléatoire discrète

Soient X une variable aléatoire à valeurs dans $\mathbb{N}^*$ de loi de probabilité la famille $(p_n)_{n \in \mathbb{N}^*}$ et f une application de $\mathbb{N}^*$ à valeurs dans $\mathbb{C}$. Si la série $(\sum p_n f(n))$ est absolument convergente, on a

$$\mathbb{E}(f(X)) = \sum_{n=1}^{+\infty} p_n f(n).$$

En particulier la fonction génératrice de X est la fonction de la variable complexe z définie par :

$$G_X(z) = \mathbb{E}(z^X) = \sum_{n=1}^{+\infty} p_n z^n.$$

Comme la série à termes positifs $(\sum p_n)$ est convergente, la série entière G_X admet un rayon de convergence au moins égal à 1. On peut donc dériver la série entière G_X terme à terme à l'intérieur du disque de rayon 1 :

$$G_X'(z) = \sum_{n=1}^{+\infty} n p_n z^{n-1}, \quad |z| < 1.$$

Si de plus l'espérance de X est finie, la série à termes positifs $(\sum np_n)$ est convergente, et la série de fonctions $(\sum(x \mapsto np_n x^{n-1}))$ est normalement convergente dans l'intervalle $[-1, 1]$. En conséquence, on a

$$\lim_{\substack{x \to 1^- \\ x \in \mathbb{R}}} \sum_{n=1}^{+\infty} np_n x^{n-1} = \sum_{n=1}^{+\infty} np_n.$$

On récupère ainsi l'espérance de X en dérivant la fonction génératrice en 1 :

$$\mathbb{E}(X) = G_X'(1).$$