Nombre d'états visités par une marche aléatoire

mots-clés : marche aléatoire, transience, séries génératrices.

Soit $d \geq 1$ un entier, et μ une mesure de probabilité sur le réseau $\mathbb{Z}^d$ à support fini : $\mathrm{card}\,(\{x \in \mathbb{Z}^d \mid \mu(x) > 0\}) < +\infty$. La *marche aléatoire* de loi des pas μ est la suite de vecteurs aléatoires $(X_n)_{n\in\mathbb{N}}$ définie par

$$X_0 = (0, \ldots, 0) \qquad ; \qquad \forall n \geq 1,\;\; X_n = T_1 + T_2 + \cdots + T_n,$$

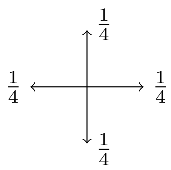
où les pas $T_{i\geq 1}$ sont des vecteurs aléatoires indépendants et tous de loi μ :

$$\forall x \in \mathbb{Z}^d,\;\; \forall i \geq 1,\;\; \mathbb{P}[T_i = x] = \mu(x).$$

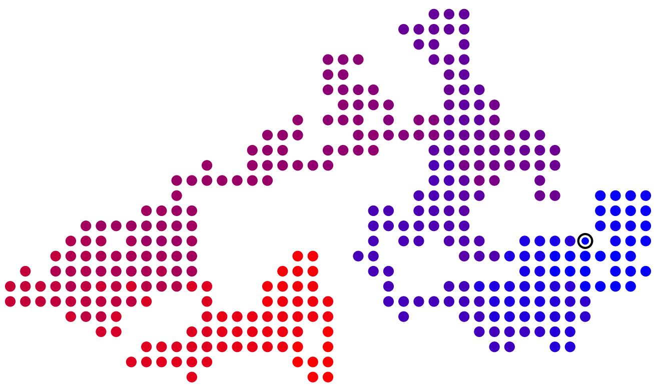
Par exemple, si $d = 1$ et $\mu = \frac{1}{2}\,\delta_1 + \frac{1}{2}\,\delta_{-1}$, alors on retrouve avec $(X_n)_{n\in\mathbb{N}}$ la marche aléatoire simple symétrique sur l'ensemble des entiers $\mathbb{Z}$. Pour $n \geq 1$, on pose

$$C_n = \mathrm{card}\,(\{X_1, X_2, \ldots, X_n\}).$$

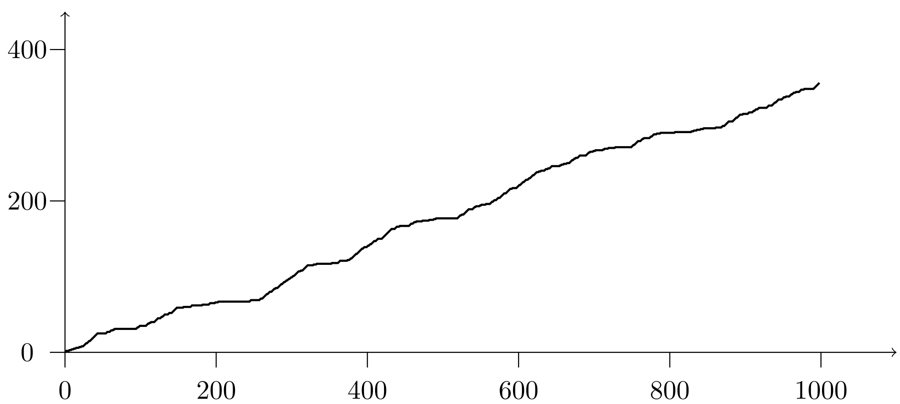
On a $C_n \leq n$ pour tout $n \geq 1$. L'objectif de ce texte est de comprendre le comportement asymptotique de la suite aléatoire $(C_n)_{n\geq 1}$ dans diverses situations. Si $d = 2$ et si $\mu = \frac{1}{4}\,(\delta_{(1,0)} + \delta_{(-1,0)} + \delta_{(0,1)} + \delta_{(0,-1)})$ est la distribution de pas représentée ci-dessous :



alors on obtient la trajectoire suivante pour $(X_n)_{0\leq n\leq 1000}$:



Les états visités sont coloriés uniquement lors de leur première visite, avec une couleur qui dépend du temps de première visite (bleu au temps $n = 0$, et de plus en plus rouge jusqu'au temps $n = 1000$). La courbe suivante représente alors $(C_n)_{0\leq n\leq 1000}$:



À première vue, $(C_n)_{n\in\mathbb{N}}$ semble croître linéairement avec n (avec bien sûr des fluctuations aléatoires autour de ce comportement moyen). Nous verrons plus loin que c'est bien le cas pour les marches aléatoires transientes, mais qu'il y a une correction logarithmique à la croissance linéaire dans le cas de la marche bidimensionnelle présentée ci-dessus.

1 Taux de croissance linéaire

Dans cette section, d et μ sont arbitraires. On note $R_0^+ = \inf\{n \geq 1 \mid X_n = (0,\dots,0)\}$, avec par convention $R_0^+ = +\infty$ si la marche aléatoire $(X_n)_{n\in\mathbb{N}}$ ne repasse jamais par l'origine. Soit $\alpha = \mathbb{P}[R_0^+ = +\infty]$. L'origine $0 = (0,\dots,0)$ est un état transient si et seulement si $\alpha > 0$.

Théorème 1 (Kesten–Spitzer–Whitman). *La suite $(C_n)_{n\geq 1}$ vérifie :*

$$\frac{C_n}{n} \to_{(n\to\infty,\,\textit{probabilité})} \alpha.$$

Démonstration. Estimons les deux premiers moments de C_n. La différence $C_{n+1} - C_n$ vaut 1 si $X_{n+1} \notin \{X_1,\dots,X_n\}$, et 0 sinon. On a donc pour tout $n \geq 1$:

$$C_n = \sum_{i=1}^n 1_{(X_i\notin\{X_1,\dots,X_{i-1}\})} = \sum_{i=1}^n 1_{(X_i-X_1\neq 0,\dots,X_i-X_{i-1}\neq 0)}.$$

En prenant les espérances, on obtient :

$$\mathbb{E}\left[\frac{C_n}{n}\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{P}[T_i+\cdots+T_2\neq 0,\dots,T_i+T_{i-1}\neq 0,\ T_i\neq 0] = \frac{1}{n}\sum_{i=1}^n \mathbb{P}[X_1\neq 0,\ X_2\neq 0,\dots,X_{i-1}\neq 0].$$

L'espérance de $\frac{C_n}{n}$ est donc la moyenne de Cesàro d'une suite tendant vers α, et $\mathbb{E}[\frac{C_n}{n}] \to_{n\to\infty} \alpha$. Pour le second moment, on développe le carré de C_n :

$$(C_n)^2 = \sum_{1\leq i,j\leq n} 1_{(X_i\notin\{X_1,\dots,X_{i-1}\})}\, 1_{(X_j\notin\{X_1,\dots,X_{j-1}\})} = 2\sum_{1\leq i<j\leq n} 1_{(X_i\notin\{X_1,\dots,X_{i-1}\},\ X_j\notin\{X_1,\dots,X_{j-1}\})} + C_n.$$

L'espérance de chaque variable de Bernoulli dans la somme ci-dessus peut être majorée comme suit :

$$\begin{aligned}
&\mathbb{P}[X_i\notin\{X_1,\dots,X_{i-1}\},\ \ X_j\notin\{X_1,\dots,X_{j-1}\}]\\
&\leq \mathbb{P}[X_i\notin\{X_1,\dots,X_{i-1}\},\ \ X_j\notin\{X_i,\dots,X_{j-1}\}]\\
&\leq \mathbb{P}[T_i\neq 0,\ T_i+T_{i-1}\neq 0,\dots,T_i+\cdots+T_2\neq 0,\ T_j\neq 0,\ T_j+T_{j-1}\neq 0,\ T_j+\cdots+T_{i+1}\neq 0]\\
&\leq \mathbb{P}[T_i\neq 0,\ T_i+T_{i-1}\neq 0,\dots,T_i+\cdots+T_2\neq 0]\ \mathbb{P}[T_j\neq 0,\ T_j+T_{j-1}\neq 0,\ T_j+\cdots+T_{i+1}\neq 0] = \alpha_i\,\alpha_{j-i},
\end{aligned}$$

avec $\alpha_i = \mathbb{P}[X_1\neq 0,\ X_2\neq 0,\dots,X_{i-1}\neq 0]$. Comme $\alpha_i\to\alpha$ et $\alpha_{j-i}\to\alpha$ lorsque i et $j-i$ tendent vers l'infini, $\mathbb{E}[(\frac{C_n}{n})^2]$ est majoré par $\frac{\mathbb{E}[C_n]}{n^2} = O(n^{-1})$ plus une moyenne de Cesàro de termes proches de α^2. Ainsi,

$$\mathbb{E}\left[\left(\frac{C_n}{n}\right)^2\right] \leq \alpha^2 + o(1). \tag{1}$$

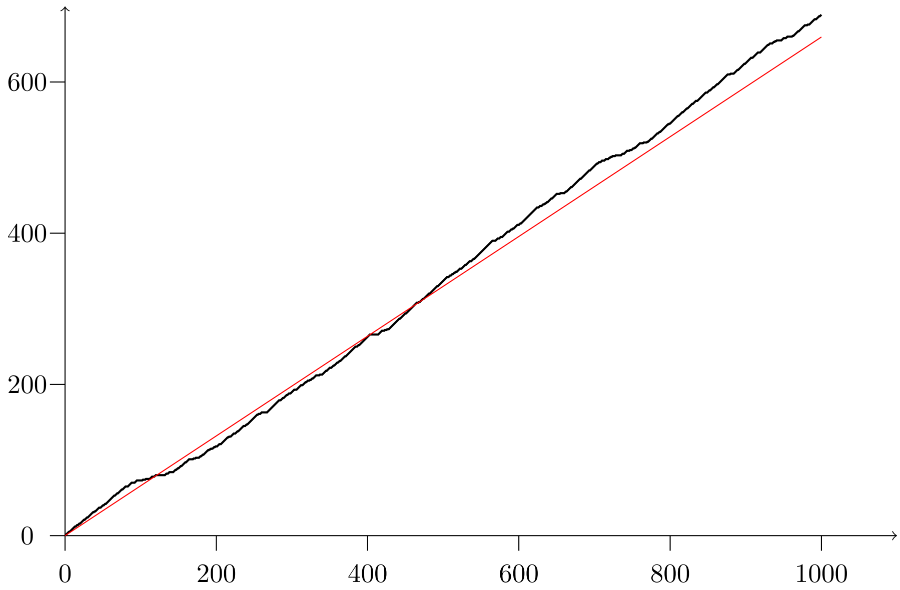
En retranchant le carré de l'espérance, on obtient donc $\mathrm{var}(\frac{C_n}{n}) = o(1)$, et l'inégalité de Bienaymé–Chebyshev donne maintenant

$$\mathbb{P}\left[\left|\frac{C_n}{n} - \mathbb{E}\left[\frac{C_n}{n}\right]\right| \geq \varepsilon\right] \leq \frac{\mathrm{var}(\frac{C_n}{n})}{\varepsilon^2} \to_{n\to\infty} 0$$

pour tout $\varepsilon > 0$. On conclut en utilisant le fait que $\mathbb{E}[\frac{C_n}{n}] \to_{n\to\infty} \alpha$. $\square$

Considérons par exemple la marche aléatoire sur $\mathbb{Z}^3$ dont la loi des pas est uniforme sur les six directions $(\pm 1,0,0)$, $(0,\pm 1,0)$ et $(0,0,\pm 1)$. Cette marche est transiente, avec un coefficient α pour lequel il existe une formule exacte (obtenue notamment par Montroll en 1964) :

$$\alpha = \frac{(2\pi)^3}{3\int_{(-\pi,\pi)^3} \frac{d\theta\,d\phi\,d\psi}{3-\cos\theta-\cos\phi-\cos\psi}} \simeq 0.6595\dots \tag{2}$$



La courbe ci-dessus représente dans ce cas une trajectoire de $(C_n)_{0\leq n\leq 1000}$, ainsi que la croissance linéaire αn (en rouge).

2 Récurrence de la marche bidimensionnelle

D'après ce qui précède, si l'origine est un état récurrent, alors $\frac{C_n}{n}$ tend vers 0, et le nombre d'états visités croît sous-linéairement. Pour étudier plus précisément le comportement de C_n dans ce cas, on a besoin de savoir à quelle vitesse $\mathbb{P}[R_0^+ > n]$ tend vers 0.

Proposition 2. *Pour la marche aléatoire bidimensionnelle standard,*

$$\mathbb{P}[X_{2n} = (0,0)] = \frac{1}{\pi n} + O\left(\frac{1}{n^2}\right).$$

Démonstration. La parité de la somme de l'abscisse et l'ordonnée change à chaque pas de la marche bidimensionnelle, ce qui explique pourquoi on ne calcule que la probabilité d'un retour en 0 à un temps pair $2n$. Cette probabilité est $\frac{1}{4^{2n}}$ fois le nombre de chemins de longueur $2n$ reliant $(0,0)$ à $(0,0)$. Un tel chemin est constitué de n_1 pas vers le haut, n_1 pas vers le bas, n_2 pas vers la droite et n_2 pas vers la gauche avec $n_1+n_2=n$. On a donc :

$$\mathbb{P}[X_{2n} = (0,0)] = \frac{1}{4^{2n}} \sum_{n_1+n_2=n} \frac{(2n)!}{(n_1!)^2\,(n_2!)^2} = \frac{1}{4^{2n}}\binom{2n}{n} \sum_{n_1+n_2=n} \binom{n}{n_1}\binom{n}{n_2}.$$

L'identité de Vandermonde assure que la somme des produits de coefficients binomiaux vaut $\binom{2n}{n}$, donc

$$\mathbb{P}[X_{2n} = (0,0)] = \frac{1}{4^{2n}}\binom{2n}{n}^2. \tag{3}$$

L'estimée se déduit maintenant immédiatement de la formule de Stirling $n! = (\frac{n}{\mathrm{e}})^n\sqrt{2\pi n}\,(1+O(n^{-1}))$. Notons que l'estimée implique que la marche aléatoire bidimensionnelle standard est récurrente : en effet, $\sum_{n=0}^{\infty}\mathbb{P}[X_n = (0,0)] = +\infty$. ☐

On se concentre dans ce qui suit sur la marche aléatoire bidimensionnelle. Posons $G_1(z) = \sum_{n=0}^{\infty}\mathbb{P}[X_{2n} = (0,0)]\,z^n$. C'est une série entière de rayon de convergence 1, et lorsque z s'approche de 1 en restant dans le disque unité,

$$G_1(z) = 1 + \frac{1}{\pi}\sum_{n=1}^{\infty}\frac{z^n}{n} + O\left(\sum_{n=1}^{\infty}\frac{|z|^n}{n^2}\right) = -\frac{1}{\pi}\log(1-z) + O(1)$$

car la série $\sum_{n=1}^{\infty}\frac{1}{n^2}$ est convergente. Posons ensuite $G_2(z) = \sum_{n=1}^{\infty}\mathbb{P}[R_0^+ = 2n]\,z^n$. On introduit plus généralement le temps de k-ième retour

$$R_0^{(k)} = \inf\{n > R_0^{(k-1)} \mid X_n = (0,0)\},$$

étant entendu que $R_0^{(k)} = +\infty$ s'il y a moins de k retours en l'origine. La propriété de Markov forte permet de montrer par récurrence sur $k \geq 1$ l'identité suivante :

$$\forall k \geq 1,\ G_{2,k}(z) = \sum_{n=1}^{\infty} \mathbb{P}[R_0^{(k)} = 2n]\, z^n = (G_2(z))^k. \tag{4}$$

Proposition 3. *Pour tout z nombre complexe de module plus petit que 1, $G_1(z) = \frac{1}{1-G_2(z)}$.*

Démonstration. Si $X_{2n} = (0,0)$, alors $2n$ est l'un des temps de retour en 0. On a donc :

$$G_1(z) = 1 + \sum_{n=1}^{\infty} \mathbb{P}[X_{2n} = (0,0)]\, z^n = 1 + \sum_{k=1}^{\infty}\left(\sum_{n=1}^{\infty} \mathbb{P}[R_0^{(k)} = 2n]\, z^n\right) = 1 + \sum_{k=1}^{\infty}(G_{2,k}(z)),$$

et l'équation (4) permet de conclure. □

En particulier, puisque $G_1(z)$ tend vers l'infini lorsque z tend vers 1, $G_2(z)$ doit tendre vers 1 et ceci implique que $\mathbb{P}[R_0^+ < +\infty] = G_2(1) = 1$; la marche aléatoire bidimensionnelle est donc récurrente. Par le Théorème 1, on a donc dans ce contexte la convergence en probabilité $\frac{C_n}{n} \to 0$. Considérons finalement $G_3(z) = \sum_{n=0}^{\infty} \mathbb{P}[R_0^+ > 2n]\, z^n$. Comme $\mathbb{P}[R_0^+ > 2n] = \sum_{k>n} \mathbb{P}[R_0^+ = 2k]$, une interversion de sommes donne la relation

$$G_3(z) = \frac{G_2(z) - 1}{z - 1} = \frac{1}{(1-z)\, G_1(z)}. \tag{5}$$

Ceci implique qu'au voisinage de $z = 1$, la série entière $G_3(z)$ s'écrit

$$G_3(z) \sim_{z \to 1} -\frac{\pi}{(1-z)\log(1-z)}.$$

Dans ce contexte, on peut utiliser le :

Théorème 4 (Hardy–Littlewood, 1914). *Soit $G(z) = \sum_{n=0}^{\infty} a_n\, z^n$ une série entière à coefficients positifs et décroissants, de rayon de convergence 1, avec $G(z) \simeq -\frac{1}{1-z}\,\frac{1}{\log(1-z)}$ lorsque z tend vers 1 par valeurs inférieures. Alors, $a_n \simeq_{n\to+\infty} \frac{1}{\log n}$.*

Nous admettrons ce résultat asymptotique, qui est un cas particulier de la théorie de transfert des singularités des séries entières en des comportements asymptotiques de leurs coefficients. Ceci implique l'estimée

$$\alpha_n = \mathbb{P}[R_0^+ \geq n] \simeq \frac{\pi}{\log n}$$

lorsque n tend vers l'infini.

3 Croissance sous-linéaire du nombre d'états pour la marche bidimensionnelle

L'estimée fine $\mathbb{P}[R_0^+ \geq n] \simeq \frac{\pi}{\log n}$ permet de préciser le comportement asymptotique de $(C_n)_{n\in\mathbb{N}}$ pour la marche aléatoire bidimensionnelle :

Théorème 5 (Dvoretzky–Erdös, 1951). *Pour la marche aléatoire bidimensionnelle standard, on a la convergence*

$$\frac{(\log n)\, C_n}{n} \to_{(n\to\infty,\, \textit{probabilité})} \pi.$$

Démonstration. On reprend les mêmes arguments que dans la preuve du Théorème 1. L'espérance de C_n s'écrit

$$\mathbb{E}[C_n] = \sum_{i=1}^{n} \mathbb{P}[X_1 \neq 0,\ X_2 \neq 0, \ldots, X_{i-1} \neq 0] = \sum_{i=1}^{n} \alpha_i,$$

elle est donc équivalente à $\pi\frac{n}{\log n}$. Puis,

$$\mathbb{E}[(C_n)^2] \leq 2 \sum_{1\leq i<j\leq n} \alpha_i\, \alpha_{j-i} + \mathbb{E}[C_n].$$

La somme double est équivalente à $(\frac{\pi n}{\log n})^2$, donc la variance de C_n est un $o(\frac{n}{\log n})^2$. L'inégalité de Bienaymé–Chebyshev permet de conclure. □

Questions

1. Écrire un programme qui calcule la suite $(C_n)_{1\leq n\leq N}$ pour une marche aléatoire $(X_n)_{n\in\mathbb{N}}$ sur $\mathbb{Z}^d$ avec $d \in \{1,2,3,4\}$, et avec les pas choisis uniformément parmi les 2, 4, 6 ou 8 directions $\pm e_i$ possibles, e_i désignant le i-ième vecteur de la base canonique. Dans toutes les questions qui suivent, les marches aléatoires considérées sur ces réseaux auront cette distribution de pas (marche aléatoire *standard*).

2. Donner des estimations du paramètre α lorsque $d=3$ ou 4. Pour $d=3$, comparer avec la formule (2). Lorsque $d=2$, écrire un programme en rapport avec le théorème de Dvoretzky–Erdös.

3. Lorsque $d=1$, on note $M_n = \max(\{X_1,\ldots,X_n\})$ et $m_n = \min(\{X_1,\ldots,X_n\})$. Quelle est la relation entre C_n, M_n et m_n ? Montrer que pour tout $b\in\mathbb{N}$,
$$\mathbb{P}[C_n \geq 2b+1] \leq 2\,\mathbb{P}[M_n \geq b].$$

4. On suppose toujours $d=1$. Soit b et a deux entiers, avec $b\geq 0$. Établir par un décompte de chemins l'identité
$$\mathbb{P}[M_n \geq b \text{ et } X_n = a] = \begin{cases} \mathbb{P}[X_n = a] & \text{si } a \geq b,\\ \mathbb{P}[X_n = 2b-a] & \text{si } a < b.\end{cases}$$
Si $a<b$, on pourra associer à tout chemin $C = (0 = x_0, x_1, \ldots, x_m, \ldots, x_n = a)$ qui atteint ou dépasse b avant le temps n et qui s'arrête en a le chemin $C' = (0 = x_0, x_1, \ldots, x_m, 2b - x_{m+1}, \ldots, 2b - x_n = 2b-a)$ qui commence de la même façon que C, mais qui est réfléchi par rapport à la droite $y=b$ à partir du temps $m\leq n$ où C atteint le niveau b. Faire un dessin de cette transformation, et identifier les ensembles de chemins qu'elle met en bijection.

   Déduire de ce qui précède que $\mathbb{P}[M_n \geq b] = \mathbb{P}[X_n \geq b] + \mathbb{P}[X_n \geq b+1]$, puis que la suite de variables aléatoires $(\frac{C_n}{\sqrt{n}})_{n\geq 1}$ est *tendue* :
$$\lim_{L\to+\infty}\left(\sup_{n\in\mathbb{N}} \mathbb{P}\left[\frac{C_n}{\sqrt{n}} \geq L\right]\right) = 0.$$

5. On admet que $\frac{C_n}{\sqrt{n}}$ admet une loi limite lorsque $d=1$ et n tend vers l'infini ; la question précédente est un premier pas vers ce résultat. Illustrer cette loi limite par des programmes.

6. Démontrer complétement l'estimée (1). On demande en particulier de préciser l'argument de somme de Cesàro, qui est utilisé avec une somme double sur des indices i et j (indication : on pourra mettre de côté dans la somme les indices i et j tels que i ou $j-i$ est trop petit).

7. La preuve de la Proposition 2 repose en particulier sur l'identité de Vandermonde
$$\sum_{k=0}^{l} \binom{n}{k}\binom{m}{l-k} = \binom{n+m}{l}$$
valable pour tous entiers $n,m,l\geq 0$ (dans le texte, on utilise le cas particulier $n=m$). Donner une preuve de cette identité.

8. Si $(X_n^{(1)})_{n\in\mathbb{N}}$ et $(X_n^{(2)})_{n\in\mathbb{N}}$ désignent respectivement les marches aléatoires standards en dimensions $d=1$ et $d=2$, alors l'équation (3) se réécrit :
$$\mathbb{P}[X_{2n}^{(2)} = (0,0)] = \left(\mathbb{P}[X_{2n}^{(1)} = 0]\right)^2.$$
Donner une preuve directe de cette factorisation, en identifiant deux marches aléatoires indépendantes dans $(X_n^{(2)})_{n\in\mathbb{N}}$ (indication : tourner la tête de 45 degrés).

9. Démontrer les identités (4) et (5) (pour la seconde identité, on pourra utiliser le résultat de la Proposition 3).

10. Question bonus. On revient à la question 5. et on note ν la loi limite de $\frac{C_n}{\sqrt{n}}$ en dimension $d=1$, qui est une loi sur $\mathbb{R}_+$. Proposer des programmes qui illustrent le fait suivant : pour toute loi de pas μ sur $\mathbb{Z}$ telle que $\sum_{k\in\mathbb{Z}} \mu(k)\,k = 0$ et $\sum_{k\in\mathbb{Z}} \mu(k)\,k^2 = 1$ (donc, pas forcément avec $\mu = \frac{1}{2}\,\delta_1 + \frac{1}{2}\,\delta_{-1}$), on a la même loi limite ν pour $\frac{C_n}{\sqrt{n}}$.