Marches aléatoires sur $\mathbb{Z}$

Nous nous intéressons ici à des *marches aléatoires* sur $\mathbb{Z}$. Un tel objet est une suite $(S_n)_{n\in N}$ d'entiers, il est décrit par la donnée d'un entier relatif S_0 et d'une suite de variables aléatoires $(X_n)_{n\geq 1}$ indépendantes et de même loi de probabilité, toutes à valeurs dans $\{-1, 1\}$. On définit, pour $n \geq 1$,

$$S_n = S_0 + \sum_{k=1}^{n} X_k.$$

c'est-à-dire que, dans un tel processus, la valeur au temps $n+1$ est celle au temps n, plus le $n+1$-ième pas (la variable X_{n+1}), qui vaut plus ou moins un, selon que ce pas est effectué « vers la gauche » ou « vers la droite ». Dire que les variables aléatoires $(X_n)_{n\geq 1}$ sont indépendantes, c'est exprimer le fait qu'au temps n, on effectue un pas vers la gauche ou vers la droite indépendamment de ce qui s'est passé auparavant.

Exemple : On peut ainsi modéliser une suite de lancers de pile-ou-face : on assigne la valeur 1 au côté pile, la valeur -1 au côté face, la variable X_k représentant alors le k-ième lancer. En prenant $S_0 = 0$, la valeur de S_n nous renseigne alors sur le nombre de fois où la pièce est tombée côté pile : la valeur de S_n est la différence $P - F$, où P est le nombre de « piles » au cours des n premiers lancers, et F le nombre de « faces ». Comme bien sûr $P + F = n$, on obtient $P = (S_n + n)/2$.

Des marches aléatoires peuvent intervenir pour modéliser de nombreuses situations. Ainsi, nous allons voir que de tels raisonnements peuvent nous aider dans la résolution des deux problèmes suivants :

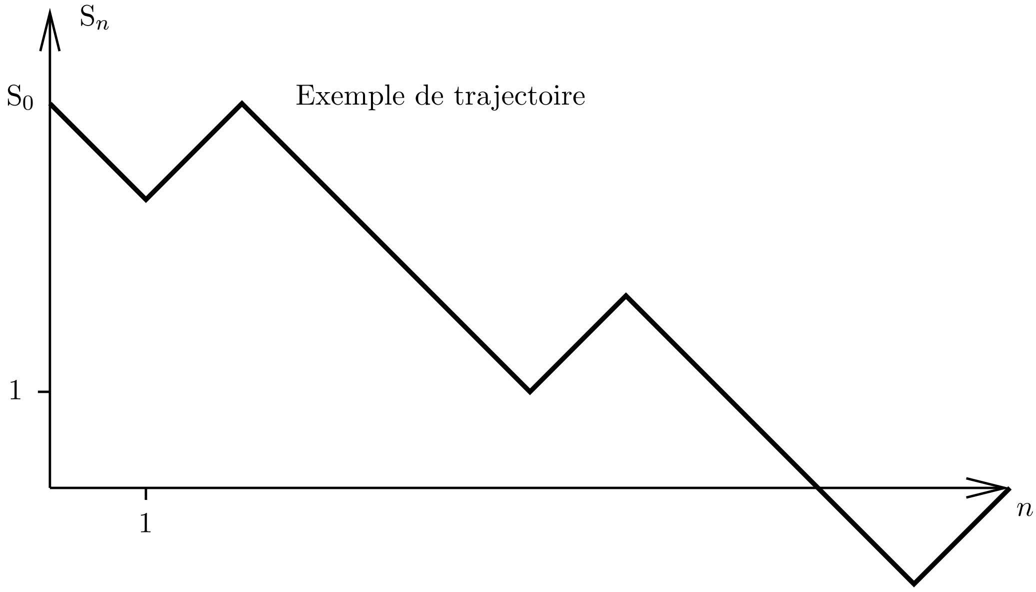
Problème 1 *Au cours d'un scrutin opposant deux candidats* A *et* B*, le candidat* A *obtient* 600 *voix et le candidat* B 400*. Quelle est la probabilité pour que* A *ait été majoritaire (au sens large) tout au long du dépouillement ?*

Problème 2 100 *personnes font la queue à un guichet de cinéma. La place coûte* 5 *Euros.* 60 *personnes ont un billet de* 5 *Euros, tandis que les* 40 *autres ont des billets de* 10 *Euros. Combien faut-il prévoir de billets de* 5 *Euros en caisse pour qu'avec une probabilité d'au moins* 95%*, tous les spectateurs soient servis, dans l'ordre dans lequel ils se présentent à la caisse ?*

Par la suite, sauf précision contraire, nous nous intéresserons à des marches aléatoires *symétriques*, c'est-à-dire telles que la probabilité de faire $+1$ soit la même que celle de faire -1 :

$$\forall n \geq 1, \quad \mathbb{P}[X_n = 1] = \mathbb{P}[X_n = -1] = \frac{1}{2}.$$

On peut représenter dans le plan la marche aléatoire $(S_n)_{0\leq n\leq N}$ à l'aide d'une ligne brisée joignant les $N+1$ points de coordonnées $(i, S_i)_{i=0,\ldots,N}$. La *trajectoire* obtenue est l'une des 2^{N} lignes brisées possibles partant du point $(0, \mathrm{S}_0)$ (on a deux choix possibles à chaque étape), toutes ces trajectoires étant équiprobables dans le cas d'une marche aléatoire symétrique. On parlera de « montée » à l'étape n lorsque l'on a $\mathrm{X}_n = 1$, et de « descente » lorsque $\mathrm{X}_n = -1$.



1 Nombre de chemins et probabilités

1.1 Calcul du nombre de chemins

<u>Notation :</u> Dans tout ce qui suit, on appellera *chemin de* (m, a) *à* (n, b) une ligne brisée du type précédent, c'est-à-dire un suite de segments joignant des points successifs du plan $(m_0, a_0), (m_1, a_1), \ldots, (m_r, a_r)$, qui commence au point $(m_0, a_0) = (m, a)$ et se termine au point $(m_r, a_r) = (n, b)$, et tel que pour tout $0 \leq i < r$, on ait $m_{i+1} = m_i + 1$ et $a_{i+1} = a_i + 1$ ou $a_i - 1$.

Propriété 1.1 *Soient (m, a) et (n, b) deux couples d'entiers, avec $n > m$.*
- *Si $n - m$ et $b - a$ n'ont pas même parité, alors il n'existe aucun chemin de (m, a) à (n, b).*
- *Si $|b - a| > n - m$, alors il n'existe aucun chemin de (m, a) à (n, b).*
- *Sinon (si $|b - a| \leq n - m$ et si $n - m$ et $b - a$ ont même parité), alors le nombre de chemins de (m, a) à (n, b) est exactement $\mathrm{C}_{n-m}^{\frac{n-m}{2} + \frac{b-a}{2}}$.*

Démonstration : Les deux premiers points ne posent pas de problème : étant donné un chemin joignant les points $(m_0, a_0), (m_1, a_1), \ldots, (m_r, a_r)$, il est clair que la parité de m_{i+1} est l'inverse de la parité de m_i, tout comme la parité de a_{i+1} est l'inverse de celle de a_i. Si la suite $m_0, m_1, \ldots m_r$ change un nombre pair de fois de parité, il en va de même pour la suite $a_0, a_1, \ldots, a_r$ (idem dans le cas impair), et donc les extrémités (m, a) et (n, b) vérifient la propriété « $n - m$ et $b - a$ ont même parité ».

De même, si $|b - a| > n - m$, alors pour joindre (m, a) à (n, b), il faudrait au moins $|b - a|$ montées (ou descentes), en seulement $n - m$ étapes.

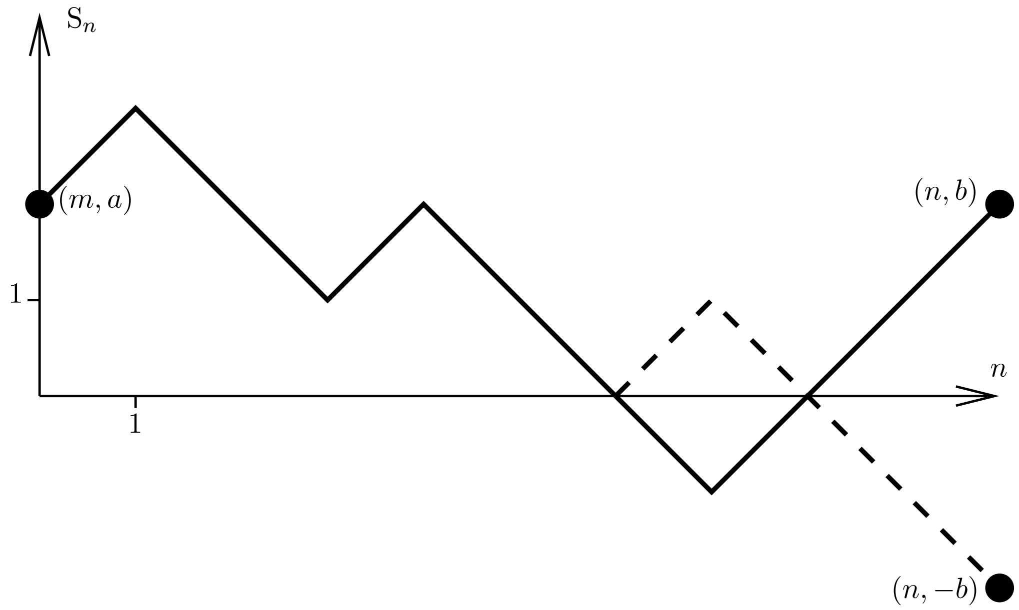
Enfin, si $|b-a| \leq n-m$ et $n-m$ et $b-a$ n'ont pas même parité, alors tout chemin joignant (m,a) à (n,b) est obtenu en « montant » $\dfrac{n-m}{2} + \dfrac{b-a}{2}$ fois et en « descendant » $\dfrac{n-m}{2} - \dfrac{b-a}{2}$.

En effet, si p désigne le nombre de montées, c'est-à-dire le nombre d'indices i tels que $a_{i+1} = a_i + 1$, et d le nombre de descentes, c'est-à-dire le nombre d'indices i tels que $a_{i+1} = a_i - 1$, alors on a $n-m = p+d$ (les $n-m$ pas du chemins étant soit des montées, soit des descentes), et $b-a = p-d$ (la progression totale est égale au nombre de fois où l'on a monté, p, fois la valeur de la montée, 1, plus le nombre de fois où l'on a descendu, d, fois la valeur de la progression dans ce cas, -1). En additionnant membre à membre ces deux égalités, on obtient bien $p = \dfrac{n-m}{2} + \dfrac{b-a}{2}$.

Un chemin joignant (m,a) à (n,b) est alors caractérisé par l'ordre dans lequel on monte et on descend, c'est-à-dire par les indices i où l'on monte. Il s'agit donc de choisir $\dfrac{n-m}{2} + \dfrac{b-a}{2}$ indices parmi $n-m$, soit un nombre de chemins de $\mathrm{C}_{n-m}^{\frac{n-m}{2}+\frac{b-a}{2}}$. $\square$

Théorème 1.2 (Principe de réflexion) *Soient a et b deux entiers strictement positifs, et $m < n$ deux entiers. Alors le nombre de chemins joignant (m,a) à (n,b) et touchant l'axe des abscisses est exactement le nombre de chemins joignant (m,a) à $(n,-b)$.*

Démonstration : La preuve de ce principe est assez simple : elle repose sur l'idée qu'un chemin joignant (m,a) à $(n,-b)$ passe nécessairement par la valeur 0. On peut alors établir une correspondance bijective entre les chemins de ces deux type par symétrie d'une partie de ces chemins (à partir du moment où l'on touche l'axe des abscisses) par rapport à ce même axe.



Plus précisément, soit $(m_0, a_0), (m_1, a_1), \ldots, (m_r, a_r)$ un chemin de (m, a) à (n, b) touchant l'axe des abscisses. Soit j le plus petit indice tel que $a_j = 0$. Alors le chemin $(m_0, a_0), \ldots, (m_j, a_j), (m_{j+1}, -a_{j+1}), \ldots, (m_r, -a_r)$ est un chemin joignant (m, a) à $(n, -b)$ (la vérification est immédiate).

Et cette opération envoie deux chemins distincts sur deux chemins distincts. En effet, si $(m_0, a_0), (m_1, a_1), \ldots, (m_r, a_r)$ et $(m_0, a'_0), (m_1, a'_1), \ldots, (m_r, a'_r)$ sont deux chemins de (m, a) à (n, b) envoyés par notre opération sur le même chemin, alors nos deux chemins de départ touchent l'axe des abscisses pour la première fois au même moment m_j. Ils coïncident nécessairement jusqu'à m_j (ils ont même image par notre transformation), mais aussi par la suite, puisqu'après m_j, ils ont par définition même symétrique par rapport à l'axe des abscisses.

On a donc défini une injection de l'ensemble des chemins de (m, a) à (n, b) touchant l'axe des abscisses dans l'ensemble des chemins de (m, a) à $(n, -b)$, et donc il y a au moins autant de chemins de de (m, a) à $(n, -b)$ que de chemin de (m, a) à (n, b) touchant l'axe des abscisses.

Inversement, la même opération transforme un chemin de (m, a) à $(n, -b)$ en un chemin de (m, a) à (n, b) touchant l'axe des abscisses, toujours de façon injective, et donc on obtient l'égalité souhaitée.

$\square$

1.2 Applications à nos deux problèmes

Problème 1 *Au cours d'un scrutin opposant deux candidats* A *et* B*, le candidat* A *obtient* 600 *voix et le candidat* B 400*. Quelle est la probabilité pour que* A *ait été majoritaire (au sens large) tout au long du dépouillement ?*

Nous allons modéliser le dépouillement par un chemin dans le plan. Lors du processus de dépouillement, on ouvre les enveloppes une à une, chacune contenant un bulletin "A" ou un bulletin "B". Si l'on affecte la valeur 1 aux bulletins "A" et la valeur -1 aux bulletins "B", alors on peut représenter le déroulement du dépouillement par un chemin qui part du point $(0, 0)$, va monter 600 fois et descendre 400 fois, donc qui arrive au point $(1000, 200)$. On considère que tous les ordres possibles d'apparition, au cours du dépouillement, des bulletins "A" et "B", sont équiprobables. D'après la propriété 1.1, le nombre total de dépouillements possibles est C_{1000}^{600}.

Un dépouillement au cours duquel le candidat "A" est majoritaire tout le temps correspond donc à un chemin de $(0, 0)$ à $(1000, 200)$ qui ne passe jamais en-dessous (strictement) de l'axe des abscisses.

Or, nous savons compter le nombre de chemins qui touchent (ou ne touchent pas) l'axe des abscisses, mais pas le nombre des chemins qui traversent (resp. ne traversent pas, mais touchent éventuellement) cet axe. Pour se ramener à la situation précédente, il suffit de tout décaler vers le haut : le nombre de chemins de $(0, 0)$ à $(1000, 200)$ qui ne traversent pas l'axe des abscisses, c'est le nombre de chemins de $(0, 0)$ à $(1000, 200)$ qui ne touchent pas la droite d'équation $y = -1$, ou encore le nombre de chemins de $(0, 1)$ à $(1000, 201)$ qui ne touchent pas l'axe des abscisses ! Et, sous cette dernière forme, nous savons le calculer.

Le nombre de chemins de $(0, 1)$ à $(1000, 201)$ qui ne touchent pas l'axe des abscisses, c'est le nombre total de chemins de $(0, 1)$ à $(1000, 201)$, à savoir C_{1000}^{600}, moins le nombre de chemins de $(0,1)$ à $(1000,201)$ qui touchent l'axe, c'est-à-dire d'après le théorème 1.2 le nombre de chemins de $(0,1)$ à $(1000,-201)$, soit C_{1000}^{399}.

La probabilité cherchée, sous l'hypothèse que tous les dépouillements possibles sont équiprobables, est donc :

$$\frac{\mathrm{C}_{1000}^{600}-\mathrm{C}_{1000}^{399}}{\mathrm{C}_{1000}^{600}} = 1 - \frac{\dfrac{(1000)!}{(399)!(601)!}}{\dfrac{(1000)!}{(400)!(600)!}} = 1 - \frac{400}{601} = \frac{201}{601} \simeq 0,334.$$

Soit environ $33,4\%$ □

Problème 2 100 *personnes font la queue à un guichet de cinéma. La place coûte* 5 *Euros.* 60 *personnes ont un billet de* 5 *Euros, tandis que les* 40 *autres ont des billets de* 10 *Euros. Combien faut-il prévoir de billets de* 5 *Euros en caisse pour qu'avec une probabilité d'au moins* 95%*, tous les spectateurs soient servis, dans l'ordre dans lequel ils se présentent à la caisse ?*

On est ici dans la situation type pour ce qui est de la modélisation : si l'on appelle S_0 le nombre de billets de 5 Euros dans la caisse au moment initial, et que l'on trace la courbe représentative du nombre de billets de 5 Euros dans la caisse en fonction du nombre de spectateurs qui ont déjà acheté leur billet, on obtient exactement un graphe du type voulu. À chaque fois qu'un spectateur passe à la caisse, soit il a un billet de 5 Euros et paye avec, auquel cas le nombre de billets de 5 Euros dans la caisse augmente d'un, soit il n'a qu'un billet de 10 Euros, auquel cas le nombre de billets de 5 Euros dans la caisse diminue d'un, qui correspond à la monnaie qu'on lui rend.

Au total, sur les 100 spectateurs, 60 donnent un billet de 5 Euros et 40 en reçoivent un, et donc on a à la fin S_0+20 billets de 5 Euros. La courbe tracée est donc un chemin de $(0,\mathrm{S}_0)$ à $(100,\mathrm{S}_0+20)$. Pour que tous les spectateurs soient servis, dans l'ordre dans lequel ils se présentent à la caisse, il faut que le chemin ne traverse pas l'axe des abscisses.

Il s'agit donc de trouver le plus petit S_0 pour lequel la probabilité de traverser l'axe des abscisses soit inférieure à 5%. Là aussi, étant donné S_0, la probabilité p_{S_0} pour que le chemin traverse l'axe des abscisses est la probabilité pour qu'un chemin de $(0,\mathrm{S}_0+1)$ à $(100,\mathrm{S}_0+21)$ touche l'axe des abscisses. Grâce au théorème 1.2, c'est le rapport du nombre de chemins de $(0,\mathrm{S}_0+1)$ à $(100,-(\mathrm{S}_0+21))$ au nombre de chemins de $(0,\mathrm{S}_0+1)$ à $(100,\mathrm{S}_0+21)$. On trouve :

$$p_{\mathrm{S}_0} = \frac{\mathrm{C}_{100}^{50-(\mathrm{S}_0+11)}}{\mathrm{C}_{100}^{60}} = \frac{(60)!(40)!}{(39-\mathrm{S}_0)!(61+\mathrm{S}_0)!} = \frac{\prod\limits_{k=40-\mathrm{S}_0}^{40} k}{\prod\limits_{k=61}^{61+\mathrm{S}_0} k}.$$

Calculons les valeurs successives de p_{S_0} lorsque S_0 varie. On trouve les valeurs approchées suivantes :

S_0	0	1	2	3	4	5
p_{S_0}	0,656	0,412	0,249	0,144	0,080	0,042

Il faut donc 5 billets au départ dans la caisse pour qu'avec une probabilité d'au moins 95%, tous les spectateurs soient servis. $\square$

N.B. La solution « déterministe » est beaucoup moins économique : pour être certain de pouvoir servir tout le monde, il faut envisager le pire, c'est-à-dire le cas où les 40 personnes qui n'ont pas de billet de 5 Euros passent en premier. Il faudra donc au départ 40 billets de 5 Euros dans la caisse pour être certain de servir tout le monde.

2 Premier retour en 0, comportement asymptotique.

Dans nos deux premiers problèmes, on s'est ramené à l'étude d'une marche aléatoire dont on connaissait à la fois le début et la fin. Il arrive également que se posent des problèmes pour lesquels on ne connaît pas l'issue... Ainsi, on peut se demander de manière générale comment va se comporter asymptotiquement une marche aléatoire sur $\mathbb{Z}$. A-t'elle tendance à revenir systématiquement en 0 ? Prend-elle des valeurs arbitrairement grandes ?

Le cas d'une marche aléatoire non symétrique ($\mathbb{P}(\mathrm{X}_1 = 1) = p \neq 1/2$) est plus simple, puisqu'alors la loi (forte) des grands nombres nous renseigne sur ce comportement asymptotique. En effet, l'espérance m de la variable aléatoire X_1 est alors non nulle, et l'on a S_n/n qui tend vers m presque sûrement (*i.e.* avec une probabilité égale à 1). Selon le signe de m, on a donc S_n qui tend presque sûrement vers plus l'infini ou vers moins l'infini.

C'est pourquoi nous allons ici nous intéresser au *cas limite*, c'est-à-dire au cas symétrique, dont nous allons étudier le comportement asymptotique.

2.1 Premier retour en 0

On se place ici dans le cadre d'une marche aléatoire symétrique partant de du point 0 (c'est-à-dire que l'on prend $\mathrm{S}_0 = 0$.) Tout d'abord, d'après la propriété 1.1, S_n a la même parité que n. C'est-à-dire que les retours éventuels en 0 ne peuvent se produire qu'au temps pairs.

Lemme 2.1 *La probabilité pour que* S_{2n} *soit nul est* $\mathbb{P}(\mathrm{S}_{2n} = 0) = \dfrac{\mathrm{C}_{2n}^n}{2^{2n}}$.

Démonstration : En effet, le nombre de chemins de $(0,0)$ à $(2n,0)$ est exactement C_{2n}^n, puisqu'il correspond au nombre de façon de choisir les n montées (et donc les n descentes). Comme il y a par ailleurs 2^{2n} trajectoires possibles de longueur $2n$ (2 choix possibles à chaque étape), et que toutes les trajectoires sont équiprobables, on a bien :

$$\mathbb{P}(\mathrm{S}_{2n} = 0) = \frac{\mathrm{C}_{2n}^n}{2^{2n}}.$$

$\square$

Définition On définit l'instant de premier retour en 0 comme :

$$\mathrm{T}_0 = \begin{cases} \inf\{n \geq 1, \mathrm{S}_n = 0\} & \text{si cet ensemble est non vide,} \\ +\infty & \text{sinon.} \end{cases}$$

(T_0, s'il est fini, est nécessairement pair d'après la propriété 1.1.)

Grâce au principe de réflexion, nous allons pouvoir calculer la probabilité, en fonction de n, pour que l'on ait $\mathrm{T}_0 = 2n$:

Théorème 2.2 *La probabilité pour que le premier retour en* 0 *ait lieu à l'instant* $2n$ *est :*

$$\mathbb{P}\left(\mathrm{T}_0 = 2n\right) = \frac{1}{2^{2n-1}}\left(\mathrm{C}_{2n-2}^{n-1} - \mathrm{C}_{2n-2}^{n-2}\right) = \frac{(2n-2)!}{2^{2n-1}n!(n-1)!}\ .$$

Démonstration : Nous allons distinguer deux cas, symétriques, selon que l'on a $\mathrm{S}_1 = 1$ ou $\mathrm{S}_1 = -1$. Si $\mathrm{S}_1 = 1$, alors il s'agit de calculer le nombre de chemins de $(1,1)$ à $(2n,0)$ qui ne touchent pas l'axe des abscisses avant l'étape $2n$. Mais le dernier pas est alors imposé : on a nécessairement $\mathrm{S}_{2n-1} = 1$, puisque l'on cherche des chemins qui aboutissent à la valeur 0 sans avoir jamais touché l'axe des abscisses. Il s'agit donc de dénombrer le nombre de chemins de $(1,1)$ à $(2n-1,1)$ ne touchant pas l'axe des abscisses, c'est-à-dire le nombre total de chemins moins le nombre de chemins qui touchent l'axe. D'après le principe de réflexion, ce dernier est égal au nombre de chemins de $(1,1)$ à $(2n-1,-1)$. Au total, on trouve :

$$\mathrm{C}_{2n-2}^{n-1} - \mathrm{C}_{2n-2}^{n-2}\ .$$

Conditionnellement à l'hypothèse $\mathrm{S}_1 = 1$, on trouve donc une probabilité pour que $\mathrm{T}_0 = 2n$ de $\frac{\mathrm{C}_{2n-2}^{n-1} - \mathrm{C}_{2n-2}^{n-2}}{2^{2n-2}} \times \frac{1}{2}$. Le premier terme du produit est la probabilité pour que la marche aléatoire passe de la valeur $\mathrm{S}_1 = 1$ à la valeur $\mathrm{S}_{2n-1} = 1$ sans toucher l'axe (rapport du nombre de trajectoires favorables sur le nombre total de possibilités), et le facteur $1/2$ correspond à la probabilité pour que S_{2n} soit nul lorsque l'on a $\mathrm{S}_{2n-1} = 1$ (il faut que le dernier pas soit de -1). Lorsque $\mathrm{S}_1 = -1$, la situation est totalement symétrique, donc on trouve exactement la même probabilité. Au total, on a trouvé :

$$\mathbb{P}\left(\mathrm{T}_0 = 2n\right) = \frac{\mathrm{C}_{2n-2}^{n-1} - \mathrm{C}_{2n-2}^{n-2}}{2^{2n-1}}.$$

En revenant aux expressions définissant les nombres, C_n^k, on obtient :

$$\begin{aligned}
\mathbb{P}\left(\mathrm{T}_0 = 2n\right) &= \frac{1}{2^{2n-1}}\left(\frac{(2n-2)!}{(n-1)!(n-1)!} - \frac{(2n-2)!}{n!(n-2)!}\right)\\
&= \frac{(2n-2)!}{2^{2n-1}n!(n-1)!}\left(n-(n-1)\right)\\
&= \frac{(2n-2)!}{2^{2n-1}n!(n-1)!}\ .
\end{aligned}$$

□

On remarque que cette dernière expression nous permet d'obtenir la relation $\mathbb{P}\left(\mathrm{T}_0 = 2n\right) = \mathbb{P}\left(\mathrm{S}_{2n-2} = 0\right) - \mathbb{P}\left(\mathrm{S}_{2n} = 0\right)$. En effet, on a :

$$\begin{aligned}
\mathbb{P}\left(\mathrm{S}_{2n-2} = 0\right) - \mathbb{P}\left(\mathrm{S}_{2n} = 0\right) &= \frac{\mathrm{C}_{2n-2}^{n-1}}{2^{2n-2}} - \frac{\mathrm{C}_{2n}^{n}}{2^{2n}}\\
&= \frac{1}{2^{2n}}\left(4\frac{(2n-2)!}{(n-1)!(n-1)!} - \frac{(2n)!}{n!n!}\right)\\
&= \frac{(2n-2)!}{2^{2n}n!n!}\left(4n^2 - 2n(2n-1)\right)\\
&= \frac{(2n-2)!}{2^{2n-1}n!(n-1)!}
\end{aligned}$$

donc
$$\mathbb{P}\left(\mathrm{T}_0 = 2n\right) = \mathbb{P}\left(\mathrm{S}_{2n-2} = 0\right) - \mathbb{P}\left(\mathrm{S}_{2n} = 0\right)\ .$$

On peut désormais aisément sommer les probabilités $\mathbb{P}(\mathrm{T}_0 = 2n)$, pour obtenir :

Théorème 2.3 *On a* $\mathbb{P}(\mathrm{T}_0 < +\infty) = 1$, *c'est-à-dire que la marche aléatoire revient en* 0 *avec une probabilité* 1.

Démonstration : En effet, la probabilité $\mathbb{P}(\mathrm{T}_0 < +\infty)$ de retour en 0 n'est rien d'autre que la somme $\sum_{n=1}^{+\infty} \mathbb{P}(\mathrm{T}_0 = 2n)$. Et cette dernière somme est télescopique : on a pour tout entier m :

$$\begin{aligned}\sum_{n=1}^{m} \mathbb{P}(\mathrm{T}_0 = 2n) &= \sum_{n=1}^{m} \mathbb{P}(\mathrm{S}_{2n-2} = 0) - \mathbb{P}(\mathrm{S}_{2n} = 0)\\ &= \mathbb{P}(\mathrm{S}_0 = 0) - \mathbb{P}(\mathrm{S}_{2m} = 0)\\ &= 1 - \frac{\mathrm{C}_{2m}^{m}}{2^{2m}}\ .\end{aligned}$$

Or, lorsque m tend vers $+\infty$, le rapport $\frac{\mathrm{C}_{2m}^{m}}{2^{2m}}$ tend lui vers 0. En effet, grâce à la formule de Stirling ($n! \simeq n^n e^{-n}\sqrt{2\pi n}$), on trouve pour équivalent de ce rapport :

$$\frac{\mathrm{C}_{2m}^{m}}{2^{2m}} = \frac{(2m)!}{2^{2m}(m!)^2} \simeq \frac{(2m)^{2m}e^{-2m}\sqrt{4\pi m}}{2^{2m}\left(m^m e^{-m}\sqrt{2\pi m}\right)^2} = \frac{1}{\sqrt{\pi m}}\ .$$

En particulier, notre rapport tend vers 0, donc en passant à la limite dans l'égalité $\sum_{n=1}^{m} \mathbb{P}(\mathrm{T}_0 = 2n) = 1 - \frac{\mathrm{C}_{2m}^{m}}{2^{2m}}$, on obtient :

$$\sum_{n=1}^{+\infty} \mathbb{P}(\mathrm{T}_0 = 2n) = 1.$$

Ceci signifie qu'avec une probabilité égale à 1, la marche aléatoire va revenir en 0 en un temps fini. □

N.B. En réalité, notre marche aléatoire revient en 0 non pas une fois mais une infinité de fois, ce avec une probabilité encore égale à 1. Ceci est une conséquence de la *propriété de Markov* : une fois revenue en 0, la marche aléatoire se comporte comme si elle recommençait au début. Elle va donc, avec une probabilité égale à 1, revenir une deuxième fois en 0, et ainsi de suite...

2.2 Tous les entiers sont atteints

Notre marche aléatoire symétrique revient donc en 0, ce avec une probabilité égale à 1. Ceci dit, intuitivement, le point 0 n'a pas grand chose de particulier dans notre problème, si ce n'est que c'est le point de départ de la marche aléatoire. Il est naturel de se demander si les autres points de $\mathbb{Z}$ sont visités. Nous allons voir que c'est effectivement le cas.

Définition Soit a un entier. On définit l'instant de premier passage en a comme :

$$\mathrm{T}_a = \begin{cases} \inf\{n \geq 1, \mathrm{S}_n = a\} & \text{si cet ensemble est non vide,}\\ +\infty & \text{sinon.}\end{cases}$$

(De la même façon que pour le retour en 0, la propriété 1.1 nous impose une condition de parité sur T_a, si celui-ci est fini.)

Théorème 2.4 *Quel que soit l'entier a, T_a est fini presque sûrement (i.e. avec une probabilité égale à 1, l'entier a est visité par la marche aléatoire).*

Démonstration : Soit donc a un entier non nul (le cas $a = 0$ ayant déjà été réglé). Nous sommes ramenés à montrer que $\mathbb{P}(\mathrm{T}_a = +\infty) = 0$. Or l'événement $\mathrm{T}_a = +\infty$ peut aussi s'exprimer par le fait que a n'est jamais atteint, c'est-à-dire sous la forme $\forall n, \mathrm{S}_n \neq a$.

Nous allons utiliser le fait que T_0 est fini presque sûrement pour décomposer cet événement :

$$\begin{aligned}\mathbb{P}(\mathrm{T}_a = +\infty) &= \mathbb{P}(\forall n, \mathrm{S}_n \neq a) \\ &= \sum_{k=1}^{+\infty} \mathbb{P}(\forall n, \mathrm{S}_n \neq a)\, \mathbb{P}(\mathrm{T}_0 = k) \\ &= \sum_{k=1}^{+\infty} \mathbb{P}(\forall n, \mathrm{S}_n \neq a \text{ et } \mathrm{T}_0 = k)\ .\end{aligned}$$

Or l'événement « $\forall n, \mathrm{S}_n \neq a$ et $\mathrm{T}_0 = k$ » peut quant à lui se reformuler sous la forme

$$« (\mathrm{S}_1 \neq a \text{ et } \mathrm{S}_1 \neq 0, \ldots, \mathrm{S}_{k-1} \neq a \text{ et } \mathrm{S}_{k-1} \neq 0, \mathrm{S}_k = 0, \mathrm{S}_{k+1} \neq a, \ldots) »$$

Autrement dit, si l'on appelle A l'événement :

$$« (\mathrm{S}_1 \neq a \text{ et } \mathrm{S}_1 \neq 0, \ldots, \mathrm{S}_{k-1} \neq a \text{ et } \mathrm{S}_{k-1} \neq 0, \mathrm{S}_k = 0) »$$

et B l'événement : « $\forall n \geq k+1, \mathrm{S}_n \neq a$ »

la probabilité recherchée est $\mathbb{P}(\mathrm{A} \text{ et } \mathrm{B})$, c'est-à-dire $\mathbb{P}(\mathrm{B}|\mathrm{A}) \cdot \mathbb{P}(\mathrm{A})$, où $\mathbb{P}(\mathrm{B}|\mathrm{A})$ désigne la probabilité conditionnelle de B sachant A.

Mais, une fois que l'on sait que $\mathrm{S}_k = 0$, la suite de la marche aléatoire se comporte comme une autre marche aléatoire symétrique partant de 0. Donc notre probabilité conditionnelle $\mathbb{P}(\mathrm{B}|\mathrm{A})$ n'est rien d'autre que la probabilité pour qu'une marche aléatoire symétrique partant de 0 n'atteigne jamais la valeur a, soit $\mathbb{P}(\mathrm{T}_a = +\infty)$.

Et la probabilité $\mathbb{P}(\mathrm{A}) = \mathbb{P}(\mathrm{S}_1 \neq a \text{ et } \mathrm{S}_1 \neq 0, \ldots, \mathrm{S}_{k-1} \neq a \text{ et } \mathrm{S}_{k-1} \neq 0, \mathrm{S}_k = 0)$ peut quant à elle s'exprimer comme la probabilité pour que $\mathrm{T}_0 = k$ et $\mathrm{T}_a > \mathrm{T}_0$. On obtient :

$$\mathbb{P}(\mathrm{T}_a = +\infty) = \sum_{k=1}^{+\infty} \mathbb{P}(\mathrm{T}_a > \mathrm{T}_0 = k) \cdot \mathbb{P}(\mathrm{T}_a = +\infty).$$

Enfin, la somme $\sum_{k=1}^{+\infty} \mathbb{P}(\mathrm{T}_a > \mathrm{T}_0 = k)$ n'est rien d'autre que la probabilité pour que l'on ait $\mathrm{T}_a > \mathrm{T}_0$, puisque l'on sait que T_0 est fini presque sûrement.

D'où $$\mathbb{P}(\mathrm{T}_a = +\infty) = \mathbb{P}(\mathrm{T}_a > \mathrm{T}_0) \cdot \mathbb{P}(\mathrm{T}_a = +\infty).$$

Or la probabilité pour que T_a soit strictement supérieur à T_0 ne peut être égale à 1 : si les $|a|$ premiers pas sont en direction de a (cas où, au début, la marche aléatoire va directement jusqu'à la valeur a), alors on a nécessairement $\mathrm{T}_a < \mathrm{T}_0$. Et cette situation arrive avec une probabilité strictement positive, égale à $1/2^{|a|}$. On en déduit que :

$$\mathbb{P}(\mathrm{T}_a = +\infty) = 0. \qquad \square$$

N.B. L'idée qui soustends ce résultat est assez naturelle. La probabilité d'atteindre a avant de revenir en 0 est non nulle (et même minorée par une constante strictement positive, par exemple $1/2^{|a|}$). Or on est sûr de revenir en 0. Et, une fois revenu en 0, la probabilité d'atteindre a avant de revenir en 0 est encore la même, et ainsi de suite. Au total, on a donc un événement de probabilité strictement positive, et l'on répète l'expérience une infinité de fois. Cet événement est en fin de compte certain !

2.3 Marche aléatoire bornée, modèle du joueur

Problème 3 *Une personne dispose d'un capital de départ c (c entier). Elle joue à pile ou face de manière répété une somme de 1, jusqu'à atteindre un but b qu'elle s'est fixée au départ, ou bien jusqu'à ce qu'elle ne puisse plus jouer (lorsque sa fortune atteint la valeur 0). Quelle est la probabilité pour qu'elle atteigne son objectif b ?*

Là aussi, la fortune du joueur en fonction du temps peut être représentée par une marche aléatoire $(\mathrm{S}_n)_{n\in\mathbb{N}}$ symétrique (les variables X_i correspondent aux piles ou face successifs, avec $\mathbb{P}(\mathrm{X}_i = -1) = \mathbb{P}(\mathrm{X}_i = 1) = 1/2$ pour tout i), dont la valeur initiale est $\mathrm{S}_0 = c$. Cette marche aléatoire est en outre bornée : si S_n prend la valeur 0 ou b, le joueur s'arrête. Or, on a vu qu'une marche aléatoire symétrique sur $\mathbb{Z}$ atteint n'importe quelle valeur en un temps fini. En particulier, notre joueur ne jouera qu'un nombre fini de fois, ce avec une probabilité 1. La question de savoir quelle est sa chance de succès a donc toute sa pertinence.

Notons p_c la probabilité de succès de notre joueur lorsque son capital de départ est c. Alors bien sûr $p_0 = 0$ (si le joueur n'a rien à miser, il ne peut gagner) et $p_b = 1$ (si le joueur a atteint son plafond avant de jouer, il ne joue pas et il a gagné...).

Soit maintenant c un capital de départ strictement compris entre 0 et b. Nous allons montrer que $p_c = (1/2)p_{c-1} + (1/2)p_{c+1}$ (c'est une propriété dite d'« harmonicité »). Si G_c désigne l'événement :

« le joueur gagne avec une mise de départ de c »

alors on peut décomposer G_c en deux événement disjoint qui sont « G_c et $\mathrm{X}_1 = 1$ » et « G_c et $\mathrm{X}_1 = -1$ ». On a donc :

$$p_c = \mathbb{P}(\mathrm{G}_c) = \mathbb{P}(\mathrm{G}_c \text{ et } \mathrm{X}_1 = 1) + \mathbb{P}(\mathrm{G}_c \text{ et } \mathrm{X}_1 = -1).$$

Mais on a $\mathbb{P}(\mathrm{G}_c \text{ et } \mathrm{X}_1 = 1) = \mathbb{P}(\mathrm{G}_c \mid \mathrm{X}_1 = 1) \cdot \mathbb{P}(\mathrm{X}_1 = 1)$, où $\mathbb{P}(\mathrm{G}_c \mid \mathrm{X}_1 = 1)$ désigne la probabilité conditionnelle pour que l'événement G_c ait lieu sachant $\mathrm{X}_1 = 1$. Or pour le joueur, gagner sachant qu'il a gagné au premier coup ($\mathrm{X}_1 = 1$), c'est exactement équivalent à gagner avec un capital de départ de $c+1$ (la suite du jeu étant indépendante du premier coup). C'est-à-dire que la probabilité $\mathbb{P}(\mathrm{G}_c \mid \mathrm{X}_1 = 1)$ est égale à p_{c+1}. De la même façon, on trouve que $\mathbb{P}(\mathrm{G}_c \mid \mathrm{X}_1 = -1)$ est égale à p_{c-1}.

D'où
$$p_c = (1/2)p_{c-1} + (1/2)p_{c+1}.$$

La suite $(p_n)_{n\in\mathbb{N}}$ vérifie donc une relation de récurrence d'ordre 2 donnée par $p_{n+2} = 2p_{n+1} - p_n$. L'équation caractéristique est $r^2 - 2r + 1 = 0$, avec 1 comme racine double. La suite p_n est donc de terme général $p_n = \mathrm{A} + \mathrm{B}n$, où A et B sont deux constantes. On détermine ces constante grâce aux conditions $p_0 = 0$ et $p_b = 1$: on trouve $\mathrm{A} = 0$ puis $\mathrm{B} = 1/b$.

On a montré : $$p_c = c/b. \qquad \square$$

N.B. Ce modèle est bien sûr assez simpliste : il suppose que la mise du joueur est toujours la même, et que le jeu est à espérance nulle ($\mathbb{P}(\mathrm{X}_i = 1) = 1/2$). Si l'on abandonne cette propriété, on trouve une relation de récurrence du type $p_n = pp_{n+1} + (1-p)p_{n-1}$, où p est la probabilité de gain à chaque coup (par exemple, $p = 18/37$ à la roulette, lorsque l'on joue pair ou impair, etc). Par la même méthode si $p \neq 1/2$, on trouve $p_c = ((1-1/p)^c - 1) \,/\, \big((1-1/p)^b - 1\big)$.

2.4 Récurrence

Nous avons vu qu'avec une probabilité 1, notre marche aléatoire symétrique revient en 0. En réalité, ceci implique qu'elle revienne en 0 une infinité de fois, ce avec une probabilité égale à 1, toujours. Ceci est une conséquence de la *propriété de Markov* : une fois revenue en 0, la marche aléatoire se comporte comme si elle recommençait au début. Elle va donc, avec une probabilité égale à 1, revenir une deuxième fois en 0, et ainsi de suite...

Cette propriété de retour en 0 une infinité de fois s'appelle la « récurrence ». Une position est *récurrente* si, avec une probabilité égale à 1, une marche aléatoire va passer une infinité de fois par cette position. Une propriété remarquable des marches aléatoires (symétriques) est qu'en dimension 1 et 2, toutes les positions sont récurrentes. En revanche, en dimension 3 (et au-delà), une marche aléatoire sur $\mathbb{Z}^3$ ne reviendra au point de départ qu'un nombre fini de fois, ce avec une probabilité 1. Une telle position est qualifiée de *transitoire* ou *transiente.*

Théorème 2.5 *On a* $\mathbb{P}(\sup \mathrm{S}_n = +\infty) = \mathbb{P}(\inf \mathrm{S}_n = -\infty) = 1$.

Autrement dit, avec des probabilités égales à 1, la marche aléatoire va prendre des valeurs arbitrairement grandes (resp. infiniment petites).

Démonstration : On a déjà presque tout dit, puisque l'on sait que tout entier est visité avec une probabilité égale à 1, c'est-à-dire que la probabilité $\mathbb{P}(\exists k, \mathrm{S}_k = n)$ est égale à 1. Mais, si la marche aléatoire atteint l'entier n, on a alors forcément $\sup\limits_{k>0} \mathrm{S}_k \geq n$.

On a donc $$\mathbb{P}\left(\sup_{k\geq 0} \mathrm{S}_k \geq n\right) = 1.$$

Reste à passer à l'intersection (sur n) de tous ces événements. Et la conjonction d'une nombre dénombrables d'événements de probabilité 1 est encore de probabilité 1. (En passant au complémentaire, ceci signifie que la réunion d'un ensemble dénombrable d'événements de probabilité nulle est encore de probabilité nulle, ce qui est une propriété intrinsèque de toute mesure de probabilité...) Autrement dit, on a :

$$\mathbb{P}\left(\forall n, \sup_{k\geq 0} \mathrm{S}_k \geq n\right) = 1.$$

On a montré $$\mathbb{P}\left(\sup_{n\geq 0} \mathrm{S}_n = +\infty\right) = 1.$$

Par symétrie $$\mathbb{P}\left(\inf_{n\geq 0} \mathrm{S}_n = -\infty\right) = 1. \qquad \square$$

N.B. Dans le cas d'une marche aléatoire non symétrique (pour $\mathbb{P}(\mathrm{X}_1 = 1) = p \neq 1/2$), la loi (forte) des grands nombres nous donne immédiatement le fait que S_n/n tend vers m presque sûrement, où m est l'espérance de la variable aléatoire X_1 ($m = 2p - 1 \neq 0$). Selon le signe de m, on en déduit que S_n tend presque sûrement vers plus ou moins l'infini.

Théorème 2.6 *Tout entier a est visité une infinité de fois.*

Démonstration : L'idée essentielle de cette démonstration est que, pour satisfaire aux deux conditions $\sup \mathrm{S}_n = +\infty$ et $\inf \mathrm{S}_n = -\infty$, la suite $(\mathrm{S}_n)_{n\in\mathbb{N}}$ doit faire une infinité de va-et-vients entre des valeurs positives et négatives de plus en plus grande. Pour ce faire, elle devra alors passer par toutes les valeurs intermédiaires, ce une infinité de fois !

Soit donc a un entier, supposons que a n'est visité qu'un nombre fini de fois. Alors il existe un entier n_0 tel que pour tout $n \geq n_0$, on ait $\mathrm{S}_n \neq a$. Mais alors, après n_0, toutes les valeurs de la suite $(\mathrm{S}_n)_{n\geq n_0}$ sont nécessairement « du même côté » de a, c'est-à-dire qu'on ne peut avoir deux entiers n_1 et n_2 tels que $\mathrm{S}_{n_1} < a$ et $\mathrm{S}_{n_2} > a$: en effet, pour passer d'une de ces deux valeurs (la première par ordre d'apparition dans la suite $(\mathrm{S}_n)_{n\in\mathbb{N}}$ à la seconde, il nous faudrait nécessairement passer par la valeur a.

Supposons alors, pour simplifier, que l'on ait pour tout $n \geq n_0$, $\mathrm{S}_n > a$ (l'autre situation est strictement équivalente). Alors la suite $(\mathrm{S}_n)_{n\in\mathbb{N}}$ est minorée par la valeur $\min(\mathrm{S}_0, \mathrm{S}_1, \ldots, \mathrm{S}_{n_0}, a)$. Mais ceci infirme la propriété $\inf \mathrm{S}_n = -\infty$.

Par contraposée, si l'on a $\sup \mathrm{S}_n = +\infty$ et $\inf \mathrm{S}_n = -\infty$, alors la suite $(\mathrm{S}_n)_{n\in\mathbb{N}}$ passe une infinité de fois par la valeur a. Comme cette condition a une probabilité égale à 1, on a montre qu'avec une probabilité égale à 1, la suite passe une infinité de fois par chaque entier.

$\square$