



## **Paradoxes probabilistes**

*- Les lois de Khintchine -*

par Bernard Beauzamy

Version révisée : mai 2016

### **Introduction**

La Théorie des Probabilités a environ 360 ans : la correspondance entre Pierre de Fermat et Blaise Pascal en 1654 en marque les débuts formels. Malgré cette ancienneté, elle a du mal à pénétrer convenablement les communautés scientifiques.

On se souvient de la remarque d'Albert Einstein à propos de la mécanique quantique "Dieu ne joue pas aux dés". Dans le même ordre d'idée, René Thom qualifiait cette même mécanique quantique de "plus grande escroquerie intellectuelle du 20<sup>ème</sup> siècle", parce que l'on n'en comprenait pas les fondements probabilistes.

Les probabilités sont un peu enseignées dans le secondaire ; il a fallu attendre 2013 pour qu'elles figurent au programme des classes préparatoires aux grandes écoles scientifiques. Les écoles d'ingénieurs ont peu d'enseignements de probabilités : il nous est arrivé de refuser des candidats à un emploi à la SCM pour cette raison, et il nous est arrivé d'écrire aux responsables des enseignements de certaines écoles.

Pourtant, les probabilités ont une utilité évidente : elles sont destinées à rendre compte de la variabilité des phénomènes ; à ce jour, c'est le meilleur moyen et en vérité le seul ! Ces variabilités sont de plusieurs natures :

- Les conditions où se déroulent les process sont variables : c'est le cas de la température ambiante, de la pression atmosphérique, etc.
- Il arrive de temps en temps des phénomènes non souhaités, mais à prendre en compte, comme des séismes, des tornades, etc.

- Le réglage des process lui-même n'est jamais parfaitement constant : ce peut être la qualité des composants chimiques, les températures de fabrication, etc.

En résumé, toute activité humaine, tout process de fabrication industrielle, présentent une certaine variabilité, qu'on le veuille ou non.

L'attitude prise par la plupart des ingénieurs consiste à vouloir réduire cette variabilité autant que possible. On y parvient dans certains cas : on peut chercher à mieux contrôler un process de fabrication. Mais on ne parviendra jamais à éliminer entièrement toute variabilité. La simple gestion du stock d'un magasin de chaussures implique un raisonnement probabiliste : on ne sait jamais à l'avance la pointure du client qui entrera.

Pour décrire un phénomène complexe, les ingénieurs aiment bien se doter de codes de calcul, aussi précis que possible. Et c'est là que la démarche devient fondamentalement "anti-probabiliste", dans son essence même. Car la prise en compte des incertitudes dans un code de calcul est impossible : il retourne un résultat précis à partir de valeurs d'entrée précises. La réponse habituellement apportée par les ingénieurs est "nous faisons plusieurs runs du code". Mais si le code comporte 40 paramètres, chacun avec 10 valeurs possibles, il faudra  $10^{40}$  runs, ce qui est impossible en pratique, quelle que soit la vitesse d'exécution du code. L'idée selon laquelle on pourrait faire ces runs au hasard et les traiter selon la "méthode de Wilks" est fondamentalement erronée, comme nous l'avons expliqué dans l'article [http://www.scmsa.eu/archives/BB\\_Wilks\\_2016\\_01\\_11.pdf](http://www.scmsa.eu/archives/BB_Wilks_2016_01_11.pdf)

Il y a donc, en particulier pour les études de sûreté, la nécessité d'une approche fondamentalement probabiliste, dans la mesure où les conditions de mise en œuvre ne sont jamais connues avec précision.

Nous entendons souvent dire "l'approche des probabilités que vous nous présentez est inapplicable dans la vraie vie. Il n'y a jamais de suites de variables indépendantes et de même loi". Cet argument est entièrement dépourvu de pertinence. Tout d'abord, il existe des énoncés probabilistes, plus sophistiqués, relatifs à des suites dont la loi évolue et présentant une dépendance. Ensuite, une présentation des règles fondamentales est nécessaire : si l'on regarde la loi de la chute des corps, elle est inapplicable sous cette forme, parce qu'il y a la résistance de l'air, la rotation de la Terre, les vents latéraux, etc. Pourtant, il faut d'abord connaître la loi de base de la chute des corps, avant de prendre en compte toutes les perturbations qui peuvent se produire en situation réelle. Il en va de même ici.

Les lois de la physique sont purement axiomatiques : on nous dit "voici la forme de la loi de la gravitation universelle, voici la forme des lois de l'électromagnétisme, etc.". Nous constatons cette forme mais nous ne savons pas l'expliquer ; elle ne résulte pas d'une démonstration.

A l'inverse, et c'est une situation unique, les lois des probabilités sont des lois de la Nature, et ces lois se démontrent, comme nous allons le voir, à partir d'une situation axiomatique très simple. Dieu joue aux dés, et il a introduit les lois découvertes en 1924 par Alexander Khintchine, connues sous le nom de "loi du logarithme itéré". Ces lois sont à la fois très belles et très complexes ; les démonstrations sont difficiles. Il est inutile de se voiler la face : la Nature est ainsi faite et tous les consensus proposant de l'ignorer sont vains.

Les lois de Khintchine partent d'une axiomatique très simple : un jeu de pile ou face, avec gains  $\pm 1$ , probabilité  $1/2$  dans chaque cas, et elle aboutit à un résultat très complexe : le comportement asymptotique des gains lorsque le nombre de parties augmente (que nous présentons ici sous forme quantitative pour la première fois). Pour prendre une comparaison physique, l'équivalent serait une présentation axiomatique des interactions entre atomes, d'où l'on pourrait déduire le mouvement des galaxies. Un tel changement d'échelle n'existe pas en physique.

Beaucoup de gens ont une compréhension fondamentalement erronée des phénomènes probabilistes. Contrairement à ce que croient les ingénieurs et les économistes, la Nature ne recherche jamais un équilibre, mais procède toujours par grandes oscillations, dont l'amplitude va croissant : les lois de Khintchine expriment ceci dans le cas le plus simple. Nous montrons les conséquences de telles lois sur les applications industrielles, économiques et sociales.

# Première Partie

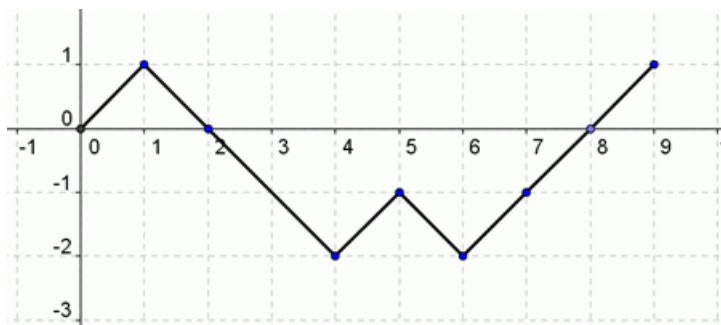
## Présentation des résultats

### et applications

# I. Présentation du problème

## A. Eléments de base

On considère un jeu de pile ou face : à l'instant  $k$ , une variable aléatoire  $X_k$  vaut  $-1$  ou  $+1$  (gain de l'un des joueurs lors de la  $k^{\text{ème}}$  partie). On note  $S_n = \sum_{k=1}^n X_k$  : c'est le gain algébrique total de l'un des joueurs : il peut être positif ou négatif. Le graphe de  $S_n$  (temps en abscisse, valeur du gain en ordonnée) représente une "marche aléatoire" dans le plan. Voici un exemple sur 9 coups :



La marche aléatoire part de l'origine. On s'intéresse à son "comportement asymptotique" lorsque le nombre de parties augmente. Nous renvoyons au livre [MPPR] pour les énoncés qualitatifs ; nous allons donner ici des énoncés quantitatifs à propos du comportement des  $S_n$ .

Ces énoncés sont de deux types :

- Les sommes  $S_n$  repassent souvent par la valeur 0 : il y a souvent égalité entre les joueurs. C'est ce que nous appellerons le rapprochement ;
- Les oscillations des  $S_n$  sont de plus en plus accentuées ; contrairement à ce que l'on croit souvent, il est faux que  $S_n \rightarrow 0$  lorsque  $n \rightarrow +\infty$ . C'est ce que nous appellerons l'éloignement.

La loi de chaque  $S_n$  en elle-même est très simple :  $S_n$  prend ses valeurs entre  $-n$  et  $n$  ; si au cours de  $n$  parties il y a eu  $k$  "Pile" et  $n-k$  "Face", la valeur de  $S_n$  est  $k - (n-k) = 2k - n$ ,  $k = 0, \dots, n$ . Pour chaque  $n$ , la loi de  $S_n$  est donc ceci :  $S_n$  prend la valeur  $2k - n$  avec probabilité

$$\frac{1}{2^n} \binom{n}{k}, \quad k = 0, \dots, n. \quad \text{Ou bien, si l'on préfère, } P(S_n = j) = \frac{1}{2^n} \binom{n}{\frac{n+j}{2}} \quad j = -n, \dots, n \text{ et } n+j \text{ pair.}$$

Par conséquent, pour tout  $c$  :

$$P(S_n \geq c) = \frac{1}{2^n} \sum_{k \geq \frac{n+c}{2}} \binom{n}{k}$$

puisque la condition  $2k - n \geq c$  équivaut à  $k \geq \frac{n+c}{2}$

### B. Les résultats

Ils consistent en cinq théorèmes ; les trois premiers sont simples. Le premier est conforme à l'intuition : il exprime que l'égalité des gains est fréquente.

**Théorème 1.** - *La probabilité que la première égalité se produise au-delà de  $2n$  vaut :*

$$g_n = \frac{1}{4^n} \binom{2n}{n} \approx \frac{1}{\sqrt{\pi n}}$$

Elle décroît donc lentement à mesure que  $n$  augmente. Un exemple est donné dans le tableau suivant :

nb de coups	10	20	50	100	200
proba pas d'égalité	0,18	0,13	0,08	0,06	0,04

Ce tableau donne la probabilité que l'on n'ait pas d'égalité de gains pendant la partie, en fonction du nombre de coups joué. Par exemple, si l'on joue 200 fois, il n'y a qu'une probabilité de 0.04 que l'égalité ne se produise jamais pendant ces 200 coups.

Nous allons maintenant évaluer la probabilité d'atteindre un seuil donné au plus tard à un instant donné : c'est la généralisation du Théorème 1.

Pour tout seuil  $s$  (par exemple 1000 Euros), pour tout  $n$ , on pose :

$$\alpha(n, s) = P(\exists k, 1 \leq k \leq n, S_k = s)$$

C'est la probabilité d'atteindre le seuil  $s$  au plus tard à l'instant  $n$ .

**Théorème 2.** – *On a, pour tout  $s \geq 1$  :*

$$\alpha(2n, 2s) = 1 - \frac{1}{4^n} \sum_{j=-s+1}^s \binom{2n}{j}$$

et donc asymptotiquement, pour  $s$  fixé, lorsque  $n \rightarrow +\infty$  :

$$\alpha(2n, 2s) \sim 1 - \frac{2s}{\sqrt{\pi n}}$$

Un autre résultat traduit l'impossibilité du confinement dans une région de largeur croissante ; ceci contredit déjà l'intuition commune.

**Théorème 3.** - Pour tout  $n \geq 1$  :

$$P\left(|S_n| > \frac{\sqrt{n}}{2}\right) \geq \frac{9}{16}.$$

Il est donc très probable que la somme algébrique des gains ne restera pas confinée dans la région  $|y| \leq \frac{\sqrt{x}}{2}$ .

Mais un tel résultat porte sur la valeur absolue du gain, et ne dit pas dans quelle mesure le gain peut grandir. Les résultats les plus précis sont dus à Alexander Khintchine (1924).

On note :

$$\varphi(x) = \sqrt{2x \text{Log}(\text{Log}(x))},$$

qui est une fonction définie pour tout  $x$  réel,  $x > e$ , et que nous appellerons "courbe de Khintchine".

**Théorème 4 (première loi de Khintchine).** - Soit  $\varepsilon > 0$  et :

$$\eta(m, \varepsilon) = \frac{1}{(\text{Log}(1 + \varepsilon))^{1+\varepsilon}} \frac{1}{\varepsilon(m-1)^\varepsilon}$$

Notons :

$$B(m, \varepsilon) = \left\{ \exists n \geq (1 + \varepsilon)^m, S_n > (1 + \varepsilon)\varphi(n) \right\}$$

Alors, pour tout  $m$  :

$$P(B(m, \varepsilon)) \leq \eta(m, \varepsilon)$$

L'ensemble  $B(m, \varepsilon)$  est constitué des chemins qui dépassent la "bande de sécurité"  $(1 + \varepsilon)\varphi$  au moins une fois après l'instant  $(1 + \varepsilon)^m$ . Le Théorème 4 dit que la probabilité d'un tel événement tend vers 0 lorsque  $m$  grandit. Autrement dit, pour une largeur de bande de sécurité fixée, c'est à dire  $\varepsilon > 0$  fixé, il devient de moins en moins probable de passer au dessus de la courbe  $(1 + \varepsilon)\varphi$  à mesure que l'on joue.

Prenons par exemple  $\varepsilon = 1$ . Le Théorème 4 donne l'estimation :

$$P\{\exists n \geq 2^m, S_n > 2\varphi(n)\} \leq \frac{1}{(\text{Log}(2))^2} \frac{1}{m-1}$$

En particulier, si l'on veut que le second membre soit  $\leq 0.05$ , on trouve  $m = 42$  et par conséquent la probabilité d'avoir un  $n \geq 2^{42}$  pour lequel  $S_n > 2\varphi(n)$  est inférieure à 0.05.

Le Théorème 5 (seconde loi de Khintchine) est le plus remarquable de tous. Il dit que, au fil des parties, le gain de chaque joueur viendra à "tangenter" la courbe de Khintchine et ce d'aussi près que l'on veut ; il donne à ce phénomène des estimations quantitatives.

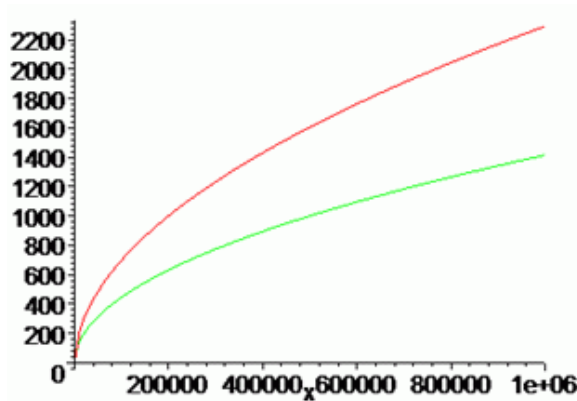
**Théorème 5 (seconde loi de Khintchine).** – Soient  $\varepsilon > 0$ ,  $\alpha > 0$ . Il existe un nombre  $N_0(\varepsilon, \alpha)$  explicitement calculé tel que :

$$P\left(\exists k \leq N_0 \text{ tel que } \frac{S_k}{\varphi(k)} > 1 - \varepsilon\right) \geq 1 - \alpha$$

Si par exemple nous prenons  $\varepsilon = 1/2$  et  $\alpha = 1/2$ , la seconde loi de Khintchine nous dit que, si nous jouons un nombre de parties  $N \geq 3 \times 10^{55}$ , nous avons une chance sur deux que, avant ce moment, nos gains (et nos pertes) à l'instant  $n$  aient atteint la valeur  $\frac{1}{2} \sqrt{2n \text{Log}(\text{Log}(n))}$ .

Les démonstrations des théorèmes sont données dans la Seconde Partie.

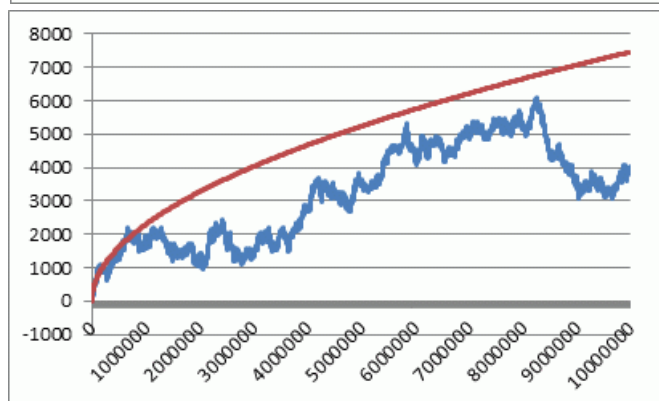
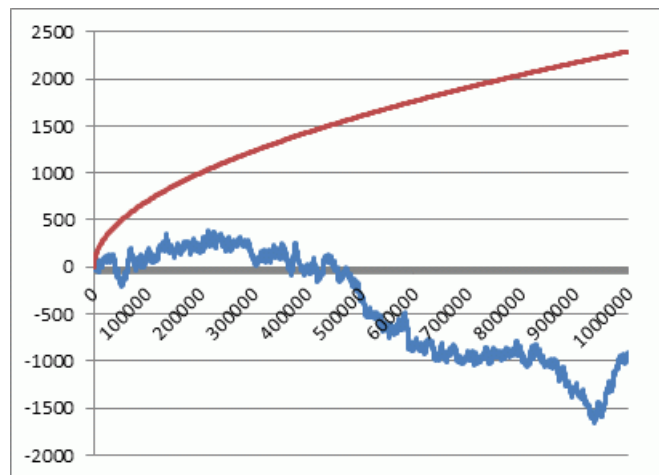
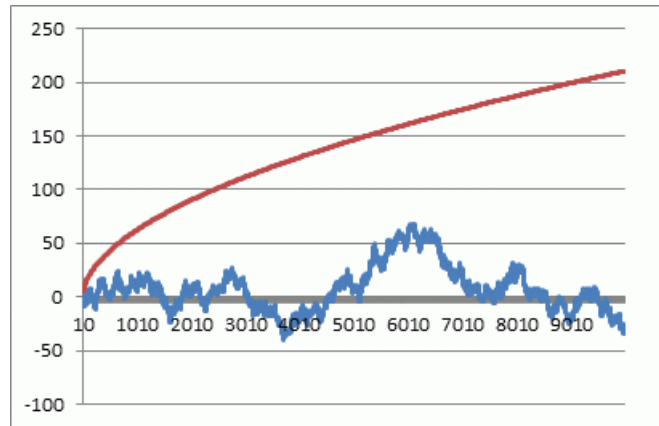
La courbe de Khintchine  $y = \sqrt{2x \text{Log}(\text{Log}(x))}$  est à croissance très proche de celle de  $\sqrt{2x}$  ; voici les graphes entre 10 et  $10^6$  (en vert  $\sqrt{x}$ , en rouge Khintchine) :





### C. Exemples :

Ces exemples sont générés sous Excel. On calcule toutes les valeurs des  $S_n$  mais (pour simplifier l'affichage) on ne fait figurer que les multiples de 1000. La courbe de Khintchine est en rouge.



La programmation de ces résultats en VBA sous Excel est facile ; elle utilise la fonction `Rand()`, générateur de nombres aléatoires. On peut vérifier qu'un générateur de nombres aléatoires fonctionne de manière satisfaisante en comparant les résultats qu'il produit aux énoncés théoriques, sur de longues périodes.

### D. Erreur commise dans les raisonnements

D'où vient l'erreur commise habituellement, qui fait penser que  $S_n \rightarrow 0$  ?

On voit souvent écrire que  $\frac{S_n}{n} = \frac{1}{n} \sum_1^n X_k$  suit une loi de Gauss (ce qui n'est pas vrai pour chaque  $n$  ; ce n'est vrai qu'asymptotiquement), et que par conséquent :

$$P\left\{-a \leq \frac{S_n}{n} \leq a\right\} = \int_{-a}^a \exp\left(\frac{-t^2}{2}\right) \frac{dt}{\sqrt{2\pi}} \quad (1)$$

ce qui est faux, ou encore :

$$P\{-an \leq S_n \leq an\} = \int_{-a}^a \exp\left(\frac{-t^2}{2}\right) \frac{dt}{\sqrt{2\pi}}$$

Posant  $b = an$  :

$$P\left\{-b \leq S_n \leq b\right\} = \int_{\frac{-b}{n}}^{\frac{b}{n}} \exp\left(\frac{-t^2}{2}\right) \frac{dt}{\sqrt{2\pi}} \rightarrow 0 \text{ lorsque } n \rightarrow \infty$$

L'égalité (1) est fautive. Ce qui est correct est le fait que, pour tout  $a$  :

$$\left| P\left\{-a \leq \frac{S_n}{n} \leq a\right\} - \int_{-a}^a \exp\left(\frac{-t^2}{2}\right) \frac{dt}{\sqrt{2\pi}} \right| \rightarrow 0, n \rightarrow \infty$$

mais cette convergence dépend de  $a$ .

## II. Bien comprendre les énoncés

Les énoncés probabilistes sont toujours très difficiles à comprendre ; comme nous le disons dans le livre [MPPR], le vocabulaire généralement utilisé est peu approprié.

Les théorèmes ci-dessus ne sont pas des résultats qui guident un jeu de pile ou face en particulier. Si vous jouez avec votre voisin, il n'y a pas d'autre règle que  $\pm 1$  à chaque coup, avec probabilité  $1/2$ . Sur une partie de  $N$  coups, tous les chemins ont probabilité  $\frac{1}{2^N}$ , y compris celui qui monte constamment et celui qui descend constamment, qui ne sont en rien moins pro-

ables que les autres. Il n'existe aucune loi occulte qui dit que, tôt ou tard, Dieu va guider votre gain vers la courbe de Khintchine décrite plus haut.

Les théorèmes ci-dessus sont en réalité des théorèmes de dénombrement de chemins. Si vous avez joué pendant  $N$  coups, la plupart des  $2^N$  chemins vont venir tangenter la courbe de Khintchine au sens mentionné plus haut. Ou bien, en d'autres termes, si d'innombrables joueurs jouent en même temps à pile ou face, pour la plupart d'entre eux on observera le phénomène.

On remarquera que les démonstrations des trois derniers théorèmes sont purement probabilistes et non combinatoires : on ne passe pas par un dénombrement de chemins, même si le résultat est en définitive un dénombrement.

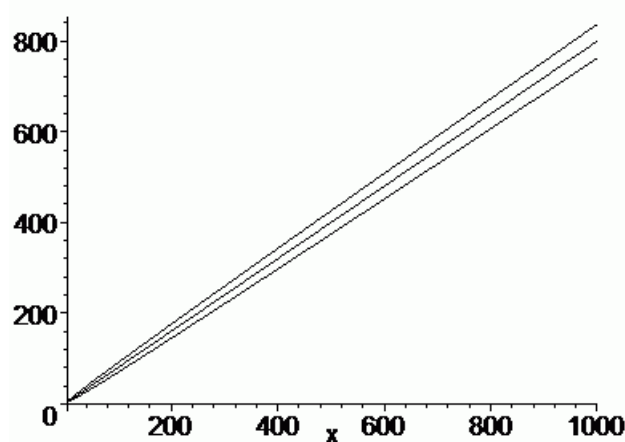
### III. Extensions

Nous ne traitons ici que le cas de variables prenant les valeurs  $\pm 1$  avec probabilités  $1/2$ . La loi du logarithme itéré est valable de manière plus générale : si les  $X_k$  sont des variables indépendantes, de même loi, d'espérance  $\mu$  et d'écart-type  $\sigma$ , la somme  $S_n$  reste entre les bornes :

$$n\mu \pm \sigma\sqrt{2n\text{Log}(\text{Log}(n))}$$

et presque sûrement les bornes sont atteintes, comme précédemment.

Par exemple, pour une v.a. prenant les valeurs  $-1$  (probabilité  $1/10$ ) et  $+1$  (probabilité  $9/10$ ), on a  $\mu = \frac{4}{5}$ ,  $\sigma = \frac{3}{5}$ . Voici la zone où se trouvera la marche aléatoire :



A partir du moment où le jeu n'est pas équilibré ( $\mu \neq 0$ ), à partir d'un certain moment le gain restera toujours dans le demi plan supérieur (cas  $\mu > 0$ ) ou dans le demi-plan inférieur (cas  $\mu < 0$ ). En effet, la fluctuation  $\sigma\sqrt{2n\text{Log}(\text{Log}(n))}$  est petite devant le terme principal  $n\mu$  lorsque  $n \rightarrow +\infty$ .

## IV. L'utilisation du hasard

### A. Généralités

Les Industriels pensent souvent qu'il est bon de restreindre le rôle du hasard. Cette idée n'est pas nécessairement correcte ; pis, elle est en général fausse.

Rappelons (voir le livre [MPPR]) que l'on désigne sous le nom générique de "hasard" toutes les causes, toutes les influences, que l'on ne peut pas connaître ou que l'on ne souhaite pas connaître.

Par exemple, un marchand de chaussures a besoin d'un stock ; pour prévoir le nombre de paires de telle pointure, il considérera que l'entrée d'un client et son choix résultent du hasard. Il n'est pas possible de faire un recensement des pointures aux alentours, sans oublier les touristes.

On peut considérer que le débit d'une rivière résulte du hasard et constituer des lois de probabilité de ce débit en se servant d'un historique. On peut aussi essayer de remonter aux causes, au moyen de modèles "pluies-débit", qui vont associer le débit de la rivière aux pluies pendant les jours précédents, mais de tels modèles sont très difficiles à construire (beaucoup d'autres paramètres interviennent, comme l'état des sols) et encore plus difficiles à valider. Dans le cas présent, l'approche "hasard" est beaucoup plus pertinente.

Un industriel qui fabrique des radiateurs considérera valablement que le dysfonctionnement d'un appareil résulte du hasard et il en recherchera la loi de probabilité : quel pourcentage ? Eventuellement, il peut vouloir des informations complémentaires : selon la saison, selon la région ; ce seront des lois de probabilité conditionnelle. Mais il s'en remet au hasard, en ce sens qu'il ne recherche pas l'explication physique du phénomène qui, dans certains cas, peut être très difficile à trouver.

S'en remettre au hasard signifie donc que l'on renonce à chercher les causes intimes d'un phénomène ; par contre, on va réclamer un historique, de manière à en apprécier les occurrences.

### B. Surveillance d'un échantillon

La situation du problème est la suivante : un industriel veut mettre un produit nouveau sur le marché (par exemple des radiateurs électriques) ; il se donne un "taux de rejet" maximum acceptable, par exemple 1/100. Cela signifie, dans son idée, que s'il vend 10 000 radiateurs, il aura des retours (appels à la garantie) pour environ 100 d'entre eux. L'industriel veut tester un échantillon et, en fonction du nombre d'appareils défectueux dans l'échantillon, prendre une décision : mise sur le marché ou non.

Deux approches sont possibles :

## 1. Approche globale

Il teste tout l'échantillon et fait à la fin le quotient  $\frac{D}{N}$ , où  $D$  est le nombre d'appareils défectueux et  $N$  est la taille de l'échantillon. Si  $\frac{D}{N} \geq \frac{1}{100}$ , le produit est considéré comme non satisfaisant, et la mise sur le marché est abandonnée.

## 2. Approche séquentielle

Les appareils sont rangés dans un ordre quelconque ; à chaque fois qu'un nouvel appareil est testé, on regarde le quotient  $\frac{d}{n}$  ( $d$  : nombre d'appareils défectueux,  $n$  : nombre d'appareils testés); si à un moment quelconque ce quotient est  $\geq \frac{1}{100}$  le produit est considéré comme non-satisfaisant.

**Remarque.** – *Les mêmes approches sont possibles pour le test d'un médicament, d'un vaccin, etc., sur une population.*

Voyons les différences entre les deux approches.

**Cas 1.** – *L'échantillon total se compose de  $N = 100$  appareils.*

En ce cas, le produit sera accepté si on ne trouve aucun appareil défectueux parmi les 100 testés ; en effet, si on en trouve 1, on a l'égalité  $\frac{D}{100} = \frac{1}{100}$  et il y a rejet. La conclusion est la même, que l'on travaille en global ou en séquentiel.

**Cas 2.** – *L'échantillon total se compose de  $N = 200$  appareils.*

En global, le produit sera accepté si  $D = 0$  ou 1. Notons  $p$  la probabilité (élémentaire) qu'un appareil soit défectueux ; cette probabilité est supposée être la même pour tous les appareils.

Alors :

$$P_{Global} = (1-p)^{200} + \binom{200}{1} p (1-p)^{199}$$

En séquentiel, le produit sera accepté si :

- Les 100 premiers tests ne donnent rien ;
- Les 100 suivants donnent 0 ou 1 défauts.

On a donc :

$$\begin{aligned}
P_{\text{séquentiel}} &= (1-p)^{100} \left( (1-p)^{100} + \binom{100}{1} p (1-p)^{99} \right) \\
&= (1-p)^{200} + \binom{100}{1} p (1-p)^{199}
\end{aligned}$$

Il est donc évident que  $P_{\text{séquentiel}} < P_{\text{global}}$  : on a moins de chances de satisfaire le test si on le fait en séquentiel.

Généralisons ceci immédiatement.

**Cas 3.** – *L'échantillon total se compose de  $N = 1000$  appareils.*

En global, l'échantillon sera accepté si  $D \leq 9$  : on peut se permettre 9 appareils défectueux, où qu'ils soient dans l'échantillon.

En séquentiel, il faut aussi que, au total,  $D \leq 9$ , mais on a des contraintes supplémentaires. Notons  $D_k$  le nombre de produits défectueux dans la  $k^{\text{ème}}$  tranche, c'est-à-dire entre  $100(k-1)+1$  et  $100k$ . On doit avoir :

$$\begin{aligned}
D_1 &= 0, \\
D_1 + D_2 &\leq 1, \\
D_1 + D_2 + D_3 &\leq 2, \\
&\text{etc., jusqu'à :} \\
D_1 + \dots + D_{10} &\leq 9.
\end{aligned}$$

Par exemple la répartition  $D_1 = 9$ ,  $D_2 = \dots = D_9 = 0$  (9 appareils défectueux dans la première tranche, aucun ensuite) sera acceptable pour l'approche globale, mais non pour l'approche séquentielle.

Il est donc complètement évident que l'approche séquentielle est plus pénalisante pour l'industriel que l'approche globale.

### 3. Lien avec la marche aléatoire dans le plan.

Notons  $X_k$  la variable aléatoire qui vaut 1 si le  $k^{\text{ème}}$  appareil est défectueux (probabilité  $p$ ) et 0 sinon, on a  $S_n = \sum_{k=1}^n X_k$  ; la condition de succès du test séquentiel se traduit donc par le fait que la marche aléatoire  $S_n$  reste constamment au-dessous de la droite  $y = \frac{x}{100}$ .

Ce sont les premiers tests qui sont les plus dangereux, puisque, asymptotiquement, la marche aléatoire  $S_n$  vérifie :

$$np - \alpha p(1-p)\sqrt{2n \log \log(n)} \leq S_n \leq np + \alpha p(1-p)\sqrt{2n \log \log(n)}$$

d'après la loi de Khintchine. Comme  $p = \frac{1}{200}$ ,  $np$  est petit devant  $\frac{n}{100}$ .

#### 4. Un exemple numérique

Traisons complètement un exemple numérique. On teste 500 appareils. La probabilité élémentaire de panne, pour chacun d'eux, est  $p = 1/200$ . On rejette un lot si  $\frac{D}{N} \geq \frac{1}{100}$ , comme précédemment.

Si on fait un test global, la probabilité de succès est

$$\sum_{k=0}^4 \binom{500}{k} p^k (1-p)^{500-k} \approx 0.89$$

Si on fait un test séquentiel, la probabilité de succès est obtenue en caractérisant toutes les situations favorables et en calculant la probabilité de chacune. On trouve 0.528 : c'est beaucoup plus faible.

### V. Application générale des lois du hasard

Contrairement à ce que l'on croit généralement, les lois du hasard ne réclament pas l'équilibre, l'uniformité. Bien au contraire, la Nature impose des déséquilibres dont l'ampleur va croissant : c'est ce qu'illustrent les lois de Khintchine, dans une situation très simple.

Voici des exemples d'application, qui tous contredisent l'intuition :

- Si on observe dix villes en apparence identiques, il n'y a pas de raison qu'elles recèlent le même nombre de cancers ; bien au contraire, les disparités iront en croissant.
- Si des partis politiques ont chacun une certaine "attractivité", il est faux que le comportement des électeurs se traduira par un équilibre reflétant cette attractivité (x% pour le premier, y% pour le second, z% pour le troisième, etc.) ; bien au contraire, quelle que soit l'attractivité, les effectifs des partis vont fluctuer, et les disparités iront en croissant.
- Les hypothèses d'équilibre économique souvent adoptées sont fausses. Si par exemple des produits de grande consommation se partagent un marché, les effectifs pour chacun varieront et les disparités iront croissant.

## Seconde Partie

### Démonstration des Théorèmes



Nous reprenons les résultats et en indiquons la démonstration.

**Théorème 1.** - *La probabilité que la première égalité se produise au-delà de  $2n$  vaut :*

$$g_n = \frac{1}{4^n} \binom{2n}{n} \approx \frac{1}{\sqrt{\pi n}}$$

Elle décroît donc lentement à mesure que  $n$  augmente.

Notre présentation suit celle du livre de J.C. Kalbfleisch "Probability and Statistical Inference", volume 1.

Soient  $A$  et  $B$  deux points du plan ; on note  $Nc(A, B)$  le nombre de chemins partant de  $A$  et finissant en  $B$ . Avec les coordonnées, on a en particulier :

$$Nc((0,0), (1,0)) = 1$$

et de même :

$$Nc((0,0), (2,0)) = 2, \quad Nc((0,0), (2,1)) = 0$$

Plus généralement, si  $n$  est pair,  $n = 2m$ , partant de 0 on ne peut atteindre qu'un point d'ordonnée paire et si  $n$  est impair, on ne peut atteindre qu'un point d'ordonnée impaire.

On remarque également que les points accessibles à partir de l'origine vérifient nécessairement  $|y| \leq x$ .

Si le point initial  $A$  est l'origine et si le point final  $B$  est donné par ses coordonnées  $(x, y)$ , nous noterons simplement  $N(x, y) = Nc(O, B)$ .

Nous aurons besoin de plusieurs lemmes.

**Lemme 2.** – *Le nombre de chemins entre 0 et le point de coordonnées  $(x, y)$  vaut :*

$$N(x, y) = \binom{x}{\frac{x+y}{2}}$$

## Démonstration du lemme 2

Si l'on veut rejoindre le point  $(x, y)$ , il y aura  $x$  parties à jouer et le gain final doit être  $y$ . Notons  $g$  le nombre de gains (cas  $X = 1$ ) et  $l$  le nombre de pertes (losses, cas  $X = -1$ ), on a  $g + l = x$  et  $g - l = y$ . Le nombre de chemins est le nombre de répartitions de  $g$  parmi  $g + l$ :

$$N(x, y) = \binom{g+l}{g} = \binom{x}{\frac{x+y}{2}}$$

ce qui prouve le Lemme 2.

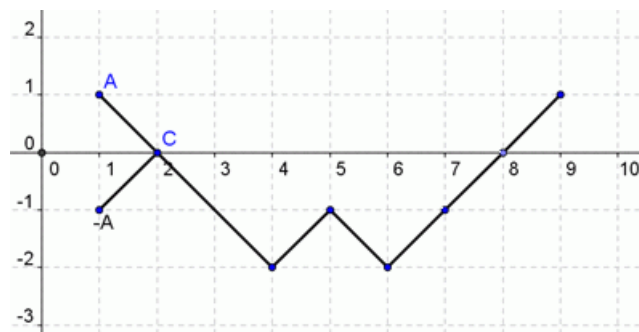
**Lemme 3.** – *Le nombre de chemins entre  $A = (a_1, a_2)$  et  $B = (b_1, b_2)$  vaut  $N(b_1 - a_1, b_2 - a_2)$ .*

Démonstration : ce lemme est évident ; par translation on peut toujours ramener le point de départ à l'origine.

**Lemme 4.** – *On suppose  $a_1 > 0$  et  $b_1 > 0$ . Le nombre de chemins entre  $A = (a_1, a_2)$  et  $B = (b_1, b_2)$  qui touchent l'axe des  $x$  ou qui le franchissent est égal à  $N(b_1 - a_1, b_2 + a_2)$ .*

## Démonstration du Lemme 4

Soit un chemin partant de  $A$ , finissant à  $B$ , touchant l'axe des  $x$  ; soit  $C$  le premier point de contact avec cet axe (il peut y en avoir plusieurs). Considérons la portion de chemin symétrique entre  $-A$  et  $C$  (voir figure) : on obtient ainsi un chemin biunivoquement associé au précédent. Donc le nombre de chemins est le même, et la formule est établie.



**Corollaire 5.** – *On suppose  $a_1 > 0$  et  $b_1 > 0$ . Le nombre de chemins entre  $A = (a_1, a_2)$  et  $B = (b_1, b_2)$  qui ne touchent pas l'axe des  $x$  est égal à  $N(b_1 - a_1, b_2 - a_2) - N(b_1 - a_1, b_2 + a_2)$ .*

Notons  $N_+(x, y)$  le nombre de chemins partant de l'origine, atteignant le point  $B = (x, y)$  (avec  $y > 0$ ) et qui ne coupent jamais l'axe des  $x$ , c'est-à-dire pour lesquels  $S_k > 0$  pour  $k > 0$ .

**Lemme 6.** – On suppose  $b_1 > 0$  et  $b_2 > 0$ . On a :

$$N_+(b_1, b_2) = \frac{b_2}{b_1} N(b_1, b_2).$$

**Démonstration du lemme 6**

Un tel chemin part du point  $A = (1,1)$  et est entièrement au dessus de  $Ox$ . Par conséquent, d'après le Corollaire 4, le nombre de chemins est  $N(b_1 - 1, b_2 - 1) - N(b_1 - 1, b_2 + 1)$ . Or :

$$\frac{N(b_1 - 1, b_2 - 1)}{N(b_1, b_2)} - \frac{N(b_1 - 1, b_2 + 1)}{N(b_1, b_2)} = \frac{\binom{b_1 - 1}{\frac{b_1 + b_2 - 1}{2}}}{\binom{b_1}{\frac{b_1 + b_2}{2}}} - \frac{\binom{b_1 - 1}{\frac{b_1 + b_2}{2}}}{\binom{b_1}{\frac{b_1 + b_2}{2}}} = \frac{b_2}{b_1}$$

ce qui prouve le Lemme 6.

**Lemme 7.** – On a, pour tout  $n \geq 1$  :

$$N_+(2n, 0) = \frac{1}{n} \binom{2n - 2}{n - 2}.$$

**Démonstration du Lemme 7**

Un tel chemin arrive nécessairement au point  $A = (2n - 1, 1)$ . D'après le Lemme 6, le nombre de chemins entièrement au dessus de  $Ox$  est :

$$\frac{1}{2n - 1} N(2n - 1, 1) = \frac{1}{2n - 1} \binom{2n - 1}{n} = \frac{(2n - 1)!}{(2n - 1)n!(n - 1)!} = \frac{(2n - 2)!}{n!(n - 1)!} = \frac{1}{n} \binom{2n - 2}{n - 1}$$

ce qui prouve le Lemme 7.

**Lemme 8.** – La probabilité d'avoir égalité pour la première fois à la partie  $2n$  est :

$$f_n = \frac{2}{n} \binom{2n - 2}{n - 1} 2^{-2n}$$

### Démonstration du Lemme 8

Par symétrie, il y a le même nombre de chemins entièrement au dessus et entièrement au dessous de  $Ox$ , atteignant le point  $(2n,0)$ . Leur nombre est au total  $\frac{2}{n} \binom{2n-2}{n-2}$  d'après le Lemme 7. La probabilité est obtenue en divisant par le nombre total de chemins, à savoir  $4^n$ .

**Lemme 9.** – *Posons :*

$$g_n = \binom{2n}{n} 2^{-2n}$$

Alors :

$$f_n = g_{n-1} - g_n$$

### Démonstration du Lemme 9

On a :

$$\begin{aligned} g_{n-1} - g_n &= \binom{2n-2}{n-1} 2^{-2n+2} - \binom{2n}{n} 2^{-2n} \\ &= 4 \binom{2n-2}{n-1} 2^{-2n} - \binom{2n}{n} 2^{-2n} \\ &= 4 \binom{2n-2}{n-1} 2^{-2n} - \frac{2n(2n-1)}{n^2} \binom{2n-2}{n-1} 2^{-2n} \\ &= \frac{2}{n} \binom{2n-2}{n-1} 2^{-2n} \\ &= f_n \end{aligned}$$

ce qui prouve le Lemme 9.

Démontrons maintenant le Théorème 1 : la probabilité que l'égalité se produise à  $2n$  ou avant est  $f_1 + f_2 + \dots + f_n = 1 - g_n$  d'après le Lemme 9. La formule de Stirling donne :

$$g_n = \binom{2n}{n} 2^{-2n} \approx \frac{1}{\sqrt{\pi n}}$$

et le Théorème 1 est démontré.

**Corollaire.** – *La probabilité d'avoir constamment  $S_k > 0$  (ou constamment  $S_k < 0$ ), pour  $k = 1, \dots, 2n$ , vaut :*

$$P(S_k > 0, k = 1, \dots, 2n) = \frac{1}{2^{2n+1}} \binom{2n}{n}$$

En effet, les chemins se décomposent en trois groupes disjoints : ceux qui sont constamment au-dessus de l'axe des  $x$ , ceux qui sont constamment au-dessous, et ceux qui le coupent. Les deux premiers groupes ont la même probabilité et la probabilité du troisième groupe est donnée par le Théorème 1.

Nous allons maintenant nous intéresser à l'impossibilité du confinement dans une bande donnée, ou dans une région donnée. En premier lieu, nous allons donner une estimation simple, qui montre que la marche aléatoire ne peut rester confinée dans une bande  $-s, s$  :

**Proposition.** - La probabilité qu'à l'instant  $n$  la somme  $S_n$  vérifie  $|S_n| \leq s$  est donnée par la formule :

$$P(|S_n| \leq s) = \sum_{j=-s}^s P(S_n = j) = \frac{1}{2^n} \sum_{j=-s}^s \binom{n}{\frac{n+j}{2}}$$

et, pour  $s$  fixé, cette quantité tend vers 0 lorsque  $n \rightarrow \infty$ .

Fixons par exemple  $s = 10$ . Pour  $n = 20$ , la probabilité d'avoir  $|S_n| \leq 10$  est 0.58 ; pour  $n = 100$ , elle vaut  $0.15 \times 10^{-16}$ .

Nous allons maintenant évaluer la probabilité d'atteindre un seuil donné au plus tard à un instant donné.

Pour tout seuil  $s$  (par exemple 1000 Euros), pour tout  $n$ , on pose :

$$\alpha(n, s) = P(\exists k, 1 \leq k \leq n, S_k = s)$$

C'est la probabilité d'atteindre le seuil  $s$  au plus tard à l'instant  $n$ .

La quantité  $1 - \alpha(n, s)$  est la probabilité :

$$1 - \alpha(n, s) = P(\forall k, 1 \leq k \leq n, S_k \neq s) = P(\forall k, 1 \leq k \leq n, S_k < s)$$

car la marche aléatoire ne peut prendre aucune valeur  $> s$  sans être d'abord passée par la valeur  $s$ .

**Théorème 2.** - On a, pour tout  $s \geq 1$  :

$$\alpha(2n, 2s) = 1 - \frac{1}{4^n} \sum_{j=-s+1}^s \binom{2n}{n+j}$$

## Démonstration du Théorème 2

Elle reprend essentiellement les ingrédients de la démonstration du Théorème 1.

**Lemme 1.** - *Le nombre de chemins partant de  $(0, s)$ , atteignant le point  $B = (n, b)$ ,  $b > 0$ , et ne touchant pas l'axe des  $x$  est :*

$$N_+((0, s), (n, b)) = N(n, b - s) - N(n, b + s)$$

## Démonstration du Lemme 1

Ceci est une application immédiate du Corollaire 5 ci-dessus : on prend  $a_1 = 0$ ,  $a_2 = s$ ,  $b_1 = n$ ,  $b_2 = b$ .

**Lemme 2.** - *Le nombre de chemins partant de  $(0, s)$ , atteignant n'importe quel point  $B = (n, b)$ ,  $b > 0$ , et ne touchant pas l'axe des  $x$  est :*

$$N_+(s, n) = \sum_{b>0} N(n, b - s) - \sum_{b>0} N(n, b + s)$$

Le nombre  $N(n, b)$  est nul si  $b > n$ ; par conséquent, la somme ci-dessus s'étend de  $b = 1$  à  $b - s \leq n$  pour le premier terme et à  $b + s \leq n$  pour le second :

$$N_+(s, n) = \sum_{b=1}^{n+s} N(n, b - s) - \sum_{b=1}^{n-s} N(n, b + s)$$

Nous remarquons ensuite que :

- Le nombre de chemins partant de  $(0, s)$  et ne touchant pas  $y = 0$  est égal au nombre de chemins partant de  $(0, 2s)$  et ne touchant pas  $y = s$  (translation vers le haut).
- Le nombre de chemins partant de 0 et ne touchant pas  $y = s$  est égal au nombre de chemins partant de  $(0, 2s)$  et ne touchant pas  $y = s$  (symétrie par rapport à  $y = s$ ).

Nous en déduisons :

**Lemme 4.** - *Le nombre de chemins partant de 0 et n'atteignant pas le seuil  $s$  avant le temps  $n$  est :*

$$N_+(s, n) = \sum_{b=1}^{n+s} N(n, b - s) - \sum_{b=1}^{n-s} N(n, b + s) = \sum_{b=1}^{n+s} \binom{n}{\frac{n+b-s}{2}} - \sum_{b=1}^{n-s} \binom{n}{\frac{n+b+s}{2}}$$

Remplaçons  $n$  par  $2n$  et  $s$  par  $2s$ . Le nombre de chemins partant de 0 et n'atteignant pas le seuil  $2s$  avant le temps  $2n$  est, avec  $s \leq n$  :

$$\begin{aligned} N_+(2s, 2n) &= \sum_{b=1}^{n+s} \binom{2n}{n+b-s} - \sum_{b=1}^{n-s} \binom{2n}{n+b+s} \\ &= \binom{2n}{n-s+1} + \binom{2n}{n-s+2} + \cdots + \binom{2n}{2n} - \left( \binom{2n}{n+s+1} + \binom{2n}{n+s+2} + \cdots + \binom{2n}{2n} \right) \\ &= \sum_{j=-s+1}^s \binom{2n}{n+j} \end{aligned}$$

La probabilité de ne pas atteindre le seuil  $2s$  avant le temps  $2n$  est donc :

$$p = \frac{1}{4^n} \sum_{j=-s+1}^s \binom{2n}{n+j}$$

et le Théorème 2 est démontré. La somme ci-dessus comporte  $2s$  termes ; chacun est équivalent à  $\frac{1}{\sqrt{\pi n}}$  par la formule de Stirling, d'où l'équivalent annoncé, lorsque  $n \rightarrow +\infty$ .

Il n'est pas non plus possible que la marche aléatoire reste indéfiniment confinée à la zone  $|y| \leq \frac{\sqrt{x}}{2}$  :

**Théorème 3.** – *Pour tout  $n \geq 1$  :*

$$P\left(|S_n| > \frac{\sqrt{n}}{2}\right) \geq \frac{9}{16}$$

### Démonstration du Théorème 3

Nous suivons la présentation de Michael Damron :

[https://web.math.princeton.edu/~mdamron/teaching/F12/Junior\\_Seminar/lect\\_1.pdf](https://web.math.princeton.edu/~mdamron/teaching/F12/Junior_Seminar/lect_1.pdf) (Lemma 0.8).

Calculons le moment d'ordre 4 :

$$E(S_n^4) = \sum_{i_1, i_2, i_3, i_4=1}^n E(X_{i_1} X_{i_2} X_{i_3} X_{i_4}) = \sum_{i_1, i_2=1}^n E(X_{i_1}^2 X_{i_2}^2) = n^2$$

Donc :

$$E(S_n^2) = E(S_n^2 1_{S_n^2 > n/4}) + E(S_n^2 1_{S_n^2 \leq n/4}) \leq (E(S_n^4))^{1/2} \sqrt{P\left(S_n^2 > \frac{n}{4}\right)} + \frac{n}{4}$$

Comme  $E(S_n^2) = n$ , le résultat annoncé s'en déduit immédiatement.

Nous allons maintenant aborder les deux lois de Khintchine.

Notre démonstration adapte la présentation de Y. Velenik, Université de Genève "Chapitres choisis de Théorie des Probabilités" :

<http://www.unige.ch/math/folks/velenik/papers/LN-CC1.pdf>

On note :

$\varphi(x) = \sqrt{2x \text{Log}(\text{Log}(x))}$ , fonction définie pour tout  $x$  réel,  $x > e$ , appelée "courbe de Khintchine".

**Théorème 4 (première loi de Khintchine)** - Soient  $X_n$  des variables indépendantes à valeur  $\pm 1$  avec  $P\{X_n = 1\} = P\{X_n = -1\} = \frac{1}{2}$ . Notons :

$$S_n = X_1 + \dots + X_n$$

Soit  $\varepsilon > 0$  et :

$$\eta(m, \varepsilon) = \frac{1}{(\text{Log}(1 + \varepsilon))^{1+\varepsilon}} \frac{1}{\varepsilon(m-1)^\varepsilon}$$

Notons :

$$B(m, \varepsilon) = \left\{ \exists n \geq (1 + \varepsilon)^m, S_n > (1 + \varepsilon)\varphi(n) \right\}$$

Alors, pour tout  $m$  :

$$P(B(m, \varepsilon)) \leq \eta(m, \varepsilon)$$

L'ensemble  $B(m, \varepsilon)$  est constitué des chemins qui dépassent la "bande de sécurité"  $(1 + \varepsilon)\varphi$  au moins une fois après l'instant  $(1 + \varepsilon)^m$ . Le théorème dit que la probabilité d'un tel événement tend vers 0 lorsque  $m$  grandit. Autrement dit, pour une largeur de bande de sécurité fixée, c'est à dire  $\varepsilon > 0$  fixé, il devient de moins en moins probable de passer au dessus de la courbe  $(1 + \varepsilon)\varphi$  à mesure que l'on joue.



Prenons par exemple  $\varepsilon = 1$ . La probabilité de l'ensemble des chemins qui dépassent  $2\varphi$  après l'instant  $2^m$  est au plus  $\frac{1}{(\text{Log}(2))^2} \frac{1}{m-1}$ .

L'estimation symétrique  $S_n < -(1+\varepsilon)\varphi(n)$  est évidemment identique. On en déduit :

**Corollaire 2.** – *Pour tout  $m \geq 2$  et tout  $n \geq 2^m$ , on a :*

$$P(\exists n > 2^m \text{ tel que } |S_n| > 2\varphi(n)) < \frac{1}{m-1} \frac{2}{(\text{Log}(2))^2} < \frac{4.2}{m-1}$$

Ce corollaire nous sera utile lors de la démonstration du Théorème 5.

### Démonstration du Théorème 4

La démonstration se décompose en plusieurs étapes.

**Lemme 1.** – *Pour tout  $n \geq 1$  :*

$$P(S_n \geq 0) > \frac{1}{2}$$

### Démonstration du Lemme 1

On a en effet :

$$P(S_n > 0) = P(S_n < 0) ;$$

$$P(S_n < 0) + P(S_n = 0) + P(S_n > 0) = 1 \text{ et}$$

$$P(S_n \geq 0) = P(S_n > 0) + P(S_n = 0)$$

ce qui prouve le Lemme 1.

En termes de dénombrements, le lemme 1 est tout à fait évident : si on regarde l'altitude atteinte, à l'instant  $n$ , par un chemin partant de 0, cette altitude peut être positive, négative, ou nulle. Comme il y a autant de chemins pour lesquels  $S_n > 0$  que de chemins pour lesquels  $S_n < 0$ , le nombre de chemins pour lesquels  $S_n \geq 0$  est supérieur à la moitié du nombre total de chemins (qui vaut  $2^n$ ).

Le lemme 2 ci-dessous relie la probabilité de dépasser une valeur  $x$  à un instant donné  $n$  à la probabilité d'avoir déjà dépassé cette même valeur à un instant antérieur.

**Lemme 2.** – *Pour tout  $x$  réel et tout  $n \geq 1$ , on a :*

$$P(\exists k \leq n, S_k > x) \leq 2P(S_n > x)$$

## Démonstration du Lemme 2

Pour  $x$  réel fixé, posons :

$$\sigma_x = \inf \{k \geq 0, S_k > x\}$$

C'est le premier instant où la suite  $S_k$  dépasse la valeur  $x$ . Les événements  $\{\sigma_x = k\}$  sont deux à deux disjoints, et l'on a :

$$P\{\exists k \leq n, S_k > x\} = \sum_{k=1}^n P\{\sigma_x = k\}$$

Pour  $k = 1, \dots, n$ , introduisons l'événement :

$$U_k = \{S_k \leq S_n\}$$

Les événements  $\{\sigma_x = k\} \cap U_k$ ,  $k = 1, \dots, n$ , forment une partition de l'événement  $\{S_n > x\}$ ; en effet, ils sont disjoints et leur réunion recouvre bien  $\{S_n > x\}$  : si  $S_n > x$ , il y a bien un premier  $k$ ,  $1 \leq k \leq n$  tel que  $S_k = x + 1$  et donc  $S_k \leq S_n$ .

Par conséquent :

$$P\{S_n > x\} = \sum_{k=1}^n P(\{\sigma_x = k\} \cap U_k) = \sum_{k=1}^n P(\sigma_x = k)P(U_k)$$

en effet, l'événement  $\sigma_x = k$  dépend de  $X_1, \dots, X_k$  tandis que l'événement  $U_k$  s'écrit  $S_n - S_k > 0$ , c'est-à-dire  $X_{k+1} + \dots + X_n > 0$  ; il est donc indépendant de  $X_1, \dots, X_k$ .

On a :

$$\sum_{k=1}^n P(\sigma_x = k)P(U_k) \geq \min_{k=1, \dots, n} P(U_k) \sum_{k=1}^n P(\sigma_x = k)$$

Mais  $P(U_k) = P(X_{k+1} + \dots + X_n > 0) > \frac{1}{2}$  d'après le Lemme 1 : toutes les sommes partielles ont la même loi. Par conséquent :

$$\sum_{k=1}^n P(\sigma_x = k)P(U_k) \geq \frac{1}{2} \sum_{k=1}^n P(\sigma_x = k) = \frac{1}{2} P(\cup(\sigma_x = k))$$

Or la réunion  $\cup(\sigma_x = k)$  se décrit comme le fait qu'il existe un  $k$ ,  $1 \leq k \leq n$ , tel que  $S_k > x$  et le lemme 2 est démontré.

Ici encore, on peut réfléchir en termes de dénombrements de chemins. Si  $S_n > x$ , il existe un premier indice  $k_0$  tel que  $S_{k_0} > x$ ; cela signifie  $S_j \leq x$  si  $j = 1, \dots, k_0 - 1$ ,  $S_{k_0-1} = x$ ,  $S_{k_0} = x + 1$ . Entre cet instant  $k_0$  et l'instant  $n$ , l'altitude va nécessairement augmenter (au sens large), puisque  $S_{k_0} = x + 1$  et  $S_n \geq x + 1$ . Or la probabilité que l'altitude augmente (au sens large) est au moins  $1/2$  d'après le lemme 1.

Mais, par la suite, on ne peut plus raisonner en termes de dénombrements de chemins. Il faut adopter un raisonnement purement probabiliste abstrait, qui est évidemment difficile à suivre, faute d'intuition pour le guider.

**Lemme 3.** - Soit  $\varepsilon > 0$  et soit  $\gamma = 1 + \varepsilon$ . Pour tout  $k > 1$ , entier, posons  $n_k = \gamma^k$ . Posons :

$$C_k = \left\{ \exists n, n_k \leq n < n_{k+1}, S_n > (1 + \varepsilon) \sqrt{\varphi(n_k)} \right\}$$

Alors :

$$P(C_k) \leq (k \text{Log}(\gamma))^{-(1+\varepsilon)}$$

### Démonstration du Lemme 3

D'après le lemme 2 :

$$P(C_k) \leq 2P\left\{S_{n_{k+1}} > (1 + \varepsilon) \sqrt{\varphi(n_k)}\right\}$$

Nous utiliserons l'inégalité :

$$P(S_n \geq x) \leq e^{-x^2/(2n)} \tag{1}$$

valable pour tout  $n$  et tout  $x$ ,  $0 \leq x \leq n$  ; elle sera démontrée en Annexe.

On a :

$$P\left\{S_{n_{k+1}} > (1 + \varepsilon) \sqrt{\varphi(n_k)}\right\} \leq \exp\left(- (1 + \varepsilon)^2 \frac{n_k}{n_{k+1}} \text{Log}(\text{Log}(n_k))\right)$$

d'après (1). Or :

$$\frac{n_{k+1}}{n_k} = \frac{\gamma^{k+1}}{\gamma^k} = \gamma = 1 + \varepsilon$$

On en déduit, puisque  $n_k \geq \gamma^k$  :

$$P(C_k) \leq 2 \exp\left(- (1 + \varepsilon) \text{Log}(\text{Log}(n_k))\right) = (\text{Log}(n_k))^{-(1+\varepsilon)} \leq (k \text{Log}(\gamma))^{-(1+\varepsilon)}$$

ce qui achève la démonstration du Lemme 3.

Posons :

$$D_k = \left\{ \exists n, n_k \leq n < n_{k+1}, S_n > (1 + \varepsilon) \sqrt{2n \text{Log}(\text{Log}(n))} \right\}$$

**Lemme 4.** – Pour tout  $k$ ,  $D_k$  est inclus dans  $C_k$ .

#### Démonstration du Lemme 4

En effet, s'il existe un  $n$  tel que l'inégalité  $S_n > (1 + \varepsilon) \sqrt{2n \text{Log}(\text{Log}(n))}$  soit satisfaite, on a a fortiori  $S_n > (1 + \varepsilon) \sqrt{2n_k \text{Log}(\text{Log}(n_k))}$ , puisque  $n \text{Log}(\text{Log}(n)) \geq n_k \text{Log}(\text{Log}(n_k))$ . Ceci prouve le Lemme 4.

Il résulte du lemme 4 que l'inégalité (7) vaut a fortiori pour  $D_k$ .

Posons  $B_m = \bigcup_{k \geq m} D_k$  ; alors les ensembles  $B_m$  sont évidemment décroissants. On a :

**Lemme 5.** – Pour tout  $m$ , on a :

$$P(B_m) \leq \frac{1}{(\text{Log}(1 + \varepsilon))^{1+\varepsilon}} \frac{1}{\varepsilon(m-1)^\varepsilon}$$

#### Démonstration du Lemme 5

On a par définition :

$$P(B_m) \leq \sum_{k=m}^{+\infty} P(D_k) \leq \sum_{k=m}^{+\infty} P(C_k)$$

et le Lemme 4 résulte immédiatement du Lemme 3 et de l'inégalité :

$$\sum_{k=m}^{+\infty} \frac{1}{k^{1+\varepsilon}} \leq \int_{m-1}^{+\infty} \frac{dx}{x^{1+\varepsilon}} = \frac{1}{\varepsilon(m-1)^\varepsilon}$$

Achevons maintenant la démonstration du Théorème 4.

L'ensemble  $B(m, \varepsilon)$  est par définition :

$$B(m, \varepsilon) = \bigcup_{k \geq m} \{ \exists n, n_k \leq n < n_{k+1}, S_n > (1 + \varepsilon)\varphi(n) \} = \{ \exists n \geq (1 + \varepsilon)^m, S_n > (1 + \varepsilon)\varphi(n) \}$$

$B(m, \varepsilon)$  est donc l'ensemble des chemins qui dépassent la bande de Khintchine  $(1 + \varepsilon)\varphi$  au moins une fois après l'instant  $(1 + \varepsilon)^m$ .

On a d'après le Lemme 5 :

$$P(B(m, \varepsilon)) \leq \eta(m, \varepsilon)$$

avec :

$$\eta(m, \varepsilon) = \frac{1}{(\text{Log}(1 + \varepsilon))^{1+\varepsilon}} \frac{1}{\varepsilon(m-1)^\varepsilon} \rightarrow 0 \text{ lorsque } m \rightarrow +\infty.$$

et le Théorème 4 est démontré.

Nous allons maintenant montrer le Théorème 5 : la marche aléatoire s'approche de la courbe de Khintchine.

On va montrer qu'il existe une suite d'entiers  $n_k$  telle que  $\frac{S_{n_k}}{\varphi(n_k)} \rightarrow 1$  lorsque  $k \rightarrow +\infty$ . Ceci signifie que, pour tout  $\varepsilon > 0$ , la probabilité que  $S_n < (1 - \varepsilon)\varphi(n)$  pour tout  $n = 1, \dots, N$  tend vers 0 avec  $N$  : c'est cette probabilité qu'il s'agit de quantifier.

Soit  $\alpha > 0$  (petit). On va montrer qu'il existe  $N_0(\varepsilon, \alpha)$  tel que si  $N > N_0(\varepsilon, \alpha)$  :

$$P\left(\forall n = 1, \dots, N ; \frac{S_n}{\varphi(n)} < 1 - \varepsilon\right) < \alpha.$$

D'où l'énoncé précis :

**Théorème 5.** – Soient  $\varepsilon > 0$ ,  $\alpha > 0$ . Posons  $C(\varepsilon) = \frac{1}{e} \sqrt{\frac{2}{\pi}} \frac{1}{(\text{Log}(16/\varepsilon^2))^{1-\varepsilon/4}}$  et  $k_1 = \exp\left(\frac{32}{\varepsilon^2}\right)$ .

Soit  $N_0 = \left(k_1^4 + \frac{\varepsilon}{4C(\varepsilon)} \text{Log}\left(\frac{2}{\alpha}\right)\right)^{\frac{4}{\varepsilon}}$ . Alors :

$$P\left(\exists k \leq N_0 \text{ tel que } \frac{S_k}{\varphi(k)} > 1 - \varepsilon\right) \geq 1 - \alpha$$

En d'autres termes, avec probabilité grande (au moins  $1 - \alpha$ ), il existe un moment avant  $N_0$  où le quotient  $\frac{S_k}{\varphi(k)}$  dépasse  $1 - \varepsilon$ .

Par exemple, avec  $\varepsilon = \frac{1}{2}$  et  $\alpha = \frac{1}{2}$ , on trouve  $N \geq 3 \times 10^{55}$

### Démonstration du Théorème 5

On pose comme précédemment :

$$\varphi(x) = \sqrt{2x \text{Log}(\text{Log}(x))}$$

Soit  $k$  un entier positif. On pose  $\gamma = \frac{16}{\varepsilon^2}$ , et  $n_k = \gamma^k$ .

On a écrit le quotient  $\frac{S_{n_k}}{\varphi(n_k)}$  sous la forme de la somme de deux termes, qui seront traités séparément.

$$\frac{S_{n_k}}{\varphi(n_k)} = \frac{S_{n_k} - S_{n_{k-1}}}{\varphi(n_k)} + \frac{S_{n_{k-1}}}{\varphi(n_{k-1})} \frac{\varphi(n_{k-1})}{\varphi(n_k)}$$

Posons  $Y_k = \frac{S_{n_k} - S_{n_{k-1}}}{\varphi(n_k)}$  et  $Z_k = \frac{S_{n_{k-1}}}{\varphi(n_{k-1})} \frac{\varphi(n_{k-1})}{\varphi(n_k)}$ . Le principe de la démonstration consistera à montrer que  $Y_k$  est grand avec forte probabilité, tandis que  $Z_k$  est petit.

Commençons par nous occuper de  $Z_k$ .

**Lemme 1.** – On a, pour tout  $k \geq 2$  :

$$\frac{\varphi(n_{k-1})}{\varphi(n_k)} \leq \frac{1}{\sqrt{\gamma}} = \frac{\varepsilon}{4}$$

## Démonstration du Lemme 1

Ceci est évident puisque :

$$\frac{\varphi(n_{k-1})}{\varphi(n_k)} = \frac{\sqrt{\gamma^{k-1} \text{Log}((k-1) \text{Log}(\gamma))}}{\sqrt{\gamma^k \text{Log}(k \text{Log}(\gamma))}}$$

et d'après le choix de  $\gamma$ .

**Lemme 2.** – Soit  $k_0 = \frac{2}{\alpha}$ , on a :

$$P\left(\forall k \geq k_0, \frac{|S_{n_{k-1}}|}{\varphi(n_{k-1})} < 2\right) \geq 1 - \frac{\alpha}{2}$$

## Démonstration du Lemme 2

Soit  $\delta = \frac{2}{(\text{Log}(2))^2}$ . Nous observons que la condition  $k \geq \frac{2}{\alpha}$  implique  $k \geq 1 + \left(2 + \frac{\delta}{\alpha}\right) \frac{\text{Log}(2)}{\text{Log}(\gamma)}$ .

En effet :

$$1 + \left(2 + \frac{\delta}{\alpha}\right) \frac{\text{Log}(2)}{\text{Log}(\gamma)} \leq 1 + \left(2 + \frac{\delta}{\alpha}\right) \frac{\text{Log}(2)}{\text{Log}(16)} \leq \frac{2}{\alpha}$$

Utilisons le Corollaire 2 du Théorème 3. Choisissons  $m = 2 + \frac{2\delta}{\alpha}$  ; on a  $\frac{\delta}{m-1} < \frac{\alpha}{2}$ . Alors, d'après le Corollaire 2, th. 3, on a :

$$P(\exists n > 2^m, |S_n| > 2\varphi(n)) < \frac{\alpha}{2}$$

et donc :

$$P\left(\forall n > 2^m, \frac{|S_n|}{\varphi(n)} \leq 2\right) \geq 1 - \frac{\alpha}{2}$$

Ici,  $n = \gamma^{k-1}$  et la condition  $n \geq 2^m$  se traduit par  $k \geq 1 + \left(2 + \frac{\delta}{\alpha}\right) \frac{\text{Log}(2)}{\text{Log}(\gamma)}$ , qui est satisfaite, et le Lemme 2 est démontré.

**Proposition 3.** - Si  $k \geq k_0 = \frac{2}{\alpha}$ , on a :

$$P\left(\forall k \geq k_0, |Z_k| < \frac{\varepsilon}{2}\right) \geq 1 - \frac{\alpha}{2}$$

### Démonstration de la Proposition 3

En effet, pour tout  $k$ , on aura simultanément  $\frac{|S_{n_{k-1}}|}{\varphi(n_{k-1})} < 2$  et  $\frac{\varphi(n_{k-1})}{\varphi(n_k)} < \frac{\varepsilon}{4}$  sur un ensemble de probabilité  $\geq 1 - \frac{\alpha}{2}$ . Ceci prouve la Proposition 3.

Occupons-nous maintenant du terme  $Y_k = \frac{S_{n_k} - S_{n_{k-1}}}{\varphi(n_k)}$ . Nous aurons besoin de plusieurs énoncés techniques.

**Proposition 4.** - Si  $n \geq 3$  et si  $k \leq n^{2/3}$ , on a :

$$P(S_n = k) \geq e^{-\frac{k^2}{2n}} \frac{1}{e} \sqrt{\frac{2}{\pi n}}$$

### Démonstration de la Proposition 4

On a :

$$P(S_n = k) = \frac{1}{2^n} \binom{n}{\frac{n+k}{2}} = \frac{1}{2^n} \frac{n!}{\left(\frac{n+k}{2}\right)! \left(\frac{n-k}{2}\right)!}$$

On applique la formule de Stirling aux différents termes :



$$\begin{aligned}
A &= \frac{1}{2^n} \frac{n!}{\left(\frac{n+k}{2}\right)! \left(\frac{n-k}{2}\right)!} \\
&\approx \frac{1}{2^n} \frac{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}}{\left(\frac{\frac{n+k}{2}}{e}\right)^{\frac{n+k}{2}} \sqrt{2\pi \left(\frac{n+k}{2}\right)} \left(\frac{\frac{n-k}{2}}{e}\right)^{\frac{n-k}{2}} \sqrt{2\pi \left(\frac{n-k}{2}\right)}} \\
&= \frac{n^n}{(n+k)^{\frac{n+k}{2}} (n-k)^{\frac{n-k}{2}}} \sqrt{\frac{2n}{\pi(n^2-k^2)}}
\end{aligned}$$

et donc :

$$\begin{aligned}
A &= \frac{1}{\left(1+\frac{k}{n}\right)^{\frac{n+k}{2}} \left(1-\frac{k}{n}\right)^{\frac{n-k}{2}}} \sqrt{\frac{2n}{\pi(n^2-k^2)}} \\
&= \frac{1}{\left(1+\frac{k}{n}\right)^{\frac{n}{2}} \left(1-\frac{k}{n}\right)^{\frac{n}{2}} \left(1+\frac{k}{n}\right)^{\frac{k}{2}} \left(1-\frac{k}{n}\right)^{-\frac{k}{2}}} \sqrt{\frac{2n}{\pi(n^2-k^2)}} \\
&= \frac{1}{\left(1-\frac{k^2}{n^2}\right)^{\frac{n}{2}} \left(1+\frac{k}{n}\right)^{\frac{k}{2}} \left(1-\frac{k}{n}\right)^{-\frac{k}{2}}} \sqrt{\frac{2n}{\pi(n^2-k^2)}}
\end{aligned}$$

Nous aurons besoin d'estimations différentes pour les différents termes :

Posons  $B = \left(1 - \frac{k^2}{n^2}\right)^{\frac{n}{2}}$ . On a  $\text{Log}(B) = \frac{n}{2} \text{Log}\left(1 - \frac{k^2}{n^2}\right)$

Or  $\text{Log}(1-x) \leq -x$  si  $0 \leq x < 1$ , et donc  $\text{Log}(B) \leq -\frac{k^2}{2n}$  et  $B \leq e^{-\frac{k^2}{2n}}$ .

Soit  $C = \left(1 + \frac{k}{n}\right)^{\frac{k}{2}}$ . On a  $\text{Log}(C) = \frac{k}{2} \text{Log}\left(1 + \frac{k}{n}\right)$

Or  $\text{Log}(1+x) \leq x$ ,  $x \geq 0$  et donc  $\text{Log}(C) \leq \frac{k^2}{2n}$  et  $C \leq e^{\frac{k^2}{2n}}$ .

Soit enfin  $D = \left(1 - \frac{k}{n}\right)^{\frac{k}{2}}$ .

**Lemme 5.** – Pour tout  $x$ ,  $0 \leq x \leq \frac{3}{4}$  on a  $\text{Log}(1-x) \geq -x - 2x^2$ .

### Démonstration du Lemme 5

On veut montrer que  $y = 2x^2 + x + \text{Log}(1-x) \geq 0$ .

C'est vrai si  $x=0$ . On a  $y' = 4x + 1 - \frac{1}{1-x}$ , du signe de  $(4x+1)(1-x) - 1 = x(3-4x)$ ,  
et donc  $y' \geq 0$  si  $0 \leq x \leq \frac{3}{4}$ . Ceci prouve le Lemme 5.

Nous revenons à la démonstration de la Proposition 4.

On a, si  $\frac{k}{n} \leq \frac{3}{4}$  :

$$\text{Log}(D) = \frac{k}{2} \text{Log}\left(1 - \frac{k}{n}\right) \geq \frac{k}{2} \left(-\frac{k}{n} - \frac{2k^2}{n^2}\right) = -\frac{k^2}{2n} - \frac{k^3}{n^2}$$

et donc :

$$D \geq e^{-\frac{k^2}{2n} - \frac{k^3}{n^2}}$$

Au total :

$$A = \frac{D}{BC} \sqrt{\frac{2n}{\pi(n^2 - k^2)}} \geq e^{\frac{k^2}{2n} - \frac{k^2}{2n}} e^{-\frac{k^2}{2n} - \frac{k^3}{n^2}} \sqrt{\frac{2n}{\pi(n^2 - k^2)}} \geq e^{-\frac{k^2}{2n} - \frac{k^3}{n^2}} \sqrt{\frac{2n}{\pi(n^2 - k^2)}}$$

Or, si  $k \leq n^{2/3}$ , on a  $\frac{k^3}{n^2} \leq 1$  et  $e^{-\frac{k^3}{n^2}} \geq \frac{1}{e}$ .

La condition  $k \leq \frac{3n}{4}$  est réalisée dès que  $k \leq n^{2/3}$  si  $n^{2/3} \leq \frac{3n}{4}$ , ou  $n \geq \left(\frac{4}{3}\right)^3 \approx 2.37$ .

Par ailleurs,  $\sqrt{\frac{2n}{\pi(n^2 - k^2)}} \geq \sqrt{\frac{2}{\pi n}}$  et la Proposition 4 est démontrée.

**Corollaire 6.** - Si  $u \leq n^{2/3} - \sqrt{n}$ , on a :

$$P(S_n \geq u) \geq \frac{1}{e} \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2} \left( \frac{u}{\sqrt{n}} + 1 \right)^2}$$

En effet, on écrit :

$$P(S_n \geq u) = \sum_{j \geq u} P(S_n = j) \geq \sum_{j=u}^{u+\sqrt{n}} P(S_n = j) \geq \sum_{j=u}^{u+\sqrt{n}} e^{-\frac{j^2}{2n}} \frac{1}{e} \sqrt{\frac{2}{\pi n}} \geq \frac{1}{e} \sqrt{\frac{2}{\pi}} e^{-\frac{(u+\sqrt{n})^2}{2n}} ;$$

puisque cette somme comporte  $\sqrt{n}$  termes, tous minorés par le dernier ; pour celui-ci, on a bien  $j \leq n^{2/3}$ .

Nous allons maintenant pouvoir aborder l'estimation concernant le terme  $Y_k = \frac{S_{n_k} - S_{n_{k-1}}}{\varphi(n_k)}$ .

Soient  $\varepsilon, \alpha$  comme précédemment. On note :

$$D_k = \left\{ \frac{S_{n_k} - S_{n_{k-1}}}{\varphi(n_k)} > 1 - \frac{\varepsilon}{2} \right\}$$

**Proposition 7.** – Pour tout  $\varepsilon > 0$  et pour tout entier  $k \geq \exp(32 / \varepsilon^2)$ , on a :

$$P(D_k) \geq u(k, \varepsilon)$$

en posant :

$$u(k, \varepsilon) = \frac{1}{e} \sqrt{\frac{2}{\pi}} \frac{1}{\left( \text{Log}(16 / \varepsilon^2) \right)^{1-\varepsilon/4}} \frac{1}{k^{1-\varepsilon/4}}$$

### Démonstration de la Proposition 7

Comme la condition porte sur des différences consécutives  $S_{n_k} - S_{n_{k-1}}$ , les événements  $D_k$  sont indépendants. On a :

$$P(D_k) = P \left\{ S_{n_k} - S_{n_{k-1}} > \left( 1 - \frac{\varepsilon}{2} \right) \varphi(n_k) \right\} = P \left\{ S_{n_k - n_{k-1}} > \left( 1 - \frac{\varepsilon}{2} \right) \varphi(n_k) \right\}$$

parce que  $S_{n_k} - S_{n_{k-1}}$  et  $S_{n_k - n_{k-1}}$  ont la même loi.

Commençons par montrer que la condition  $u \leq n^{2/3} - \sqrt{n}$  est satisfaite si nous remplaçons  $n$  par  $n_k - n_{k-1}$  et  $u$  par  $\left(1 - \frac{\varepsilon}{2}\right) \sqrt{2n_k \text{Log}(\text{Log}(n_k))}$ .

Cette condition s'écrit :

$$\left(1 - \frac{\varepsilon}{2}\right) \sqrt{2n_k \text{Log}(\text{Log}(n_k))} \leq (n_k - n_{k-1})^{2/3} - \sqrt{n_k - n_{k-1}} \quad (1)$$

ou encore :

$$\left(1 - \frac{\varepsilon}{2}\right) \sqrt{2n_k \text{Log}(\text{Log}(n_k))} + \sqrt{n_k - n_{k-1}} \leq (n_k - n_{k-1})^{2/3}$$

Nous avons :

$$(n_k - n_{k-1})^{2/3} = \gamma^{2k/3} \left(1 - \frac{1}{\gamma}\right)^{2/3}$$

et  $\left(1 - \frac{1}{\gamma}\right)^{2/3} \geq \frac{1}{2}$  dès que  $1 - \frac{1}{2^{3/2}} \geq \frac{1}{\gamma}$  ; cette condition est vérifiée puisque  $\gamma = \frac{16}{\varepsilon^2}$ . Pour montrer (1), il nous suffit donc de montrer que :

$$\sqrt{2n_k \text{Log}(\text{Log}(n_k))} + \sqrt{n_k} \leq \frac{1}{2} n_k^{2/3}$$

et ceci sera réalisé dès que :

$$4\sqrt{n_k \text{Log}(\text{Log}(n_k))} \leq n_k^{2/3}$$

ou encore :

$$16\text{Log}(\text{Log}(n_k)) \leq n_k^{1/3}$$

Or l'inégalité  $16\text{Log}(\text{Log}(x)) \leq x^{1/3}$  est vraie pour tout  $x \geq 480$ . Comme  $n_k = \gamma^k$  et  $\gamma = \frac{16}{\varepsilon^2}$ , la condition (1) sera réalisée dès que :

$$k \geq 3 \geq \frac{\text{Log}(480)}{\text{Log}(16)} \geq \frac{\text{Log}(480)}{\text{Log}(16/\varepsilon^2)}$$

qui est réalisée par hypothèse.

Remplaçons, dans l'estimation donnée par le Corollaire 6 ci-dessus,  $n$  par  $n_k - n_{k-1}$  et  $u$  par  $\left(1 - \frac{\varepsilon}{2}\right)\sqrt{2n_k \text{Log}(\text{Log}(n_k))}$ . Il vient :

$$P\left(S_{n_k - n_{k-1}} \geq \left(1 - \frac{\varepsilon}{2}\right)\sqrt{2n_k \text{Log}(\text{Log}(n_k))}\right) \geq \frac{1}{e} \sqrt{\frac{2}{\pi}} \exp\left(-\frac{1}{2} \left(\frac{\left(1 - \frac{\varepsilon}{2}\right)\varphi(n_k)}{\sqrt{n_k - n_{k-1}}} + 1\right)^2\right)$$

$$\geq \frac{1}{e} \sqrt{\frac{2}{\pi}} \exp\left(-\left(\frac{\left(1 - \frac{\varepsilon}{2}\right)\sqrt{n_k \text{Log}(\text{Log}(n_k))}}{\sqrt{n_k - n_{k-1}}} + \frac{1}{\sqrt{2}}\right)^2\right)$$

Or  $\frac{n_k - n_{k-1}}{n_k} = \frac{\gamma^k - \gamma^{k-1}}{\gamma^k} = \frac{\gamma - 1}{\gamma} > 1 - \frac{\varepsilon}{2}$ , d'après le choix de  $\gamma$ . Donc :

$$P\left\{S_{n_k - n_{k-1}} > \left(1 - \frac{\varepsilon}{2}\right)\sqrt{2n_k \text{Log}(\text{Log}(n_k))}\right\} \geq \frac{1}{e} \sqrt{\frac{2}{\pi}} \exp\left(-\left(\sqrt{\left(1 - \frac{\varepsilon}{2}\right) \text{Log}(\text{Log}(n_k))} + \frac{1}{\sqrt{2}}\right)^2\right)$$

**Lemme 8.** – Pour tout  $\varepsilon > 0$  et tout  $x \geq \frac{32}{\varepsilon^2}$ , on a :

$$\left(\sqrt{\left(1 - \frac{\varepsilon}{2}\right)x} + \frac{1}{\sqrt{2}}\right)^2 \leq \left(1 - \frac{\varepsilon}{4}\right)x$$

**Démonstration du lemme 8**

La condition  $\left(\sqrt{\left(1 - \frac{\varepsilon}{2}\right)x} + \frac{1}{\sqrt{2}}\right)^2 \leq \left(1 - \frac{\varepsilon}{4}\right)x$  équivaut à :

$$\sqrt{1 - \frac{\varepsilon}{2}} + \frac{1}{\sqrt{2}\sqrt{x}} \leq \sqrt{1 - \frac{\varepsilon}{4}}$$

ou encore :

$$\frac{1}{\sqrt{2x}} \leq \sqrt{1 - \frac{\varepsilon}{4}} - \sqrt{1 - \frac{\varepsilon}{2}} = \frac{\varepsilon/4}{\sqrt{1 - \frac{\varepsilon}{4}} + \sqrt{1 - \frac{\varepsilon}{2}}}$$

elle est satisfaite dès que  $\frac{1}{\sqrt{x}} \leq \frac{\varepsilon}{8}$ , ce qui prouve le lemme 8.

Revenons à la démonstration de la Proposition 7. D'après le Lemme 8, si  $\text{Log}(\text{Log}(n_k)) \geq \frac{32}{\varepsilon^2}$ , on a :

$$\begin{aligned} P \left\{ S_{n_k - n_{k-1}} > \left(1 - \frac{\varepsilon}{2}\right) \sqrt{2n_k \text{Log}(\text{Log}(n_k))} \right\} &\geq \frac{1}{e} \sqrt{\frac{2}{\pi}} e^{-\left(1 - \frac{\varepsilon}{4}\right) \text{Log}(\text{Log}(n_k))} \\ &= \frac{1}{e} \sqrt{\frac{2}{\pi}} (\text{Log}(n_k))^{-(1-\varepsilon/4)} \\ &= \frac{1}{e} \sqrt{\frac{2}{\pi}} (\text{Log}(\gamma))^{-(1-\varepsilon/4)} \frac{1}{k^{1-\varepsilon/4}} \end{aligned}$$

et ceci achève la démonstration de la Proposition 7.

La condition  $\text{Log}(\text{Log}(n_k)) \geq \frac{32}{\varepsilon^2}$  s'écrit, puisque  $n_k = \gamma^k$  :

$$\text{Log}(k) + \text{Log}(\text{Log}(\gamma)) \geq \frac{32}{\varepsilon^2}$$

Elle est satisfaite a fortiori dès que :

$$k \geq \exp(32 / \varepsilon^2)$$

Nous posons  $k_1 = \exp(32 / \varepsilon^2)$ .

**Proposition 9.** – Soit  $D_k$  comme précédemment ; notons  $D_k^c$  leurs complémentaires. Alors :

$$P \left( \bigcap_{k=k_1}^N D_k^c \right) \leq e^{-\sum_{k=k_1}^N u(k, \varepsilon)}$$

### Démonstration de la Proposition 9

Puisque les événements  $D_k$  sont indépendants, on a, pour tout  $N$  :

$$P \left( \bigcap_{k=k_1}^N D_k^c \right) = \prod_{k=k_1}^N (1 - P(D_k)) \leq e^{-\sum_{k=k_1}^N P(D_k)}$$

en utilisant l'inégalité  $1 - x \leq e^{-x}$ ,  $0 \leq x \leq 1$ . Cette inégalité peut être considérée comme une version quantitative du second lemme de Borel-Cantelli. L'estimation finale résulte immédiatement de la Proposition 7 ; ceci achève la démonstration de la Proposition 9.

Comme la série de terme général  $u(k, \varepsilon)$ , pour  $\varepsilon > 0$  fixé, est divergente, la somme  $\sum_{k=k_1}^N u(k, \varepsilon)$  peut être rendue arbitrairement grande, en choisissant  $N$  assez grand. Dans ces conditions,

$e^{-\sum_{k=k_1}^N u(k, \varepsilon)}$  sera proche de 0. Donc l'intersection  $\bigcap_{k=k_1}^N D_k^c$  aura une probabilité très petite. Mais

cette intersection est l'ensemble des chemins pour lesquels  $\left\{ \frac{S_{n_k} - S_{n_{k-1}}}{\varphi(n_k)} \leq 1 - \frac{\varepsilon}{2} \right\}$  pour tout  $k = k_1, \dots, N$ . Nous allons rendre cette démarche quantitative.

**Proposition 10.** – *On a :*

$$P\left(\frac{S_{n_k} - S_{n_{k-1}}}{\varphi(n_k)} \leq 1 - \frac{\varepsilon}{2}, \forall k = k_1, \dots, N\right) < \frac{\alpha}{2}$$

$$\text{dès que } N \geq N_0 = \left(k_1^{\frac{\varepsilon}{4}} + \frac{\varepsilon}{4C(\varepsilon)} \text{Log}\left(\frac{2}{\alpha}\right)\right)^{\frac{4}{\varepsilon}}$$

$$\text{où } C(\varepsilon) = \frac{1}{e} \sqrt{\frac{2}{\pi}} \frac{1}{\left(\text{Log}(16/\varepsilon^2)\right)^{1-\varepsilon/4}} \text{ et } k_1 = \exp(32/\varepsilon^2).$$

### Démonstration de la Proposition 10

D'après la Proposition 9, il suffit de choisir  $N$  assez grand pour que :

$$e^{-\sum_{k=k_1}^N u(k, \varepsilon)} < \frac{\alpha}{2}$$

C'est-à-dire :

$$\sum_{k_1}^N u(k, \varepsilon) > \text{Log}\left(\frac{2}{\alpha}\right)$$

D'après la définition de  $C(\varepsilon)$ , on a  $u(k, \varepsilon) = \frac{C(\varepsilon)}{k^{\frac{1-\varepsilon}{4}}}$ , et donc :

$$\sum_{k_1}^N u(k, \varepsilon) = C(\varepsilon) \sum_{k_1}^N k^{-1+\frac{\varepsilon}{4}} \geq \frac{4C(\varepsilon)}{\varepsilon} \left( N^{\frac{\varepsilon}{4}} - k_1^{\frac{\varepsilon}{4}} \right)$$

Il suffit donc que :

$$\frac{4C(\varepsilon)}{\varepsilon} \left( N^{\frac{\varepsilon}{4}} - k_1^{\frac{\varepsilon}{4}} \right) > \text{Log} \left( \frac{2}{\alpha} \right)$$

ou encore :

$$N \geq \left( k_1^{\frac{\varepsilon}{4}} + \frac{\varepsilon}{4C(\varepsilon)} \text{Log} \left( \frac{2}{\alpha} \right) \right)^{\frac{4}{\varepsilon}}$$

ce qui prouve la Proposition 10.

Achevons la démonstration du Théorème 4. On a :

$$\frac{S_{n_k}}{\varphi(n_k)} = \frac{S_{n_k} - S_{n_{k-1}}}{\varphi(n_k)} + \frac{S_{n_{k-1}}}{\varphi(n_{k-1})} \frac{\varphi(n_{k-1})}{\varphi(n_k)}.$$

Notons  $E_1$  l'ensemble :

$$E_1 = \left\{ Y_k < 1 - \frac{\varepsilon}{2}, \forall k = k_1, \dots, N \right\}.$$

Alors, d'après la Proposition 10,  $P(E_1) < \frac{\alpha}{2}$ .

Notons  $E_2$  l'ensemble :

$$E_2 = \left\{ \forall k \geq k_0, |Z_k| < \frac{\varepsilon}{2} \right\}$$

Alors  $P(E_2) > 1 - \frac{\alpha}{2}$ . On a :

$$P\left(\left(E_1^c \cap E_2\right)^c\right) = P\left(E_1 \cup E_2^c\right) \leq P\left(E_1\right) + P\left(E_2^c\right) = P\left(E_1\right) + 1 - P\left(E_2\right) < \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$$

Par conséquent :

$$P\left(E_1^c \cap E_2\right) \geq 1 - \alpha$$



Or l'ensemble  $E_1^c \cap E_2$  est caractérisé par le fait qu'il existe un  $k \leq N$  (plus précisément  $k_1 \leq k \leq N$ ) pour lequel  $Y_k \geq 1 - \frac{\varepsilon}{2}$  et pour lequel  $|Z_k| < \frac{\varepsilon}{2}$ . On en déduit :

$$Y_k + Z_k > Y_k - |Z_k| > 1 - \varepsilon.$$

Ceci achève la démonstration du Théorème 5.

## Remarque méthodologique

### Comparaison entre méthodes probabilistes et méthodes déterministes.

Nous avons vu la démonstration des Théorèmes 4 et 5, par des méthodes probabilistes vraiment difficiles ; à l'inverse, le Théorème 1 a été démontré par une méthode déterministe : dénombrement de chemins. Voyons si nous parvenons à démontrer le Théorème 1 par une méthode probabiliste. Nous cherchons simplement à démontrer qu'il est impossible que la marche aléatoire reste constamment au-dessous de l'axe des  $x$  :

$$P(S_k < 0, k = 1, \dots, n) \rightarrow 0, \text{ lorsque } n \rightarrow +\infty.$$

Fixons  $n$  entier quelconque. Nous observons que la loi de  $S_N$  connaissant la valeur de  $S_n$  est, lorsque  $N \rightarrow +\infty$ , pratiquement la même que celle de  $S_N$  sans information. Si nous jouons un million de parties, savoir que nous sommes partis avec 10 Euros ou avec 0 Euro ne change pas grand'chose. Par conséquent :

$$\frac{P(S_N < 0 \text{ et } S_n < 0)}{P(S_n < 0)} = P(S_N < 0 | S_n < 0) \approx P(S_N < 0) < \frac{1}{2}$$

et donc, si  $n_2$  est grand devant  $n_1$  :

$$P(S_{n_1} < 0 \text{ et } S_{n_2} < 0) < \frac{1}{2} P(S_{n_1} < 0) < \frac{1}{4}$$

De la même façon, si  $n_3$  est grand devant  $n_2$  :

$$P(S_{n_3} < 0 | S_{n_2} < 0 \text{ et } S_{n_1} < 0) \approx P(S_{n_3} < 0) < \frac{1}{2}$$

et donc :

$$P(S_{n_3} < 0 \text{ et } S_{n_2} < 0 \text{ et } S_{n_1} < 0) < \frac{1}{2} P(S_{n_2} < 0 \text{ et } S_{n_1} < 0) < \frac{1}{2^3}$$

Nous obtenons donc le résultat suivant : si la suite  $n_j$  est suffisamment croissante :

$$P(S_{n_j} < 0, j = 1, \dots, k) < \frac{1}{2^k}$$

Ceci répond à notre question, car :

$$P(S_l < 0, l = 1, \dots, n_k) < P(S_{n_j} < 0, j = 1, \dots, k)$$

Les chemins qui sont constamment au dessous de l'axe des  $x$  sont évidemment moins nombreux que ceux qui sont au-dessous de l'axe des  $x$  aux points désignés  $n_1, \dots, n_k$ .

Cette remarque illustre bien la différence entre méthodes déterministes et méthodes probabilistes. Ici, la démonstration probabiliste est très simple, mais ne donne pas un résultat quantitatif aussi précis que la méthode de dénombrement.

## Références

[MPPR] Bernard Beauzamy : Méthodes probabilistes pour l'étude des phénomènes réels, ISBN : 2-9521458-0-6, Editions de la SCM, mars 2004.

J.C. Kalbfleisch : Probability and Statistical Inference, volume 1 : Probability. Springer Texts in Statistics, 1985.

A. Khinchine. "Über einen Satz der Wahrscheinlichkeitsrechnung", Fundamenta Mathematica, 6:9-20, 1924.

## Annexe

Démonstration de l'inégalité :

$$P(|S_n| \geq x) \leq e^{-x^2/(2n)}$$

On remarque d'abord que  $E(S_n) = 0$  et  $\text{var}(S_n) = n$ . On écrit :

$P(S_n \geq x) = P(e^{\lambda S_n} \geq e^{\lambda x}) \leq e^{-\lambda x} E(e^{\lambda S_n})$  d'après l'inégalité de Markov :

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}$$

On a en outre :

$$e^{-\lambda x} E(e^{\lambda S_n}) = e^{-\lambda x} E\left(\prod_1^n e^{\lambda X_k}\right) = e^{-\lambda x} (E e^{\lambda X_1})^n$$

Or :

$$E(e^{\lambda X_1}) = \frac{e^{-\lambda} + e^{\lambda}}{2} \leq e^{\lambda^2/2} \quad (1)$$

Démontrons (1). On a :

$$\frac{e^{-\lambda} + e^{\lambda}}{2} = \sum_{k=0}^{+\infty} \frac{\lambda^{2k}}{(2k)!}, \quad e^{\lambda^2/2} = \sum_{k=0}^{+\infty} \frac{\lambda^{2k}}{2^k k!}$$

Mais  $2^k k! \leq (2k)!$  et (1) est démontré.

On déduit de (1), pour tout  $\lambda$  :

$$P(S_n \geq x) \leq e^{-\lambda x} e^{n\lambda^2/2}$$

et le choix de  $\lambda = \frac{x}{n}$  donne l'estimation voulue.