Lycée Pasteur de Calgary
Introduction à l'exposé de Marie Théret
Hammersley, feux de forêt, porosité et réseaux
à la BNF, conférences *Un texte, un mathématicien.*
14 avril 2021.

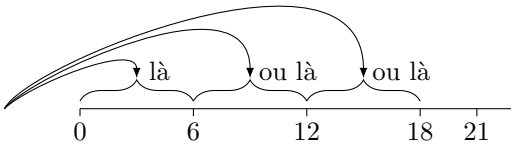
Notes d'exposé

1 Introduction : de la modélisation aléatoire.

Comment représenter le réel et toutes ses occurrences possibles. Un aspect de la gestion de la complexité. Une possibilité souvent adéquate : faire intervenir des probabilités. Aboutit à la construction d'objets mathématiques à la fois très élaborés et très "naturels". Evoquer la modélisation des gaz en physique statistique et le mouvement brownien.

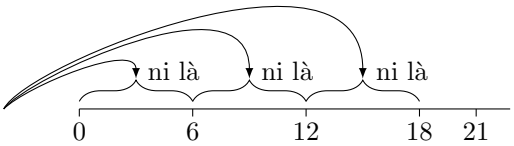
2 Du singe

Jeu de pile ou face équilibré. On écrit le mot des tirages successifs : $M = PPFPFFFFFPF\cdots$ Avec quelle probabilité le motif $\mu = PFFPPF$ apparaît-il avant le $21^{\text{ième}}$ lancer ? Le calcul exact n'est pas si simple, mais on peut faire une minoration (assez) grossière :

si μ est 

, alors μ apparaît avant le $21^{\text{ième}}$ lancer.

Autrement dit,

si μ n'apparaît pas avant le $21^{\text{ième}}$ lancer, alors μ n'est 

En termes de probabilités, cela se traduit par

$$1 - \mathbf{P}\left(\mu \text{ apparaît avant le } 21^{\text{ième}} \text{ lancer}\right) \leq \left(1 - \frac{1}{2^6}\right)^3 .$$

Plus généralement, on montre par le même raisonnement que si on fabrique un mot très long en tirant des lettres aléatoires indépendantes et uniformément équidistribuées dans un alphabet de L lettres, la probabilité qu'apparaisse un motif μ donné, de longueur m, avant le $N^{\text{ième}}$ tirage vérifie

$$\mathbf{P}\left(\mu \text{ apparaît avant le } N^{\text{ième}} \text{ tirage}\right) \geq 1 - \left(1 - \frac{1}{L^m}\right)^{\lfloor \frac{N}{m} \rfloor} ,$$

où $\lfloor x \rfloor$ désigne la partie entière du réel x (si $N = mq + r$ est la division euclidienne de N par m, alors $q = \lfloor \frac{N}{m} \rfloor$).

Prenons alors le poème “Marthe” de René Char♫, extrait du recueil *Fureur et mystère* :

> “Marthe que ces vieux murs ne peuvent pas s’approprier, fontaine où se mire ma monarchie solitaire, comment pourrais-je jamais vous oublier puisque je n’ai pas à me souvenir de vous : vous êtes le présent qui s’accumule. Nous nous unirons sans avoir à nous aborder, à nous prévoir comme deux pavots font en amour une anémone géante. Je n’entrerai pas dans votre cœur pour limiter sa mémoire. Je ne retiendrai pas votre bouche pour l’empêcher de s’ouvrir sur le bleu de l’air et la soif de partir. Je veux être pour vous la liberté et le vent de la vie qui passe le seuil de toujours avant que la nuit ne devienne introuvable.”

Oublions les accents et la ponctuation, mais pas les espaces : on considère ce poème comme écrit dans un alphabet de 27 lettres. Il est alors constitué de 625 lettres. Prenons aussi un singe et installons-le devant une machine à écrire, ou un clavier d’ordinateur qui n’aurait que les touches de ces 27 lettres. On considère qu’il écrit un texte en frappant les touches au hasard. La probabilité qu’il écrive le poème “Marthe” avant la $N^{\text{ième}}$ lettre est alors supérieure à $1-\left(1-\frac{1}{27^{625}}\right)^{\lfloor \frac{N}{625} \rfloor}$. Comme $1-\frac{1}{27^{625}}<1$, cette probabilité tend vers 1 lorsque N tend vers $+\infty$.

Ainsi, avec probabilité 1*, le singe écrira “Marthe” en un temps fini.*

Par exemple, notre calcul garantit que pour que cette probabilité soit strictement supérieure $\frac{1}{2}$, il suffit que $\left(1-\frac{1}{27^{625}}\right)^{\lfloor \frac{N}{625} \rfloor}<\frac{1}{2}$. On passe au logarithme : il suffit que

$$N > 625 \times \frac{-\ln 2}{\ln\left(1-\frac{1}{27^{625}}\right)} \approx 625 \times \ln 2 \times 27^{625} \approx 10^{900}$$

pour que la probabilité que le singe écrive “Marthe” soit supérieure à $\frac{1}{2}$. Hum, il faut attendre longtemps longtemps pour que cela se passe ! Pour donner une idée, on estime l’âge de l’Univers à 13 milliards d’années. Non, décidément, écrire au hasard n’est pas une bonne méthode de production littéraire.

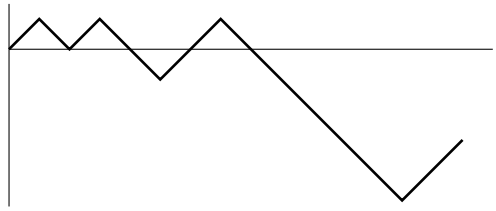
Au delà de l’anecdote, ce type de calcul a des conséquences sur l’écriture décimale des nombres ou sur la philosophie des sciences (théorie de l’évolution, par exemple : le “hasard”, au sens de l’équidistribution et de l’indépendance, ne suffit pas à expliquer l’émergence de la vie).

3 Marche aléatoire

On prend un autre modèle : une particule se déplace sur $\mathbb{Z}$, aux instants 0, 1, 2, etc, de la façon suivante : elle part de 0 puis, à chaque instant, si elle est au point $x \in \mathbb{Z}$, elle saute en $x+1$ avec probabilité $\frac{1}{2}$ et en $x-1$ avec probabilité $\frac{1}{2}$, sans que ces probabilités ne dépendent de rien d’autre (établir le lien avec pile ou face, dessin).

On se pose la question de la *récurrence* de cette marche : avec quelle probabilité la marche revient-elle en 0 (en un temps fini) ?

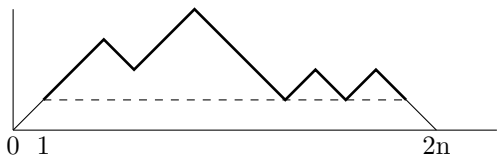
On note S_n la position de la particule au temps n. Ainsi, $S_0=0$, $S_1=1$ avec probabilité $\frac{1}{2}$, etc. On représente la marche par le graphe d’une fonction : en abscisse, le temps n, en ordonnée, la position S_n et on interpole ces points par des segments de droite. Exemple du chemin dont les tirages successifs sont *hbhbbhhbbbbbbhh* (h pour *haut* qui correspond au saut $x \rightsquigarrow x+1$ et b pour *bas* qui correspond au saut $x \rightsquigarrow x-1$). Son point d’arrivée est $S_{15}=-3$.



♫René Char, 1907 – 1988.

Si $n \geq 1$, on note p_n la probabilité pour que, pour la première fois, la marche retourne en 0 à l'instant n. On calcule les premières valeurs en dessinant les chemins qui conviennent : $p_1 = 0$, $p_2 = \frac{2}{2^2} = \frac{1}{2}$, $p_3 = 0$, $p_4 = \frac{2}{2^4} = \frac{1}{8}$, $p_5 = p_7 = 0$ (comme tous les p_{2n+1}), $p_6 = \frac{4}{2^6} = \frac{1}{16}$, $p_8 = \frac{10}{2^8} = \frac{5}{128}$, etc. On remarque que les nombres de chemins convenables sont pairs par symétrie par rapport à l'axe des abscisses (si un chemin convient, son symétrique aussi).

Un *chemin de Dyck*♫ à n pas est une graphe de ce type joignant $(0,0)$ à $(n,0)$ en restant positif. On note D_n le nombre de chemins de Dyck à n pas. Bien sûr, si n est impair, $D_n = 0$. Une marche dont le premier pas est $+1$ revient en 0 pour la première fois à l'instant $2n$ si, et seulement si le graphe de la marche entre les instants 1 et $2n-1$ est un chemin de Dyck à $2n-2$ pas.



Ainsi, la probabilité que la marche retourne à l'origine pour la première fois à l'instant $2n$ est

$$p_{2n} = \frac{2D_{2n-2}}{2^{2n}}.$$

Calculer cette probabilité revient donc à calculer le nombre de chemins de Dyck à n pas. On pourrait le faire et donner une formule explicite à l'aide des nombres de Catalan♫ ou des coefficents du binôme, mais on n'en a pas ici besoin.
On cherche à savoir avec quelle probabilité la marche aléatoire revient en 0 en un temps fini. Comme les p_n mesurent les probabilités d'événements disjoints, il s'agit de calculer la somme infinie

$$p_1 + p_2 + p_3 + p_4 + \cdots = p_2 + p_4 + p_6 + \cdots = \sum_{n\geq 1} p_{2n}.$$

[Commentaire sur le sens de cette somme, limite d'une suite croissante – on ajoute des nombres positifs – et majorée – par 1, c'est une probabilité.] Pour cela, on note

$$D(x) = D_0 + D_1 x + D_2 x^2 + D_3 x^3 + D_4 x^4 + \cdots = \sum_{n\geq 0} D_n x^n.$$

Puisqu'un chemin de Dyck a un nombre pair de pas (il retourne en 0), cette somme s'écrit à vrai dire

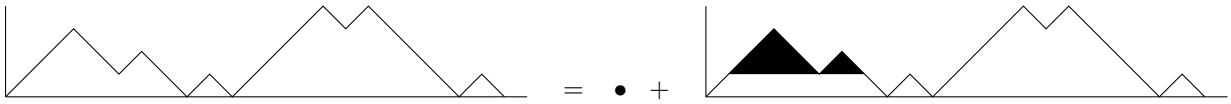
$$D(x) = D_0 + D_2 x^2 + D_4 x^4 + D_6 x^6 \cdots = \sum_{n\geq 0} D_{2n} x^{2n}.$$

[Commentaire sur cette nouvelle somme infinie. Si $x \in [0, \frac{1}{4}]$, elle est encore définie comme tout à l'heure puisque remplacer x par $\frac{1}{2}$ dans la somme donne comme plus haut une majoration des sommes finies.] On peut alors écrire la somme cherchée de la façon suivante :

$$\sum_{n\geq 1} p_n = \sum_{n\geq 1} p_{2n} = 2\sum_{n\geq 1} \frac{D_{2n-2}}{2^{2n}} = \frac{1}{2}\sum_{n\geq 0} \frac{D_{2n}}{2^{2n}} = \frac{1}{2} D\left(\frac{1}{2}\right). \tag{1}$$

La somme $D(x)$, fonction de x lorsque x est assez petit, s'écrit explicitement. Les arguments qui suivent, qui sont ici évoqués sans être bien justifiés, trouvent une légitimation entière en mathématiques, pourvu qu'on travaille un peu plus.
Un chemin de Dyck peut se décomposer comme le dessin qui suit l'illustre, en isolant le premier instant de retour à 0 (comme dans notre affaire de marche).

♫Walther von Dyck, 1856 – 1934.
♫Eugène Charles Catalan, 1814 – 1894.



Cette décomposition se traduit sur la fonction D par $D(x) = 1 + [xD(x)x] \times D(x)$, c'est-à-dire par

$$x^2 D^2 - D + 1 = 0.$$

On voit que $D(x)$ est solution d'une équation de degré 2 que l'on résout (en écartant l'autre racine par la valeur en 0 qui doit être finie) :

$$D(x) = \frac{1 - \sqrt{1 - 4x^2}}{2x^2}.$$

En particulier, $D\left(\frac{1}{2}\right) = 2$, ce qui montre avec (1) que la probabilité que la marche retourne au point d'origine en un temps fini égale 1. On dit que la marche est *récurrente*. Elle repasse par l'origine une infinité de fois avec probabilité 1.

On pourrait étudier une marche de ce type sur $\mathbb{Z}^2$, $\mathbb{Z}^3$, ou plus généralement sur $\mathbb{Z}^d$. On obtiendrait le résultat suivant, dû à Pólya♫ : sur $\mathbb{Z}^2$, la marche est encore récurrente (elle retourne au point de départ avec probabilité 1). En revanche, si $d \geq 3$, la marche n'est plus récurrente. On montre que la probabilité pour qu'elle revienne au point de départ est $\approx 0,34$ pour $d = 3$, $\approx 0,19$ pour $d = 4$, $\approx 0,10$ pour $d = 6$, $\approx 0,07$ pour $d = 8$, etc.

On peut aussi étendre le modèle en modifiant l'aléa (abandonner l'équidistribution, ou l'indépendance), en remplaçant $\mathbb{Z}^d$ par un graphe qui peut lui-même devenir aléatoire, etc. Ce type de sujet est très actif dans la recherche mathématique contemporaine.

4 Percolation de premier passage

Le modèle est le suivant (dessin) : à chaque arête a du graphe $\mathbb{Z}^2$, on associe un nombre réel positif ou nul T_a, aléatoire (les T_a sont indépendants et tous de même loi dont l'espérance est finie). Si $c = a_1 a_2 \ldots a_n$ est un chemin (une suite d'arêtes consécutives), le temps de passage $T(c)$ est le nombre (aléatoire)

$$T(c) = T_{a_1} + T_{a_2} + \cdots + T_{a_n}.$$

Si $(x, y) \in \mathbb{Z}^2$, la *distance* (de percolation) de $(0, 0)$ à (x, y) est définie comme

$$D(x, y) = \inf_{\{c,\ \text{chemin } (0,0) \to (x,y)\}} T(c).$$

On peut voir cela comme un modèle de propagation d'un liquide en milieu poreux – on injecte le liquide en $(0, 0)$ et $D(x, y)$ mesure le premier instant où le point (x, y) est mouillé – ou d'un feu de forêt en Australie, ou d'un coronavirus sur la planète.

Les questions qu'on peut se poser : quelle est la distance de l'origine à un point très éloigné ? Sa moyenne ? Sa distribution ? Quelle est la forme d'une boule pour cette distance ? D'une grande boule ? A quoi ressemblent les géodésiques (les chemins qui réalisent le plus court temps de passage) ? Etc.

Les inventeurs de ce modèle passionnant sont (probablement) Hammersley♫ et Welsh♫, dont Marie Théret parlera.

Un aperçu très rapide de premiers résultats connus : on appelle X_n la distance de l'origine au point $(n, 0)$. C'est une variable aléatoire positive ou nulle. Un résultat en moyenne : la suite réelle $E(X_n)/n$ admet une limite qu'on appelle μ. Ce nombre dépend du temps de passage (aléatoire) des arêtes qu'on a noté T_a. On le voit, la propagation, en moyenne, est au plus proportionnelle au temps. Mais elle peut être beaucoup plus lente, lorsque $\mu = 0$. Kesten♫ a montré, par exemple, que $\mu = 0$ si, et seulement si $\mathbf{P}(T_a = 0) > \frac{1}{2}$. Il reste énormément de questions ouvertes sur ce modèle pourtant très simple et très utile.

♫George Pólya, 1887 – 1985.
♫John Michael Hammersley, 1920 – 2004
♫James Anthony Dominic Welsh, né en 1938
♫Harry Kesten, 1931 – 2019