Peut-on imiter le hasard ?

Nicole Vogel

Depuis que statistiques et probabilités ont pris une large place dans les programmes de mathématiques, on nous propose souvent de petites expériences pour tester notre perception du hasard. Mais cela se termine généralement par l'énoncé d'une propriété sans démonstration, ce qui peut laisser penser qu'il n'en existe pas de preuve élémentaire. J'ai choisi de présenter ici une de ces situations et d'en proposer une démonstration utilisant le cours de spécialité de terminale ES, en particulier les graphes.

A - Une expérience

Demandez à quelques personnes de lancer cent fois une pièce de monnaie et de noter la suite des résultats des cent lancers en indiquant P pour pile et F pour face.

Évidemment, elles peuvent remplacer la pièce par un dé (en général plus facile à lancer qu'une pièce) et noter P lorsque le résultat est pair et F lorsque le résultat est impair.

Demandez à quelques autres personnes de noter une suite inventée de cent P ou F imitant les résultats de cent lancers de pièce.

Comparez ensuite les résultats des deux groupes…

Recherchez dans chaque suite la séquence la plus longue, c'est-à-dire le nombre maximum de résultats identiques qui se suivent, et notez sa longueur.

Par exemple, dans :

F F P F F F F P F P P F P F P P P F F P F P F F P P P P F F P F F F F P P P F F F
P F F P F P P F F P F F P P F F F P F F F F F F F F F F F P P P F F P P P F P P F
P P P P P F F P F P F F P F F F F P F,
la plus longue séquence est F F F F F F F F F F F. Sa longueur est 11.

Vous constaterez que dans le groupe qui a imité le hasard, les séquences maximales sont en général nettement plus courtes que dans le groupe qui a réellement fait l'expérience (à condition bien sûr que les imitateurs ne soient pas des statisticiens qu'on ne piège pas aussi facilement !).

On peut même adopter le critère suivant pour distinguer une suite expérimentale d'une suite inventée : s'il n'y a pas de séquence de longueur au moins 6, on se trompe peu en décidant que la suite est inventée (surtout si elle n'a pas non plus de séquence de longueur 5) et s'il y a une séquence de longueur 6 ou plus, on peut penser que la suite est expérimentale.

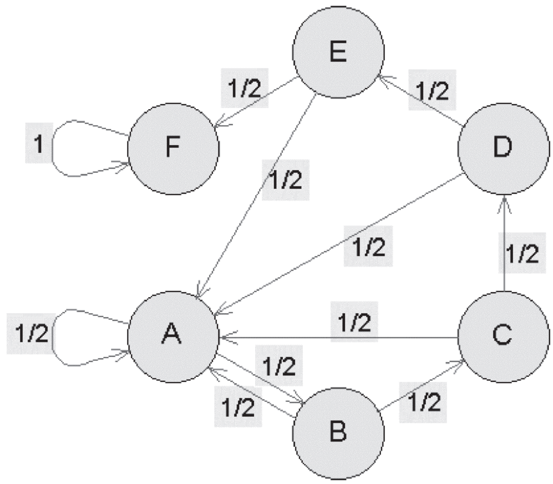
Les imitateurs ont une mauvaise intuition du hasard : ils imaginent qu'une suite au hasard distribue assez régulièrement les piles et les faces, ils n'osent donc pas mettre 6 résultats identiques consécutifs.

Nous allons voir que la probabilité d'avoir une séquence de longueur au moins 6 dans une suite de cent lancers de pièce au hasard est d'environ 81 %, et que la probabilité d'avoir une séquence de longueur au moins 5 est d'environ 97 %…

B - Une démonstration

Pour modéliser cela, considérons six états possibles d'une suite de n lancers de pièce :
– État A : elle ne contient pas de séquence de longueur au moins 6 et finit par une séquence de longueur 1.
– État B : elle ne contient pas de séquence de longueur au moins 6 et finit par une séquence de longueur 2.
– État C : elle ne contient pas de séquence de longueur au moins 6 et finit par une séquence de longueur 3.
– État D : elle ne contient pas de séquence de longueur au moins 6 et finit par une séquence de longueur 4.
– État E : elle ne contient pas de séquence de longueur au moins 6 et finit par une séquence de longueur 5.
– État F : elle contient une séquence de longueur au moins 6.
Ces six états forment une partition de l'ensemble des situations possibles pour cette suite.

Les changements d'états lorsqu'on relance la pièce pour passer d'une suite de n lancers à une suite de $n + 1$ lancers peuvent alors être résumés par le graphe probabiliste suivant :



Par exemple, lorsqu'une suite est à l'état C, elle finit par une séquence de longueur 3.

Lorsqu'on relance la pièce, elle tombe ou bien du même côté que dans cette séquence avec une probabilité 1/2 et passe à l'état D, ou bien du côté opposé avec une probabilité 1/2 et repasse à l'état A.

La matrice de transition d'une suite de n lancers à une suite de $n + 1$ lancers est donc :

→	A	B	C	D	E	F
A	0.5	0.5	0	0	0	0
B	0.5	0	0.5	0	0	0
C	0.5	0	0	0.5	0	0
D	0.5	0	0	0	0.5	0
E	0.5	0	0	0	0	0.5
F	0	0	0	0	0	1

Posons

$$M = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Notons a_n, b_n, c_n, d_n, e_n, f_n les probabilités respectives qu'une suite de lancers soit à l'état A, B, C, D, E, F après n lancers.
On a donc :

$$(a_{n+1}, b_{n+1}, c_{n+1}, d_{n+1}, e_{n+1}, f_{n+1}) = (a_n, b_n, c_n, d_n, e_n, f_n) \times M.$$

D'où :

$$(a_n, b_n, c_n, d_n, e_n, f_n) = (a_1, b_1, c_1, d_1, e_1, f_1) \times M^{n-1}.$$

Or après un lancer, la suite est toujours à l'état A. Donc

$$a_1 = 1$$

et

$$b_1 = c_1 = d_1 = e_1 = f_1 = 0.$$

On en déduit :

$$(a_n, b_n, c_n, d_n, e_n, f_n) = (1, 0, 0, 0, 0, 0) \times \mathrm{M}^{n-1}.$$

En particulier :

$$(a_{100}, b_{100}, c_{100}, d_{100}, e_{100}, f_{100}) = (1, 0, 0, 0, 0, 0) \times \mathrm{M}^{99}.$$

Le calcul avec une calculatrice (ou avec un logiciel de mathématiques) donne :
$(a_{100}, b_{100}, c_{100}, d_{100}, e_{100}, f_{100})$

$$\approx (1,0,0,0,0,0) \times \begin{pmatrix} 0.0982 & 0.0499 & 0.0254 & 0.0129 & 0.00657 & 0.806 \\ 0.0948 & 0.0482 & 0.0245 & 0.0124 & 0.00635 & 0.813 \\ 0.0883 & 0.0449 & 0.0228 & 0.0116 & 0.00591 & 0.826 \\ 0.0753 & 0.0383 & 0.0195 & 0.00992 & 0.00504 & 0.851 \\ 0.0499 & 0.0254 & 0.0129 & 0.00657 & 0.00334 & 0.901 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

On en déduit :

$$f_{100} \approx 0{,}806.$$

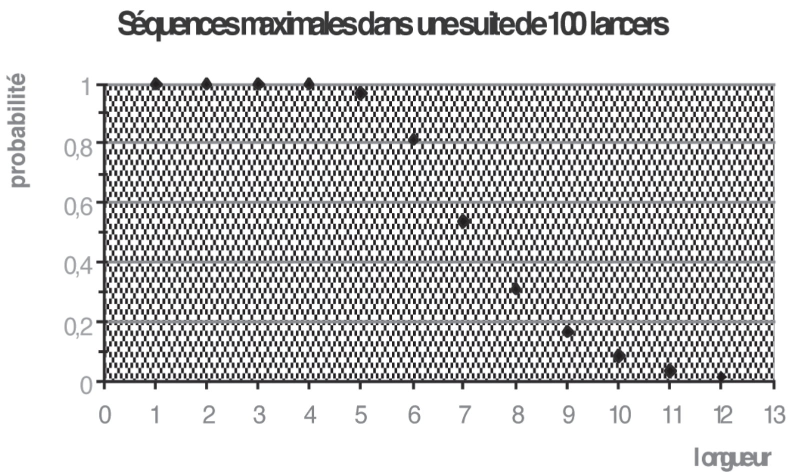
Cela signifie que la probabilité de trouver une séquence de longueur au moins 6 dans une suite de 100 lancers est environ égale à 81 %.
Un raisonnement analogue montre que la probabilité de trouver une séquence de longueur au moins 5 dans une suite de 100 lancers est environ égale à 97 % car :

$$\begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}^{99} \approx \begin{pmatrix} 0.0146 & 0.00761 & 0.00395 & 0.00205 & 0.971 \\ 0.0136 & 0.00706 & 0.00366 & 0.00190 & 0.973 \\ 0.0115 & 0.00600 & 0.00311 & 0.00161 & 0.977 \\ 0.00761 & 0.00395 & 0.00205 & 0.00106 & 0.985 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Le calcul fournit les résultats suivants pour une suite de 100 lancers :

longueur de la plus longue séquence : au moins	4	5	6	7	8	9	10	11	12
probabilité arrondie au centième	1,00	0,97	0,81	0,54	0,31	0,17	0,09	0,04	0,02

Ces valeurs donnent le graphique ci-dessous(1) :



On voit que la forte probabilité de certains événements est peu intuitive...

On se rend ainsi compte qu'il n'est pas du tout évident pour un être humain d'imiter le hasard.

(1) Sur l'intervalle [4, 12] cette courbe évoque, dans un contexte très différent du théorème de la limite centrée et de la densité de Laplace-Gauss, une (demi-)courbe en cloche. On observe, comme dans de nombreuses situations (dates d'anniversaire, battage de cartes, …), un phénomène de « rupture » : la probabilité chute de 0,9 à 0,1 sur le court (par rapport à 100) intervalle [5, 10].