Lancer de pièces: nombre maximum de résultats consécutifs identiques

28 février 2021

Le lancer d'une pièce de monnaie peut être assimilé à une suite finie (X_n) de variables aléatoires indépendantes sur $\{0,1\}$ telle que $p(X_n = 0) = \frac{1}{2}$ et $p(X_n = 1) = \frac{1}{2}$.

Soit (Y_n) la suite définie par :

$$Y_1 = 1$$
$$Y_n = \begin{cases} N & \text{si } Y_{n-1} = N \\ Y_{n-1}+1 & \text{si } Y_{n-1} \neq N \text{ et } X_n = X_{n-1} \\ 1 & \text{si } Y_{n-1} \neq N \text{ et } X_n \neq X_{n-1} \end{cases}$$

(Y_n) est à valeurs dans $\{1,2,\ldots,N\}$ et Y_n est ne dépend que de l'état antérieur Y_{n-1} puisque (X_n) est une suite de variables aléatoires indépendantes ; une telle suite est encore appelée chaîne de Markov.

Ce qui nous intéresse ici c'est $p(Y_n = N)$. Ce n'est autre que la probabilité pour que le nombre maximum de coups consécutifs égaux soit supérieur ou égal à N. Il est clair que N ne peut être supérieur au nombre total de lancers.

Première approche par les matrices de transition

On a $p(Y_1 = 1) = 1$ et on note $t(i,j)$, la probabilité conditionnelle $p(Y_n = j|Y_{n-1} = i)$.
t est la matrice de transition et on a pour $i = 1,\ldots,N-1$:
$t(i,1) = p(Y_n = 1|Y_{n-1} = i) = \frac{1}{2}$ (3^{e} cas de la définition de Y_n),
$t(i,i+1) = \frac{1}{2}$.
De plus : $t(N,j) = 0$ pour $j \neq N$ et $t(N,N) = 1$.

Par exemple pour $N = 4$ la matrice de transition sera : $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

On démontre alors que :

$$p(Y_1 = 1, Y_2 = k_2, Y_3 = k_3, \ldots, Y_n = N) = t(1,k_2)\,t(k_2,k_3)\ldots t(k_{n-1},N)$$

et par suite que :

$$p(Y_n = N) = \sum_{k_2=1}^{N}\left(t(1,k_2)\sum_{k_3=1}^{N}\left(t(k_2,k_3)\ldots\sum_{k_{n-1}=1}^{N} t(k_{n-1},N)\right)\right)$$

Utilisation d'une suite numérique

Posons par exemple $N = 6$ et $p_n = p(Y_n = 6)$.
Il est clair que l'on a : $p_1 = p_2 = p_3 = p_4 = p_5 = 0$ et $p_6 = \dfrac{2}{2^6}$.
Si on lance n fois la pièce avec $n > 6$ alors $Y_n = 6$ se produit dans les cas suivants :

- lorsque $Y_{n-1} = 6$
- dans l'un des deux cas suivants :
  - on vient d'obtenir 6 fois 1 (événement A_n) et $X_{n-6} = 0$ et $Y_{n-6} < 6$
  - on vient d'obtenir 6 fois 0 (événement B_n) et $X_{n-6} = 1$ et $Y_{n-6} < 6$

D'après la formule des probabilités totales on a donc pour tout $n > 6$:

$$p_n = p(Y_{n-1} = 6) + p(A_n \cap \{X_{n-6} = 0\} \cap \{Y_{n-6} < 6\}) + p(B_n \cap \{X_{n-6} = 1\} \cap \{Y_{n-6} < 6\})$$

Comme A_n est indépendant de $\{X_{n-6} = 0\} \cap \{Y_{n-6} < 6\}$ et B_n est indépendant de $\{X_{n-6} = 1\} \cap \{Y_{n-6} < 6\}$:

$$p_n = p_{n-1} + \frac{1}{2^6} \times \left(p(\{X_{n-6} = 0\} \cap \{Y_{n-6} < 6\}) + p(\{X_{n-6} = 1\} \cap \{Y_{n-6} < 6\})\right)$$

Or $p(\{X_{n-6} = 0\} \cap \{Y_{n-6} < 6\}) + p(\{X_{n-6} = 1\} \cap \{Y_{n-6} < 6\}) = p(\{Y_{n-6} < 6\}) = 1 - p_{n-6}$.
Pour tout $n > 6$ on a donc :

$$p_n = p_{n-1} + \frac{1}{2^6}(1 - p_{n-6})$$

On peut ainsi calculer de proche en proche p_n pour $n > 6$ sur tableur ou calculatrice. On trouve les valeurs suivantes :

n	10	20	50	100	150	200
p_n	0,094	0,237	0,544	0,806	0,918	0,965

La probabilité pour que le nombre maximum de résultats consécutifs identiques soit égal à N s'obtient alors par soustraction.

Remarque : (p_n) est une suite croissante majorée par 1, elle converge donc vers une limite l vérifiant $l = l + \dfrac{1}{2^6}(1 - l)$. Ce qui conduit à $l = 1$. En d'autres termes, la probabilité d'obtenir au moins six résultats consécutifs égaux tend en croissant vers 1 quand n tend vers $+\infty$.

Simulation

La méthode la plus naturelle consisterait à mettre en évidence les 2^n suites différentes et équiprobables de 0 et de 1 que l'on obtient lorsqu'on lance n fois une pièce. Il suffit ensuite de rechercher à chaque fois le nombre maximum de résultats consécutifs identiques. Il reste que 2^n devient beaucoup trop grand et il est préférable de chercher ce nombre maximum sur un échantillon de taille 1000 par exemple obtenu par simulation. L'erreur commise reste faible si la taille de cet échantillon est assez grande.