Un exemple de marche aléatoire

Yves Coudène, 1/9/04

Certains problèmes de la vie courante peuvent être modélisés par des marches aléatoires. On montre comment, par des raisonnements élémentaires, on peut étudier le comportement d'une de ces marches.

1 Le problème du collectionneur

Nicolas, 10 ans, se lance dans une collection de cartes Pokémon. Chaque carte représente un Pokémon, et il existe 150 Pokémons différents. Les cartes sont vendues dans des paquets scellés, si bien que lorsque Nicolas achète une carte, il peut tomber sur n'importe quel Pokémon ; en particulier sur un Pokémon qu'il possède déjà.
En supposant qu'à chaque achat, on ait même probabilité d'obtenir chacun des 150 Pokémons, quel est le nombre moyen de cartes que Nicolas doit acheter afin d'obtenir tous les Pokémons ?

Un raisonnement simple permet d'obtenir la réponse. Traitons le cas général où le nombre total de Pokémons est quelconque, égal à n. Notons t_k le nombre moyen de cartes à se procurer pour terminer la collection, sachant que l'on est déjà en possession de k Pokémons distincts.

Supposons que nous avons k Pokémons, tous différents, et achetons une carte supplémentaire. Avec probabilité $p_k = k/n$, cette carte était déjà en notre possession ; il faut donc encore acheter en moyenne t_k cartes pour terminer la collection. Avec probabilité $1 - p_k$, c'est une nouvelle carte ; il faut donc acheter en moyenne t_{k+1} cartes supplémentaires. Par conséquent :

$$t_k = p_k t_k + (1 - p_k) t_{k+1} + 1$$

Ce qui donne : $t_k = t_{k+1} + \dfrac{1}{(1 - p_k)}$, soit $t_k = t_n + \displaystyle\sum_{l=k}^{n-1} \frac{n}{n-l}$.
Le nombre de cartes nécessaires pour terminer la collection, sachant que l'on possède déjà n Pokémons distincts, est égal à zéro : $t_n = 0$. Le nombre moyen total de cartes à se procurer pour posséder l'ensemble des n Pokémons est égal à t_0, avec :

$$t_0 = n \sum_{j=1}^{n} \frac{1}{j}$$

Une comparaison série-intégrale, appliquée à la fonction $1/x$, donne l'encadrement $\ln(n) < \sum_{j=1}^{n} 1/j < \ln(n) + 1$. La différence $\sum 1/j - \ln(n)$ est décroissante ; elle converge donc vers une certaine constante γ, appelée constante d'Euler. Si bien que le nombre recherché est de l'ordre de $n(\ln(n) + \gamma)$. En première approximation, on peut prendre $\gamma = 0.577$. Pour $n = 150$, on obtient $t_0 \simeq 838$. Nicolas peut espérer terminer sa collection après avoir acheté 838 cartes.

2 Interprétation en terme de marche aléatoire

Cherchons à formaliser les raisonnements précédents. Pour cela, il faut se donner un univers, composé des résultats qui peuvent être obtenus à l'issue de l'épreuve, d'une probabilité sur cet ensemble, et d'une variable aléatoire qui représente la quantité à étudier.

L'univers Ω

Dans chacun des deux exemples, on peut modéliser la situation par une suite de nombres $X_0, X_1...$ compris entre 0 et n. Dans le premier exemple, le terme de rang i de la suite, noté X_i, correspond au nombre de Pokémons différents en possession de Nicolas, après l'achat de i cartes.

L'univers Ω de tous les résultats possibles correspond donc à l'ensemble des suites indexées par les entiers, à valeurs dans $\{0..n\}$:

$$\Omega = \{0..n\}^{\mathbf{N}}$$

La variable aléatoire $X_i : \Omega \to \mathbf{R}$ est donnée par la projection sur la i^{eme} coordonnée.

La probabilité P

Cherchons à définir sur Ω une probabilité P rendant compte du phénomène étudié. Rappelons que le cylindre $[X_0 = k_0, X_1 = k_1, ..X_l = k_l]$ est le sous-ensemble de Ω composé des suites qui débutent par $k_0, ..k_l$. Pour définir une probabilité P sur Ω, il suffit de spécifier sa valeur sur les cylindres $[X_0 = k_0, X_1 = k_1, ..X_l = k_l]$, et ce pour tout l et tout l-uplet $k_0, ..k_l$.
Nous allons définir la probabilité P par une récurrence sur la taille des cylindres. Pour cela, rappelons la notion de *probabilité conditionnelle* : soit A, B deux sous-ensembles de Ω, et P une probabilité sur Ω. La probabilité de A sachant B est donnée par :

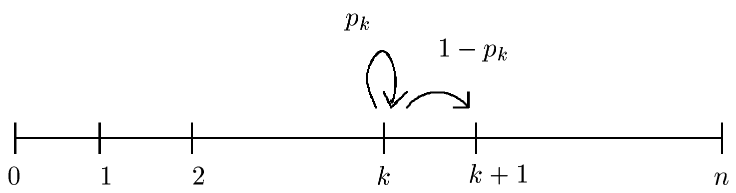
$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Prenons pour A l'événement $(X_{i+1} = k_{i+1})$ et pour B l'événement $[X_0 = k_0, X_1 = k_1, ..X_i = k_i]$. L'intersection de A et de B correspond à l'événement $[X_0 = k_0, X_1 = k_1, ..X_{i+1} = k_{i+1}]$, si bien qu'il suffit de spécifier les valeurs $P(X_{i+1} = k_{i+1}|X_0 = k_0, X_1 = k_1, ..X_{i+1} = k_{i+1})$ pour déterminer P.

Dans nos exemples, la valeur prise par X_{i+1} ne dépend pas des valeurs prises par $X_0...X_{i-1}$, mais juste de la valeur prise par X_i (C'est la *propriété de Markov*). En effet, si la valeur de X_i est connue, disons égale à k, la variable aléatoire X_{i+1} ne peut prendre que les deux valeurs k et $k+1$, et les probabilités associées à ces deux valeurs sont complètement déterminées :

$$\begin{array}{rcll} P(X_{i+1} = l|X_i = k) & = & p_k & \text{si l=k} \\ & = & 1 - p_k & \text{si l=k+1} \\ & = & 0 & \text{sinon} \end{array}$$

avec $p_k = k/n$.



Nous décidons donc de munir Ω de la probabilité P définie par les relations précédentes, car nous pensons qu'elle correspond aux situations que nous sommes en train d'étudier.
Remarquons que la probabilité P n'est pas entièrement déterminée par les relations données plus haut ; les quantités $P(X_0 = k_0)$ n'ont pas été spécifiées. Elles n'interviendront pas dans la suite.

La variable aléatoire T

On s'intéresse au temps T nécessaire pour atteindre la valeur n. Cette quantité est une variable aléatoire définie sur Ω qui peut s'exprimer en fonction des X_i :

$$T \;=\; Card\{i \in \mathbf{N} \mid X_i < n\} \;=\; \sum_{i=0}^{\infty} \mathbf{1}_{\{X_i<n\}}.$$

Rappelons maintenant comment est défini le concept d'*espérance conditionnelle* : soit T une variable aléatoire définie sur Ω, prenant des valeurs entières, et soit A un sous-ensemble de Ω. L'espérance conditionnelle de T sachant A est donnée par une des deux expressions équivalentes suivantes :

$$E(T|A) = \frac{E(T\mathbf{1}_A)}{P(A)} \;=\; \sum_{i=0}^{\infty} iP(T=i|A).$$

La quantité que nous cherchons à calculer correspond au temps moyen nécessaire pour atteindre la valeur n, sachant que nous sommes parti de la valeur 0. Il s'agit donc de l'espérance de T, sachant que $X_0 = 0$:

$$t_0 = E(T|X_0 = 0)$$

3 Calcul des t_k

On peut maintenant définir précisément les quantités t_k et dériver la relation de récurrence qui permet de les calculer. Les quantités t_k correspondent au temps moyen nécessaire pour atteindre la valeur n, sachant qu'on part de la valeur k. On a donc :

$$t_k = E(T|X_0 = k)$$

Le résultat suivant se déduit aisément de la définition de l'espérance conditionnelle.

Théorème : Soit T une variable aléatoire définie sur Ω, A un sous-ensemble de Ω. Soit $B_0, ...B_n$ une partition de Ω : les B_i sont des sous-ensembles de Ω disjoints deux à deux, et l'union de tous ces sous-ensembles est égale à Ω. On a alors la relation suivante :

$$E(T|A) \;=\; \sum_l E(T|B_l \cap A)\; P(B_l|A).$$

Soit k un entier strictement inférieur à n ; considérons l'événement $A = (X_0 = k)$ et les événements $B_l = (X_1 = l)$. Posons $T' = \sum_{i\geq 1} \mathbf{1}_{\{X_i<n\}}$, si bien que $T = \mathbf{1}_{\{X_0<n\}} + T'$. La relation précédente devient :

$$\begin{aligned} E(T|X_0=k) &= \textstyle\sum_l\; E(T|X_1=l, X_0=k)\; P(X_1=l|X_0=k) \\ &= \textstyle\sum_l\; E(\mathbf{1}_{\{X_0<n\}} + T'|X_1=l, X_0=k)\; P(X_1=l|X_0=k) \\ &= \textstyle\sum_l\; E(T'|X_1=l, X_0=k)\; P(X_1=l|X_0=k) \quad +1 \end{aligned}$$

La quantité T' ne dépend que des variables $X_1, X_2...$ mais pas de la variable X_0. Par conséquent, on a l'égalité $E(T'|X_1 = l, X_0 = k) = E(T'|X_1 = l)$, ce qui donne :

$$E(T|X_0=k) = \sum_l\; E(T'|X_1=l)\; P(X_1=l|X_0=k) \quad +1$$

La quantité $E(T'|X_1 = l)$ correspond au temps moyen nécessaire pour atteindre n, sachant qu'on est parti de l. Elle est donc égale à t_l (Il s'agit du caractère *stationnaire* de la marche aléatoire). L'équation précédente devient :

$$t_k = p_k t_k + (1-p_k)t_{k+1} + 1$$

C'est la relation attendue.