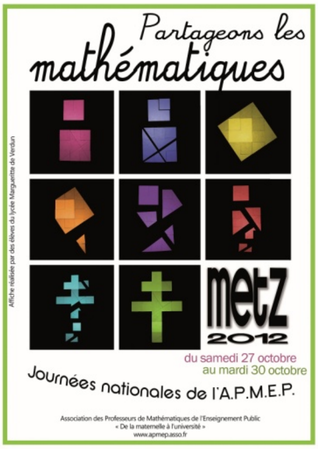


JOURNÉES NATIONALES A.P.M.E.P.
METZ
27-30 octobre 2012

Atelier P3-36
MARCHE ALÉATOIRE VERS UN PARADOXE DU CALCUL DES PROBABILITÉS
Jacques Faisant

Plan

Rappel du thème de l'atelier : *si deux candidats, A et B, se présentent à une élection, autant d'électeurs étant favorables à A qu'à B, il est très probable que, lors du dépouillement d'une urne standard, l'un d'eux soit en tête pendant plus des trois quarts du temps. la probabilité que, lors du dépouillement d'une urne standard, l'un d'eux soit en tête pendant plus des trois quarts du temps est 0,67.*

- Représentation du déroulement du dépouillement
- Les chemins sont équiprobables
- Le théorème du scrutin
- Retours à l'origine et probabilité
- Le premier passage à l'origine
- Le dernier passage à l'origine
- Le temps de séjour au-dessus de l'axe des t (notre objectif !)
- La loi de probabilité discrète de l'Arcsinus d'ordre n

Notations

Notations et formules

(nombre de tous les chemins $(s_0, s_1, ..., s_n)$ avec $s_0 = 0$ et $s_n = x) = N_{n,x} = C_n^{\frac{n+x}{2}}$

$\mathrm{P}(S_n = x) = p_{n,x} = \frac{C_n^{\frac{n+x}{2}}}{2^n}$

$\mathrm{P}(S_{2.\nu} = 0) = u_{2.\nu}$ $\qquad u_{2.\nu} = \frac{C_{2.\nu}^{\nu}}{2^{2.\nu}}$ (1)

(probabilité que le premier passage à l'origine ait lieu à l'instant $2.n) = f_{2.n} = u_{2.n-2} - u_{2.n} = \frac{u_{2.n}}{2.n-1}$

$$\mathrm{P}(S_1 \neq 0 \text{ et ... et } S_{2.n} \neq 0) = u_{2.n} \tag{2}$$

$$\mathrm{P}(S_1 \geq 0 \text{ et ... et } S_{2.n} \geq 0) = u_{2.n} \tag{3}$$

(probabilité que la dernière visite à l'origine avant l'instant $2.n$ ait lieu à l'instant $2.k$ avec $k \leq n) = \alpha_{2.k,2.n} = u_{2.k}.u_{2.n-2.k}$

$$u_{2.n} = f_2.u_{2.n-2} + f_4.u_{2.n-4} + \ \dots\ + f_{2.n}.u_0 \tag{4}$$

(probabilité que, dans l'intervalle de temps de 0 jusqu'à $2.\nu$, la différence des scores soit positive pendant exactement $2.k$ unités de temps) $= b_{2.k,2.\nu} = \alpha_{2.k,2.\nu}$

Représentation

Représentation du déroulement du dépouillement

- Nous considérons seulement les instants correspondant exactement à l'ouverture d'une enveloppe de l'urne et nous disons que ces instants correspondent aux temps repérés par les entiers naturels de 1 jusqu'à *au plus* N.
- Pour chacun de ces instants , nous calculons la différence algébrique des voix obtenues par A et B. Cette différence est placée sur l'axe des ordonnées d'un repère du plan, ce qui nous donne une *marche aléatoire en dimension 1.*
- En fait, une marche aléatoire en dimension 1 se représente en dimension 2, par un *chemin* dessiné dans un repère du plan.

Les chemins

Soient n et x des entiers naturels, n étant non nul. Un chemin $(s_0, s_1, ..., s_n)$ de l'origine au point de coordonnées (n, x) est une ligne polygonale dont les sommets ont comme abscisses *0, 1, ... ,* n et comme ordonnées $s_0, s_1, s_2, ..., s_n$ satisfaisant, pour tout entier naturel non nul k inférieur ou égal à n, à $s_k - s_{k-1} \in \{-1; 1\}$, $s_0 = 0$ *et* $s_n = x$ (Et on peut changer de repère ...).

Propriété

Le nombre de tous les chemins de ce type est noté $N_{n,x}$.

Lorsqu'il n'est pas nul, ce nombre est égal à $C_n^{\frac{n+x}{2}}$.

Démonstration. Le nombre de ces chemins est C_{p+q}^p où p et q sont les entiers naturels tels que $p+q = n$ et $p-q = x$ s'ils existent. (Sinon, $N_{n,x} = 0$). On a alors $2.p = n + x$. □

Probabilités

Probabilités et chemins

Les chemins de longueur n sont au nombre total de 2^n et sont équiprobables. (Voir document *Equiprobables.*)

S_n étant la v.a.r. qui a s_n pour valeur, on note $\mathrm{P}(S_n = x) = p_{n,x}$.

Propriété

D'après ce qui précède, cette probabilité est égale à $\frac{C_n^{\frac{n+x}{2}}}{2^n}$ si p et q existent et est nulle sinon.

Les retours à l'origine

On dit qu'un retour à l'origine a lieu à l'instant k ssi $s_k = 0$.

Pour cela, il est nécessaire (mais pas suffisant !) que k soit pair.

On note $\mathrm{P}(S_{2.\nu} = 0) = u_{2.\nu}$.

On sait que cette probabilité est égale à

$$\frac{C_{2.\nu}^{\nu}}{2^{2.\nu}} \ (1).$$

Le théorème du scrutin

Théoreme

Supposons que, lors d'un scrutin, p personnes aient voté pour le candidat A et q pour le candidat B, avec $p > q$.

La probabilité que, pendant tout le dépouillement, A soit strictement devant B est égale à $\frac{p-q}{p+q}$.

Le lemme de réflexion

Soit A de coordonnées (a, α) et B de coordonnées (b, β) des points à coordonnées entières telles que $b > a \geq 0$, $\alpha > 0$ et $\beta > 0$.

On dit que le point obtenu par réflexion de A par rapport à l'axe des t est A', de coordonnées $(a, -\alpha)$.

Lemme : le nombre des chemins de A à B qui touchent ou qui traversent l'axe des t est égal au nombre total de chemins de A' à B.

Démonstration. Considérons l'ensemble $\mathcal{E}$ des chemins de A à B $(s_a = \alpha, s_{a+1}, ..., s_b = \beta)$ dont un ou plusieurs sommets sont sur l'axe des t . Pour un tel chemin, notons t l'abscisse du premier de ces sommets. On a $s_a > 0$ et ... et $s_{t-1} > 0$ et $s_t = 0$. A tout chemin de $\mathcal{E}$, on peut associer $(-s_a, -s_{a+1}, ..., -s_{t-1}, s_t = 0, s_{t+1}, ..., s_b)$ qui appartient à l'ensemble $\mathcal{F}$ des chemins de A' à B. Cette application ϕ de $\mathcal{E}$ à $\mathcal{F}$ est visiblement bijective. □

Démonstration du théorème du scrutin

Déduisons du lemme de réflexion le théorème du scrutin, qui peut aussi s'écrire :

soient n et x des entiers naturels non nuls. Il existe exactement $\frac{x}{n}.N_{n,x}$ chemins $(s_0, s_1, ..., s_n)$ de l'origine au point de coordonnées (n, x) tels que $s_1 > 0$ et ... et $s_n > 0$

Démonstration. Les chemins qui conviennent passent tous par le point de coordonnées $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. D'après le lemme, leur nombre est donc $N_{n-1,x-1} - N_{n-1,x+1} = C_{p+q-1}^{p-1} - C_{p+q-1}^{p} = \frac{(p+q-1)!}{(p-1)!.q!} - \frac{(p+q-1)!}{p!.(q-1)!} = (p+q-1)!.\left(\frac{p-q}{p!.q!}\right) = \frac{p-q}{p+q}.C_{p+q}^{p}$ où p et q sont tels que $p+q = n$ et $p-q = x$ et que $p-1+q = n-1$, $p-1-q = x-1$, $p+(q-1) = n-1$ et $p-(q-1) = x+1$. □

Propriété des retours à l'origine

La probabilité qu'aucun retour à l'origine ne se produise jusqu'à l'instant $2.n$ compris est égale à la probabilité qu'un tel retour ait lieu à l'instant $2.n$ exactement, c-à-d : $\mathrm{P}(S_1 \neq 0$ et ... et $S_{2.n} \neq 0) = u_{2.n}$ (2), et ceci revient à $\mathrm{P}(S_1 > 0$ et ... et $S_{2.n} > 0) = \frac{1}{2}.u_{2.n}$ et encore à $\mathrm{P}(S_1 \geq 0$ et ... et $S_{2.n} \geq 0) = u_{2.n}$ (3).

Démonstration. $\mathrm{P}(S_1 > 0$ et ... et $S_{2.n} > 0) = \sum\limits_{r=1}^{n} \mathrm{P}(S_1 > 0$ et ... et $S_{2.n} = 2.r)$

$= \frac{1}{2^{2.n}}.\sum\limits_{r=1}^{n}(N_{2.n-1,2.r-1} - N_{2.n-1,2.r+1}) = \frac{1}{2^{2.n}}.N_{2.n-1,1} = \frac{C_{2.n-1}^{n}}{2^{2.n}} = \frac{(2.n)!}{2^{2.n}.n!.(n-1)!.2.n} = \frac{1}{2}.\frac{C_{2.n}^{n}}{2^{2.n}}.$

Enfin, en utilisant aussi le repère d'origine le point de coordonnées $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, on voit que $2.\mathrm{P}(S_1 > 0$ et ... et $S_{2.n} > 0) = \mathrm{P}(S_1 \geq 0$ et ... et $S_{2.n-1} \geq 0)$. Comme $2.n - 1$ est impair, c'est aussi $\mathrm{P}(S_1 \geq 0$ et ... et $S_{2.n} \geq 0)$. □

Le premier passage à l'origine

Propriété

La probabilité que le premier passage à l'origine ait lieu à l'instant $2.n$ est
$\mathrm{P}(S_1 \neq 0$ et ... et $S_{2.n-1} \neq 0$ et $S_{2.n} = 0)$.

On la note $f_{2.n}$ et sa valeur est $u_{2.n-2} - u_{2.n} = \frac{u_{2.n}}{2.n-1}$.

Démonstration. $u_{2.n-2} - u_{2.n} = \frac{C_{2.n-2}^{n-1}}{2^{2.n-2}} - \frac{C_{2.n}^{n}}{2^{2.n}} = \frac{(2.n-2)!}{2^{2.n}}.\frac{2^2.n^2-(2.n-1).2.n}{n!.n!} = \frac{C_{2.n}^{n}}{2.n-1}.\frac{1}{2^{2.n}}$ selon (2) □

Remarque : $\mathrm{P}(S_1 > 0$ et ... et $S_{2.n-1} > 0$ et $S_{2.n} = 0) = \frac{1}{2}.f_{2.n}$ par symétrie.

Propriété

Si $n \geq 1$, $u_{2.n} = f_2.u_{2.n-2} + f_4.u_{2.n-4} + \ ... \ + f_{2.n-2}.u_2 + f_{2.n}.u_0$ (4).

Démonstration. Utilisons la formule des probabilités totales. Un passage à l'origine à l'instant $2.n$ peut être le premier. Sinon, ce premier passage a lieu à l'instant $2.k$ avec $0 < k < n$, évènement que nous notons F_k , et est suivi, notamment, par un passage à l'origine $2.n - 2.k$ instants plus tard, évènement que nous notons G_k . On a alors, pour chaque k tel que $0 < k < n$, $\mathrm{P}(F_k) = f_{2.k}$ et $\mathrm{P}_{F_k}(G_k) = u_{2.n-2.k}$. □

Le dernier passage à l'origine

Propriété

La probabilité que la dernière visite à l'origine jusqu'à l'instant $2.n$ ait lieu à l'instant $2.k$ $(k \leq n)$ est notée $\alpha_{2.k,2.n}$ et est égale à $u_{2.k}.u_{2.n-2.k}$ ($k = 0$: *aucune visite*) .

Démonstration. L'évènement étudié est l'intersection des évènements E : "une visite à l'origine a lieu à l'instant $2.k$ et, éventuellement, avant cet instant" et F : "aucune visite n'a lieu entre l'instant $2.k+1$ et l'instant $2.n$ compris". Or $\mathrm{P}(E) = u_{2.k}$, d'après (1), et $\mathrm{P}_E(F) = u_{2.n-2.k}$, d'après (2) et ceci par changement d'abscisse de l'origine du repère utilisé..

La probabilité recherchée est donc $\mathrm{P}(E).\mathrm{P}_E(F) = u_{2.k}.u_{2.n-2.k}$. □

Temps de séjour

Le temps de séjour

Propriété

La probabilité que, dans l'intervalle de temps de 0 jusqu'à $2.n$, la différence des scores soit positive ou nulle pendant exactement $2.k$ unités de temps est égale à $\alpha_{2.k,2.n}$.

Démonstration. Considérons les chemins de longueur $2.n$. Notons $b_{2.k,2.n}$ la probabilité de l'évènement E : « exactement $2.k$ segments se situent au-dessus de l'axe des t ». On sait que, d'après (3), $b_{2.n,2.n} = u_{2.n}$. Par symétrie, on a aussi $b_{0,2.n} = u_{2.n}$. Prenons maintenant le cas où $1 \leq k \leq n-1$. Dans ce cas, un premier retour à l'origine doit se produire à un instant $2.r$ tel que $0 < 2.r < 2.n$.
– Ou bien les $2.r$ unités de temps avant le premier retour à l'origine correspondent à une position au-dessus de l'axe. Notons F_r cet évènement, qui est d'intersection vide avec E si on n'a pas $r \leq k \leq n-1$. Notons G_r l'évènement : « au-delà de l'instant $2.r$, exactement $2.k - 2.r$ segments sont au-dessus de l'axe des t ». On a alors $\mathrm{P}(F_r) = \frac{1}{2}.f_{2.r}$ et $\mathrm{P}_{F_r}(G_r) = b_{2.k-2.r,2.n-2.r}$.
– Ou bien les $2.r$ unités de temps avant le premier retour à l'origine correspondent à une position au-dessous de l'axe. Notons H_r cet évènement, qui est d'intersection vide avec E si on n'a pas $k \leq n-r$. Notons I_r l'évènement : « au-delà de l'instant $2.r$, exactement $2.k$ segments sont au-dessus de l'axe des t ». On a alors $\mathrm{P}(H_r) = \frac{1}{2}.f_{2.r}$ et $\mathrm{P}_{H_r}(I_r) = b_{2.k,2.n-2.r}$.
Donc, d'après la formule des probabilités totales, lorsque $1 \leq k \leq n-1$,

$$b_{2.k,2.n} = \frac{1}{2}.\sum_{r=1}^{k} f_{2.r}.b_{2.k-2.r,2.n-2.r} + \frac{1}{2}.\sum_{r=1}^{n-k} f_{2.r}.b_{2.k,2.n-2.r}.$$

Nous pouvons maintenant démontrer par récurrence sur $n \in \mathbb{N}^*$ que $b_{2.k,2.n} = \alpha_{2.k,2.n} = u_{2.k}.u_{2.n-2.k}$. Ceci est vrai si $n = 1$. Supposons-le vrai pour tout naturel $\nu < n$. Cette supposition, jointe au résultat précédent (vrai pour n si $k < n$) nous donne, en ayant choisi les valeurs $\nu = n - r$, $b_{2.k,2.n} = \frac{1}{2}.u_{2.n-2.k}.\sum_{r=1}^{k} f_{2.r}.u_{2.k-2.r} + \frac{1}{2}.u_{2.k}.\sum_{r=1}^{n-k} f_{2.r}.u_{2.n-2.k-2.r} = \frac{1}{2}.u_{2.n-2.k}.u_{2.k} + \frac{1}{2}.u_{2.k}.u_{2.n-2.k}$ d'après (4).

On conclut alors cette récurrence. □

Et notre objectif ?

La probabilité que le candidat B soit en tête pendant plus des trois quarts du temps est, si l'urne contient 1000 bulletins, $\sum_{0 \leq k < 125} \alpha_{2.k,1000}$

On peut, par exemple, calculer une valeur approchée de cette probabilité avec le logiciel bc (peu connu, mais bien pratique). On trouve 0,66850 avec une très bonne précision de calcul.

Loi de l'Arcsinus

La loi de probabilité discrète de l'Arcsinus d'ordre n

Propriété de convergence en loi

Si l'entier n est suffisamment grand, alors, pour chaque réel x strict[t] compris entre 0 et 1, $\sum_{0 \leq \frac{k}{n} \leq x} \alpha_{2.k,2.n} \cong \frac{2}{\pi}.\mathrm{Arcsin}\,(\sqrt{x})$.

C'est pourquoi on a la définition :

La loi de probabilité d'une v.a.r. X telle que, pour tout naturel k inférieur ou égal à n, $\mathrm{P}(X = 2.k) = \alpha_{2.k,2.n}$ s'appelle la loi de probabilité discrète de l'*Arcsinus* d'ordre n.

(Paul Lévy a trouvé cette loi en 1939 pour le mouvement brownien.)

Propriété

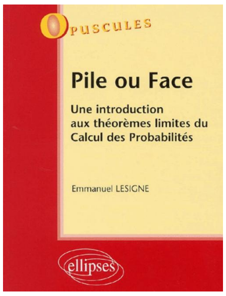
La dérivée de $x \mapsto \frac{2}{\pi}.\mathrm{Arcsin}\,(\sqrt{x})$ est $x \mapsto \frac{1}{\pi.\sqrt{x.(1-x)}}$.

Convergence de la loi discrète de l'Arcsinus d'ordre n

La démonstration part de la dérivée indiquée puis s'obtient en utilisant la formule de Stirling, des sommes de Riemann et des convergences *uniformes*.

Voir « Pile ou Face : Une introduction aux théorèmes limites du Calcul des Probabilités »

par Emmanuel Lesigné, éditions Ellipses, pages 51 à 54.



Nous avons démontré que la loi de probabilité du temps de séjour au-dessus de l'axe des t est la loi discrète de l'Arcsinus d'ordre n.
Comme $2.\frac{2}{\pi}.\text{Arcsin}\left(\sqrt{0.25}\right) = \frac{2}{3}$, il est donc vrai que la probabilité que l'un des candidats soit en tête pendant plus des trois quarts du temps est environ 0,67.

GeoGebra

On peut trouver des indications sur l'illustration de ce paradoxe avec le logiciel de géométrie dynamique **GeoGebra** sur le site

http ://jacques.faisant.pagesperso-orange.fr/JN2012/