DEUX ANS DE PARIS AVEC UNE VARIABLE ALEATOIRE ETONNANTE

J F Kentzel – Enseignant au lycée Pardailhan à Auch (32) - jkentzel@ac-toulouse.fr

n est un entier, $n \geq 1$. On lance n fois une pièce, c'est-à-dire qu'on réalise n expériences de Bernoulli identiques et indépendantes dont les issues sont désignées par succès, noté 1, de probabilité p, $0 \prec p \prec 1$, et échec, noté 0.

On désigne par $L_n^{(p)}$ la variable aléatoire : « longueur de « la » plus longue suite de succès consécutifs obtenue à l'issue de la $n^{\text{ième}}$ expérience », présentée par A de Moivre en 1738. Exemple : avec 0101110100, $L_{10}^{(p)} = 3$.

Tout ce qui suit est traité en détails sur la page [1] (voir cette page pour comprendre le titre de cet article !)

1) UNE FORMULE DE RECURRENCE

Pour tout n fixé, on peut donner la loi de $L_n^{(p)}$, c'est-à-dire calculer $P\left(L_n^{(p)} = k\right)$ pour tout entier k entre 0 et n grâce à la formule de récurrence qui va suivre.

Le calcul simple $P\left(L_n^{(p)} = 0\right)$ est à part : $P\left(L_n^{(p)} = 0\right) = (1-p)^n$. Soit k un entier fixé, $k \geq 1$.

Pour tout entier n valant au moins 1, soit $s_n^{(p)(k)} = \underline{s_n^{(k)}}$ la probabilité de l'événement { $\underline{L_n^{(p)} \geq k}$ }.

La suite $(s_n^{(k)})_n$ est définie par ses $(k+1)$ valeurs initiales faciles à vérifier :

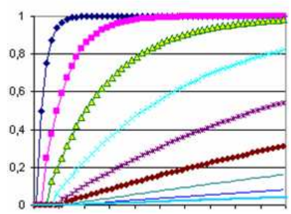
si $1 \leq n \leq k-1, s_n^{(k)} = 0$ (cas évidemment exclu si $k = 1$) ; $s_k^{(k)} = p^k$; $s_{k+1}^{(k)} = 2(1-p)p^k + p^{k+1}$;

et par la relation de récurrence valable si $n \geq k+2$: $s_n^{(k)} = s_{n-1}^{(k)} + (1-p)p^k\left(1 - s_{n-k-1}^{(k)}\right)$, plus lisible sous la forme : $\underline{s_n = s_{n-1} + (1-p)p^k\left(1 - s_{n-k-1}\right)}$.

preuve de cette relation : soit on avait déjà une suite de k succès consécutifs à l'issue de la $(n-1)^{\text{ème}}$ expérience, soit une telle suite apparaît pour la première fois à la $n^{\text{ème}}$ expérience sous la forme $(\underbrace{\ldots\ldots\ldots\ldots}_{\text{"RIEN"}}\ 0\underbrace{11111\ldots11}_{k\ \text{fois}})$.

(où « rien » signifie pas de suite de k succès consécutifs, c'est-à-dire $\{L_{n-k-1} \prec k\}$)

On a représenté ci-contre chacune des suites $(s_n^{(k)})_{1 \leq n \leq 50}$ pour k entre 1 et 9 dans le cas $p = 1/2$. Ces suites $(s_n^{(k)})_{1 \leq n}$ sont (strictement) croissantes et convergent vers 1.

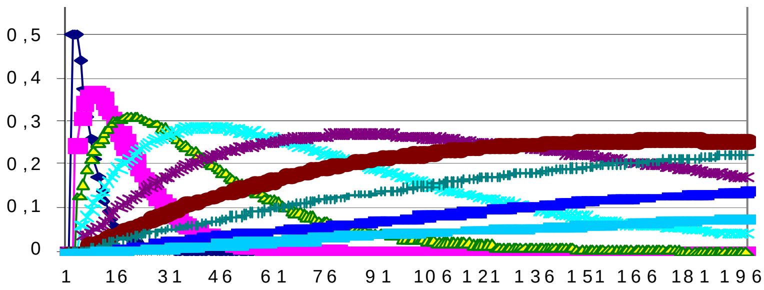


La ligne qui suit montre que l'allure des courbes de la figure 2 donnée ci-dessous n'est pas étonnante. **FIGURE 1**

Si $k \geq 1$, $\underline{P\left(L_n^{(p)} = k\right) = s_n^{(k)} - s_n^{(k+1)}}$ car { $L_n^{(p)} \geq k+1$ } $\subset$ { $L_n^{(p)} \geq k$ }.

On s'intéresse surtout au cas $p = 1/2$ jusqu'à la fin du paragraphe 2.

On peut, par soustraction des colonnes $(s_n^{(k)})_{1 \leq n}$ si on a un tableur, obtenir une représentation des suites $(P(\{L_n = k\}))_n$, dans le cas $p = 1/2$, pour $1 \leq n \leq 200$ et $1 \leq k \leq 9$: **FIGURE 2**



Bien sûr, c'est de gauche à droite on voit successivement les « vagues » $k = 1$, $k = 2$ …etc

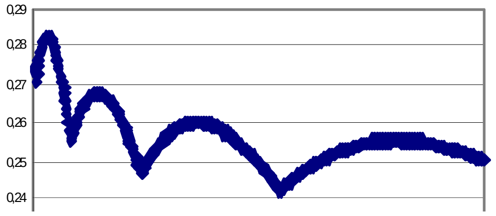
Un résultat annoncé en 1978 et assez lisible sur ce dessin est $P(L_{200}^{(1/2)} \geq 5) \approx 0{,}96$ **(c'est-à-dire** $P(L_{200}^{(1/2)} \leq 4) \approx 0{,}04$**). On peut en déduire des activités à mener en classe** (voir à la fin).

2) UN PARI SUR LA VALEUR DE $L_n^{(p)}$

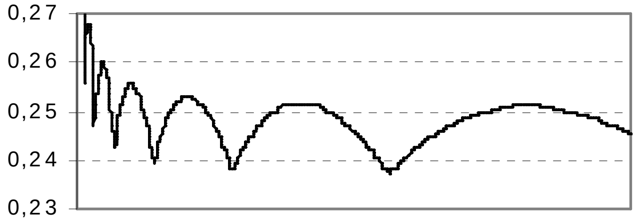
$L_n^{(p)}$ prend toutes les valeurs k entre 1 et n et on s'intéresse à u_n : Maximum sur k des $P\left(L_n^{(p)}=k\right)$, c'est à dire qu'on se demande, pour un n donné, sur quelle valeur de $L_n^{(p)}$ parier. La figure ci-dessus donne la réponse pour $p=1/2$ et $n \le 200$. On gagne alors avec une probabilité proche de 1 / 4.
On observe que l' « arche » correspondant à « la plus longue suite la plus probable est de longueur k » est obtenue, approximativement, pour n dans l'intervalle $\left[2^{k+1}; 2^{k+2}\right]$, autrement dit, on a approximativement : si $n \in \left[2^{k+1}; 2^{k+2}\right]$, c'est-à-dire $(k+1)\, Ln\,(2) \le Ln\,(n) \le (k+2)\, Ln\,(2)$,
la valeur la plus probable de $L_n^{(1/2)}$ est $Ent\,(Ln\,(n)/Ln\,(2))-1$ où Ent est la partie entière.

Sur le dessin ci-dessus, on voit intervenir pour n les valeurs 15, 31, 61…On va voir que « la » formule donnant les abscisses des « pieds des arches » n'est pas $x=2^{k+1}$ mais est plutôt $x = Ent\,(C.2^{k+1})$ avec $C \approx 0{,}96$. Les feuilles de tableur sont sur le site [1].

Les deux représentations de (u_n) qui suivent semblent rendre moins vraisemblable une probabilité de gagner proche de 1 / 4 quand n grandit …



$30 \le n \le 430$: les quatre « arches » visibles correspondent à k valant 4, 5, 6 et 7.



$30 \le n \le 3500$: les quatre « arches » les plus à droite correspondent à k valant 7, 8, 9 et 10.

FIGURE 3

3) UNE APPROXIMATION DE $P\left(L_n^{(p)}=k\right)$

Soit p un nombre fixé vérifiant $0 \prec p \prec 1$. Soit n un entier fixé, assez grand...
En désignant par S_r l'événement : succès lors de la $r^{ème}$ épreuve, une suite de n épreuves peut être représentée sous la forme : .... $\overline{S_1}$ .... $\overline{S_2}$ .... $\overline{S_3}$ .... $\overline{S_N}$ ....

longueur l_1 *longueur* l_2 l_3 l_N

où les points représentent des succès et la variable aléatoire $N = N\,(n)$ est le nombre d'échecs.

La loi de $L_n^{(p)}$ est alors proche de celle de $\underset{1\le i\le N}{Max}\ (\,l_i\,)$.

Les longueurs l_i sont des variables aléatoires indépendantes (valant éventuellement 0) et vérifient évidemment pour tout entier x, $0 \le x \le n$: $P(l_i \ge x) \approx p^x$.

Puisque les l_i sont indépendantes, on a

$$P\,(L_n \prec x) \approx P\,(\bigcap_{1\le i\le N}(l_i \prec x)) = P(l_1 \prec x)\,.\,P(l_2 \prec x).....P(l_N \prec x) = \left(1-p^x\right)^N$$

donc $P\,(L_n \prec x) \approx \left(1-p^x\right)^N = \exp\left(N.\,Ln\,(1-p^x)\right)$.
$N(n)$ suit la loi binomiale $B(n\,;1-p)$ donc $E(N) = n\,(1-p)$, autrement dit $N \approx n\,(1-p)$.
Par ailleurs si x est assez grand, p^x est assez petit et $Ln\,(1-p^x) \approx -p^x$ ($Ln\,(A) \approx A-1$ si $A \approx 1$).

On a donc : $P\left(L_n^{(p)} \prec k\right) \approx e^{-n\,p^k(1-p)}$.

Cette approximation est justifiée correctement pour n entier assez grand depuis au moins 1944, c'est un résultat de Goncharov dont je n'ai pas trouvé de trace. Je remercie Anne Bauval (Université de Toulouse 3) qui m'a donné une majoration de $\left|P\left(L_n^{(p)}=k\right)-d_k(n)\right|$ assez efficace (mais trop longue pour figurer ici) pour prouver tout ce qui suit.

$P\left(L_n^{(p)}=k\right)=P\left(L_n^{(p)}\prec k+1\right)-P\left(L_n^{(p)}\prec k\right)$ donc en désignant par $d_{p\,;\,k}=d_k$ la fonction définie sur IR par d_k : $x\mapsto e^{-x(1-p)p^{k+1}}-e^{-x(1-p)p^k}$, on a, pour n entier assez grand : $P\left(L_n^{(p)}=k\right)\approx d_k(n)$.

4) UNE ETUDE DES FONCTIONS d_k

On visualise la démarche d'approximation évoquée à la fin de 3):

On a vu que pour tous k et p fixés on peut facilement, grâce à une formule de récurrence, **calculer** $P\left(L_n^{(p)}=k\right)$ pour tout n.

Considérons pour ce résumé l'exemple $p=0{,}4$.

Par exemple, pour k entre 1 et 6 et n entre 1 et 200, on obtient à l'aide d'un tableur ce dessin :

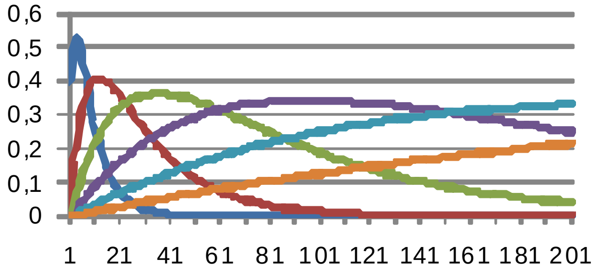
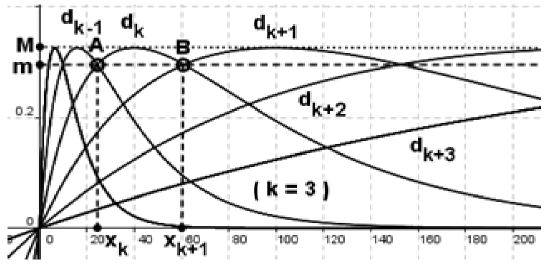


FIGURE 4

Le cas particulier simple $k=0$ est traité à part.

Pour les mêmes valeurs de k et n que ci-dessus, on obtient pour $d_{k;0,4}$ les six courbes ci-contre (appelées « arches »).

On peut prouver que $d_k(n)$ est une (bonne) **approximation** de $P(L_n=k)$ dans un cadre précis (k (« donc n ») assez grand et $x_k\le n\le x_{k+1}$). **FIGURE 5**



On calcule facilement les nombres M, x_k et m visibles sur la figure 3. M et m sont indépendants de k. Une explication de ce qui suit est que sur cette figure 3, on passe d'une courbe à une autre par une affinité (disons un « étirement vers la droite » quand k augmente ?) : en effet, puisque pour tout x, $d_k(x)=e^{-x(1-p)p^{k+1}}-e^{-x(1-p)p^k}$, on a bien $d_{p\,;\,k+1}(x)=d_{p\,;\,k}(px)$.

La résolution de l'équation $d_k(x)=d_{k-1}(x)$ donne $x_{p\,;\,k}=x_k=\dfrac{Ln(y_p)}{(1-p)^2\,p^{k-1}}$ où y_p désigne la racine réelle autre que 1 de l'équation $y^{p+1}-2y+1=0$. On a alors $m_p=m=d_k(x_k)=y_p^{\frac{p}{p-1}}-y_p^{\frac{1}{p-1}}$.

[exemple vu au 2) : $p=1/2$, $y_{1/2}=(3+\sqrt{5})/2$, $C=Ln\,(y_{1/2})\approx 0{,}96$ et $m_{1/2}=\sqrt{5}-2\approx 0{,}236$]

5) PRINCIPE DES PARIS SUR LA VALEUR DE $L_n^{(p)}$ (pour n supposé assez grand)

Ce principe est simple : pour chaque n, on parie sur le k de « la courbe d_k la plus haute »

$(x_{p\,;\,k})_k$ est une suite géométrique de raison $1/p$ donc tout entier n vérifiant $n\ge x_{p\,;\,1}=x_1$ est dans un intervalle du type $[x_k\,;\,x_{k+1}[$ pour une valeur de k qu'on peut facilement calculer en fonction de n.

On appellera $k_0(p\,;n)=k_0$ cette valeur de k.

Sur $[x_k\,;\,x_{k+1}[$, d_k est minorée par m cependant que $d_k(n)$ est une approximation de $P(L_n=k)$.

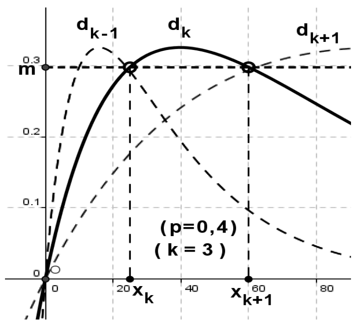


FIGURE 6

On calcule facilement $k_0 = k_0(p\,;n) = Ent\left(\ Ln\left(\dfrac{(1-p)^2}{Ln\,(y_p)}\,n\right) / \ Ln\left(\dfrac{1}{p}\right)\right) + 1$.

On prouve alors facilement que si $p \le 1/2$, pour tout $\varepsilon \succ 0$, on peut calculer un entier $N_{p\,;\varepsilon}$ tel que : pour tout entier n valant au moins $N_{p;\varepsilon}$, on a $P\left(L_n^{(p)} = k_0(p\,;n)\right) \succ m_p - \varepsilon$.

6) QUELQUES REMARQUES

a) La précision de la majoration de Anne Bauval permet pour $p \le 1/2$ et $\varepsilon \succ 0$ donnés de calculer $N_{p\,;\varepsilon}$. On peut alors à l'aide d'un ordinateur montrer que pour certaines valeurs de $p \le 1/2$ et ε, l'inégalité $P\left(L_n^{(p)} = k_0(p\,;n)\right) \succ m_p - \varepsilon$ est vraie pour presque tout n (une ou deux exceptions). Le nombre k_0 donne alors une formule unique permettant de parier sur la valeur de $L_n^{(p)}$ pour tout n.

b) Voir ci-contre quelques points $(p\,;m_p)$ dont les points-limite D et E.
On a aussi tracé en pointillés quelques points $(p\,;M_p)$.

m_p est, à ε près, la probabilité minimum, **grande si p est petit**, de gagner un pari sur $L_n^{(p)}$ en pariant que $L_n^{(p)} = k_0$.
(et, accessoirement, M_p en est la probabilité maximum)

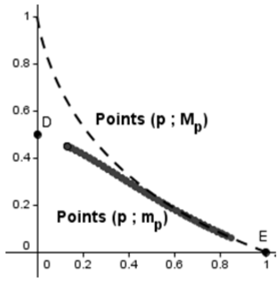


FIGURE 7

On peut aussi avec la même méthode faire des paris sur deux valeurs de $L_n^{(p)}$, du type : on parie que $L_n^{(p)}$ vaut k_1 ou k_1+1. On a alors des probabilités de gagner très importantes (assez proches de $2m_p$).

c) Le k_0 donné ici améliore (très légèrement !) la valeur bien connue $\alpha = Ln\,(n\,(1-p)) / Ln\,(1/p)$ pour parier sur la valeur de $L_n^{(p)}$. $Var\,(L_n^{(p)})$ est petite et $E(L_n^{(p)}) \approx \alpha$ sont des résultats connus aussi.

d) **Une activité en classe** Voir la fiche-élève téléchargeable sur [1]
Premier acte : chaque élève tape =ent(alea()*2) en A1 puis glisse-copie jusqu'en A200 et, dans la colonne C, « invente » une suite de deux cents 0 ou 1 en les tapant un par un.
L'objectif est d'empêcher « l'adversaire » de reconnaître quelle est la série inventée. Pour ça on cache les formules aléa() avec une recopie et on **peut** inverser les colonnes A et C.
Entracte : on explique aux élèves (preuve ou, plutôt, simulation) que $P\ (L_{200}^{(1/2)} \ge 5) \approx 0{,}965$.
Acte deux : les élèves changent de place. Leur nouvel objectif est repérer la série inventée : la A ou la C. Il y arrivent dans 75 ou 80 % des cas. Pourquoi ? (il y a un facteur psychologique)
La lecture de ce qui suit est impérative si vous utilisez la fiche-élève évoquée ci-dessus.

e) Soit $L_n^{(1/2)\prime} = L_n'$ la nouvelle variable aléatoire : longueur de « la » plus longue suite de **résultats consécutifs identiques** obtenue à l'issue du n-ème lancer d'une pièce équilibrée.
Certes, $P\,(L_n' = k)$ vaut « souvent » « un peu moins du double » de $P\,(L_n^{(1/2)} = k)$ mais il y a mieux :
On peut prouver assez facilement que pour tous n et k, $\underline{P(L_n^{(1/2)} = k)} = \underline{P(L_{n+1}' = k+1)}$.
Pour une pièce bien équilibrée ($p = 1/2$), les lois de L_n' et $L_n^{(1/2)}$ sont définies par des suites identiques à « un double décalage près ».
Dans la fiche-élève évoquée au e), c'est la variable L_n' qui est utilisée avec $P\,(L_{200}' \ge 6) \approx 0{,}965$.
A la fin du 1) et au e), on aurait en fait du écrire : $P\,(L_{199}^{(1/2)} \ge 5) \approx 0{,}965$.

Référence [1] Site du lycée Pardailhan à Auch (32) http://pardailhan.entmip.fr/
Cliquer sur / Rubriques des disciplines / Mathématiques / Documents pour les enseignants