

# Les urnes de Polya en classe de seconde

Gilles Aldon

21 janvier 2008

## Présentation du problème

Dans une urne sont placées une boule noire et une boule blanche. On choisit une boule au hasard ; on la replace alors dans l'urne avec une autre boule de la même couleur. Il s'agit soit d'étudier la dynamique de l'urne, soit la composition de l'urne après  $n$  tirages.

## Une solution

Si l'on considère  $X_n$  le nombre de boules blanches dans l'urne à l'étape  $n$ , et  $M_n$  la proportion de boules blanches dans l'urne :

$$M_n = \frac{X_n}{n+2}$$

1.  $X_n$  est uniformément distribuée sur  $\{1, 2, \dots, n+1\}$
2.  $M_n$  converge presque sûrement vers une variable aléatoire uniformément distribuée sur  $[0, 1]$

Autrement dit, en recommençant plusieurs fois l'expérience jusqu'à une étape donnée, toutes les compositions d'urne possibles sont équiprobables (c'est le résultat 1) mais, si l'on prolonge une expérience, la proportion de boules blanches (ou noires) de l'urne va se stabiliser vers une valeur<sup>1</sup>. Ces deux résultats peuvent sembler incompatibles, et en tout cas dirigent le problème dans deux directions différentes.

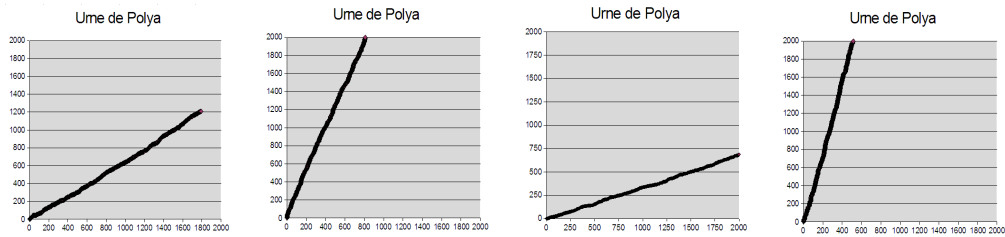
On peut illustrer ces phénomènes par les simulations faites sur tableur :

Pour illustrer le premier point de vue, on recommence 3000 fois la simulation jusqu'à l'étape 6 et on note les fréquences d'apparition de chacune des compositions de l'urne (il y a alors sept compositions d'urne possibles).

	A	B	C	D	E	F	G	H	I	J
		1	2	3	4	5	6	7		
		1B7R	2B6R	3B5R	4B4R	5B3R	6B2R	7B1R		
Effectif		403	434	398	458	465	443	399		3000
Fréquence		0.13	0.14	0.13	0.15	0.16	0.15	0.13		
Fréquence théorique		0.14	0.14	0.14	0.14	0.14	0.14	0.14		

Pour illustrer le second point de vue, on simule quatre expériences jusqu'à une grande valeur de  $n$  (ici  $n = 3000$ ) et on représente la proportion de boules noires et blanches dans l'urne :

<sup>1</sup>En fait, vers l'une quelconque des valeurs de  $[0, 1]$



Expérience 1

Expérience 2

Expérience 3

Expérience 4

Il y a donc deux approches de cette situation : une approche « longitudinale » qui cherche à étudier l'évolution de la composition de l'urne à long terme et une approche « transversale » qui étudie les probabilités des compositions de l'urne à une étape donnée. C'est sans doute la confusion entre ces deux approches qui rend difficile le travail en classe comme on le verra dans la suite.

## Prolongements

Cette situation mathématique est intéressante en elle-même, mais aussi de par ses applications et ses généralisations ; citons par exemple, une première généralisation avec une composition initiale de l'urne avec  $k$  boules blanches et  $k'$  boules noires ; c'est une situation dont l'étude expérimentale peut se faire à l'aide d'un tableur.

Avec des boules de  $k$  couleurs différentes et  $\alpha_i$  boules de couleur «  $i$  ».

Mais aussi, la possibilité pour les boules de se dupliquer aléatoirement dans un intervalle de temps  $dt$ . Je ne donnerai pas de précisions supplémentaires concernant ces généralisations mais on pourra se reporter à l'article [Pemantle 2007] dans lequel elles sont présentées.

On peut également faire le lien entre l'étude de la dynamique de l'urne de Polya et le comportement des consommateurs devant un choix a priori de plusieurs produits ; si on consulte ses amis sur le choix possible, les premiers avis peuvent être déterminant pour la popularité d'un produit. Cette situation peut être modélisée par une urne de Polya à trois boules.

## Un énoncé en seconde

En classe de seconde, cette situation mathématique nous a paru intéressante pour présenter le concept de fluctuation des échantillonnages. La situation suivante a donc été élaborée avec l'objectif de faire prendre conscience de la variabilité des résultats obtenus par des expériences limitées et de la « convergence »<sup>2</sup> des fréquences.

Par ailleurs, nous avons tenu à mettre les élèves en situation d'expérimenter réellement, les simulations sur ordinateurs n'arrivant que plus tard dans une phase d'institutionnalisation. Nous avons donc utilisé des urnes en tissu et des billes<sup>3</sup> pour permettre aux élèves d'expérimenter.

Enfin le choix d'étudier la composition de l'urne à l'étape six permet de construire un modèle de dé à sept faces. Cette variable a été retenue pour justifier auprès des élèves une procédure complexe de choix entre sept éventualités.

## Enoncé

Pour pouvoir choisir au hasard entre 7 alternatives différentes, un mathématicien procède de la manière suivante :

- Il place une boule rouge et une boule noire dans un sac.
- Il choisit au hasard une boule ;
- si cette boule est rouge, il remet deux boules rouges dans l'urne

<sup>2</sup>Dans un sens difficile à expliciter rigoureusement en seconde

<sup>3</sup>indiscernables au toucher !

- si cette boule est noire, il remet deux boules noires dans l'urne.
- Il s'arrête à la sixième étape : il y a alors 8 boules dans l'urne
- Quelles sont les 7 compositions possibles de l'urne ?
- A t'on les mêmes chances d'obtenir chacune de ces compositions ?

## Déroulement prévu

- Donner et lire l'énoncé.
- Laisser réfléchir individuellement à la première question.
- Mettre en commun les résultats et coder les résultats de la même manière pour tous.
- Faire faire l'expérience par groupe et noter les résultats sur la feuille de résultats.
- Réponse par groupe, puis mise en commun des résultats et réponse de la classe.
- Institutionnalisation des résultats appuyée par une simulation sur orsinateur.

## Compte-rendu du déroulement

La classe est séparée en 7 groupes (6 groupes de 4 et un groupe de deux)

J'ai commencé par décrire l'expérience de Polya aux élèves ; nous avons ensuite passer en revue différentes procédures de choix : entre deux alternatives, le lancer d'une pièce de monnaie est un procédé qui donne à chaque choix la même chance d'apparaître ; entre 6 alternatives et un élève a proposé le dé. J'ai alors demandé si le procédé décrit (les urnes de Polya) pouvait permettre de choisir entre 7 personnes et comment il le pourrait.

Après avoir distribué les sujets, j'ai laissé un temps de réflexion individuelle sur la première question pendant cinq minutes, puis j'ai repris la main pour demander les réponses proposées. S'en est suivi un dialogue avec la classe qui a permis de dégager les 7 résultats possibles et le codage que l'on allait adopter : choisir entre 7 alternatives revient à attribuer à chacun un nombre de boules rouges et celui qui est choisi est celui qui porte le numéro correspondant au nombre de boules rouges. J'ai posé la question à la classe : est-ce que tous les résultats ont la même chance d'apparaître.

Le sentiment général était que oui (confère remarques). J'ai alors dit que pour en décider, chaque groupe allait faire l'expérience 20 fois et noter les résultats dans la feuille distribuée.

Expérimentation... En circulant dans la classe, de groupes en groupes, je demandais si les élèves avaient une idée de la réponse à la question ; beaucoup de non réponses ou d'essai d'interprétation longitudinale<sup>4</sup>. J'ai noté la réflexion d'un élève : « dès qu'on a tiré une boule, il y a des résultats qui sont impossibles, donc au début c'est important ».

Dès qu'un groupe finissait ses 20 expériences, ils venaient au tableau noter les résultats :

	1R	2R	3R	4R	5R	6R	7R
	5	2	0	2	3	5	3
	2	3	3	2	3	3	4
	4	2	5	2	3	2	2
	0	7	2	3	2	0	6
	7	2	1	2	3	3	2
	1	3	3	4	2	4	2
	4	2	5	2	3	2	2
Totaux	23	21	19	17	19	19	21

Fin de la première séance : les élèves sont repartis avec les résultats globaux de la classe.

<sup>4</sup>Au sens indiqué plus haut

## Deuxième séance

Comparaison des résultats individuels et cumulés recopiés au tableau. Devant la question « est ce que tous les résultats ont la même chance d'apparaître ? » les élèves semblent dubitatifs et n'apportent pas de réponses ;

je demande quelle serait la fréquence théorique si tous avait la même chance : réponse  $20/140 = 0,14$

je propose alors de multiplier les expériences en faisant une simulation sur ordinateur ce qui ne réveille pas de curiosité particulière ni de demande d'explications sur le pourquoi de la multiplication des expériences<sup>5</sup>.

Je montre d'abord que l'ordinateur simule les six étapes et affiche les résultats d'une expérience. Puis je montre les deux graphiques : fréquence d'un échantillon, fréquence cumulée. Enfin, je lance les simulations ; chaque appui sur le bouton lance une centaine d'expériences. Les résultats sont donc comparables à ceux réalisés dans la classe. En cumulant ces expériences, on fait apparaître une « convergence » des fréquences vers la valeur théorique.

## Analyse de la séquence

Je reste sur l'idée que ce n'est pas la meilleure situation pour parler de fluctuations d'échantillonnage du fait de la distance et des difficultés mêmes du problème posé. La question de savoir si les résultats étaient équiprobables n'est pas judicieuse et peut renforcer l'idée, très fortement présente, que s'il y a 7 résultats, il y a une chance sur 7 que chacun apparaisse. L'expérimentation effective ne me semble, là encore pas ou peu judicieuse du fait encore une fois de la distance des élèves au problème. Ce n'est pas une question qu'ils se posent, et s'ils ont fait l'expérience c'est pour me faire plaisir et non pas pour tenter de répondre à une question. En d'autres termes l'expérience et le problème n'ont pas été corrélés fortement dans ma classe. Peut être aurait il fallu plus de temps et attendre que l'idée de l'expérience vienne des élèves eux-mêmes. Dans mon cas, je l'ai imposé sans qu'il soit évident que les élèves en aient vu la nécessité ni même l'intérêt. Le problème dans un premier temps est plus vu de façon longitudinal que transversal et quelques réflexions d'élèves<sup>6</sup> me laissent penser que c'était leur vision première du problème et que certaines difficultés d'interprétation venaient de ce décalage entre l'idée que les élèves se faisaient de la tâche et la tâche elle-même.

Peut-être, pour utiliser une telle situation en seconde, faudrait il la faire précéder par d'autres de même type mais plus proche des élèves et dont le résultat peut surprendre de prime abord. Je pense à la classique somme des deux dés qui soulève plus de questions et d'interrogations de type probabiliste, ou bien la situation des punaises qui montre la nécessité de modéliser la probabilité en dehors de toute possibilité de raisonnement physique ou géométrique. Il est cependant intéressant de mettre en évidence une situation d'équiprobabilité dans laquelle la « géométrie » et la « symétrie » du hasard ne sont pas évidentes. En faisant l'expérience proposée dans l'énoncé donné, dès le premier tirage, une des compositions de l'urne est éliminée ; nous sommes dans une situation du même type que la situation de la courte-paille : il y a  $n$  paille de longueur donnée et une paille plus courte. le gagnant est celui qui tire la courte paille. Parmi les  $n + 1$  personnes présentes, on choisit à tour de rôle une paille : quelle est la meilleure place pour choisir la paille ? Il semble qu'il y ait une dissymétrie et pourtant, la place n'a pas d'importance<sup>7</sup>

D'une façon plus générale, il me semble que l'expérimentation en mathématiques ne se décrète pas, mais apparaît comme un contrôle d'une idée ou une expérience de pensée pour tenter de faire

---

<sup>5</sup>Il y a une hypothèse sous jacente admettant comme évident que la multiplication d'expériences va permettre de faire apparaître un « meilleur » résultat, ce qui n'est pas !

<sup>6</sup>Comme celle citée plus haut

<sup>7</sup>En effet, le premier qui choisit a une probabilité de  $\frac{1}{n+1}$  de gagner, le second gagnera si le premier ne l'a pas fait et donc la probabilité de gagner sera de  $\frac{n}{n+1} \times \frac{1}{n} = \frac{1}{n+1}$ , etc.

apparaître un résultat inaccessible. Elle se met en place en lien avec un contexte précis et dans une problématique explicitée.

### 0.1 Quelques éléments de bibliographie

- Pemantle R., 2007, *A survey of random processes with reinforcement*, Probability Surveys Vol. 4, pp 1-79.
- Peccati G., 2003, *Hoeffding decompositions for exchangeable sequences and chaotic representation of functionals of Dirichlet processes*, Comptes rendus. Mathématique (C. r., Math.) ISSN 1631-073X, vol. 336, no10, pp. 845-850.
- Pólya G., 1931, *Sur quelques points de la théorie des probabilités*. Ann.Inst. H. Poincaré, 1 :117-161.