Les urnes de Polya tricolores

Claire Berthéol et Marion Béral [1]

Cet article est extrait d'un mémoire de 1ère année du Master Enseignement des Mathématiques soutenu le 26 mai 2011 à l'Université Blaise Pascal (Clermont-Ferrand 2).
Il comporte deux versions : l'une destinée au Bulletin et l'autre sur le site de l'APMEP plus complète avec les démonstrations, ainsi que les animations.

L'emploi de modèles d'urne est très ancien. Les premiers travaux remontent à Daniel Bernoulli et Laplace dans la deuxième moitié du XVIIIème siècle. L'ouvrage de Johnson et Kotz « Urn models and their applications » montre l'impressionnante quantité de résultats pouvant être démontrés grâce à ces modèles régis pourtant par des règles simples. On effectue deux types d'actions, piocher des boules dans l'urne ou au contraire en rajouter.

Le modèle de l'urne de Polya utilise les deux. Il a été introduit dans un article paru en 1923 par Polya lui-même et Eggenberger dans le but de modéliser des phénomènes de contagion (maladie ou opinion) dans une population.

On considère une seule urne qui contient des boules de différentes couleurs. L'expérience consiste à tirer au hasard une boule et à la remettre dans l'urne accompagnée d'un nombre fixé de boules de la même couleur. On réitère ce procédé un certain nombre de fois.

Depuis, les urnes bicolores ont fait l'objet de nombreuses études (cf. bibliographie).

I- Etude du modèle

Nous étudierons ici une urne avec des boules de trois couleurs différentes. Cela nous permettra dans la prochaine partie, de représenter graphiquement dans le plan l'évolution de l'urne. On considère donc que l'urne contient a boules rouges, b boules bleues et c boules vertes. A chaque tirage, la boule sera remise dans l'urne accompagnée de λ boules de la même couleur.
Nous supposerons $\lambda \geq 1$, les cas $\lambda=0$ (tirages avec remise) et $\lambda=-1$ (tirages sans remise) devant être étudiés à part.

Comme dans le cas bicolore, on démontre :
1- Par récurrence, que si $a=b=c=\lambda$, les compositions possibles de l'urne à un instant donné sont équiprobables.
2- Que la probabilité d'une suite de tirages ne dépend que du nombre de tirages où chaque couleur est sortie et non de l'ordre de ces sorties.
3- Que la suite des proportions de boules est une chaine de Markov et une martingale. En conséquence, elle converge presque sûrement.

On trouvera sur le site de l'APMEP les démonstrations en annexe.

[1] claire.bertheol@gmail.com ; marion.beral@gmail.com

II- Représentation graphique des tirages dans une urne tricolore

Considérons une urne qui contient a rouges, b bleues et c vertes.
On peut lui associer le point n de $\mathbb{N}^3$ de coordonnées (a,b,c).
Nous appellerons n le <u>contenu</u> de l'urne et poserons a+b+c=t.
Comme (a,b,c) prend des valeurs de plus en plus grandes, cette représentation ne permet pas de visualiser l'évolution de l'urne au-delà de quelques pas.

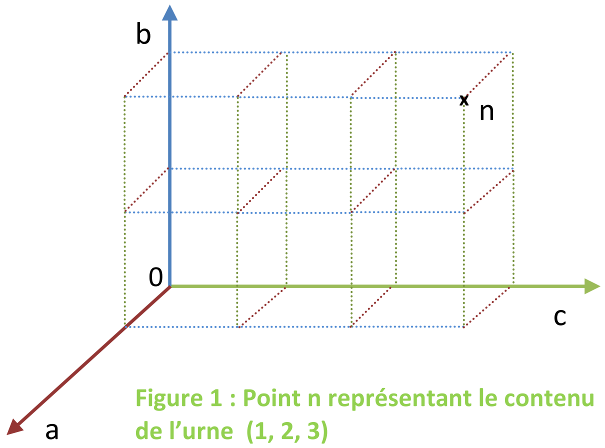


Figure 1 : Point n représentant le contenu de l'urne (1, 2, 3)

Il est donc naturel d'introduire le point m de $\mathbb{R}^3$ de coordonnées $\left(\frac{a}{t}, \frac{b}{t}, \frac{c}{t}\right)$, proportions de boules de chaque couleur.
On a donc $\overrightarrow{\mathrm{Om}} = \frac{1}{t}\,\overrightarrow{\mathrm{On}}$.
Nous appellerons m la <u>composition</u> de l'urne.

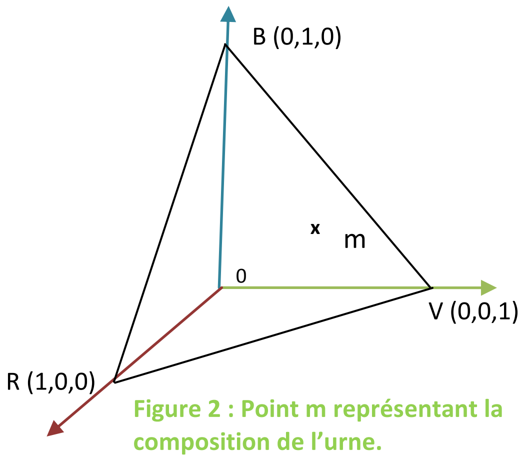


Figure 2 : Point m représentant la composition de l'urne.

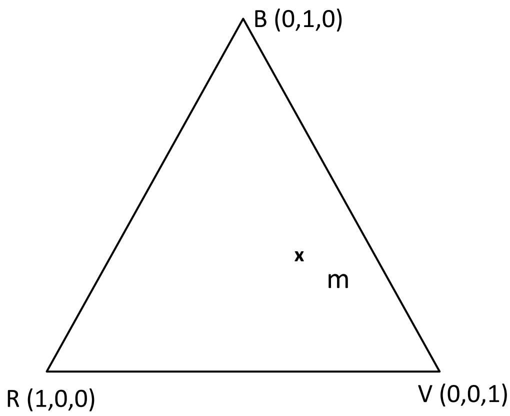


Figure 3 : le triangle RBV en vraie grandeur

Le point m se situe dans le triangle (équilatéral si le repère de départ est orthonormé) défini par $x \geq 0, y \geq 0, z \geq 0$ et $x + y + z = 1$ et de sommets R (1,0,0), B (0,1,0) et V (0,0,1).
On peut écrire (x, y, z) = x (1,0,0) + y (0,1,0) + z (0,0,1) ou $\overrightarrow{\mathrm{Om}} = x\overrightarrow{\mathrm{OR}} + y\overrightarrow{\mathrm{OB}} + z\overrightarrow{\mathrm{OV}}$.
m est donc le barycentre des trois sommets R, B, V affectés des masses x, y, z.
On dit que x, y et z sont les coordonnées barycentriques de m.
On a aussi $a\,\overrightarrow{\mathrm{mR}} + b\,\overrightarrow{\mathrm{mB}} + c\,\overrightarrow{\mathrm{mV}} = \vec{0}$.

Remarquons que si nous ne connaissons pas le nombre total de boules t, nous ne pouvons pas repasser de m à n.

Montrons que les coordonnées barycentriques de m sont proportionnelles à ses distances aux trois côtés du triangle RBV :

Soit I le barycentre de R et de V affectés des masses x et z :
Par associativité des barycentres on obtient que m est barycentre de { (I, x+z) ; (B, y) } donc I appartient à (mB).

On en déduit $\frac{\overline{Im}}{\overline{IB}} = \frac{y}{x+y+z}$, et, par Thalès $\frac{\overline{H_1 m}}{\overline{H_2 B}} = \frac{y}{x+y+z}$ en notant H_1 la projection orthogonale de m sur (RV) et H_2 la projection orthogonale de B sur (RV).

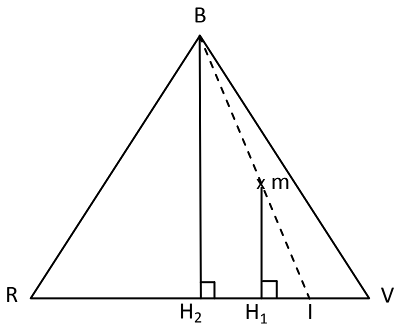


Figure 4

Considérons maintenant deux contenus consécutifs de l'urne représentés par les points n et n'.

On a $\overrightarrow{On'} = \overrightarrow{On} + \lambda\overrightarrow{OE}$ où $\overrightarrow{OE}$ est l'un des trois vecteurs :

$\overrightarrow{OR} = (1,0,0)$; $\overrightarrow{OB} = (0,1,0)$; $\overrightarrow{OV} = (0,0,1)$

Considérons maintenant les points m et m' qui représentent les compositions des deux urnes :

On a $\overrightarrow{Om} = \frac{1}{t}\overrightarrow{On}$ et $\overrightarrow{Om'} = \frac{1}{t+\lambda}\overrightarrow{On'}$ ou $(t+\lambda)\overrightarrow{Om'} = \overrightarrow{On'} = \overrightarrow{On} + \lambda\overrightarrow{OE}$

D'où $(t+\lambda)\overrightarrow{mm'} = t\overrightarrow{Om} + \lambda\overrightarrow{OE} - (t+\lambda)\overrightarrow{Om} = \lambda\left(\overrightarrow{OE} - \overrightarrow{Om}\right) = \lambda\overrightarrow{mE}$

Les vecteurs $\overrightarrow{mm'}$ et $\lambda\overrightarrow{mE}$ sont donc colinéaires, m' appartient au segment [mE] et

$\overrightarrow{mm'} = \frac{\lambda}{t+\lambda}\overrightarrow{mE}$ est d'autant plus petit que t est plus grand.

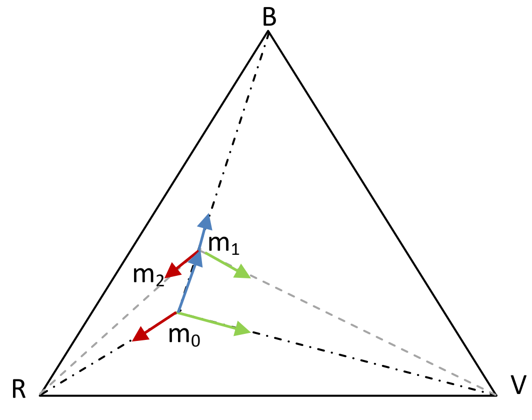


Figure 5 :
- A l'instant initial : m_0= Bar {(R,3) ; (B,1) ; (C,1)}
- Si on tire une boule bleue, m_1 sera sur le segment [m_0B].
- Ensuite, si on tire une boule rouge, m_2 sera le segment [m_1R].

Représentation d'une suite de tirages

Supposons que nous ayons fait p tirages dans une urne de contenu initial (a, b, c) et soient a_p, b_p, c_p le nombre de tirages rouges, bleus et verts (donc $a_p+b_p+c_p=p$).

Le contenu de l'urne avant le $(p+1)^{ième}$ tirage est représenté par le point n_p de $\mathbb{N}^3$ de coordonnées $a+\lambda a_p$, $b+\lambda b_p$, $c+\lambda c_p$.
Les coordonnées de n_p tendent vers l'infini et une telle représentation ne permet pas d'observer la convergence de la suite $\{n_p\}$.
C'est la raison pour laquelle nous associons à cette suite, la suite $\{m_p\}$ des compositions successives de l'urne, définie par $\overrightarrow{Om_p} = \frac{1}{t+\lambda p}\overrightarrow{On_p}$. On peut dire aussi que le point m_p a pour coordonnées : $\frac{a+\lambda a_p}{t+\lambda p}$, $\frac{b+\lambda b_p}{t+\lambda p}$, $\frac{c+\lambda c_p}{t+\lambda p}$.

Remarque : On pourrait aussi s'intéresser au point l_p de coordonnées a_p, b_p, c_p. On a $\overrightarrow{On_p} = \overrightarrow{On_0} + \lambda\overrightarrow{Ol_p}$
La suite $\{n_p\}_{p\geq 1}$ se déduit de la suite $\{l_p\}_{p\geq 0}$ par une homothétie de rapport λ suivie d'une translation de vecteur $\overrightarrow{On_0}$, tandis que la suite $\{l_p\}_{p\geq 0}$ se déduit de la suite $\{n_p\}_{p\geq 1}$ par une translation de vecteur $-\overrightarrow{On_0}$, suivie d'une homothétie de rapport $\frac{1}{\lambda}$.

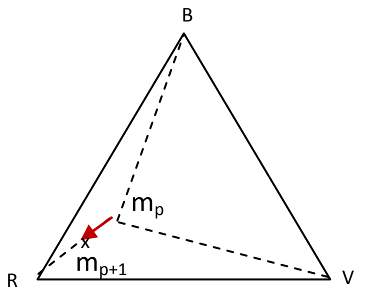
Soit m_p le point représentatif de la composition de l'urne avant le $(p+1)^{ième}$ tirage et M_{p+1} le point (aléatoire) représentatif de la composition avant le tirage suivant.
On a : $\overrightarrow{m_pM_{p+1}} = \overrightarrow{OM_{p+1}} - \overrightarrow{Om_p} = \frac{\lambda}{t+\lambda p}\overrightarrow{OE}$
et le vecteur $\overrightarrow{m_pM_{p+1}}$ peut prendre les trois valeurs $\frac{\lambda}{t+\lambda p}\overrightarrow{OR}$, $\frac{\lambda}{t+\lambda p}\overrightarrow{OB}$ ou $\frac{\lambda}{t+\lambda p}\overrightarrow{OV}$, avec les probabilités $\frac{a+\lambda a_p}{t+\lambda p}$, $\frac{b+\lambda b_p}{t+\lambda p}$, $\frac{c+\lambda c_p}{t+\lambda p}$.

Pour simuler le processus, on utilise le logiciel Scilab. Celui-ci va générer une suite de nombres pseudo-aléatoires $(U_n)_{n>0}$. Cela signifie que cette suite possède certaines propriétés d'une suite de tirages indépendants d'une variable uniforme sur [0 ;1]. En effet, un ordinateur est incapable de hasard. Il génère donc une suite complètement déterministe. Scilab utilise le modèle suivant : $U_{n+1}=(843314861.U_n + 453816693)$ modulo 2^{31}.
Pour que U_0 soit aléatoire, on lui donnera la date système de l'ordinateur à l'instant où la simulation est lancée (sauf si l'on veut refaire la même simulation que la fois précédente).
On compare alors U_{p+1} aux deux nombres $\frac{a+\lambda a_p}{t+\lambda p}$ et $1-\frac{c+\lambda c_p}{t+\lambda p}$.

Figure 6 : Couleur de la boule en fonction de la place du nombre aléatoire dans l'intervalle [0;1]

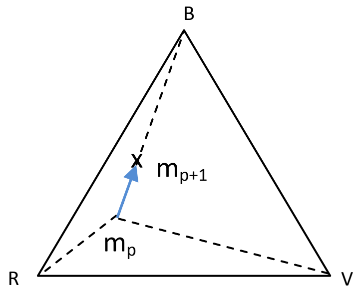
Si $U_{p+1} \leq \frac{a+\lambda a_p}{t+\lambda p}$ (tirage d'une boule rouge) :

$$\begin{cases} a_{p+1} = a_p + 1 \\ b_{p+1} = b_p \\ c_{p+1} = c_p \end{cases}$$



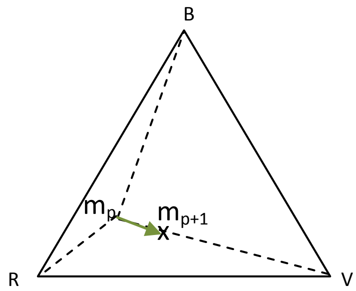
Si $\frac{a+\lambda a_p}{t+\lambda p} < U_{p+1} \leq 1 - \frac{c+\lambda c_p}{t+\lambda p}$ (tirage d'une boule bleue) :

$$\begin{cases} a_{p+1} = a_p \\ b_{p+1} = b_p + 1 \\ c_{p+1} = c_p \end{cases}$$



Si $1 - \frac{c+\lambda c_p}{t+\lambda p} < U_{p+1} \leq 1$ (tirage d'une boule verte) :

$$\begin{cases} a_{p+1} = a_p \\ b_{p+1} = b_p \\ c_{p+1} = c_p + 1 \end{cases}$$



III- Application en seconde

D'après le **programme officiel** (2009), les élèves de seconde devront être capables de « concevoir, mettre en œuvre et exploiter des simulations de situations concrètes à l'aide du tableur ou d'une calculatrice. »

Un des objectifs du programme « est d'amener les élèves à un questionnement lors des activités suivantes :

- l'estimation d'une proportion inconnue à partir d'un échantillon ;
- la prise de décision à partir d'un échantillon. »

Il nous a donc paru intéressant d'introduire l'urne de Polya en seconde dans l'objectif de travailler sur les notions d'échantillon, de fluctuation d'échantillonnage et de fréquence. Un des objectifs de ce travail sera de mettre en évidence la variabilité des résultats obtenus par des expériences limitées. Par ailleurs, les élèves pourront simuler sur tableur l'échantillon et analyser les résultats ce qui fait partie des nouvelles compétences exigées par le programme de 2009.

Objectifs :

Dans une urne sont placées une boule rouge, une boule bleue et une boule verte. On tire une boule au hasard que l'on remet dans l'urne accompagnée d'une boule de la même couleur. L'objectif est d'étudier la dynamique de l'urne c'est-à-dire comment évolue la composition de l'urne après n tirages.

Nous avons vu plus haut, qu'en effectuant l'expérience n fois, toutes les compositions possibles de l'urne sont équiprobables. Aussi, si l'on continue les tirages, la proportion de boules rouges dans l'urne va se stabiliser vers une valeur quelconque comprise entre 0 et 1. Nous allons étudier avec les élèves ces deux approches de la situation.

L'étude que nous allons soumettre aux élèves ne peut être réalisée dans de bonnes conditions que si les élèves ont préalablement travaillé sur des situations modélisées par une urne. On pourra par exemple travailler sur l'urne de Bernoulli dans un problème du type : « On choisit au hasard un élève dans une classe comportant 12 garçons et 15 filles. Quelle est la probabilité que ce soit un garçon ? ».

1ère étude : étude des probabilités des compositions de l'urne à une étape donnée

Déroulement de la séquence :

1ère séance :

Premier énoncé donné aux élèves:

> *Problème :*
> ***3** savonnettes identiques mais de senteur différente (anis, vanille, jasmin) sont sur le marché. A la suite d'une campagne publicitaire, il y a **1** acheteur pour chaque parfum. Ensuite tout nouvel acheteur se fie au conseil du premier acheteur rencontré.*
> *Quel produit va être le plus vendu ?*
>
> *1ère partie : Comment modéliser le problème ?*

Le premier travail des élèves sera de réfléchir par groupes de trois élèves sur la modélisation du problème. Le professeur discutera ensuite des modélisations proposées avec chaque groupe et institutionnalisera la modélisation suivante :

Pour modéliser ce problème, nous allons introduire une urne contenant des boules de 3 couleurs différentes représentant les 3 savonnettes : rouge, bleue et verte.
Au départ, nous plaçons seulement une boule de chaque couleur dans l'urne.
On choisit au hasard une boule : si cette boule est rouge (respectivement bleue, verte), on remet dans l'urne la boule tirée accompagnée d'une boule rouge (respectivement bleue, verte). La boule rajoutée à chaque tirage représente le nouvel acheteur influencé par le précédent.

Deuxième énoncé donné aux élèves :

<u>2ème partie : étude de l'urne après 3 tirages :</u>
1°) Combien y a-t-il de boules dans l'urne avant le quatrième tirage ?
2°) Quelles sont les différentes compositions possibles de l'urne ?
3°) A-t-on les mêmes chances d'obtenir chacune de ces compositions ?

Les élèves devront répondre aux questions suivantes : ils répondront aux deux premières questions individuellement et ensuite les élèves discuteront de la troisième question.
Après, chaque élève devra effectuer l'expérience 5 fois avec des billes. Les résultats seront inscrits dans le tableau ci-dessous dans la partie « effectif »:

	1	2	3	4	5	6	7	8	9	10
Compostions de l'urne juste avant le 4ème tirage	(1R,1B,4V)	(1R,2B,3V)	(1R,3B,2V)	(1R,4B,1V)	(2R,1B,3V)	(2R,2B,2V)	(2R,3B,1V)	(3R,1B,2V)	(3R,2B,1V)	(4R,1B,1V)
Effectif										
Fréquence										
Fréquence théorique	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1

Figure 8 : Tableau à remplir par les élèves

Ensuite, ils devront calculer les fréquences d'apparition de chaque configuration à l'aide de leurs résultats.

<u>2ème séance</u> :

Dans la deuxième séance, les élèves devront analyser les résultats obtenus la séance d'avant : ceci passera par le calcul de la fréquence théorique si toutes les compositions de l'urne avaient la même probabilité.
Les élèves pourront ensuite simuler une expérience sur tableur qui consiste à tirer 3 fois dans l'urne et relever la configuration finale de l'urne.

état de l'urne	boules rouges	boules bleues	boules vertes
à l'état initial	1	1	1
après le1er tirage	1	2	1
après le 2ème tirage	2	2	1
après le 3ème tirage	3	2	1

Proportion de rouges	Proportion de bleue	Proportion de vertes
0,333333333	0,333333333	0,333333333
0,25	0,5	0,25
0,4	0,4	0,2
0,5	0,333333333	0,166666667

Figure 9 : Simulation de 3 tirages

Le professeur pourra ensuite simuler sur tableur mille expériences et relever les résultats dans ce tableau :

	1	2	3	4	5	6	7	8	9	10
Compostions de l'urne juste avant le 4ème tirage	(1R,1B,4V)	(1R,2B,3V)	(1R,3B,2V)	(1R,4B,1V)	(2R,1B,3V)	(2R,2B,2V)	(2R,3B,1V)	(3R,1B,2V)	(3R,2B,1V)	(4R,1B,1V)
Effectif	72	111	93	98	108	101	99	111	105	102
Fréquence	0,07	0,11	0,09	0,10	0,11	0,10	0,10	0,11	0,11	0,10
Fréquence théorique	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1

Figure 10 : tableau des fréquences pour 1000 expériences

En cumulant ces expériences, on fait apparaître une « convergence » des fréquences vers la valeur théorique.

Il est intéressant de voir le cas où l'urne de départ contient 2 boules rouges, 2 boules bleues et 2 boules vertes : en effet, nous avons vu que contre toutes attentes, les configurations de l'urne à un instant donné n'étaient pas équiprobables dans cette situation.
Nous avons donc prévu de poser la question suivante aux élèves : « Et avec une urne de départ avec 2 boules rouges, 2 boules bleues et 2 boules vertes, pensez-vous que les configurations possibles de l'urne sont toutes équiprobables ? »
On attend que les élèves nous répondent qu'ils pensent que oui.
Ils procèdent alors comme précédemment, en simulant cette expérience sur tableur :

état de l'urne	boules rouges	boules bleues	boules vertes
à l'état initial	2	2	2
après le1er tirage	3	2	2
après le 2ème tirage	3	2	3
après le 3ème tirage	3	2	4

Proportion de rouges	Proportion de bleue	Proportion de vertes
0,333333333	0,333333333	0,333333333
0,428571429	0,285714286	0,285714286
0,375	0,25	0,375
0,333333333	0,222222222	0,444444444

Figure 11 : Simulation de 3 tirages avec 2 boules de chaque couleur

Le professeur simule ensuite un millier d'expériences et relève les résultats dans ce tableau :

	1	2	3	4	5	6	7	8	9	10
Compostions de l'urne juste avant le 4ème tirage	(2R,2B,5V)	(2R,3B,4V)	(2R,4B,3V)	(2R,5B,2V)	(3R,2B,4V)	(3R,3B,3V)	(3R,4B,2V)	(4R,2B,3V)	(4R,3B,2V)	(5R,2B,2V)
Effectif	58	100	105	76	113	129	111	113	135	60
Fréquence	0,06	0,10	0,11	0,08	0,11	0,13	0,11	0,11	0,14	0,06

Figure 12 : Résultat de 100 séries de 3 tirages

On fait ainsi apparaître que les fréquences obtenues sont éloignées de celle qu'on pourrait attendre si les compositions possibles de l'urne étaient équiprobables (c'est-à-dire 0,1).

2ème étude : étude de l'évolution de la composition de l'urne à long terme

L'objectif de la séquence est d'étudier la composition de l'urne au bout d'un grand nombre de tirages. Nous allons plus particulièrement nous intéresser à la proportion de boules rouges et voir s'il y a « convergence ».
D'après le programme officiel de 2009, les élèves de seconde devront être capables de « concevoir, mettre en œuvre et exploiter des simulations de situations concrètes à l'aide du tableur ou d'une calculatrice. »
L'étude de l'urne de Polya est propice à l'utilisation de ces outils : en effet, la simulation de l'urne à long terme est possible sur calculatrice et sur tableur.
Ainsi en fonction des moyens à disposition, le professeur aura le choix entre ces deux outils.

Sur calculatrice :

Enoncé donné aux élèves :

> Objectif : observer l'évolution de la proportion des boules rouge dans l'urne après de nombreux tirages.
>
> 1°) Ecrire un algorithme sur calculatrice permettant de simuler l'expérience autant de fois que l'on veut avec la composition initiale de l'urne au choix.
> 2°) Observer la suite de proportions de boules rouge obtenues. Que remarquez-vous ?
> Est-ce toujours le cas ?
> 3°) Que se passe t-il lorsque la composition initiale de l'urne change ?

On note R_i la proportion de boules rouges dans l'urne après le i^{ième} tirage.
La simulation sur calculatrice consiste à réaliser n tirages et retourner la suite $(R_1, R_2, \dots, R_n)$ des proportions de rouges dans l'urne.

Algorithme :

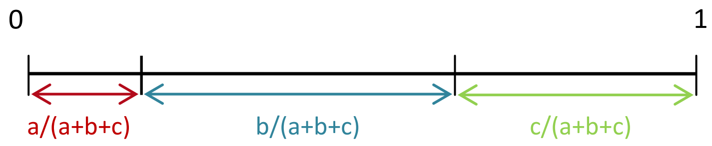


Figure 13 : Subdivision de l'intervalle [0;1] en 3 parties

Entrer a, b et c ; ← nombres de boules dans l'urne au départ

Entrer le nombre n de tirages ;

Pour k variant de 1 à n faire

    Si aléa ≤ a/(a+b+c) faire

        a ← a+1 ;

    Sinon

        Si [aléa ≥ a/(a+b+c) et aléa ≤ (a+b)/(a+b+c)] faire

            b ← b+1 ;

        Sinon faire

            c ← c+1 ;

Afficher a/(a+b+c) ; ← cela retourne la suite des proportions de boules rouges dans l'urne

Fin ;

Cet algorithme permet de choisir la composition initiale de l'urne et le nombre de tirages que l'on veut effectuer. Si l'on choisit a=b=c=1, et n=60, l'urne de départ contient une boule de chaque couleur, et on fait 60 tirages. On peut ainsi observer la suite de proportions obtenues $(R_1, R_2, \dots, R_{60})$: on remarquera que cette suite se stabilise autour d'une valeur.
Les élèves pourront donc conjecturer que la suite converge.

Ensuite, on exécutera plusieurs fois ce programme pour voir si la suite se stabilise toujours autour de la même valeur (ce qui ne sera pas le cas).

Le professeur pourra poser aux élèves la question :
« Que se passe t-il lorsque la composition initiale de l'urne change ?»
Les élèves exécuteront alors l'algorithme en changeant la composition initiale de l'urne et observeront que la suite converge toujours.

Sur tableur :

Enoncé donné aux élèves :

Objectif : *observer l'évolution de la proportion des boules rouges dans l'urne après de nombreux tirages.*

1°) Simuler sur tableur l'expérience qui consiste à tirer un grand nombre de fois une boule dans l'urne.
2°) Observer la suite de proportions de boules rouge obtenues. Que remarquez-vous ?
Est-ce toujours le cas ?
3°) Comment se répartissent ces valeurs « limites » ?
4°) Que se passe t-il lorsque la composition initiale de l'urne change ?

Les élèves devront comme dans la première étude simuler l'expérience mais cette fois-ci sur une grande suite de tirages. Ils vont ensuite observer la suite de proportion de boules rouges obtenues et la représenter sur un graphique.

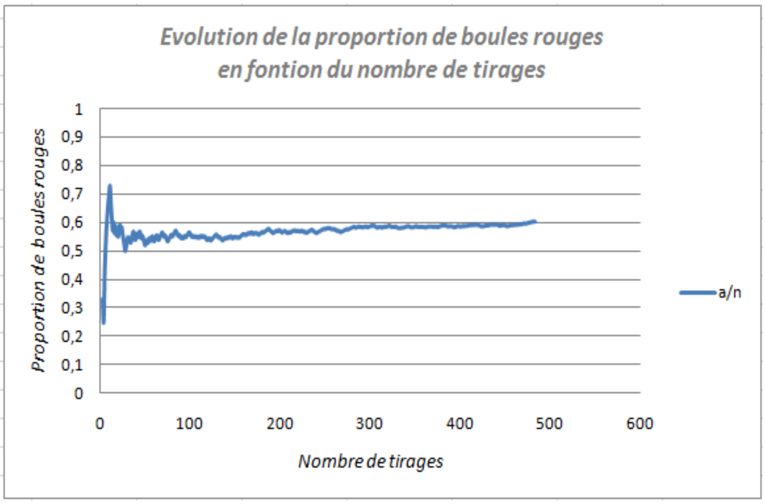


Figure 14 : Graphique associé à la feuille Excel précédente

Remarque : si on étudie seulement l'évolution de la proportion de boules d'une seule couleur, on se rend compte que ce résultat est semblable à celui d'une urne bicolore (en rassemblant les 2 couleurs restantes).

Ensuite, pour répondre à la question n°3 sur la répartition des valeurs « limites », les élèves pourront exécuter plusieurs fois la simulation et noter les valeurs « limites ».
Une mise en commun sera faite et le professeur pourra ensuite représenter leur répartition par un diagramme en bâtons.

Compte tenu du nombre insuffisant de « limites » obtenues, le professeur pourra simuler lui-même (grâce à une macro sur Excel) 250 séries de tirages. Il montrera ensuite aux élèves le diagramme en bâton obtenu avec ces 250 limites.

Animations Scilab

Nous avons créé deux programmes Scilab différents utilisant la modélisation barycentrique. Tous deux génèrent eux-mêmes des nombres aléatoires. Le premier permet de visualiser la trajectoire du point M au fur et à mesure des tirages, alors que le second affiche les points « limites » de plusieurs séries de nombreux tirages.
Il est important que les élèves puissent eux-mêmes manipuler ces programmes : les élèves, ouvrent les fichiers des animations grâce au logiciel Scilab. Dans les deux animations, les élèves doivent renseigner les conditions de l'expérience.

Dans un premier temps, nous nous intéressons à la trajectoire du barycentre pour des séries de 10 tirages, où la remise est de 1 boule. (cf. figure 15)
Les élèves, devront se rendre compte, après avoir réalisé quelques fois l'expérience dans les mêmes conditions, que la trajectoire et les propriétés asymptotiques de celles-ci ne sont pas les mêmes à chaque fois.
Ensuite, il leur sera demandé de réaliser des séries de 150 tirages. (cf. figure 16)

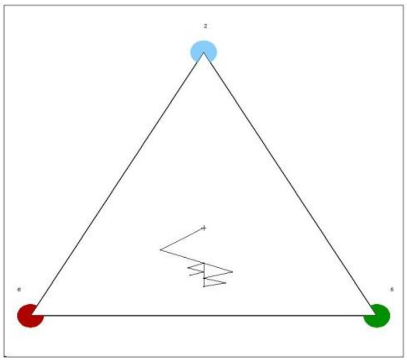


Figure 15 : trajectoire du point M pour une série de 10 tirages

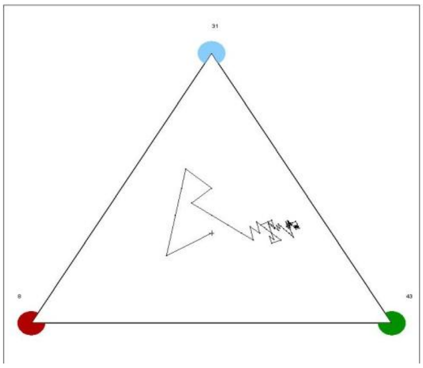


Figure 16 : trajectoire pour une série de 150 tirages

On remarquera qu'il y a à chaque fois convergence vers un point « limite ». Même si la notion de convergence n'est pas approfondie en lycée, on peut évoquer le terme de "point limite", assez intuitif sur les figures.

Dans un troisième temps, on se concentre sur ces points « limites ». Il faudra donc utiliser la deuxième animation. Là aussi les élèves renseigneront les conditions de l'expérience.
Voici par exemple le résultat de 500 séries de 150 tirages dans le cas où l'urne contient au départ une boule de chaque couleur en rajoutant une boule à chaque tirage :

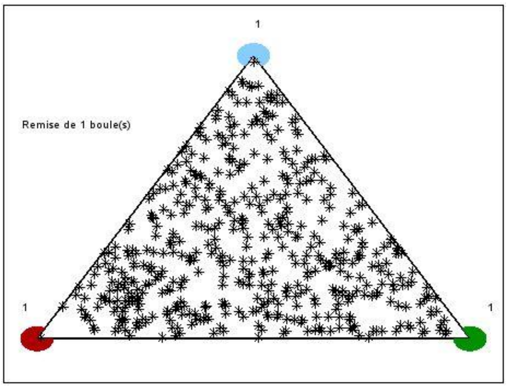


Figure 17 : Points "limites" de 500 séries de 150 tirages avec une boule de chaque couleur au départ et ajout d'une boule

Animation Flash

Nous avons aussi créé une animation Flash qui récapitule toutes les expériences menées en classe. Elle inclut des résultats avec des nombres de tirage plus élevés qui nécessitent trop de temps (environ une heure) pour une application en classe.
Les figures suivantes représentent par exemple les points limites de 3000 séries de 150 tirages, avec des conditions initiales différentes.

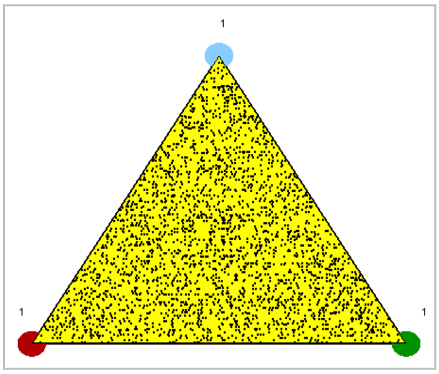


Figure 18 : Points limites de 3000 séries de 150 tirages avec 1 boule de chaque couleur au départ, et ajout d'une boule

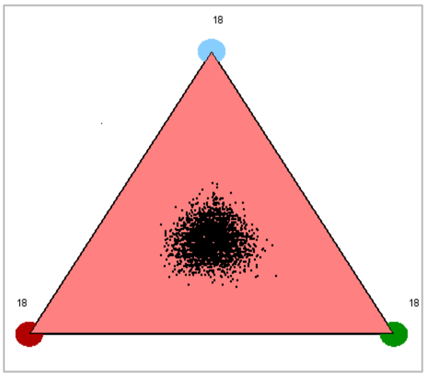


Figure 19: Points limites de 3000 séries de 150 tirages avec 18 boules de chaque couleur au départ et ajout d'une boule

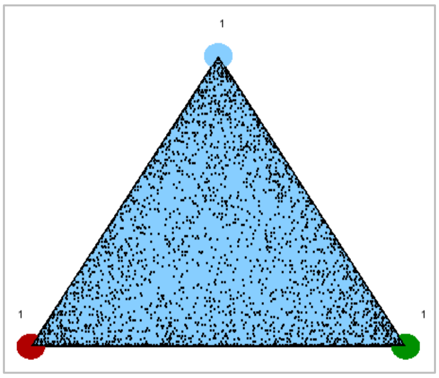


Figure 20 : Points limites de 3000 séries de 150 tirages avec 1 boule de chaque couleur au départ, et ajout de 2 boules

Cette animation comporte quatre séquences :

- Les tirages et leur représentation (présentation de l'expérience de l'urne de Polya tricolore).
- Evolution à partir de 3 boules distinctes (répétition d'une même expérience).
- Influence de la composition initiale.
- Deux d'un coup (évolution de l'urne lorsque l'on rajoute deux boules à chaque tirage).

Conclusion :

On peut aussi introduire l'urne de Polya en classe de première dans l'objectif de travailler sur des points du programme officiel (2009) tels que les lois de probabilité, la représentation de l'expérience par un arbre pondéré, la simulation de l'expérience, la fluctuation d'échantillonnage, les fréquences.
La séquence présentée en classe de seconde peut bien évidemment être reprise en classe de première. Par ailleurs, il serait intéressant d'inclure dans le travail des élèves, la représentation de l'expérience par un arbre pondéré (cf annexe 1).

En outre, si l'on décide de réaliser cette séquence dans une classe de terminale, il sera intéressant de rajouter des calculs de probabilités conditionnelles et pour la filière scientifique un travail sur des variables aléatoires avec des calculs d'espérance et de variance.

Bibliographie

- Urn models and their applications, Norman Lloyd Johnson, Samuel Kotz, Wiley, 1977

- Sur quelques points de la théorie des probabilités, G. POLYA
  Annales de l'I.H.P, tome 1, n°2 (1930), p117-161.
  Polya y évoque son modèle d'urne.

- Modèles d'urnes et phénomènes de seuils en combinatoire analytique : Thèse de Vincent Puyhaubert, INRIA Rocquencourt, mars 2005
  http://www.imprimerie.polytechnique.fr/Theses/Files/Puyhaubert.pdf

- « Les urnes de Polya » de Pierre GRIHON du lycée Montaigne-BORDEAUX
  2009 Bulletin de l'APMEP. Num. 485. p. 737-746
  Cet article résume le travail d'élèves de seconde et de première sur une urne de Polya bicolore.

- « Les urnes de Polya en classe de seconde », Gilles Aldon, 21 janvier 2008
  http://irem-fpb.univ-lyon1.fr/feuillesprobleme/feuille11/dansnosclasses/article.pdf
  Cet article présente une séquence d'enseignement sur les urnes de Polya bicolores en seconde.

- 4 films « URNE DE POLYA » de Josèphe Badrikian et Paul-Louis Hennequin, IREM de Clermont-Ferrand, avril 1976.

- Document d'accompagnement des 4 films « URNE DE POLYA » à l'usage des enseignants, Paul-Louis Hennequin, IREM de Clermont-Ferrand, avril 1976.