

Ressources pour la classe de première
générale et technologique

Statistiques et probabilités

Ces documents peuvent être utilisés et modifiés librement dans le cadre des activités d'enseignement scolaire, hors exploitation commerciale.

Toute reproduction totale ou partielle à d'autres fins est soumise à une autorisation préalable du Directeur général de l'enseignement scolaire.

La violation de ces dispositions est passible des sanctions édictées à l'article L.335-2 du Code de la propriété intellectuelle.

février 2012

Introduction

La place des probabilités et des statistiques dans l'enseignement des mathématiques en collège et en lycée s'est considérablement accrue depuis ces dernières années. Pour les élèves entrant en classe de première, l'apprentissage des probabilités débute désormais dès la classe de troisième.

Au collège, l'objectif de cet enseignement est de développer une réflexion sur l'aléatoire en général et de sensibiliser les élèves au fait que les situations aléatoires peuvent faire l'objet d'un traitement mathématique. Un vocabulaire spécifique est introduit et quelques règles du calcul des probabilités sont mises en place.

La classe de seconde est l'occasion pour l'élève d'approfondir la formalisation de ces notions en dégagant notamment la notion de modèle probabiliste, et d'être sensibilisé, à travers des situations de prise de décision ou d'estimation d'une proportion, aux premiers éléments de statistique inférentielle comme la notion d'intervalle de fluctuation et celle d'intervalle de confiance, introduites sous des conditions de validité qui les rendent rapidement opérationnelles.

Avec la notion de variable aléatoire et la découverte de la loi binomiale, le programme de Première fournit les outils mathématiques qui permettent, en prenant appui sur la réflexion initiée en Seconde autour de la prise de décision, de construire un intervalle de fluctuation et d'établir une démarche de prise de décision valables en toute généralité pour une proportion et une taille d'échantillon quelconques. Ce thème se prête en particulier à la mise en œuvre d'algorithmes et de raisonnements logiques et, au-delà, à une adaptation de ces raisonnements au domaine de l'aléatoire et de l'incertain.

En Terminale, la problématique de prise de décision sera travaillée à nouveau, et la réflexion initiée en Seconde sur l'estimation sera approfondie avec l'introduction d'outils mathématiques supplémentaires.

Dans ce document ressource, le professeur trouvera des compléments théoriques et un ensemble de situations développées dans le cadre du programme officiel. L'accent est surtout mis sur les notions nouvelles par rapport aux précédents programmes de Première : répétition d'expériences identiques et indépendantes, loi géométrique tronquée, loi binomiale, échantillonnage et prise de décision avec la loi binomiale.

Les exemples d'application ont été choisis pour montrer la variété, la richesse et l'actualité des applications possibles des probabilités et de la statistique. Ils ne prétendent pas à l'exhaustivité et ne sont pas conçus comme des activités pédagogiques « clé en main », tout comme le plan adopté pour les exposer ne se veut pas une progression pédagogique. Ces situations visent plutôt à ouvrir des pistes de travail susceptibles d'être exploitées par le professeur ; c'est pourquoi elles sont traitées de façon suffisamment détaillée afin de permettre au professeur de s'en inspirer pour élaborer, à partir de la connaissance de sa classe et de sa pratique professionnelle, des activités pédagogiques ajustées au niveau de ses élèves.

Enfin les points présentés dans les annexes du document ne sont pas des attendus du programme. Ils doivent être considérés comme des compléments d'information à l'attention du professeur sur les notions introduites. Ils permettent de mieux situer le cadre mathématique plus général dans lequel s'inscrivent les notions au programme.

Table des matières

I.	Statistique descriptive, analyse de données.....	4
II.	Variables aléatoires discrètes	5
III.	Utilisation des arbres pondérés.....	7
A.	Exemple d'expérience aléatoire à deux épreuves.....	7
B.	Justification de l'arbre des probabilités.....	9
1.	Étape 1	9
2.	Étape 2	9
3.	Étape 3	10
C.	Généralisation et exploitation en Première.....	11
IV.	Loi géométrique tronquée	12
A.	Étude de la loi géométrique tronquée.....	12
1.	Approche de la loi géométrique tronquée	12
2.	Définition de la loi géométrique tronquée	13
3.	Expression de la loi géométrique tronquée	13
4.	Algorithme de simulation	13
5.	Représentation graphique.....	15
6.	Espérance de la loi géométrique tronquée	16
B.	Exemples d'activités	17
1.	Limitation des naissances	17
2.	Le paradoxe de Saint-Petersbourg	19
V.	Loi binomiale	22
A.	Définitions.....	22
1.	Approche de la loi binomiale.....	22
2.	Définition de la loi binomiale	24
3.	Coefficients binomiaux.....	24
B.	Propriétés.....	25
1.	Expression de la loi binomiale.....	25
2.	Propriétés des coefficients binomiaux	25
3.	Représentation graphique.....	27
4.	Espérance et écart-type	27
C.	Exemples d'activités	30
1.	Avec la loi de probabilité.....	30
2.	Avec l'espérance mathématique	31
VI.	Échantillonnage et prise de décision	32
A.	Intervalle de fluctuation avec la loi binomiale	32
B.	Aspect général de la prise de décision avec la loi binomiale	34
C.	Détermination de l'intervalle de fluctuation à l'aide d'un algorithme	34
D.	Exemples d'activités	36
1.	Exemple 1 : politique dans un pays lointain	36
2.	Exemple 2 : discrimination	37
3.	Exemple 3 : sécurité au carrefour	38
4.	Exemple 4 : gauchers	38

5.	Exemple 5 : ségrégation sexiste à l'embauche.....	39
6.	Exemple 6 : générateur de nombres aléatoires.....	39
E.	Lien avec l'intervalle de fluctuation exploité en classe de Seconde	39
VII.	Annexes.....	42
A.	Couple d'indicateurs et problèmes de minimisation	42
1.	Position du problème	42
2.	Étude des trois fonctions d_1 , d_2 et d_∞	42
B.	Loi faible des grands nombres.....	44
C.	Espérance de la loi géométrique tronquée : approches expérimentales	45
1.	Calculatrice	45
2.	Algorithme modifié sur Algobox.....	45
3.	Point de vue numérique	46
4.	Point de vue graphique.....	46
D.	Loi géométrique	47
1.	Quelques propriétés de la loi géométrique.....	47
2.	Précautions en classe de Première	47
3.	Une approche à l'aide de l'algorithmique.....	47
4.	Langage naturel.....	47
5.	Sur calculatrice.....	47
6.	Sous Algobox.....	48
7.	Sous Scilab.....	48
E.	Quelques outils de calcul avec la loi binomiale	48
1.	Tableur	48
2.	Un logiciel : Scilab.....	49
3.	Deux modèles de calculatrice	50
F.	Coefficients binomiaux et quadrillage	51
1.	De l'arbre au quadrillage	51
2.	Interprétation des coefficients binomiaux.....	53
3.	Formule de Pascal	53
4.	Application : le problème des pilules.....	53
5.	Simulation de l'expérience sur Algobox.....	54
6.	Traitement mathématique	54
G.	Compléments sur la prise de décision	56
1.	L'affaire Woburn	56
2.	Radioactivité ou bruit de fond ?.....	61
3.	Cartes de contrôle	63

I. Statistique descriptive, analyse de données

L'étude et la comparaison de séries statistiques menées en classe de seconde se poursuivent avec la mise en place de nouveaux outils. Dans un premier temps, les caractéristiques de dispersion (variance, écart-type) sont déterminées à l'aide de la calculatrice ou d'un tableur. Afin d'utiliser de façon appropriée les deux couples d'indicateurs usuels (moyenne/écart-type et médiane/écart interquartile) qui permettent de résumer une série statistique, il semble utile de rappeler le lien entre ces couples (position/dispersion) et un problème de minimisation (voir annexe 1). Il convient aussi de rappeler que l'utilisateur d'un outil statistique doit prendre en compte la situation réelle et les objectifs visés pour effectuer le choix des indicateurs de façon pertinente.

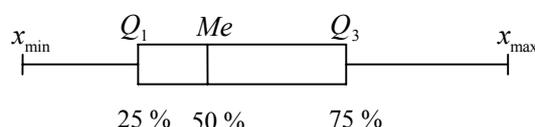
Les exemples de séries statistiques amènent à utiliser l'un des deux couples à notre disposition. Ils suscitent une réflexion sur le choix d'un résumé statistique. Dans les exemples proposés en classe, il est important de faire remarquer que deux séries de même écart-type (et de même moyenne et médiane) peuvent avoir une distribution très différente. C'est alors l'occasion de rappeler l'intérêt d'un graphique, qui peut être plus « parlant » qu'un simple résumé numérique.

Il n'existe pas de règle (au sens mathématique) qui indiquerait quel type d'indicateur statistique utiliser par rapport à une situation donnée. Le choix des indicateurs dépend de ce qu'on veut en faire et de la réalité de la situation. On peut juste proposer quelques remarques qui conduisent à privilégier tel couple plus que tel autre. Le couple (médiane, écart interquartile), sans apporter les mêmes renseignements que le couple (moyenne, écart-type), est peu sensible aux valeurs extrêmes. Dans de nombreux domaines il est privilégié et souvent associé à une représentation graphique en boîte à moustaches. De manière générale, la moyenne arithmétique est peu significative quand l'influence des valeurs extrêmes est trop forte. Quant à la médiane, elle ne se prête pas aux calculs algébriques, c'est pourquoi, dans le cas où la série statistique est formée de divers sous-ensembles homogènes, on lui préfère la moyenne.

Le diagramme en boîte (ou boîte à moustaches) est une représentation graphique qui permet d'avoir une bonne vision d'une série statistique. En effet, beaucoup d'informations sont disponibles sur ce diagramme (médiane, écart interquartile et valeurs extrêmes), ce qui en fait un très bon outil pour comparer deux séries statistiques. Il faut noter qu'il n'existe pas de définition commune (au sens mathématiques du terme) du diagramme en boîte, mais il semble assez répandu d'utiliser les conventions suivantes :

- la « boîte » est un rectangle limité par le premier et le troisième quartile où figure la médiane ;
- les « moustaches » en revanche peuvent s'achever aux valeurs extrêmes (le minimum et le maximum de la série) ou aux premier et neuvième déciles ¹. D'autres conventions sont quelquefois utilisées.

On obtient alors un diagramme comme suit :



Au-delà de la réalisation d'un diagramme en boîte, il est surtout important de savoir interpréter et d'utiliser ces diagrammes pour des comparaisons pertinentes de deux séries statistiques.

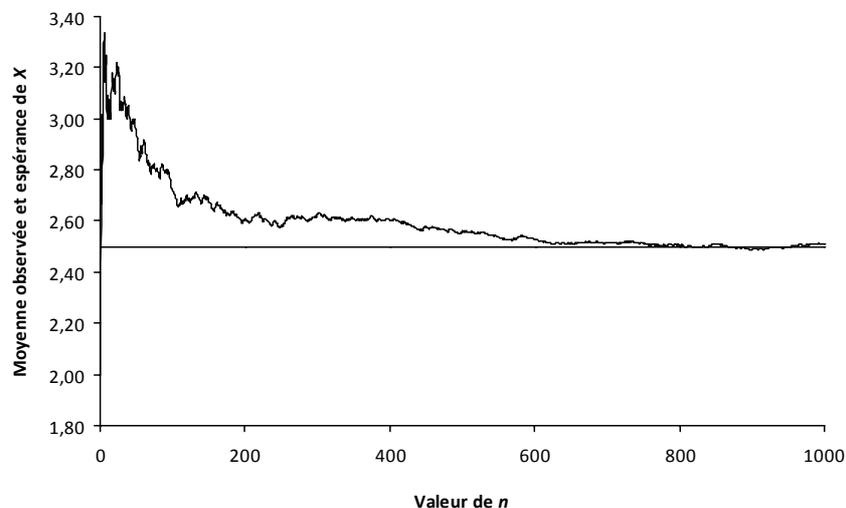
¹ Définition du décile D_k : pour k de 1 à 9, le k^{e} décile noté D_k est la plus petite valeur d'une série statistique telle qu'au moins $(k \times 10)$ % des valeurs de la série sont inférieures ou égales à D_k .

II. Variables aléatoires discrètes

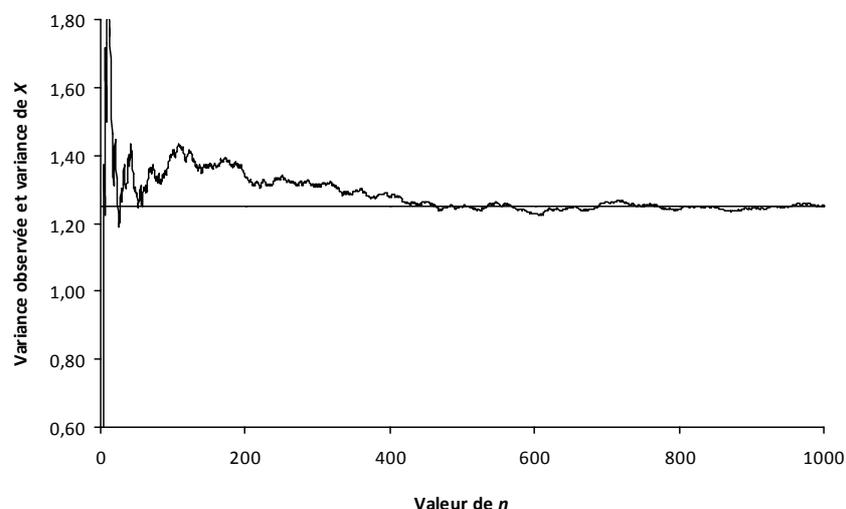
Afin d'interpréter l'espérance comme la valeur moyenne dans le cas d'un grand nombre de répétitions, on considère l'expérience aléatoire consistant à lancer un dé supposé équilibré à six faces et à noter le numéro observé. On considère ensuite la variable aléatoire discrète notée X qui prend la valeur 1 si on observe 1, la valeur 2 si on observe 2, 3 ou 4 et enfin la valeur 4 si on observe 5 ou 6. Son espérance est

$E(X) = 1 \times P(X = 1) + 2 \times P(X = 2) + 4 \times P(X = 4) = 1 \times 1/6 + 2 \times (1/6 + 1/6 + 1/6) + 4 \times (1/6 + 1/6) = 15/6$, soit $E(X) = 2,5$.

À l'aide d'une simulation, on répète un grand nombre de fois cette expérience aléatoire à l'identique et on peut ainsi observer un grand nombre de réalisations de la variable aléatoire X . Le graphique suivant montre l'évolution de la moyenne observée en fonction du nombre n de répétitions.



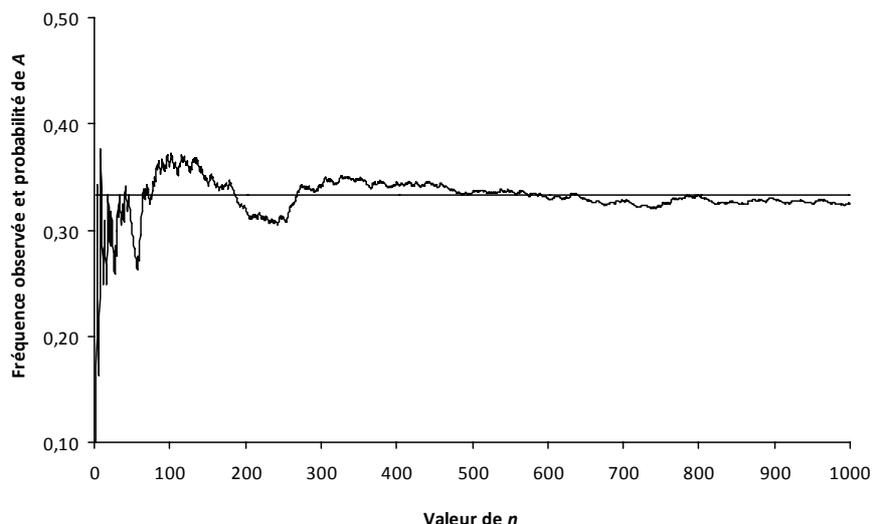
On remarque que les moyennes observées se stabilisent autour de l'espérance mathématique de la variable aléatoire X . On peut aussi représenter l'évolution de la variance des observations et remarquer que lorsque le nombre de lancers augmente, la variance observée se stabilise vers la variance de la variable aléatoire X qui vaut 1,25.



Ces observations constituent une approche heuristique de la loi des grands nombres². Celle-ci permet de justifier le phénomène de stabilisation des fréquences autour de la probabilité d'un événement.

Plus généralement, on se place dans un modèle probabiliste ; on considère un événement A de probabilité $P(A)$ et la variable aléatoire X qui prend la valeur 1 si on observe A et 0 sinon. La variable aléatoire X suit la loi de Bernoulli de paramètre p qui est égale à $P(A)$. Une simulation permet d'observer le phénomène de stabilisation de la suite des fréquences³ observées f_n de réalisation de l'événement A , lors de n répétitions de la même expérience aléatoire, vers l'espérance de X qui est égale à $P(A)$.

La simulation qui a donné le graphique suivant a été réalisée pour un événement A de probabilité $p=1/3$.



Ainsi le phénomène de stabilisation (expression du registre du langage courant pour dire qu'une suite de réels converge) est l'illustration de la loi des grands nombres et ce « phénomène » n'est justifiable que lorsque le modèle probabiliste est donné.

² Un énoncé et une preuve de la loi faible des grands nombres sont proposés dans l'annexe 2.

³ Ces fréquences peuvent être interprétées comme des moyennes, c'est-à-dire la moyenne des valeurs observées.

III. Utilisation des arbres pondérés

A. Exemple d'expérience aléatoire à deux épreuves

On se donne :

- une urne contenant quatre boules indistinguables au toucher dont trois boules bleues, notées b_1 , b_2 et b_3 , portant respectivement les numéros 1, 2 et 3, et une boule rouge unique, notée r .
- un jeu de six cartes identiques portant chacune un chiffre en couleur : une carte avec un chiffre "1" en vert, une carte avec un chiffre "2" en rouge, une carte avec un chiffre "2" en bleu, une carte avec un chiffre "2" en vert, une carte avec un chiffre "3" en rouge, une carte avec un chiffre "3" en bleu.

On considère l'expérience aléatoire suivante : on prélève de façon équiprobable une boule dans l'urne puis une carte du jeu. On note, dans l'ordre, la couleur de la boule extraite et le numéro inscrit sur la carte.

On rappelle qu'un modèle associé à cette expérience aléatoire est défini par la donnée :

- de l'ensemble Ω de toutes les issues possibles de l'expérience ;
- d'une probabilité P déterminée par ses valeurs pour chacun des événements élémentaires définis par ces issues.

La liste de toutes les issues possibles peut être trouvée en utilisant l'arbre des possibles ci-dessous. Les issues possibles pour cette expérience aléatoires sont les couples $(R,1)$; $(R,2)$; $(R,3)$; $(B,1)$; $(B,2)$; $(B,3)$ où B désigne la couleur « Bleu » et R la couleur « Rouge ».

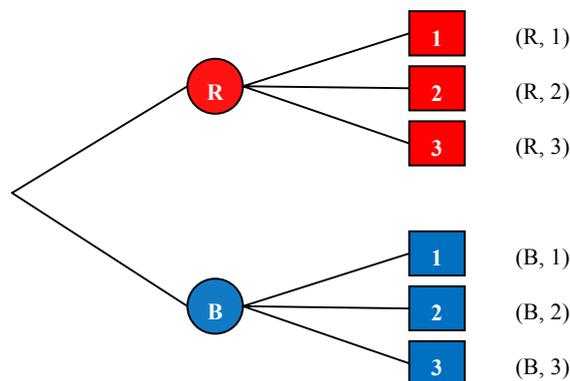


Figure 1

Une fois les issues toutes identifiées, il s'agit de trouver la probabilité des événements élémentaires déterminés par chacune des issues. Il est clair que l'équiprobabilité n'est pas une réponse possible. En effet, on a des raisons de penser que la couleur « Bleu » sera plus probable que la couleur « Rouge » et que le chiffre "2" a plus de chances de sortir que les autres ; en conséquence, l'issue $(B,2)$ a plus de chances de sortir que l'issue $(R,1)$.

Pour affecter une probabilité à chacune des issues, nous allons considérer un autre modèle (qualifié par la suite de **modèle intermédiaire**) qui prend en compte, pour la boule extraite, sa couleur et aussi son numéro éventuel, et pour la carte, le chiffre mentionné mais aussi sa couleur. On peut recenser tous les résultats par l'arbre représentant les issues possibles ci-après.

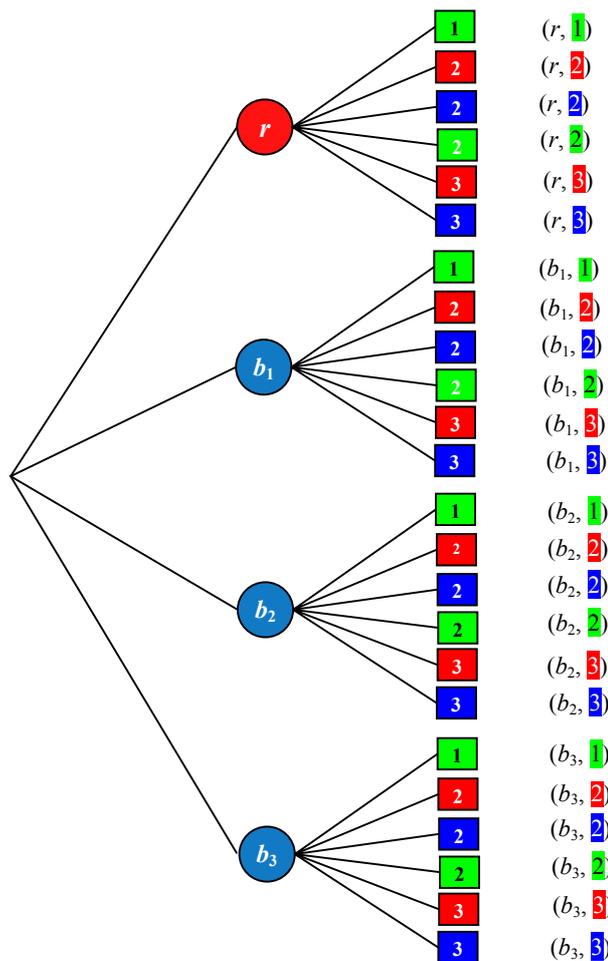


Figure 2

On obtient 4×6 résultats possibles. On peut les noter de la façon suivante : $(r,1)$; $(r,2)$; $(r,2)$; $(r,2)$; $(r,3)$; $(r,3)$; $(b_1,1)$; $(b_1,2)$; $(b_1,2)$; $(b_1,2)$; $(b_1,3)$; $(b_1,3)$; $(b_2,1)$; $(b_2,2)$; $(b_2,2)$; $(b_2,2)$; $(b_2,3)$; $(b_2,3)$; $(b_3,1)$; $(b_3,2)$; $(b_3,2)$; $(b_3,2)$; $(b_3,3)$; $(b_3,3)$.

Chaque branche de l'arbre représente une issue, et compte tenu des conditions du tirage équiprobable de la boule, puis du tirage équiprobable de la carte, il n'y a pas de raison de penser qu'une branche de l'arbre ait plus de chances d'être parcourue qu'une autre. On peut donc considérer que chacune des issues précédentes a la même probabilité, égale à $\frac{1}{24}$, d'être réalisée.

Dans le modèle intermédiaire, par exemple, l'événement « Tirer une boule bleue puis une carte portant le chiffre "2" » se représente mathématiquement par le sous-ensemble des issues $\{(b_1,2)$; $(b_1,2)$; $(b_1,2)$; $(b_2,2)$; $(b_2,2)$; $(b_2,2)$; $(b_3,2)$; $(b_3,2)$; $(b_3,2)\}$. Par suite, la probabilité de cet événement sera égale à $\frac{9}{24}$. Revenant alors au premier modèle où l'événement « Tirer une boule bleue puis une carte portant le chiffre "2" » se représente mathématiquement par l'événement élémentaire $\{(B,2)\}$, on prendra $\frac{9}{24}$ pour la probabilité d'obtenir l'issue $(B,2)$. On peut faire de même pour les cinq autres issues : $(R,1)$; $(R,3)$; $(B,1)$; $(B,2)$; $(B,3)$.

Ce qui conduit au tableau ci-dessous donnant les probabilités affectées à chaque issue du premier modèle :

ω	(R,1)	(R,2)	(R,3)	(B,1)	(B,2)	(B,3)
$P(\{\omega\})$	$\frac{1}{24}$	$\frac{3}{24}$	$\frac{2}{24}$	$\frac{3}{24}$	$\frac{9}{24}$	$\frac{6}{24}$

B. Justification de l'arbre des probabilités

Si on revient à l'arbre (cf. figure 2) utilisé pour trouver toutes les issues possibles du modèle intermédiaire, on constate que cet arbre est très fastidieux à dessiner. Dans la mesure où on ne s'intéresse qu'à la couleur de la boule et au chiffre inscrit sur la carte, on peut alléger sa construction, moyennant quelques conventions de lecture, pour retrouver l'arbre (cf. figure 1) des issues possibles du premier modèle pondéré par les probabilités et justifier la règle des produits de la façon suivante :

1. Étape 1

Partant de l'arbre de la figure 2, dans la mesure où on ne s'intéresse qu'à la couleur de la boule (et non à son numéro éventuel) et qu'au chiffre inscrit sur la carte (et non à sa couleur) on peut convenir de représenter chaque branche de l'arbre de la figure 2 aboutissant à la même couleur de boule, par une seule branche comprenant autant de traits parallèles qu'il y a de boules physiques de cette même couleur. On procède de même en représentant chaque branche de l'arbre de la figure 2 aboutissant à un même chiffre de carte, par une seule branche comprenant autant de traits parallèles qu'il y a de cartes physiques avec ce même chiffre inscrit avec des couleurs différentes. On obtient ainsi l'arbre plus simple de la figure 3 ci-après qui contient cependant autant de branches que celui de la figure 2 tout se rapprochant de l'allure de l'arbre de la figure 1.

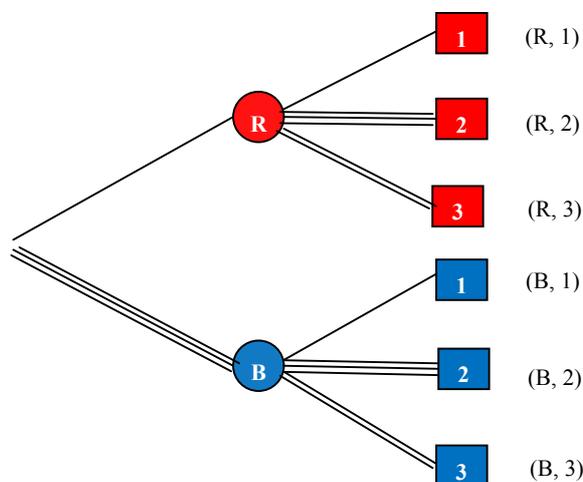


Figure 3

2. Étape 2

On peut alors simplifier davantage l'arbre de la figure 3, en représentant chaque branche par un seul trait pondéré par le nombre de traits composant la branche correspondante dans l'arbre de la figure 3. On obtient alors l'arbre pondéré de la figure 4 qui suit :

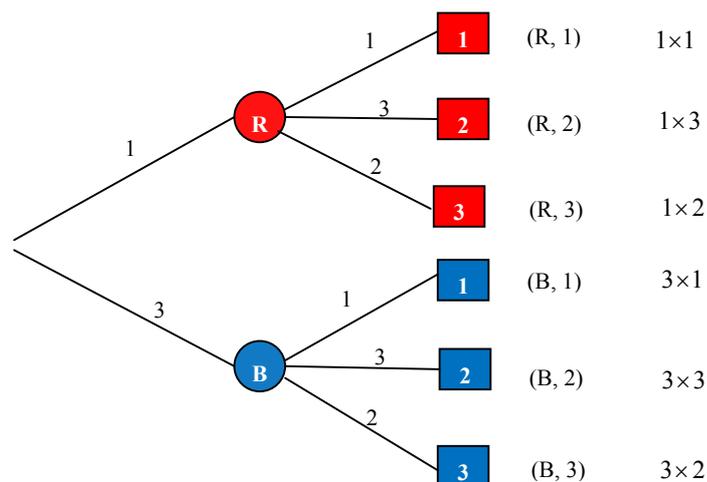


Figure 4

On remarque alors que le produit des nombres rencontrés le long d'un chemin représentant une issue du premier modèle est égal au nombre de chemins de l'arbre de la figure 2 qui réalisent l'événement correspondant dans le modèle intermédiaire. Ainsi, pour l'événement « Tirer une boule bleue puis une carte portant le chiffre 2 », c'est-à-dire (B, 2), le produit 3×3 est égal au nombre de chemins dans le modèle intermédiaire, soit 9.

3. Étape 3

Cette étape consiste à pondérer chaque branche de l'arbre, non plus avec le nombre de traits composant la branche correspondante dans l'arbre de la figure 3, mais avec le quotient de ce nombre par le nombre total de branches d'un même niveau. On obtient ainsi l'arbre pondéré suivant :

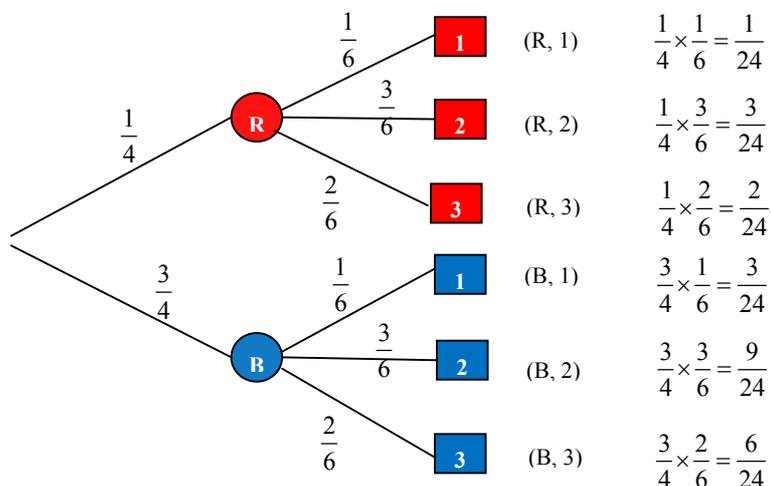


Figure 5

On remarque alors que le produit des quotients affectés aux diverses branches d'un chemin aboutissant à une issue donnée du premier modèle, par exemple pour (B,2) le produit $\frac{3}{4} \times \frac{3}{6}$, est égal à la probabilité, dans cet exemple $\frac{9}{24}$, que cette issue se réalise. Cette remarque est valable pour toutes les branches de l'arbre. Au final, on peut constater que l'arbre de la figure 5 n'est rien d'autre que l'arbre de probabilités associé à l'arbre des possibles de la figure 1.

C. Généralisation et exploitation en Première

La méthode utilisée plus haut peut se généraliser aisément au cas de la succession de n expériences aléatoires. Considérons n expériences aléatoires, $E_1, E_2, E_3, \dots, E_n$, comportant chacune un nombre fini d'issues (non nécessairement le même pour chaque expérience). Considérons l'expérience aléatoire E obtenue par la réalisation successive (dans cet ordre) de ces n expériences aléatoires. On peut alors dessiner l'arbre de probabilités de l'expérience E dont chaque chemin représente une issue⁴ (indiquée en bout de branche) de l'expérience E . La probabilité qu'une issue se réalise est égale au produit des probabilités rencontrées le long du chemin représentant cette issue.

En classe de troisième et de seconde, on s'est intéressé à la succession de deux expériences (éventuellement trois), pas nécessairement identiques. Ces activités ont permis à l'élève de se familiariser avec les arbres de probabilités construits intégralement.

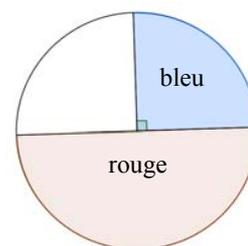
En Première, on s'intéresse surtout à la répétition d'une même expérience aléatoire, un certain nombre n de fois. Contrairement aux classes précédentes, ce nombre n peut alors être éventuellement grand, notamment lorsqu'il s'agit de réinvestir les arbres de probabilités dans le cadre de la loi binomiale.

Un exemple d'activité sur la répétition d'une même expérience aléatoire à trois issues est proposé ci-dessous. Il permettra à l'élève de réinvestir ses connaissances acquises dans les classes précédentes et de le préparer à manipuler des arbres de probabilités, non nécessairement construits intégralement du fait du grand nombre de répétitions de la même expérience aléatoire, lorsqu'il abordera l'étude des propriétés des coefficients binomiaux.

La justification proposée précédemment pour les règles de calcul sur les arbres ne fonctionne que pour des valeurs rationnelles des probabilités et l'on admet que ces règles restent valables pour des valeurs réelles quelconques.

Exemple : répétition d'une expérience à trois issues

On fait tourner la roue de loterie présentée ci-contre : on obtient la couleur « Rouge » avec la probabilité 0,5, la couleur « Bleu » avec la probabilité 0,25 et la couleur « Blanc » avec la probabilité 0,25. Ensuite, on fait tourner une deuxième fois, puis une troisième fois la même roue dans des conditions identiques, et on note les couleurs obtenues.



1. Un joueur est gagnant lorsqu'il obtient dans cet ordre les couleurs « Bleu », « Blanc », « Rouge ». Quelle est la probabilité de gagner à ce jeu ?
2. Quelle est la probabilité que le joueur obtienne dans le désordre les couleurs « Bleu », « Blanc », « Rouge » ?

La réalisation d'un arbre pondéré permet de visualiser les calculs de probabilité demandés.

Pour la première question, il suffit de considérer le seul chemin Bleu-Blanc-Rouge qui a donc pour probabilité $0,25 \times 0,25 \times 0,5$.

Pour la deuxième question, il reste à considérer les 5 chemins qui comportent les trois couleurs dans le désordre. La probabilité de chaque chemin est $0,25 \times 0,25 \times 0,5$ donc la réponse est $5 \times 0,25 \times 0,25 \times 0,5$.

À travers cette activité on rencontre ainsi des chemins de même probabilité, situation qui sera reprise au moment de l'introduction de la loi binomiale.

⁴ Une issue de l'expérience E est une suite $(\omega_1, \omega_2, \dots, \omega_k, \dots, \omega_n)$ où ω_k est une issue de l'expérience E_k .

IV. Loi géométrique tronquée

Les situations de répétition d'une même expérience aléatoire, reproduite dans des conditions identiques constituent un élément fort du programme de Première.

L'introduction de la *loi géométrique tronquée* présente de nombreux avantages :

- travailler des répétitions d'une expérience de Bernoulli ;
- envisager ces répétitions sous l'angle algorithmique ;
- présenter une situation d'arbre pour lequel tous les chemins n'ont pas la même longueur ;
- exploiter hors de l'analyse les propriétés des suites géométriques ;
- exploiter hors du cadre habituel des résultats relatifs à la dérivation ;
- travailler les variables aléatoires.

A. Étude de la loi géométrique tronquée

1. Approche de la loi géométrique tronquée

La probabilité qu'un atome se désintègre par unité de temps est 0,07. On décide d'observer cette désintégration en limitant le temps d'attente à 100 unités de temps, et l'on convient de noter 0 lorsque, après 100 unités de temps, l'atome n'est pas encore désintégré. On distingue ainsi cette situation de la désintégration lors de la 100^e unité de temps.

On peut concevoir un algorithme qui affiche une série de 200 temps d'attente avant la désintégration, ainsi que le temps moyen d'attente calculé à partir de ces 200 valeurs.

On constate que les temps d'attente avant désintégration sont, individuellement, extrêmement imprévisibles. En revanche, la moyenne sur 200 expériences est relativement stable avec des valeurs autour de 13, 14 ou 15. Il est donc sans doute intéressant d'étudier de plus près la loi de la variable aléatoire « temps d'attente ». On montre que l'espérance de cette variable aléatoire vaut $\frac{100}{7} [1 - 8 \times 0,93^{100}]$, soit environ 14,2.

On a relevé ci-dessous, sur tableur, 10 séries de 200 temps d'attente. La moyenne et l'écart-type de chaque série sont affichés. Cela permet de constater combien la dispersion des valeurs individuelles est grande alors que celle des moyennes est petite.

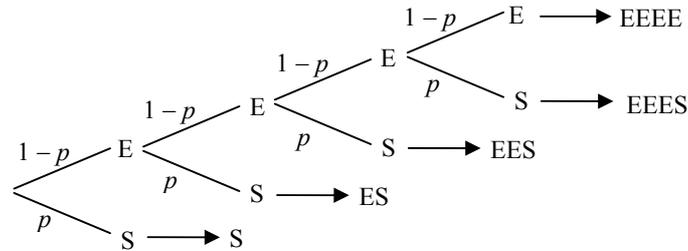
	A	B	C	D	E	F	G	H	I	J	K	L
1	numéro/série	1	2	3	4	5	6	7	8	9	10	
2	1	9	2	1	42	27	14	13	9	10	18	
3	2	10	27	4	9	2	8	2	4	1	53	
4	3	1	6	16	11	6	6	5	18	15	3	
5	4	12	42	23	2	71	13	4	15	41	5	
6	5	22	4	15	3	23	13	5	3	1	3	
7	6	18	26	46	19	35	1	3	15	3	17	
8	7	6	1	27	19	32	10	12	11	4	1	
9	8	2	5	5	5	12	2	16	4	2	12	
10	9	18	46	12	29	61	8	18	19	3	4	
11	10	1	1	4	28	7	3	6	5	2	2	
191	190	12	11	4	5	16	8	2	3	40	28	
192	191	5	11	7	18	1	9	15	3	4	28	
193	192	36	16	6	18	5	9	1	3	9	1	
194	193	18	25	15	20	5	10	7	41	16	40	
195	194	7	58	17	12	4	5	9	4	15	13	
196	195	3	4	7	1	18	14	5	20	19	2	
197	196	44	2	5	8	44	26	3	3	11	15	
198	197	10	16	1	20	29	30	8	19	11	8	
199	198	92	45	2	10	4	18	20	5	37	22	
200	199	9	27	3	56	17	24	5	15	4	29	
201	200	2	5	15	10	34	16	5	4	11	10	
202												
203	emps moyen d'attent	13.415	14.75	13.75	13.375	14.745	14.535	12.685	13.37	13.945	14.905	ecart type des moyennes 0.715521663
204												
205	ecart type de la série	13.321891	13.751636	12.985665	12.110094	14.024977	12.268202	11.234579	14.093371	12.625053	13.493553	

2. Définition de la loi géométrique tronquée

Soit p un réel de l'intervalle $]0,1[$ et n un entier naturel non nul. On considère l'expérience aléatoire qui consiste à répéter dans des conditions identiques une expérience de Bernoulli de paramètre p avec au maximum n répétitions et arrêt du processus au premier succès.

On appelle loi géométrique tronquée de paramètres n et p la loi de la variable aléatoire X définie par :

- $X = 0$ si aucun succès n'a été obtenu ;
- pour $1 \leq k \leq n$, $X = k$ si le premier succès est obtenu à l'étape k .



3. Expression de la loi géométrique tronquée

L'arbre permet de déterminer la loi de la variable aléatoire X décrite ci-dessus, c'est-à-dire la loi géométrique tronquée de paramètres n et p , où n un entier naturel non nul et p un réel de l'intervalle $]0,1[$.

- si aucun succès n'a été obtenu, $X = 0$ et $P(X = 0) = (1 - p)^n$;
- pour $1 \leq k \leq n$, le premier succès est obtenu à l'étape k pour le chemin qui présente dans l'ordre $k - 1$ échecs et un succès, d'où : $P(X = k) = (1 - p)^{k-1} p$.

On vérifie facilement que $\sum_{k=0}^n P(X = k) = 1$ (exploitation des sommes de suites géométriques).

4. Algorithme de simulation

Le processus lié à la loi géométrique tronquée est aisé à mettre en œuvre avec un algorithme. Il suffit de remarquer que l'instruction **ent(NbrAléat + p)** génère un nombre aléatoire entier qui vaut 1 avec la probabilité p , et 0 avec la probabilité $1 - p$.

a) En langage naturel

Entrées : valeur de n
valeur de p

Initialisations : a prend la valeur 0
 k prend la valeur 0

Traitement : Tant que $a = 0$ et $k < n$
 a prend la valeur ent(NbrAléat + p)
 k prend la valeur $k + 1$
Fin de la boucle "tant que"

Sortie : Si $a = 0$
 Alors afficher message " $X =$ "
 valeur de a
 Sinon afficher message " $X =$ "
 valeur de k
Fin de l'instruction conditionnelle

⁵ La convention, $X = 0$ si aucun succès n'a été obtenu, permet d'assurer les mêmes valeurs pour $P(X = k)$ et $P(Y = k)$ pour $k \in [1, n_1]$ si X suit la loi géométrique tronquée de paramètres n_1 et p , Y la loi géométrique tronquée de paramètres n_2 et p , avec $n_1 < n_2$.

b) Avec une calculatrice (modèle TI 84+)

L'instruction "et" se trouve dans le catalogue.

```
PROGRAM:GEOMNP
:Prompt N,P
:0→A
:0→K
:While A=0 et K<
N
:ent(NbrAléat+P)
→A
:K+1→K
:End
:If A=0
:Then
:Disp "X=",A
:Else
:Disp "X=",K
:End
```

c) Avec le logiciel AlgoBox

```
1  VARIABLES
2  n EST_DU_TYPE NOMBRE
3  p EST_DU_TYPE NOMBRE
4  k EST_DU_TYPE NOMBRE
5  a EST_DU_TYPE NOMBRE
6  DEBUT_ALGORITHME
7  LIRE n
8  LIRE p
9  a PREND_LA_VALEUR 0
10 k PREND_LA_VALEUR 0
11 TANT_QUE (a==0 ET k<n) FAIRE
12   DEBUT_TANT_QUE
13   a PREND_LA_VALEUR floor(random()+p)
14   k PREND_LA_VALEUR k+1
15   FIN_TANT_QUE
16 SI (a==0) ALORS
17   DEBUT_SI
18   AFFICHER "X="
19   AFFICHER a
20   FIN_SI
21 SINON
22   DEBUT_SINON
23   AFFICHER "X="
24   AFFICHER k
25   FIN_SINON
26 FIN_ALGORITHME
```

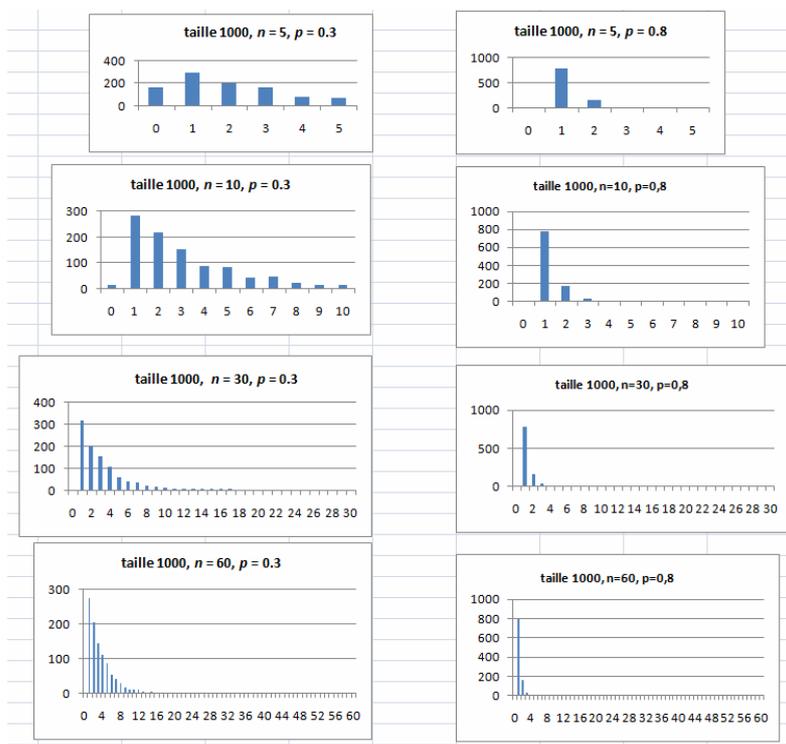
d) Avec le logiciel Scilab

```
1 //loi geometrique tronquee
2 //X=rang du premier succes si succes, 0 sinon
3 k=0;
4 a=0;
5 n=input("donner le nombre de lancers : ");
6 p=input("donner la probabilite d'un succes : ");
7 while (a==0 & k<n)
8   a=floor(rand()+p);
9   k=k+1;
10  afficher(["a="+string(a), "k="+string(k) ])
11  end
12 if a==1 then
13   afficher("X= "+string(k) )
14 else afficher("X=0")
15 end
```

Il est possible de modifier cet algorithme pour obtenir une série de valeurs de la variable X (voir annexe 3). Grâce à ces données, on peut alors visualiser une représentation graphique de la distribution de la loi géométrique tronquée.

5. Représentation graphique

Les diagrammes ci-dessous sont obtenus pour des séries de 1000 valeurs avec d'une part $p = 0,3$ et d'autre part $p = 0,8$. Il est frappant de noter que lorsque n devient grand les histogrammes ont des allures semblables.



L'étude de l'expression de la loi géométrique tronquée va permettre d'expliquer en partie ces observations.

La suite de terme général $p(1 - p)^{k-1}$ est décroissante, donc l'allure générale des diagrammes (hormis le bâton correspondant à $k = 0$) se trouve confirmée.

Pour $p = 0,3$, on obtient en fonction de n :

$$P(X = 0) = (0,7)^n \text{ et pour } 1 \leq k \leq n, P(X = k) = 0,3 \times (0,7)^{k-1}.$$

Il est facile de vérifier avec une calculatrice que :

$$(0,7)^n < 0,005 \text{ pour } n > 14 \text{ et } 0,3 \times (0,7)^{k-1} < 0,005 \text{ pour } k > 12.$$

Ainsi, pour les diagrammes correspondant aux valeurs $n = 30$ et $n = 60$, il n'est pas surprenant de ne voir figurer aucune réalisation de la valeur 0. De même, les bâtons correspondant aux valeurs de k supérieures ou égales à 13 ont une hauteur pratiquement nulle.

Pour $p = 0,8$, on obtient en fonction de n :

$$P(X = 0) = (0,2)^n \text{ et pour } 1 \leq k \leq n, P(X = k) = 0,8 \times (0,2)^{k-1}.$$

$$(0,2)^n < 0,002 \text{ pour } n > 3 \text{ et } 0,8 \times (0,2)^{k-1} < 0,0005 \text{ pour } k > 5.$$

Les diagrammes correspondants sont compatibles avec ces valeurs seuil. En particulier, pour $n = 5$, on n'observe pas de réalisation de la valeur 0.

6. Espérance de la loi géométrique tronquée

Au niveau de la classe de Première, la détermination de l'espérance de la loi géométrique tronquée de paramètres n et p mobilise à la fois les suites géométriques et la dérivation.

Sans être exigible, cette activité peut faire l'objet d'un travail de recherche.

Activité :

Pour tout l'exercice, X désigne une variable aléatoire de loi géométrique tronquée de paramètres n et p . On pose : $q = 1 - p$.

1. Montrer que $E(X) = p \sum_{k=1}^n k(1-p)^{k-1} = p \sum_{k=1}^n kq^{k-1} = p [1 + 2q + 3q^2 + \dots + nq^{n-1}]$.

2. Soit f la fonction définie sur l'intervalle $]0,1[$ par : $f(x) = 1 + x + x^2 + \dots + x^n$.

a. Pour tout réel x de l'intervalle $]0,1[$, écrire $f(x)$ sous la forme d'un quotient.

b. Vérifier que la fonction f est dérivable sur l'intervalle $]0,1[$ et calculer deux expressions différentes de $f'(x)$ pour tout réel x élément de l'intervalle $]0,1[$.

c. En déduire le calcul de la somme $1 + 2x + 3x^2 + \dots + nx^{n-1} = \sum_{k=1}^n k x^{k-1}$ pour tout réel x de l'intervalle $]0,1[$.

3. Prouver l'égalité $E(X) = \frac{1}{p} [1 - (1 + np)(1 - p)^n]$.

4. Utiliser un outil numérique ou graphique pour émettre une conjecture sur la limite de $E(X)$ lorsque n tend vers l'infini.

Remarque :

La limite de $E(X)$ semble être égale à $\frac{1}{p}$ (voir les illustrations en *annexe 3*).

Pour démontrer ce résultat, la principale difficulté est de calculer la limite en $+\infty$ de la suite (u_n) de terme général $u_n = n(1 - p)^n$.

Pour cela, on peut considérer la suite (v_n) définie par $v_n = \frac{u_{n+1}}{u_n}$. Elle converge vers $1 - p$ qui est strictement inférieur à 1. On obtient la limite de la suite (u_n) par comparaison avec une suite géométrique de limite nulle.

B. Exemples d'activités

1. Limitation des naissances

(D'après Claudine ROBERT, *Contes et décomptes de la statistique*, Éd. Vuibert. Voir aussi le document ressource de 2000 rédigé par Claudine ROBERT.)

a) Énoncé

Pour limiter le nombre de filles dans un pays (imaginaire ?), on décide que :

chaque famille aura au maximum 4 enfants ;

chaque famille arrêtera de procréer après la naissance d'un garçon.

On considère que chaque enfant a une chance sur deux d'être un garçon ou une fille et que, pour chaque couple de parents, le sexe d'un enfant est indépendant du sexe des précédents.

Ce choix a-t-il la conséquence attendue, à savoir de diminuer le nombre de filles dans la population ?

Il n'est pas inintéressant de solliciter d'abord une réponse a priori, c'est une façon d'entrer dans le problème et de motiver son étude.

b) Simulation de l'expérience sur un tableur

Les naissances d'une famille se simulent sur une ligne. On passe facilement à la simulation pour 1000 familles en recopiant les formules.

On entre en A4 : `=ENT(ALEA()+0,5)`

et en B4 : `=SI(OU(A4=1;A4="");"";A4+ENT(ALEA()+0,5))`, que l'on recopie jusqu'à D4.

On décompte le nombre d'enfants en E4 : `=NB(A4:D4)` et le nombre de garçons en F4 : `=NB.SI(A4:D4;1)`.

Il reste à recopier les formules de la ligne 4 jusqu'à la ligne 1003. Le calcul de N , G et P est alors immédiat.

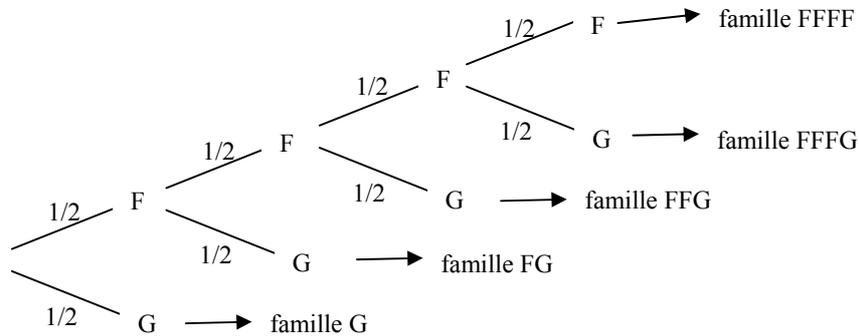
	A	B	C	D	E	F	G	H	I	J	K
1	Chaque colonne représente une naissance. Le nombre indiqué est le nombre de garçons.										
2											
3	Naissances				Enfants	Garçons		Nombre total d'enfants : N =		1896	
4	0	0	0	1	4	1		Nombre total de garçons : G =		939	
5	1				1	1		Proportion de garçons : p =		0,495	
6	1				1	1					
7	0	1			2	1					
8	1				1	1					
9	1				1	1					
10	0	1			2	1					
11	0	1			2	1					
12	1				1	1					
13	1				1	1					
14	0	1			2	1					
15	1				1	1					
16	0	0	1		3	1					
17	1				1	1					
18	1				1	1					
19	0	0	0	0	4	0					

La simulation montre clairement que la proportion de garçons semble bien rester voisine de 0,5. La politique nataliste mise en place n'aurait donc aucun effet sur la modification de cette proportion.

On observerait la même chose lorsque la probabilité de naissance d'un garçon est égale à p .

c) Représentation à l'aide d'un arbre

Le traitement mathématique à l'aide de l'arbre permet de valider les conjectures émises avec le tableau.



$$E(N) = \frac{15}{8}, \quad E(G) = \frac{15}{16}, \quad \text{donc} \quad \frac{E(G)}{E(N)} = \frac{1}{2}.$$

Nombre total d'enfants N	4	4	3	2	1
Nombre total de garçons G	0	1	1	1	1
Probabilité	1/16	1/16	1/8	1/4	1/2

Cette situation peut se prêter à une différenciation pédagogique selon que l'on envisage la valeur $p=0,5$ ou p quelconque, selon que l'on s'en tienne à 4 enfants au plus ou que l'on généralise à n enfants au plus.

d) Généralisation

On considère l'expérience aléatoire qui consiste à répéter dans des conditions identiques une expérience de Bernoulli de paramètre p avec au maximum n répétitions et arrêt du processus au premier succès. On note toujours X la variable aléatoire qui représente le rang du 1^{er} succès et qui vaut 0 si aucun succès n'a été obtenu. La variable aléatoire X suit la loi géométrique tronquée de paramètres n et p .

On considère les variables aléatoires A , nombre de succès et B , nombre d'étapes du processus aléatoire.

La loi de la variable aléatoire A est très simple, elle ne prend que deux valeurs 0 et 1 avec :

$$P(A=0) = P(X=0) = (1-p)^n$$

$$P(A=1) = \sum_{k=1}^{k=n} P(X=k) = 1 - (1-p)^n$$

L'espérance de la variable aléatoire A est donc : $E(A) = 1 - (1-p)^n$.

La variable aléatoire B prend des valeurs entre 1 et n avec :

$$\text{- pour } 1 \leq k \leq n-1, \quad P(B=k) = P(X=k) = (1-p)^{k-1} p ;$$

$$\text{- pour } k=n, \quad P(B=n) = P(X=0) + P(X=n) = (1-p)^n + (1-p)^{n-1} p.$$

L'espérance de la variable aléatoire B est donc :

$$E(B) = \sum_{k=1}^{n-1} k p (1-p)^{k-1} + n [p(1-p)^{n-1} + (1-p)^n],$$

soit $E(B) = \sum_{k=1}^n k p (1-p)^{k-1} + n(1-p)^n$, ou encore $E(B) = E(X) + n(1-p)^n$.

On obtient après simplification $E(B) = \frac{1}{p} [1 - (1-p)^n]$.

Conclusion : si l'on répète un grand nombre de fois ce processus de n étapes au maximum (ceci quelle que soit la valeur de l'entier n), on obtient en moyenne un nombre de succès égal à $1 - (1-p)^n$ pour un nombre moyen d'étapes égal à $\frac{1}{p} [1 - (1-p)^n]$. Ainsi, en moyenne, la proportion de succès est égale à p . Il est remarquable de retrouver cette probabilité de succès, quel que soit le nombre maximal d'étapes du processus.

2. Le paradoxe de Saint-Pétersbourg

Formulé par Nicolas Bernoulli en 1713, ce problème a été approfondi par son cousin Daniel Bernoulli dans l'ouvrage *Les transactions de l'Académie de Saint-Pétersbourg*, ce qui lui a valu son nom.

a) Énoncé

Un joueur joue contre la banque au jeu de « pile ou face », en misant toujours sur « face ». Il adopte la stratégie suivante : il mise un euro au premier coup, et s'il perd, double la mise au coup suivant, tant que « face » ne sort pas. S'il gagne, il récupère sa mise augmentée d'une somme équivalente à cette mise. Le joueur dispose d'une fortune limitée, qui lui permet de perdre au maximum n coups consécutifs et, si « pile » sort n fois de suite, le joueur ne peut plus miser et arrête le jeu. La fortune de la banque, elle, n'est pas limitée.

Une partie consiste pour le joueur à jouer, si sa fortune le lui permet, jusqu'à ce que « face » sorte.

Il s'agit de déterminer la probabilité qu'a le joueur de gagner une partie, son gain algébrique moyen par partie, et d'analyser l'intérêt pour le joueur de jouer à ce jeu.

b) Traitement mathématique

Pour modéliser la situation, on suppose que le joueur lance la pièce n fois : si « face » sort avant le n -ième coup, le joueur ne mise rien les coups suivants. Lorsqu'il joue n fois de suite à « pile ou face », on note :

- A_n l'événement « le joueur obtient n piles » ; $G = \overline{A_n}$ l'événement : « le joueur gagne la partie » ;
- X la variable aléatoire qui comptabilise le rang de la première face, et l'on convient que ce rang est égal à 0 si « face » ne sort pas ;
- Y la variable aléatoire qui donne le gain algébrique du joueur.

On envisage d'abord le cas où le joueur dispose d'une fortune limitée, par exemple à 1000 €.

Le joueur double sa mise tant qu'il perd. Sa fortune lui permet de tenir n coups, où il mise successivement (en euro) 1, 2, 2^2 , ..., 2^{n-1} , tant que $1 + 2 + \dots + 2^{n-1} \leq 1000$. La formule sommatoire sur les suites géométriques simplifie cette inégalité en : $2^n - 1 \leq 1000$. D'où $n = 9$.

On obtient $P(A_9) = \frac{1}{2^9} \approx 0,002$, d'où $P(G) = 1 - P(A_9) = 1 - \frac{1}{2^9} \approx 0,998$.

La variable aléatoire X suit la loi géométrique tronquée de paramètres 9 et $\frac{1}{2}$. Elle prend les valeurs entières de 0 à 9, avec $P(X=0) = P(A_9) = \frac{1}{2^9}$ et, pour k compris entre 1 et 9 :

$$P(X=k) = \frac{1}{2^{k-1}} \left(1 - \frac{1}{2}\right) = \frac{1}{2^k}.$$

On vérifie bien que $\sum_{k=0}^9 P(X=k) = \left(\frac{1}{2} + \frac{1}{2^2} + \dots + \frac{1}{2^9}\right) + \frac{1}{2^9} = 1$.

Déterminons les valeurs de Y .

Si « face » sort pour la première fois au k -ième coup (avec $1 \leq k \leq n$), le joueur a misé au total une somme en euro égale à $1 + 2 + \dots + 2^{k-1}$, il gagne le double de sa dernière mise, soit $2 \times 2^{k-1}$. Son gain algébrique est donc égal à $2^k - (1 + 2 + \dots + 2^{k-1})$, c'est-à-dire 1 €.

Si « face » ne sort pas, le joueur a perdu toutes ses mises, soit (en euro) $1 + 2 + \dots + 2^8 = 2^9 - 1 = 511$.

On en tire la loi de la variable aléatoire Y et son espérance mathématique :

Valeurs de Y	+1	$-(2^9 - 1)$
Probabilités	$1 - \frac{1}{2^9}$	$\frac{1}{2^9}$

$$E(Y) = 1 \times \left(1 - \frac{1}{2^9}\right) - (2^9 - 1) \times \frac{1}{2^9} = 0.$$

c) Simulation de 1000 parties en 9 coups au plus sur un tableur

On code la sortie de « face » par « 1 », celle de « pile » par « 0 ».

On place en A1 la formule `=ENT(2*ALEA())`,

puis en B1 la formule `=SI(OU(A1=1;A1="");"";ENT(2*ALEA()))`, que l'on recopie jusqu'en I1 ;

enfin on place en K1 la formule `=SI(SOMME(A1:I1)=0;"PERDU";"GAGNE")`.

Les formules précédentes sont recopiées jusqu'à la ligne 1000.

Il reste alors en décompter en M1 le nombre de parties perdues, avec la formule :

$$\text{=NB.SI(K1:K1000;"PERDU")}$$

Quoique faible, la probabilité de perdre n'est pas négligeable. Sur la simulation précédente, on s'aperçoit que le joueur perd effectivement 6 parties sur 1000. Il perd donc six fois 511 €, soit 3066 €. Il a gagné 994 parties qui lui rapportent chacune 1 €, soit un gain total de 994 €. Il a donc perdu 2972 euros sur 1000 parties, soit environ 3 euros par partie en moyenne.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	1										GAGNE	Nombre de parties perdues	6	
2	1										GAGNE			
3	0	0	0	1							GAGNE			
4	0	1									GAGNE			
5	1										GAGNE			
6	0	1									GAGNE			
7	1										GAGNE			
8	0	1									GAGNE			
9	1										GAGNE			
10	1										GAGNE			
11	1										GAGNE			
12	0	0	0	0	0	1					GAGNE			
13	0	0	0	0	1						GAGNE			
14	1										GAGNE			
15	0	0	0	0	0	0	0	0	0		PERDU			
16	0	1									GAGNE			
17	1										GAGNE			
18	0	1									GAGNE			

d) Conclusions de l'étude : deux paradoxes

Chaque partie gagnée rapporte 1 € au joueur. Si sa fortune était illimitée (ou simplement très grande), la probabilité de gagner, égale à $1 - \frac{1}{2^n}$, aurait pour limite 1, et permettrait au joueur de gagner chaque partie. Il semble donc que la stratégie du joueur constitue une « martingale » infaillible. Le jeu semble favorable au joueur.

Cependant, puisque $E(Y) = 0$, le jeu est honnête. La stratégie mise en place donne une espérance de gain identique à celle du simple jeu de pile ou face. C'est un premier paradoxe.

Par ailleurs, ce problème montre la limite de la notion d'espérance pour juger si un jeu est favorable. En effet, la simulation précédente a révélé que la perte est importante, et qu'elle se produit plusieurs fois sur 1000 parties. Peu de joueurs s'aventureraient dans un jeu pourtant honnête où l'on risque de perdre gros, même si ce risque est faible, alors que l'on gagne peu. C'est là le deuxième paradoxe.

La notion de risque, liée à celle de la dispersion de la variable aléatoire « gain », est un élément décisif d'appréciation d'un jeu. Le paradoxe de Saint-Pétersbourg est l'un des problèmes ayant donné naissance à la théorie de la décision en économie. Dans cette théorie, on formalise en particulier la notion de *fonction d'utilité*, qui mesure le degré de satisfaction d'un consommateur.

V. Loi binomiale

A. Définitions

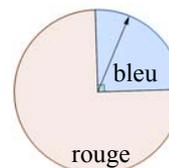
1. Approche de la loi binomiale

Les exemples suivants proposent, en s'appuyant sur les outils déjà disponibles, une découverte de la loi binomiale avant sa formalisation mathématique. Ces activités sont conçues de façon à faciliter une formalisation progressive de ces notions.

Exemple 1 : mise en place du vocabulaire

On fait tourner la roue de loterie présentée ci-contre : on obtient la couleur « rouge » avec la probabilité 0,75 et la couleur « bleu » avec la probabilité 0,25.

Le joueur est gagnant lorsque la flèche s'arrête sur la zone bleue comme sur la figure ci-contre.



On décide de noter S (comme succès) cette éventualité et de noter E (comme échec) l'éventualité contraire c'est-à-dire « la flèche tombe sur la zone rouge ».

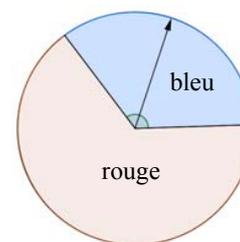
Une expérience à deux issues, succès ou échec, est appelée « épreuve de Bernoulli ». La loi de Bernoulli de paramètre p est la loi de la variable aléatoire qui prend la valeur 1 en cas de succès et 0 en cas d'échec, où p désigne la probabilité du succès.

Consigne aux élèves : on joue trois fois de suite dans des conditions identiques et on désigne par X la variable aléatoire qui donne le nombre de succès obtenus. Réaliser un arbre pondéré représentant cette situation et en déduire la loi de la variable aléatoire X puis son espérance mathématique.

On parlera de « schéma de Bernoulli » lorsqu'on effectue une répétition d'épreuves de Bernoulli identiques et indépendantes.

Exemple 2 : schéma de Bernoulli pour un paramètre p quelconque

On fait maintenant tourner la roue de loterie présentée ci-contre : on obtient la couleur « Bleu » avec une probabilité qui dépend de l'angle indiqué sur la figure et qui est notée p . On obtient donc la couleur « Rouge » avec une probabilité de $1 - p$.



On décide encore de noter S (comme succès) cette éventualité et de noter E (comme échec) l'éventualité contraire c'est-à-dire « la flèche tombe sur la zone rouge ».

1° On répète quatre fois cette épreuve de Bernoulli de paramètre p . Représenter cette répétition par un arbre pondéré à quatre niveaux.

2° On définit alors la variable aléatoire X égale au nombre de succès obtenus à l'issue des quatre répétitions.

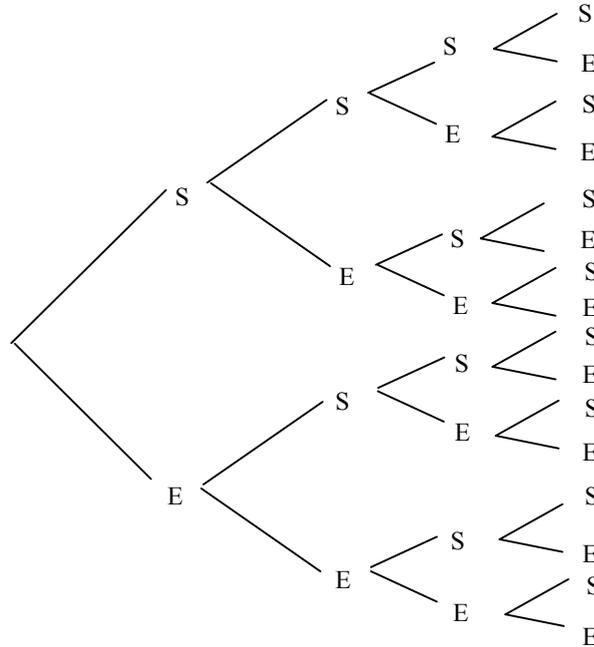
En utilisant l'arbre, déterminer la probabilité des événements suivants :

$$\{X = 0\} \quad \{X = 4\} \quad \{X = 1\} \quad \{X = 2\}.$$

Il faut observer que les probabilités $P(X = 1)$ et $P(X = 2)$ s'obtiennent en comptant les chemins qui conduisent respectivement à 1 et à 2 succès.

On note $\binom{4}{1}$ et on lit « 1 parmi 4 » le nombre de chemins qui conduisent à 1 succès exactement. Ici, $\binom{4}{1} = 4$.

On note $\binom{4}{2}$ et on lit « 2 parmi 4 » le nombre de chemins qui conduisent à 2 succès exactement. Ici, $\binom{4}{2} = 6$.



Exemple 3 : *utiliser une représentation mentale de l'arbre pondéré*

On décide maintenant de répéter cinq fois cette épreuve de Bernoulli et on note toujours X le nombre de succès obtenus à l'issue des cinq répétitions. La réalisation de l'arbre pondéré devient fastidieuse.

1°) Sans réaliser l'arbre, mais en s'inspirant de ce qui a déjà été fait, déterminer la probabilité des événements $\{X = 0\}$ et $\{X = 5\}$.

2°) On s'intéresse dans cette question à la probabilité de l'événement $\{X = 2\}$.

a) Quelle est la probabilité d'un chemin conduisant à exactement deux succès ?

b) On note $\binom{5}{2}$ et on lit « 2 parmi 5 » le nombre de chemins qui conduisent à 2 succès.

Déterminer ce nombre en utilisant l'arbre déjà réalisé pour 4 répétitions.

Il y a deux façons d'obtenir 2 succès selon qu'à la dernière étape on obtient un succès ou un échec.

- Si la dernière étape donne un échec, il faut compter les chemins qui au niveau précédent conduisaient déjà à 2 succès. Avec l'arbre déjà réalisé, on sait que 6 chemins sont dans ce cas.
- Si la dernière étape donne un succès, il faut compter les chemins qui au niveau précédent conduisaient à un seul succès. Avec l'arbre déjà réalisé, on sait que 4 chemins sont dans ce cas.

En conclusion, $6 + 4 = 10$ chemins de l'arbre des 5 répétitions conduisent à 2 succès, soit avec les notations introduites : $\binom{5}{2} = \binom{4}{2} + \binom{4}{1}$.

Enfin la réponse attendue est : $P(X = 2) = 10p^2(1 - p)^3$ ou $P(X = 2) = \binom{5}{2}p^2(1 - p)^3$.

2. Définition de la loi binomiale

On considère une épreuve de Bernoulli de paramètre p . Un schéma de Bernoulli associé à n répétitions de cette épreuve peut être représenté par un arbre pondéré qui comporte n niveaux.

Par définition, la loi binomiale de paramètres n et p , notée $\mathcal{B}(n, p)$, est la loi de la variable aléatoire X qui donne le nombre de succès dans la répétition de n épreuves de Bernoulli de paramètre p .

Quelques cas particuliers

Calcul de $P(X = 0)$ et de $P(X = n)$

L'événement $\{X = 0\}$ est réalisé sur l'unique chemin de l'arbre qui ne comporte que des échecs, c'est-à-dire le dernier chemin de l'arbre qui est constitué de n branches qui ont toutes la probabilité $1 - p$.

D'où le résultat : $P(X = 0) = (1 - p)^n$.

L'événement $\{X = n\}$ est réalisé sur l'unique chemin de l'arbre qui ne comporte que des succès, c'est-à-dire le premier chemin de l'arbre qui est constitué de n branches qui ont toutes la probabilité p .

D'où le résultat : $P(X = n) = p^n$.

Calcul de $P(X = 1)$ et de $P(X = n - 1)$

L'événement $\{X = 1\}$ est réalisé sur les chemins de l'arbre qui comportent exactement un succès et $n - 1$ échecs. La probabilité de chacun de ces chemins est : $p(1 - p)^{n-1}$.

Il reste à déterminer combien de chemins de ce type figurent dans l'arbre pondéré. Cette question est assez simple dans la mesure où il suffit de repérer à quel niveau de l'arbre figure l'unique succès. Il y a donc n possibilités et ainsi n chemins qui réalisent l'événement $\{X = 1\}$.

D'où le résultat : $P(X = 1) = np(1 - p)^{n-1}$.

L'événement $\{X = n - 1\}$ est réalisé sur les chemins de l'arbre qui comportent exactement $n - 1$ succès et 1 échec. La probabilité de chacun de ces chemins est : $p^{n-1}(1 - p)$.

Il reste à déterminer combien de chemins de ce type figurent dans l'arbre pondéré. Comme précédemment, il suffit de repérer à quel niveau de l'arbre figure l'unique échec. Il y a donc encore n possibilités et ainsi n chemins qui réalisent l'événement $\{X = n - 1\}$.

D'où le résultat : $P(X = n - 1) = np^{n-1}(1 - p)$.

3. Coefficients binomiaux

Pour déterminer par exemple, $P(X = 2)$ on procéderait de la même façon : la probabilité de chaque chemin qui réalise exactement deux succès est : $p^2(1 - p)^{n-2}$. Il faut ensuite multiplier cette probabilité par le nombre de chemins qui présentent exactement deux succès. Ce nombre est noté $\binom{n}{2}$ et on lit « 2 parmi n ». Il peut être obtenu avec une calculatrice ou avec un tableur.

Plus généralement :

Si n est un entier naturel et si k est un entier compris entre 0 et n , on note $\binom{n}{k}$ et on lit « k parmi n » le nombre de chemins qui réalisent exactement k succès dans l'arbre à n niveaux, associé à un schéma de Bernoulli. Ces nombres sont appelés coefficients binomiaux.

Ces nombres $\binom{n}{k}$ sont par construction des entiers et l'étude précédente nous fournit quelques valeurs :

- quel que soit n , entier naturel : $\binom{n}{0} = 1$; $\binom{n}{n} = 1$,
- quel que soit n , entier naturel non nul : $\binom{n}{1} = n$; $\binom{n}{n-1} = n$,
- $\binom{4}{1} = 4$; $\binom{4}{2} = 6$; $\binom{5}{2} = 10$.

B. Propriétés

1. Expression de la loi binomiale

La probabilité de chacun des chemins qui réalisent exactement k succès est $p^k(1-p)^{n-k}$. On obtient donc :

Soient un entier naturel n et un réel p de l'intervalle $[0, 1]$

La variable aléatoire X égale au nombre de succès dans la répétition de n épreuves de Bernoulli de paramètre p suit la loi binomiale $\mathcal{B}(n, p)$, avec pour tout entier k compris entre 0 et n :

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Remarque : les coefficients binomiaux $\binom{n}{k}$ interviennent comme coefficients dans la formule générale ci-dessus, mais aussi dans la formule du binôme de Newton qui donne le développement de $(a+b)^n$ pour tous réels a et b .

2. Propriétés des coefficients binomiaux

Symétrie

Le nombre de chemins réalisant $n-k$ succès est aussi le nombre de chemins réalisant k échecs. Par symétrie, on obtient autant de chemins réalisant k succès que de chemins réalisant k échecs.

Si n est un entier naturel et si k est un entier compris entre 0 et n , alors $\binom{n}{k} = \binom{n}{n-k}$.

Formule de Pascal

Il s'agit ici de calculer un coefficient binomial associé à $n + 1$ répétitions à partir des coefficients calculés sur l'arbre réalisé au niveau n .

Le coefficient binomial $\binom{n+1}{k+1}$ donne le nombre de chemins qui réalisent exactement $k + 1$ succès.

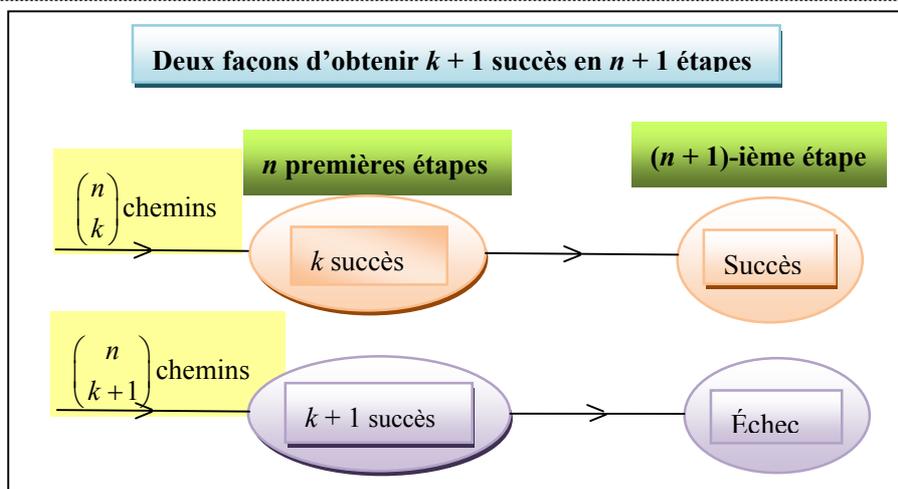
Il y a deux façons d'obtenir $k + 1$ succès suivant qu'à la dernière étape on obtient un succès ou un échec.

- Si la dernière étape donne un échec, il faut compter les chemins qui au niveau précédent conduisaient déjà à $k + 1$ succès. On sait que $\binom{n}{k+1}$ chemins sont dans ce cas.

- Si la dernière étape donne un succès, il faut compter les chemins qui au niveau précédent conduisaient à exactement k succès. On sait que $\binom{n}{k}$ chemins sont dans ce cas.

En conclusion, $\binom{n}{k} + \binom{n}{k+1}$ chemins de l'arbre des $n + 1$ répétitions conduisent à $k + 1$ succès, d'où le résultat :

Si n est un entier naturel et si k est un entier compris entre 0 et $n - 1$, alors $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$.



Somme des coefficients binomiaux

En ajoutant tous les coefficients binomiaux obtenus sur un arbre de n répétitions, on obtient le nombre total de chemins de l'arbre. Or cet arbre comporte n niveaux et à chaque niveau on multiplie le nombre de chemins existants par 2. Le nombre total de chemins est donc 2^n . On obtient la relation :

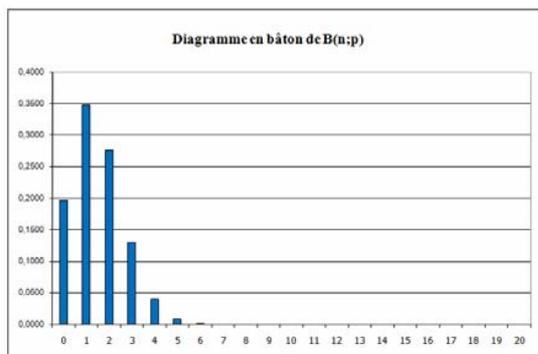
$$\sum_{k=0}^n \binom{n}{k} = \binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{n} = 2^n .$$

On pourra se reporter à l'**annexe 5** pour l'utilisation de quelques outils de calcul avec la loi binomiale.

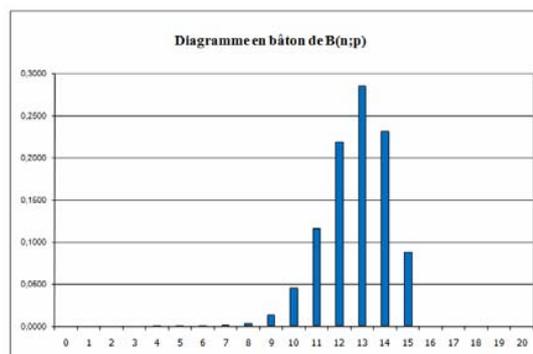
3. Représentation graphique

Dans ce document, on parle de « grande binomiale » si $n \geq 25$ et $0,2 < p < 0,8$ - conditions énoncées dans le programme de la classe de seconde –, et dans le cas contraire, on parle de « petite binomiale ».

Petites binomiales

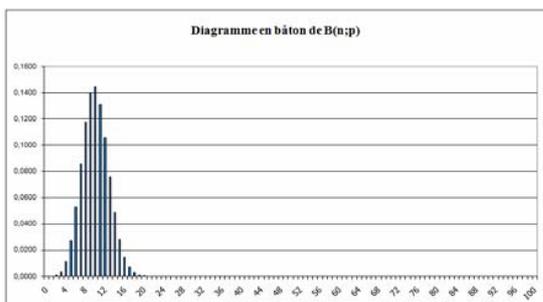


$n = 10 ; p = 0,15$

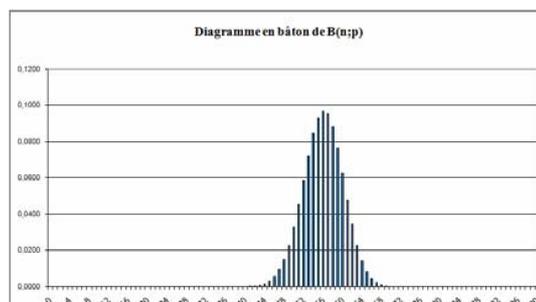


$n = 15 ; p = 0,85$

Grandes binomiales



$n = 40 ; p = 0,25$



$n = 80 ; p = 0,85$

L'observation des différentes représentations graphiques permet de constater les comportements suivants :

- déplacement vers la droite du diagramme à n fixé en fonction de la croissance de p ; constatation analogue si p est fixé et n augmente ;
- allure symétrique « en cloche » des grandes binomiales ; il est facile de démontrer l'exacte symétrie de la représentation lorsque $p = 0,5$;
- dispersion maximale lorsque $p = 0,5$.

4. Espérance et écart-type

Il s'agit ici de proposer une activité conduisant à une conjecture sur l'expression de l'espérance et de l'écart-type d'une loi binomiale.

On utilisera le tableur pour calculer, à l'aide de l'instruction SOMMEPROD, l'espérance et la variance de la loi $\mathcal{B}(n, p)$ pour différentes valeurs de n et p . La variance est obtenue à partir de la relation $V(X) = E(X^2) - E(X)^2$ (cette formule n'est pas un attendu du programme).

Pour la copie d'écran ci-dessous la valeur de p est 0,2 et les valeurs de n vont de 5 à 50 avec un pas de 5 unités.

La feuille de calcul est conçue pour admettre des valeurs de n entre 0 et 100.

Dans les cellules de B3 à K3 on a saisi : $=\$B1$.

Dans la cellule B7 on a saisi : $=SI(\$A7<=B\$4;LOI.BINOMIALE(\$A7;B\$4;B\$3;0);" ")$ pour demander l'affichage de la probabilité de $\{X = k\}$, uniquement lorsque $k \leq n$ et une cellule vide dans l'autre alternative.

Dans la cellule B110 on a saisi : $=SOMMEPROD(\$A7:\$A107;B7:B107)$ pour obtenir l'espérance de la variable aléatoire X de loi binomiale de paramètres n et p . L'adressage absolu sur la première colonne autorise la recopie vers la droite de cette formule.

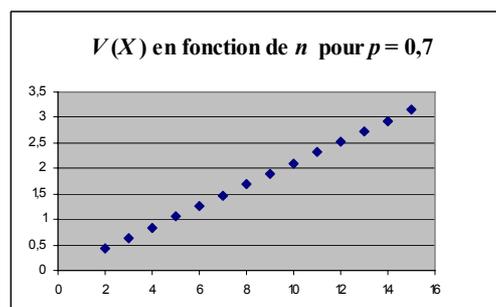
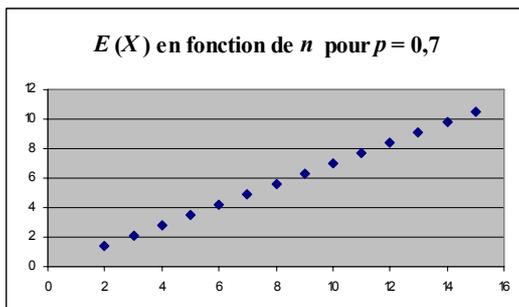
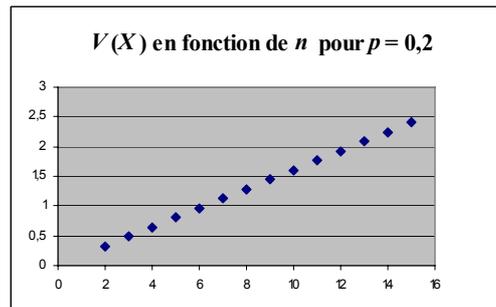
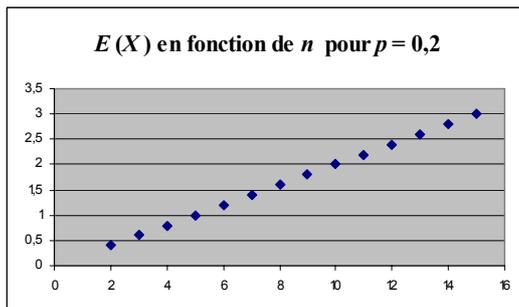
Dans la cellule B112 on a saisi : $=SOMMEPROD((\$A7:\$A107)^2;B7:B107)$ pour obtenir l'espérance de la variable X^2 .

La variance s'obtient alors en B116 avec la formule $=B112-B113$.

On peut aussi obtenir la variance de X comme espérance de $(X - E(X))^2$ en saisissant dans la cellule B116 la formule : $=SOMMEPROD((\$A7:\$A107-B110)^2;B7:B107)$.

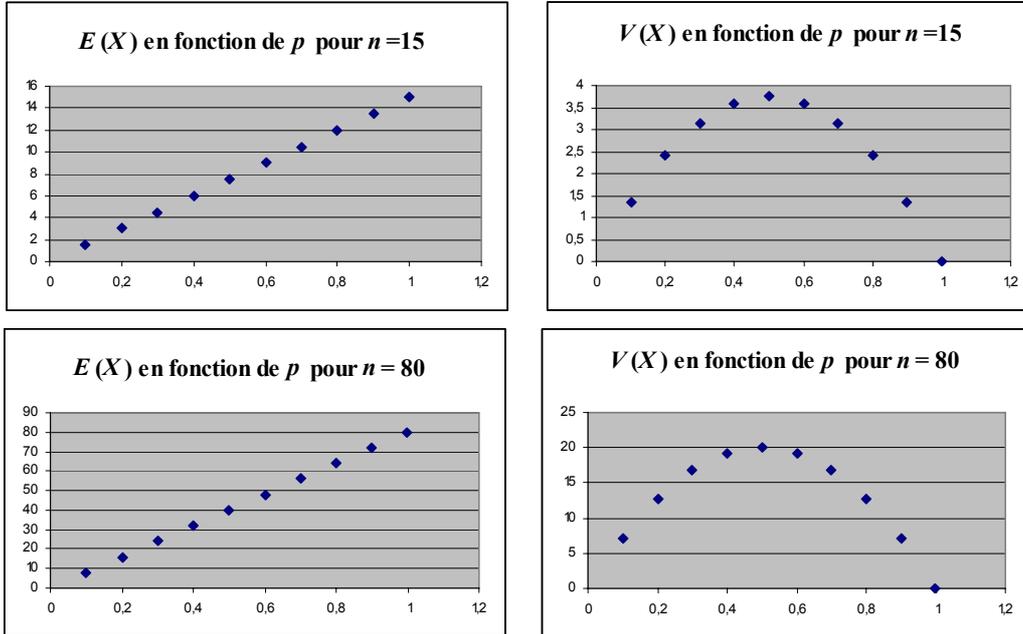
	A	B	C	D	E	F	G	H	I	J	K
1	p	0,2									
2											
3	p	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
4	n	5	10	15	20	25	30	35	40	45	50
5											
6	k	$P(X=k)$									
7	0	0,32768	0,107374	0,035184	0,011529	0,003778	0,001238	0,000406	0,000133	4,36E-05	1,43E-05
8	1	0,4096	0,268435	0,131941	0,057646	0,023612	0,009285	0,003549	0,001329	0,00049	0,000178
9	2	0,2048	0,30199	0,230897	0,136909	0,070835	0,033656	0,015085	0,00648	0,002695	0,001093
10	3	0,0512	0,201327	0,250139	0,205364	0,135768	0,078532	0,041484	0,02052	0,009657	0,004371
11	4	0,0064	0,08808	0,187604	0,218199	0,186681	0,132522	0,082968	0,047452	0,02535	0,01284
12	5	0,00032	0,026424	0,103182	0,17456	0,196015	0,172279	0,1286	0,085414	0,051968	0,029531
13	6		0,005505	0,042993	0,1091	0,163346	0,179457	0,16075	0,124563	0,086613	0,055371
107	100										
108											
109	n	5	10	15	20	25	30	35	40	45	50
110	$E(X)$	1	2	3	4	5	6	7	8	9	10
111											
112	$E(X^2)$	1,8	5,6	11,4	19,2	29	40,8	54,6	70,4	88,2	108
113	$E(X)^2$	1	4	9	16	25	36	49	64	81	100
114											
115	n	5	10	15	20	25	30	35	40	45	50
116	$V(X)$	0,8	1,6	2,4	3,2	4	4,8	5,6	6,4	7,2	8
117											

Observations pour p fixé :



Lorsque la valeur de p est fixée, on observe que l'espérance de la loi binomiale, aussi bien que sa variance, semblent être des fonctions linéaires de n .

Observations pour n fixé :



Lorsque la valeur de n est fixée, on observe que l'espérance de la loi binomiale semble aussi être une fonction linéaire de p . De plus, on peut noter que la valeur obtenue avec le cas particulier $p=1$ correspond à la valeur de n qui a été fixée. D'où la conjecture $E(X) = np$ que le professeur validera dans le cours.

En revanche, la variance se comporte comme une fonction du second degré en p . On peut noter aussi que la variance semble bien être maximale pour $p=0,5$ comme l'observation des représentations graphiques le laissait prévoir.

Un réinvestissement des notions d'analyse permet, par exemple, de déterminer les fonctions polynômes du second degré qui sont maximales en $0,5$ et qui s'annulent en 0 et 1 . La linéarité selon la variable n incite à chercher un coefficient « multiple » de n et quelques essais permettent d'aboutir à l'expression $V(X) = np(1-p)$ que le professeur validera dans le cours.

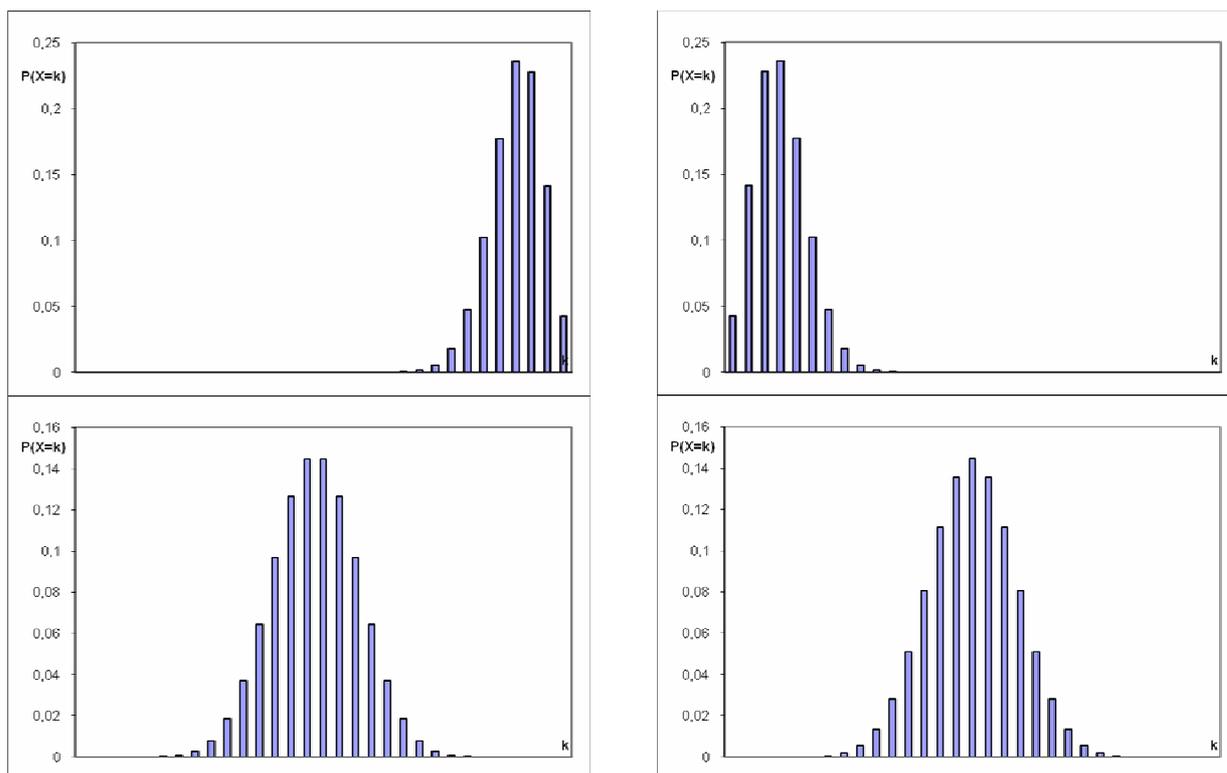
C. Exemples d'activités

1. Avec la loi de probabilité

a) Appariement

On a représenté ci-dessous la distribution de probabilité de quatre variables aléatoires suivant les lois binomiales $\mathcal{B}(30; 0,1)$, $\mathcal{B}(30; 0,5)$, $\mathcal{B}(30; 0,9)$, $\mathcal{B}(29; 0,5)$.

Associer chaque loi à son graphique.



b) Le quorum

Une association comprenant 30 adhérents organise chaque année une assemblée générale. Les statistiques montrent que chaque adhérent assiste à l'assemblée avec la probabilité 80 %. Les décisions prises par l'assemblée n'ont de valeur légale que lorsque plus de la moitié des adhérents assiste à l'assemblée.

Quelle est la probabilité que, lors de la prochaine assemblée, le quorum soit atteint ?

c) Paradoxe ?

Paul affirme : « Avec un dé régulier, on a autant de chance d'obtenir au moins un six en 4 lancers que d'obtenir au moins deux six avec 8 lancers ».

Sara objecte : « Pas du tout. Dans le premier cas, la probabilité est supérieure à 0,5, dans le deuxième cas, elle est inférieure à 0,5. ».

Qui a raison ?

d) Le tir à l'arc

À chaque tir, un archer atteint sa cible avec une probabilité égale à 0,7.

Combien de tirs doit-il effectuer pour que, avec une probabilité supérieure ou égale 0,99, il atteigne la cible au moins deux fois ? Au moins trois fois ?

e) Lancers de pièce

On lance une pièce équilibrée n fois. On s'intéresse à la probabilité d'obtenir « face » dans 60 % des cas ou plus.

Envisager les cas $n = 10$, puis $n = 100$, puis $n = 1000$.

Donner d'abord, sans calcul, une estimation spontanée du résultat, puis solliciter la calculatrice ($n = 10$) ou un algorithme de calcul ($n = 100$ et $n = 1000$).

2. Avec l'espérance mathématique

a) Contrôle de production

Une entreprise fabrique chaque jour 10 000 composants électroniques. Chaque composant présente un défaut avec la probabilité 0,002. Si le composant est repéré comme étant défectueux, il est détruit par l'entreprise, et chaque composant détruit fait perdre 1 € à l'entreprise.

- Les composants sont contrôlés un à un, et chaque contrôle coûte 0,1 €. Quel est le coût moyen journalier pour l'entreprise (contrôles et destruction des composants défectueux) ?
- Les composants sont regroupés par lots de 10, et on effectue un unique contrôle automatique de chaque lot, qui coûte lui aussi 0,1 €. À l'issue de ce contrôle, le lot est accepté si tous les composants sont sains, et globalement détruit si l'un au moins des 10 composants présente un défaut. Quel est le coût moyen journalier pour l'entreprise de ce nouveau dispositif (contrôles et destruction des composants défectueux) ?

b) Le QCM

Un QCM comporte 20 questions. Pour chaque question, quatre réponses sont proposées dont une seule est juste. Chaque réponse juste rapporte un point et il n'y a pas de pénalité pour une réponse fausse. Un candidat répond au hasard à chaque question.

Quel nombre total de points peut-il espérer ?

Quelle pénalité doit-on attribuer à une réponse fausse pour que le total espéré, en répondant entièrement au hasard, soit égal à 2 sur 20 ?

c) Correction de fautes

Un texte contient n erreurs. Lors d'une relecture, on considère que chaque erreur a 80 % de chances d'être corrigée.

Peut-on prévoir, en moyenne, le nombre d'erreurs restantes après une relecture, ..., après k relectures, k étant un entier supérieur à 1 ?

VI. Échantillonnage et prise de décision

La prise de décision apparaît pour la première fois dans le programme de Seconde. La démarche s'appuie sur la notion d'intervalle de fluctuation dont une définition est donnée et une expression proposée sous réserve de satisfaire aux conditions de validité, $n > 25$ et $0,2 < p < 0,8$ (n est la taille de l'échantillon prélevé et p est la proportion dans la population du caractère étudié). Avec la notion de variable aléatoire et la découverte de la loi binomiale, le programme de Première fournit les premiers outils qui permettent, en prenant appui sur la réflexion initiée en Seconde autour de la prise de décision, de construire l'intervalle de fluctuation déterminé à l'aide de la loi binomiale, et une démarche de prise de décision, valable en toute généralité pour une proportion p et une taille n d'échantillon.

La **partie A** synthétise les définitions, la **partie B** présente la problématique de la *prise de décision*.

La **partie C** développe une approche possible en classe.

A. Intervalle de fluctuation avec la loi binomiale

L'intervalle de fluctuation d'une fréquence au seuil de 95% a été défini dans le programme de Seconde de la façon suivante :

« L'intervalle de fluctuation au seuil de 95%, relatif aux échantillons de taille n , est l'intervalle centré autour de p , proportion du caractère dans la population, où se situe, avec une probabilité égale à 0,95, la fréquence observée dans un échantillon de taille n . »

La loi binomiale permet de calculer très exactement les probabilités des différentes fréquences observables dans un échantillon de taille n , à savoir les valeurs $\frac{k}{n}$, avec $0 \leq k \leq n$, probabilités qui peuvent être représentées à l'aide d'un diagramme en bâtons. On peut également calculer, à l'aide d'un tableur, les probabilités (cumulées) des événements suivants : « La fréquence observée dans l'échantillon prélevé de taille n est comprise entre 0 et $\frac{k}{n}$ », événement qu'on peut aussi écrire $\left\{0 \leq F \leq \frac{k}{n}\right\}$, si F désigne la variable aléatoire qui à tout échantillon de taille n associe la fréquence observée dans l'échantillon prélevé.

En faisant varier les paramètres n et p , on observe que le diagramme n'est pas toujours symétrique, et non exactement centré sur p . Par ailleurs, le caractère discret de la loi binomiale fait qu'il n'est pas possible de déterminer précisément un intervalle où la fréquence observée se situe avec une probabilité égale à 0,95. La définition donnée en seconde pour l'intervalle de fluctuation suppose en effet, de manière implicite, que la fréquence suit une loi continue.

Pour ces différentes raisons, on est amené à construire un intervalle qui approxime l'intervalle de fluctuation défini plus haut en adoptant la définition suivante :

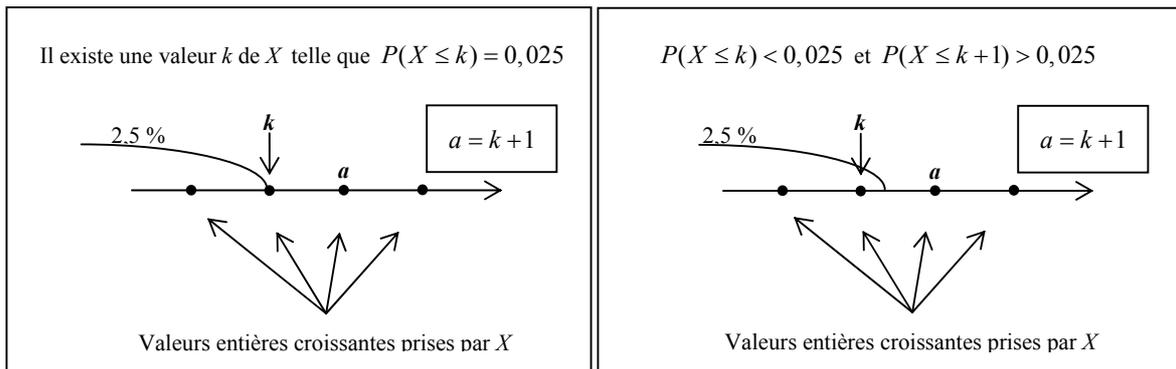
L'intervalle de fluctuation au seuil de 95 % d'une fréquence F , correspondant à la réalisation, sur un échantillon aléatoire de taille n , de la variable aléatoire X égale à nF et de loi binomiale de paramètres n et p , est l'intervalle $\left[\frac{a}{n}, \frac{b}{n}\right]$ défini par le système de conditions suivant :

- a est le plus grand entier tel que $P(X < a) \leq 0,025$,
- b est le plus petit entier tel que $P(X > b) \leq 0,025$.

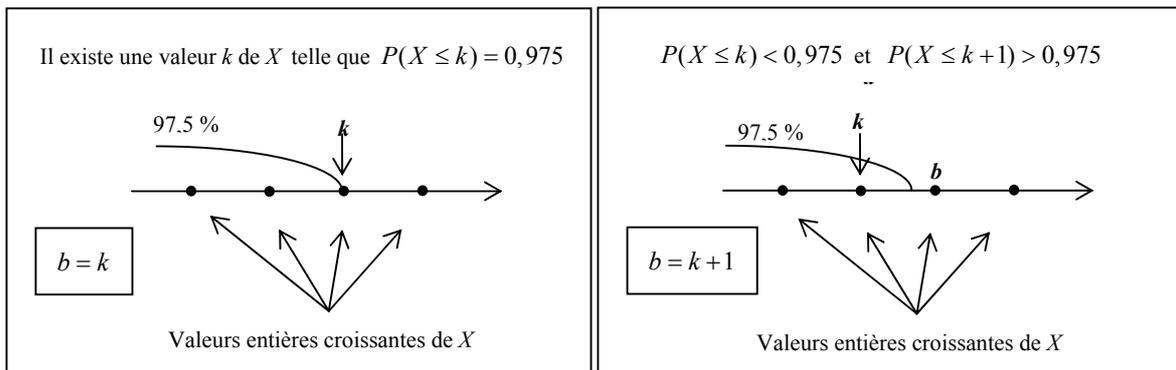
ou encore par le système de conditions équivalent :

- a est le plus petit entier tel que $P(X \leq a) > 0,025$,
- b est le plus petit entier tel que $P(X \leq b) \geq 0,975$.

Détermination de a



Détermination de b



Remarques :

1. Ce jeu sur les inégalités strictes est dû au caractère discret de la variable aléatoire X considérée (on pourra s'en convaincre dans la partie C).
2. Les entiers a et b dépendent de la taille n de l'échantillon.

La connaissance de la loi binomiale de la variable aléatoire X rend maintenant possible le calcul de la probabilité $P\left(\frac{a}{n} \leq F \leq \frac{b}{n}\right) = P(a \leq X \leq b)$.

On remarque que l'intervalle $\left[\frac{a}{n}, \frac{b}{n}\right]$ est quasiment centré sur p dès que n est « assez grand » et que

l'intervalle $\left[\frac{a}{n}, \frac{b}{n}\right]$ est « quasiment » le même que l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ donné dans le programme de seconde pour les « grandes binomiales » ($n > 25$ et $0,2 < p < 0,8$ où n est la taille de l'échantillon prélevé et p est la proportion dans la population du caractère étudié, conditions énoncées dans le programme de seconde).

On trouvera des exemples dans le sous-paragraphe C ci-après.

L'intérêt de l'intervalle $\left[\frac{a}{n}, \frac{b}{n}\right]$ (qu'il conviendrait de noter $\left[\frac{a_n}{n}, \frac{b_n}{n}\right]$ pour être précis), calculé à partir de la loi binomiale, est de fournir un intervalle convenable **pour toutes les valeurs de n et de p** , alors que l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ **n'est pas adapté** pour les « petites binomiales ».

B. Aspect général de la prise de décision avec la loi binomiale

On considère une population dans laquelle on suppose que la proportion d'un certain caractère est p . Pour juger de cette hypothèse, on y prélève, au hasard et avec remise, un échantillon de taille n sur lequel on observe une fréquence f du caractère.

On rejette l'hypothèse selon laquelle la proportion dans la population est p lorsque la fréquence f observée est trop éloignée de p , dans un sens ou dans l'autre. On choisit de fixer le **seuil à 95 %** de sorte que la probabilité de rejeter l'hypothèse, alors qu'elle est vraie, soit inférieure à 5 %.

Hypothèse :
la proportion est p .

Échantillon taille n
Observation :
fréquence f

Lorsque la proportion dans la population vaut p , la variable aléatoire X correspondant au nombre de fois où le caractère est observé dans un échantillon aléatoire de taille n , suit la loi binomiale de paramètres n et p .

La **règle de décision** adoptée est la suivante :

- si la fréquence observée f appartient à l'intervalle de fluctuation au seuil de 95 % $\left[\frac{a}{n}, \frac{b}{n}\right]$, on considère que l'hypothèse selon laquelle la proportion est p dans la population n'est pas remise en question et on l'accepte ;
- sinon, on rejette l'hypothèse selon laquelle cette proportion vaut p .

Cette prise de décision repose sur le raisonnement suivant : si la proportion vaut p on a, en gros, au moins 95 % de chances que le prélèvement d'un échantillon de taille n conduise à une fréquence f du caractère dans cet échantillon située dans $\left[\frac{a}{n}, \frac{b}{n}\right]$. On sait bien que dans ce cas, compte tenu du hasard, la fréquence réellement observée f n'est pas nécessairement égale à p , mais qu'elle **fluctue** dans un voisinage de p , appelé justement **intervalle de fluctuation**. Un intervalle de fluctuation est donc un intervalle où l'on « s'attend » à trouver la fréquence observée f , si l'hypothèse que la proportion est p est la bonne.

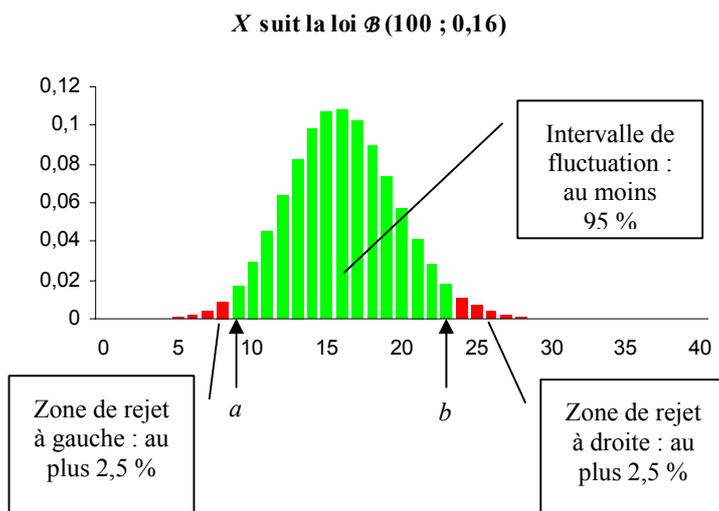
En conséquence, si la proportion vaut p , il y a très peu de chances (environ au plus 5% des échantillons) que cette fréquence observée f soit hors de l'intervalle de fluctuation $\left[\frac{a}{n}, \frac{b}{n}\right]$. Donc si elle est à l'extérieur de l'intervalle $\left[\frac{a}{n}, \frac{b}{n}\right]$, il est cohérent de penser que ce n'est plus le seul fait du hasard cette fois-ci, mais que c'est bien plutôt le signe que l'hypothèse que la proportion est p n'est pas la bonne.

C. Détermination de l'intervalle de fluctuation à l'aide d'un algorithme

Considérons l'exemple suivant. Un médecin de santé publique veut savoir si, dans sa région, le pourcentage d'habitants atteints d'hypertension artérielle est égal à la valeur de 16 % récemment publiée pour des populations semblables. En notant p la proportion d'hypertendus dans la population de sa région, le médecin formule l'hypothèse $p = 0,16$. Pour vérifier cette hypothèse, le médecin constituera un échantillon de $n = 100$ habitants de la région ; il déterminera la fréquence f d'hypertendus (l'échantillon est prélevé au hasard et la population est suffisamment importante pour considérer qu'il s'agit de tirages avec remise).

Lorsque la proportion dans la population vaut $p = 0,16$, la variable aléatoire X correspondant au nombre d'hypertendus observé dans un échantillon aléatoire de taille $n = 100$, suit la loi binomiale de paramètres $n = 100$ et $p = 0,16$.

On cherche à partager l'intervalle $[0;100]$, où X prend ses valeurs, en trois intervalles $[0, a - 1]$, $[a, b]$ et $[b + 1, 100]$ de sorte que la variable aléatoire X prenne ses valeurs dans chacun des intervalles extrêmes avec une probabilité proche de 0,025, sans dépasser cette valeur. On recherche donc le plus grand entier a tel que $P(X < a) \leq 0,025$ et le plus petit entier b tel que $P(X > b) \leq 0,025$.



D'un point de vue **algorithmique**, il est plus efficace de travailler avec les probabilités cumulées croissantes, que la calculatrice ou le tableur fournissent facilement. En tabulant les probabilités cumulées $P(X \leq k)$, pour k allant de 0 à 100, il suffit de déterminer le plus petit entier a tel que $P(X \leq a) > 0,025$ et le plus petit entier b tel que $P(X \leq b) \geq 0,975$.

Le calcul, à l'aide de la loi binomiale, de l'intervalle de fluctuation au seuil de 95 %, $\left[\frac{a}{n}, \frac{b}{n} \right]$, de la fréquence des échantillons aléatoires de taille n , correspondant à la zone d'acceptation d'une hypothèse sur une proportion, peut ainsi faire l'objet d'une recherche d'algorithme.

On peut, par exemple, procéder sur un tableur comme le montre l'image d'écran ci-dessous.

La cellule B3 contient la valeur de n , taille de l'échantillon. La cellule D3 contient la valeur de p , proportion supposée dans la population.

On a entré en B6 la formule `=SI(A6<=B$3;LOI.BINOMIALE(A6;B$3;D$3;VRAI);"")`

pour tabuler les probabilités $P(X \leq k)$ lorsque X suit la loi binomiale de paramètres n et p .

On a entré en C6 la formule `=SI(B6>0,025;A6/B$3;"")`

pour afficher les valeurs de k telles que $P(X \leq k)$ dépasse strictement 0,025.

On a entré en D6 la formule `=SI(B6>=0,975;A6/B$3;"")`

pour afficher les valeurs de k telles que $P(X \leq k)$ égale ou dépasse 0,975.

Ces trois formules ont été ici recopiées vers le bas jusqu'à la ligne 1 006 (l'algorithme fonctionne pour une valeur maximale de n égale à 1 000, mais on peut, en cas de besoin, recopier plus bas).

L'intervalle de fluctuation au seuil de 95 % est affiché en cellules F8 et G8 contenant les formules `=MIN(C6:C1006)` et `=MIN(D6:D1006)`.

A	B	C	D	E	F	G	H
Intervalle de fluctuation à 95 % d'une fréquence donné par la loi binomiale							
n	taille de l'échantillon	p	proportion supposée dans la population				
	n = 100	p = 0,16					
k	P(X ≤ k)	recherche a	recherche b	Intervalle de fluctuation à 95 %			
0	2,67873E-08			(selon la loi binomiale)			
1	5,37021E-07						
2	5,3478E-06				0,09	0,23	
3	3,52815E-05						
4	0,000173547						
5	0,000679203						
6	0,002204197						
7	0,006104862						
8	0,014742048						
9	0,031559427	0,09					
10	0,060709551	0,1					
11	0,106138316	0,11					
12	0,170315459	0,12					
13	0,25306401	0,13					
14	0,351011275	0,14					
15	0,457975907	0,15					
16	0,566213927	0,16					
17	0,688085005	0,17					
18	0,757559073	0,18					
19	0,831111691	0,19					
20	0,887852282	0,2					
21	0,929024593	0,21					
22	0,95718574	0,22					
23	0,975376792	0,23	0,23				
24	0,986493546	0,24	0,24				

Dans le cas de l'exemple choisi, on a $n = 100$ et $p = 0,16$. L'algorithme fournit $\frac{a}{n} = 0,09$ et $\frac{b}{n} = 0,23$.

La règle de décision, pour le médecin, sera la suivante :

- si la fréquence observée f appartient à l'intervalle de fluctuation $[0,09 ; 0,23]$, on considère que l'hypothèse selon laquelle la proportion d'hypertendus dans la population est $p = 0,16$ n'est pas remise en question et on l'accepte ;
- sinon, on rejette l'hypothèse selon laquelle cette proportion vaut $p = 0,16$.

D. Exemples d'activités

1. Exemple 1 : politique dans un pays lointain

Monsieur Z, chef du gouvernement d'un pays lointain, affirme que 52 % des électeurs lui font confiance. On interroge 100 électeurs au hasard (la population est suffisamment grande pour considérer qu'il s'agit de tirages avec remise) et on souhaite savoir à partir de quelles fréquences, au seuil de 95 %, on peut mettre en doute le pourcentage annoncé par Monsieur Z, dans un sens, ou dans l'autre.

1. On fait l'hypothèse que Monsieur Z dit vrai et que la proportion des électeurs qui lui font confiance dans la population est 0,52. Montrer que la variable aléatoire X , correspondant au nombre d'électeurs lui faisant confiance dans un échantillon de 100 électeurs, suit la loi binomiale de paramètres $n = 100$ et $p = 0,52$.

2. On donne ci-contre un extrait de la table des probabilités cumulées $P(X \leq k)$ où X suit la loi binomiale de paramètres $n = 100$ et $p = 0,52$.

a. Déterminer a et b tels que :

- a est le plus petit entier tel que $P(X \leq a) > 0,025$;
- b est le plus petit entier tel que $P(X \leq b) \geq 0,975$.

b. Comparer l'intervalle de fluctuation au seuil de 95 %,

$\left[\frac{a}{n}, \frac{b}{n} \right]$, ainsi obtenu grâce à la loi binomiale, avec l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}} \right]$.

k	$P(X \leq k) \approx$
40	0,0106
41	0,0177
42	0,0286
43	0,0444
...	...
61	0,9719
62	0,9827
63	0,9897
64	0,9941

3. Énoncer la règle décision permettant de rejeter ou non l'hypothèse que la proportion des électeurs qui font confiance à Monsieur Z dans la population est 0,52, selon la valeur de la fréquence f des électeurs favorables à Monsieur Z obtenue sur l'échantillon.

4. Sur les 100 électeurs interrogés au hasard, 43 déclarent avoir confiance en Monsieur Z. Peut-on considérer, au seuil de 95 %, l'affirmation de Monsieur Z comme exacte ?

Éléments de réponse

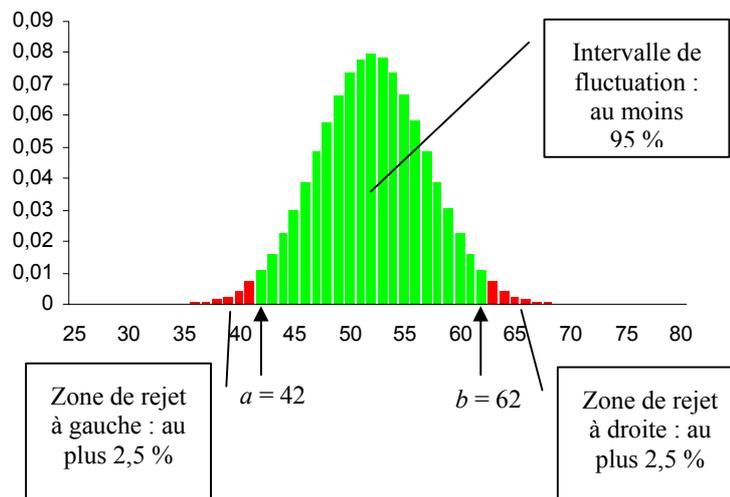
2. a. On lit $a = 42$ et $b = 62$.

b. Les intervalles sont identiques.

3. Si f appartient à l'intervalle $[0,42 ; 0,62]$, l'hypothèse que la proportion des électeurs qui font confiance à Monsieur Z dans la population est 0,52 est acceptable, sinon, l'hypothèse est rejetée, au seuil de 95 %.

4. On considère que l'affirmation de Monsieur Z est exacte.

X suit la loi $\mathcal{B}(100 ; 0,52)$



Remarque : la recherche de l'intervalle de fluctuation peut-être illustrée par le diagramme en bâtons de la loi binomiale de paramètres $n = 100$ et $p = 0,52$.

2. Exemple 2 : discrimination

(d'après le document ressource mathématiques pour les baccalauréats professionnels)

En Novembre 1976 dans un comté du sud du Texas, Rodrigo Partida est condamné à huit ans de prison. Il attaque ce jugement au motif que la désignation des jurés de ce comté est, selon lui, discriminante à l'égard des Américains d'origine mexicaine. Alors que 80 % de la population du comté est d'origine mexicaine, sur les 870 personnes convoquées pour être jurés lors des années précédentes, il n'y a eu que 339 personnes d'origine mexicaine.

Devant la Cour Suprême, un expert statisticien produit des arguments pour convaincre du bien fondé de la requête de l'accusé. En vous situant dans le rôle de cet expert, pouvez-vous décider si les Américains d'origine mexicaine sont sous-représentés dans les jurys de ce comté ?

Éléments de réponse

On suppose que les 870 jurés sont tirés au sort dans la population du comté (la population étant très importante, on peut considérer qu'il s'agit de tirages avec remise). Sous cette hypothèse, la variable aléatoire X correspondant au nombre de jurés d'origine mexicaine suit la loi binomiale de paramètres $n = 870$ et $p = 0,8$.

On peut alors rechercher, en utilisant la loi binomiale, l'intervalle de fluctuation au seuil de 95 % correspondant.

Une tabulation de la loi binomiale de paramètres $n = 870$ et $p = 0,8$ fournit les résultats suivants :

k	$P(X \leq k)$	fréquence k/n
672	0,0245	0,772
673	0,0296	0,774
...
718	0,9733	0,825
719	0,9783	0,826

L'intervalle de fluctuation au seuil de 95 % de la fréquence des jurés d'origine mexicaine est : $[0,774; 0,826]$.

La fréquence observée est $f = \frac{339}{870} \approx 0,39$. Cette valeur ne se situe pas dans l'intervalle de fluctuation.

La différence est significative au seuil de 95 % et l'hypothèse $p = 0,8$, avec un tirage aléatoire des jurés, est rejetée.

De fait, l'accusé a obtenu gain de cause et a été rejugé par un autre jury.

3. Exemple 3 : sécurité au carrefour

Un groupe de citoyens demande à la municipalité d'une ville la modification d'un carrefour en affirmant que 40 % des automobilistes tournent en utilisant une mauvaise file.

Un officier de police constate que sur 500 voitures prises au hasard, 190 prennent une mauvaise file.

1. Déterminer, en utilisant la loi binomiale sous l'hypothèse $p = 0,4$, l'intervalle de fluctuation au seuil de 95 %.

2. D'après l'échantillon, peut-on considérer, au seuil de 95 %, comme exacte l'affirmation du groupe de citoyens ?

Éléments de réponse

1. $[0,358 ; 0,444]$.

2. $f = 0,38$. L'affirmation est considérée comme exacte.

4. Exemple 4 : gauchers

Dans le monde, la proportion de gauchers est 12 %.

Soit n le nombre d'élèves dans votre classe.

1. Déterminer, à l'aide de la loi binomiale, l'intervalle de fluctuation au seuil de 95 % de la fréquence des gauchers sur un échantillon aléatoire de taille n .

2. Votre classe est-elle « représentative » de la proportion de gauchers dans le monde ?

Éléments de réponse

1. En prenant $n = 30$, l'intervalle de fluctuation au seuil de 95 % est $[0,033 ; 0,233]$ (entre 1 et 7 gauchers).

En prenant $n = 25$, l'intervalle de fluctuation au seuil de 95 % est $[0 ; 0,24]$ (entre 0 et 6 gauchers).

2. Cela revient à situer la fréquence observée dans la classe par rapport à l'intervalle de fluctuation.

5. Exemple 5 : ségrégation sexiste à l'embauche

(d'après document ressource pour la classe de Seconde. Ce problème peut être revisité à l'aide de la loi binomiale.)

Deux entreprises recrutent leur personnel dans un vivier comportant autant d'hommes que de femmes. Voici la répartition entre hommes et femmes dans ces deux entreprises :

	Hommes	Femmes	Total
Entreprise A	57	43	100
Entreprise B	1350 (54%)	1150 (46%)	2500

Peut-on suspecter l'une des deux de ne pas respecter la parité hommes-femmes à l'embauche ?

6. Exemple 6 : générateur de nombres aléatoires

1. Sur un tableur, on entre dans la cellule A1 la formule $\boxed{=ENT(ALEA()*2)}$, que l'on recopie vers le bas jusqu'en A100. On dénombre alors que le nombre 1 apparaît 58 fois dans la plage de cellules A1 : A100. Au seuil de 95 %, peut-on rejeter l'hypothèse selon laquelle le générateur de nombres aléatoires du tableur fonctionne bien ?

2. Même question si l'on recopie la formule vers le bas jusqu'en A1000, et que l'on dénombre 580 occurrences du nombre 1 dans la plage de cellules A1 : A1000.

E. Lien avec l'intervalle de fluctuation exploité en classe de Seconde

On considère une population où la proportion d'un caractère est p , dans laquelle on prélève, au hasard et avec remise, un échantillon de taille n . La variable aléatoire X correspondant au nombre d'observations du caractère sur un échantillon suit la loi binomiale de paramètres n et p .

On peut calculer à l'aide de la loi binomiale, notamment à l'aide de l'algorithme du paragraphe VII-C-1, l'intervalle $\left[\frac{a_n}{n}, \frac{b_n}{n} \right]$ de fluctuation au seuil de 95 % de la fréquence observée sur un échantillon de taille n , où a_n est le plus petit entier tel que $P(X \leq a_n) > 0,025$ et b_n est le plus petit entier tel que $P(X \leq b_n) \geq 0,975$.

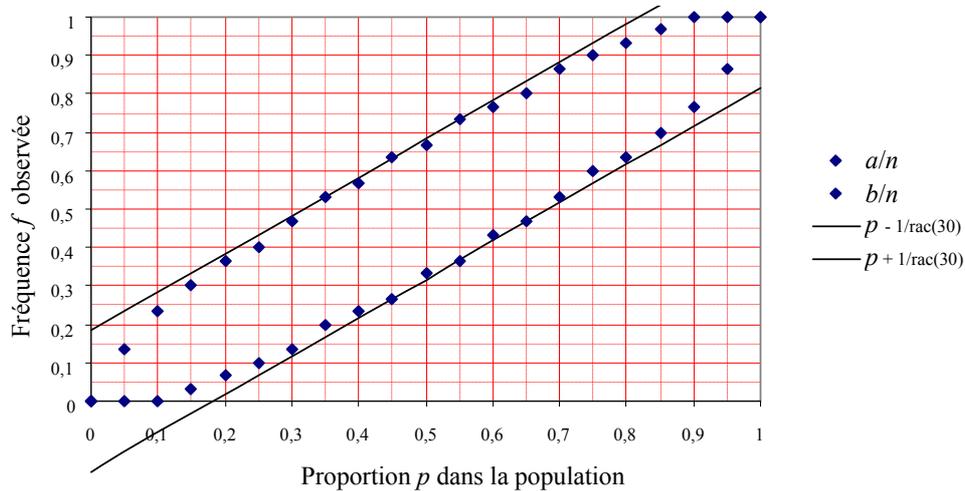
La notation $\left[\frac{a_n}{n}, \frac{b_n}{n} \right]$ retenue ici rappelle que les entiers a_n et b_n dépendent de l'entier n .

Les programmes demandent de comparer, pour une taille de l'échantillon importante, cet intervalle avec l'intervalle de fluctuation $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}} \right]$ exploité en classe de seconde. Ce dernier

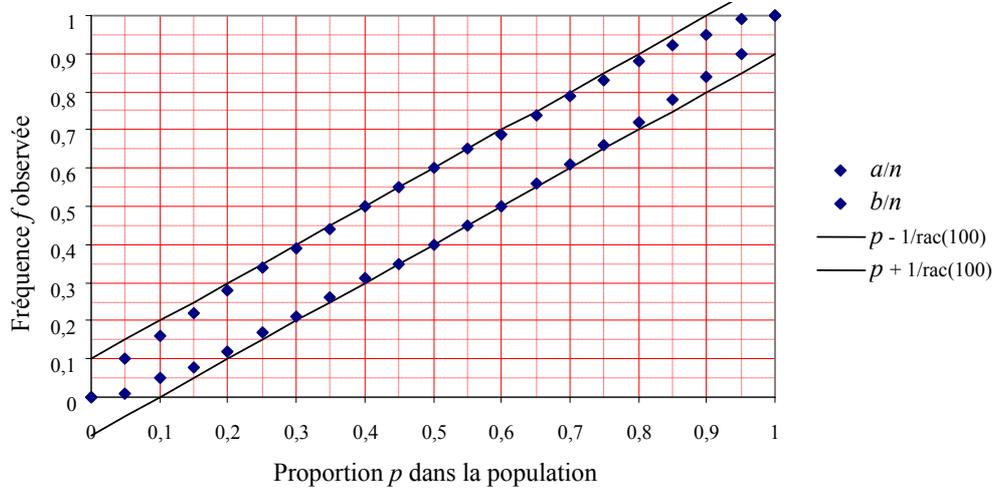
intervalle, plus facilement calculable, résulte d'approximations⁶, alors que la loi binomiale est la loi exacte correspondant à la situation.

En fixant différentes valeurs de n , il est possible de calculer les deux types d'intervalles pour p variant de 0 à 1. Les graphiques suivants ont été réalisés pour $n = 30$, $n = 100$ et $n = 1\,000$, en faisant varier, dans chaque cas, p de 0 à 1 avec un pas de 0,05.

Intervalle de fluctuation au seuil de 95 % pour des échantillons de taille 30

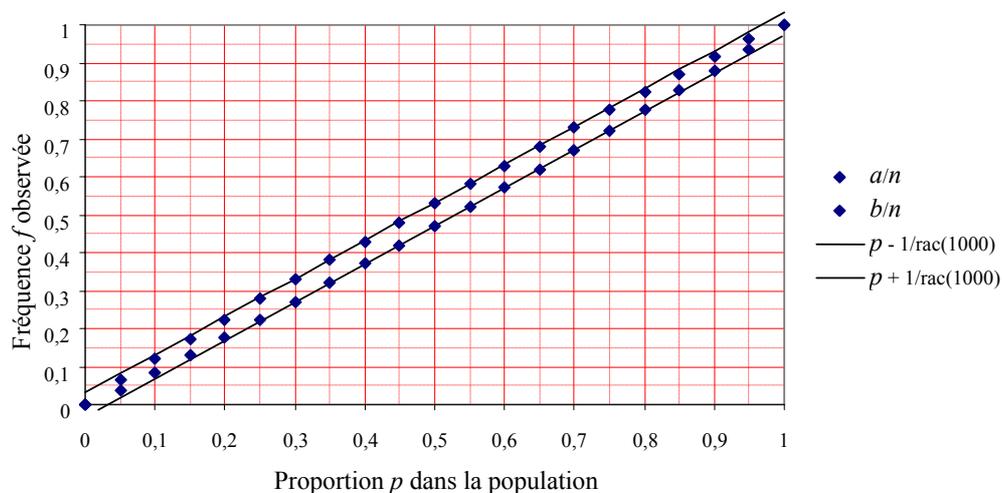


Intervalle de fluctuation au seuil de 95 % pour des échantillons de taille 100



⁶ Voir l'annexe 8.

Intervalle de fluctuation au seuil de 95 % pour des échantillons de taille 1 000



On observe que l'intervalle de fluctuation $\left[\frac{a_n}{n}, \frac{b_n}{n} \right]$ est sensiblement le même que l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}} \right]$ lorsque n est assez grand et p ni trop petit ni trop grand : pour $n = 30$, c'est le cas lorsque p est compris entre 0,3 et 0,7 ; pour $n = 100$, c'est le cas lorsque p est compris entre 0,2 et 0,8 ; pour $n = 1000$, c'est le cas lorsque p est compris entre 0,05 et 0,95.