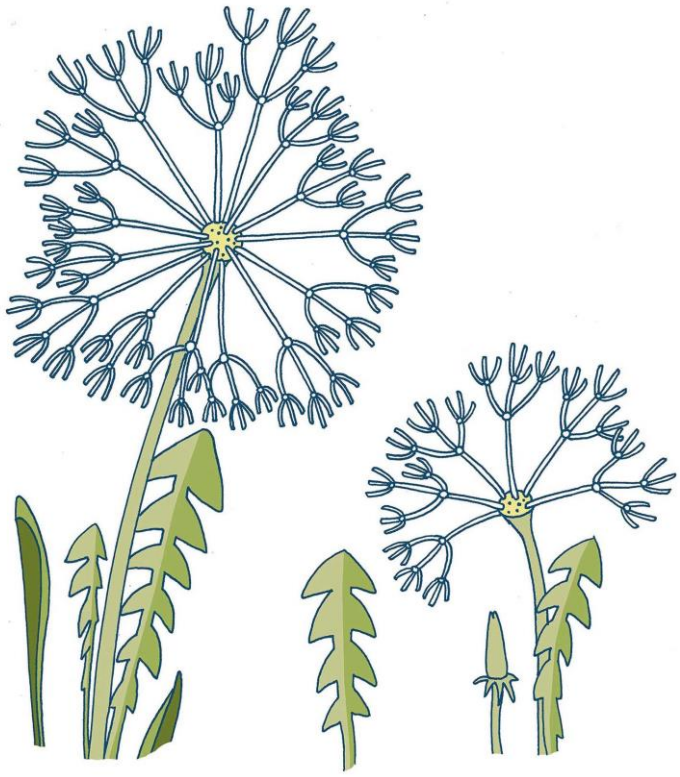
Que proposer aux élèves en statistiques et probabilités du collège au lycée ?

Annette Corpart et Nelly Lassalle



Avril 2016

IREM
Campus Universitaire des Cézeaux – 3 place Vasarely
TSA 60026 – CS 60026 – 63178 Aubière cedex
Tél. : **04 73 40 70 98** – Mail : **irem@univ-bpclermont.fr** – Site : **www.irem.univ-bpclermont.fr**

Préface

Probabilités ! Statistiques ! L'évocation de ces deux mots suscite souvent bien des craintes chez nos étudiants. Elles sont justifiées : probabilités et statistiques mêlent les habituelles difficultés techniques et logiques des mathématiques au délicat problème de la modélisation de situations concrètes. Quel mathématicien, même le plus chevronné, n'a pas un jour fait un raisonnement probabiliste erroné ?

Pourtant, au gré des réformes des programmes, ces deux disciplines ont pris une importance de plus en plus grande dans l'enseignement secondaire. On ne peut que s'en réjouir car les probabilités et les statistiques sont utiles et nécessaires à tout citoyen éclairé et critique sur les masses de données dont nous abreuvent les médias.

L'ouvrage d'Annette Corpart et Nelly Lassalle vient répondre à l'appréhension ressentie par nos collègues enseignants des collèges et lycées lorsqu'ils doivent aborder ces chapitres que parfois eux-mêmes n'ont pas, ou peu, rencontrés pendant leurs études. Il propose un corpus d'activités couvrant tout le(s) programme(s), de la 3ème à la Terminale. Toutes ces activités correspondent à de vraies situations concrètes, avec des données réelles, et pour lesquelles la manipulation des outils probabilistes et statistiques apporte un éclairage nouveau et décisif. Les professeurs y trouveront donc une formidable ressource de problèmes à étudier avec leurs élèves. Cet ouvrage mérite en réalité de dépasser le premier cercle auquel il est destiné, tant la pertinence du choix des situations évoquées est frappante et tant leur analyse permet de réfléchir aux « vérités » qui nous sont assénées. Je suis sûr qu'il atteindra son but : développer l'esprit critique de nos jeunes générations.

Frédéric Bayart,
Membre de l'Institut Universitaire de France,
Professeur au Laboratoire de Mathématiques de l'Université Blaise Pascal

Avant propos

Depuis 2000, nous sommes membres du groupe IREM probabilités/statistiques de l'académie de Clermont-Ferrand et de la commission inter-IREM qui réunit à Paris trois fois par an des animateurs IREM et des universitaires de différentes académies.

Nous avons été sollicitées par nos inspecteurs régionaux pour assurer la formation continue des enseignants en probabilités/statistiques lors des nouveaux programmes de collège et de lycée (de 2009 à 2014).

Nous avons donc été amenées à travailler sur l'introduction des notions de probabilités/statistiques de la troisième à la terminale. Partant d'expériences « à la main » ou de simulations avec les TICE, nous avons créé de nombreuses activités. Signalons que la plupart de ces activités ont été testées dans nos classes de lycée en cours de mathématiques.

Nous vous proposons de les découvrir.

Nous remercions les membres du groupe IREM liaison lycée / BTS pour la relecture de cet ouvrage ainsi que Florine pour ses illustrations.

Sommaire

Statistiques

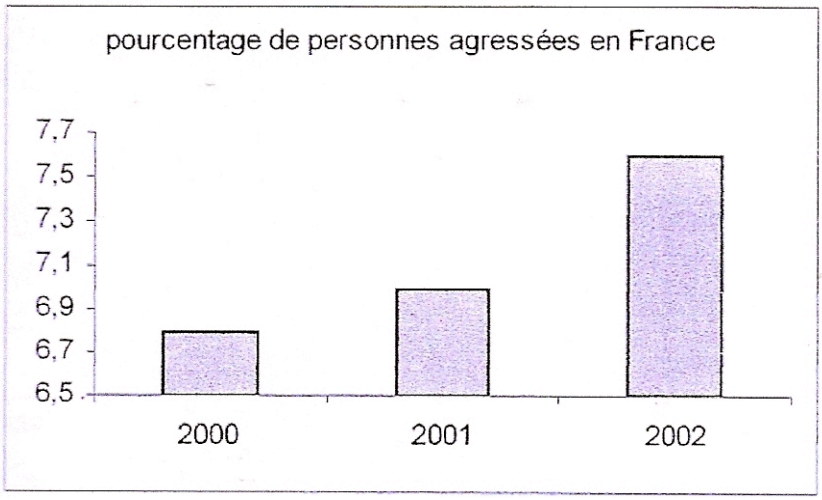
Lecture de graphiques

- **Niveau :** collège et seconde.

- **Objectif :** développer un regard critique.

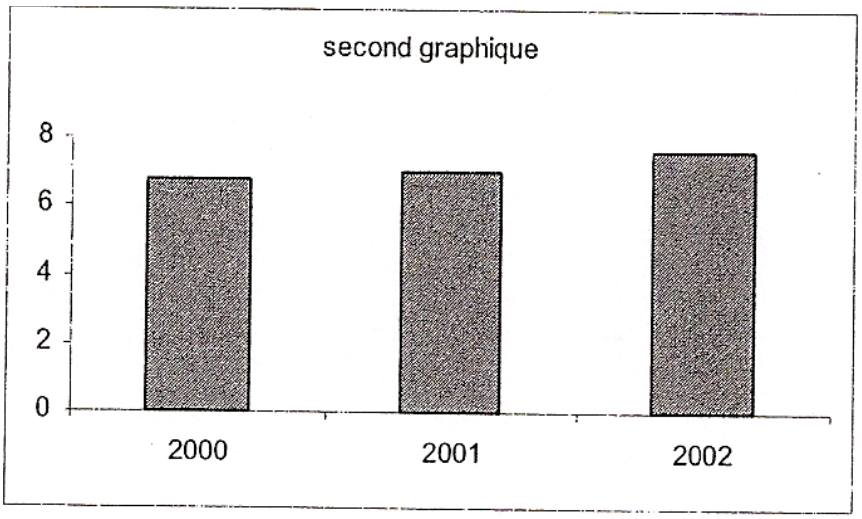
- **Exercice 1 :**
  Lu dans un magazine : L'explosion de l'insécurité en France continue entre 2000 et 2002. Depuis 2000, le pourcentage de personnes agressées augmente de manière dramatique. Le graphique ci-dessous, qui donne le pourcentage de personnes agressées chaque année depuis 2000, montre bien l'explosion de violence dont nous sommes aujourd'hui victime.

  (*Source : Insee, enquêtes permanentes sur les conditions de vie des ménages*).



1. Refaire un histogramme dont l'axe des ordonnées est gradué à partir de 0.
2. Critiquer.

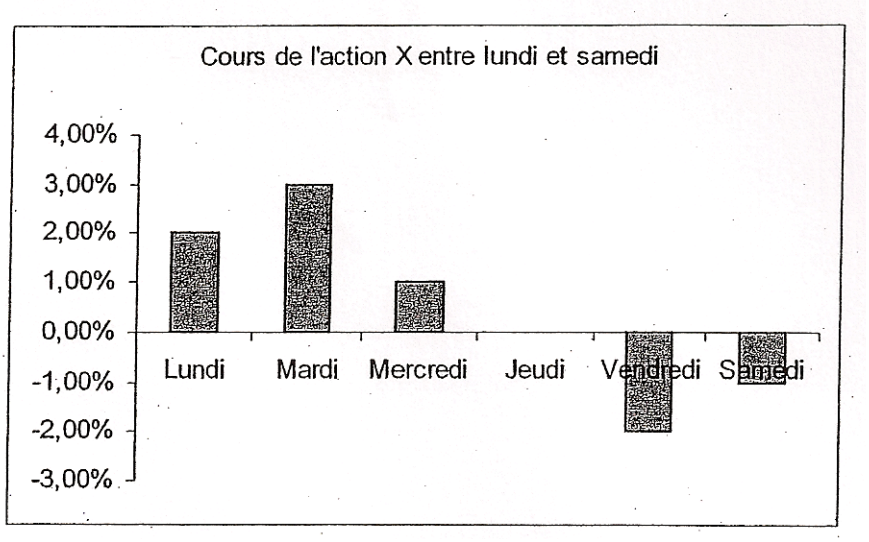
Corrigé :



L'explosion de l'insécurité est beaucoup moins nette avec le second histogramme !

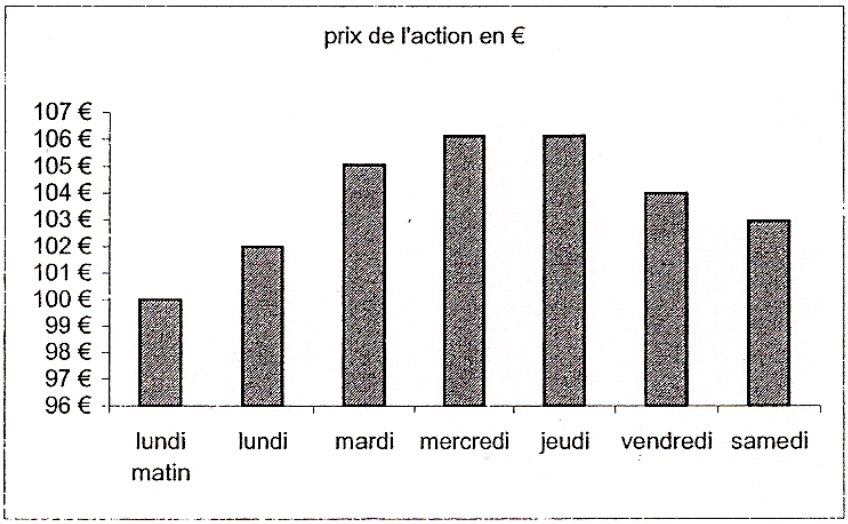
- **Exercice 2 :**

Un graphique d'une revue boursière donne les variations du cours d'une action, au jour le jour, pendant une semaine donnée : au vu de ce graphique, on pourrait penser que la bourse a « dégringolé » entre lundi et samedi. Qu'en est-il ?



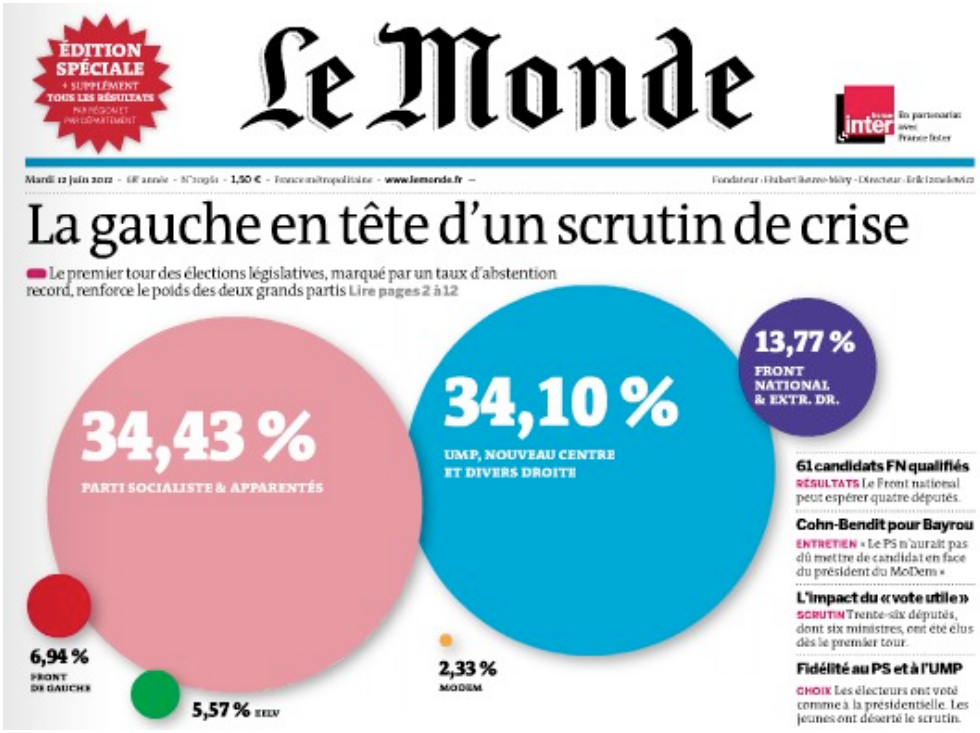
1. En supposant que l'action coûte à l'ouverture le lundi 100 €, refaire un histogramme donnant le prix de l'action en fonction des jours de la semaine.
2. Combien coûte l'action à la fermeture le samedi ? Critiquer.

Corrigé :



Le prix de l'action est plus élevé à la fermeture le samedi que le lundi à l'ouverture : l'action vaut 100 € le lundi et 103 € le samedi !

- **Exercice 3 :**
  Le journal *Le Monde*, dans son édition du 12 juin 2012, a consacré sa « Une » aux résultats du premier tour des législatives, à l'aide d'une infographie :

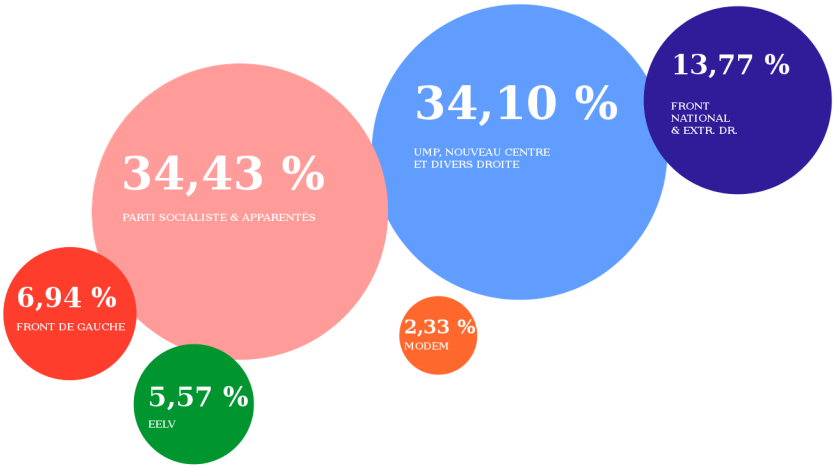


1. On rappelle que dans un tel graphique la surface des disques doit être proportionnelle au pourcentage de vote. L'infographie est-elle réalisée de façon correcte ?
   Les résultats du premier tour pour chaque parti politique sont redonnés dans le tableau ci-dessous :

PS	UMP	FN	Front de Gauche	Les Verts	MoDem
34,43%	34,10%	13,71%	6,94%	5,57%	2,33%

2. Proposer une représentation correcte des résultats de cette élection.

Corrigé :
La superficie des cercles, censée symboliser le poids respectif de chaque parti, n'est pas proportionnelle aux suffrages obtenus par les différents partis politiques : le cercle du PS devrait par exemple être environ 5 fois plus gros que celui du Front de Gauche ; or il est représenté environ 25 fois plus gros !
Si on respecte une superficie des cercles proportionnelle aux suffrages, on obtient le graphique suivant :



Dans *Le Monde,* ce sont les diamètres des cercles qui sont (à peu près) proportionnels aux résultats (sauf pour le cercle du MoDem). Par cette bévue, *Le Monde* a amplifié visuellement le poids réel du bipartisme en France (et a donc contribué à amplifier ce phénomène dans l'esprit des lecteurs).

Moyenne et médiane

- **Niveau :** collège et seconde.

- **Objectif :** illustrer la différence de comportement entre moyenne et médiane.

- **Enoncé :** Dans une entreprise familiale, les salaires sont répartis comme suit :

  Direction : le patron : 7 000 €
  son frère : 4 500 €
  6 parents : 1 500 €
  Personnel : 5 contremaîtres : 1 200 €
  10 ouvriers : 800 €

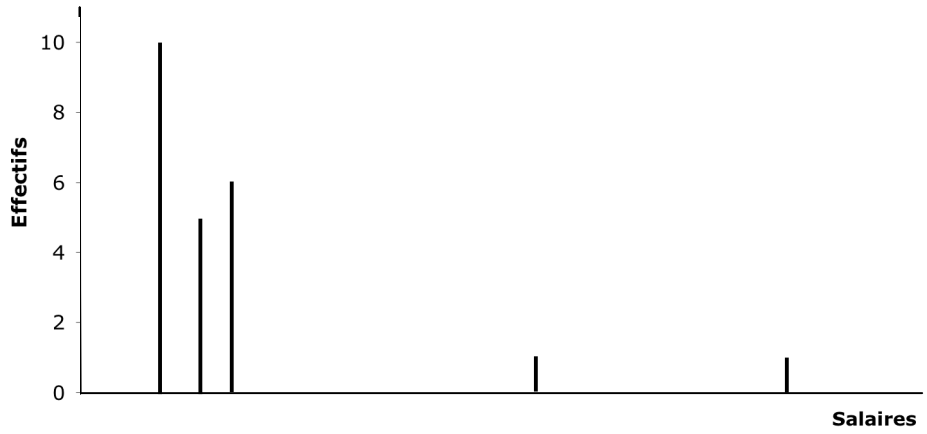
  1. Tracer le diagramme en bâtons de cette série statistique.

  2. La C.G.T. déclare que le salaire "moyen" est de 800 €. Les contremaîtres s'estiment, eux, dans la "moyenne". Quant au patron, il prétend que ses parents sont dans la "moyenne".
  Qui a raison ? Comment justifier les différences de position de chacun ?

  3. Que deviennent la moyenne et la médiane si le salaire du patron est de 15 000 € ?

- **Corrigé :**

  1. Diagramme en bâtons :



  2. Pour cette série statistique : 800 € : le mode
  1 200 € : la médiane
  1 500 € : la moyenne

  3. La médiane ne bouge pas alors que la moyenne passe à 1 848 € (on a alors 21 personnes sur 23 qui ont un salaire inférieur au salaire moyen et 2 personnes sur 23 avec un salaire supérieur au salaire moyen).

On dit que la médiane est un paramètre « robuste » alors que la moyenne est une mesure très sensible aux valeurs extrêmes.

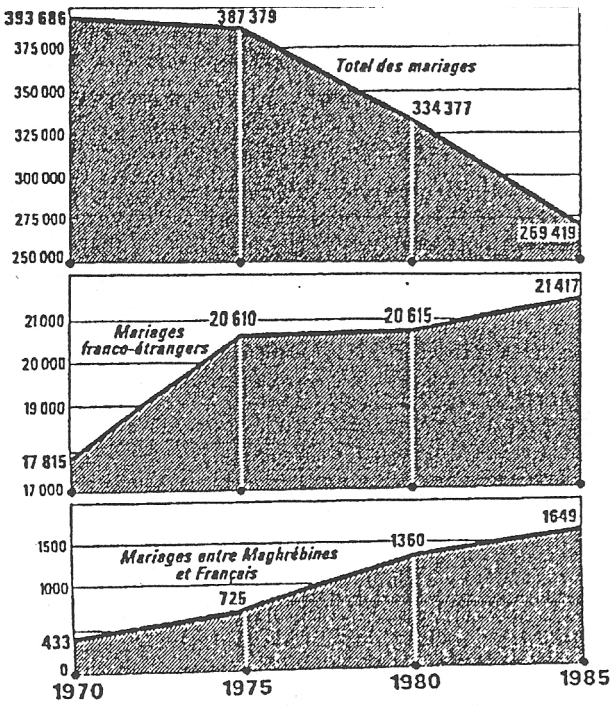
Une histoire de mariages

- **Niveau :** seconde

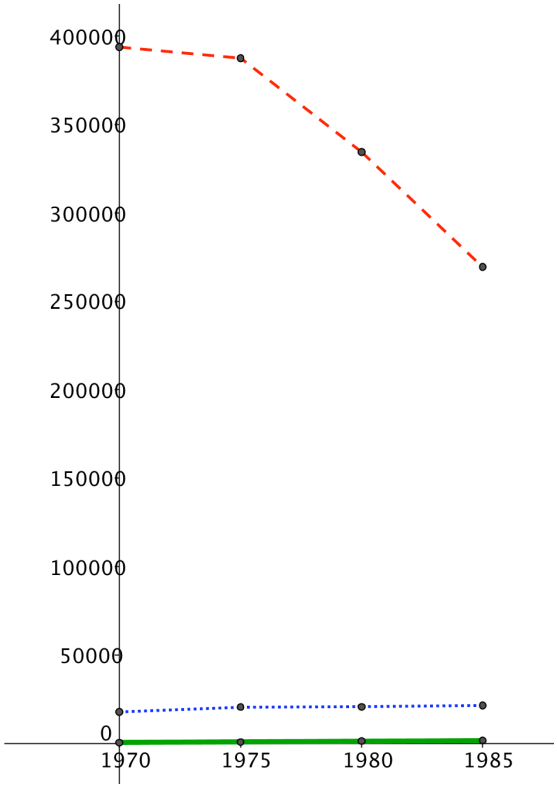
- **Objectifs :** – montrer l'influence des échelles graphiques.
  – calculer des indices (base 100) en liaison avec d'autres disciplines.

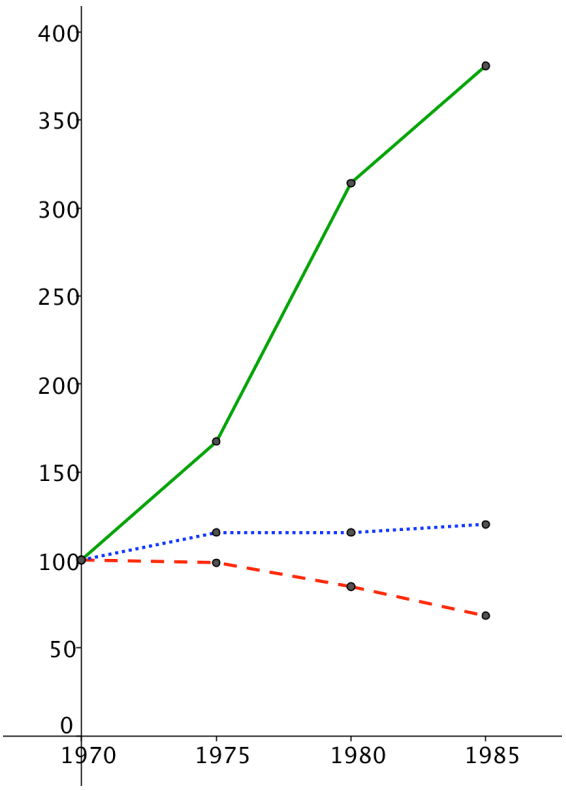
- **Enoncé :**
  Le journal *Le Monde* a donné dans une de ses éditions les graphiques ci-contre.
  1. Refaire les courbes (dans un même repère) en commençant à 0 l'échelle sur l'axe des ordonnées.
  2. Construire (toujours dans un même repère) les trois courbes, les valeurs étant rapportées à la base 100 en 1970.
     Pour cela, on complétera d'abord le tableau suivant :

	1970	1975	1980	1985
Total des mariages	100			
Mariages franco-étrangers	100			
Mariages maghrébines/français	100			



- **Corrigé :**





Total des mariages
Mariages franco-étrangers
Mariages maghrébines/français

	1970	1975	1980	1985
Total des mariages	100	98,4	84,9	68,4
Mariages franco-étrangers	100	115,7	115,7	120,2
Mariages maghrébines/français	100	167,4	314,1	380,8

Construction et lecture de diagrammes en bâtons

- **Niveau :** seconde et première.

- **Objectifs :** – comparer deux séries.
  – aller plus loin dans la notion de dispersion en seconde.
  – introduire l'écart-type en première.

- **Enoncé :**
  1. Deux séries différentes sont données. Pour chacune des deux séries, calculer les différents paramètres statistiques demandés. Peut-on conclure que ces deux séries sont semblables d'un point de vue statistique ?
  2. Construire les diagrammes en bâtons correspondants de ces deux séries. Que constate-t-on ?
  3. En admettant que ces deux séries représentent les résultats de deux groupes de cent étudiants à un QCM comportant 120 items, quel est le groupe le plus homogène ?

Série A :

Valeurs	14	25	37	50	62	78	85	103	118
Effectifs	2	7	10	17	24	15	12	9	4

Série B :

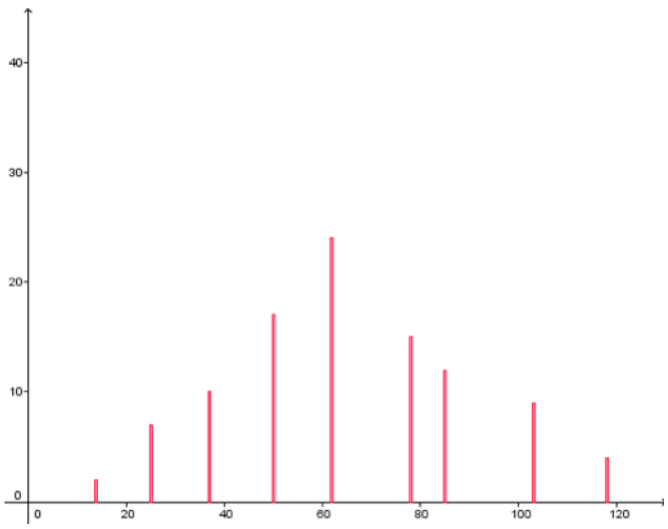
Valeurs	14	25	37	50	62	78	85	103	118
Effectifs	1	1	4	19	40	23	9	2	1

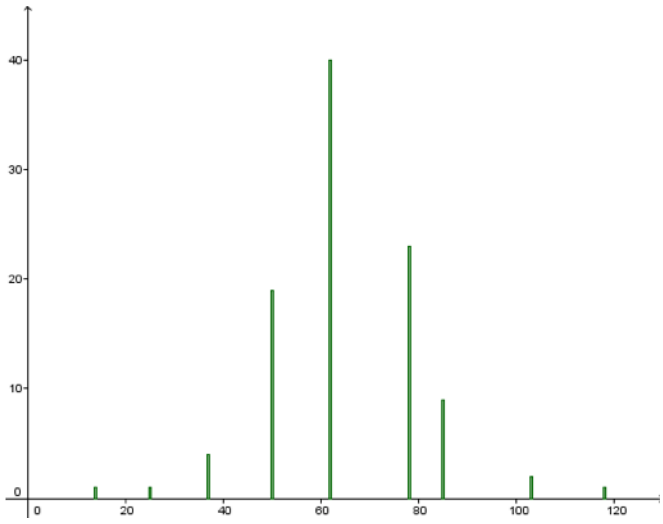
Caractéristiques Série A	Caractéristiques Série B
Effectif : Etendue : Mode : Moyenne : Médiane : Quartile Q_1 = Quartile Q_3 =	Effectif : Etendue : Mode : Moyenne : Médiane : Quartile Q_1 = Quartile Q_3 =

- **Corrigé :**

Série A
Effectif : **100**
Etendue : **104**
Mode : **62**
Moyenne : **65**
Médiane : **62**
Quartile Q_1 = **50**
Quartile Q_3 = **78**

Série B
Effectif : **100**
Etendue : **104**
Mode : **62**
Moyenne : **65**
Médiane : **62**
Quartile Q_1 = **50**
Quartile Q_3 = **78**





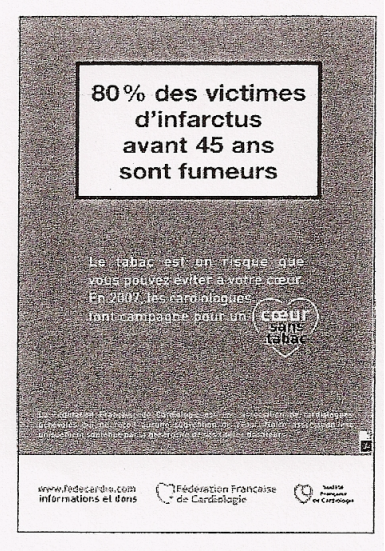
- **Remarques :**
  – Ces résultats ont été obtenus en utilisant la définition du programme pour les quartiles. Avec les listes statistiques d'une calculatrice TI, on obtient des résultats un peu différents :
  Q_1 = **50 et** Q_3 = **81,5** pour la série A ; Q_1 = **56** et Q_3 = **78** pour la série B.
  – La majorité des élèves considèrent que les deux séries sont semblables. Quelques-uns cependant remarquent que la répartition des effectifs n'est pas la même avant même d'avoir fait les diagrammes en bâton.
  – Les calculatrices donnent comme écart-type 24,46 pour la série A et 15,86 pour la série B.

Tabac et risque d'infarctus

- **Niveau :** collège et seconde.

- **Objectifs :** – calcul de proportions, calcul littéral.
  – interprétation des résultats, réflexion sur un sujet de société : montrer que des connaissances mathématiques permettent de mieux comprendre des statistiques médicales.

- **Activité :**



Dans une campagne de presse, la Fédération française de cardiologie donne l'information suivante : « 80 % des victimes d'infarctus avant 45 ans sont fumeurs ».
Cette donnée donne à penser que fumer augmente le risque d'infarctus. Peut-on calculer l'augmentation de ce risque ?

1. Parmi les victimes d'infarctus ayant moins de 45 ans, montrer qu'il y a 4 fois plus de fumeurs que de non–fumeurs.

2. Dans la population des moins de 45 ans, on souhaite calculer l'augmentation du risque d'avoir un infarctus si on est fumeur.
On peut estimer qu'en France, parmi les moins de 45 ans, il y a environ 40 % de fumeurs (ou d'anciens fumeurs).

a) On note n le nombre de personnes de moins de 45 ans et i le nombre de cas d'infarctus observés chez les moins de 45 ans.
Remplir les cases non grisées du tableau ci-dessous, représentatif de la population des moins de 45 ans :

	Nombre de fumeurs	Nombre de non-fumeurs	Total
Nombre de victimes d'infarctus			
Nombre de non victimes			
Total			n

b) En déduire la proportion q d'infarctus parmi les fumeurs et la proportion q' d'infarctus parmi les non-fumeurs.

c) Montrer que $\frac{q}{q'} = 6$. Interpréter ce résultat.

- **Corrigé :**

1. Parmi les victimes d'infarctus ayant moins de 45 ans, il y a 80 % de fumeurs et donc 20 % de non-fumeurs.

2.b) $q = \frac{0{,}8 \times i}{0{,}4 \times n}$; $q' = \frac{0{,}2 \times i}{0{,}6 \times n}$.

c) $\frac{q}{q'} = 6$: on peut interpréter ce résultat en disant que pour les moins de 45 ans, un fumeur a 6 fois plus de risques d'avoir un infarctus qu'un non-fumeur.

- **Commentaires :**

Le résultat obtenu n'est pas intuitif.
On peut remarquer que le « risque » d'infarctus lui-même, qui correspond à q pour un fumeur et à q' pour un non fumeur n'a pas été calculé et est sans doute assez faible (on peut penser que la population n est grande par rapport au nombre d'infarctus i.) Pour un risque relativement faible, le message « risque multiplié par 6 » est par contre frappant. Par ailleurs, si le risque individuel est relativement faible, il n'en est pas de même globalement pour la société.

Le rapport de chances

- **Niveau :** lycée et post-bac.

- **Objectifs :** – comparer différentes mesures statistiques.
  – introduire de nouveaux indicateurs statistiques.

- **Activité :**
  Il y a un demi-siècle, 45 % des enfants de cadres obtenaient le baccalauréat, contre seulement 5 % des enfants d'ouvriers. Désormais, 90 % des enfants de cadres l'obtiennent, contre 45 % des enfants d'ouvriers. Que nous disent ces données sur l'évolution des inégalités sociales d'accès à ce diplôme de fin d'enseignement secondaire ? La conclusion dépend largement de l'indicateur que l'on utilise.

  1. **Première mesure : la différence entre proportions**
     Il y a 50 ans, quelle est la différence entre les proportions de bacheliers chez les enfants de cadres et les enfants d'ouvriers ?
     Même question pour aujourd'hui. Conclure.

  2. **Deuxième mesure : le rapport entre proportions**
     Il y a 50 ans, quel est le rapport de bacheliers entre les enfants de cadres et les enfants d'ouvriers ?
     Même question pour aujourd'hui. Conclure.

  3. **Troisième mesure : le taux de variation par rapport au maximum de variation possible**
     Cette mesure consiste à comparer la variation réelle des pourcentages à la longueur du chemin qui restait à parcourir pour atteindre la proportion maximale de 100 %.
     Ainsi, en 50 ans, les enfants de cadres ont amélioré leur taux d'obtention du bac de 90 % – 45 % = 45 %, alors qu'ils pouvaient l'améliorer au maximum de 100 % – 45 % = 55 %. Le taux de bacheliers s'est donc amélioré de $\frac{45}{55} = 82\,\%$ du maximum de variation possible.
     Calculer de même le taux de variation pour les enfants d'ouvriers. Conclure.

  4. **Quatrième mesure : le rapport de chances**
     Il y a 50 ans, les enfants de cadres étaient 45 % à avoir le bac et donc 55 % à ne pas l'avoir. La « chance relative » est le rapport entre la fréquence d'avoir le bac et celle de ne pas l'avoir, soit $\frac{0{,}45}{0{,}55} = 0{,}82$. Pour les enfants d'ouvriers, leur chance relative d'avoir le bac était : $\frac{0{,}05}{0{,}95} = 0{,}053$.
     On obtient enfin le rapport de chances relatives des enfants de cadres d'avoir le bac et celles des enfants d'ouvriers : $\frac{0{,}82}{0{,}053} = 15{,}5$: il y a 50 ans, les enfants de cadres avaient 15,5 fois plus de chances que les enfants d'ouvriers d'obtenir le baccalauréat plutôt que de ne pas l'obtenir.
     Calculer le rapport de chances aujourd'hui. Conclure.

- **Corrigé :**
  1. Il y a 50 ans, le taux d'obtention du baccalauréat était de 45 % – 5 % = 40 % plus élevé chez les enfants de cadres que chez les enfants d'ouvriers. Aujourd'hui, il est de 90 % – 45 % = 45 % en faveur des enfants de cadres. Par conséquent, l'écart s'est accru : avec cet indicateur, les inégalités d'obtention du baccalauréat ont **augmenté** au cours des 50 dernières années.

  2. Il y a 50 ans, les enfants de cadres étaient 9 fois plus nombreux à être bacheliers que les enfants d'ouvriers ; ils ne sont plus que 2 fois plus nombreux aujourd'hui.
     Avec cet indicateur, les inégalités d'obtention du baccalauréat ont **diminué** au cours des 50 dernières années.

3. Les enfants d'ouvriers ont amélioré leur taux d'obtention du baccalauréat de : $\frac{45-5}{100-5} = 42\ \%$ du maximum de variation possible.
Les enfants de cadres ont réduit plus vite que les enfants d'ouvriers la distance qui les séparait de l'idéal des 100 % d'une classe d'âge au baccalauréat. Avec cet indicateur, les inégalités d'obtention du baccalauréat ont **augmenté** au cours des 50 dernières années.

4. Aujourd'hui, le rapport de chances vaut : $\frac{\frac{0,9}{0,1}}{\frac{0,45}{0,55}} = \frac{9}{0,82} \approx 11$: les enfants de cadres ont 11 fois plus de chances que les enfants d'ouvriers d'obtenir le baccalauréat plutôt que de ne pas l'obtenir.
Avec cet indicateur, les inégalités d'obtention du baccalauréat entre enfants de cadres et enfants d'ouvriers ont **diminué** au cours des 50 dernières années, de 15 fois plus de chances relatives à 11 fois plus de chances relatives.

- **Commentaires :**
  Quelle mesure faut-il utiliser ? Nous disposons d'au moins quatre mesures, qui donnent des conclusions contradictoires : soit « noires » (les inégalités ont augmenté), soit au contraire « roses » (elles ont diminué). Aucune de ces mesures n'est meilleure ou plus valable que les autres.

  Le rapport de chances est une mesure souvent utilisée en sociologie ou en épidémiologie.

  Il y a un autre phénomène à prendre en compte lorsqu'on mesure l'évolution générale des inégalités scolaires : la structure sociale se modifie et la distribution des élèves entre les différents milieux sociaux change. Par exemple, il y a de plus en plus d'enfants de cadres et de moins en moins d'enfants d'ouvriers.

- **Prolongement :**
  Pour les enfants nés avant 1929, 35 % des enfants de cadres obtenaient le bac contre 1 % des enfants d'ouvriers. Calculer le rapport de chances pour cette génération.

Probabilités

En introduction

Lorsque des questions de probabilité sont posées aux élèves en termes de « chances de gagner ou de perdre », avant tout enseignement théorique, les élèves ne refusent jamais de répondre sous prétexte que cela ne leur a jamais été enseigné. Autrement dit, la plupart ont une conception a priori de la notion de probabilité, ce que l'on peut constater en leur proposant dès le collège le questionnaire ci-dessous.

❶ Je jette une pièce de monnaie non truquée. Combien ai-je de chances d'avoir « Pile » ?

Question : Peux–tu répondre à la question posée ? Oui Non

Si oui, réponds :

……

Si non, pourquoi ?

……

❷ Je lance un dé classique (non truqué). **a)** Combien ai-je de chances d'avoir « 2 » ?

b) Combien ai-je de chances d'avoir un numéro pair ?

Questions : ✦ Peux–tu répondre à la question **a)** ? Oui Non

Si oui, réponds :

……

Si non, pourquoi ?

……

✦ Peux–tu répondre à la question **b)** ? Oui Non

Si oui, réponds :

……

Si non, pourquoi ?

……

❸ Une urne contient 3 boules jaunes et 4 boules rouges. Les boules sont indiscernables au toucher. Je tire une boule (sans regarder !) **a)** Combien ai-je de chances de tirer une boule jaune ?

b) Combien ai-je de chances de tirer une boule rouge ?

Questions : ✦ Peux–tu répondre à la question **a)** ? Oui Non

Si oui, réponds :

……

Si non, pourquoi ?

……

✦ Peux–tu répondre à la question **b)** ? Oui Non

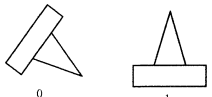
Si oui, réponds :

……

Si non, pourquoi ?

……

❹ Je lance une punaise. Combien ai-je de chances que la punaise tombe sur sa tête (position 1) ?



Question : Peux–tu répondre à la question posée ? Oui Non

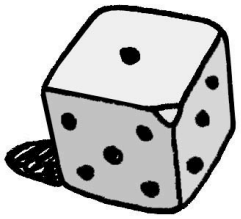
Si oui, réponds :

……

Si non, pourquoi ?

……

Lancer d'un dé



- **Niveau :** collège.

- **Objectifs :**
  - lors d'une expérience aléatoire, s'interroger sur la liste des résultats possibles, leur variabilité, leur imprévisibilité, leurs fréquences, l'existence d'une « répartition idéale » ;
  - réfléchir à l'indépendance des résultats vue sous l'angle d'absence de mémoire (le dé ne sait pas ce qui s'est passé avant) car les dés seront peu à peu remplacés par des « dés électroniques » ;
  - observer que des situations familières sont interchangeables pour décrire un modèle.

- **Déroulement de la séance :**
  - **expérimentation** : lancer un dé, tirer une carte, une boule ou un jeton. Cette première phase est indispensable.
  - **simulation** : utiliser le tableur et les fonctions ALEA.ENTRE.BORNES et NB.SI
    Il s'agit de mettre en évidence intuitivement la convergence des fréquences et de constater la nécessité de nombreuses expériences.

- **Remarque :**
  La convergence dans la loi des grands nombres est « lente ». Ce n'est pas parce que le hasard s'assagit avec le temps : quand on calcule une fréquence, on met le nombre n de lancers au dénominateur, si bien que l'effet des sautes d'humeur du numérateur sont amoindries lorsque n est grand.

Fiche élève

1. On va lancer un dé. Pour cela, on met le dé dans un gobelet qu'on agite avant de le lancer sur un plateau à rebord. Le dé va rouler sur le plateau avant de s'arrêter.

Utilise la feuille des tableaux.

a) Note le résultat du premier lancer dans la première case du tableau A. Pouvais-tu prévoir ce résultat ?
b) Explique pourquoi le résultat du 2ème lancer ne peut pas être prévu et ne dépend pas du résultat précédent.
c) Complète le tableau A en effectuant 60 lancers.
d) Compte les effectifs de 1, de 2, …… , de 6 du tableau A. Complète alors le tableau B. On exprimera les fréquences avec des nombres à deux décimales.
e) Pourquoi dit-on que le tableau B est un résumé du tableau A ? Quelles informations ont été perdues ?
f) Combien vaut la somme des fréquences du tableau B ? ***Démontre*** que cette somme est 1. Si tu as trouvé un résultat différent, quelle en est la raison ?
g) Compare tes tableaux avec ceux de tes voisins ? Sont-ils identiques ? Est-ce normal ?
h) Complète le tableau C pour avoir selon toi une répartition idéale des 1, 2, 3, 4, 5 et 6.

2. On va maintenant cumuler les résultats obtenus par tous les élèves de ta classe dans un tableau récapitulatif.

a) Choisis une face du dé (de 1 à 6) et remplis le tableau D.
b) Trace sur une feuille de papier millimétré les points :
   – d'abscisse : l'effectif cumulé des lancers.
   – d'ordonnée : la fréquence d'apparition correspondante pour la face choisie.
c) Que constates-tu ?

3. On joue maintenant à tirer au hasard une carte d'un paquet de 6 cartes de même couleur, composé de l'as, du 2, du 3, du 4, du 5 et du 6. On recommence en procédant toujours de la même manière : quand une carte est tirée et qu'on a noté son numéro, on la remet dans le paquet que l'on bat avant de tirer une nouvelle carte.
Effectue 60 tirages et note tes résultats dans le tableau E.

a) Peux-tu dire que, comme dans le lancer d'un dé, le résultat de chaque tirage est imprévisible et ne dépend pas des tirages précédents ?
b) En supposant que tu tires 10 fois de suite la carte 2, que pourrais-tu dire du résultat du 11ème tirage ?
c) Chaque carte a-t-elle autant de chances de sortir que les autres ? Déduis une répartition idéale des tirages.

4. On place dans un sac opaque 6 boules ou 6 jetons, indiscernables au toucher, numérotés de 1 à 6. Le jeu consiste à secouer le sac pour bien mélanger les boules, à plonger la main dans le sac, sortir une boule, noter son numéro et la remettre dans le sac. On répète 60 fois ce tirage.
Chaque boule a-t-elle autant de chances de sortir que les autres ? Déduis une répartition idéale des tirages.

5. Si on voulait tirer 1000 fois de suite un dé, une carte ou une boule comme précédemment, ce serait très long ! La solution consiste à simuler ces tirages à l'aide de certaines fonctions **d'un tableur**.

a) Ouvre l'assistant fonction et lis la notice de la fonction ALEA.ENTRE.BORNES(1;6). Cette fonction produit un nombre entier au hasard entre 1 et 6 exactement comme si on lançait un dé ou si on tirait une carte ou une boule comme ci-dessus. **Simule ainsi 1000 lancers d'un dé** des cellules A1 à J100.
b) Tu vas **construire l'histogramme en bâtons des fréquences d'apparition de chaque numéro**. Pour cela, présente ton travail en écrivant : *numéros des faces* en K1, *effectifs* en L1, *fréquences* en M1, *total* en K8. Remplis les cellules K2 à K7 avec les nombres de 1 à 6.
   Ouvre l'assistant fonction et lis la notice de la fonction NB.SI. On voit que le rôle de cette fonction est de compter les effectifs ; insère alors la formule =NB.SI($A$1:$J$100;K2) dans la cellule L2. Que compte-t-on avec cette formule ? Quel est le rôle du double dollar ? Recopie la formule jusqu'à la cellule L7. Vérifie que la somme de ces effectifs est 1000 en L8.

   Calcule les fréquences correspondantes de M2 à M7 puis leur somme en M8.

   Trace l'histogramme des fréquences avec l'assistant graphique.

   Refais plusieurs simulations en utilisant la touche F9. Que constates-tu ?

6. Recommencer la question 5 avec 10 0000 lancers.

Feuille des tableaux

Tableau A (60 lancers)

Tableau B (résumé des 60 lancers)

Résultat	1	2	3	4	5	6
Effectif						
Fréquence						

Tableau C (répartition idéale)

Résultat	1	2	3	4	5	6
Effectif						
Fréquence						

Tableau D (mise en commun des lancers) en choisissant d'étudier la face …. et en se rapportant au tableau B récapitulatif

Lancers cumulés	60	120	180	240	300	360	420	480	540	600	660	720
Cumul des effectifs de la face choisie												
Fréquence correspondante												

Lancers cumulés	780	840	900	960	1020	1080	1140	1200	1260	1320	1380	1440
Cumul des effectifs de la face choisie												
Fréquence correspondante												

Tableau E (60 tirages)

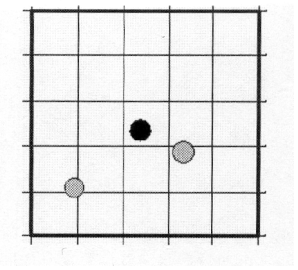
Probabilités et Géométrie

Jeu de « Franc Carreau »

- **Niveau :** lycée.

- **Objectifs :**
  - approche d'une probabilité (sans que cette valeur soit connue a priori, contrairement au jeu de pile ou face avec une pièce bien équilibrée ou au lancer d'un dé non pipé) à partir d'un grand nombre d'expériences.
  - justification géométrique.

- **Activité :**
  Dans le jeu de Franc Carreau, on lance un palet rond sur un parquet quadrillé, on fait « Franc Carreau » si le palet s'immobilise à l'intérieur d'un carreau :



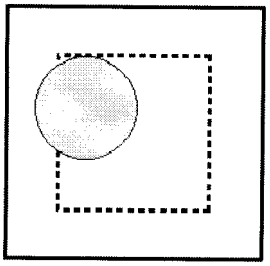
Prendre une pièce de monnaie de 1 centime d'euro et la lancer sur le plateau de jeu distribué[1]. On fait « Franc Carreau » (et on gagne) si la pièce tombe sur une seule case (elle peut toucher les bords mais pas empiéter sur une autre case). Sinon on perd.

1. Avant de jouer, peut-on savoir si on a davantage de chance de gagner que de perdre ?

2. Effectuer 30 fois ce jeu et compter le nombre de fois où on a fait « Franc Carreau ».

3. Mise en commun des résultats de tous les élèves de la classe : compter le nombre de « Franc Carreau » obtenus parmi les $30 \times N$ (avec N : nombre d'élèves de la classe).
Calculer la fréquence des « Franc Carreau ».

- **Calcul de la probabilité de gagner :**
  On peut déterminer cette probabilité à l'aide de considérations géométriques en la calculant à l'aide du rapport des aires de deux carrés.
  On lance un disque de rayon $r = 0{,}8$ cm (rayon de la pièce de 1 centime) sur une portion de plan pavée par des carrés de côtés $l = 3{,}2$ cm. Il y a « Franc Carreau » si le centre du disque s'immobilise à une distance supérieure à r des côtés du carré, donc dans un carré intérieur de $l - 2r = 1{,}6$ cm.



La probabilité de gagner est égale au rapport « surface favorable sur surface possible », donc au rapport des surfaces des deux carrés. On a :

$$\text{P(Franc Carreau)} = \frac{(l-2r)^2}{l^2} = \frac{1{,}6^2}{3{,}2^2} = 0{,}25$$

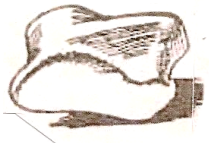
[1] Le plateau de jeu est à photocopier au verso.

Et si on ne sait pas calculer la probabilité d'un évènement ?

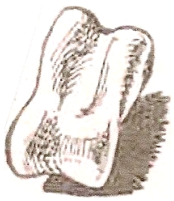
- **Niveau :** collège et seconde.

- **Objectif :** approche fréquentiste de la probabilité dans un cas où il n'y a pas d'autre approche possible : lorsque rien ne permet de calculer la probabilité d'un événement élémentaire, la stabilisation des fréquences conduit à une estimation de cette probabilité.

- **Activité :**
  Si on lance un osselet, il peut retomber selon quatre positions : la position « Bosse », la position « Creux », la position « I » et la position « J ».



Position « B »



Position « C »

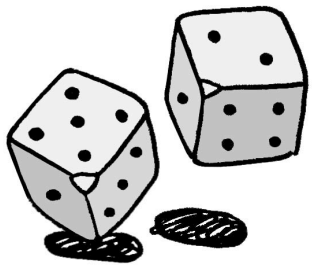
Position « I »

Position « J »

1. Peut-on attribuer a priori une probabilité à la position « C » ?
2. Mettre un osselet dans un gobelet. Agiter et effectuer 20 lancers. Pour chaque lancer, compter le nombre de fois où l'on obtient la position « C ».
3. Mise en commun des résultats de tous les élèves de la classe : compter le nombre de positions « C » obtenues parmi les $20 \times N$ (avec N : nombre d'élèves de la classe).
   Calculer la fréquence de la position « C ».
   On estime que la probabilité de la position « C » est donnée par la valeur de cette fréquence.

- **Remarque :**
  En réalisant des expériences aléatoires, on fait des statistiques et on calcule des fréquences d'apparition de telle ou telle issue. On ne détermine pas la probabilité (ou la loi de probabilité), mais on l'estime. Ensuite, une fois le modèle accepté, les raisonnements faits sur les probabilités sont tout à fait rigoureux.

Lancer de deux dés



- **Niveau :** collège et seconde.

- **Objectifs :** – déconstruire les idées préconçues des élèves.
  – se poser la question de l'équiprobabilité dans un univers.

- **Activité :** problème ouvert.
  On lance simultanément deux dés de couleurs différentes à six faces. On calcule la somme des nombres affichés sur les faces de dessus et on se pose la question suivante : « j'obtiens 10 en ajoutant les nombres *5* et *5* ou *4* et *6* ; de même, j'obtiens 9 en ajoutant les nombres *3* et *6* ou *4* et *5* ; à chaque fois, il y a deux sommes possibles. On me demande de jouer avec deux dés en pariant sur 10 ou sur 9 : mes chances d'obtenir 10 ou 9 sont-elles identiques ? »
  Proposer des solutions pour résoudre ce dilemme.

- **Solutions attendues :** – expérimenter ;
  – lister toutes les issues de l'expérience.

- **Compte-rendu d'expériences :**
  Les élèves ne proposent pas d'emblée de lancer les dés pour se faire une idée, même si le professeur a demandé de les apporter en classe.
  Ils proposent les réponses suivantes :

  – *On a autant de chance d'obtenir 10 ou 9 car le lancer de dé est du pur hasard.*

  – *On a autant de chance d'obtenir 10 ou 9.*
  En effet, pour obtenir 10, on a deux possibilités : 10 = 4 + 6 = 5 + 5 et pour obtenir 9, on a aussi deux possibilités : 9 = 3 + 6 = 4 + 5.

  – *Quand on lance les deux dés et qu'on additionne les faces, on a 11 sommes réalisables, de 2 à 12.*
  Pour obtenir 10, on a deux possibilités, ce qui fait une probabilité de 2/11.
  Pour obtenir 9, on a deux possibilités, ce qui fait une probabilité de 2/11.
  On a donc autant de chance de faire 9 que 10.

  – *Quand on lance les deux dés et qu'on additionne les faces, on a 11 sommes réalisables, de 2 à 12.*
  Pour obtenir 10, on a trois possibilités (4 et 6, 6 et 4, 5 et 5), ce qui fait une probabilité de 3/11.
  Pour obtenir 9, on a quatre possibilités (3 et 6, 6 et 3, 4 et 5, 5 et 4), ce qui fait une probabilité de 4/11.
  On a donc plus de chance de faire 9 que 10.

  – *J'ai pris deux stylos de couleur et j'ai écrit 36 combinaisons possibles pour les dés.*
  Pour obtenir 10, on a trois possibilités ce qui fait une probabilité de 3/36.
  Pour obtenir 9, on a quatre possibilités, ce qui fait une probabilité de 4/36.
  On a donc plus de chance de faire 9 que 10.

  Les propositions sont écrites au tableau et discutées. Les élèves n'arrivant pas souvent à se mettre d'accord, ils peuvent alors proposer de valider ou non leur solution en lançant deux dés de nombreuses fois.

- **Remarque :**
  Il est beaucoup plus simple de valider les 36 issues équiprobables si les dés sont de couleur différente. Certains élèves continuent ensuite de penser qu'avec deux dés parfaitement identiques, les résultats seraient différents, d'où la nécessité de les laisser manipuler.

Jeu des cartons

- **Niveau :** collège.

- **Objectifs :**
  – approche d'une probabilité (sans que cette valeur soit connue au départ) à partir d'un grand nombre d'expériences.
  – calcul de cette probabilité à partir d'un tableau à double entrée.

- **Activité :**
  On dispose de dix cartons. Sur chacun figure un nombre. Cinq de ces nombres sont positifs, les cinq autres sont négatifs. Vaut-il mieux faire le pari d'obtenir un nombre négatif en tirant un seul carton ou d'obtenir un produit négatif en tirant successivement et sans remise deux cartons ?

  1. Quelle est la probabilité p_1 d'obtenir un nombre négatif en tirant un seul carton ?

  2. On va estimer la probabilité p_2 d'obtenir un produit négatif en tirant successivement et sans remise deux cartons. Pour cela, on procède par expérimentation :
  – Tirer successivement et sans remise deux cartons. Reproduire 10 fois cette expérience.
  Compter le nombre de produits négatifs (sur les 10) obtenus avec les 2 cartons tirés.
  – Mise en commun des résultats de tous les élèves de la classe : compter le nombre de produits négatifs obtenus parmi les $10 \times N$ (avec N : nombre d'élèves de la classe).
  – Calculer la fréquence des produits négatifs.

  3. On va calculer la probabilité p_2 à partir d'un tableau à double entrée.
  Compléter le tableau suivant :

$2^{ème}$ tirage \ 1^{er} tirage										

  Qu'en est-il des « cases » de la diagonale ?
  Compter le nombre de « cases » où le produit est positif, celui où le produit est négatif.
  En déduire la probabilité cherchée.

- **Conclusion :**
  Considérons l'expérience aléatoire : « Tirer successivement et sans remise 2 cartons ».
  Notons Ω l'univers et A l'événement : « le produit obtenu est négatif ».

$$p_2 = \mathrm{P(A)} = \frac{\mathrm{Card\ A}}{\mathrm{Card\ \Omega}} = \frac{50}{90} = \frac{5}{9} > \frac{1}{2}$$

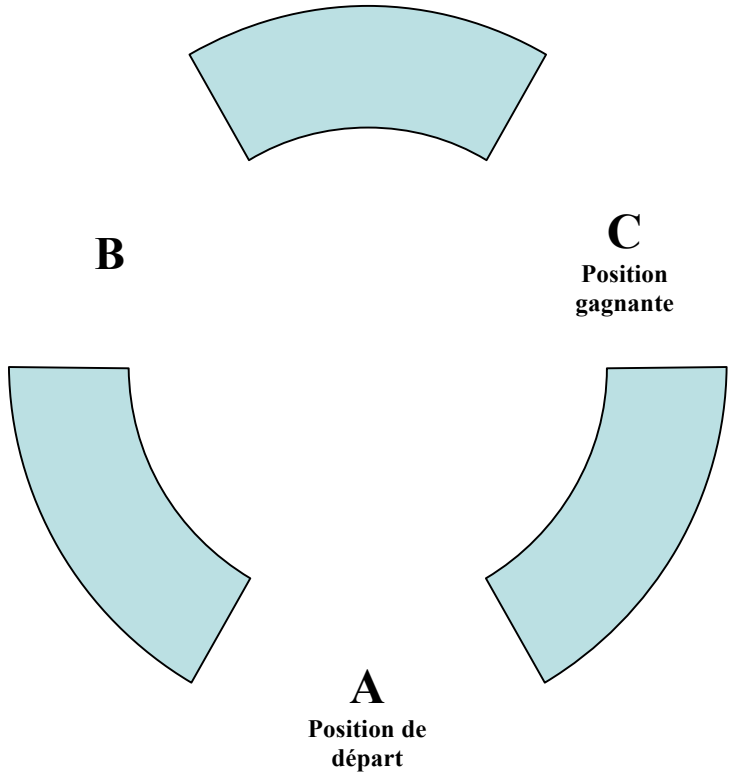
3	5	8	12	7
-1	-2	-9	-4	-6
3	5	8	12	7
-1	-2	-9	-4	-6
3	5	8	12	7
-1	-2	-9	-4	-6

Tourn'en rond

- **Niveau :** seconde.

- **Objectif :** introduire des arbres de probabilité à partir des arbres de fréquence.

- **Activité :**
  Se munir d'un dé, d'un jeton (une pièce de monnaie peut convenir) et du plateau de jeu ci-dessous.
  Place le jeton sur le point A. Un copain te propose deux règles de jeu:



Jeu n° 1	Jeu n° 2
❖ Lance le dé : si tu obtiens 5 ou 6, tourne et déplace le jeton en B; si tu obtiens 1, 2, 3 ou 4, tourne dans l'autre sens et déplace le jeton en C. ❖ Si tu es arrivé en C, tu as gagné la partie en un coup. Si tu es arrivé en B, relance le dé: si tu obtiens encore 5 ou 6, tu arrives alors en C et tu as gagné la partie en deux coups; sinon, tu retournes en A et la partie est perdue.	❖ Lance le dé : si tu obtiens 1, 2, 3, 4 ou 5, tourne et déplace le jeton en B; si tu obtiens 6, tourne dans l'autre sens et déplace le jeton en C. ❖ Si tu es arrivé en C, tu as gagné la partie en un coup. Si tu es arrivé en B, relance le dé: si tu obtiens encore 1, 2, 3, 4 ou 5, tu arrives alors en C et tu as gagné la partie en deux coups; sinon, tu retournes en A et la partie est perdue.

Si tu veux jouer avec ton copain, chacun d'entre vous devra mettre en C une sucette.
Celui qui gagne la partie empoche les deux sucettes.

En supposant que tu es gourmand, quel jeu vas-tu choisir pour t'assurer le maximum de chance de gagner ?

1. Simulation de l'expérience (à faire à la maison)

Effectue 40 parties avec le jeu n°1 et note dans le tableau ci-dessous tes 40 résultats en cochant chaque partie dans la bonne case.

Parties gagnées en un seul coup	Parties gagnées en deux coups	Parties perdues
Total $n_1 =$	Total $n_2 =$	Total $n_3 =$

Vérifie que $n_1 + n_2 + n_3 = 40$.

Recommence 40 parties avec le jeu n°2 et note dans le tableau ci-dessous tes 40 résultats en cochant chaque partie dans la bonne case.

Parties gagnées en un seul coup	Parties gagnées en deux coups	Parties perdues
Total $n_1' =$	Total $n_2' =$	Total $n_3' =$

Vérifie que $n_1' + n_2' + n_3' = 40$.

Quel jeu te paraît le plus favorable après cette simulation ?

2. Exploitation des résultats expérimentaux (étude statistique)

Mettons en commun les résultats de tous les élèves de la classe dans le tableau récapitulatif suivant :

Nom de l'élève	Jeu n°1			Jeu n°2		
	Parties gagnées en un seul coup	Parties gagnées en deux coups	Parties perdues	Parties gagnées en un seul coup	Parties gagnées en deux coups	Parties perdues
Total	$N_1 =$	$N_2 =$	$N_3 =$	$N_1' =$	$N_2' =$	$N_3' =$

On peut vérifier que $N_1 + N_2 + N_3 = N_1' + N_2' + N_3' = 40 \times N$, si N est le nombre d'élèves de la classe.

A partir de ce tableau, nous allons nous intéresser à la proportion de parties gagnées parmi les parties jouées, que nous appellerons ***la fréquence des parties gagnées***.

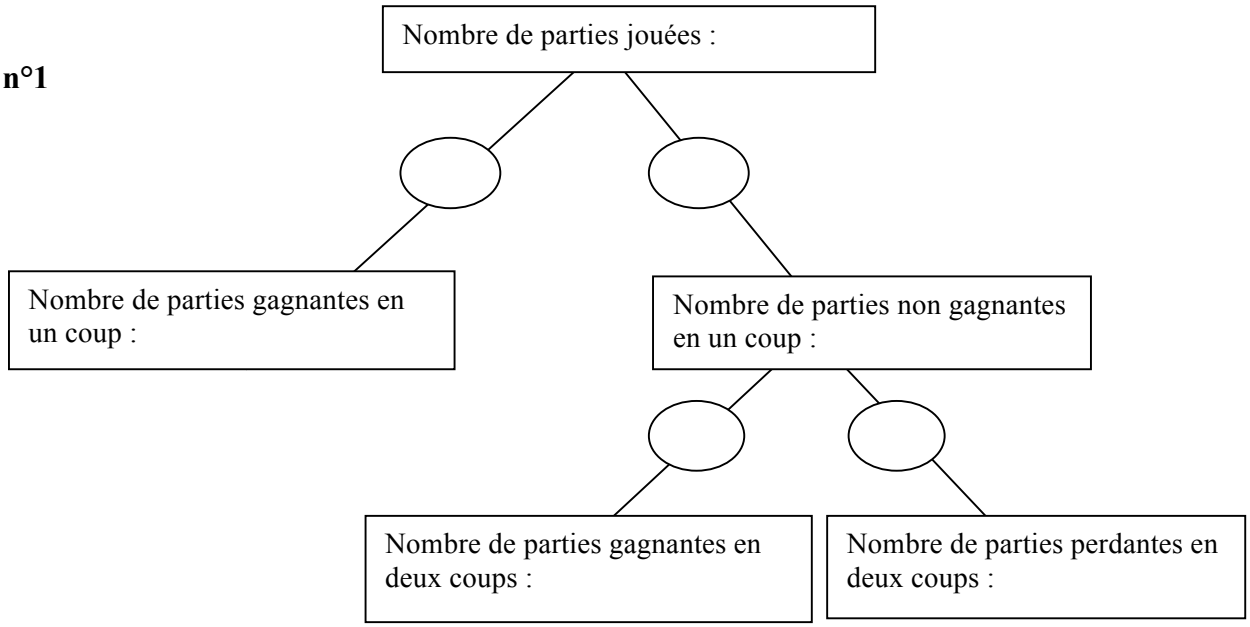
Calculer cette fréquence pour chaque jeu :

$f =$

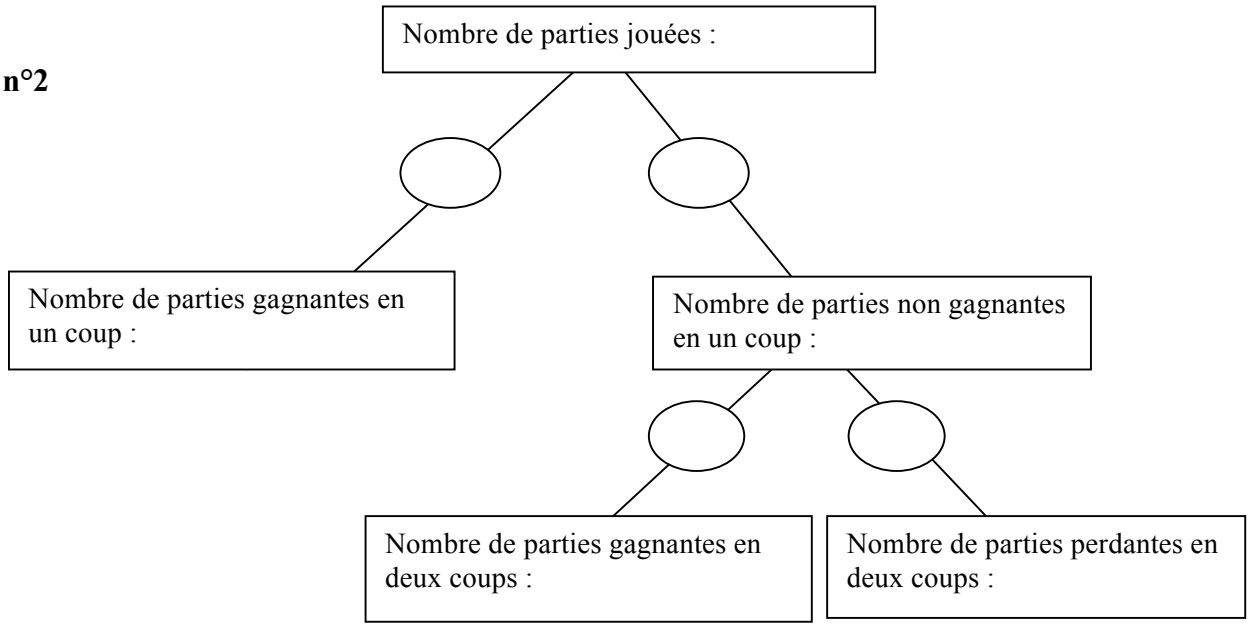
$f' =$

Résumons à présent ces résultats en complétant les schémas "en arbre" ci-dessous. Remplir les rectangles puis indiquer dans les ronds les *fréquences* correspondant à chacune des branches.

Jeu n°1



Jeu n°2



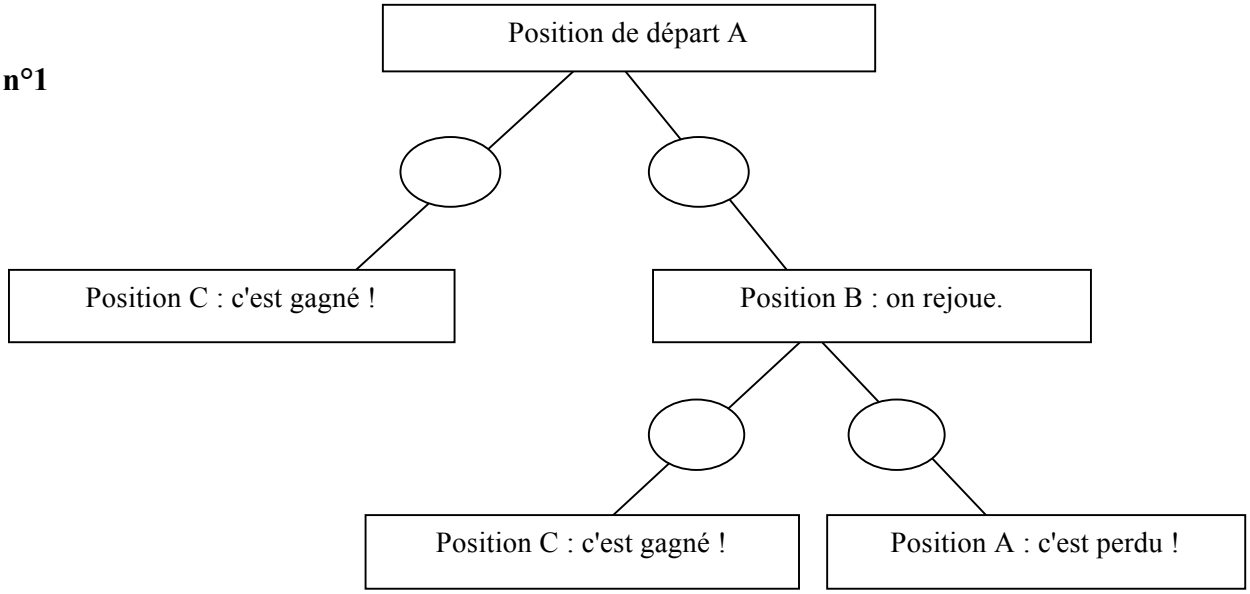
Comment peut-on retrouver *la fréquence des parties gagnées* à chaque jeu, en utilisant uniquement *les quatre fréquences* figurant dans le schéma correspondant ? De la même manière, déterminer *la fréquence des parties perdues*. Quelle relation existe-t-il entre ces deux résultats ?

3. Mesurer le hasard (domaine des probabilités)

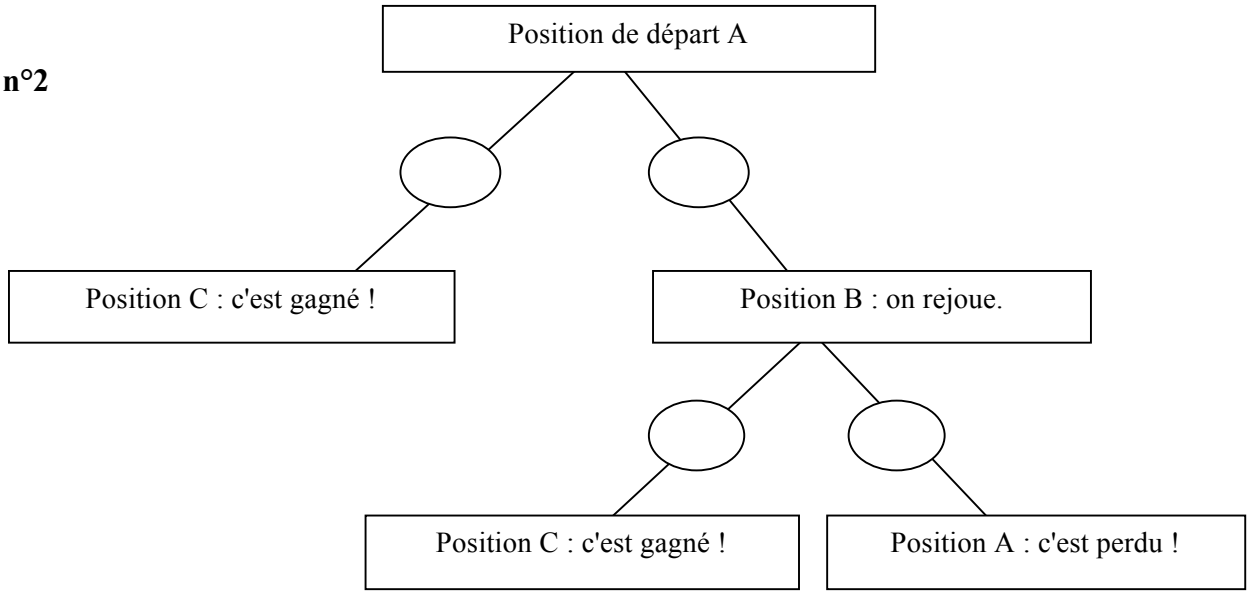
Monsieur Hasardus, savant probabiliste bien connu, prétend qu'il n'était pas nécessaire de réaliser les expériences car "*il est évident que la probabilité de gagner est de 7/9 au jeu n°1 et de 31/36 au jeu n°2 : il suffit de réaliser des arbres similaires aux précédents mais en écrivant des* ***probabilités*** *à la place des fréquences"*.

Saurez-vous faire aussi bien que lui ?

Jeu n°1



Jeu n°2

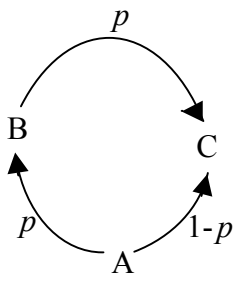


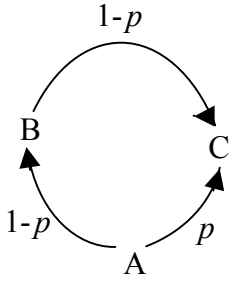
- **Commentaires et prolongements :**

  – On peut faire comparer les fréquences expérimentales avec les probabilités théoriques. La stabilisation autour de la valeur théorique est relativement lente.

  – En exercice ou en contrôle, on peut demander aux élèves de choisir eux-mêmes les faces du dé qui mènent à C puis de faire les calculs correspondants pour déterminer la probabilité de gagner.

  – On peut faire démontrer que, si p désigne la probabilité de tourner d'un cran dans un sens donné, deux jeux "symétriques" donnent la même probabilité de gagner.





  En effet, cette probabilité vaut $p^2 + (1-p)$ dans le premier cas et $(1-p)^2 + p$ dans le deuxième.

Jeu de l'attaquant

- **Niveau :** seconde.

- **Objectif :** introduire des arbres de probabilité à partir des arbres de fréquence.
  Cette activité a le même objectif que le tourn' en rond mais avec des arbres plus complexes.

- **Activité :**
  Au cours d'un jeu de société, deux joueurs s'affrontent : l'un dispose de six cartons jaunes portant les numéros 1,2,3,2,2,3 ; l'autre de six cartons bleus portant les numéros 1,2,3,1,2,2.
  Tous les cartons sont de taille identique, posés à l'envers sur la table devant chacun des joueurs et mélangés.
  Le joueur qui possède les jaunes attaque la partie : il choisit au hasard un de ses cartons et le retourne. Son adversaire retourne alors au hasard un de ses cartons bleus.
  L'attaquant gagne la partie si son carton marque **plus de points** que celui de son adversaire.
  A-t-on plus de chance de gagner si on joue avec les cartons jaunes ou avec les cartons bleus ?

1. Expérimentation

Tu vas jouer 2 fois 30 parties avec un camarade: 30 avec les cartons jaunes et 30 avec les cartons bleus. N'oublie pas de bien mélanger tes cartons avant toute nouvelle partie.
Ecris les résultats des 30 parties où tu joues en position d'attaquant (cartons jaunes) dans le tableau ci-dessous :

n° du carton jaune										
n° du carton bleu										
Carton gagnant (écrire J ou B)										

n° du carton jaune										
n° du carton bleu										
Carton gagnant (écrire J ou B)										

n° du carton jaune										
n° du carton bleu										
Carton gagnant (écrire J ou B)										

Complète :

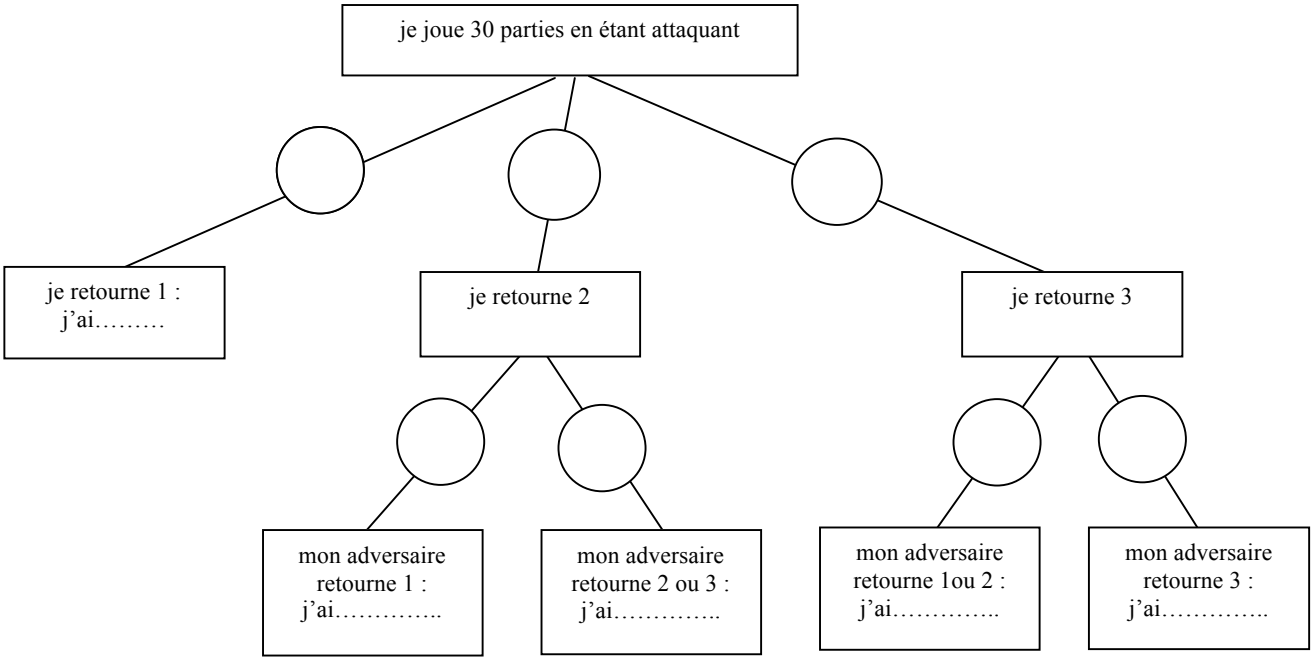
Le nombre de parties que tu as gagnées en étant attaquant est $n_1 =$

Le nombre de parties gagnées par ton adversaire est $n_2 =$

Vérifie que $n_1 + n_2 = 30$.

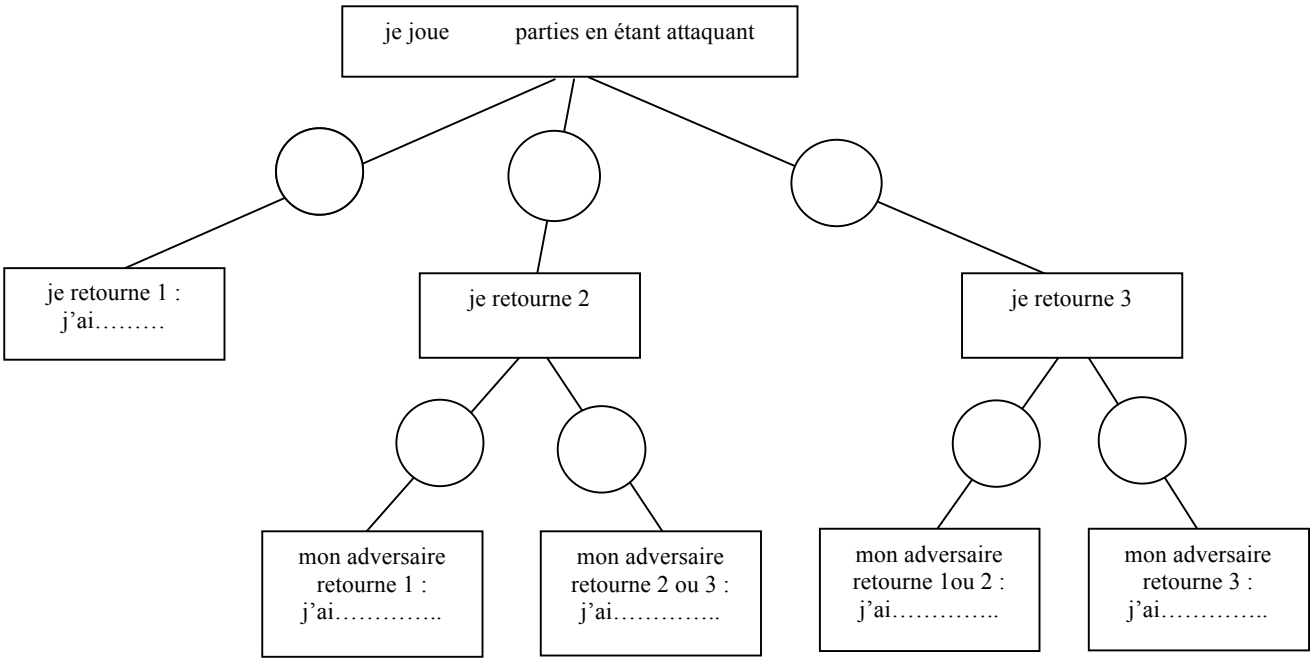
A cette étape, quelle couleur de carton choisir pour avoir plus de chance de gagner ?

Tu vas maintenant résumer les résultats des 30 parties précédentes en complétant l'arbre ci-dessous : remplis les rectangles par *« gagné »* ou *« perdu »* et indique dans les ronds l'effectif correspondant à chaque branche.



2. Exploitation des résultats expérimentaux (domaine des statistiques)

On va mettre en commun les résultats de toute la classe : complète **l'arbre des effectifs** suivant à partir du tableau récapitulatif fourni par le professeur.

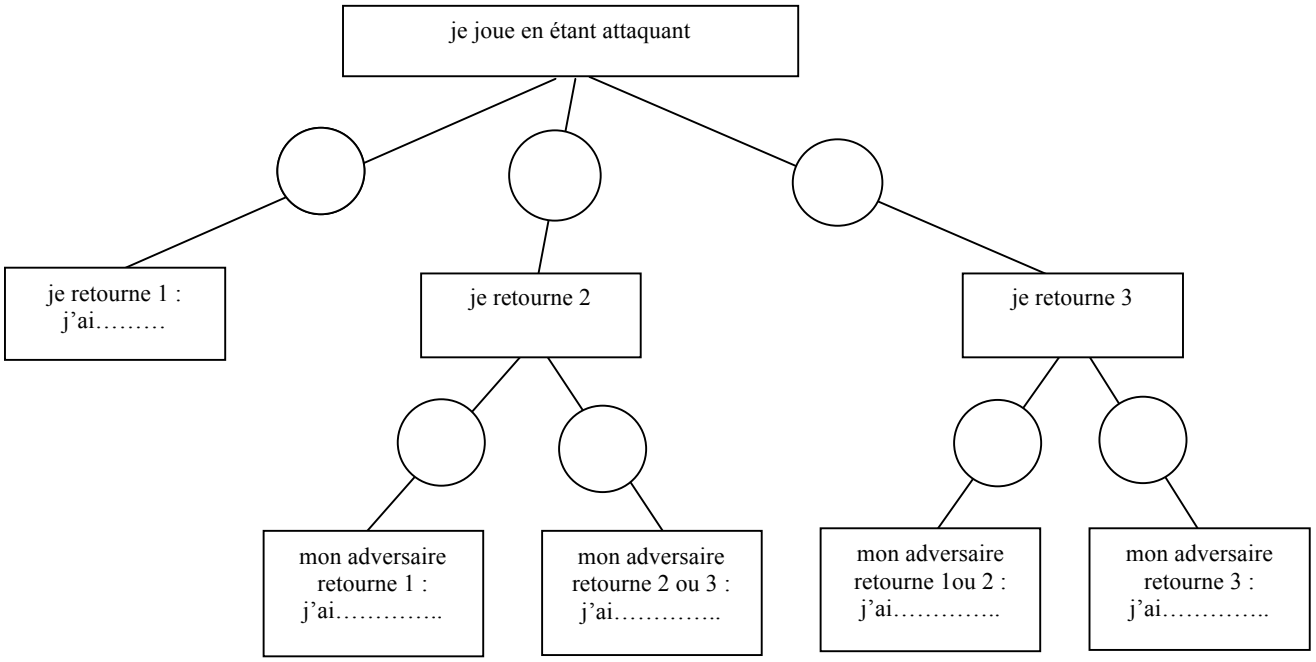


En position d'attaquant, le nombre de parties gagnées est $n_1 =$, ce qui correspond à une proportion de parties gagnées parmi les parties jouées de $f_1 =$. Nous appellerons ce nombre la ***fréquence des parties gagnées par l'attaquant***.

De même, pour l'adversaire, on peut calculer $n_2 =$ et $f_2 =$

Remarque : on a $f_1 + f_2 =$ …..

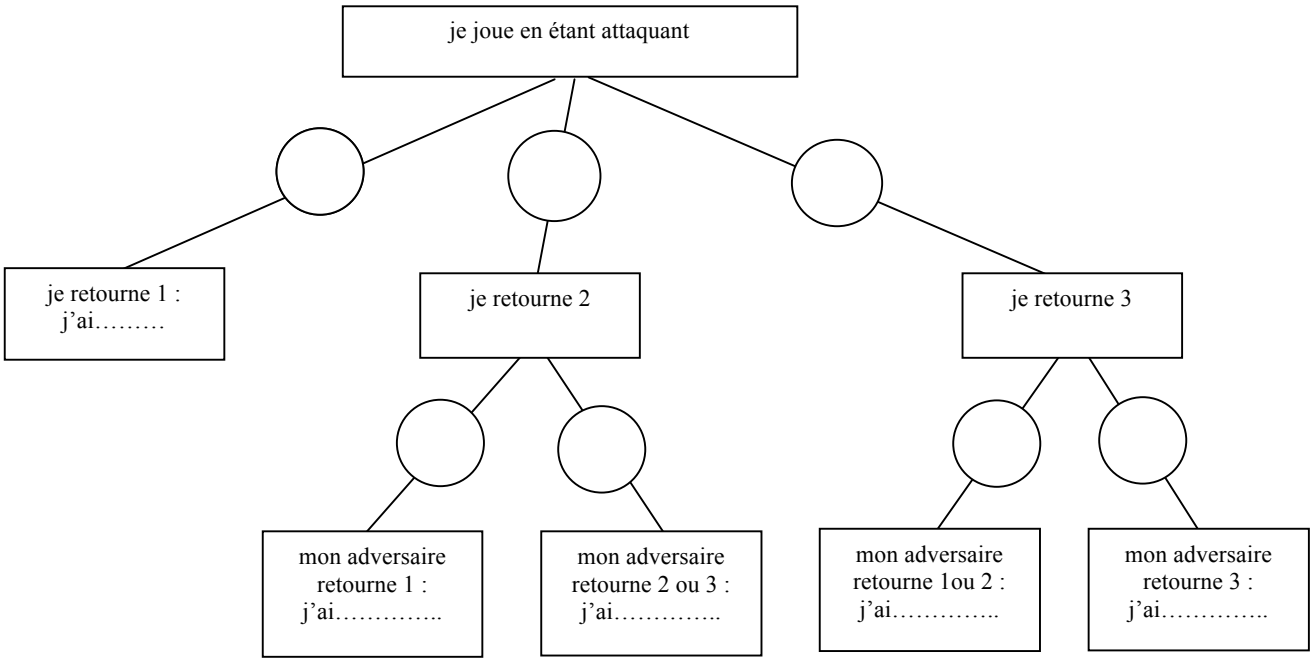
Complète **l'arbre des fréquences**, en indiquant cette fois dans les ronds les fréquences correspondant à chaque branche.



Comment peut-on retrouver f_1 et f_2 en utilisant les fréquences figurant dans les ronds de cet arbre ?

3. Mesurer le hasard (domaine des probabilités)

Monsieur Hasardus prétend qu'il n'était pas nécessaire de réaliser les expériences car *« il est évident que le jeu est défavorable à l'attaquant ; il suffit pour faire les calculs de réaliser des arbres similaires aux précédents en indiquant dans les ronds des probabilités à la place des fréquences. »*
Sauras-tu faire aussi bien que lui ?



Ascenseur social

- **Niveau :** terminale et post-bac.

- **Objectifs :** illustrer le fait qu'il faut comparer ce qui est comparable et en particulier tenir compte de l'évolution de la constitution de la population sous-jacente.

- **Activité :**
  Le recrutement dans les grandes Ecoles s'est-il démocratisé entre les années 1950 et 1990 ?

Proportion d'élèves reçus à l'École polytechnique	1950	1990
d'origine populaire	21 %	7,8 %
enfants de cadres et d'enseignants	79 %	92,2 %

D'après le tableau précédent, le verdict semble sans appel... l'ascenseur social est en panne.
Mais pour savoir si la discrimination sociale est plus forte dans les années 1990 que dans les années 1950, il faut tenir compte de l'évolution de la composition de la société française entre ces deux périodes.

Proportion de la population des 20-24 ans	1950	1990
d'origine populaire	90,8 %	68,2 %
enfants de cadres et d'enseignants	9,2 %	31,8 %

1. Que signifie 21 % dans le premier tableau ? Que signifie 90,8 % dans le second tableau ?

2. On prélève au hasard un jeune entre 20 et 24 ans dans la population française de 1950.
On note A l'événement « le jeune est d'origine populaire ».
On note B l'événement « le jeune est enfant de cadre ou d'enseignant ».
On note X l'événement « le jeune a été reçu à l'Ecole polytechnique ».

a) Ecrire en terme de probabilités les quatre nombres des colonnes 1950.

b) Justifier que $\mathrm{P}_A(X) = \dfrac{\mathrm{P}(X) \times \mathrm{P}_X(A)}{\mathrm{P}(A)}$ puis établir une formule analogue pour $\mathrm{P}_B(X)$.

c) En déduire que $\mathrm{P}_B(X) \approx 37 \times \mathrm{P}_A(X)$. Interpréter ce résultat par une phrase.

3. En s'inspirant de la question 2., montrer qu'en 1990 la probabilité d'être reçu à l'Ecole polytechnique était environ 25 fois plus grande pour un jeune enfant de cadre ou d'enseignant que pour un jeune d'origine populaire.

4. Que peut-on penser au sujet de « l'ascenseur social » entre les années 1950 et 1990 ?

- **Remarque :** l'indicateur $t = \dfrac{\mathrm{P}_B(X)}{\mathrm{P}_A(X)}$ s'appelle le « risque relatif de B » et il est très utilisé.

Par exemple, en médecine, si $t = 20$ pour une maladie X, la probabilité d'avoir cette maladie est 20 fois plus grande chez les B que chez les A.

- **Prolongement :** on peut calculer t pour d'autres grandes écoles.

Proportion d'élèves d'origine populaire	1951 - 1955	1989 - 1993
ENA (École nationale d'administration)	18,3 %	6,1 %
ENS (Écoles normales supérieures)	23,9 %	6,1 %
HEC (Hautes études commerciales)	38,2 %	11,8 %
Grandes écoles (avec entrée sur concours)	29 %	8,6 %

Accident nucléaire : une certitude statistique ?

- **Niveau :** lycée (première, terminale).

- **Objectif :** développer un regard critique.

- **Activité :**
  Le journal Libération a publié le 3 juin 2011 un article intitulé « Accident nucléaire : une certitude statistique », dont voici des extraits :

  Le parc actuel de réacteurs des centrales nucléaires cumule 14 000 années-réacteur, ce qui correspond à environ 450 réacteurs fonctionnant durant trente et un ans. ..... Sur ce parc, cinq réacteurs ont connu un accident grave (un à Three Mile Island, un à Tchernobyl et trois à Fukushima), dont quatre sont des accidents majeurs (Tchernobyl et Fukushima). .....
  L'Union européenne compte actuellement un parc de 143 réacteurs en fonctionnement. Sur la base du constat des accidents majeurs survenus ces trente dernières années, la probabilité d'un accident majeur sur ce parc serait donc de plus de 100%. Autrement dit, on serait statistiquement sûr de connaître un accident majeur dans l'Union européenne au cours de la vie du parc actuel. .....

  1. Calculer la probabilité (estimée à partir des observations statistiques) qu'un réacteur nucléaire pris au hasard ait un accident majeur par an ?

  2. Expliquer comment les auteurs de l'article sont arrivés à leur conclusion. Avez vous des commentaires ?

  3. On suppose que la probabilité d'un accident majeur par an est de 0,0003 et que les accidents majeurs sont indépendants les uns des autres.

  a) Calculer la probabilité qu'il n'y ait pas d'accident pour un réacteur en une année.

  b) Calculer la probabilité qu'il n'y ait pas d'accident pour les 143 réacteurs européens pendant les 30 années à venir.

  c) En déduire la probabilité qu'il y ait un accident majeur en Europe dans les 30 prochaines années.

- **Corrigé :**

  1. Parmi les $450 \times 31 \approx 14000$ années-réacteurs, il y a eu quatre accidents majeurs, soit une probabilité d'accident majeur d'environ 0,0003 par an pour chaque réacteur.

  2. Les auteurs ont alors calculé que, pour les 143 réacteurs européens, on aurait $143 \times 30 \times 0{,}0003 \approx 1{,}29$ accident majeur dans les 30 prochaines années. Ils en déduisent que « la probabilité d'un accident majeur serait de plus de 100% ». Comment une probabilité pourrait-elle dépasser 1 ? Ce serait plus qu'une certitude !!!

  3. En supposant que les accidents majeurs sont indépendants les uns des autres (hypothèse discutable puisque les accidents de Fukushima ne sont pas vraiment indépendants) :
  a) la probabilité qu'il n'y ait pas d'accident pour un réacteur en une année est de $(1-0{,}0003)$,
  b) la probabilité qu'il n'y ait pas d'accident pour les 143 réacteurs européens pendant les 30 années à venir est de $(1-0{,}0003)^{143\times 30} \approx 0{,}28$,
  c) la probabilité qu'il y ait un accident majeur en Europe dans les 30 prochaines années est donc de 0,72, ce qui n'est pas une certitude mais représente cependant un risque important…

Comment bien choisir au hasard ?

- **Niveau :** collège et lycée.

- **Objectifs :**
  – illustrer l'ambiguïté de l'expression « choisir au hasard ».
  – mettre en évidence l'importance de préciser les conditions de l'expérience.

- **Activité :**
  On considère une population constituée de cercles de différents diamètres dessinés sur une feuille de papier : des grands cercles, des cercles moyens et des petits cercles. Le but de l'activité est d'estimer la proportion des trois catégories de cercles à partir du prélèvement d'échantillons de 5 cercles.

  Dans chacun des deux cas, on demande aux élèves de choisir 5 cercles d'une certaine manière, et de calculer les proportions obtenues pour chacune des trois catégories de cercles de leur échantillon. On met en commun les résultats de tous les élèves de la classe afin de les comparer. On peut calculer les trois proportions moyennes obtenues.

  A la fin, le professeur donne les proportions réelles de chaque type de cercles dans la population, ce qui permet de comparer avec les estimations obtenues par les élèves.

  Echantillon empirique : pour prélever cet échantillon, on ne demande aucune méthode particulière, l'élève fait comme il veut : soit en fermant les yeux, soit en choisissant d'après lui au hasard…

  Cette méthode est très mauvaise car le regard des élèves est attiré par les grands cercles qui ont plus tendance à être choisis que les petits, ce qui implique une forte influence sur la proportion observée des grands cercles.

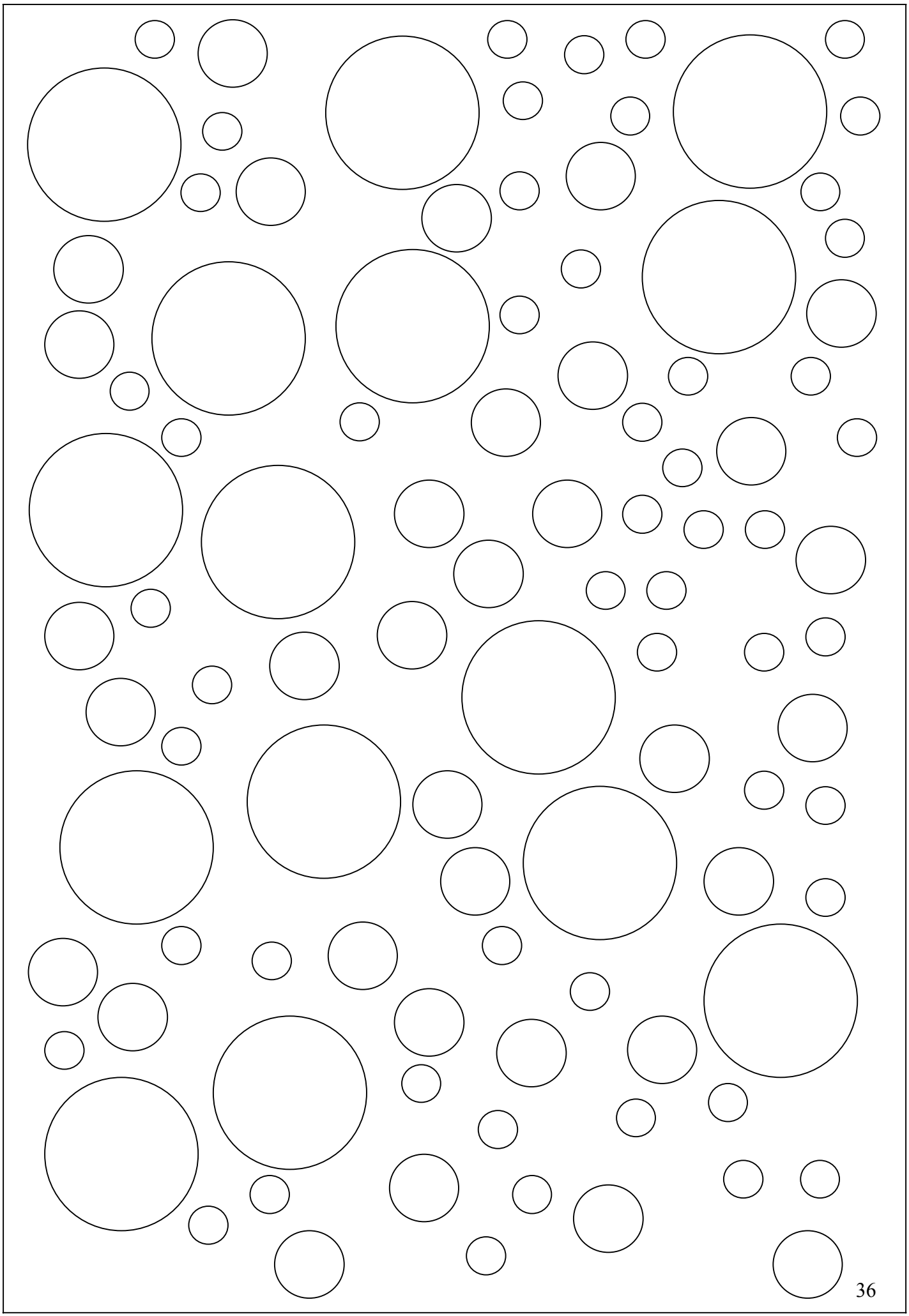
  Cette forme de sondage correspond aux sondages où l'enquêteur choisit les personnes à interroger (comme les « micro-trottoirs ») : il choisira les personnes les plus visibles ou les plus disponibles. Mais il ne rencontrera pas les personnes qui sortent peu (les malades, les personnes âgées, certains handicapés…) ou qui n'ont pas envie de répondre. Elle est évidemment très facile à mettre en œuvre mais on dit qu'elle est « biaisée ».

  Echantillon aléatoire : pour prélever cet échantillon, on commence par numéroter les cercles de 1 à 100. On utilise une table de chiffres au hasard (ou la touche « Rand » d'une calculatrice ou la fonction « Alea » d'un tableur) pour obtenir 5 nombres compris entre 1 et 100 afin de sélectionner les 5 cercles constituant l'échantillon.

  La théorie montre que cette méthode est « sans biais ». En effet, dans un sondage aléatoire simple, chaque individu statistique a une probabilité égale d'être sélectionné pour constituer l'échantillon. Cette méthode nécessite d'avoir une base de sondage complète (c'est-à-dire la liste entière des membres de la population étudiée). Elle a l'avantage d'être basée sur une théorie solide permettant de calculer la précision et la fiabilité des résultats. Elle est facile à appliquer, mais elle peut se révéler très coûteuse.

- **Corrigé :**
  La population est constituée de 15 grands cercles (de 3,2 cm de diamètre), 33 cercles moyens (de 1,4 cm de diamètre) et 52 petits cercles (de 0,8 cm de diamètre), soit des fréquences respectives de 0.15, 0.33 et 0.52.

- **Remarque :**
  Il existe une activité plus complète pour le lycée avec comme objectif supplémentaire de comparer les différentes méthodes d'échantillonnage (échantillon empirique / échantillon aléatoire simple / échantillon aléatoire stratifié / échantillon à deux degrés).
  Cette activité est présentée dans la brochure « Activités de probabilités », Janvier 2008, IREM de Clermont-Ferrand.



La corde de Bertrand

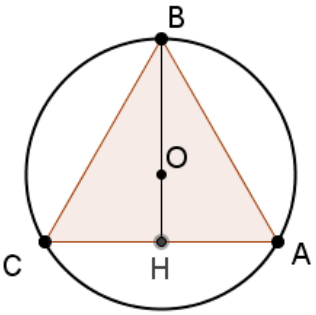
- **Niveau :** terminale et post-bac.

- **Objectifs :** – sensibiliser au fait qu'on ne doit pas choisir le hasard au hasard (importance du choix d'un modèle et de l'univers qui en découle).
  – comprendre des algorithmes et les faire tourner pour obtenir des simulations.
  – travailler sur la loi uniforme.

- **Activité :**
  On choisit au hasard une corde [*MN*] dans un cercle de rayon 1.

  Quelle est la probabilité que sa longueur soit supérieure à celle du côté d'un triangle équilatéral inscrit dans ce cercle ?

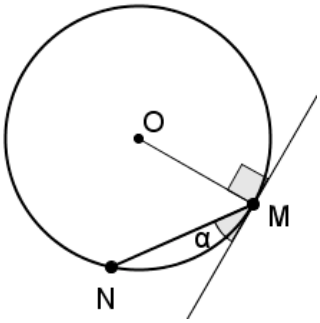
  Questions préliminaires.



a) Montrer que le côté d'un triangle équilatéral inscrit dans un cercle de centre *O* et de rayon 1 mesure $\sqrt{3}$.

b) calculer *OH*.

1. Modèle n°1



Par invariance du cercle par rotation, on peut fixer le point *M* sur le cercle.

Une corde [*MN*] est alors déterminée par la donnée au hasard de l'angle α qu'elle forme avec la tangente en *M* au cercle.

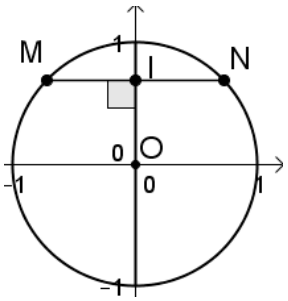
Montrer que $MN = 2 \times \sin\alpha$

Etudier l'algorithme suivant, où N est le nombre d'expériences réalisées, S le nombre de succès :

```
1   Afficher « entrer  N »
2   Saisir N
3   S prend la valeur 0
4   Pour K de 1 à N faire
5       Donner à M la valeur d'un nombre au hasard entre 0 et 180
6       Donner à L la valeur de  2 sin M
7       Si  L >  √3  alors
8           S prend la valeur S+1
9       Fin Si
10  Fin pour
11  Afficher « la fréquence de succès est »,  S/N
```

Que représente la variable M ? La variable L ? Expliquer les lignes *5*, *6*, *7* et *8*.
Programmer cet algorithme sur calculatrice ou ordinateur et faire plusieurs simulations.
Répondre alors au problème en donnant une valeur à la probabilité cherchée.

2. Modèle n°2



Par invariance du cercle par rotation, on peut fixer un diamètre du cercle et on s'intéresse aux cordes perpendiculaires à ce diamètre.

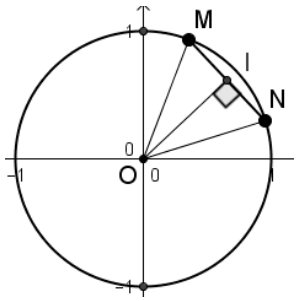
Une corde [*MN*] est alors déterminée par la donnée au hasard d'un point *I* sur ce diamètre.

Etudier l'algorithme suivant, où N est le nombre d'expériences réalisées, S le nombre de succès :

```
1   Afficher « entrer  N »
2   Saisir N
3   S prend la valeur 0
4   Pour K de 1 à N faire
5       Donner à I la valeur d'un nombre au hasard entre -1 et 1
6       Donner à L la valeur de 2√(1 – I²)
7       Si  L > √3 alors
8           S prend la valeur S+1
9       Fin Si
10  Fin pour
11  Afficher « la fréquence de succès est »,  S/N
```

Que représente la variable I ? Expliquer les lignes *5* et *6*.
Programmer cet algorithme sur calculatrice ou ordinateur et faire plusieurs simulations.
Répondre alors au problème en donnant une valeur à la probabilité cherchée.

3. Modèle n°3



Une corde [*MN*] est déterminée par la donnée au hasard d'un point *I* à l'intérieur du cercle, qui sera le milieu de [*MN*].

Etudier l'algorithme suivant, où N est le nombre d'expériences réalisées, S le nombre de succès :

```
1   Afficher « entrer N »
2   Saisir N
3   S prend la valeur 0
4   D prend la valeur 0
5   Tant que D < N
6       Donner à X la valeur d'un nombre au hasard entre -1 et 1
7       Donner à Y la valeur d'un nombre au hasard entre -1 et 1
8       Si  X² + Y² ≤ 1 alors
9           D prend la valeur D+1
10          Donner à L la valeur de 2√(1 – X² – Y²)
11          Si L > √3 alors
12              S prend la valeur S+1
13          Fin Si
14      Fin Si
15  Fin tant que
16  Afficher « la fréquence de succès est »,  S/N
```

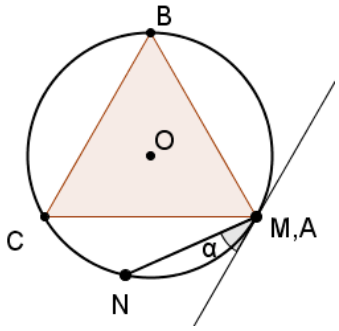
Justifier que la boucle « Tant que » génère des points dans un carré de centre *O* jusqu'à avoir N points dans le disque. Expliquer la ligne *10*.
Programmer cet algorithme sur calculatrice ou ordinateur et faire plusieurs simulations.
Répondre alors au problème en donnant une valeur à la probabilité cherchée.

Remarque : Pour programmer les lignes 6 et 7, on remarquera que si ALEA est un réel au hasard dans l'intervalle $[0\ ;\ 1[$, $2\times \text{ALEA} - 1$ est un réel de l'intervalle $[-1\ ;\ 1[$.

- **Commentaire :**
  L'expression « prendre au hasard une corde dans un cercle » n'a pas d'interprétation unique.
  Pour simuler l'expérience comme pour entamer un calcul de probabilités, **il est nécessaire de choisir des hypothèses de départ**.
  Toute simulation s'appuie sur un modèle et vise à explorer ce modèle.
  Par conséquent, les résultats des trois simulations sont tous corrects ; on a choisi trois modèles différents de lois uniformes.

- **Prolongement :** calcul des probabilités.

Modèle n°1 : Le point N est choisi au hasard en supposant que l'angle α est une variable aléatoire distribuée selon la loi uniforme sur [0 ; 180].

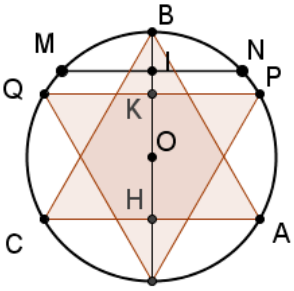


La condition $MN > \sqrt{3}$ est réalisée si la mesure en degrés de l'angle α est comprise entre 60 et 120.

En déduire la probabilité que la longueur d'une corde soit supérieure à celle du côté du triangle équilatéral.

Réponse : la probabilité demandée est $p_1 = \dfrac{120-60}{180} = \dfrac{1}{3}$

Modèle n°2 : Le point I est choisi au hasard en supposant que la variable « milieu de la corde » a une distribution uniforme sur le diamètre du cercle.

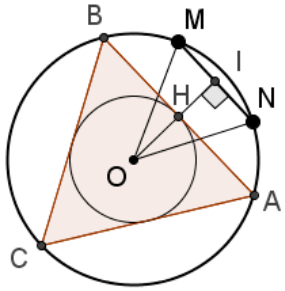


La condition $MN > \sqrt{3}$ est réalisée si le point I est sur le segment $[KH]$.

En déduire la probabilité que la longueur d'une corde soit supérieure à celle du côté du triangle équilatéral.

Réponse : la probabilité demandée est égale au rapport de la longueur $KH = 2OH$ *et de la longueur du diamètre, soit* $p_2 = \dfrac{2\times 0{,}5}{2} = \dfrac{1}{2}$

Modèle n°3 : Le point I est choisi au hasard en supposant que la variable « milieu de la corde » a une distribution uniforme sur la surface intérieure du cercle.



La condition $MN > \sqrt{3}$ est réalisée si $OI < OH$ donc si le point I est à l'intérieur du cercle inscrit dans le triangle équilatéral ABC.

En déduire la probabilité que la longueur d'une corde soit supérieure à celle du côté du triangle équilatéral.

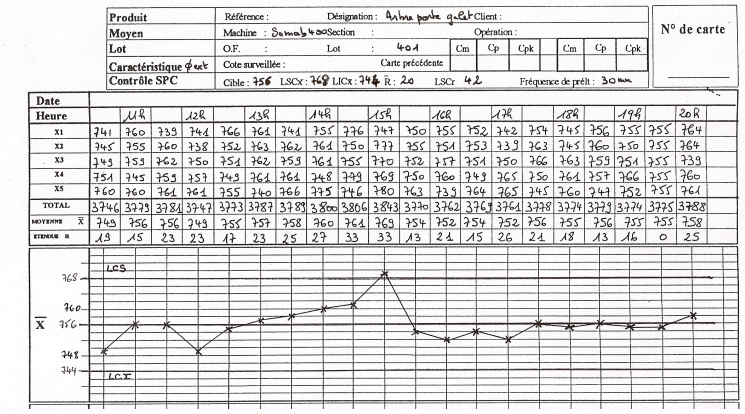
Réponse : la probabilité demandée est égale au rapport de l'aire du cercle inscrit et de l'aire du cercle de rayon 1, soit $p_3 = \dfrac{\pi\times\left(\frac{1}{2}\right)^2}{\pi\times 1^2} = \dfrac{1}{4}$

Les cartes de contrôle

- **Niveau :** seconde.

- **Objectif :** calcul d'une probabilité dans une situation concrète.

- **Activité :**
  Lors de certains contrôles de qualité dans l'industrie, on utilise des cartes de contrôle reposant sur la procédure suivante : la moyenne de la cote surveillée (le diamètre d'une pièce par exemple) est calculée sur des échantillons aléatoires prélevés régulièrement en cours de fabrication. Ces moyennes sont reportées sur la carte de contrôle.
  Si une série de **7** points consécutifs se trouve du même côté de la « moyenne attendue » (la norme visée), le processus de fabrication doit être arrêté pour déceler une éventuelle « dérive ».



  L'explication du choix du nombre 7 se trouve dans la résolution du problème de probabilités suivant : une pièce de monnaie équilibrée est lancée 7 fois, quelle est la probabilité de l'événement A : « la pièce est tombée 7 fois sur pile » ?

  1. Simulation :
  Estimer la valeur de cette probabilité à l'aide de simulations avec une pièce de monnaie ou sur un tableur.

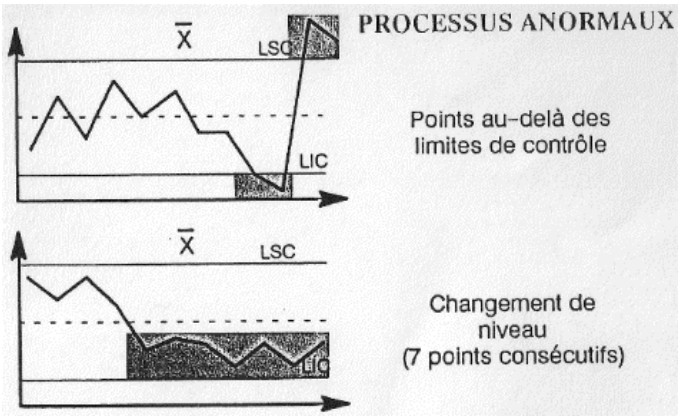
  2. Calcul de la probabilité.

  3. Pourquoi arrête-t-on le processus de fabrication dans une telle situation ?

- **Corrigé :**

  2. La probabilité de l'évènement A est $\left(\frac{1}{2}\right)^7 \approx 0,008$.

  3. On peut considérer qu'une série de 7 points consécutifs se trouvant du même côté de la « moyenne attendue » ne se produit de manière aléatoire que dans 8 cas sur 1000. Cette probabilité étant très faible, on peut penser que cette situation n'est pas due au seul hasard mais peut être la conséquence d'un dérèglement dans le processus de fabrication.



Les anniversaires

- **Niveau :** lycée.
- **Objectif :** calcul d'une probabilité dans un cas où l'intuition est mauvaise conseillère.
- **Activité :**
  Dans une classe, il y a 30 élèves. Marc affirme : « il y a largement plus d'une chance sur deux que 2 élèves de la classe aient le même jour anniversaire (mais pas forcément le même âge) ». Comment savoir si son affirmation est correcte ?

1. **Deux simulations possibles :**

   ➢ **Avec 2 urnes :** – dans la première urne : 12 jetons numérotés de 1 à 12 (pour les mois) ;
   – dans la deuxième urne : 31 jetons numérotés de 1 à 31 (pour les jours).
   Effectuer 30 tirages successifs, avec remise, d'un jeton dans chaque urne. (On élimine les tirages aberrants). Compter le nombre de fois où la même date a été tirée.

   ➢ **Avec une calculatrice :**
   Avec la touche « random » de la calculatrice, produire 30 nombres aléatoires compris entre 1 et 365. Compter le nombre de fois où (au moins) 2 nombres identiques ont été obtenus.

2. **Mise en commun** des résultats de tous les élèves de la classe : compter le nombre de fois où (au moins) 2 nombres identiques ont été obtenus parmi N (avec N : nombre d'élèves de la classe).
   Calculer la fréquence des paires obtenues.

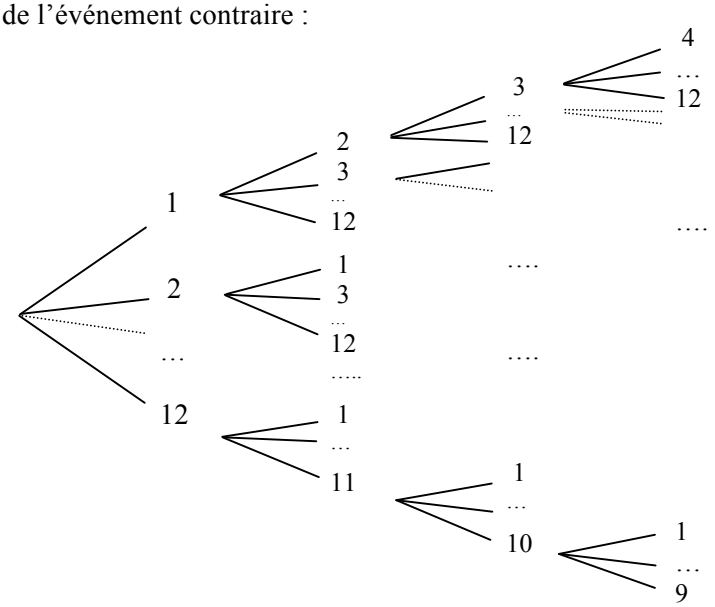
3. **L'affirmation est correcte, démonstration :**

   ➢ On peut commencer par un exemple plus simple :
   Dans un groupe de 4 personnes prises au hasard, quelle est la probabilité qu'au moins deux d'entre elles fêtent leur anniversaire le même mois ?

   On suppose que, pour chaque personne, tous les mois d'anniversaire sont équiprobables. On peut alors assimiler cette expérience à quatre tirages successifs et avec remise d'un jeton dans une urne contenant 12 jetons numérotés de 1 à 12.

   On peut compter le nombre total des issues avec un arbre comportant des pointillés, montrer la difficulté pour compter les cas favorables et alors faire réfléchir à l'intérêt d'énoncer et d'utiliser l'événement contraire.

   Arbre de l'événement contraire :



Nombre d'issues où les 4 personnes ont des mois d'anniversaire différents : $12 \times 11 \times 10 \times 9 = 11\,880$.

Probabilité de l'événement contraire : $\dfrac{12 \times 11 \times 10 \times 9}{12^4} \approx 0{,}57$.

Probabilité cherchée : $1 - 0{,}57 \approx 0{,}43$.

- Retour au problème initial :

  On considère les évènements suivants :

  A : « au moins 2 élèves de la classe ont le même jour anniversaire ».

  $\overline{A}$: « tous les élèves de la classe ont des jours anniversaires différents » (Evènement contraire).

$$P(\overline{A}) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} = \frac{365 \times 364 \times 363 \times ..... \times (365-29)}{(365)^{30}} = \frac{A_{365}^{30}}{(365)^{30}} \approx 0{,}294$$

d'où : $\boxed{P(A) = 1 - P(\overline{A}) \approx 0{,}706}$

- **Remarque :** on peut fournir des listes de classes (pour les élèves qui ne sont toujours pas convaincus !).

- **Calcul avec un algorithme :**
  La touche « Arrangement » n'étant pas utilisée dans les programmes de lycée, le calcul du nombre $\dfrac{365 \times 364 \times 363 \times ..... \times (365-29)}{(365)^{30}}$ est long à taper et souvent impossible sur les calculatrices (dépassement de capacité).

  On peut réécrire ce calcul en vue de la rédaction d'un algorithme :
  $P(\overline{A}) = \dfrac{365}{365} \times \dfrac{365-1}{365} \times \dfrac{365-2}{365} \times ..... \times \dfrac{365-29}{365}$.

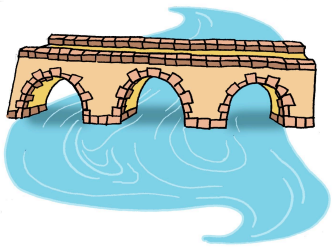
Initialisation :
P prend la valeur 1

Traitement :
Pour K allant de 1 à 29 Faire
P prend la valeur $P \times \dfrac{365-K}{365}$
Fin pour

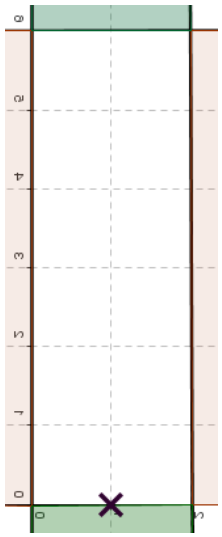
Sortie :
Afficher P

- **Prolongement :**
  On peut demander de modifier cet algorithme pour obtenir directement $P(A)$ avec 30 élèves ou *N* élèves.
  Avec 23 élèves, $P(A) \approx 0{,}507$. Avec 40 élèves, $P(A) \approx 0{,}891$.

La traversée du pont

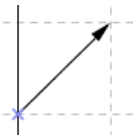


- **Niveau :** lycée.
- **Objectif :** estimation d'une probabilité dans le cas où celle-ci est difficile à calculer.
- **Activité :**
  Mr Bringue, pour rentrer chez lui, doit traverser une passerelle rectiligne de deux mètres de largeur qu'habituellement il traverse en 6 pas en ligne droite. Après une fête bien arrosée, il a une chance sur 2 de faire un pas en avant et tout droit, une chance sur 4 de faire un pas en avant et vers la droite et une chance sur 4 de faire un pas en avant et vers la gauche... Quelles sont ses chances d'arriver au bout sans tomber dans l'eau ?

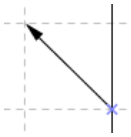


Une chance sur 2...

Une chance sur 4...

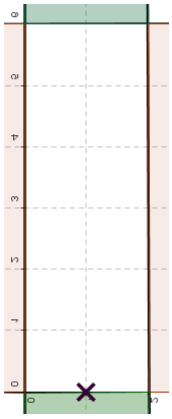


Une chance sur 4...

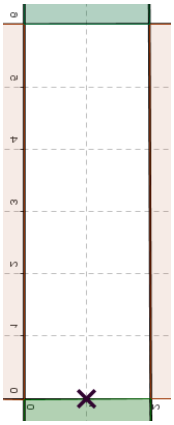


1. Donner une règle permettant de **simuler** sa traversée à l'aide de la calculatrice.
   Tracer son parcours sur le « pont ». Faire trois essais successifs.

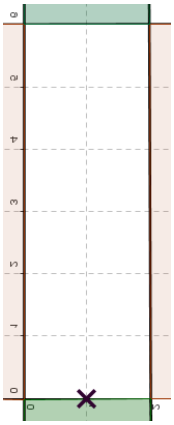
essai 1



essai 2



essai 3



2. **Mise en commun** des résultats de tous les élèves de la classe : compter le nombre de fois où on est arrivé au bout sans tomber dans l'eau parmi les $3 \times N$ (avec N : nombre d'élèves de la classe).
3. En déduire une **estimation de la probabilité** d'arriver au bout sans tomber dans l'eau.

- **Corrigé :**
  Pour la simulation, on peut adopter la règle suivante : produire un nombre aléatoire A entre 0 et 1.
  - si $A \in [0 ; 0{,}25[$, on fait un pas en avant vers la gauche ;
  - si $A \in [0{,}25 ; 0{,}75[$, on fait un pas en avant tout droit ;
  - si $A \in [0{,}75 ; 1[$, on fait un pas en avant vers la droite.

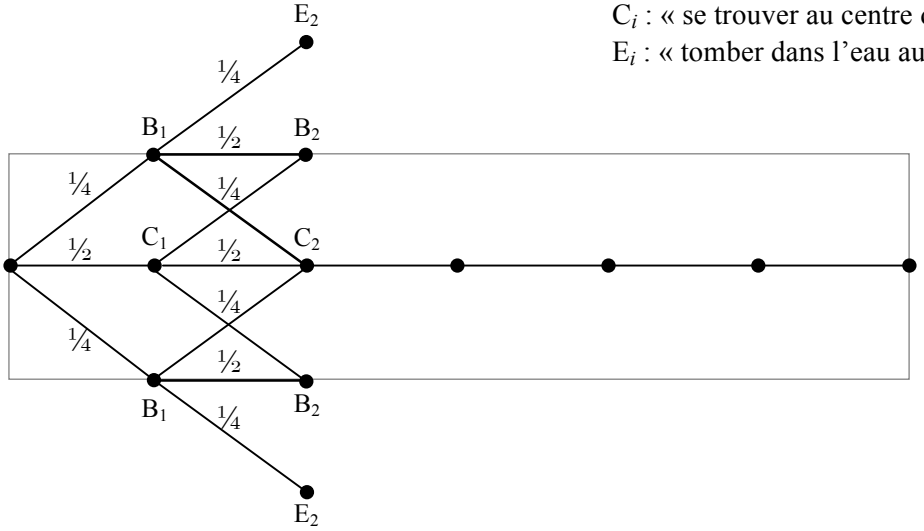
- **Algorithme** de simulation pour 1000 traversées avec Algobox :

I est le nombre de pas de l'ivrogne
P est le type de pas :
-1 pour un pas en avançant à gauche ;
0 pour un pas tout droit ;
+1 pour un pas en avançant à droite.
A est un réel aléatoire entre 0 et 1
N est le nombre d'expériences
S est le nombre de succès :
l'ivrogne n'est pas tombé dans l'eau et a traversé le pont.

```
VARIABLES
  S EST_DU_TYPE NOMBRE
  N EST_DU_TYPE NOMBRE
  I EST_DU_TYPE NOMBRE
  P EST_DU_TYPE NOMBRE
  A EST_DU_TYPE NOMBRE
DEBUT_ALGORITHME
  S PREND_LA_VALEUR 0
  POUR N ALLANT_DE 1 A 1000
    DEBUT_POUR
    I PREND_LA_VALEUR 0
    P PREND_LA_VALEUR 0
    TANT_QUE (I<6 ET P>=-1 ET P<=1) FAIRE
      DEBUT_TANT_QUE
      I PREND_LA_VALEUR I+1
      A PREND_LA_VALEUR random()
      SI (A<0.25) ALORS
        DEBUT_SI
        P PREND_LA_VALEUR P-1
        FIN_SI
      SI (A>=0.75) ALORS
        DEBUT_SI
        P PREND_LA_VALEUR P+1
        FIN_SI
      FIN_TANT_QUE
    SI (P>=-1 ET P<=1) ALORS
      DEBUT_SI
      S PREND_LA_VALEUR S+1
      FIN_SI
    FIN_POUR
  AFFICHER S
FIN_ALGORITHME
```

- **Remarque :** le calcul de la probabilité de traverser la passerelle est faisable mais difficile…

On définit les évènements suivants : B_i : « se trouver sur un bord de la passerelle au $i^{\text{ème}}$ pas » ;
C_i : « se trouver au centre de la passerelle au $i^{\text{ème}}$ pas » ;
E_i : « tomber dans l'eau au $i^{\text{ème}}$ pas ».



Il y a symétrie de l'arbre, on peut donc considérer que les deux bords gauche et droit jouent le même rôle.

Numéro du pas	1	2	3	4	5	6
Probabilité d'être sur la ligne centrale	$\frac{1}{2}$	$P(C_1)\times\frac{1}{2}+P(B_1)\times\frac{1}{4}$ $=\frac{1}{2}\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{4}=\frac{3}{8}$	$P(C_2)\times\frac{1}{2}+$ $P(B_2)\times\frac{1}{4}=\frac{5}{16}$	$P(C_3)\times\frac{1}{2}+$ $P(B_3)\times\frac{1}{4}=\frac{17}{64}$	$P(C_4)\times\frac{1}{2}+$ $P(B_4)\times\frac{1}{4}=\frac{29}{128}$	$P(C_5)\times\frac{1}{2}+$ $P(B_5)\times\frac{1}{4}=\frac{99}{512}$
Probabilité d'être sur un bord	$\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$	$P(C_1)\times\frac{1}{2}+P(B_1)\times\frac{1}{2}$ $=\frac{1}{2}\times\frac{1}{2}+\frac{1}{2}\times\frac{1}{2}=\frac{1}{2}$	$P(C_2)\times\frac{1}{2}+$ $P(B_2)\times\frac{1}{2}=\frac{7}{16}$	$P(C_3)\times\frac{1}{2}+$ $P(B_3)\times\frac{1}{2}=\frac{12}{32}$	$P(C_4)\times\frac{1}{2}+$ $P(B_4)\times\frac{1}{2}=\frac{41}{128}$	$P(C_5)\times\frac{1}{2}+$ $P(B_5)\times\frac{1}{2}=\frac{70}{256}$
Probabilité d'être tombé dans l'eau	0	$P(B_1)\times\frac{1}{4}$ $=\frac{1}{2}\times\frac{1}{4}=\frac{1}{8}$	$P(B_2)\times\frac{1}{4}$ $=\frac{1}{8}$	$P(B_3)\times\frac{1}{4}$ $=\frac{7}{64}$	$P(B_4)\times\frac{1}{4}$ $=\frac{3}{32}$	$P(B_5)\times\frac{1}{4}$ $=\frac{41}{512}$

$$P(\text{tomber dans l'eau}) = P(E_2 \cup E_3 \cup E_4 \cup E_5 \cup E_6) = \frac{1}{8}+\frac{1}{8}+\frac{7}{64}+\frac{3}{32}+\frac{41}{512} = \frac{273}{512} \approx 0{,}53$$

$$P(\text{traverser la passerelle}) = P(C_6 \cup B_6) = \frac{99}{512}+\frac{70}{256} = \frac{319}{512} \approx 0{,}47$$

Sondage détourné :
éviter les réponses biaisées grâce au hasard

- **Niveau :** terminale et post-bac.

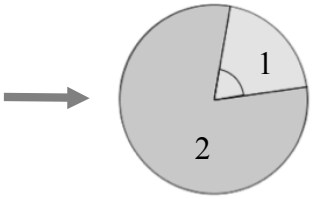
- **Objectifs :**
  – étude d'une technique qui permet d'éviter les réponses biaisées grâce au hasard.
  – modélisation du problème avec un arbre pondéré peu évident.

- **Activité :**
  Si on pose la question « Avez-vous déjà volé dans un supermarché ? » aux clients, on peut supposer que les réponses seront biaisées !! En 1965, le statisticien américain Warner propose une méthode pour obtenir un résultat non biaisé à cette enquête (Randomized Method) :

  ➢ L'enquêteur donne à chaque client interrogé la carte suivante :

  1 : j'ai déjà volé dans un supermarché

  2 : je n'ai encore jamais volé dans un supermarché

  ➢ Pour éviter les fausses déclarations, l'enquêteur préserve l'anonymat des réponses en proposant de réaliser **en secret** la procédure suivante :
  – Faire tourner la roulette, attendre son arrêt puis, selon le numéro indiqué, lire la phrase correspondante sur le carton.
  – Donner la réponse « *Vrai* » ou « *Faux* » à cette phrase.

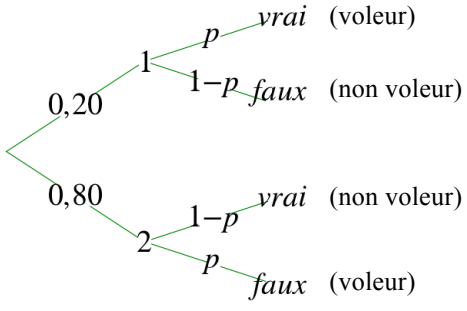
  ➢ A la fin de la journée, l'enquêteur possède un grand nombre de réponses : il établit que 60% des clients interrogés ont donné la réponse « *Vrai* ».

  En construisant un arbre pondéré modélisant ce problème, montrer comment le calcul va permettre à l'enquêteur de déterminer la fréquence réelle p de vols dans les supermarchés.
  On suppose ici que : – l'angle au centre du secteur **1** mesure 72°.
  – étant donné le grand nombre de réponses, on assimile fréquence observée et probabilité.

- **Corrigé :**
  P(« la réponse donnée est ''*Vrai*'' ») = $0{,}20 \times p + 0{,}80 \times (1-p)$

  Cette probabilité est estimée à 0,60 d'après l'étude statistique.

  En résolvant l'équation $0{,}60 = 0{,}20 \times p + 0{,}80 \times (1-p)$, on obtient : $p = \frac{1}{3}$.

- **Commentaires :**
  Cet exercice montre comment « le hasard » peut être un allié du statisticien, en permettant d'éviter certains biais.
  La méthode exposée n'est pas utilisée pour les sondages politiques. Cependant, elle est utilisée aux Etats-Unis pour les enquêtes sur des sujets touchant la vie privée ou sur des sujets sensibles (avortement, usage de stupéfiants, sexualité...).

Statistiques inférentielles

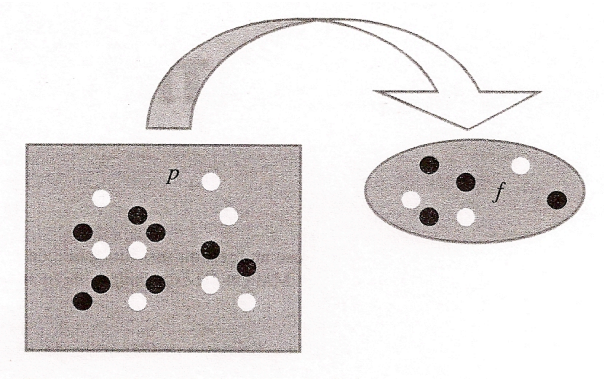
Les statistiques inférentielles sont avant tout un mode de pensée qui permet de passer :

du connu : données fournies par **des statistiques** sur un ensemble nécessairement restreint d'individus,
à l'inconnu : résultats applicables à l'ensemble des individus concernés,
par un raisonnement logique, **le raisonnement inférentiel**, reposant sur des bases théoriques : **le calcul des probabilités.**

Trois thèmes sont étudiés au lycée : fluctuation des échantillons (programme de seconde), prise de décision (test d'hypothèse, programme de première) et estimation (intervalle de confiance ou « fourchette de sondage », programme de terminale)

- **Illustration** : *(Extrait de « l'induction statistique au lycée », Philippe DUTARTE)*
  Considérons une urne « de Bernoulli » contenant une proportion p de boules blanches, dont on extrait (tirage avec remise) n boules, la proportion de boules blanches dans le tirage étant notée f.

n boules avec remise



➢ **Si p est connu**, on peut dire : dans plus de 95% des tirages, $f \in \left[p - \frac{1}{\sqrt{n}} \ ; \ p + \frac{1}{\sqrt{n}} \right]$.

(intervalle de fluctuation de 95% des échantillons)

➢ **Si p est inconnu** et on a des raisons de penser qu'il **vaut peut-être** $\hat{p}$: on procède alors à un tirage qui donne une valeur de f et l'on peut dire :

- si $f \notin \left[\hat{p} - \frac{1}{\sqrt{n}} \ ; \ \hat{p} + \frac{1}{\sqrt{n}} \right]$, on rejette l'hypothèse $p = \hat{p}$ avec un risque de 5%.
- si $f \in \left[\hat{p} - \frac{1}{\sqrt{n}} \ ; \ \hat{p} + \frac{1}{\sqrt{n}} \right]$, on accepte l'hypothèse $p = \hat{p}$ avec un risque inconnu.

(test de validité d'hypothèse)

➢ **Si p est inconnu** et on souhaite l'estimer : on procède à un tirage qui donne une valeur de f et on affirmera à juste titre que dans plus de 95% des tirages, $p \in \left[f - \frac{1}{\sqrt{n}} \ ; \ f + \frac{1}{\sqrt{n}} \right]$.

(fourchette de sondage ou intervalle de confiance)

Intervalle de fluctuation

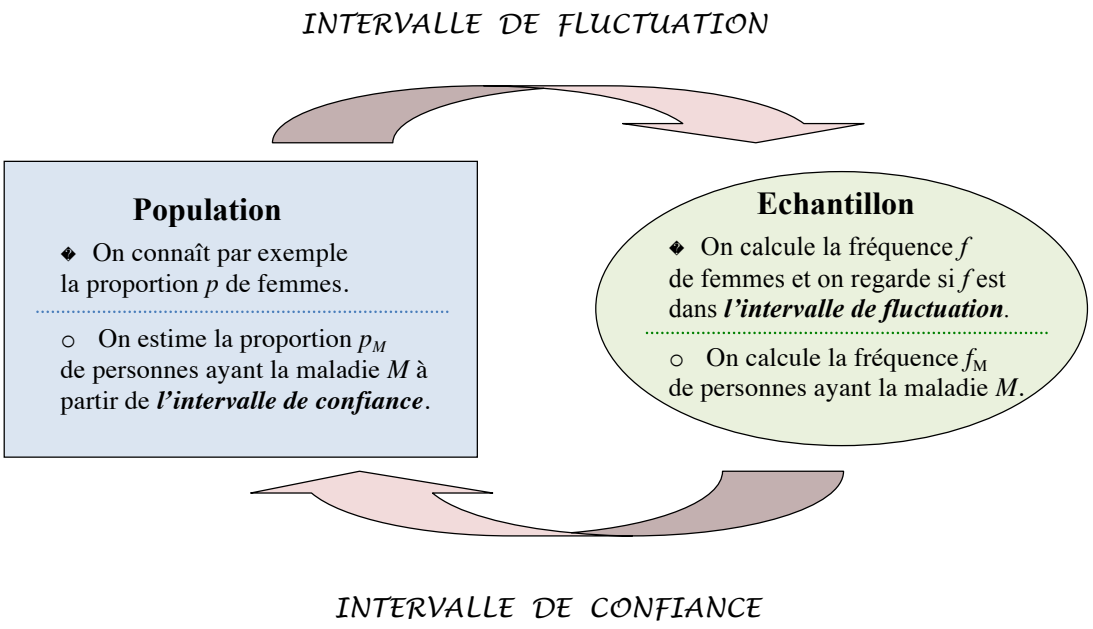
Que disent les programmes ?

Soit X_n une v.a. suivant la loi binomiale $\mathcal{B}(n\ ;\ p)$ et α un réel dans]0 ; 1[. Un intervalle de fluctuation de la variable fréquence $F_n = \dfrac{X_n}{n}$, au niveau $1 - \alpha$, est un intervalle $[f_1\ ;f_2]$ tel que $P(F_n \in [f_1\ ;f_2]) \geq 1 - \alpha$.

- Dans le programme de **seconde**, on a défini un intervalle de fluctuation approché au niveau 0,95, de F_n par $\left[p - \dfrac{1}{\sqrt{n}}\ ;\ p + \dfrac{1}{\sqrt{n}}\right]$, valable sous certaines conditions de n et de p.

- En **première**, on a déterminé l'intervalle de fluctuation $\left[\dfrac{a}{n};\dfrac{b}{n}\right]$, calculé à partir de la loi binomiale, a étant le plus petit entier tel que $P(X_n \leq a) > 0{,}025$ et b le plus petit entier tel que $P(X_n \leq b) \geq 0{,}975$.
  Cet intervalle convient pour toutes les valeurs de n et de p.

- En **terminale**, le théorème de Moivre-Laplace permet de donner un intervalle de fluctuation plus facilement calculable qu'en première, sous réserve que n soit assez grand (mais valable pour toute valeur de p).
  L'intervalle de fluctuation asymptotique au niveau $1 - \alpha$ est défini par $\left[p - u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$,
  avec u_α l'unique réel tel que $P(-u_\alpha \leq Z \leq u_\alpha) = 1 - \alpha$ où Z est une v.a. suivant la loi normale $\mathcal{N}(0\ ;\ 1)$.
  Cet intervalle contient $F_n = \dfrac{X_n}{n}$ avec une probabilité d'autant plus proche de $1 - \alpha$ que n est grand.
  On le qualifie d'intervalle de fluctuation asymptotique car il est obtenu grâce à une convergence. On convient d'utiliser cette approximation si $n \geq 30$, $np \geq 5$, $n(1 - p) \geq 5$.



Le biberon

- **Niveau :** lycée.
- **Objectif :** à partir d'une activité découpée en trois parties, introduire trois notions différentes : fluctuation d'échantillonnage en seconde, prise de décision en première et estimation d'une proportion en terminale.
- **Activité :** on part d'une situation concrète rencontrée en lycée agricole : l'étude de la répartition de graines dans une semence de gazon.
  Pour effectuer les tirages, on utilise des bouteilles opaques de lait, dans lesquelles on découpe le bouchon pour introduire une tétine transparente (d'où notre appellation de « biberon »). A l'intérieur de la bouteille, on introduit des boules de cotillon (moins bruyantes que des billes) ; ainsi le contenu de la bouteille demeure caché pendant toute l'activité.

  Le document est à distribuer aux élèves, les parties du texte en italique et les annexes étant destinées à l'usage du professeur.

  Cette activité est parue dans le bulletin vert n°500 de l'APMEP.

Activité 1 : Découvrir la fluctuation d'échantillonnage.

On considère une population constituée de 30 % de boules rouges et de 70 % de boules blanches.
Le nombre de boules rouges obtenu dans un échantillon de taille n prélevé dans la population est variable. C'est une variable aléatoire X_n. L'objectif est de comprendre et d'illustrer la variabilité de la fréquence $F_n = \frac{X_n}{n}$ d'apparition des boules rouges dans l'échantillon.

1. Expérience et recueil des données.

Chaque élève reçoit une bouteille opaque munie d'une tétine et l'enseignant donne la proportion de boules rouges à l'intérieur, soit 30%. Le bouchon est percé d'un trou suffisamment large pour laisser descendre une boule dans la tétine si on retourne la bouteille.

- On va prélever des échantillons aléatoires de taille n dans la population.

Pour cela, on retourne n fois la bouteille, sans oublier après chaque « retournement » de remettre la boule dans la bouteille et de bien secouer.
On rappelle qu'un échantillon de taille n est constitué des résultats de n répétitions indépendantes de la même expérience, l'indépendance étant assurée dans cette expérience par la remise de la boule et l'agitation de la bouteille.
On note 1 à chaque fois que la boule est rouge et 0 sinon.
Réaliser 20 tirages de boules pour générer un échantillon A de taille 20 et noter les résultats (0 ou 1) sur une fiche individuelle. De la même manière, noter les résultats d'un échantillon B obtenu avec 50 tirages.
Calculer ensuite le cumul des 1 pour chacun des deux échantillons. On a ainsi le nombre X_n de boules rouges obtenu dans chaque échantillon.
Calculer (mentalement) les fréquences de boules rouges obtenues dans les deux échantillons.

- Pour obtenir des échantillons de grande taille, on va utiliser un tableur d'ordinateur.

En tapant = **ENT(ALEA() + 0,30)**, on obtient 1 avec une probabilité de 30 % et 0 avec une probabilité de 70 %.
Avec le tableur de géogébra, on tape **floor(random() + 0.30)**)
Simuler sur un tableur des échantillons C et D de tailles respectives 100 et 500. Pour chacun de ces deux échantillons, compter le nombre de boules rouges obtenu et calculer la fréquence correspondante.

- Remplir la fiche récapitulative de la classe pour les quatre échantillons.

L'enseignant peut ensuite en donner une photocopie à chaque élève.

Fiche récapitulative des fréquences de boules rouges

Nom Prénom	Echantillon A $n = 20$	Echantillon B $n = 50$	Echantillon C $n = 100$	Echantillon D $n = 500$
HANI Hélène	0,6	0,28	0,25	0,27
Sonia Bouzidi	0,5	0,22	0,38	0,322
THUAIRE M CHARBONNEAU	0,3	0,24	0,26	0,284
AMRAUI Idriss/Mohamed	0,3	0,38	0,32	0,288
FARTARIA - CORTIER	0,35	0,36	0,34	0,294
Boutard - GOI	0,25	0,38	0,41	0,338
Ouvry - Bourdier	0,45	0,36	0,41	0,282
Bealon-Vessaire	0,45	0,32	0,25	0,292
Boucher Jérôme	0,4	0,36	0,27	0,334
Dargaux / Mevieu	0,15	0,38	0,35	0,326
Telbani / Lebeau	0,55	0,4	0,29	0,308
Rubaud / Raimbault	0,55	0,40	0,24	0,298
Dorte / Rouxeise	0,2	0,4	0,3	0,302
Cottrez / Petitjean	0,2	0,22	0,28	0,334
Vaddivia / Lamadon	0,4	0,38	0,31	0,328
Lacombe / Martin	0,2	0,36	0,36	0,316
Touni / Malondo	0,45	0,42	0,3	0,296

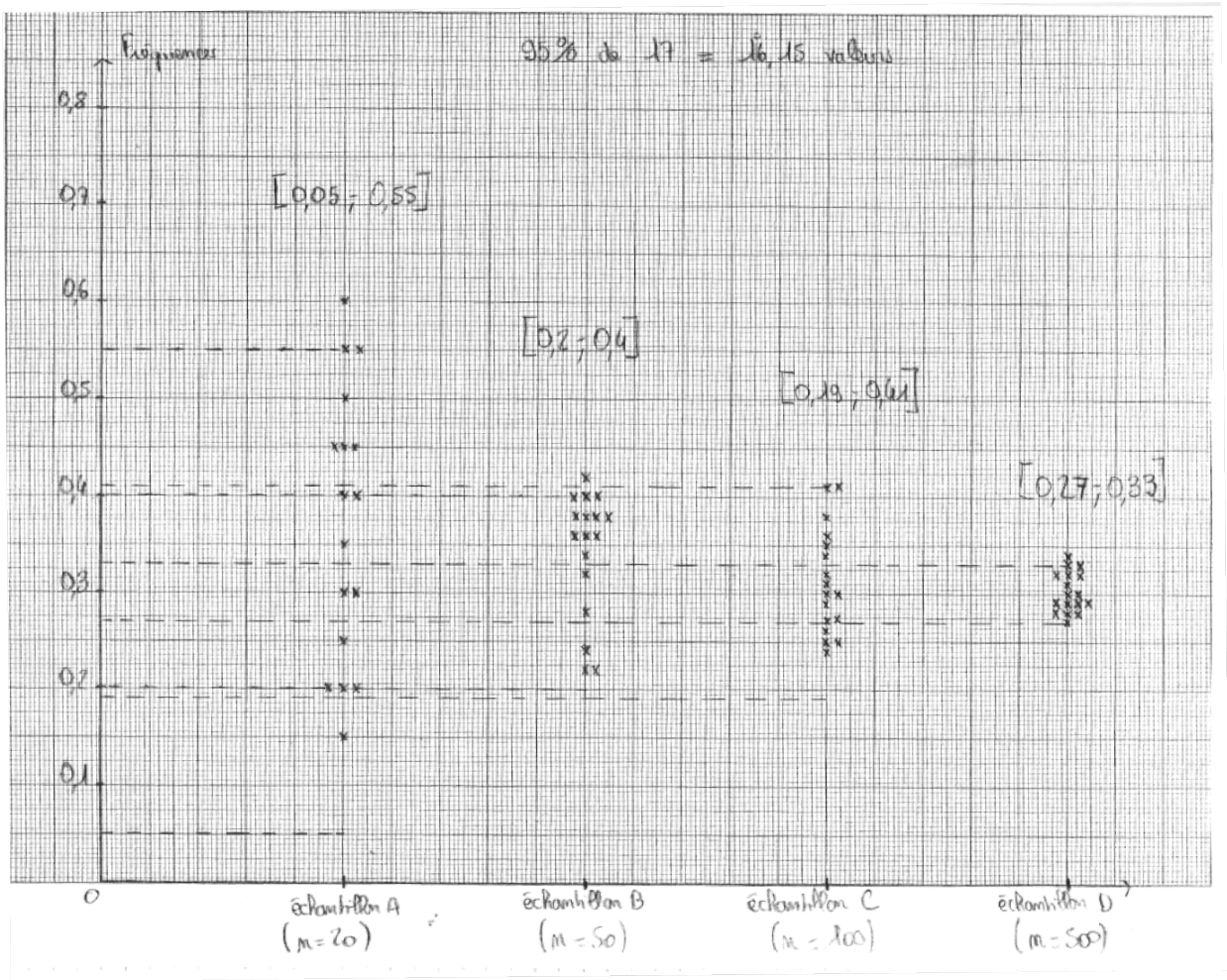
Annexe 1

2. Description des séries statistiques.

On va illustrer les fréquences $F_n = \frac{X_n}{n}$ des boules rouges par des graphiques statistiques.

- Construire sur papier millimétré les quatre nuages de points, associés aux quatre séries.

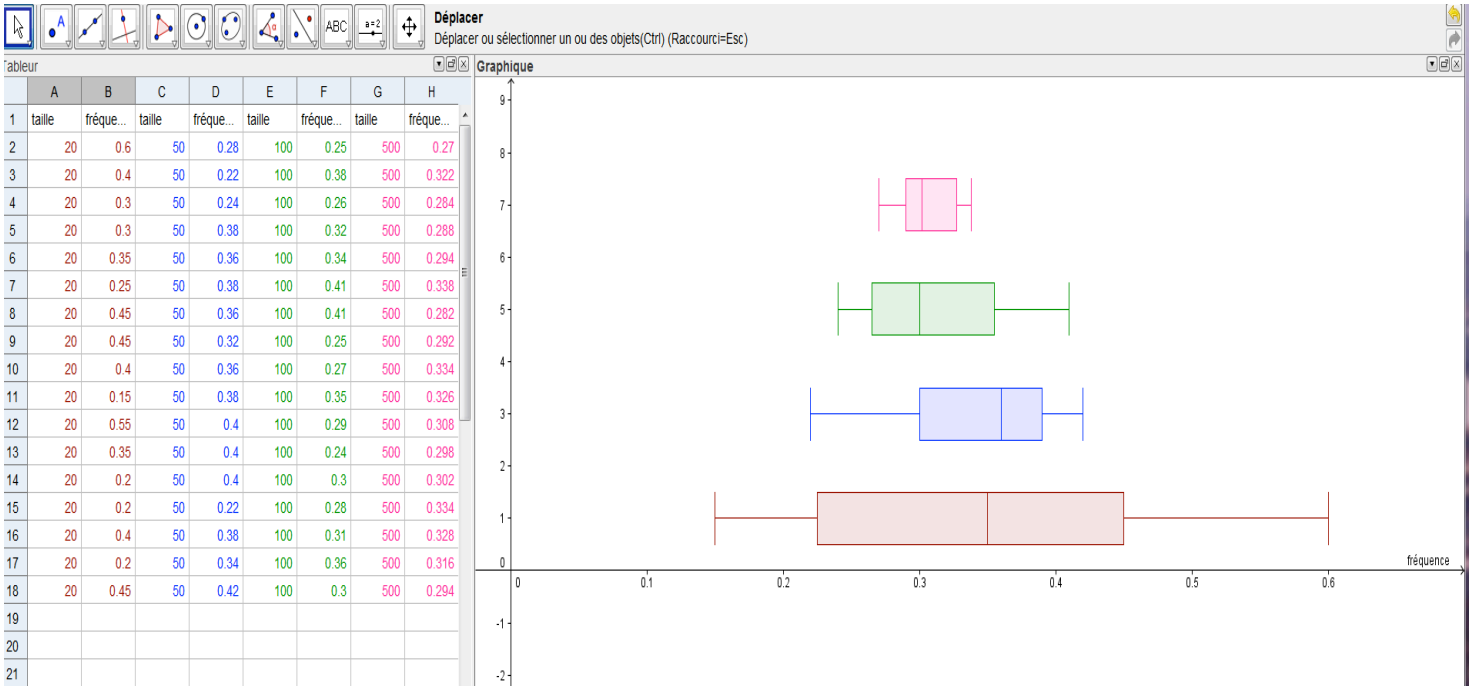
Pour les points de même ordonnée, on demande aux élèves de les distinguer en les plaçant côte à côte.



Annexe 2

- A l'aide du tableur du logiciel Géogébra (à cocher dans l'option **Affichage**), remplir les cellules de la première colonne avec les fréquences relevées pour les échantillons A. Procéder de même dans les colonnes suivantes pour les échantillons B, C et D.

Construire sur un même graphique les quatre diagrammes en boîte associés aux quatre séries. Dans **Saisir**, écrire **BoiteMoustaches** et indiquer comme syntaxe dans les crochets : valeur de l'ordonnée pour l'axe de la boîte, demi-hauteur de la boîte, plage des données statistiques.



	A	B	C	D	E	F	G	H
1	taille	fréque...	taille	fréque...	taille	fréque...	taille	fréque...
2	20	0.6	50	0.28	100	0.25	500	0.27
3	20	0.4	50	0.22	100	0.38	500	0.322
4	20	0.3	50	0.24	100	0.26	500	0.284
5	20	0.3	50	0.38	100	0.32	500	0.288
6	20	0.35	50	0.36	100	0.34	500	0.294
7	20	0.25	50	0.38	100	0.41	500	0.338
8	20	0.45	50	0.36	100	0.41	500	0.282
9	20	0.45	50	0.32	100	0.25	500	0.292
10	20	0.4	50	0.36	100	0.27	500	0.334
11	20	0.15	50	0.38	100	0.35	500	0.326
12	20	0.55	50	0.4	100	0.29	500	0.308
13	20	0.35	50	0.4	100	0.24	500	0.298
14	20	0.2	50	0.4	100	0.3	500	0.302
15	20	0.2	50	0.22	100	0.28	500	0.334
16	20	0.4	50	0.38	100	0.31	500	0.328
17	20	0.2	50	0.34	100	0.36	500	0.316
18	20	0.45	50	0.42	100	0.3	500	0.294
19								
20								
21								

Annexe 3 : ***BoiteMoustaches [1, 0.5, B2:B18]*** *pour l'échantillon A.*

Le tableur peut donner des paramètres statistiques pour les quatre séries : moyenne, médiane, quartiles…
Le graphique en nuage de points pourrait aussi être obtenu avec géogébra mais en présentant l'inconvénient que les points de même ordonnée se superposent.

3. Analyse et comparaison des quatre distributions observées de la variable $F_n = \frac{X_n}{n}$.

Afin de comparer plusieurs échantillons, la juxtaposition judicieuse des graphiques permet d'observer la dispersion et la position des quatre séries et d'émettre des hypothèses sur les facteurs qui en sont à l'origine.

Que peut-on déduire des illustrations graphiques que l'on vient de faire ?
Quel est l'effet de la taille d'échantillon ?
Tracer sur chacun des nuages de points le plus petit intervalle centré sur 0,30 contenant au moins 95 % des fréquences observées.

En classe de seconde, le programme définit l'intervalle de fluctuation d'une proportion p d'un caractère étudié dans la population comme étant le plus petit intervalle centré autour de p où se situe, avec une probabilité au moins égale à 0,95, la fréquence observée f dans un échantillon de taille n. Pour des proportions p du caractère comprises entre 0,2 et 0,8, et pour des échantillons de taille $n \geq 25$, f appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}} \ ; \ p + \frac{1}{\sqrt{n}}\right]$ avec une probabilité d'au moins 0,95.

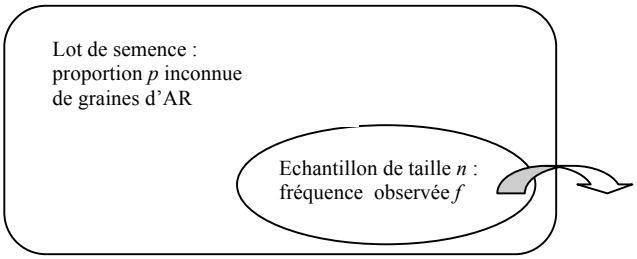
Avec la formule précédente, déterminer l'intervalle de fluctuation au niveau 95 % de la proportion de boules rouges dans un échantillon de taille 50, 100 ou 500.
Comparer avec les intervalles tracés précédemment.

Dans cette activité, la fréquence f observée est en dehors de l'intervalle de fluctuation dans environ 5 % des cas. Autrement dit, dans 95 % des cas, on n'observe pas de différence significative entre p et les valeurs de $F_n = \frac{X_n}{n}$.

Activité 2 : Prendre une décision dans un contrôle de qualité

Dans le service technique d'une commune, vous êtes chargé de réceptionner la commande d'un lot de semence (par exemple une tonne) pour engazonner les espaces verts. Le cahier des charges de la commande précise que **le mélange doit contenir 30% de graines d'Agropyrum Repens (chiendent, noté AR)**. Cette plante n'offre pas une couverture très esthétique mais elle est bien adaptée à la sécheresse et au piétinement.
La procédure de contrôle statistique, appelé plan de contrôle, fixe :

- la procédure d'échantillonnage en indiquant **la taille *n* de l'échantillon aléatoire** de graines à prélever dans le lot et les modalités de prélèvement. On compte le nombre de graines d'AR dans cet échantillon.
- **le critère de rejet du lot** : c'est le nombre de graines d'AR à partir ou en dessous duquel le lot doit être rejeté comme non conforme à la commande.



Deux erreurs peuvent survenir dans la prise de décision : on peut se tromper en rejetant un lot conforme (risque fournisseur) ou en acceptant un lot non conforme (risque client). Ces risques proviennent de la variabilité des résultats issus d'une procédure d'échantillonnage : en échantillonnant deux fois de la même façon, on peut obtenir des résultats différents (voir programme de seconde).

1. Protocole expérimental et recueil des données.

Comme un contrôle réel est impossible en classe, un lot de semence sera représenté par une bouteille opaque munie d'une tétine et contenant chacune des boules rouges et des boules blanches, en proportion inconnue.
A chaque tirage, on notera 1 si la couleur de la « graine » obtenue est rouge (c'est une graine d'AR) et 0 sinon.

Le professeur aura préparé par avance des biberons contenant par exemple 20 boules, avec des proportions de boules rouges assez variées. Chaque élève expérimente avec une bouteille numérotée, assimilée au lot de semence et dont la proportion de boules rouges lui est inconnue.
Pour un contrôle réel, on prélèverait n graines au hasard mais sans remise ; l'échantillon serait pourtant considéré comme aléatoire car obtenu parmi un nombre considérable de graines.

- Réaliser 20 tirages de graines pour générer un échantillon A de taille 20 et noter les résultats (0 ou 1) sur une fiche individuelle. De la même manière, noter les résultats d'un échantillon B obtenu avec 50 tirages. Calculer ensuite le nombre X de graines d'AR obtenues dans chacun des deux échantillons.

- Remplir la fiche récapitulative de la classe pour les deux échantillons.

L'enseignant peut ensuite en donner une photocopie à chaque élève.

2. Retour sur le contrôle de qualité.

Vous devez prendre une décision concernant l'acceptation ou le refus du lot de semence livré.
Qu'est ce qu'un lot conforme au cahier des charges ?
Quels lots vous semblent non conformes à la commande au vu des échantillons A ? Des échantillons B ?

Les enseignants qui ont testé cette activité dans les journées de formation ont eu des réactions variées : certains ont eu tendance à rejeter beaucoup de lots, d'autres pas. Ils sont alors tombés d'accord sur la nécessité d'une règle de décision.

3. Construction de l'intervalle de fluctuation de la fréquence des graines d'AR.

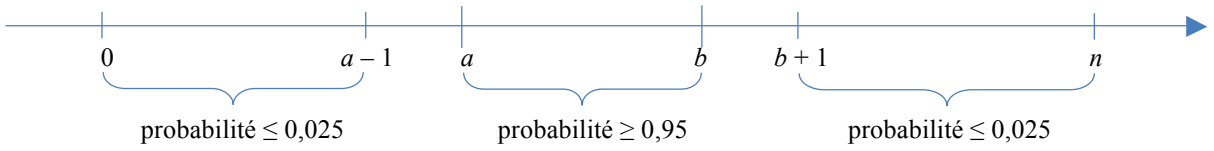
Prenons comme **hypothèse** que **la proportion *p* est de 30%** car c'est celle qui est prévue par le cahier des charges. On rejettera cette hypothèse lorsqu'on observera une différence significative entre la valeur supposée de p et la valeur observée f (donc si f est trop éloignée de p).

On rappelle que X est la variable aléatoire mesurant le nombre de graines d'AR observées dans un échantillon de taille n. Quelle est la loi suivie par X ? Quels sont ses paramètres n et p ?

En classe de première, le programme définit l'intervalle de fluctuation à 95% d'une fréquence, associé à une variable aléatoire X suivant une loi binomiale de paramètres n et p, comme étant l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$ où

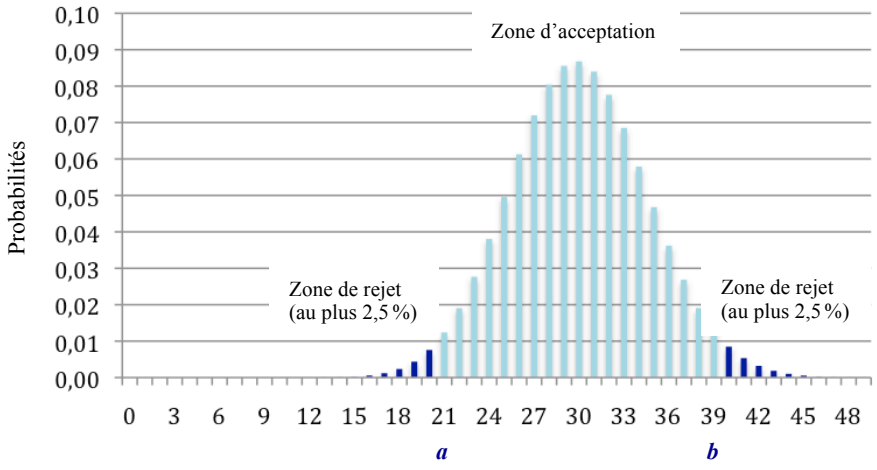
- a est le plus petit des entiers tel que $P(X \leq a) > 0{,}025$
- b est le plus petit des entiers tel que $P(X \leq b) \geq 0{,}975$

Le programme impose un seuil de risque de 5% dans la construction de l'intervalle de fluctuation : la probabilité de rejeter l'hypothèse, alors qu'elle est vraie, est au maximum de 5%.
On cherche à partager l'intervalle [0, n], où X prend ses valeurs, en trois intervalles [0 ; a – 1], [a ; b] et [b + 1 ; n] de sorte que X prenne ses valeurs dans chacun des intervalles extrêmes avec une probabilité cumulée d'au plus 0,025, ce qui est réalisé si $P(X \leq a - 1) \leq 0{,}025$ et $P(X \geq b + 1) \leq 0{,}025$.



Comme la loi de X est discrète, on peut remarquer que :
$P(X \leq a - 1) \leq 0{,}025 \Leftrightarrow P(X < a) \leq 0{,}025$ et $P(X \geq b + 1) \leq 0{,}025 \Leftrightarrow P(X > b) \leq 0{,}025$.

Exemple avec la loi binomiale $\mathcal{B}(100 ; 0{,}3)$:



Avec un tableur (ordinateur), on peut obtenir les entiers a et b : on procède de la manière suivante, décrite pour le cas $n = 50$ et $p = 0{,}30$:

- Dans la colonne A, on écrit les entiers de 0 à 50.
- Avec la formule **= LOI.BINOMIALE (A1; 50; 0,30; VRAI)** dans la cellule B1, on calcule $P(X \leq$ A1).
  Remarque : sur « Openofficecalc », il faut écrire **1** à la place de VRAI.
- On recopie cette formule en colonne B.

Avec une calculatrice, on utilise l'éditeur de fonctions, en entrant (en Y_1 par exemple) la formule :
BinomialCD(X,50,0.30) pour Casio ou **binomFRép(50,0.30,X)** pour Texas.
On obtient les probabilités cumulées dans la table de valeurs.

Compléter le tableau suivant :

Taille d'échantillon	Valeur de a	Valeur de b	Intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$
$n = 20$			
$n = 50$			

Pour n assez grand et p ni trop petit ni trop grand, on observe que l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$ *est sensiblement le même que l'intervalle de fluctuation* $\left[p-\frac{1}{\sqrt{n}}\ ;\ p+\frac{1}{\sqrt{n}}\right]$ *étudié en seconde.*

L'intérêt de l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$*, calculé à partir de la loi binomiale, est de convenir pour toutes les valeurs des paramètres n et p (en particulier pour de petits échantillons).*

4. Prise de décision.

La règle de décision est la suivante :

si la fréquence observée *f* de graines d'AR appartient à l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$**, on accepte l'hypothèse selon laquelle le lot de semence comporte 30 % d'AR ; sinon, on rejette cette hypothèse en sachant qu'on a 5 % de risque de faire une erreur.**

Quels sont les lots qui sont refusés ?

On dévoile la composition des bouteilles numérotées. Certaines ont une proportion de boules rouges égale à 30%, d'autres une proportion de boules rouges inférieure ou supérieure à 30%.

On pourra alors noter les cas où on a pris une mauvaise décision et faire sentir aux élèves qu'il peut y avoir deux types d'erreur :
on a rejeté un lot conforme (probabilité de 5%): c'est dommage pour le fournisseur.
on a accepté un lot non conforme (probabilité non connue) : c'est dommage pour notre gazon !

Activité 3 : Estimer une proportion

Une entreprise d'agronomie produit en grande quantité de la semence pour gazon, qui est un mélange de deux types de graines :

- de l'Agropyrum Repens (chiendent, noté AR), dans une proportion p. Cette graine est adaptée aux conditions sèches et au piétinement mais n'offre pas une couverture très esthétique.
- du Lolium Perenne (ray-grass en anglais), dans une proportion $1 - p$.

Le nombre de graines d'AR dans un échantillon de n graines prélevées dans la production de l'entreprise est variable. C'est une variable aléatoire notée X_n.

L'objectif est d'estimer p à partir de la fréquence $F_n = \frac{X_n}{n}$ d'apparition des graines d'AR dans un échantillon.

1. Expérience et recueil des données.

On va prélever un échantillon de n graines, successivement et avec remise, dans la production de l'entreprise contenant **une proportion p fixée mais inconnue de graines d'AR**. Cette production sera représentée par une bouteille opaque munie d'une tétine et contenant des boules rouges et des boules d'une autre couleur. A chaque tirage, on notera 1 à chaque fois que la couleur de la « graine » obtenue est rouge (c'est une graine d'AR) et 0 sinon.

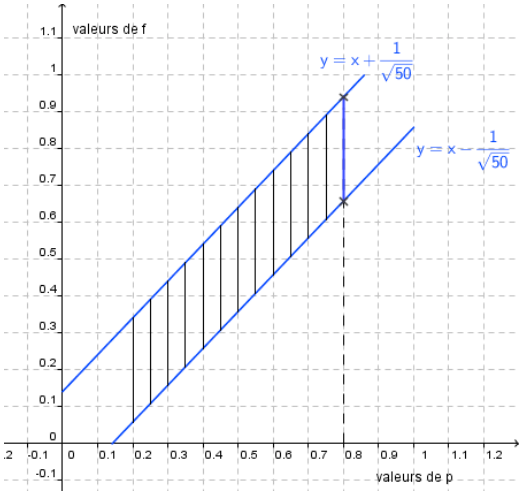
Le professeur aura préparé par avance des biberons contenant par exemple 20 boules. Tous les biberons ont la même proportion de boules rouges, par exemple 25%. Chaque élève expérimente avec une bouteille, assimilée au lot de semence et dont la proportion de boules rouges lui est inconnue.
Dans la réalité, on prélève n graines au hasard mais sans remise ; l'échantillon serait pourtant considéré comme aléatoire car obtenu parmi un nombre considérable de graines.
La reconnaissance effective des graines se fait selon leur morphologie à l'œil nu ou à la loupe (certaines filières de BTSA ont les graines de 100 espèces à reconnaître en utilisant des flores de reconnaissance).

Réaliser 50 tirages de graines pour générer un échantillon de taille 50 et noter les résultats (0 ou 1) sur une fiche individuelle.
Calculer le cumul des 1 puis la fréquence f des graines d'AR pour votre échantillon.
En comparant avec les résultats obtenus par les autres élèves, est-il raisonnable de prendre la valeur f de votre observation comme estimation de la proportion p ? Pourquoi ?

2. Estimation de p par une fourchette.

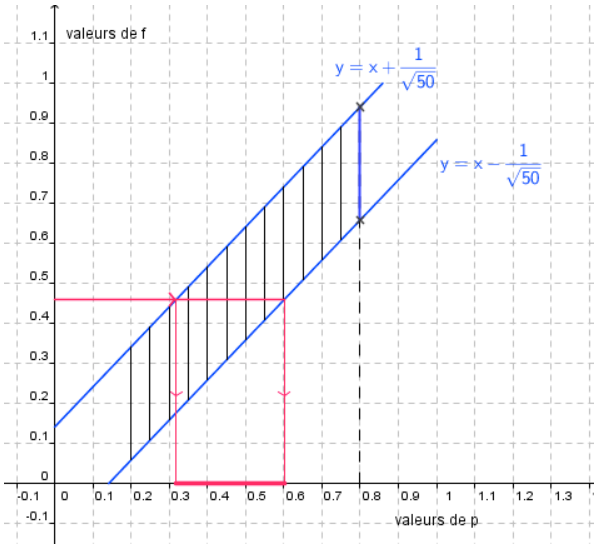
On sait depuis la classe de seconde que pour une proportion p du caractère comprise entre 0,2 et 0,8 et pour des échantillons de taille $n \geq 25$, f appartient à l'intervalle de fluctuation $\left[p - \frac{1}{\sqrt{n}} \ ; \ p + \frac{1}{\sqrt{n}}\right]$ avec une probabilité d'au moins 0,95.
Calculer ces intervalles de fluctuation pour des valeurs de p allant de 0,2 à 0,8 avec un pas de 0,05 puis représenter ces intervalles sur papier millimétré : à chaque valeur de p en abscisse, on dessinera verticalement l'intervalle associé.
On peut remarquer que les bornes de tous les intervalles de fluctuation que l'on pourrait fabriquer se situent sur les droites d'équations $y = x - \frac{1}{\sqrt{50}}$ et $y = x + \frac{1}{\sqrt{50}}$. Tracer ces droites.

Annexe 4



Réciproquement, en plaçant en ordonnée la valeur de f obtenue au 1. avec un échantillon de 50 graines, on peut déterminer graphiquement les intervalles de fluctuation contenant f avec une probabilité d'au moins 0,95. Pour cela, on cherche les abscisses a et b des points d'intersection des deux droites précédentes avec la droite d'équation $y = f$.

Annexe 5



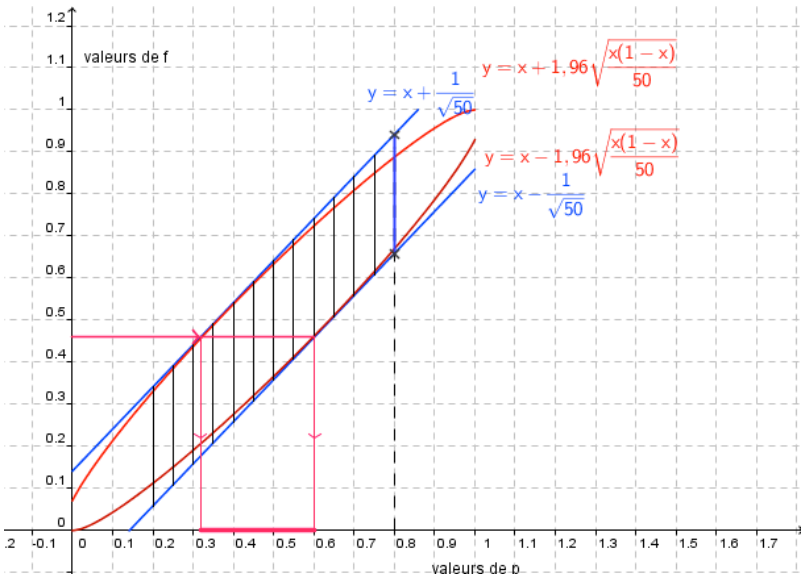
[*a* ; *b*] est appelé un intervalle de confiance de *p* au niveau 95 %. Ceci signifie que *p* est en dehors de cet intervalle pour seulement 5 % des échantillons.

Donner votre estimation de la proportion de graines d'AR dans la production de semence de gazon par un intervalle de confiance au niveau 95 %.

3. Remarques.

- Selon l'échantillon de 50 graines tiré par chaque élève, la valeur de f obtenue peut être différente. On ne peut donc pas parler de **l**'intervalle de confiance de p mais seulement d'**un** intervalle de confiance de p.

- Avec le travail sur la loi normale, on montre que l'intervalle de fluctuation de p au seuil 95 % est de la forme $\left[p - 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \ ; \ p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ dès que $n \geq 30$, $np \geq 5$ et $n(1 - p) \geq 5$. On admettra que si l'on dessine les intervalles de fluctuation pour des valeurs de p variant de 0 à 1, leurs bornes se situent sur des arcs d'ellipses.



Le professeur pourra dévoiler la composition des biberons et donner la vraie valeur de la proportion de graines d'AR même si dans l'entreprise cette opération est impossible à réaliser.
Les élèves peuvent ainsi s'apercevoir que leur propre intervalle de confiance ne contient pas forcément la valeur de p. Si on calcule le pourcentage de « bons » intervalles dans la classe, on s'aperçoit que ce résultat est proche de 95 %.

La parité, c'est quoi ?

- **Niveau :** lycée.
- **Objectif :** prise de décision à partir d'un intervalle de fluctuation.
- **Activité :**
  Deux entreprises A et B recrutent dans un bassin d'emploi où il y a autant de femmes que d'hommes, avec la contrainte du respect de la parité.

  La parité signifie que l'identité sexuelle n'intervient pas au niveau du recrutement, c'est-à-dire qu'au niveau du caractère homme ou femme, les résultats observés pourraient être obtenus par choix au hasard des individus dans la population.

  Dans l'entreprise A, il y a 100 employés dont 43 femmes; dans l'entreprise B, il y a 2500 employés dont 1150 femmes (soit 46%). Ces entreprises respectent-elles la parité ?

  Les entreprises sont assimilées à des échantillons de taille n prélevés dans une population où la proportion p de femmes est égale à 0,5.

  Déterminer les intervalles de fluctuation de p au niveau 0,95 correspondant aux entreprises A et B. Conclure.

- **Corrigé :**

  En seconde : L'intervalle de fluctuation de p au niveau 0,95 est : $\left[p-\frac{1}{\sqrt{n}}\ ;\ p+\frac{1}{\sqrt{n}}\right]$.

  L'entreprise A est un échantillon de taille 100 ; l'intervalle de fluctuation est [0,4 ; 0,6].

  L'entreprise B est un échantillon de taille 2500 ; l'intervalle de fluctuation est [0,48 ; 0,52].

  En première : l'intervalle de fluctuation de p au niveau 0,95 est : $\left[\frac{a}{n};\frac{b}{n}\right]$.

  Avec la loi binomiale $\mathcal{B}(100\ ;\ 0{,}5)$, on obtient $a = 40$ et $b = 60$, soit l'intervalle [0,4 ; 0,6].
  Avec la loi binomiale $\mathcal{B}(2500\ ;\ 0{,}5)$, on obtient $a = 1201$ et $b = 1299$, soit l'intervalle [0,4804 ; 0,5196].

  En terminale : L'intervalle de fluctuation de p au niveau 0,95 est : $\left[p-1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p+1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$.

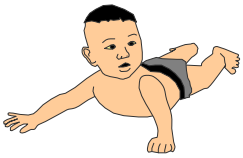
  On obtient : I_{100} = [0,402 ; 0,598] ; I_{2500} = [0,4804 ; 0,5196].

  Conclusion :
  La valeur 43% est dans l'intervalle de fluctuation pour un échantillon de taille 100 alors que 46% n'est pas dans l'intervalle de fluctuation pour un échantillon de taille 2500 !

  Autrement dit :
  - pour l'entreprise B, la proportion de 46% s'observe dans moins de 5% des échantillons de taille 2500 prélevés au hasard dans une population où il y a autant d'hommes que de femmes. Cette entreprise n'est donc pas représentative du critère « parité ».
  - pour l'entreprise A, on considère que le résultat observé est compatible avec le modèle (l'écart entre f et p est probable, au sens où il est contenu dans la fourchette que le hasard produirait avec 95% des échantillons envisageables). De ce fait, on peut admettre que l'entreprise respecte la parité.

Naissances à pile ou face





- **Niveau :** seconde et première.

- **Objectif :** prise de décision à partir d'un intervalle de fluctuation.

- **Activité :**
  Les données statistiques suivantes ont été relevées :
  - en 2000, dans le village de Xicun, en Chine, il est né 20 enfants, parmi lesquels 16 garçons,
  - dans la réserve indienne d'Aamjiwnaag, située au Canada à proximité d'industries chimiques, il est né entre 1999 et 2003, 132 enfants dont 46 garçons.

  Ces observations sont-elles le fruit du hasard ?

1. Expérimentation avec des pièces de monnaie et avec un tableur.

Lancer 20 fois une pièce de monnaie et noter le nombre de « pile ».
Comparer le résultat avec ceux des autres élèves de la classe.
Comment peut-on utiliser ces expériences pour commenter les statistiques de Xicun ?

La simulation pour le cas de 132 naissances au Canada se fera avec un tableur.

Quelle réponse peut-on apportée à la question posée ?

2. Calcul des intervalles de fluctuation

On veut rejeter (ou pas) l'hypothèse que la distribution des sexes des 20 enfants nés en 2000 à Xicun (ou des 132 enfants nés à Aamjiwnaag) n'est due qu'au seul hasard, autrement dit que les naissances de filles et de garçons sont équiprobables.

Les populations d'enfants sont assimilées à des échantillons de taille n prélevés dans une population où la proportion p de garçons est égale à 0,5.

On considère la variable aléatoire X_n « nombre de garçons dans un échantillon de n nouveau-nés ». X_n une v.a. suivant la loi binomiale $\mathcal{B}(n\ ;\ 0{,}5)$. On souhaite déterminer l'intervalle de fluctuation approché au niveau 0,95, de la variable fréquence $F_n = \dfrac{X_n}{n}$.

➢ **En seconde (pour la population d'Aamjiwnaag où $n \geq 30$)**

Déterminer l'intervalle de fluctuation approché au niveau 0,95 de la variable fréquence F_{132} .

Calculer la fréquence f de garçons observée dans la réserve indienne d'Aamjiwnaag.

Conclure.

➢ **En première (pour la population Xicun où $n < 30$)**

A l'aide de la table de la loi binomiale $\mathcal{B}(20\ ;\ 0{,}5)$, déterminer :

- la valeur du plus petit entier a tel que $P(X_n \leq a) > 0{,}025$
- la valeur du plus petit entier b tel que $P(X_n \leq b) \geq 0{,}975$.

Donner l'intervalle de fluctuation au niveau 0,95 de la variable fréquence F_{20} .

Calculer la fréquence f de garçons observée dans le village de Xicun.

Conclure.

Table de la loi binomiale $\mathcal{B}(20 ; 0{,}5)$:

k	P(X = k)	P(X ≤ k)
0	0,0000	0,0000
1	0,0000	0,0000
2	0,0002	0,0002
3	0,0011	0,0013
4	0,0046	0,0059
5	0,0148	0,0207
6	0,0370	0,0577
7	0,0739	0,1316
8	0,1201	0,2517
9	0,1602	0,4119
10	0,1762	0,5881
11	0,1602	0,7483
12	0,1201	0,8684
13	0,0739	0,9423
14	0,0370	0,9793
15	0,0148	0,9941
16	0,0046	0,9987
17	0,0011	0,9998
18	0,0002	1,0000
19	0,0000	1,0000
20	0,0000	1,0000

- **Corrigé :**

  Intervalles de fluctuation : $I_{20} = \left[\frac{6}{20} \ ; \ \frac{14}{20}\right] = [0{,}30 \ ; \ 0{,}70]$; $I_{132} = \left[0{,}5 - \frac{1}{\sqrt{132}} \ ; \ 0{,}5 + \frac{1}{\sqrt{132}}\right] = [0{,}41 \ ; \ 0{,}59]$.

  Dans les deux situations, les fréquences observées (0,8 pour Xicun et 0,35 pour d'Aamjiwnaag) n'appartiennent pas à l'intervalle de fluctuation : on rejette l'hypothèse que le nombre de garçons est dû au seul hasard (c'est peu probable). C'est une « alerte » qui doit inciter à rechercher des causes extérieures.

- **Remarque :**

  On aurait aussi pu travailler en considérant que la variable aléatoire « sexe à la naissance » prend les deux valeurs Fille ou Garçon avec les probabilités respectives 0,48 et 0,52 (issues de statistiques internationales).
  L'intervalle de fluctuation obtenu (pour $n = 20$) est [0,30 ; 0,75]. Dans ce contexte, la conclusion est la même.

- **Commentaires :**

  Il est important de préciser que les « réponses » apportées ne sont que des réponses « statistiques ». Les résultats observés sur les naissances à Xicun et Aamjiwnaag sont « bizarres » (et préoccupants). Rien de plus ne peut être dit quant aux causes, mais ces résultats doivent inciter à enquêter. C'est une obligation morale, car on a établi une « preuve statistique », rationnelle, qu'il se passe sans doute quelque chose d'inhabituel.

  Pour le cas de Xicun, la cause probable est l'acquisition dans ce village (en 1999) d'une machine à ultra-sons bon marché, permettant aux médecins de déterminer le sexe du fœtus.

  (Source : *Washington Post* du 29 mai 2001.)

  Dans le cas d'Aamjiwnaag, une enquête sanitaire est menée. En effet, depuis Seveso, le rôle de certains polluants sur les déséquilibres du sex-ratio est connu.

  (Sources : *Science et Vie* février 2006 – *Environmenthal Health Perspectives* octobre 2005)

Contester un jugement

(Source : *Prove it with figures* – H. Zeisel et D Kaye.)

- **Niveau :** lycée.

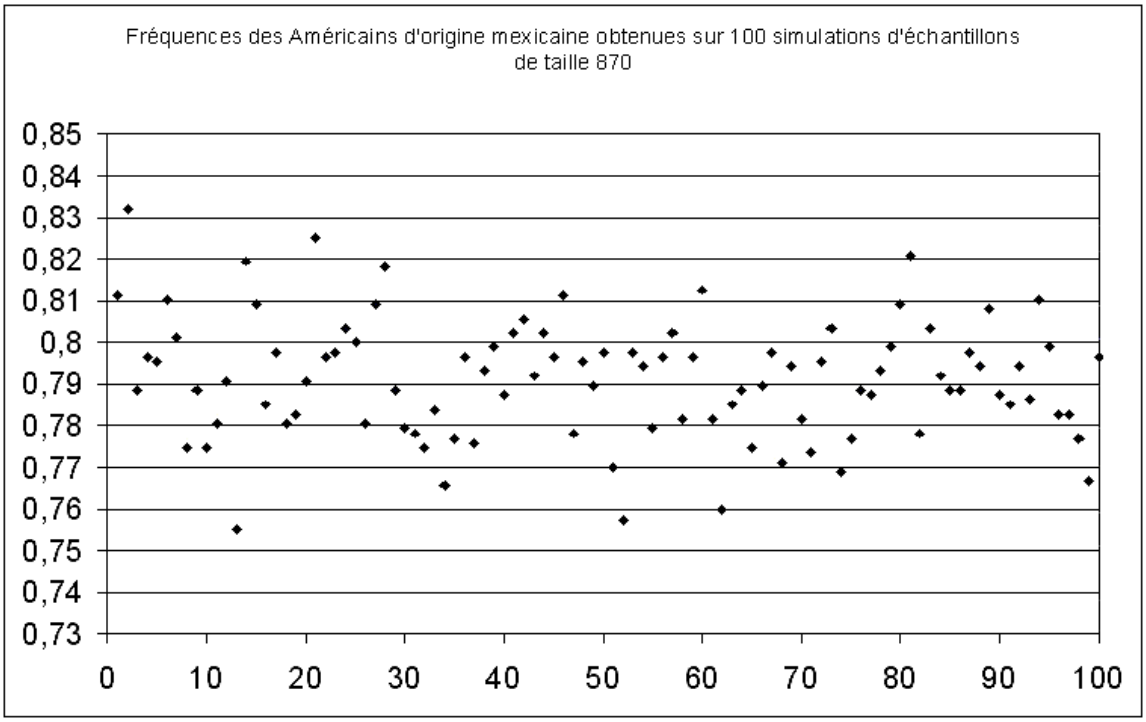
- **Objectif :** prise de décision à partir d'un intervalle de fluctuation.

- **Activité :**
  En Novembre 1976 dans un comté du sud du Texas, Rodrigo Partida était condamné à huit ans de prison. Il attaqua ce jugement au motif que la désignation des jurés de ce comté était discriminante à l'égard des Américains d'origine mexicaine. Alors que 79,1% de la population de ce comté était d'origine mexicaine, sur les 870 personnes convoqués pour être jurés lors d'une certaine période de référence, il n'y eut que 339 personnes d'origine mexicaine.

  1. Quelle est la fréquence des jurés d'origine mexicaine observée dans ce comté du Texas ?

  2. Calculer l'intervalle de fluctuation au niveau 0,95 de p.

  3. La simulation sur un tableur du prélèvement d'échantillons aléatoires de taille $n = 870$ dans une population où la proportion des habitants d'origine mexicaine est $p = 0{,}791$. Les fréquences des habitants d'origine mexicaine observées sur 100 échantillons simulés sont représentées ci-dessous.



  Quel est le pourcentage des simulations fournissant une fréquence en dehors de l'intervalle précédent ?

  Sur les simulations, est-il arrivé au hasard de fournir une fréquence d'habitants d'origine mexicaine comparable à celle des jurés d'origine mexicaine observée dans ce comté du Texas ?

  4. Comment expliquez-vous cette situation ?

- **Corrigé :**

  La fréquence observée des jurés d'origine mexicaine est $f = \frac{339}{870} \approx 0{,}39$.

  Comme intervalle de fluctuation au niveau 0,95 de p, on obtient :

  - en seconde, $\left[0{,}791 - \frac{1}{\sqrt{870}} \ ; \ 0{,}791 + \frac{1}{\sqrt{870}}\right]$ soit environ $[0{,}76 \ ; \ 0{,}82]$.
  - en terminale, $\left[0{,}791 - 1{,}96\sqrt{\frac{0{,}791 \times 0{,}219}{870}} \ ; \ 0{,}791 + 1{,}96\sqrt{\frac{0{,}791 \times 0{,}219}{870}}\right]$ soit environ $[0{,}76 \ ; \ 0{,}82]$.

  Sur le graphique, 4 points sur 100 sont en dehors de l'intervalle précédent, soit 4% des cas.
  D'autre part, la fréquence 0,39 n'a jamais été observée sur les 100 simulations et cette valeur est très éloignée de l'intervalle de fluctuation.

- **Commentaires :**
  Les données étudiées constituent une « preuve statistique » du fait que la constitution de ces jurys n'est pas totalement aléatoire, c'est-à-dire que ceux ci ne sont pas « représentatifs » de la population, du point de vue du caractère hispanique. Les calculs montrent qu'il n'est pas possible de considérer que les jurys résultent d'un tirage au sort où chaque élément de la population a les mêmes chances d'être choisi.

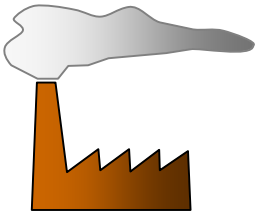
  Mais c'est tout ce que l'on peut dire et, en particulier, il n'est pas possible de se prononcer sur les causes et porter des accusations de discrimination raciale.

  L'étude statistique doit inciter à enquêter sur les conditions de constitution des jurys : au Texas, pour être juré, on doit maîtriser la langue anglaise (écrite et parlée), ce qui n'est pas le cas de la majorité de la population d'origine hispanique.

  Remarquons enfin que cette étude statistique s'est étalée sur onze années : la proportion des hispaniques dans la population durant ce temps a évolué ainsi que la proportion d'hispaniques dans les jurys.

- **Remarque :**
  Cette activité peut servir d'évaluation.

Taux anormal de cas de leucémie



- **Niveau :** seconde.

- **Objectifs :** – simulation d'échantillons sur tableur.
  – prise de décision à partir de cette simulation.

- **Activité :**
  Woburn est une petite ville industrielle du Massachusetts, au Nord-Est des États-Unis. Du milieu à la fin des années 1970, la communauté locale s'émeut d'un grand nombre de leucémies infantiles survenant en particulier chez les garçons dans certains quartiers de la ville. Les familles se lancent alors dans l'exploration des causes et constatent la présence de décharges et de friches industrielles ainsi que l'existence de polluants. Dans un premier temps, les experts gouvernementaux concluent qu'il n'y a rien d'étrange. Mais les familles s'obstinent et saisissent leurs propres experts.
  Le tableau suivant résume les données statistiques concernant les garçons de moins de 15 ans, pour la période 1969-1979 (Source : *Massachusetts Department of Public Health*)

Population des garçons de moins de 15 ans à Woburn selon le recensement de 1970	Nombre de leucémies infantiles observées chez les garçons à Woburn entre 1969 et 1979	Fréquence des leucémies aux Etats-Unis (garçons)
5 969	9	0,000 52

  Doit-on, comme l'ont alors affirmé les autorités, en accuser le hasard ?
  La question statistique qui se pose est de savoir si le hasard seul peut raisonnablement expliquer le nombre de leucémies observées chez les jeunes garçons de Woburn, considérés comme résultant d'un échantillon prélevé dans la population américaine.
  La population des États-Unis étant très grande par rapport à celle de Woburn, on peut considérer que l'échantillon résulte d'un tirage avec remise et simuler des tirages de taille n avec le tableur.

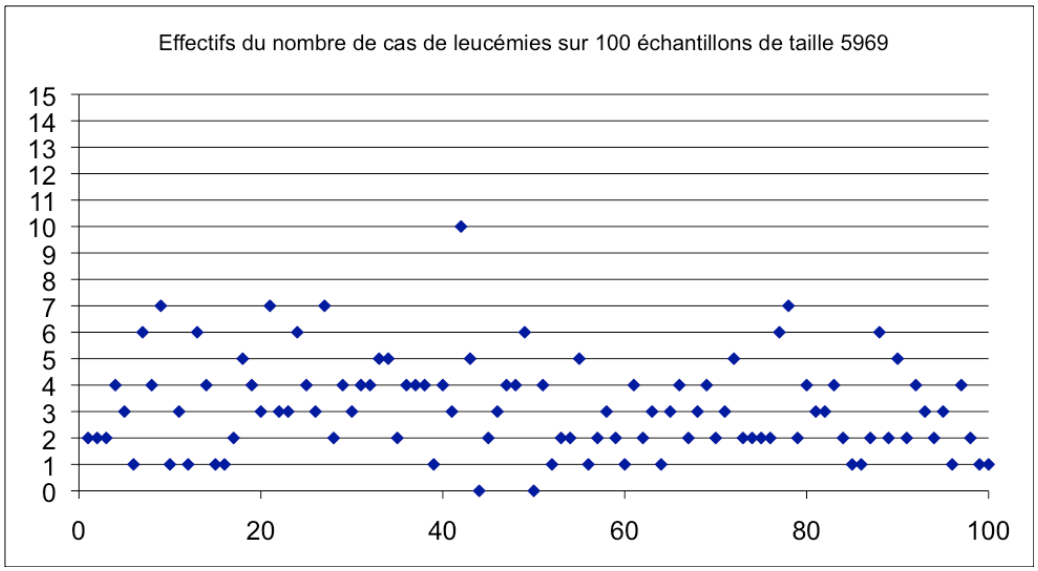
  1. **Simulation**.
     Sur tableur, on simule 100 échantillons de taille n = 5 969 prélevés au hasard dans une population de garçons où la probabilité de leucémie est $p = 0{,}00052$ (cas « normal ») en utilisant l'instruction :
     =ENT(ALEA()+0,00052).
     L'instruction =ALEA() génère un nombre aléatoire dans l'intervalle [0, 1[.
     L'instruction =ALEA()+0,000 52 génère un nombre aléatoire dans l'intervalle [0,000 52 ; 1,000 52[.
     Ainsi, l'instruction =ENT(ALEA()+0,000 52), où ENT désigne la partie entière, vaut 1 (malade) avec la probabilité 0,000 52 et 0 (non malade) sinon.
     Pour chaque échantillon, en faisant la somme des 0 et des 1, on obtient le nombre de leucémies observées.

     Représenter sur un graphique (avec un nuage de points) les 100 résultats observés.

     Calculer le pourcentage des simulations fournissant un nombre de leucémies observées supérieur ou égal à 9.

  2. **Prise de décision**.
     Peut-on penser qu'il se passe quelque chose d'étrange à Woburn ?

- **Corrigé :**
  Graphique attendu lors de la simulation :



- **Commentaires :**
  Les simulations montrent que le nombre de leucémies observées à Woburn (9 cas) est extrêmement rare (de l'ordre de 1 % des simulations sur un grand nombre d'essais), sous l'hypothèse d'une probabilité « normale » de leucémie.

  Il est donc raisonnable de penser que le nombre très « significativement » élevé des leucémies infantiles observées chez les garçons de Woburn n'est pas dû au hasard.

  Ce taux anormalement élevé de leucémies est officiellement confirmé par le Département de Santé Publique du Massachusetts en avril 1980. Les soupçons se portent alors sur la qualité de l'eau de la nappe phréatique qui, par des forages, alimente la ville. On découvre ainsi le syndrome du trichloréthylène.

Est-il nécessaire de travailler pour réussir ?

- **Niveau :** première.

- **Objectif :** étude des deux types d'erreurs dans une prise de décision.

- **Activité :**
  Un examinateur doit faire passer une épreuve de type QCM à des étudiants. Ce QCM est constitué de vingt questions indépendantes. Pour chaque question, il y a trois réponses possibles dont une seule correcte.

  On suppose qu'il y a deux sortes d'étudiants :
  - l'étudiant qui n'a pas travaillé et qui répond au hasard : il a alors une chance sur 3 d'avoir une réponse juste.
  - l'étudiant qui a travaillé : il a davantage de chance de donner une bonne réponse à chaque question mais le pourcentage de réussite est inconnu. L'examinateur l'estime cependant à 0,6.

  L'examinateur veut déterminer une valeur k telle que :
  - si le nombre de réponses correctes est supérieur ou égal à k , l'étudiant est reçu.
  - si le nombre de réponses correctes est strictement inférieur à k , l'étudiant est recalé.

  Pour cela on considère la variable aléatoire X égale au nombre de réponses correctes parmi les 20, pour un étudiant choisi au hasard. Alors $\boldsymbol{X_i}$ **suit la loi binomiale** $\boldsymbol{\mathcal{B}(20 ; p_i)}$ avec p_i : probabilité de réussite d'un étudiant.
  Si l'étudiant n'a pas travaillé, on a donc $p_1 =$ …… et X_1 suit la loi binomiale $\mathcal{B}($;).
  Si l'étudiant a travaillé, on a alors $p_2 =$ …… et X_2 suit la loi binomiale $\mathcal{B}($;).

  A l'issue des résultats de l'épreuve, quatre cas sont possibles :

  (1) l'étudiant n'a pas travaillé et il est recalé

  (2) l'étudiant a travaillé et il est reçu

  (3) l'étudiant n'a pas travaillé et il est reçu

  (4) l'étudiant a travaillé et il est recalé

  On a donc deux risques d'erreur correspondant aux cas (3) et (4) :
  - l'erreur α : on rejette l'idée que l'étudiant n'a pas travaillé alors que c'est vrai (c'est le « risque professeur ») ;
  - l'erreur β : on pense que l'étudiant n'a pas travaillé alors que c'est faux, (c'est le « risque étudiant »).

  L'examinateur cherche à contrôler ces deux erreurs α et β :
  Si on choisit $k = 10$ (au moins la moitié de réponses correctes) :
  alors $\alpha = P(X_1 \geq 10) \approx$ ……. et $\beta = P(X_2 < 10) \approx$ …….
  Si on choisit $k = 12$:
  alors $\alpha = P(X_1 \geq 12) \approx$ ……. et $\beta = P(X_2 < 12) \approx$ …….

  Comment arriver à réduire en même temps les deux erreurs ? En augmentant le nombre de questions.
  Avec par exemple 40 questions, si on considère la variable aléatoire Y_i égale au nombre de réponses correctes parmi les 40, alors Y_1 suit la loi binomiale $\mathcal{B}(40 ; \frac{1}{3})$ pour un étudiant qui n'a pas travaillé et Y_2 suit la loi binomiale $\mathcal{B}(40 ; 0{,}6)$ pour un étudiant qui a travaillé.

  Si on choisit $k = 20$ (au moins la moitié de réponses correctes) :
  alors $\alpha = P(Y_1 \geq 20) \approx$ **0,02** et $\beta = P(Y_2 < 20) \approx$ **0,07**

tables de la loi Binomiale

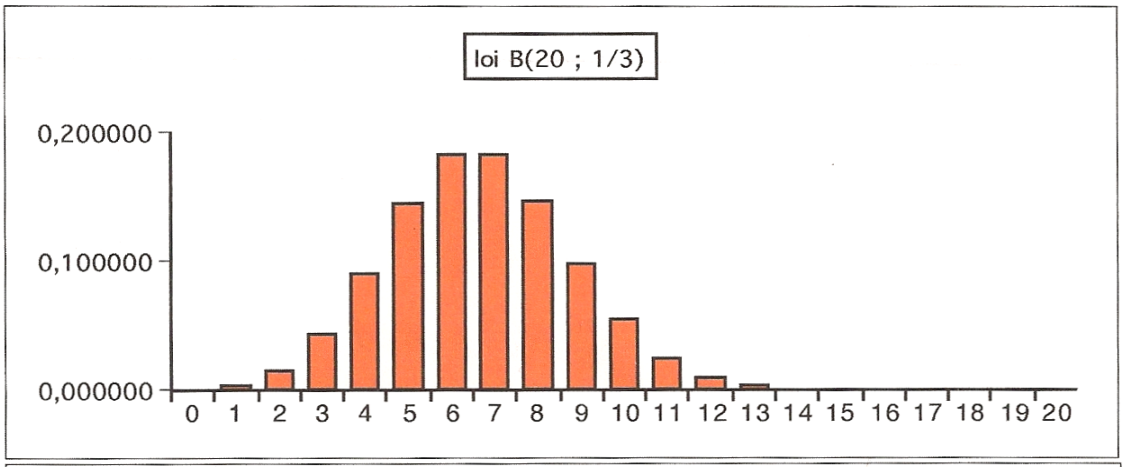
n = 20

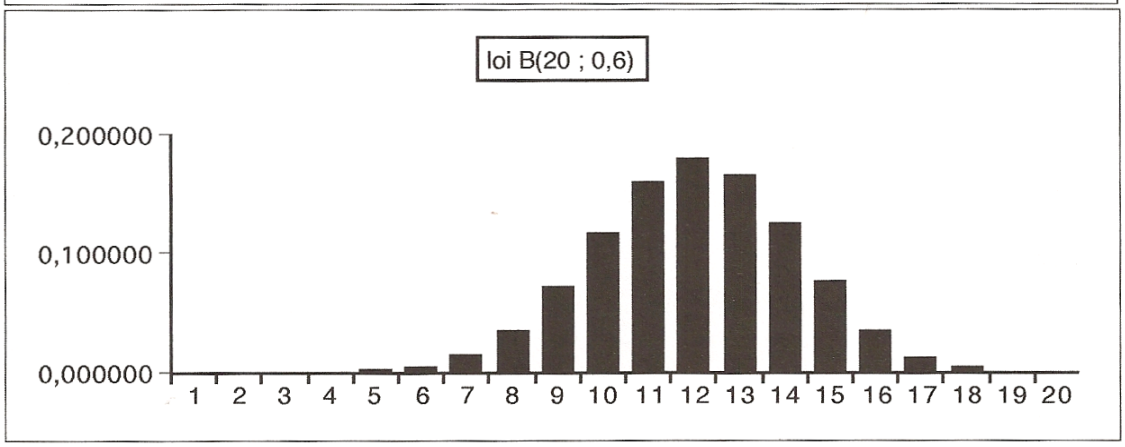
p=1/3

k	P(X=k)	P(X≥k)
0	0,000301	1,000000
1	0,003007	0,999699
2	0,014285	0,996692
3	0,042854	0,982407
4	0,091064	0,939554
5	0,145703	0,848489
6	0,182129	0,702786
7	0,182129	0,520657
8	0,147980	0,338529
9	0,098653	0,190549
10	0,054259	0,091896
11	0,024663	0,037637
12	0,009249	0,012973
13	0,002846	0,003725
14	0,000711	0,000879
15	0,000142	0,000167
16	0,000022	0,000025
17	0,000003	0,000003
18	0,000000	0,000000
19	0,000000	0,000000
20	0,000000	0,000000

p=0,6

k	P(X=k)	P(X<k)
0	0,000000	0,000000
1	0,000000	0,000000
2	0,000005	0,000000
3	0,000042	0,000005
4	0,000270	0,000047
5	0,001294	0,000317
6	0,004854	0,001612
7	0,014563	0,006466
8	0,035497	0,021029
9	0,070995	0,056526
10	0,117142	0,127521
11	0,159738	0,244663
12	0,179706	0,404401
13	0,165882	0,584107
14	0,124412	0,749989
15	0,074647	0,874401
16	0,034991	0,949048
17	0,012350	0,984039
18	0,003087	0,996389
19	0,000487	0,999476
20	0,000037	0,999963





Marge d'erreur de 3% du sondage par quotas ?

D'après un article « Statistique et sondages » de Jeanne Fine

- **Niveau :** terminale et post-bac
- **Objectif :** comparer une estimation ponctuelle et une estimation par intervalle de confiance.
- **Activité :**
  Le 18 avril 2002, l'institut CSA effectue un sondage dans la population en âge de voter pour le premier tour de l'élection présidentielle du 21 Avril 2002.
  On constitue un échantillon de 1000 personnes (inscrites sur les listes électorales) que l'on suppose choisies de manière aléatoire. Voici les résultats recueillis auprès des 1000 personnes interrogées :

	Effectif	Fréquence
Electeurs déclarant vouloir voter pour J. Chirac	195	19,5 %
Electeurs déclarant vouloir voter pour L. Jospin	180	18,0 %
Electeurs déclarant vouloir voter pour J-M. Le Pen	140	14,0 %

Si on prend ces fréquences comme estimation ponctuelle des résultats de l'élection (**non connus le 18 avril**), quels sont les deux candidats qui auraient dû s'affronter au second tour ?

Voici les résultats du premier tour de l'élection présidentielle du 21 avril 2002 :

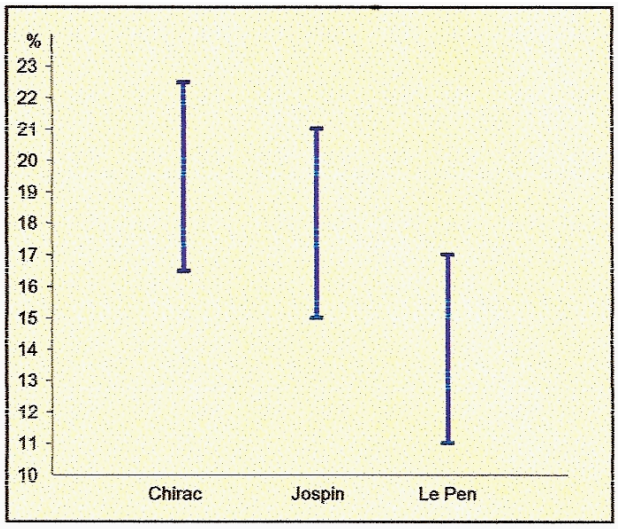
	Résultat de l'élection
Chirac	19,7 %
Jospin	16,1 %
Le Pen	16,9 %

Ces résultats correspondent-ils aux prévisions ?

Cependant, les statisticiens savent que la valeur observée f varie d'un échantillon à l'autre, de par la loi d'échantillonnage de la fréquence. Cette fluctuation est à prendre en compte et à l'estimation ponctuelle on préfèrera une estimation par intervalle de confiance.
Ainsi, à la place de pourcentages bruts, l'institut de sondage devrait donner des « fourchettes », entachées d'une marge d'erreur de $\frac{1}{\sqrt{1000}}$, soit environ 3 %.
Les résultats du sondage se présenteraient alors sous la forme suivante :

Chirac	entre 16,5 % et 22,5 % des voix
Jospin	entre 15,0 % et 21,0 % des voix
Le Pen	entre 11,0 % et 17,0 % des voix

Ce qui peut être représenté par le graphique suivant :



On constate donc que les scores réels sont compatibles avec ces intervalles, puisque, pour chacun des candidats, l'écart (en valeur absolue) entre l'estimation ponctuelle du sondage et le résultat définitif est inférieur à 3%, marge d'erreur de la technique.

Il apparaît que Le Pen peut être second … et même premier : toutes les configurations de l'ordre des trois premiers candidats étaient possibles.

Bibliographie

- Document ressource des nouveaux programmes de lycée professionnel, Avril 2009.
  Pour le télécharger : http://www.ac-grenoble.fr/maths/docresseconde/Proba_stat_LP.doc

  Les activités suivantes sont inspirées de ce document :
  « Les cartes de contrôle » (p 40),
  « Naissances à pile ou face » (p 57),
  « Contester un jugement » (p 59),
  « Taux anormal de cas de leucémie » (p 61).

- Document ressource pour les classes de lycées généraux et technologiques (probabilités et statistiques).

  L'activité « La parité, c'est quoi ? » (p 56) est extraite de ce document.

- Enseigner les probabilités en classe de Terminale, IREM de STRASBOURG, 1994.

- L'induction statistique au lycée illustrée par le tableur, Philippe DUTARTE, 2005 (éditions Didier).

- Statistiques et Citoyenneté, le citoyen face au chiffre, IREM de PARIS-NORD, brochure n°135, 2007.

  Les activités suivantes sont inspirées de cette brochure :
  « Lecture de graphiques » (exercices 1 et 2 p 5 et 6),
  « Tabac et risques d'infarctus » (p 11),
  « L'ascenseur social » (p 33).

AUTEURS : Annette CORPART
Nelly LASSALLE

TITRE : Que proposer aux élèves en statistiques et probabilités du collège au lycée ?

ÉDITEUR : IREM de CLERMONT-FERRAND.

DATE : Avril 2016.

PUBLIC CONCERNÉ : Enseignants de collège et lycée.

RÉSUMÉ : Les statistiques et les probabilités ont pris une place importante dès le collège dans l'enseignement et tout au long de la vie du citoyen. Il est donc nécessaire que l'enseignant dispose d'activités variées pour ses classes.

MOTS CLÉS : Simulation – Hasard – Statistiques – Probabilités – Graphiques – Arbres – Echantillonnage – Prise de décision – Estimation.

FORMAT A4 : 66 pages.