Probabilité statistique en seconde

Énoncé 1 : Urne de Polya.

Une urne contient une boule rouge et une boule blanche (indiscernables au toucher).
On choisit au hasard une boule :

- Si cette boule est rouge, il remet deux boules rouges dans l'urne.
- Si cette boule est blanche, il remet deux boules blanches dans l'urne.

1 Étudier la composition de l'urne après 2 tirages, 3 tirages.
2 Étudier la composition de l'urne après plusieurs tirages.
3 Étudier la dynamique de l'urne. (C'est à dire l'évolution de la proportion de boules blanches dans l'urne lorsque le nombre de tirages devient grand).

Énoncé 2 : Tirage dans une urne

Une urne contient 10 boules : 3 boules blanches, 5 boules noires et 2 boules rouges.
1) Écrire un algorithme permettant de simuler 100 tirages avec remise et étudier la distribution de fréquence d'apparition de chacune des couleurs, l'adapter pour un nombre N donné de tirages.
Programmer cet algorithme, que constate-t-on lorsque N augmente ?
2) Construire sur tableur l'histogramme des distributions de fréquences et celui de la distribution de probabilités afin de les comparer.

Énoncé 3 : Jeu du croix - pile.

Diderot et d'Alembert échangèrent une correspondance au sujet d'un problème de jeu, le problème de "Croix ou Pile".
Un joueur lance une pièce de monnaie, il gagne s'il obtient croix en jouant au maximum deux fois.
M. d'Alembert prétend que la mise est de 2 contre 1 puisque : « Si Croix arrive du premier coup, le jeu est fini. Les combinaisons Croix-Croix et Croix-Pile se réduisent donc à une. Il n'y a que trois combinaisons possibles, deux qui font gagner et une qui fait perdre, donc la mise est de 2 contre 1 ».
M. Diderot rétorque que la mise est de 3 contre 1, en effet : « Il y a quatre combinaisons différentes (Croix-Croix, Croix-Pile, Pile-Croix et Pile-Pile). Les trois premières font gagner, seule la dernière fait perdre. La mise doit être de 3 contre 1 ».

Qui de Diderot ou de d'Alembert a raison ?

Énoncé 4 : Duc de Toscane

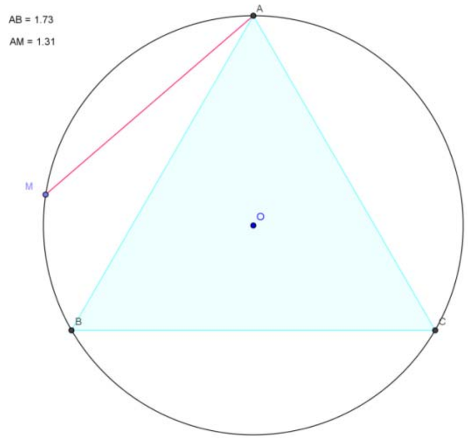
On lance trois dés. On calcule la somme des trois résultats "sortis".
A la cour de Florence, au dix-septième siècle, le Prince de Toscane avait remarqué que l'on obtenait plus souvent 10 que 9, et il écrivit à Galilée pour lui poser la question : « Comment se fait-il que le total de 10 apparaisse plus souvent que le total de 9, alors qu'il y a exactement le même nombre de façons différentes d'écrire ces deux nombres comme sommes de trois termes compris entre 1 et 6 ? »

Le Prince se trompait-il ? Quelle est votre explication ?

Énoncé 5 : Paradoxe de Bertrand

Traçons au hasard une corde dans un cercle.
Quelle est la probabilité pour qu'elle soit plus longue que le côté d'un triangle équilatéral inscrit dans ce cercle ?

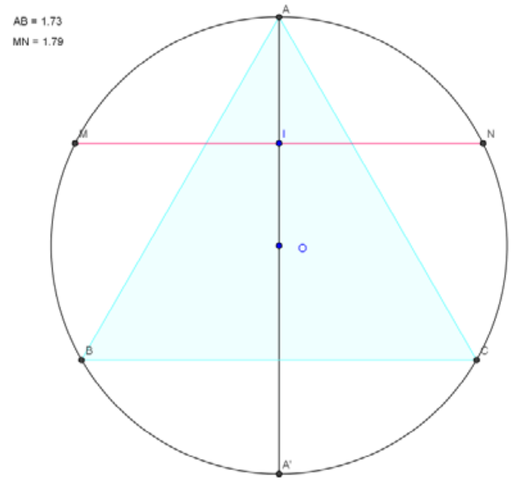
Modélisation 1



On considère un triangle équilatéral ABC inscrit dans le cercle de centre O et de rayon 1 et une corde [AM] de ce cercle : ici le point M est un point libre sur le cercle.

1. Réaliser la figure ci-contre à l'aide d'un logiciel de géométrie dynamique puis déterminer les positions de M sur le cercle pour lesquelles la longueur de la corde [AM] est supérieure à la longueur du côté [AB] ?
2. Quelle est la probabilité de choisir M au hasard sur le cercle tel que la longueur de la corde [AM] soit supérieure à la longueur du côté [AB] ?

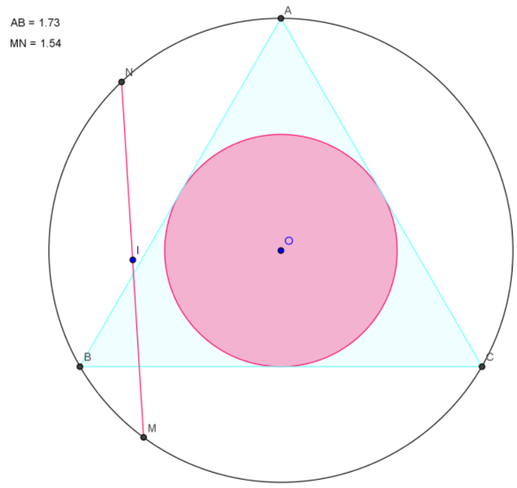
Modélisation 2



On considère un triangle équilatéral ABC inscrit dans le cercle de centre O et de rayon 1.
La longueur de la corde [AM] est déterminée par la position de son milieu I : ici I est un point libre sur le diamètre [AA'].

1. Réaliser la figure ci-contre à l'aide d'un logiciel de géométrie dynamique, puis déterminer les positions du point I pour lesquelles la longueur de la corde [MN] est supérieure à la longueur du côté [AB] ?
2. Quelle est la probabilité de choisir I au hasard sur le diamètre [AA'] tel que la longueur de la corde [MN] soit supérieure à la longueur du côté [AB] ?

Modélisation 3



On considère un triangle équilatéral ABC inscrit dans le cercle de centre O et de rayon 1.
La longueur de la corde [AM] est déterminée par la position de son milieu I : ici I est un point libre à l'intérieur du cercle circonscrit au triangle ABC.

1. Comment doit-on choisir le point I pour que la longueur de la corde [MN] soit supérieure à la longueur du côté [AB] ?
2. Quelle est la probabilité de choisir I au hasard dans le disque de centre O et de rayon 1 tel que la longueur de la corde [MN] soit supérieure à la longueur du côté [AB] ?

Comparer les trois modélisations et expliquer le paradoxe entre les résultats trouvés.

Énoncé 6 : Pile-face.

1) Écrire un algorithme puis un programme sur la calculatrice ou à l'aide d'un logiciel permettant de calculer la fréquence f des « Pile » lors de 100 lancers.

2) Les statisticiens démontrent que, pour environ 95% d'entre elles, les fréquences f d'apparition appartiennent à l'intervalle de fluctuation $\left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right]$(lorsque n supérieur ou égal à 25 et p comprise entre 0.2 et 0.8).

Calculer 20 valeurs de f avec le programme de la question1) parmi ces valeurs combien sont en dehors de l'intervalle de fluctuation au seuil de 95%?

Ce résultat est-il conforme à la définition de l'intervalle de fluctuation ?

Adapter le programme pour N lancers, quelle est l'évolution des fréquences f lorsque N augmente jusqu'à 1000 ?

b) On donne l'algorithme ci-dessous, expliquer ce qu'il affiche comme résultat.

```
▼ VARIABLES
│ ├ I EST_DU_TYPE NOMBRE
│ ├ J EST_DU_TYPE NOMBRE
│ ├ X EST_DU_TYPE NOMBRE
│ ├ L1 EST_DU_TYPE LISTE
│ └ L2 EST_DU_TYPE LISTE
▼ DEBUT_ALGORITHME
│ ▼ POUR J ALLANT_DE 1 A 50
│ │ ├ DEBUT_POUR
│ │ ├ X PREND_LA_VALEUR 0
│ │ ▼ POUR I ALLANT_DE 1 A 100
│ │ │ ├ DEBUT_POUR
│ │ │ ▼ SI (random()<0.5) ALORS
│ │ │ │ ├ DEBUT_SI
│ │ │ │ ├ X PREND_LA_VALEUR X+1
│ │ │ │ └ FIN_SI
│ │ │ └ FIN_POUR
│ │ ├ L1[J] PREND_LA_VALEUR J
│ │ ├ L2[J] PREND_LA_VALEUR X/100
│ │ ├ AFFICHER L1[J]
│ │ ├ AFFICHER L2[J]
│ │ └ FIN_POUR
└ FIN_ALGORITHME
```

c) Adapter cet algorithme puis le programmer afin de représenter le nuage de points d'abscisse le nombre de lancers et d'ordonnée la fréquence observée d'apparition de « pile ».
Combien de fréquences sont à l'extérieur de l'intervalle de fluctuation ?

Énoncé 7 : Sex-Ratio

(extrait de la brochure Statistique et citoyenneté : IREM Paris-Nord)

Le « sex-ratio » est le rapport du nombre de garçons à celui des filles à la naissance.
Il est habituellement de 105 garçons pour 100 filles.

La probabilité de naissance d'un garçon est environ égale à $p = \frac{105}{105+100}=0,512$.

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 ?

Énoncé 8 : Chaîne de production.

1) Dans une usine automobile, on contrôle les défauts de peinture de type« grains ponctuels sur le capot ». Lorsque le processus est sous contrôle, on a 20 % de ce type de défauts.

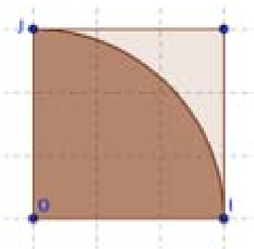
Lors du contrôle aléatoire de 50 véhicules, on observe 26 % de défauts (13 sur 50). Faut-il s'inquiéter ?

2) Dans une autre usine, le responsable de la fabrication affirme que le nombre de produits fabriqués présentant un défaut est égal à 7%.
Sur la chaine de fabrication, on a prélevé 38 produits, et on a relevé 6 produits présentant un défaut. Faut-il s'inquiéter ?

Énoncé 9 : Méthode de Monte-Carlo

On considère un quart de cercle de centre O et de rayon 1cm inscrit dans un carré de côté 1cm.

On lance une fléchette au hasard dans le carré. Soit M le point à l'intérieur du carré correspondant à l'endroit où s'est plantée la fléchette.



1) Rédiger un algorithme permettant de simuler 1000 lancers de fléchettes et d'obtenir le nombre de lancers atteignant la cible, l'adapter pour un nombre fixé *N* de lancers.
2) En programmant l'algorithme précédent à l'aide d'un logiciel ou de la calculatrice, compléter le tableau suivant.

Nombre d'essais	1000	5000	50000	100000
Fréquence f d'obtenir la fléchette à l'intérieur de la cible (arrondie à 0,01)				

3) Pour un nombre n de simulations ($n \geq 25$) et pour une fréquence observée f, l'intervalle $\left[f-\frac{1}{\sqrt{n}};f+\frac{1}{\sqrt{n}}\right]$ est appelée intervalle de confiance au niveau 0,95.

(Cela signifie que 95% des intervalles de cette forme contiennent la probabilité p d'obtenir la fléchette à l'intérieur du quart de cercle). Donner un encadrement de p le plus précis possible.
4) Combien de lancers faudrait-il simuler pour obtenir un encadrement de p (toujours avec un niveau de confiance de 95%) à 0,001 près ?
5) Calculer la probabilité p d'obtenir la fléchette à l'intérieur du quart de cercle ?
Comparer et commenter avec le dernier résultat de la simulation.

Énoncé 9 : Sondages

(extrait de la brochure Statistique et citoyenneté : IREM Paris-Nord)

Lors du premier tour des élections présidentielles en 2002, le dernier sondage publié par l'institut B.V.A., effectué sur 1000 électeurs le vendredi 19/04/02 prévoyait :

Jacques Chirac	**19%**
Lionel Jospin	**18%**
Jean-Marie Le Pen	**14%**

La surprise a été grande le dimanche 21/04/02 au vu des résultats, puisque Jean-Marie Le Pen figurait au second tour :

Jacques Chirac	**19,88%**
Lionel Jospin	**16,18%**
Jean-Marie Le Pen	**16,86%**

Les statisticiens démontrent que :
pour un échantillon de taille 1000 si f désigne la proportion d'électeurs favorables au candidat X alors l'intervalle $\left[f-\frac{1}{\sqrt{1000}};f+\frac{1}{\sqrt{1000}}\right]$ est appelé fourchette de sondage au niveau 0,95.
Cela signifie que 95% des intervalles de cette forme contiennent la proportion réelle d'électeurs qui voteront pour le candidat X.

1) Calculer les trois fourchettes de sondages à partir du sondage B.V.A. et les représenter sur un graphique. Peut-on « prévoir » l'ordre des candidats au premier tour de l'élection ?
Doit-on considérer que ce sondage était « faux » ?
2) Un autre sondage aurait-il pu prévoir le résultat de ces élections ?
On procède à une simulation d'un sondage de taille 1000 en utilisant les pourcentages obtenus à l'issue des résultats.
Pour cela on simule le tirage au hasard d'un nombre entre 1 et 100 :

- Si le nombre tiré est compris entre 1 et 20 alors on convient qu'il s'agit d'un électeur de J Chirac.
- Si le nombre tiré est compris entre 20+1 =21 et 20+16=36 alors il s'agit d'un électeur de L Jospin.
- Si le nombre tiré est compris entre 36+1 =37 et 36+17=53 alors il s'agit d'un électeur de J-M Le Pen.

Ouvrir le fichier « elections.eleves » et compléter les cellules E2, E3, E4, F2, F3, F4 et G2, G3, G4 afin de calculer pour chaque candidat les valeurs de *f*, de « borne inférieure » : $f-\frac{1}{\sqrt{1000}}$ et de « borne supérieure » : $f+\frac{1}{\sqrt{1000}}$.
A l'aide de la touche F9 observer plusieurs sondages de taille 1000.
Observez-vous des sondages analogues au dernier sondage B.V.A. qui donnait l'ordre Chirac - Jospin - Le Pen ?