



I/ Simulation d'une expérience aléatoire

Simuler une expérience, c'est la remplacer par une autre expérience qui permet d'obtenir des résultats similaires à ceux de la première, tout en offrant l'avantage d'être plus simple à réaliser.

Par exemple :

- lors du TP «Jeu de dés», on a simulé le lancé d'un dé grâce à la commande : $=\text{ENT}(\text{ALEA}())*6+1$ (on rappelle que $\text{ALEA}()$ est une fonction qui permet d'obtenir "aléatoirement" un nombre dans l'intervalle $[0 ; 1]$, tandis que $\text{ENT}()$ permet d'obtenir la partie entière d'un nombre);
- avec le TP «Un point dans un carré», on a simulé le choix d'un point quelconque dans un carré;
- le TP «Naissances à pile ou face» nécessitait la commande $=\text{ENT}(\text{ALEA}()+0,5)$, simulant le lancé d'une pièce de monnaie (pile ou face / 0 ou 1).

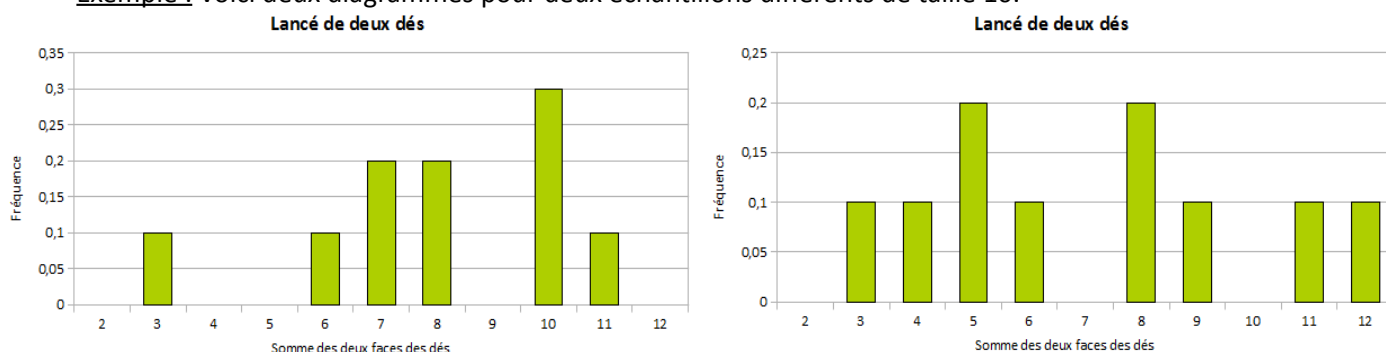
II/ Fluctuation d'échantillonnage

Lorsqu'on répète n fois une même expérience sous les mêmes conditions, on obtient un échantillon de taille n .

On peut alors calculer les fréquences d'apparition des issues qui ressortent de cet échantillon : on obtient ainsi une distribution de fréquences, associée à cet échantillon particulier de taille n .

Mais les distributions des fréquences varient d'un échantillon à l'autre pour une même expérience : c'est ce qu'on appelle la fluctuation d'échantillonnage.

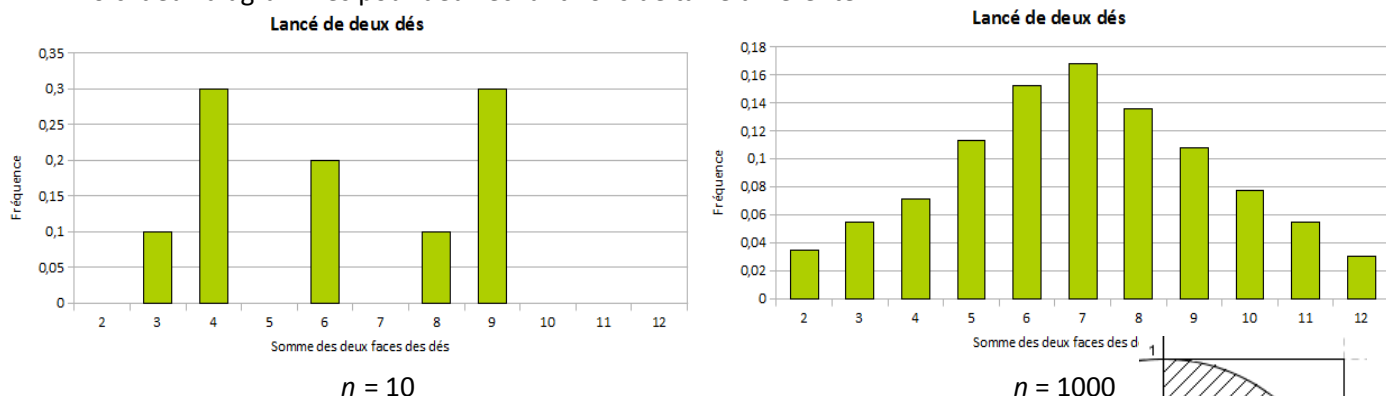
Exemple : Voici deux diagrammes pour deux échantillons différents de taille 10.



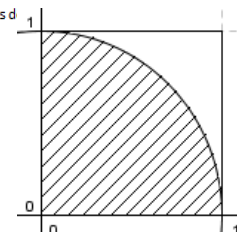
Plus la taille n de l'échantillon est grande, plus les fréquences tendent à se stabiliser : l'analyse de l'échantillon sera alors plus pertinente.

Exemple : Lorsqu'on lance deux dés, le cours de probabilités nous dit que le nombre qui a la plus forte probabilité d'être obtenu est le 7.

Voici deux diagrammes pour deux échantillons de taille différente.



III/ Intervalle de fluctuation (TP «Un point dans un carré»)



Seconde 07-09 - Année Scolaire 2009 - 2010

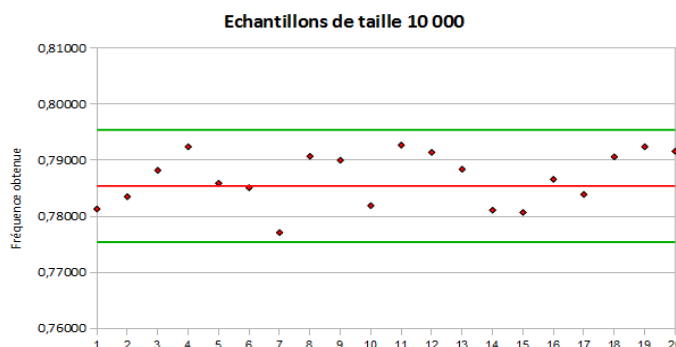
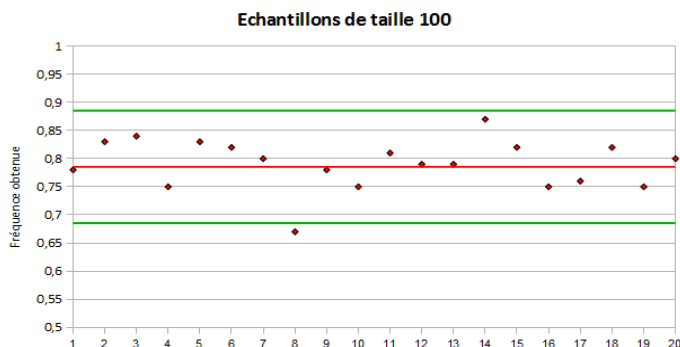
Chapitre n°3 : Notions d'échantillonnage ;

Intéressons-nous à un caractère, par exemple le fait qu'un point aléatoirement choisi dans un carré se situe dans le quart de disque comme sur la figure ci-contre.

Les calculs concluent à dire que ce point a une probabilité $p = \frac{\pi}{4}$ de se situer dans la zone hachurée.

On a réalisé 20 échantillons de taille $n = 100$, puis de taille $n = 10000$, de l'expérience consistant à choisir ce point aléatoirement.

A chaque échantillon, on attribue une fréquence (nombre de points dans le quart de disque/ n), qu'on enregistre de sorte à obtenir un nuage de points.



On se rend compte que les valeurs des fréquences obtenues sont assez proches de la valeur $\frac{\pi}{4} \approx 0,785$.

Plus précisément, sur les graphiques on observe qu'il y a au moins 95% de ces échantillons pour lesquels la fréquence est comprise entre : $p - 0,1$ et $p + 0,1$ (diagramme de gauche); $p - 0,01$ et $p + 0,01$ (diagramme de droite).

Théorème : Si on analyse un grand nombre d'échantillons de taille n ($n \geq 25$), et que l'on observe à chaque fois la fréquence d'apparition f de l'issue choisie, on s'aperçoit que pour une probabilité p comprise entre 0,2 et 0,8, au moins 95% des fréquences se situent dans un intervalle donné par :

$$\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right].$$

Cet intervalle est appelé intervalle de fluctuation.

IV/ Prise de décision (TP «Naissances à pile ou face»)

Étudions un échantillon : par exemple l'étude des naissances dans la réserve indienne d'Aamjiwnaag.

La proportion de garçon y est égale à $f = \frac{46}{132} \approx 0,35$, et la taille de l'échantillon est $n = 132$.

La probabilité naturelle pour une femme d'obtenir un garçon à la naissance est estimée autour de $p = 0,5$. Donc selon le théorème précédent, il y a normalement 95% de chances que f appartienne à l'intervalle :

$$\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right] = \left[0,5 - \frac{1}{\sqrt{132}}; 0,5 + \frac{1}{\sqrt{132}} \right], \text{ soit environ } [0,41; 0,59].$$

f n'appartient pas à cet intervalle de fluctuation : cela semble suspect; on est en droit de se poser la question de savoir s'il n'y a pas une cause extérieure expliquant ce taux exceptionnellement faible de naissance de garçons dans la réserve indienne d'Aamjiwnaag.

TP d'échantillonnage : Jeu de dés

On lance deux dés équilibrés à six faces et on fait la somme des deux nombres obtenus. Plutôt que d'utiliser deux dés réels, on va simuler cette expérience à l'aide du tableur.



I/ Étude d'un échantillon de taille n = 10

Pour simuler le lancé d'un dé équilibré à six faces, on génère les numéros 1, 2, 3, 4, 5 et 6 en calculant $ALEA() \times 6 + 1$, ce qui fournit un nombre compris entre 1 et 7 (7 non compris), puis en prenant sa partie entière.

Dans les cellules B3 et C3, rentrez la commande : $=ENT(ALEA()) \times 6 + 1$

Calculez par une formule adéquate la somme des deux dés dans la cellule D3.

Étirez enfin les formules, de sorte à simuler 10 lancers de deux dés.

On souhaite maintenant compléter le tableau récapitulant les résultats obtenus. Pour cela, on va utiliser la commande NB.SI(). Cette commande permet de compter le nombre de cellules d'une plage de données satisfaisant à un critère donné.

Par exemple, on souhaite indiquer dans la cellule G3 le nombre de fois que le nombre 2 est apparu dans la colonne D. On rentre alors la commande :

$=NB.SI(D3:D12;2)$
Plage de données Nombre à chercher dans la plage

Complétez le reste de la ligne en adaptant la formule, puis le reste du tableau.

Le diagramme décrivant la distribution des fréquences se crée au fur et à mesure.

Recopiez le tableau obtenu donnant la distribution des fréquences associée à cet échantillon.

CR : Après avoir recopié ce tableau, vous calculerez tous les paramètres (médiane, moyenne, étendue, 1^{er} et 3^{ème} quartiles) de cette série statistique.

II/ Fluctuation d'échantillonnage

En appuyant sur la touche F9 (ou sur la touche  correction selon les versions du logiciel) du clavier, les dés "se relancent" : l'ordinateur recalcule d'autres valeurs aléatoires.

Les distributions des fréquences varient d'un échantillon à l'autre pour une même expérience : c'est ce qu'on appelle la fluctuation d'échantillonnage.

Remarque : Selon les versions du logiciel, le graphique s'adapte ou non aux nouvelles valeurs. Si le diagramme reste inchangé, on peut retaper une formule simple (par exemple, celle en G4) pour forcer le logiciel à réadapter le graphique.

CR : Relancez plusieurs fois les dés. A partir de vos observations sur le diagramme, commentez l'expérience (y a-t-il des issues qui apparaissent plus/moins fréquemment que d'autres? la simulation est-elle assez riche de renseignements pour conclure?...)

III/ Échantillonnage de plus grande taille (n = 100, 1 000)

On se propose d'étudier un échantillon de taille $n = 100$, c'est-à-dire 100 lancers de dés. Pour cela, il vous suffit d'étirer les formules simulant les dés.

Adaptez les formules du tableau de façon à étudier l'échantillon de taille 100.

CR : Reprenez l'étude de l'échantillon de taille $n = 10$ pour la confronter avec cette nouvelle expérience. Observez-vous des changements notables?

Poursuivez votre étude avec un échantillon de taille 1 000.

IV/ Le modèle théorique

CR : Si vous lancez deux dés équilibrés à six faces, et que vous calculez la somme des deux nombres obtenus, quelle issue a théoriquement la plus grande probabilité de se réaliser? Vous argumenterez votre réponse.

Vous confronterez ce résultat aux résultats obtenus au cours du TP.

TP d'échantillonnage : Naissances à pile ou face

I/ Explications du sujet

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 Communauté Aamjiwnaang, 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?

CR : Déterminer la proportion de garçons (en %) à la naissance de chacune de ces deux localités. Cela vous paraît-il normal?

II/ Simulation d'une naissance : garçon ou fille?

Dans la cellule B3, rentrez la formule =ENT(ALEA()+0,5). L'ordinateur renvoie comme valeur soit 0, soit 1. On adopte alors la convention suivante :

- si on obtient le chiffre 0, on dira que cela simule la naissance d'une fille,
- sinon, on obtient le chiffre 1 et on dira que cela simule la naissance d'un garçon.

CR : Expliquez ce qu'effectue la formule =ENT(ALEA()+0,5). Pourquoi cette formule renvoie de façon équiprobable le chiffre 0 et le chiffre 1. Sur Wikipédia, cherchez la signification du terme sex-ratio. A combien l'estime-t-on à la naissance? En déduire pourquoi le modèle expérimental est cohérent avec ce que propose la réalité. Proposez une autre façon, plus expérimentale, permettant d'effectuer la même simulation que ce que produit la formule du tableur.

III/ Dan le village de Xicun

Étirez la formule rentrée en B3, de sorte à simuler la naissance de 20 bébés.

Dans une cellule à part, calculez par la méthode de votre choix la proportion de garçons nés, via cette simulation, par rapport au nombre total de naissances. Vous obtenez ainsi une fréquence d'apparition. Enregistrez celle-ci dans une autre cellule, puis relancez la simulation en appuyant sur la touche F9 ou sur la touche  correction.

Vous obtenez une nouvelle simulation, et donc une nouvelle fréquence. Enregistrez cette nouvelle fréquence dans une cellule adjacente à l'autre conservée, puis répétez l'opération de façon à obtenir une vingtaine de fréquence.

CR : Parmi les fréquences obtenues, retrouvez-vous la proportion calculée dans la partie I/ pour Xicun? La situation de Xicun vous paraît-elle normale ou alarmante?

IV/ Dans la Communauté Autochtone Aamjiwnaang des Premières Nations

Simulez la naissance de 132 bébés selon le même mode opératoire que précédemment, puis enregistrez 20 fréquences associées.

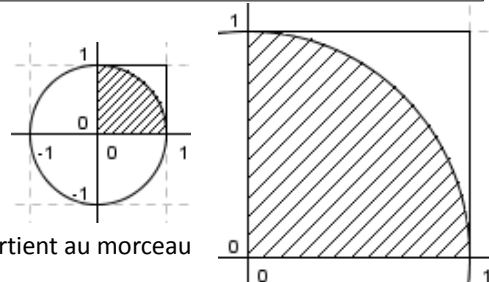
A partir de l'assistant graphique, tracez un diagramme en «nuage de points» recensant toutes les fréquences obtenues. Légendez les axes et y indiquez un titre, puis sollicitez moi pour imprimer votre production.

CR : Sur ce graphique, tracez la droite d'équation $y = 0,5$. Calculez $\sqrt{132}$, puis tracez les droites d'équation $y = 0,5 - \frac{1}{\sqrt{132}}$ et $y = 0,5 + \frac{1}{\sqrt{132}}$. Combien de points (sur les 20 représentés) se situent entre ces deux droites? Quel pourcentage cela représente-t-il? La proportion calculée dans la partie I/ pour Aamjiwnaang se situe-t-elle entre ces deux droites? La situation d'Aamjiwnaang vous paraît-elle normale ou alarmante?

TP d'échantillonnage : Un point dans un carré

I/ Explications du sujet

Soit un disque de rayon 1.
On considère le quart supérieur droit du disque.
On place aléatoirement un point $M(x; y)$ dans ce carré.
Soit il se trouve dans le secteur circulaire hachuré, soit il appartient au morceau blanc du carré restant.
On va estimer la probabilité que ce point M soit situé dans le secteur circulaire hachuré du carré.



II/ Simulation de l'expérience aléatoire

A partir du logiciel Albox, vous allez simuler le choix aléatoire de n points dans ce carré.

- la fonction `random()` permet d'attribuer au point M des coordonnées aléatoires dans l'intervalle $[0; 1[$
- on calcule la distance du point M à l'origine du repère
- si le point est dans la zone hachurée, alors on compte le point
- la fréquence correspond au nombre de points dans la zone hachurée par rapport au nombre total de points considérés.

Codez cet algorithme.

CR : Vous expliquerez comment la formule (2) a été obtenue.
Pourquoi a-t-on choisi la condition "distance ≤ 1 " pour comptabiliser les points?

```

DEBUT_ALGORITHME
  compteur PREND_LA_VALEUR 0
  n PREND_LA_VALEUR 100
  POUR i ALLANT_DE 1 A n
    DEBUT_POUR
      (1) x PREND_LA_VALEUR random()
      (2) y PREND_LA_VALEUR random()
      (2) distance PREND_LA_VALEUR sqrt(x*x+y*y)
      SI (distance <= 1) ALORS
        (3) DEBUT_SI
              compteur PREND_LA_VALEUR compteur+1
            FIN_SI
        (4) FIN_POUR
      fréquence PREND_LA_VALEUR compteur/n
    AFFICHER fréquence
  FIN_ALGORITHME
  
```

III/ Étude d'échantillons de taille $n = 100$

Exécutez le programme et notez la fréquence obtenue. Répétez cette opération plusieurs fois de façon à obtenir 20 fréquences.

Reportez ces valeurs dans le fichier "Un point dans un carré.ods / feuille n = 100". Celles-ci s'inscrivent automatiquement dans le diagramme.

Remarque : Attention, sur ce logiciel, un nombre décimal s'écrit avec un virgule et non un point.

Sur ce diagramme apparaissent aussi trois valeurs de référence.

La centrale, en rouge, est notée p dont on calculera la valeur exacte plus tard.

Elle est encadrée par les deux vertes, égales à $p \pm 0,1 = p \pm \frac{1}{\sqrt{n}}$.

Copiez le graphique dans un fichier de traitement de texte (que vous imprimerez avec le graphique de la partie suivante).

CR : Combien de points, sur les 20 enregistrées, se situent dans la zone encadrée par les barres vertes, c'est-à-dire dans l'intervalle $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$? Quel pourcentage cela représente-t-il?

IV/ Étude d'échantillons de taille $n = 10\,000$

Reprenez l'étude précédente avec $n = 10\,000$, en utilisant l'autre feuille de calcul "feuille n = 10 000".
Attention : n'oubliez pas de modifier l'algorithme afin qu'il traite désormais 10 000 points.

Ici, l'encadrement change aussi : $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right] = [p - 0,01 ; p + 0,01]$.

TSVP

Seconde 07-09 - Année Scolaire 2009 - 2010

Chapitre n°3 : Notions d'échantillonnage ;

Copiez le nouveau graphique sur la même feuille de traitement de texte que précédemment, puis imprimez-la.

CR : Combien de points, sur les 20 enregistrées, se situent dans la zone encadrée par les barres vertes, c'est-à-dire dans l'intervalle $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$? Quel pourcentage cela représente-t-il ?

V/ Le modèle théorique

On considère un point $M(x ; y)$ situé dans le carré ci-contre.

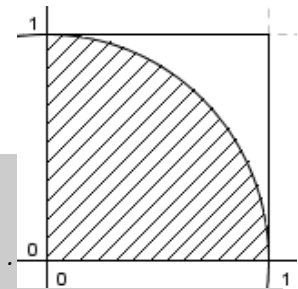
La probabilité que ce point se situe, ou non, dans le secteur circulaire est déterminée par l'aire de chaque régionnement du carré.

CR : Calculer l'aire $\mathcal{A}_1$ du carré. Ce carré représente alors l'univers.

Calculer l'aire $\mathcal{A}_2$ du secteur circulaire hachuré.

La probabilité que le point M soit dans la zone hachurée est égale au rapport $\frac{\mathcal{A}_2}{\mathcal{A}_1}$.

Quel lien faites-vous avec le nombre p ?



CR : Ce procédé, dit probabiliste, est appelé méthode de Monte-Carlo. Expliquez pourquoi celui-ci permet de donner une valeur approchée assez précise de la constante π .