Échantillonnage.

I Intervalle de fluctuation.

Considérons l'expérience aléatoire consistant en un pile-ou-face avec une pièce équilibrée.

Si je joue une fois et que j'obtiens Pile puis-je en déduire que le résultat de cette expérience est toujours Pile ? Non. Pour analyser le résultat d'une expérience aléatoire il faut la recommencer n fois, le nombre n restant à fixer.

Les n résultats successivement obtenus en répétant l'expérience à l'identique et de façon indépendante, constituent un *échantillon de taille n*.

Si je dois recommencer les n expériences j'obtiendrai un échantillon différent. Ce phénomène est appelé la *fluctuation d'échantillonnage*.

Exercice 1

L'algorithme ci-contre modélise des lancers de pièces. *La fonction* Aleat$(0;1)$ *désigne un choix équiprobable entre le nombre* 0 *et le nombre* 1.

1. Programmez cet algorithme sur votre Ti82.
2. Les deux variables qui apparaissent dans ce programme ont un rôle de compteur. Précisez ce qu'elle comptent.
3. Interprétez N et C sachant qu'il s'agit de modéliser des lancers de pièces.
4. Quel est l'échantillon dans cet expérience ? Quel est sa taille ?
5. Nous nous intéressons dorénavant au fluctuations de cet échantillonnage. Relancez plusieurs fois le programme. Obtenez-vous 50 ou 0 ? Quelle valeur peut-on espérer ?

```
Entrée
    N ← 0
    C ← 0

Traitement
    Tant que N < 50 faire
        R ← Aleat(0;1)
        Si R = 1 alors
            C ← C + 1
        FinSi
        N ← N + 1
    Fin Tant que

Sortie
    Afficher C
```

6. Modifiez la sortie comme ci-contre.
   (a) Relancez le programme dix fois et notez les résultats obtenus.
   (b) Comment interpréter ce résultat ?

Sortie
  Si $0{,}5 - \frac{1}{\sqrt{50}} \leq \frac{C}{50}$ et $\frac{C}{50} \leq 0{,}5 + \frac{1}{\sqrt{50}}$, alors
    Afficher « Dans l'intervalle »
    Sinon
    Afficher « Hors de l'intervalle »
  FinSi

Proposition 1

Soit p la probabilité d'un événement d'une expérience aléatoire avec $0{,}2 \leq p \leq 0{,}8$.
Lorsque l'expérience est réalisée et répétée $n \geq 25$ fois (à l'identique et de façon indépendante), il y a 95 % de chances que la fréquence d'apparition de l'événement soit dans l'intervalle

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

Remarques

1. La démonstration de ce résultat sera vu l'année prochaine avec la loi de probabilité dite binomiale.
2. Autrement dit il y a 5 % de chance que la fréquence d'apparitions de l'événement ne soit pas dans cet intervalle.
3. Ce résultat permet de tester la cohérence entre la fréquence d'apparition d'un résultat (« obtenir Pile ») lorsqu'une expérience aléatoire est renouvelée n fois et la probabilité théorique ($p = \frac{1}{2}$).
   Si la fréquence d'apparition f n'est pas dans l'intervalle c'est qu'il y a une erreur : soit p est faux, soit l'échantillon a été mal choisi.
4. L'intervalle est appelé l'*intervalle de fluctuation des fréquences au seuil de 95%*.

Exercice 2

Un slogan publicitaire affirme : « 55% des personnes ayant suivi la méthode ARETABA ont arrêté de fumer ! ». Dans un échantillon de 700 personnes ayant suivi cette

méthode, prises au hasard avec remise, 382 ont arrêté de fumer.
À partir de cet échantillon, le slogan publicitaire est-il acceptable, au seuil de 5% ?

Exercice 3 pour s'entraîner.

Exercice 18 page 35 (Sesamath). Exercice corrigé

Exercice 4 pour s'entraîner.

Exercice 19 page 35 (Sesamath). Échantillon représentatif ?

Exercice 5 pour s'entraîner.

Exercice 23 page 36 (Sesamath).

Exercice 6 pour s'entraîner.

Exercice 25 page 36 (Sesamath). Vérifier une modélisation probabiliste.

II Intervalle de confiance.

L'intervalle de fluctuation permet de mesurer la dispersion des valeurs de l'échantillon autour d'une valeur théorique (proportion ou probabilité).

Cependant lorsqu'un sondage avec une question fermée est effectué, nous ne savons pas à priori la proportion de personnes qui vont répondre favorablement. Et le résultat du sondage ne peut-être considéré comme exact (transposable directement à la population totale) à cause du phénomène de fluctuation.

Nous pourrons uniquement donner un encadrement de la valeur théorique.

Exercice 7

Soient $p \in [0;1]$, $n \in \mathbb{N}^*$.
Démontrez :

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

Le précédent exercice montre qu'il est possible d'inférer de l'échantillon un encadrement de la valeur théorique p. Par exemple, dans le domaine des sondages, nous n'aurons plus uniquement le résultat du sondage mais aussi un encadrement de l'erreur ; ce que les sondeurs appellent des marges d'erreurs.

Proposition 2

Soit f la la fréquence d'apparition d'un événement d'une expérience aléatoire avec $0{,}2 \leq f \leq 0{,}8$.
Lorsque l'expérience est réalisée et répétée $n \geq 25$ fois (à l'identique et de façon indépendante), il y a 95 % de chances que la probabilité de l'événement soit dans l'intervalle

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

Exercice 8

Une entreprise de sondage interroge 1 200 personnes avec la question suivante : « Allez-vous voter pour Dugenoux à la prochaine élection ? »

571 personnes ont répondu oui.

Dugenoux peut-t-il être élu ?

III Exercices.

Exercice 9

Une maladie attaque une plante dans 70 % des cas lorsque cette plante n'a pas été traitée. On traite un échantillon de 150 de ces plantes et on constate alors que 28 % de ces plantes sont malades. Peut-on en conclure que le produit est efficace au seuil de 95 % ?

Exercice 10 pour s'entraîner.

Une entreprise fabrique des composants mécaniques en très grande quantité dans deux ateliers A et B. Le cahier des charges indique que l'entreprise tolère que 7 % des composants produits soient défectueux.

On accepte, dans cet exercice, que la proportion théorique p soit inférieure à 0,20.

Lors du contrôle hebdomadaire, on a testé 400 composants dans l'atelier A et 650 dans l'atelier B. Pour l'atelier A on a trouvé 33 composants défectueux et pour l'atelier B on en a trouvé 78.

Que peut-on en conclure concernant la production dans chacun des ateliers ?

Exercice 11

1. Rédigez un algorithme qui calcule les bornes de l'intervalle de fluctuation des fréquences.
2. Complétez l'algorithme de sorte qu'il vérifie les conditions de la proposition 1 avant de déterminer l'intervalle de fluctuation.
3. Complétez l'algorithme de sorte qu'il indique si la fréquence appartient ou non à l'intervalle de fluctuation des fréquences.

Exercice 12

On considère un caractère dont la proportion dans une population donnée est p avec $0{,}2 \leq p \leq 0{,}8$.

On prélève un échantillon de taille n de cette population et on note f la fréquence du caractère dans cet échantillon.

On note I l'intervalle $\left[p - \dfrac{1}{\sqrt{n}} ; p + \dfrac{1}{\sqrt{n}}\right]$.

1. Exprimer l'amplitude de l'intervalle I (distance entre ses deux bornes) en fonction de la taille n de l'échantillon.
2. Déterminer la plus petite valeur de n pour que l'amplitude de l'intervalle I soit
   (a) inférieure ou égale à 0,1 (10 %),
   (b) inférieure ou égale à 0,3 (30 %).
3. (a) Conjecturez le tableau de variation de $f : \left\{ \begin{array}{ccc}]0;+\infty[& \rightarrow & \mathbb{R} \\ x & \mapsto & \frac{2}{\sqrt{x}} \end{array} \right.$.
   (b) D'après le tableau de variation précédent, si on suppose $n \geq 25$, quelle est l'amplitude maximale de I.
4. On considère l'algorithme suivant :

Entrer un nombre a tel que $0 < a \leq 0{,}4$.
Calculer $x = \frac{4}{a^2}$.
Si x est un entier afficher x.
Sinon afficher l'entier suivant x.

   (a) Exécuter l'algorithme pour $a = 0{,}1$.
   (b) Quel est le but de cet algorithme ?
   (c) Programmer cet algorithme sur la calculatrice. *« x est un entier » si et seulement si « partie entière de x égale x ».*
   (d) Une entreprise de sondage souhaite savoir le nombre de personnes qu'elle doit interroger pour confirmer un résultat avec une marge d'erreur de 2,5 % puis 1 %. Utilisez le précédent programme pour répondre à cette question.

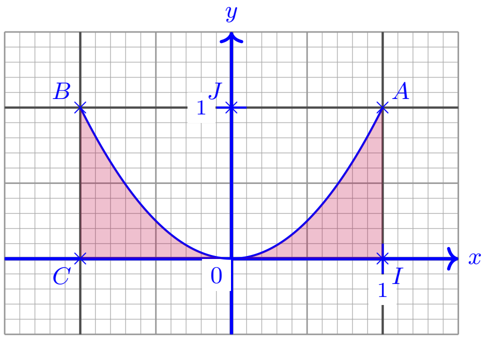
Exercice 13

Il s'agit d'un problème de quadrature, *i.e.* un problème de calcul d'aire.

On considère l'arc de parabole $\mathscr{P}$ qui est la courbe représentative de la fonction $f : \left\{ \begin{array}{ccc} [-1,1] & \rightarrow & \mathbb{R} \\ x & \mapsto & x^2 \end{array} \right.$ dans un repère orthonormé (O,I,J).

Notons $A(1;1)$, $B(-1;1)$ et $C(-1;0)$.

L'objectif de cet exercice est d'obtenir une valeur approchée de l'aire comprise entre $\mathscr{P}$ et l'axe des abscisses (partie colorée sur la figure ci-dessous).



Nous allons faire apparaître 100 points $M(x;y)$ au hasard dans le rectangle $IAJO$ et regarder la proportion de points qui se trouvent dans la partie colorée. Le point $M(x;y)$ est dans la partie colorée si et seulement si $y < x^2$.

1. Justifiez que $A \in \mathscr{P}$ et $B \in \mathscr{P}$.
2. Nous utilisons le programme suivant (rédigé en python) qui crée et teste 100 points dans le carré $IAJO$.

```
import random

def point():
  if random.random()<random.random()**2:
    return(1)
  else:
    return(0)

n=0
sous=0
while n<100:
  sous=sous+point()
  n=n+1
print(sous)
```

Remarque : random.random() est un nombre aléatoire entre 0 et 1.

   (a) Dans quel cas la fonction point() renvoie-t-elle 1 ? 0 ? Comment l'interpréter pour le graphique ?
   (b) Combien vaut n à la fin de l'algorithme ? Qu'est-ce que n ?
   (c) Que désigne la variable *sous* ?
   (d) Programmer ce programme sur votre calculatrice. Notez et interprétez ce que renvoie ce programme.
3. Le précédent programme génère un échantillon de taille 100. Déterminez l'intervalle de confiance correspondant.

4. Interprétez l'intervalle de confiance précédemment déterminé.
5. Modifiez le programme pour obtenir un échantillon de taille 1 000 et déterminez l'intervalle de confiance correspondant.
6. Répondez à la question initiale de cet exercice.

Exercice 14