9.1 Intervalle de fluctuation

9.1.1 Du vocabulaire

DÉFINITION
En statistique, un échantillon de taille n est la liste des résultats obtenus par n répétitions indépendantes de la même expérience.

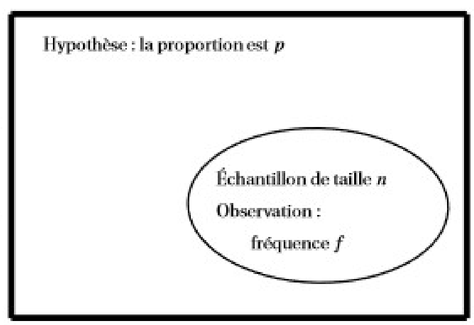
Exemples et remarques :

- Par exemple, un échantillon de taille 100 du lancer de la même pièce est la liste des résultats pile ou face obtenus successivement en répétant 100 fois le lancer de cette pièce.
- Un échantillon de taille n est constitué des résultats de n tirages successifs avec remise.
- Un prélèvement sans remise peut être envisagé dans une prise d'échantillon lorsque la taille de la population est jugée suffisamment grande par rapport à celle de l'échantillon.
  Par exemple, dans la population française, on sélectionne au hasard un échantillon de taille 1000 sans interroger deux fois la même personne. Il n'y a donc pas de tirage avec remise. Par contre, la taille de l'échantillon est suffisamment petite par rapport à celle de la population. On peut donc apparenter ce modèle à un tirage avec remise.
- Dans tous les cas, il y a répétitions d'une même épreuve à deux issues possibles (épreuve de Bernoulli) de façon indépendante.
  On dit qu'un tel échantillon relève du **modèle de Bernoulli**.

On s'intéresse à un caractère de proportion p dans une population donnée.
On considère la variable aléatoire X_n qui, à chaque échantillon aléatoire de taille n de cette population, associe le nombre d'individus de cet échantillon ayant ce caractère et la variable aléatoire $F_n = \frac{X_n}{n}$ qui, à ce même échantillon, associe la fréquence du caractère étudié dans cet échantillon.

DÉFINITION
On appelle intervalle de fluctuation de F_n au seuil de 95 %, tout intervalle $[\alpha\ ;\ \beta]$ tel que $p(F_n \in [\alpha\ ;\ \beta]) \geqslant 0,95$.

9.1.2 En première



Situation :
Dans une population, on suppose que la proportion d'un certain caractère est p.
On souhaite accepter ou refuser cette hypothèse.
On prélève dans la population au hasard et avec remise un échantillon de taille n sur lequel on calcule la fréquence f observée du caractère.

Démarche :
La loi binomiale de paramètres n et p permet de trouver **l'intervalle de fluctuation I au seuil de 95 %** de la fréquence observée :

$$\mathrm{I} = \left[\frac{a}{n}\ ;\ \frac{b}{n}\right] \text{ avec } \begin{cases} a \text{ est le plus petit entier tel que } P(X \leqslant a) > 0,025 \\ b \text{ est le plus petit entier tel que } P(X \leqslant b) \geqslant 0,975 \end{cases}$$

Règle de prise de décision :

- Si $f \in \mathrm{I}$, au seuil de 95 %, on accepte l'hypothèse que la proportion dans la population est p ;
- Sinon, on rejette l'hypothèse selon laquelle la proportion vaut p.
  La probabilité de rejeter l'hypothèse alors qu'elle est vraie est inférieure à 5 %.
  En encore :
  On choisit donc de fixer le seuil de décision de telle sorte que la probabilité de rejeter l'hypothèse, alors qu'elle est vraie, soit inférieure à 5 %

Remarques :

- Ce nouvel intervalle de fluctuation fonctionne pour toutes les valeurs de n et p.
- Sous les conditions $n > 25$ et $0,2 < p < 0,8$, l'intervalle $\left[\frac{a}{n}\ ;\ \frac{b}{n}\right]$ est « quasiment » le même que l'intervalle $\left[p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}}\right]$.

9.1.3 Un exemple

Un médecin de santé publique veut savoir si, dans sa région , le pourcentage d'habitants atteints d'hypertension artérielle est égal à la valeur de 16 % récemment publiée pour des populations semblables.Le médecin formule l'hypothèse $p = 0,16$. Pour vérifier cette hypothèse, le médecin constitue un échantillon de 100 habitants de la région. 8 habitants de cet échantillon sont atteints d'hypertension artérielle.
L'échantillon est prélevé au hasard et la population est suffisamment importante pour considérer qu'il s'agit de tirages avec remise.

1. Soit X la variable aléatoire dénombrant le nombre d'hypertendus observés dans un échantillon aléatoire de taille n. Quelle loi est suivie par X ?
2. En utilisant le tableau des $P(X \leqslant k)$ fourni ci-contre, déterminer les plus petits entiers a et b tels que $P(X \leqslant a) > 0,025$ et $P(X \leqslant b) \geqslant 0,975$.
3. En déduire l'intervalle de fluctuation.
4. Quelles sont les conclusions du médecin ?

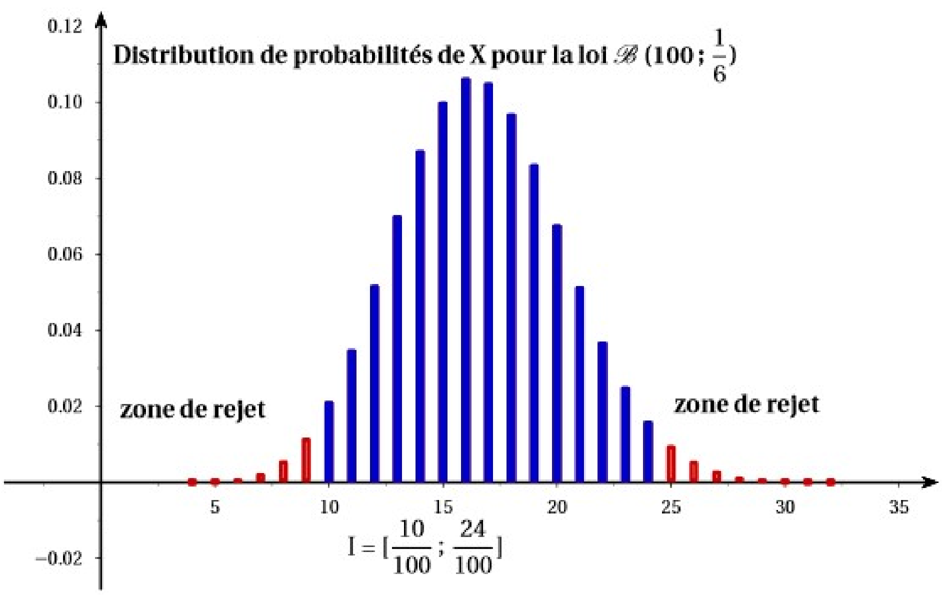
	A	B
1	k	P(x<=k)
2	6	0.0022
3	7	0.0061
4	8	0.01474
5	9	0.03156
6	10	0.06071
7	11	0.10614
8	12	0.17032
9	13	0.25306
10	14	0.35101
11	15	0.45798
12	16	0.56621
13	17	0.66809
14	18	0.75756
15	19	0.83111
16	20	0.88785
17	21	0.92902
18	22	0.95719
19	23	0.97538
20	24	0.98649
21	25	0.99293
22	26	0.99647
23	27	0.99831
24	28	0.99923

9.1.4 Une interprétation graphique

On fait l'hypothèse que la proportion d'un caractère C dans une population donnée est p. Après expérience on observe que la fréquence de ce même caractère dans un échantillon de taille n est f.
Soit I l'intervalle de fluctuation au seuil de 95% de la fréquence du caractère dans un échantillon de taille n, on propose un critère pour infirmer ou non l'hypothèse de départ.

• Si $f \notin$ I alors on rejette l'hypothèse que la proportion du caractère dans la population est p.

• Si $f \in$ I alors , au seuil de 95 %, on accepte l'hypothèse que la proportion du caractère dans la population est p.



9.1.5 Un exercice d'application directe

EXERCICE 1 (Politique à Tamalou)

Mr Toupourmoa, chef du pays imaginaire mais hypothétique Tamalou, affirme que 52 % des électeurs soutiennent sa politique. Un organisme international interroge 100 électeurs au hasard et de façon indépendante (la population étant suffisamment grande pour considérer qu'il s'agit de tirages avec remise) et il souhaite savoir , au seuil de 95 %, si on peut mettre en doute le pourcentage avancé par Mr Toupourmoa.

1. Quelle hypothèse pouvons-nous poser ?
2. Soit X la variable aléatoire correspondant au nombre d'électeurs lui faisant confiance dans un échantillon de taille 100. En justifiant, déterminer la loi suivie par X.
3. Déterminer a et b tels que :
$$\begin{cases} a \text{ est le plus petit entier tel que } P(X \leqslant a) > 0,025 \\ b \text{ est le plus petit entier tel que } P(X \leqslant b) \geqslant 0,975 \end{cases}$$
4. En déduire l'intervalle de fluctuation au seuil de 95 % obtenu à l'aide de la loi binomiale.
5. Comparer cet intervalle avec l'intervalle de fluctuation au seuil de 95 % vu en seconde.
6. Énoncer la règle de décision permettant de rejeter ou non l'hypothèse émise à la question 1) selon la valeur de la fréquence f observée sur l'échantillon prélevé.
7. Sur les 100 électeurs interrogés au hasard, 43 déclarent soutenir la politique de Mr Toupourmoa. Pouvez-vous considérer, au seuil de 95 %, l'affirmation de Mr Toupourmoa comme exacte ?

k	p(X<=k)
39	0,0061
40	0,0106
41	0,0177
42	0,0286
43	0,0444
44	0,0667
45	0,0967
46	0,1355
47	0,1838
48	0,2417
49	0,3082
50	0,3816
51	0,4596
52	0,5393
53	0,6174
54	0,6911
55	0,7579
56	0,8160
57	0,8646
58	0,9037
59	0,9338
60	0,9561
61	0,9719
62	0,9827
63	0,9897

9.1.6 Représentativité d'un échantillon

Une autre question se pose :
Pour une population donnée dont on connaît des proportions de caractères, un échantillon sélectionné est-il représentatif de la population ?
Pour ce faire, on adopte la démarche suivante :

1. Sur l'échantillon, on observe une fréquence d'un caractère connu de la population
2. On calcule un intervalle de fluctuation au seuil de 95 % .
3. On décide oui ou non de la représentativité de cet échantillon au seuil de 95 %.

Remarque : La notion d'échantillon représentatif dépasse largement les propos de ce cours.
Cependant, il ne faut pas oublier que cette question délicate est fondamentale dans le cadre d'un sondage. En effet, l'échantillon concerne des personnes.
Par contre, lorsqu'il s'agit d'un échantillon de pièces dans une chaîne de production, la question de représentativité d'un échantillon sera moins primordiale.
Dans tous les cas, pour tester la représentativité d'un échantillon, on aura recours à des intervalles de fluctuation effectués sur différents caractères connus de la population.

EXERCICE 2

Avant de procéder à un sondage sur une commune, une agence spécialisée dans le sondage effectue un test de représentativité d'un échantillon.
On sait que dans cette commune de 5 000 habitants, environ 55% sont originaires de la commune ou de ses environs proches.
On interroge quatre cents habitants de la commune choisis au hasard, 250 habitants sont originaires de la commune ou de ses environs proches.
L'échantillon sélectionné est-il représentatif de cette commune au seuil de 95% ?
Pour déterminer un intervalle de fluctuation, on donne les tableaux suivants :

k	197	198	199	200	201	202
$p(X \leqslant k)$	0,012	0,0155	0,0199	0,0252	0,0317	0,0395

k	237	238	239	240	241	242
$p(X \leqslant k)$	0,9611	0,9689	0,9753	0,9806	0,9849	0,9884