ÉCHANTILLONNAGE

Cours	**Première S**

Dans une population on sait que le caractère C est présent avec la proportion p (on pourrait dire que la probabilité de choisir un individu ayant le caractère C est p).

On prélève un échantillon de taille n dans cette population.

Deux questions peuvent se poser :

• Que peut-on dire de la fréquence f du caractère C dans cet échantillon ?

• Si on peut calculer cette fréquence f, peut-on affirmer que cet échantillon est représentatif de la population ?

1. Rappel sur l'intervalle de fluctuation

On a déterminé en classe de seconde un intervalle de fluctuation d'une proportion d'un caractère à 95 %.

Propriété 1 (vue en 2^{nde}) : Si p est la proportion d'un caractère dans une population, avec $0,2 \leq p \leq 0,8$ alors pour un échantillon de taille $n \geq 25$, la fréquence f du caractère dans l'échantillon appartient à l'intervalle $\left[p-\frac{1}{\sqrt{n}}\ ;\ p+\frac{1}{\sqrt{n}}\right]$ avec une probabilité d'au moins 95 %.

Exemple : La proportion de personnes ayant les yeux marrons dans la population française est 0,34. On prélève un échantillon de 100 enfants.
a) Donner un intervalle donnant la fréquence probable au seuil de 95%
b) Quel est le nombre de personnes aux yeux marrons probables dans cet échantillon ?

a) Un intervalle donnant la fréquence probable au seuil de 95% est
$\left[0,34-\frac{1}{\sqrt{100}}\ ;\ 0,34+\frac{1}{\sqrt{100}}\right]=[0,24\ ;\ 0,44]$.

b) On peut donc estimer que dans cet échantillon on doit avoir au seuil de 95% entre 24 et 44 personnes ayant les yeux marrons.

2. Intervalle de fluctuation à 95 %

On va améliorer le résultat précédent en utilisant une loi binomiale.

1) Propriété

Propriété 2 : On considère une population dont une proportion p des individus possède un caractère donné.
On prélève dans cette population un échantillon de taille n.
La variable aléatoire qui associe le nombre d'individus possédant ce caractère suit une loi binomiale de paramètres n et p.

Exemple : On définit la variable aléatoire X égale au nombre de personnes ayant les yeux marrons dans un échantillon de taille 100. En assimilant le choix d'une personne au hasard dans cet échantillon à un tirage avec remise, on peut supposer que X suit une loi binomiale

de paramètres $n = 100$ et $p = 0{,}34$.
X prend alors les valeurs de 0 à 100.

k	*p(X=k)*
0	9,00313E-19
1	4,63798E-17
2	1,18268E-15
3	1,99025E-14
4	2,48631E-13
5	2,45919E-12
6	2,00585E-11
7	1,3876E-10
8	8,30982E-10
9	4,37594E-09
10	2,05139E-08

On peut dresser ce tableau de valeurs (ici les 11 premières) en utilisant la calculatrice ou un tableur.

Avec la calculatrice, utilisez les tutoriels du site Garomaths.

Avec le tableur :
LibreOffice (ou OpenOffice) :

	LibreOffice	**Excel**
$p(X = k)$	=LOI.BINOMIALE (k ; n ; p ; 0)	=LOI.BINOMIALE (k ; n ; p ; FAUX)
$p(X \leq k)$	=LOI.BINOMIALE (k ; n ; p ; 1)	=LOI.BINOMIALE (k ; n ; p ; VRAI)

2) Intervalle de fluctuation au seul de 95 %

Définition : On considère une population dont une proportion p des individus possède un caractère donné.
On prélève dans cette population un échantillon de taille n.
Soit X la variable aléatoire associée au nombre d'individus possédant ce caractère.
L'intervalle de fluctuation au seuil de 95 % associée à la variable aléatoire X est l'intervalle $\left[\frac{a}{n} \;;\; \frac{b}{n}\right]$ défini par :

- a est le plus petit entier tel que : $p(X \leq a) > 0{,}025$
- b est le plus petit entier tel que : $p(X \leq b) \geq 0{,}975$

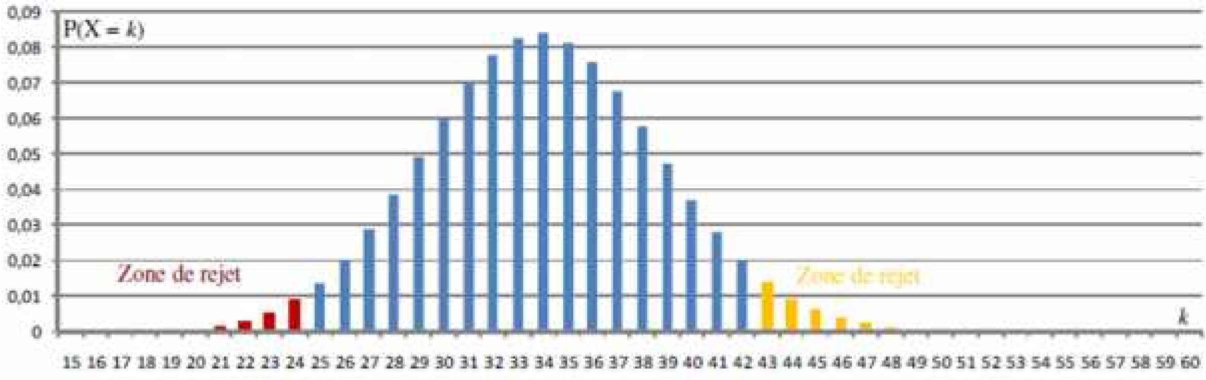
Remarques : • On obtient alors que $p(a \leq X \leq k) \geq 0{,}95$.
L'intervalle de fluctuation est aussi appelé « **intervalle de confiance** ».

• Dire que l'intervalle de fluctuation au seuil de 95 % est $\left[\frac{a}{n} \;;\; \frac{b}{n}\right]$ signifie que pour un échantillon de n personnes, il y a au moins 95 % de chance qu'il y ait entre $100\frac{a}{n}$ % et $100\frac{b}{n}$ % des individus qui possèdent le caractère donné.

Exemple : Reprenons l'exemple précédent cumulant les probabilités.

On remarque que $p(X \leq k) > 0{,}025$ lorsque $k \geq 25$, et $p(X \leq k) > 0{,}975$ lorsque $k \geq 43$.
Par suite, $a = 25$ et $b = 43$.
Par conséquent, l'intervalle de fluctuation au seuil de 95 % est $\left[\frac{25}{100} \;;\; \frac{43}{100}\right] = [0{,}25 \;;\; 0{,}43]$.

Pour un échantillon de 100 personnes, il y a au moins 95 % de chance qu'il y ait entre 25 % et 43 % des personnes ayant les yeux marrons.



Sur le graphique, la plus petite somme des probabilités supérieure ou égale à 95 % est représentée en bleue.

Remarque : On remarque que l'intervalle $[0,25 \,;\, 0,43]$ est inclus dans celui trouvé en seconde $[0,24 \,;\, 0,44]$, et que les résultats obtenus sont très voisins.
Cela conforte le résultat vu en seconde plus simple à utiliser mais plus approximatif.

3. Prise de décision à partir d'un échantillon

Règle de décision : On considère une population dont une proportion p des individus possède un caractère donné.
Après expérience, on observe f comme fréquence de ce caractère dans un échantillon de taille n.
Si f n'appartient pas à l'intervalle de fluctuation $\left[\frac{a}{n} \,;\, \frac{b}{n}\right]$, on rejette l'hypothèse au seuil de risque 5 %. Dans le cas contraire, on ne rejette pas cette hypothèse au seuil 5 %.

Remarque : « *au risque de 5 %* » signifie que la probabilité de rejeter l'hypothèse, alors qu'elle est vraie, est inférieure à 5 %.

Exemple : *Un laboratoire annonce qu'un médicament sauve 40 % des patients atteint d'une maladie rare.*
Pour contrôler cette affirmation, l'autorité de surveillance demande le test sur 100 patients atteints de cette maladie.
Soit X le nombre de malades sauvés par ce médicament dans un échantillon aléatoire de malades et assimilé à un tirage avec remise de taille 100.
1) Quelle loi suit X ?
2) Déterminer les plus petits entiers a et b tels que P (X ≤ a) > 0,025 et P (X ≤ b) ≥ 0,975.
3) Énoncer la règle de décision permettant de rejeter ou non l'hypothèse p = 0,40 selon la valeur de la fréquence f des malades sauvés dans l'échantillon.
4) Sur les 100 malades auxquels on a administré ce traitement, on en a sauvé 30.
Au seuil de risque 5 %, que peut-on dire de l'annonce faite par le laboratoire ?

1) L'épreuve de Bernoulli « un patient est sauvé ou non par le médicament » est répétée 100 fois de manière identique et indépendante, donc X suit la loi binomiale de paramètres $n = 100$ et $p = 0,4$.

2) À l'aide de la calculatrice ou d'un tableur, on trouve que $a = 31$ et $b = 50$.

3) L'intervalle de fluctuation à 95 % de la fréquence dans les échantillons de taille 100 est

$\left[0,31\,;\,0,5\right]$.

On en déduit la règle de décision : si f appartient à l'intervalle $\left[0,31\,;\,0,5\right]$, l'hypothèse $p=0,4$ est acceptable, sinon elle est rejetée, au seuil de risque 5 %.

4) La fréquence observée est de $\frac{30}{100}=0,3$, et $0,3\notin\left[0,31\,;\,0,5\right]$, donc, au seuil de risque 5 %, on rejette l'hypothèse selon laquelle le médicament sauve 40 % des malades.