I - Intervalle de fluctuation

1) Retour sur une propriété vue en Seconde.

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

Exemple :

> On sait que 26% des Français sont allergiques aux pollens des fleurs. On décide d'interroger un échantillon de 400 personnes : on trouve 120 personnes allergiques.
> **Question :** Le nombre d'allergiques dans cet échantillon est-il anormal ?

Solution :

2) Lien avec la loi binomiale.

L'intervalle de fluctuation précédent peut être amélioré à l'aide de la loi binomiale. En effet, si on prélève un échantillon aléatoire de taille n et qu'on appelle X la variable aléatoire donnant le nombre d'individus de cet échantillon ayant ce caractère, alors il est clair que X suit une loi binomiale de paramètres n la taille de l'échantillon et p la probabilité que l'individu ait le caractère étudié.
On peut alors en déduire le résultat suivant :

Définition : On s'intéresse à un caractère de proportion p dans une population.
On prélève un échantillon de taille n et on considère la variable aléatoire X donnant le nombre d'individus de cet échantillon ayant ce caractère.
Alors X suit une loi binomiale de paramètres n et p.
L'intervalle de fluctuation à 95% de la fréquence correspondant à X est l'intervalle $\left[\dfrac{a}{n} ; \dfrac{b}{n}\right]$ où :
a est le **plus petit** entier tel que $P(X \leqslant a) > 0,025$.
b est le **plus petit** entier tel que $P(X \leqslant b) \geqslant 0,975$.

Exemple :

> La proportion de personnes ayant les yeux marron dans la population française est 0,34. Déterminer un intervalle de fluctuation de la fréquence f des personnes ayant les yeux marron dans un échantillon de taille 100.

Solution :
Soit X la variable aléatoire égale .... dans un échantillon de taille ....
Alors, en assimilant le choix d'une personne au hasard dans un échantillon à un tirage avec remise, on peut supposer que X ....

Les valeurs prises par X sont donc tous les entiers compris entre .... et ....

Pour déterminer toutes les probabilités de la forme $P(X = k)$ avec k compris entre .... et ...., on utilise le tableur :

	A k	B pk	C sommeproba	D	E	F	G	H
◆	=seq(n,n,0,100)	=binompdf(100,0.34,a[])						
1	0	9.00313068484E-19	9.00313068484E-19					
2	1	4.6379764134E-17	4.72800772025E-17					
3	2	1.18268398542E-15	1.22996406262E-15					
4	3	1.99025406031E-14	2.11325046657E-14					
5	4	2.48630980716E-13	2.69763485382E-13					
6	5	2.45918642744E-12	2.72894991283E-12					
7	6	2.00585155572E-11	2.278746547E-11					
8	7	1.38759774287E-10	1.61547239757E-10					
9	8	8.30981830108E-10	9.92529069865E-10					
10	9	4.37594472151E-9	5.36847379138E-9					
11	10	2.05138984369E-8	2.58823722283E-8					
12	11	8.64635388663E-8	1.12345911095E-7					
13	12	3.30351854305E-7	4.42697765399E-7					
14	13	1.15199620988E-6	1.59469397528E-6					
15	14	3.6878839706E-6	5.28257794588E-6					
16	15	1.08922956869E-5	1.61748736328E-5					
17	16	2.98094077037E-5	4.59842813365E-5					
18	17	7.58784923368E-5	1.21862773673E-4					
19	18	1.80243354692E-4	3.02106128366E-4					
20	19	4.00732434515E-4	7.02838562881E-4					
21	20	8.36073579284E-4	0.001538912142					
F32								

Colonne A : ....
Colonne B : ....
Colonne C : ....

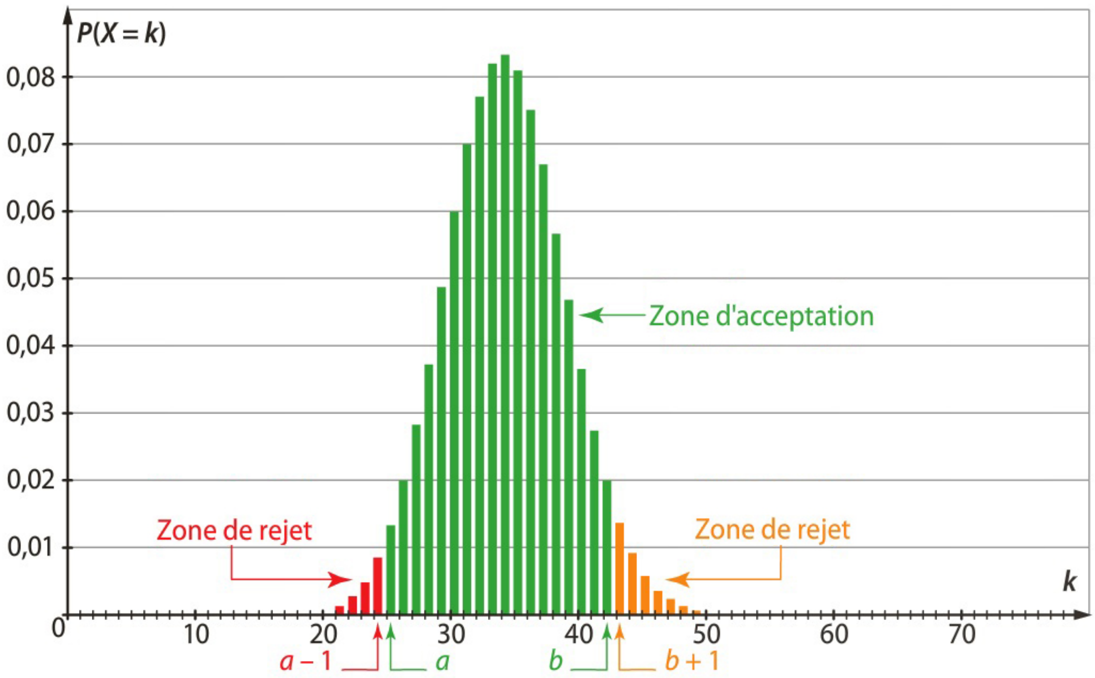
Parmi ces valeurs, on cherche donc deux entiers a et b tels que la probabilité que X soit dans l'intervalle $[a; b]$ est au moins égale à ....

a est la première valeur de k pour laquelle la somme des probabilités dépasse ...., donc $a =$ ....

b est la première valeur de k pour laquelle la somme des probabilités dépasse ....„ donc $b =$ ....

Conclusion :
La fréquence f appartient à l'intervalle [.........; ....] avec une probabilité au moins égale à ....

On peut également réaliser le diagramme en bâtons correspondant :



II - Prise de décision

La détermination d'un intervalle de fluctuation permet de prendre une décision lorsque l'on fait une hypothèse sur une proportion dans une population.

En effet, en faisant une hypothèse sur la proportion p d'un caractère dans une population, on peut déterminer un intervalle de fluctuation I à 95% de la fréquence de ce caractère dans un échantillon de taille n.

Ainsi, on établira une règle de décision : si la fréquence observée dans l'échantillon n'appartient pas à I, comme cela n'a qu'une probabilité de 0,05 de se produire, alors on rejettera l'hypothèse faite sur p, avec un risque de se tromper de 5%.

On en déduit la propriété suivante :

Propriété 2 :
On considère une population dans laquelle on **suppose** que la proportion d'un caractère est p. Après expérience, on **observe** f comme fréquence de ce caractère dans un échantillon de taille n.

Soit l'hypothèse : « La proportion de ce caractère dans la population est p ».

Si I est l'intervalle de fluctuation à 95% de dans un échantillon de taille n, alors :

Si $f \notin I$: on rejette cette hypothèse au seuil de risque de 5%.

Sinon, on ne rejette pas cette hypothèse au seuil de risque de 5%.

Exemple :

Dans un article, un journaliste affirme que 42% des jeunes qui aiment la lecture préfèrent lire le soir.
On interroge au hasard 140 jeunes qui aiment lire et on assimile le sondage à un tirage successif avec remise : on observe que 49 jeunes déclarent lire le soir.
Peut-on mettre en doute, avec un risque de se tromper de 5%, l'affirmation du journaliste ?

Solution :