I. Contexte :

Dans une population, on veut **vérifier** que la proportion d'un caractère est égale à p.

Exemple :

On considère qu'une machine à former des gélules fonctionne de façon satisfaisante si le pourcentage de gélules défectueuses est de 1 pour 1 000. On veut savoir si la machine est bien réglée.

Pour cela, on prélève un échantillon. Sur un échantillon de 5 000 gélules, on a trouvé 9 gélules défectueuses.

Peut-on en penser que la machine est bien réglée ?

On va voir par la suite comment répondre à cette question.

Dans notre exemple :

- la taille de l'échantillon est : $n = 5\ 000$
- la proportion à valider est : $p = \dfrac{1}{1000} = 0,001$
- la fréquence observée est : $f = \dfrac{9}{5000}$ notée aussi f_n

II. Procédé :

1. A partir de la proportion p supposée et à valider, on définit un **intervalle de fluctuation asymptotique** I qui dépend de p et de la taille n de l'échantillon, si les conditions suivantes sont vérifiées :

$$\boxed{n \geqslant 30} \qquad \boxed{np \geqslant 5} \qquad \boxed{n(1-p) \geqslant 5}$$

Cet intervalle est défini pour un **seuil de confiance** de $1-\alpha$ ou un **seuil de risque** de α.

Soit $I_n = \left[p - u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$

L'intervalle I_n contient la fréquence f_n obtenue sur un échantillon de taille n avec une probabilité qui tend vers $1-\alpha$ quand n tend vers $+\infty$.

L'intervalle I_n est appelé **intervalle de fluctuation asymptotique de F_n au seuil de** $1-\alpha$

Remarques

- L'intervalle est centré en p
- L'intervalle à une amplitude de : $2 \times u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}$

2. Intervalle à connaître : l'intervalle de fluctuation asymptotique au seuil de 95 %

Rappel : $\boxed{u_{0,05} = 1,96}$ donc l'intervalle de fluctuation est : $\boxed{\left[p - 1,96\ \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1,96\ \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]}$

3. Interprétation d'un seuil de confiance de 95 % :

Si I est un **intervalle de fluctuation avec un seuil de confiance de 95 %**, cela signifie que si on étudie par exemple 10 000 échantillons, parmi les 10 000 fréquences obtenues, 95 % d'entres elles appartiennent à l'intervalle I.

Pour un échantillon étudié, la probabilité que la fréquence obtenue ne soit pas dans I étant très improbable, on décide de rejeter l'hypothèse de départ, à savoir la valeur de p supposée, si la fréquence n'est pas dans I.

Mais sachant que pour une valeur de p donnée, 5% des fréquences se situent en dehors de I, il y a donc un risque de 5 % de se tromper en rejetant l'hypothèse.

Exercice 1 Réponse au problème initial

On considère qu'une machine à former des gélules fonctionne de façon satisfaisante si le pourcentage de gélules défectueuses est de 1 pour 1 000. On veut savoir si la machine est bien réglée.

Sur un échantillon de 5 000 gélules, on a trouvé 9 gélules défectueuses.

Peut-on en penser que la machine est bien réglée ?

1. Déterminer un intervalle de fluctuation asymptotique au niveau de confiance 0,95 de la fréquence de pilules défectueuses dans les échantillons de taille 5 000, prélevés au hasard et avec remise.

   (On arrondira les bornes de l'intervalle à 10^{-4} près).

2. Enoncer la règle de décision permettant d'accepter ou non l'hypothèse $p = 0,001$ au seuil de confiance 0,95. Conclure

Exercice 2

Le fabricant d'un médicament breveté affirme qu'il est efficace à 90 % pour guérir une allergie en huit heures.

1. Déterminer un intervalle de fluctuation au niveau de confiance 0,95 de la fréquence des personnes guéries en 8 h dans un échantillon de taille 300. (On arrondira les bornes de l'intervalle à 0,01 près).

2. Sur un échantillon de 300 personnes, 250 personnes ont été guéries de cette allergie en 8 h : ce résultat remet-il ou non en cause l'affirmation du fabricant, au seuil de confiance 95 % ?

3. Quelle doit être la taille de l'échantillon choisi afin que l'amplitude de l'intervalle de fluctuation au seuil de confiance 0,90 soit inférieure à 0,01 ?