

CHAPITRE 10 : Echantillonnage - Estimation

1	Effectifs et fréquences avec la loi binomiale.....	2
2	La valeur de la proportion p dans la population est connue. On prédit un intervalle de fluctuation I_n contenant la fréquence F_n	5
2.1	Définition de l'intervalle de fluctuation à 95 % avec la loi binomiale	5
2.2	Définition de I_n l'intervalle de fluctuation à 95 % avec la loi normale	5
2.3	Intervalle I_n de fluctuation asymptotique au seuil de $(1-\alpha)\%$	9
	RÉSUMÉ : Échantillonnage dans une population où p est connue	11
3	La valeur de la proportion p dans la population est une hypothèse qu'on veut tester.....	12
4	La valeur de la proportion p dans la population est une inconnue qu'on estime par un intervalle de confiance I_c	14
	RÉSUMÉ : Estimation par un intervalle de confiance de la proportion p qui reste inconnue.....	18

CHAPITRE 10 : Echantillonnage - Estimation

1 Effectifs et fréquences avec la loi binomiale

Notation :

Prélever un échantillon de taille $n \in \mathbb{N}^*$ dans une population très nombreuse d'individus possédant un caractère considéré comme « succès » avec la probabilité p , revient à effectuer n épreuves de Bernoulli identiques et indépendantes.

- On note X_n la variable aléatoire qui donne le nombre k de succès dans l'échantillon :
 $X_n = k$, avec $k \in \mathbb{N}$ et $0 \leq k \leq n$.
- On note F_n la variable aléatoire qui donne la fréquence $\frac{k}{n}$ de succès dans l'échantillon :
 $F_n = \frac{k}{n}$ avec $k \in \mathbb{N}$ et $0 \leq \frac{k}{n} \leq 1$.

Exemple 1 : Fréquence F_n dans un échantillon

Soit une population d'individus telle que 5 personnes sur 100 aient la grippe.

- 1) On prend un échantillon de taille $n = 220$ personnes dans cette population. Déterminer la probabilité d'observer un nombre X_{220} de malades dans l'échantillon tel que $7 \leq X_{220} \leq 15$.

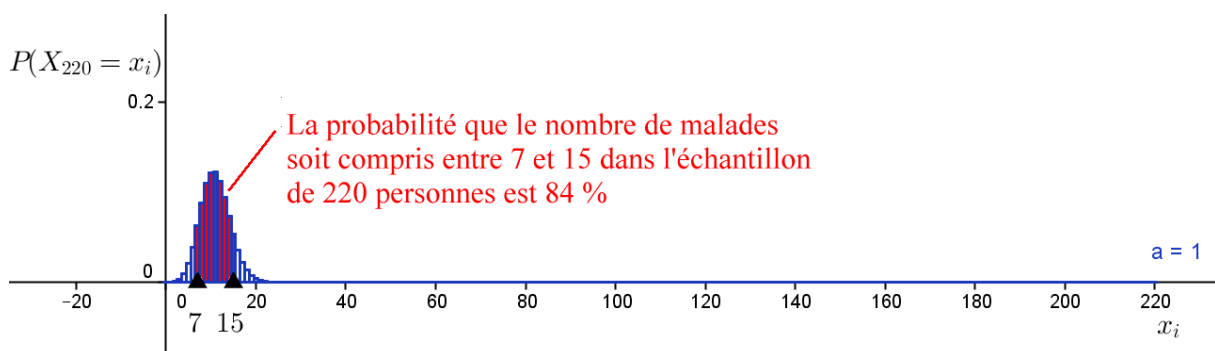
Réponse :

- Choisir 220 personnes dans une population importante revient à répéter 220 **épreuves de Bernoulli identiques et indépendantes**. On considère comme « succès » l'évènement « la personne a la grippe » (probabilité $p = 0,05$) et comme « échec » l'évènement « la personne n'a pas la grippe » (probabilité $q = 0,95$).
- Donc le nombre observé X_{220} dans l'échantillon de taille $n = 220$ peut **varier entre 0 et 220** et X_n **suit la loi binomiale $\mathcal{B}(220 ; 0,05)$** .

Le calcul de $P(7 \leq X_{220} \leq 15)$ se fait à la calculatrice avec la fonction **binomFRép** :

$$P(7 \leq X_{220} \leq 15) = P(X_{220} \leq 15) - P(X_{220} \leq 6)$$

$$P(7 \leq X_{220} \leq 15) = 0,84 \text{ à } 10^{-2} \text{ près.}$$



- 2) Toujours dans l'échantillon de 220 personnes, on pose F_{220} la fréquence observée des malades de la grippe. Déterminer la probabilité de trouver une fréquence F_{220} observée de malades dans l'échantillon telle que $\frac{7}{220} \leq F_{220} \leq \frac{15}{220}$.

Réponse :

On a $F_{220} = \frac{X_{220}}{220}$.

$$\frac{7}{220} \leq F_{220} \leq \frac{15}{220}$$

équivalent successivement à

$$\frac{7}{220} \leq \frac{X_{220}}{220} \leq \frac{15}{220}$$

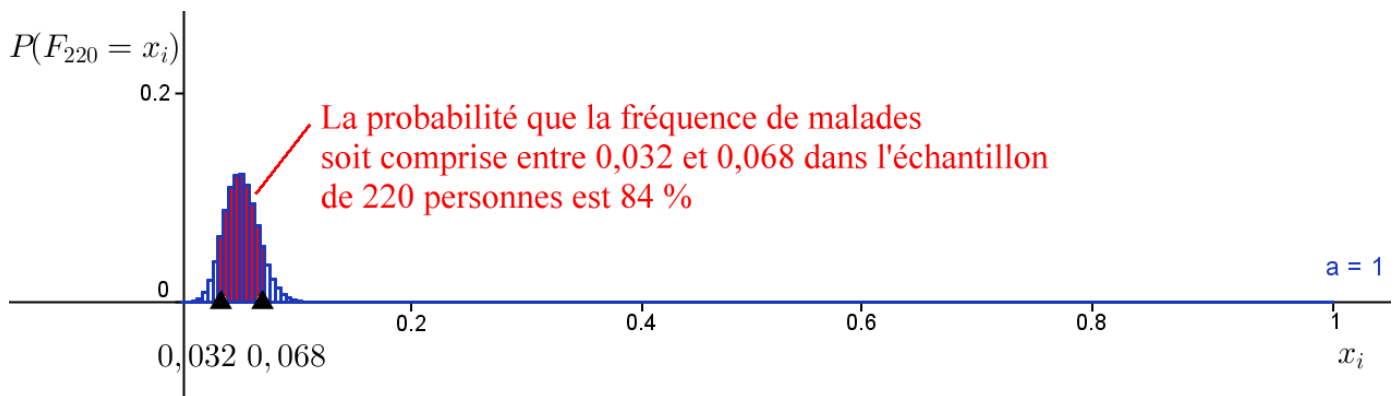
$$7 \leq X_{220} \leq 15$$

D'où :

$$P\left(\frac{7}{220} \leq F_{220} \leq \frac{15}{220}\right) = P(7 \leq X_{220} \leq 15) = 0,84 \text{ à } 10^{-2} \text{ près.}$$

ou encore, avec des valeurs approchées : $\frac{7}{220} \approx 0,032$ et $\frac{15}{220} \approx 0,068$ donc

$$P(0,032 \leq F_{220} \leq 0,068) = 0,84 \text{ à } 10^{-2} \text{ près.}$$



Propriété :

Si F_n est la fréquence d'observation d'un caractère binaire (chaque individu possède ou ne possède pas une certaine caractéristique) sur un échantillon de taille n , alors :

$$P\left(\frac{\alpha}{n} \leq F_n \leq \frac{\beta}{n}\right) = P(\alpha \leq X_n \leq \beta)$$

où X_n suit la loi binomiale $\mathcal{B}(n; p)$ et où α et β sont des entiers de l'intervalle $[0; n]$.

Exemple 2 : Intervalle de fluctuation à 95 % de la fréquence observée

- 1) En reprenant la même population que dans l'exemple 1, déterminer les entiers a et b tels que :

a soit le plus petit entier tel que $P(X_{220} \leq a) > 0,025$.

b soit le plus petit entier tel que $P(X_{220} \leq b) \geq 0,975$.

Réponse :

On cherche $P(X_{220} \leq x_i)$ avec les probabilités cumulées de la loi $\mathcal{B}(220 ; 0,05)$:

- Soit a le premier entier tel que $P(X_{220} \leq a) > 0,025$
- Soit b le premier entier tel que $P(X_{220} \leq b) \geq 0,975$

A la calculatrice, on utilise les listes L_1 et L_2 pour calculer les **probabilités cumulées** de la loi binomiale à l'aide de la fonction `binomFRép(220, 0.05, ...)`

- Appuyer sur la touche Stats, puis dans le menu EDIT choisir EffListe L_1, L_2
- Appuyer sur Stats, puis dans le menu EDIT choisir Edite.

Remplissage de la liste L_1 avec les entiers x_i compris entre 0 et 220 :

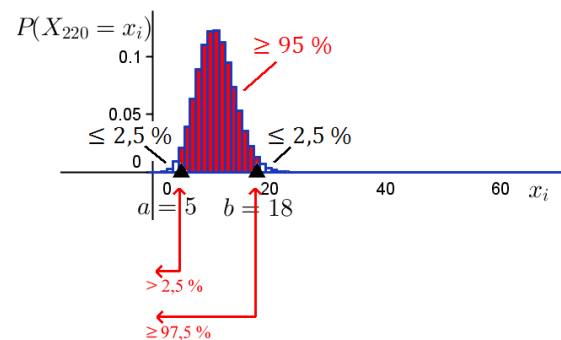
- Sélectionner le titre de la colonne L_1 , entrée, et saisir la formule $L_1 = \text{suite}(X, X, 0, 220)$
Pour trouver la fonction suite, appuyer sur `(2nde)` `[listes]` et aller dans le menu OPS.

Remplissage de la liste L_2 avec les probabilités $P(X_{220} \leq x_i)$:

- Sélectionner le titre de la colonne L_2 , entrée, et saisir $L_2 = \text{binomFRép}(220, 0.05, L_1)$
Pour trouver la fonction `binomFRép`, appuyer sur `(2nde)` `[distrib]` et descendre dans le menu DISTRIB.

Après quelques dizaines de secondes de calcul, les listes sont prêtes :

L_1	L_2
4	0,0134
5	0,0342 est la 1 ^{ère} valeur supérieure à 2,5%
...	...
...	...
17	0,9714
18	0,9848 est la 1 ^{ère} valeur supérieure à 97,5%
...	...



$a = 5$ est le premier entier tel que $P(X_{220} \leq a) > 0,025$
et $b = 18$ est le premier entier tel que $P(X_{220} \leq b) \geq 0,975$.

La probabilité que l'on observe un nombre de malades X_{220} entre 5 et 18 dans l'échantillon est :

$$0,9848 - 0,0134 = 0,9714 \text{ soit environ } 0,97.$$

2) Que peut-on en déduire pour la fréquence F_{220} de malades dans l'échantillon de taille $n = 220$?

Réponse :

D'après la propriété précédente :

$$P\left(\frac{5}{220} \leq F_{220} \leq \frac{18}{220}\right) = P(5 \leq X_{220} \leq 18)$$

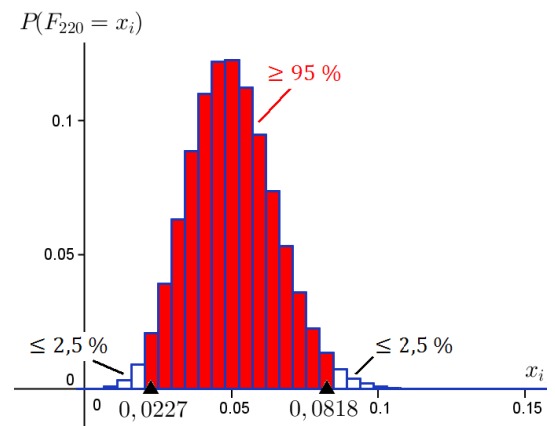
Avec des valeurs approchées, on a :

$$P(0,0227 \leq F_{220} \leq 0,0818) \approx 0,9848 - 0,0134$$

$$P(0,0227 \leq F_{220} \leq 0,0818) \approx 0,97$$

La probabilité que l'on observe une fréquence entre 0,0227 et 0,0818 de malades dans un échantillon de 220 personnes est

$0,97 \approx 0,95$. L'intervalle $I_{220} = \left[\frac{5}{220} ; \frac{18}{220}\right]$ est l'**intervalle de fluctuation de F_{220} au seuil 95 %**.



2 La valeur de la proportion p dans la population est connue. On prédit un intervalle de fluctuation I_n contenant la fréquence F_n

On suppose que la taille de la population est très grande, de sorte que les prélèvements des individus soient considérés comme une succession d'épreuves de Bernoulli identiques et indépendantes à 2 issues : l'individu a le caractère avec la probabilité p « succès » ou il ne l'a pas avec la probabilité $q = 1 - p$ « échec ». On **prédit** un intervalle de fluctuation I_n dans lequel **il est probable** que se situe la fréquence F_n des individus qui présentent ce caractère observée sur cet échantillon de taille n .

2.1 Définition de l'intervalle de fluctuation à 95 % avec la loi binomiale

Soit X_n une variable aléatoire qui suit une loi binomiale $\mathcal{B}(n; p)$ et $F_n = \frac{X_n}{n}$ la variable aléatoire qui représente la fréquence observée des « succès » dans un échantillon de taille n .

Un intervalle de fluctuation de F_n au seuil de 95 % est un intervalle :

- I_n est de la forme $\left[\frac{a}{n}; \frac{b}{n}\right]$ où a et b sont des entiers compris entre 0 et n
- I_n est tel que $P\left(\frac{a}{n} \leq F_n \leq \frac{b}{n}\right) \geq 0,95$ ce qui équivaut à $P(a \leq X_n \leq b) \geq 0,95$

En pratique¹, on cherche avec **binomFRép** le plus petit entier a tel que $P(X_n \leq a) > 0,025$ et le plus petit entier b tel que $P(X_n \leq b) \geq 0,975$.

2.2 Définition de I_n l'intervalle de fluctuation à 95 % avec la loi normale

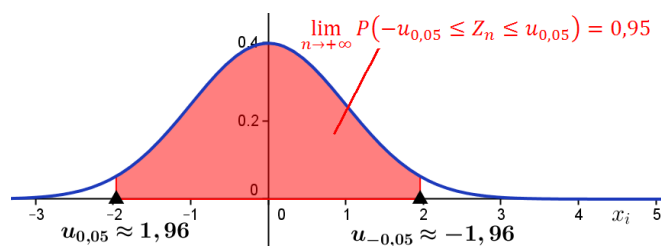
1. Théorème de Moivre-Laplace pour la suite définie pour tout $n \in \mathbb{N}^*$ par $P(-u_{0,05} \leq Z_n \leq u_{0,05})$

Soit X_n la variable aléatoire qui donne le nombre d'individus possédant le caractère étudié dans un échantillon de taille n . X_n suit la loi $\mathcal{B}(n; p)$. Soit $Z_n = \frac{X_n - \mu}{\sigma}$ la variable X_n centrée réduite.

Pour connaître I_n l'intervalle de fluctuation avec la loi normale au seuil 95 %, on écrit le théorème de Moivre-Laplace avec les bornes $-u_{0,05}$ et $u_{0,05}$:

$$\lim_{n \rightarrow +\infty} P(-u_{0,05} \leq Z_n \leq u_{0,05}) = \int_{-u_{0,05}}^{u_{0,05}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

$$\lim_{n \rightarrow +\infty} P(-u_{0,05} \leq Z_n \leq u_{0,05}) = 0,95.$$



FracNormale(0,975) donne $U_{0,05} \approx 1,96$

Conclusion : $\lim_{n \rightarrow +\infty} P(-u_{0,05} \leq Z_n \leq u_{0,05}) = 0,95.$

2. On écrit une équivalence de $-u_{0,05} \leq Z_n \leq u_{0,05}$ qui fasse intervenir $F_n = \frac{X_n}{n}$

$-u_{0,05} \leq Z_n \leq u_{0,05}$ Equivaut successivement à : $-u_{0,05} \leq \frac{X_n - \mu}{\sigma} \leq u_{0,05}$

¹ Ces définitions de a et b viennent du fait qu'il faut au moins 95 % des probabilités cumulées lorsque $a \leq X_n \leq b$ et donc on a $P(X_n \leq a - 1) \leq 0,025$ et $P(X_n \geq b + 1) \leq 0,025$.

$$\begin{aligned}
-u_{0,05} &\leq \frac{X_n - np}{\sqrt{npq}} \leq u_{0,05} \\
-u_{0,05}\sqrt{npq} &\leq X_n - np \leq u_{0,05}\sqrt{npq} \\
np - u_{0,05}\sqrt{npq} &\leq X_n \leq np + u_{0,05}\sqrt{npq} \\
\frac{1}{n}(np - u_{0,05}\sqrt{npq}) &\leq \frac{X_n}{n} \leq \frac{1}{n}(np + u_{0,05}\sqrt{npq}) \\
p - \frac{u_{0,05}\sqrt{npq}}{n} &\leq \frac{X_n}{n} \leq p + \frac{u_{0,05}\sqrt{npq}}{n} \\
p - \frac{u_{0,05}\sqrt{n}\sqrt{pq}}{n} &\leq \frac{X_n}{n} \leq p + \frac{u_{0,05}\sqrt{n}\sqrt{pq}}{n} \\
p - \frac{u_{0,05}\sqrt{n}\sqrt{pq}}{\sqrt{n} \times \sqrt{n}} &\leq \frac{X_n}{n} \leq p + \frac{u_{0,05}\sqrt{n}\sqrt{pq}}{\sqrt{n} \times \sqrt{n}} \\
p - \frac{u_{0,05}\sqrt{pq}}{\sqrt{n}} &\leq \frac{X_n}{n} \leq p + \frac{u_{0,05}\sqrt{pq}}{\sqrt{n}} \\
p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} &\leq F_n \leq p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}}
\end{aligned}$$

Conclusion :

$$-u_{0,05} \leq Z_n \leq u_{0,05} \quad \Leftrightarrow \quad p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \leq F_n \leq p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}}$$

3. On conclut sur $\lim_{n \rightarrow +\infty} P\left(p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \leq F_n \leq p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}}\right)$

On a démontré $\left\{ \begin{array}{l} \lim_{n \rightarrow +\infty} P(-u_{0,05} \leq Z_n \leq u_{0,05}) = 0,95 \quad (\text{partie 1}) \\ -u_{0,05} \leq Z_n \leq u_{0,05} \quad \Leftrightarrow \quad p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \leq F_n \leq p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \quad (\text{partie 2}) \end{array} \right.$

Conclusion :

$$\lim_{n \rightarrow +\infty} P\left(p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \leq F_n \leq p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}}\right) = 0,95$$

ou encore : $\lim_{n \rightarrow +\infty} P(F_n \in I_n) = 0,95$ en posant $I_n = \left[p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} ; p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}}\right]$.

On appelle I_n l'**intervalle de fluctuation asymptotique à 95 %** car, quand la taille n de l'échantillon tend vers $+\infty$, la probabilité que F_n se situe dedans tend vers 0,95.

En pratique, si $n \geq 30$, avec $np \geq 5$ et $nq \geq 5$, la probabilité que F_n soit dedans est proche de 0,95.

Exemple :

Sur une population d'individus, 5 personnes sur 100 ont la grippe. On prend un échantillon de $n = 220$ personnes. Déterminer l'intervalle de fluctuation I_{220} au seuil de 95 % avec la loi normale.

Réponse :

La population présente une proportion $p = 0,05$ d'individus atteints par la grippe. On a comme taille d'échantillon $n = 220$. L'intervalle de fluctuation asymptotique est :

$$I_{220} = \left[p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{220}} ; p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{220}} \right]$$

$$I_{220} = \left[0,05 - u_{0,05} \frac{\sqrt{0,05 \times 0,95}}{\sqrt{220}} ; 0,05 + u_{0,05} \frac{\sqrt{0,05 \times 0,95}}{\sqrt{220}} \right]$$

On a $u_{0,05} \approx 1,96$

$$0,05 - 1,96 \times \frac{\sqrt{0,05 \times 0,95}}{\sqrt{220}} \approx \mathbf{0,0212}$$

$$0,05 + 1,96 \times \frac{\sqrt{0,05 \times 0,95}}{\sqrt{220}} \approx \mathbf{0,0788}$$

D'où $I_{220} = [0,021 ; 0,079]$ en arrondissant les bornes de l'intervalle de fluctuation à 10^{-3} près.

Remarques :

- On avait trouvé l'intervalle de fluctuation en utilisant la loi binomiale ($0,0227 \leq F_{220} \leq 0,0818$) $\approx 0,95$. On n'a donc pas retrouvé la même chose exactement car $n = 220$ n'est pas l'infini. Cependant, l'intervalle de fluctuation obtenu avec la loi normale est assez proche.
- Si n est très grand alors l'approximation est très bonne. Cela est dû au fait qu'on assimile $P(-u_{0,05} \leq Z_n \leq u_{0,05})$ où Z_n est une variable aléatoire discrète, à l'intégrale de la densité de la loi normale centrée réduite qui est égale à $\lim_{n \rightarrow +\infty} P(-u_{0,05} \leq Z_n \leq u_{0,05})$.
-  En pratique, on considère que si les conditions suivantes sont vraies :

$$\begin{cases} n \geq 30 \\ np \geq 5 \\ nq \geq 5 \end{cases}$$

alors on peut utiliser l'intervalle de fluctuation asymptotique I_n donné par la loi normale centrée réduite. Cet intervalle de fluctuation est d'autant plus précis que n la taille de l'échantillon tend vers $+\infty$.

- On donne l'intervalle de fluctuation en arrondissant la borne inférieure par défaut et la borne supérieure par excès, de façon à garantir au moins 95% de probabilité pour que fréquence observée F_n soit dans cet intervalle.
- On peut définir l'intervalle de fluctuation asymptotique des valeurs de α autres que $\alpha = 0,05$. Par exemple $\alpha = 0,01$; $\alpha = 0,10$ etc.
- La proportion p dans la population étant fixée alors q est aussi fixé. Comme $u_{0,05}$ est aussi fixé, on voit que l'amplitude (c'est-à-dire la distance entre ses bornes) ne dépend que de la taille de l'échantillon. Cette distance est :

$$\left(p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \right) - \left(p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \right) = 2 \times u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}}$$

En posant la constante $k = 2 \times u_{0,05} \sqrt{pq}$, on a :

$$\left(p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \right) - \left(p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \right) = k \frac{1}{\sqrt{n}}$$

L'amplitude de l'intervalle de fluctuation I_n est donc inversement proportionnelle à $\sqrt{n}$. **Pour diviser par 2** l'amplitude de l'intervalle de fluctuation, on doit augmenter la taille de l'échantillon de n à n' .

$$k \frac{1}{\sqrt{n'}} = \frac{1}{2} \times k \frac{1}{\sqrt{n}}$$

$$\frac{1}{\sqrt{n'}} = \frac{1}{2\sqrt{n}}$$

$$\sqrt{n'} = 2\sqrt{n}$$

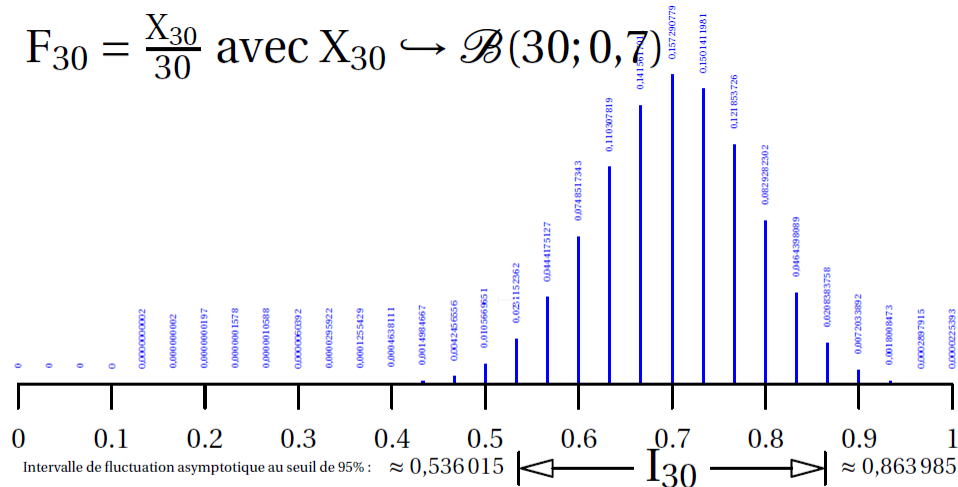
$$(\sqrt{n'})^2 = (2\sqrt{n})^2$$

$$n' = 4n$$

Donc en **quadruplant** la taille de l'échantillon, on divise par 2 l'amplitude de l'intervalle de fluctuation de la fréquence observée.

Exemple :

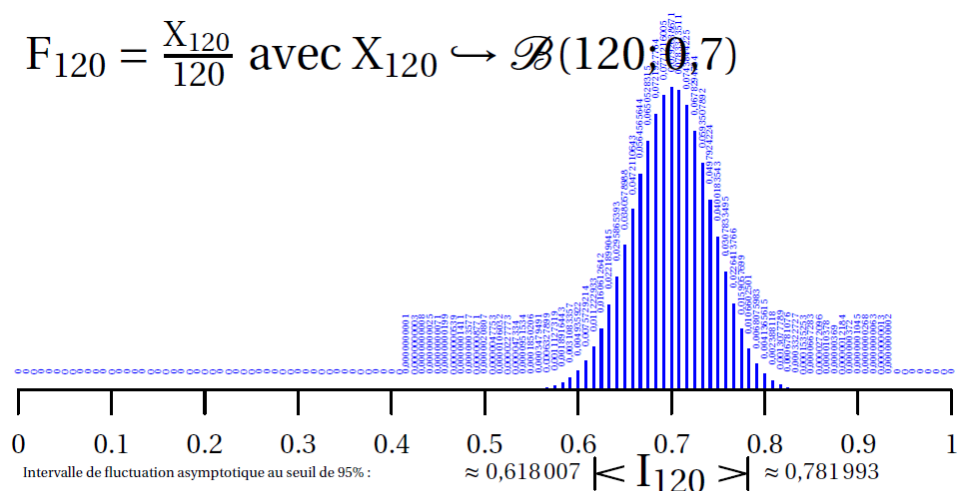
$$F_{30} = \frac{X_{30}}{30} \text{ avec } X_{30} \hookrightarrow \mathcal{B}(30; 0,7)$$



$$P(F_{30} \in I_{30}) \approx 0.95$$

$$I_{30} = \left[0,7 - u_{0,05} \frac{\sqrt{0,7 \times 0,3}}{\sqrt{30}} ; 0,7 + u_{0,05} \frac{\sqrt{0,7 \times 0,3}}{\sqrt{30}} \right] \text{ amplitude de } I_{30} \approx 0,328$$

$$F_{120} = \frac{X_{120}}{120} \text{ avec } X_{120} \hookrightarrow \mathcal{B}(120; 0,7)$$



$$P(F_{120} \in I_{120}) \approx 0.95$$

$$I_{120} = \left[0,7 - u_{0,05} \frac{\sqrt{0,7 \times 0,3}}{\sqrt{120}} ; 0,7 + u_{0,05} \frac{\sqrt{0,7 \times 0,3}}{\sqrt{120}} \right] \text{ amplitude de } I_{120} \approx 0,164$$

2.3 Intervalle I_n de fluctuation asymptotique au seuil de $(1-\alpha)\%$

Théorème :

Si la variable aléatoire X_n suit la loi $\mathcal{B}(n; p)$, alors pour tout réel $\alpha \in]0; 1[$, on a :

$$\lim_{n \rightarrow +\infty} \left(P \left(p - u_\alpha \frac{\sqrt{pq}}{\sqrt{n}} \leq F_n \leq p + u_\alpha \frac{\sqrt{pq}}{\sqrt{n}} \right) \right) = 1 - \alpha$$

Autre écriture :

$$\lim_{n \rightarrow +\infty} \left(P \left(F_n \in \left[p - u_\alpha \frac{\sqrt{pq}}{\sqrt{n}} ; p + u_\alpha \frac{\sqrt{pq}}{\sqrt{n}} \right] \right) \right) = 1 - \alpha$$

Démonstration (exigible) :

- Soit X_n la variable aléatoire qui donne le nombre d'observations dans un intervalle de taille n . X_n suit la loi binomiale $\mathcal{B}(n; p)$. X_n a comme valeurs les entiers $x_i \in [0; n]$

Soit $Z_n = \frac{X_n - \mu}{\sigma}$ la variable X_n centrée réduite. Z_n a comme valeurs les réels $\frac{x_i - \mu}{\sigma} \in \left[\frac{0 - \mu}{\sigma} ; \frac{n - \mu}{\sigma} \right]$

- Alors, d'après le **théorème de Moivre-Laplace**, pour tous réels a et b tels que $a < b$, on a :

$$\lim_{n \rightarrow +\infty} P(a \leq Z_n \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

- Pour connaître I_n l'intervalle de fluctuation asymptotique au seuil de $(1 - \alpha) \%$, on écrit le théorème de Moivre-Laplace avec $a = -u_\alpha$ et $b = u_\alpha$:

$$\lim_{n \rightarrow +\infty} P(-u_\alpha \leq Z_n \leq u_\alpha) = \int_{-u_\alpha}^{u_\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

- $-u_\alpha \leq Z_n \leq u_\alpha$

équivalent successivement à :

$$-u_\alpha \leq \frac{X_n - \mu}{\sigma} \leq u_\alpha$$

$$-u_\alpha \leq \frac{X_n - np}{\sqrt{npq}} \leq u_\alpha$$

$$-u_\alpha \sqrt{npq} \leq X_n - np \leq u_\alpha \sqrt{npq}$$

$$np - u_\alpha \sqrt{npq} \leq X_n \leq np + u_\alpha \sqrt{npq}$$

$$\frac{1}{n} (np - u_\alpha \sqrt{npq}) \leq \frac{X_n}{n} \leq \frac{1}{n} (np + u_\alpha \sqrt{npq})$$

$$p - \frac{u_\alpha \sqrt{npq}}{n} \leq \frac{X_n}{n} \leq p + \frac{u_\alpha \sqrt{npq}}{n}$$

$$p - \frac{u_\alpha \sqrt{n} \sqrt{pq}}{n} \leq \frac{X_n}{n} \leq p + \frac{u_\alpha \sqrt{n} \sqrt{pq}}{n}$$

$$p - \frac{u_\alpha \sqrt{n} \sqrt{pq}}{\sqrt{n} \times \sqrt{n}} \leq \frac{X_n}{n} \leq p + \frac{u_\alpha \sqrt{n} \sqrt{pq}}{\sqrt{n} \times \sqrt{n}}$$

$$p - \frac{u_\alpha \sqrt{pq}}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + \frac{u_\alpha \sqrt{pq}}{\sqrt{n}}$$

$$p - u_\alpha \frac{\sqrt{pq}}{\sqrt{n}} \leq F_n \leq p + u_\alpha \frac{\sqrt{pq}}{\sqrt{n}}$$

- Conclusion :

Dans un échantillon de taille n prélevé dans une population avec une proportion p d'individus présentant un certain caractère, on observera que la fréquence F_n de ce caractère est dans l'intervalle :

$$I_n = \left[p - u_\alpha \frac{\sqrt{pq}}{\sqrt{n}} ; p + u_\alpha \frac{\sqrt{pq}}{\sqrt{n}} \right]$$

avec une probabilité de $1 - \alpha$.

C'est l'intervalle de fluctuation asymptotique au seuil de $1 - \alpha$ % d'une fréquence F_n

Exemple :

Sur une population d'individus, 5 personnes sur 100 ont la grippe.

On prend un échantillon de $n = 220$ personnes. Déterminer l'intervalle de fluctuation asymptotique I_{220} au seuil de 99 %.

Réponse :

La population présente une proportion $p = 0,05$ d'individus atteints par la grippe. On a comme taille d'échantillon $n = 220$. L'intervalle de fluctuation avec la loi normale est :

$$I_{220} = \left[p - u_{0,01} \frac{\sqrt{pq}}{\sqrt{220}} ; p + u_{0,01} \frac{\sqrt{pq}}{\sqrt{220}} \right]$$

$$I_{220} = \left[0,05 - u_{0,01} \frac{\sqrt{0,05 \times 0,95}}{\sqrt{220}} ; 0,05 + u_{0,01} \frac{\sqrt{0,05 \times 0,95}}{\sqrt{220}} \right]$$

On a $u_{0,01} \approx 2,58$ donné par $\text{FracNormale}(0.995)$

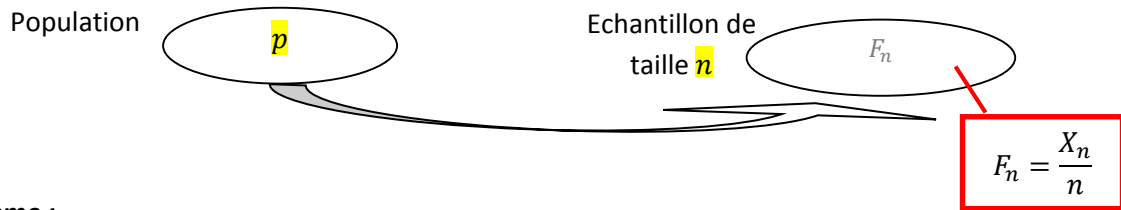
$$0,05 - 2,58 \times \frac{\sqrt{0,05 \times 0,95}}{\sqrt{220}} \approx \mathbf{0,0121}$$

$$0,05 + 2,58 \times \frac{\sqrt{0,05 \times 0,95}}{\sqrt{220}} \approx \mathbf{0,0879}$$

D'où $I_{220} = [\mathbf{0,012} ; \mathbf{0,088}]$ en arrondissant les bornes de l'intervalle de fluctuation à 10^{-3} près.

Ainsi, dans un échantillon de 220 personnes, on peut prédire avec une probabilité de 99 % que la fréquence de personnes atteintes par la grippe sera entre 1,2 % et 8,8 %.

RÉSUMÉ : Échantillonnage dans une population où p est connue

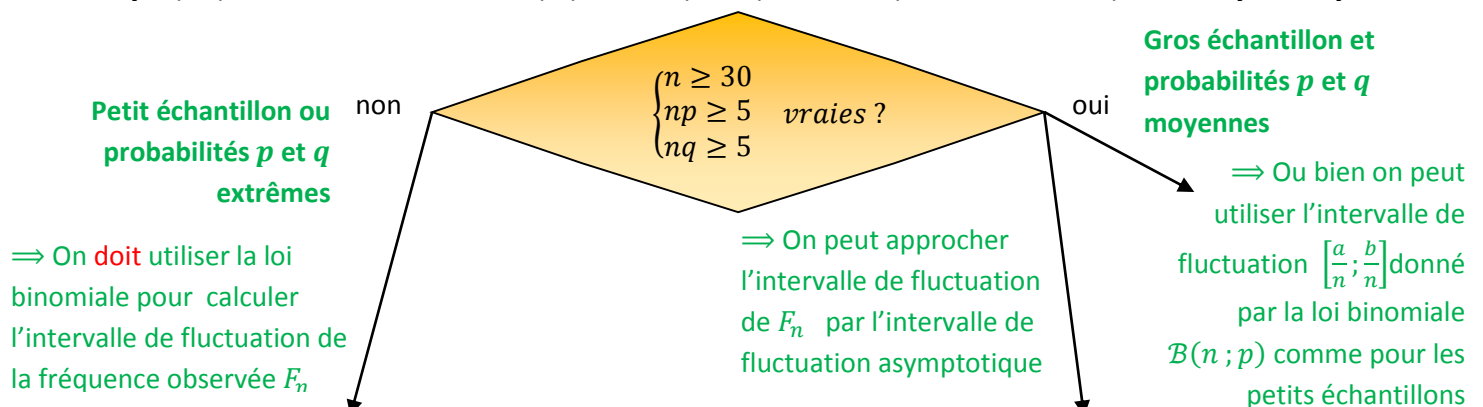


Position du problème :

- La proportion p d'individus présentant une certaine caractéristique dans la population est connue.
- On prélève un échantillon de n individus. On cherche à prédire par un intervalle de fluctuation, les valeurs probables de la fréquence F_n observée de cette caractéristique dans un échantillon de taille n connue.

$$F_n = \frac{X_n}{n} \quad \text{où } X_n \text{ est le nombre d'individus observés présentant la caractéristique dans l'échantillon de taille } n.$$

On note q la proportion d'individus dans la population qui ne présentent pas la caractéristique. On a $q = 1 - p$.



L'intervalle de fluctuation, au seuil 95% de la fréquence F_n correspondant à la réalisation sur un échantillon de taille n d'une variable aléatoire X_n suivant une loi binomiale, est :

$$I_n = \left[\frac{a}{n}; \frac{b}{n} \right]$$

$F_n \in I_n$ avec une probabilité d'au moins 0,95.

a et b sont les plus petits entiers tels que $\begin{cases} P(X_n \leq a) > 0,025 \\ P(X_n \leq b) \geq 0,975 \end{cases}$

Pour trouver a et b on utilise binomFRép sur la calculatrice.

Exemple : Dans une population, la proportion d'un caractère est $p = 0,3$. Donner l'intervalle de fluctuation, au coefficient 95% de la fréquence F_n du caractère sur un échantillon de taille 50.

$L_2 = \text{binomFRép}(50, 0,3, L_1)$ donne $a = 9$ et $b = 22$.

$$I_n = \left[\frac{9}{50}; \frac{22}{50} \right] \approx [0,18; 0,44].$$

La fréquence F_n dans l'échantillon suit approximativement la loi $\mathcal{N}\left(p; \frac{pq}{n}\right)$

L'intervalle de fluctuation, au seuil $1 - \alpha$ % de la fréquence F_n correspondant à la réalisation sur un échantillon de taille n d'une variable aléatoire X_n suivant une loi binomiale, est approximativement :

$$I_n = \left[p - u_\alpha \frac{\sqrt{pq}}{\sqrt{n}}; p + u_\alpha \frac{\sqrt{pq}}{\sqrt{n}} \right]$$

$F_n \in I_n$ avec une probabilité d'au moins $1 - \alpha$.

u_α est le réel positif tel que

$$P(X_n \in]-\infty; -u_\alpha[\cup]u_\alpha; +\infty[) = \alpha$$

Pour trouver u_α on utilise FracNormale sur la calculatrice.

Exemple : Avec $n = 50$, $np = 50 \times 0,3 = 15$ et $nq = 50 \times 0,7 = 35$ les conditions pour utiliser l'intervalle de fluctuation asymptotique sont vérifiées.

$$u_{0,05} = \text{FracNormale}(0,975) \approx 1,960.$$

$$I_n = \left[0,3 - 1,96 \sqrt{\frac{0,3 \times 0,7}{50}}; 0,3 + 1,96 \sqrt{\frac{0,3 \times 0,7}{50}} \right] \approx [0,173; 0,427].$$

3 La valeur de la proportion p dans la population est une hypothèse qu'on veut tester

Dans ce paragraphe, on ne connaît pas la proportion p d'individus dans la population à présenter un caractère donné.

- Mais on fait **une hypothèse (conjecture) sur la valeur de p** .

Après avoir conjecturé la valeur de p , on calcule l'intervalle de fluctuation I_n , ou bien par la loi binomiale, ou bien l'intervalle de fluctuation asymptotique, d'après la taille connue n de l'échantillon et d'après la valeur du risque α qu'on a choisie (par exemple $\alpha = 0,05$ c'est-à-dire 5 %).

- On prélève un échantillon de taille n dans la population et on observe la fréquence F_n d'individus dans l'échantillon qui présentent le caractère.

- **Règle de décision :**

Le test de l'hypothèse sur la valeur de p repose sur la règle de décision suivante :

- La fréquence F_n observée appartient à l'intervalle de fluctuation. Donc cela veut dire que l'observation est compatible avec l'hypothèse faite sur la valeur de p . On accepte donc l'hypothèse ce qui veut dire qu'elle reste une hypothèse plausible.
- La fréquence F_n observée n'appartient pas à l'intervalle de fluctuation. Donc cela veut dire que l'observation est incompatible avec l'hypothèse faite sur la valeur de p . Donc on rejettera l'hypothèse.

Exemple :

Un laboratoire annonce qu'un médicament sauve 40 % des patients atteints d'une maladie. Pour contrôler cette affirmation, on le teste sur 100 patients atteints de cette maladie.

Soit X_{100} le nombre de malades sauvés par ce médicament dans cet échantillon aléatoire de malades et assimilé à un tirage avec remise de taille $n = 100$.

On fait l'hypothèse : « La proportion de malades sauvés dans la population de malades est $p = 0,4$ »

- 1) Quelle est la loi suivie par X_{100} ?
- 2) Déterminer les plus petits entiers a et b tel que $P(X_{100} \leq a) > 0,025$ et $P(X_{100} \leq b) \geq 0,975$.
- 3) Énoncer la règle de décision permettant de rejeter ou non l'hypothèse " $p = 0,4$ ", selon la valeur de la fréquence F_{100} des malades sauvés dans l'échantillon.
- 4) Sur les 100 malades auxquels on a administré ce médicament, on en a sauvé 30. Au seuil de risque 5 %, que peut-on dire de l'annonce faite par le laboratoire ?

Réponse :

- 1) Il s'agit d'une succession de 100 épreuves de Bernoulli identiques et indépendantes (car on assimile le prélèvement de l'échantillon à un tirage avec remise). On appelle « succès » l'évènement « Le malade a été sauvé » avec la probabilité $p = 0,4$.

Donc X_{100} suit la loi binomiale $\mathcal{B}(100 ; 0,4)$.

- 2) A la calculatrice, on utilise les listes L_1 et L_2 pour calculer les **probabilités cumulées** de la loi binomiale à l'aide de la fonction $\text{binomFRép}(100, 0.4, \dots)$
- Appuyer sur la touche Stats, puis dans le menu EDIT choisir EffListe² L_1, L_2
 - Appuyer sur Stats, puis dans le menu EDIT choisir Edite.
 - Sélectionner le titre de la colonne L_1 , entrée, et saisir la formule $L_1 = \text{suite}(X, X, 0, 100)$

Pour trouver la fonction suite, appuyer sur 2nde [listes] et aller dans le menu OPS.

- Sélectionner le titre de la colonne L_2 , entrée, et saisir $L_2 = \text{binomFRép}(100, 0.4, L_1)$

Pour trouver la fonction binomFRép , appuyer sur 2nde [distrib] et descendre dans le menu DISTRIB.

Après quelques secondes de calcul, les listes sont prêtes. On peut alors déterminer :

- Le plus petit entier a tel que $P(X \leq a) > 0,025$. On trouve $a = 31$.
- Le plus petit entier b tel que $P(X \leq b) \geq 0,975$. On trouve $b = 50$.

3) Règle de décision :

L'intervalle de fluctuation à 95% de la fréquence F_{100} est $I_{100} = \left[\frac{a}{n} ; \frac{b}{n} \right]$

$$I_{100} = \left[\frac{31}{100} ; \frac{50}{100} \right]$$

$$I_{100} = [0,31 ; 0,5]$$

- Si la fréquence observée dans l'échantillon de malades sauvés F_{100} appartient à l'intervalle $I_{100} = [0,31 ; 0,5]$ alors on accepte l'hypothèse :
" la proportion de malades sauvés dans la population est $p = 0,4$ "
- Si la fréquence observée dans l'échantillon de malades sauvés F_{100} n'appartient pas à l'intervalle $I_{100} = [0,31 ; 0,5]$ alors on rejette l'hypothèse :
" la proportion de malades sauvés dans la population est $p = 0,4$ "

- 4) Sur les 100 malades auxquels on a administré ce médicament, on en a sauvé 30. Donc la fréquence observée est $F_{100} = 0,3$.

Donc $F_{100} \notin I_{100}$

Donc on rejette l'hypothèse H_0 .

Mais le rejet se fait **au risque 5 % de se tromper** car il y a une probabilité de 0,05 que F_{100} soit à l'extérieur de l'intervalle $I_{100} = [0,31 ; 0,5]$ **du seul fait de la fluctuation d'échantillonnage**.

Remarque :

Dans cet exemple, on a :

$$n = 100 ; \quad np = 100 \times 0,4 = 40 ; \quad nq = 100 \times 0,6 = 60$$

Donc les trois conditions : $\begin{cases} n \geq 30 \\ np \geq 5 \\ nq \geq 5 \end{cases}$ sont vérifiées.

Donc, à la place d'utiliser l'intervalle de fluctuation I_{100} donné par a et b de la loi $\mathcal{B}(100 ; 0,4)$, on

aurait pu utiliser l'intervalle de fluctuation asymptotique $I_n = \left[p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} ; p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \right]$.

$$I_{100} = \left[0,4 - 1,96 \frac{\sqrt{0,4 \times 0,6}}{\sqrt{100}} ; 0,4 + 1,96 \frac{\sqrt{0,4 \times 0,6}}{\sqrt{100}} \right] \quad I_{100} = [0,304 ; 0,4960].$$

² Pour toute nouvelle utilisation des fonctions statistiques, penser à effacer les listes précédentes. Ainsi l'ancien contenu ne sera pas pris en compte dans les nouveaux calculs.

4 La valeur de la proportion p dans la population est une inconnue qu'on estime par un intervalle de confiance I_c



Le problème de l'estimation est le problème « inverse » de celui de la recherche d'un intervalle de fluctuation de la fréquence observée F_n étudié au paragraphe 2.

A partir de la fréquence F_n observée sur un échantillon de taille n , dans quel intervalle de confiance I_c peut se situer la proportion p des individus d'une population (de taille très grande) présentant un certain caractère ? L'estimation se fait toujours à un niveau de confiance donné.

Dans ce paragraphe 4, dans un souci de simplification :

- On se place toujours au niveau de confiance 95 % ce qui signifie que la proportion p inconnue a une probabilité de 0,95 de se trouver dans l'intervalle de confiance I_c qu'on va calculer.
- Plutôt que d'utiliser l'intervalle de fluctuation de F_n au seuil 95 %, démontré au §2.3 : $I_n = \left[p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} ; p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \right]$, on utilisera l'intervalle de fluctuation « simplifié » :

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

Propriété 1 : F_n est dans un intervalle centré sur p

Soit X_n une variable aléatoire de loi binomiale $\mathcal{B}(n ; p)$ et $F_n = \frac{X_n}{n}$ la fréquence observée d'individus présentant un certain caractère sur un échantillon de taille n .

Pour tout réel $p \in]0 ; 1[$, il existe $n_0 \in \mathbb{N}$ tel que si $n \geq n_0$ alors :

$$P \left(F_n \in \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right] \right) \geq 0,95$$

Démonstration :

- Etablissons un 1^{er} résultat : **Pour tout $p \in]0 ; 1[$, $4p(1 - p) \leq 1$.**

Pour cela résolvons sur $]0 ; 1[$ l'inéquation :

$$4p(1 - p) \leq 1$$

$$-4p^2 + 4p - 1 \leq 0$$

$$4p^2 - 4p + 1 \geq 0$$

$$(2p - 1)^2 \geq 0$$

Un carré étant toujours positif, **$4p(1 - p) \leq 1$ est vrai pour tout $p \in]0 ; 1[$.**

- Etablissons un 2^e résultat : $F_n \in \left[p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} ; p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \right] \Rightarrow F_n \in \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$.

$F_n \in \left[p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} ; p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \right]$ implique successivement :

$$p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \leq F_n \leq p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}}$$

$$-u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \leq F_n - p \leq u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}}$$

$$|F_n - p| \leq u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \quad \text{Rappel : } |x| \leq r \text{ équivaut à } -r \leq x \leq r.$$

$$|F_n - p| \leq 1,96 \frac{\sqrt{pq}}{\sqrt{n}} \quad u_{0,05} \approx 1,96 \text{ est une valeur à connaitre.}$$

$$|F_n - p| \leq 2 \frac{\sqrt{pq}}{\sqrt{n}}$$

$$|F_n - p| \leq 2 \sqrt{\frac{pq}{n}}$$

$$|F_n - p|^2 \leq 4 \frac{pq}{n}$$

$$|F_n - p|^2 \leq 4 \frac{p(1-p)}{n}$$

$$|F_n - p|^2 \leq \frac{1}{n} \times 4p(1-p)$$

Or d'après le 1^{er} résultat préalable, $4p(1-p) \leq 1, \forall p \in]0; 1[,$ d'où :

$$|F_n - p|^2 \leq \frac{1}{n} \times 4p(1-p) \leq \frac{1}{n} \times 1$$

$$|F_n - p|^2 \leq \frac{1}{n}$$

$$|F_n - p| \leq \frac{1}{\sqrt{n}}$$

$$-\frac{1}{\sqrt{n}} \leq F_n - p \leq \frac{1}{\sqrt{n}}$$

$$p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}}$$

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

Conclusion :

$$\text{Si } F_n \in \left[p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} ; p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \right] \text{ alors } F_n \in \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$$

- Reprenons le résultat du théorème du §2.3 :

On a montré que :

$$\lim_{n \rightarrow +\infty} \left(P \left(F_n \in \left[p - u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} ; p + u_{0,05} \frac{\sqrt{pq}}{\sqrt{n}} \right] \right) \right) = 0,95$$

Cela implique donc, en utilisant le deuxième résultat préalable :

$$\lim_{n \rightarrow +\infty} \left(P \left(F_n \in \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right] \right) \right) \geq 0,95$$

En utilisant la définition d'une limite finie d'une suite, on obtient **la propriété 1** :

Pour tout réel $p \in]0 ; 1[$, il existe $n_0 \in \mathbb{N}$ tel que si $n \geq n_0$ alors :

$$P \left(F_n \in \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right] \right) \geq 0,95$$

Propriété 2 : p est dans un intervalle centré sur F_n

On part de la **propriété 1**

Pour tout réel $p \in]0 ; 1[$, il existe $n_0 \in \mathbb{N}$ tel que si $n \geq n_0$ alors :

$$P \left(p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}} \right) \geq 0,95$$

$$P \left(-\frac{1}{\sqrt{n}} \leq F_n - p \leq \frac{1}{\sqrt{n}} \right) \geq 0,95$$

$$P \left(-F_n - \frac{1}{\sqrt{n}} \leq -p \leq -F_n + \frac{1}{\sqrt{n}} \right) \geq 0,95$$

$$P \left(F_n + \frac{1}{\sqrt{n}} \geq p \geq F_n - \frac{1}{\sqrt{n}} \right) \geq 0,95$$

$$P \left(F_n - \frac{1}{\sqrt{n}} \leq p \leq F_n + \frac{1}{\sqrt{n}} \right) \geq 0,95$$

On peut donc conclure :

Propriété 2 : Pour tout réel $p \in]0 ; 1[$, il existe $n_0 \in \mathbb{N}$ tel que si $n \geq n_0$ alors :

$$P \left(p \in \left[F_n - \frac{1}{\sqrt{n}} ; F_n + \frac{1}{\sqrt{n}} \right] \right) \geq 0,95$$

Définition :

Soit une population où la **proportion p** des individus présentant un certain caractère **est inconnue**.

Soit F_n la **fréquence observée** de ce caractère dans un échantillon de taille n .

Alors, dans au moins 95 % des cas, la proportion p appartient à l'intervalle, dit **intervalle de**

confiance au niveau de confiance 95 % $I_c = \left[F_n - \frac{1}{\sqrt{n}} ; F_n + \frac{1}{\sqrt{n}} \right]$



Cet intervalle de confiance peut être utilisé lorsque les conditions suivantes sont remplies :

$$\begin{cases} n \geq 30 \\ nF_n \geq 5 \\ n(1 - F_n) \geq 5 \end{cases}$$

Exemple

Dans une urne contenant des boules blanches et bleues en proportions inconnues, on effectue des tirages au hasard avec remise.

- 1) Après avoir effectué 100 tirages, on compte 52 boules blanches (et donc 48 boules bleues).
Donner l'intervalle de confiance à 95% de la proportion de boules blanches p dans l'urne.
- 2) Combien faudrait-il, au minimum, effectuer de tirages pour obtenir un intervalle de confiance I_c au niveau de confiance 95% d'amplitude inférieure ou égale à 2.10^{-2} ?

Réponse :

- 1) On a un échantillon de taille $n = 100$.

La fréquence du caractère « la boule est blanche » est $F_{100} = 0,52$.

Les conditions pour l'utilisation de l'intervalle de confiance

$$\begin{cases} 100 \geq 30 \\ 100 \times 0,48 \geq 5 \\ 100 \times 0,52 \geq 5 \end{cases} \quad \text{Sont vérifiées.}$$

$$\text{Donc } I_c = \left[F_{100} - \frac{1}{\sqrt{n}} ; F_{100} + \frac{1}{\sqrt{n}} \right]$$

$$F_{100} - \frac{1}{\sqrt{n}} = 0,52 - \frac{1}{10} \quad \text{et} \quad F_{100} + \frac{1}{\sqrt{n}} = 0,52 + \frac{1}{10}$$

Donc, l'intervalle de confiance au niveau de confiance 95% est $I_c = [0,42 ; 0,62]$.

Autrement dit, il y a au moins 95 chances sur 100 pour que la proportion de boules blanches dans l'urne soit comprise entre 0,42 et 0,62.

- 2) On cherche $n \in \mathbb{N}$ tel que $\left(F_{100} + \frac{1}{\sqrt{n}} \right) - \left(F_{100} - \frac{1}{\sqrt{n}} \right) \leq 2.10^{-2}$

$$\frac{2}{\sqrt{n}} \leq 2.10^{-2}$$

$$\frac{2}{2.10^{-2}} \leq \sqrt{n}$$

$$\sqrt{n} \geq 10^2$$

$$n \geq 10^4$$

En tirant 10^4 boules, l'amplitude de l'intervalle de confiance au niveau de confiance 95 % est 0,02.

Remarque :

Dans certains domaines, on utilise un intervalle de confiance plus précis :

$$I_c = \left[F_n - 1,96 \sqrt{\frac{F_n(1-F_n)}{n}} ; F_n + 1,96 \sqrt{\frac{F_n(1-F_n)}{n}} \right].$$

RÉSUMÉ : Estimation par un intervalle de confiance de la proportion p qui reste inconnue

Echantillon E
de taille n

F_n

Population P

p

Position du problème : Un échantillon E est connu

Par sa taille n et par F_n la fréquence d'apparition d'un caractère dans l'échantillon.

On cherche à estimer, par un intervalle de confiance, les valeurs probables de la proportion p des individus de la population P qui présentent ce caractère.

Petit échantillon ou
fréquences F_n et
 $1 - F_n$ extrêmes

non

⇒ Hors
programme
de terminale

$$\begin{cases} n \geq 30 \\ nF_n \geq 5 \\ n(1 - F_n) \geq 5 \end{cases} ?$$

oui

Gros
échantillon
et
probabilités
 F_n et $1 - F_n$
moyennes

⇒ On peut
utiliser une loi
normale pour p

La proportion p de la population suit la loi :

$$\mathcal{N}\left(F_n; \frac{F_n(1 - F_n)}{n}\right)$$

L'intervalle de confiance au niveau **95%** de p est :

$$I_c = \left[F_n - 1,96 \sqrt{\frac{F_n(1 - F_n)}{n}}; F_n + 1,96 \sqrt{\frac{F_n(1 - F_n)}{n}} \right]$$

$p \in I_c$ au risque 5% qu'elle n'y soit pas.

Ou de façon moins précise mais simplifiée :

Dans au moins **95%** des cas, la proportion p dans la population appartient à l'intervalle

$$I_c = \left[F_n - \frac{1}{\sqrt{n}}; F_n + \frac{1}{\sqrt{n}} \right]$$

$p \in I_c$ avec un risque inférieur à 5% qu'elle n'y soit pas.

Exemple :

On a observé sur un échantillon de 500 personnes que 21 % étaient cardiaques. Calculer l'intervalle de confiance au niveau de confiance 95% dans lequel peut se situer la proportion p de personnes cardiaques parmi toute la population.

Réponse :

Ici, on a $\begin{cases} n = 500 \\ nF_n = 500 \times 0,21 = 105 \\ n(1 - F_n) = 500 \times 0,79 = 395 \end{cases}$ Donc les conditions $\begin{cases} n \geq 30 \\ nF_n \geq 5 \\ n(1 - F_n) \geq 5 \end{cases}$ sont vérifiées.

- Méthode précise :

$$I_c = \left[0,21 - 1,96 \sqrt{\frac{0,21 \times 0,79}{500}}; 0,21 + 1,96 \sqrt{\frac{0,21 \times 0,79}{500}} \right]$$

$$I_c = [0,21 - 0,036; 0,21 + 0,036]$$

$$I_c = [0,174; 0,246]$$

La proportion de cardiaques dans la population $p \in [0,174; 0,246]$ avec un niveau de confiance de 95%.

- Méthode simplifiée :

$$I_c = \left[0,21 - \frac{1}{\sqrt{500}}; 0,21 + \frac{1}{\sqrt{500}} \right]$$

$$I_c = [0,21 - 0,045; 0,21 + 0,045]$$

$$I_c = [0,165; 0,255]$$

La proportion de cardiaques dans la population $p \in [0,165; 0,255]$ avec un niveau de confiance de 95% (au moins)