Chapitre 9- INTERVALLE DE FLUCTUATION ET ESTIMATION

I – ECHANTILLONNAGE ET PRISE DE DECISION.

1- INTERVALLE DE FLUCTUATION ASYMPTOTIQUE.

Soit X_n la variable aléatoire suivant la loi binomiale B(n ; p).

Soit F_n la variable aléatoire qui associe la fréquence du caractère étudié dans l'échantillon aléatoire de taille n.

On a $F_n = \dfrac{X_n}{n}$. F_n représente alors la fréquence de succès pour un schéma de Bernoulli de paramètres n et p.

On admet que, pour n suffisamment grand, X_n suit approximativement la loi normale d'espérance mathématique $\mu = np$ et d'écart-type $\sigma = \sqrt{np(1-p)}$.

Dans le chapitre précédent que, on a vu : Si X_n suit une loi normale N(μ,σ^2),

PARTIE

A LIRE ET

A COMPRENDRE

MAIS NE PAS

RETENIR

$Z_n = \dfrac{X_n - \mu}{\sigma}$ suit la loi normale centrée réduite N(0,1) et

$P(-1{,}96 < Z_n < 1{,}96) \approx 0{,}95 \Leftrightarrow P\left(-1{,}96 < \dfrac{X_n - \mu}{\sigma} < 1{,}96\right) \approx 0{,}95$

Donc

$$P(-1{,}96\sigma < X_n - \mu < 1{,}96\sigma) \approx 0{,}95$$
$$P(\mu - 1{,}96\sigma < X_n < \mu + 1{,}96\sigma) \approx 0{,}95$$
$$P\left(\frac{\mu - 1{,}96\sigma}{n} < \frac{X_n}{n} < \frac{\mu + 1{,}96\sigma}{n}\right) \approx 0{,}95$$
$$P\left(\frac{np - 1{,}96\sqrt{np(1-p)}}{n} < \frac{X_n}{n} < \frac{np + 1{,}96\sqrt{np(1-p)}}{n}\right) \approx 0{,}95$$

D'où :
$$P\left(p - 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}} < F_n < p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \approx 0{,}95.$$

Soit F_n la variable aléatoire qui à tout échantillon de taille n associe la fréquence d'un caractère.

Posons $I_n = \left[\, p - 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \;;\; p + 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \,\right]$ où p désigne la proportion de ce caractère dans la population.

Alors ***F**$_n$* **prend ses valeurs dans I$_n$ avec une probabilité qui s'approche de 0,95 quand *n* devient grand.**

On pratique cette approximation dès que ***n* ≥ 30**, ***np* ≥ 5** et ***n*(1-*p*) ≥ 5.**

L'intervalle $I_n = \left[p - 1{,}96\,\frac{\sqrt{p(1-p)}}{\sqrt{n}} \;;\; p + 1{,}96\,\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ est appelé **intervalle de fluctuation asymptotique au seuil de 95%** .

A RETENIR

Aller voir vidéo : **https://www.youtube.com/watch?v=ycWtHDGTLmM&t=241s**

Application : (Révision des 3 intervalles de fluctuation étudiés en lycée)

On considère un dé cubique supposé équilibré. On désigne par X la variable aléatoire associant à chaque échantillon de 100 lancers le nombre de 6 obtenus. On pose $F = \dfrac{X}{100}$ la variable aléatoire fréquence.

1. Déterminer l'intervalle de fluctuation de la variable aléatoire F au seuil de 95 % vu en Première.

2. Déterminer l'intervalle de fluctuation asymptotique de la variable aléatoire F au seuil de 95 % (arrondir les bornes à 10^{-3}). Interpréter le résultat obtenu.

3. Déterminer l'intervalle de fluctuation de la variable aléatoire F au seuil de 95 % vu en classe de Seconde. (Arrondir les bornes à 10^{-2}).

4. Comparer les trois intervalles.

ATTENTION : Pour les valeurs approchées des bornes : règle :pour la borne inférieure, on prend la valeur approchée par défaut et pour la borne supérieure, la valeur approchée par excès.

SOLUTION

1. X suit la loi binomiale de paramètres $n = 100$ et $p = \frac{1}{6}$.

La calculatrice ou un logiciel (ici un tableur) permet de constater que :

- $P(X \leqslant 9) \approx 0{,}021$ et $P(X \leqslant 10) \approx 0{,}043$
- $P(X \leqslant 23) \approx 0{,}962$ et $P(X \leqslant 24) \approx 0{,}978$.

Le plus petit entier a tel que $P(X \leqslant a) > 0{,}025$ est donc $a = 10$ et le plus petit entier b tel que $P(X \leqslant b) \geqslant 0{,}975$ est $b = 24$. L'intervalle de fluctuation de la variable aléatoire F au seuil de 95 % est donc $\left[\frac{a}{n} ; \frac{b}{n}\right] = [0{,}1 ; 0{,}24]$.

	A	B
9	8	0,0095
10	9	0,0213
11	10	0,0427
12	11	0,0777
13	12	0,1297
23	22	0,9369
24	23	0,9621
25	24	0,9783
26	25	0,9881

2. $p - 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}} = \frac{1}{6} - 1{,}96\frac{\sqrt{\frac{1}{6} \times \frac{5}{6}}}{\sqrt{100}} \approx 0{,}094$ et de même $p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \approx 0{,}240$.

L'intervalle de fluctuation asymptotique de la variable aléatoire F au seuil de 95 % est donc, en arrondissant les bornes à 10^{-3}, l'intervalle $[0{,}093 ; 0{,}240]$.

3. $p - \frac{1}{\sqrt{n}} = \frac{1}{6} - \frac{1}{\sqrt{100}} \approx 0{,}067$ et $p + \frac{1}{\sqrt{n}} = \frac{1}{6} + \frac{1}{\sqrt{100}} \approx 0{,}267$. En classe de Seconde, on aurait donc obtenu l'intervalle $[0{,}067 ; 0{,}267]$ comme intervalle de fluctuation de la variable F au seuil de 95 %.

4. On constate que les intervalles des classes de Première et de Terminale sont très proches, le second étant plus simple à déterminer que le premier. L'intervalle de la classe de Seconde est certes très simple à calculer mais il est bien plus large

2- PRISE DE DECISION A PARTIR D'UN ECHANTILLON.

Voir la vidéo : https://www.youtube.com/watch?v=QZ0YFthGI0Y

La détermination d'un intervalle de fluctuation permet de prendre une décision lorsque **la proportion du caractère étudié dans la population est supposée être égale à p.**
La prise de décision consiste, à partir d'un échantillon de taille n, à valider ou non, cette hypothèse faite sur p.
En pratique :

- On calcule la fréquence observée f du caractère étudié dans l'échantillon.
- Puis, **si $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$**, on détermine l'intervalle I de fluctuation asymptotique au seuil de 95% défini précédemment. (*Les conditions seront toujours vérifiées pour utiliser cet intervalle !!!*)

- Enfin, on applique la règle suivante :
  REGLE DE DECISION :
  Soit I l'intervalle de fluctuation de la fréquence à 95% dans les échantillons de taille n
  - Si la fréquence observée f appartient à I : on accepte l'hypothèse faite sur la proportion p au seuil de risque de 5%.

  - Si la fréquence observée f n'appartient pas à I, on rejette l'hypothèse faite sur la proportion p au seuil de 5%.

Application :

L'organisateur d'une tombola prétend que la proportion des tickets gagnants est $p = 0{,}1$.
Pour vérifier son hypothèse, on prélève au hasard un échantillon de 400 tickets.
On considère que le nombre total de tickets est suffisamment grand pour que le tirage soit assimilé à un tirage avec remise.

1. Sous l'hypothèse $p = 0{,}1$, déterminer l'intervalle de fluctuation asymptotique au seuil de 95 % de la variable aléatoire correspondant à la fréquence des tickets gagnants dans un échantillon de taille 400. (Arrondir les bornes de l'intervalle à 10^{-3}).

2. Énoncer la règle de décision permettant de rejeter, ou non, l'hypothèse $p = 0{,}1$, au seuil de décision de 5 % sur un échantillon aléatoire de taille 400.

3. L'échantillon prélevé comporte 25 tickets gagnants. Que peut-on conclure ?

SOLUTION

1. Avec $p_0 = 0{,}1$ et $n = 400$, on a $p_0 - 1{,}96\dfrac{\sqrt{p_0(1-p_0)}}{\sqrt{n}} \approx 0{,}070$ et $p_0 + 1{,}96\dfrac{\sqrt{p_0(1-p_0)}}{\sqrt{n}} \approx$ **0,130**

Sous l'hypothèse $p = 0{,}1$, l'intervalle de fluctuation asymptotique au seuil de 95 % de la variable aléatoire donnant la fréquence des tickets gagnants sur un échantillon de taille 400 est l'intervalle : **I = [0,070 ; 0,130]**

2. La règle de décision, au seuil de décision de 5 %, est la suivante : on calcule la fréquence f des tickets gagnants sur un échantillon aléatoire de taille 400. Si $f \notin$ I, on rejette l'hypothèse $p = 0{,}1$; si $f \in$ I, on ne la rejette pas.

3. Comme $f = \dfrac{25}{400} = 0{,}0625$ n'appartient pas à l'intervalle I, on rejette l'hypothèse $p = 0{,}1$ au seuil de décision 5 %.
(Ou au seuil de 95%)

II- ESTIMATION D'UNE PROPORTION.

L'**estimation** consiste à déterminer le mieux possible la proportion p inconnue d'un caractère dans la population toute entière, connaissant la fréquence f observée de ce caractère dans un échantillon extrait de cette population.
Ce problème d'estimation se pose en particulier dans les sondages.

Aller voir vidéo : https://www.youtube.com/watch?v=cU5cJlCVAM8

Propriété : Soit F_n la variable aléatoire fréquence, qui, à tout échantillon de taille n extrait d'une population dans lequel la proportion est p, associe la fréquence donnée.
Alors l'intervalle $\left[F_n - \dfrac{1}{\sqrt{n}} ; F_n + \dfrac{1}{\sqrt{n}}\right]$ contient, pour n assez grand, la proportion p avec une probabilité au moins égale à 0,95.

Définition : Soit f la fréquence observée d'un caractère dans un échantillon de taille n extrait d'une population dans laquelle la proportion de ce caractère est p. Alors l'intervalle $\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$ est un **intervalle de confiance** de la proportion p **au niveau de confiance 0,95.** **A RETENIR**

Remarques : 1) Cet intervalle est soumis aux 3 mêmes conditions que celles de l'intervalle de fluctuation. On considèrera qu'elles seront vérifiées.

2) Cet intervalle a une amplitude de $\frac{2}{\sqrt{n}}$ (En effet : $f + \frac{1}{\sqrt{n}} - (f - \frac{1}{\sqrt{n}}) = \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n}} = \frac{2}{\sqrt{n}}$)

Application : Sur un échantillon de 400 personnes, 73 ont vu un certain spot publicitaire à la télévision.

1) Au niveau de confiance de 95%, dans quel intervalle se trouve le pourcentage de ceux qui ont vu le spot dans la population totale ?

$f = \frac{73}{400}$; $n = 400$ donc $I_C = \left[\frac{73}{400} - \frac{1}{\sqrt{400}} ; \frac{73}{400} + \frac{1}{\sqrt{400}}\right]$; **I_C = [0,1325 ; 0,2325]**

Voir la vidéo : https://www.youtube.com/watch?v=ogmMVpkBVgs

2) Combien faudrait-il interroger de personnes pour que cet intervalle ait une amplitude inférieure à 2% ?

On résout : $\frac{2}{\sqrt{n}} < 0{,}02 \Leftrightarrow 2 < 0{,}02\sqrt{n} \Leftrightarrow 100 < \sqrt{n} \Leftrightarrow 1000 < n$

Pour que l'intervalle ait une amplitude inférieure à 2%, il faut interviewer **au moins 1000 personnes.**

POUR Y VOIR CLAIR ENTRE INTERVALLE DE FLUCTUATION ET INTERVALLE DE CONFIANCE :

Aller voir la vidéo : **https://www.youtube.com/watch?v=97vzxWsyie8**

CONCLUSION : (Tout est résumé ici)

REGLE GENERALE :

- On utilise un **intervalle de fluctuation** lorsque la **proportion *p* dans la population est connue** ou si l'on fait une **hypothèse sur sa valeur.**
- On utilise un **intervalle de confiance** lorsque l'on veut estimer une **proportion *p* inconnue** dans une population.

BILAN :

Bien que ces deux notions soient voisines, il ne faut pas confondre proportion *p* et fréquence *f*:
• La **proportion** est une valeur statistique (théorique), **sur l'ensemble de la population**
• La **fréquence** est la proportion observée **sur un échantillon de la population**

Intervalle de fluctuation

Ce qui est connu ou supposé:

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

- probabilité (proportion) *p*,
- taille de l'échantillon *n*

But : Estimer une fréquence *f* à partir d'une probabilité.
Construire un intervalle à l'aide de la probabilité *p* centré. contenant les fréquences observées à 95%.

La proportion dans la population est connue (ou supposée).
On tire un échantillon. La proportion dans l'échantillon ne sera que par hasard égale à celle de la population. En général la proportion dans l'échantillon sera différente, mais la plupart du temps (dans 95 % des cas par exemple) elle ne sera pas trop différente de la proportion dans la population.
L'intervalle de fluctuation est l'intervalle dans lequel on trouve (pour 95 % des cas) la proportion dans l'échantillon ; cet intervalle est calculé à partir de la proportion dans la population qui est connue.

Intervalle de confiance

Ce qui est connu :

$$\left[f - \frac{1}{\sqrt{n}}\ ;\ f + \frac{1}{\sqrt{n}}\right]$$

- fréquence *f*
- taille de l'échantillon *n*

But : Estimer une proportion (probabilité) *p* à partir des observations.
Construire un intervalle à l'aide de la fréquence *f* contenant la probabilité (ou la proportion inconnue) avec un niveau de confiance de 95%.

La proportion dans la population est inconnue.
On tire un échantillon. On compte la proportion dans l'échantillon. On se demande s'il est possible d'estimer à partir de cette valeur connue par l'échantillon, la valeur inconnue de la proportion dans la population. On sait bien qu'il est tout à fait possible que l'estimation soit fausse. On calcule un intervalle pour la proportion dans la population : c'est l'intervalle de confiance. Avec la formule que tu emploies on trouve un intervalle qui encadre bien la valeur inconnue de la population dans 95 % des cas.