

Intervalle de fluctuation

1 Loi binomiale et intervalle de fluctuation au seuil de $1 - \alpha$

1.1 Définition et premier exemple

Un médecin de santé publique veut savoir si, dans sa région, le pourcentage d'habitants atteints d'hypertension artérielle est égale à la valeur de 16% récemment publiée pour des populations semblables.

On note p la proportion d'hypertendus dans la population de sa région, ici on suppose donc que $p = 0,16$.

La question qu'on peut alors se poser est comment valider ou non cette hypothèse?

Le médecin va donc constituer un échantillon de $n = 100$ personnes choisies au hasard (la population est suffisamment importante pour considérer qu'il s'agit de tirages avec remise).

1. Soit X la variable aléatoire correspondant au nombre d'hypertendus observé dans un échantillon de taille $n = 100$. Déterminer la loi de X .

On cherche à déterminer l'intervalle de fluctuation au seuil de $1 - \alpha$, relatif à cet échantillon de taille $n = 100$, c'est à dire l'intervalle centré autour de p qui contient la fréquence observée f avec une probabilité égale à $1 - \alpha$. Dans notre exemple, on prend $\alpha = 0,05$ soit $1 - \alpha = 0,95$.

2. Déterminer le nombre d'hypertendus correspondant à la proportion d'hypertendus pour un échantillon de taille $n = 100$.
3. Déterminer le plus petit entier a tel que $P(X \leq a) > 0,025$.
4. Déterminer le plus petit entier b tel que $P(X \leq b) \geq 0,975$.
5. En déduire une minoration de $P(a \leq X \leq b)$.

Ainsi, la fréquence f observée par le médecin a une probabilité de 0,95 de vérifier $a \leq nf \leq b$ soit $\frac{a}{n} \leq f \leq \frac{b}{n}$. On en déduit que l'intervalle de fluctuation au seuil de 95% de la proportion d'hypertendus $p = 0,16$ est l'intervalle

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

6. Le médecin recense 24 hypertendus sur son échantillon de taille $n = 100$. Quelle décision peut-il prendre au seuil de 95%?
7. Le médecin réalise un nouvel échantillon de taille $n = 1000$ et recense 179 hypertendus. Ce résultat confirme-t'il le précédent au seuil de 95%?
8. Faire la même étude au seuil de 99%.

1.2 Cours

Au sein d'une population, on suppose que la proportion d'un certain caractère est p . On souhaite juger cette hypothèse. Pour cela on prélève dans la population au hasard et avec remise, un échantillon de taille n sur lequel on observe une fréquence f de ce caractère.

Définition:

L'intervalle de fluctuation au seuil de $1 - \alpha$ d'une fréquence, sur un échantillon aléatoire de taille n , selon la loi binomiale de paramètres n et p est :

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

avec a le plus petit entier tel que $P(X \leq a) > \frac{\alpha}{2}$ et b le plus petit entier tel que $P(X \leq b) \geq 1 - \frac{\alpha}{2}$.

Règle:

- Si $f \in \left[\frac{a}{n}; \frac{b}{n} \right]$, l'hypothèse est acceptée au seuil de $1 - \alpha$;
- Si $f \notin \left[\frac{a}{n}; \frac{b}{n} \right]$, l'hypothèse est rejetée au seuil de $1 - \alpha$.

1.3 Applications

Exercice 1:

Monsieur M , chef du gouvernement d'un pays lointain, affirme que 52% des électeurs lui font confiance. On interroge 100 électeurs au hasard (la population est suffisamment importante pour considérer qu'il s'agit de tirages avec remise) et on souhaite savoir à partir de quelles fréquences, au seuil de 95%, on peut mettre en doute le pourcentage annoncé par Monsieur M , dans un sens, ou dans l'autre.

1. On fait confiance à l'hypothèse de Monsieur M et soit X la variable aléatoire correspondant au nombre d'électeurs qui lui font confiance dans un échantillon de taille $n = 100$. Déterminer la loi de X .
2. Déterminer l'intervalle de fluctuation au seuil de 95% obtenu grâce à la loi binomiale.
3. Comparer avec l'intervalle de fluctuation au seuil de 95% donnée en seconde.
4. Sur les 100 électeurs interrogés au hasard, 43 déclarent avoir confiance en Monsieur M . Peut-on considérer, au seuil de 95%, l'affirmation de Monsieur M comme exacte ?

Exercice 2:

En novembre 1976 dans un comté du sud du Texas, Rodrigo Partida est condamné à huit ans de prison. Il attaque ce jugement au motif que la désignation des jurés de ce comté est, selon lui, discriminante à l'égard des Américains d'origine mexicaine. Alors que 80% de la population du comté est d'origine mexicaine, sur les 870 personnes convoquées pour être jurés lors des années précédentes, il n'y a eu que 339 personnes d'origine mexicaine.

Devant la Cour Suprême, un expert statisticien produit des arguments pour convaincre du bien fondé de la requête de l'accusé. En vous situant dans le rôle de cet expert, pouvez-vous décider si les Américains d'origine mexicaine sont sous-représentés dans les jurys de ce comté ?

Exercice 3:

Voici les résultats du premier tour de l'élection présidentielle de 2002, où un total de 28.498.471 bulletins de votes furent comptabilisés :

Nom	CHIRAC	LE PEN	JOSPIN	BAYROU	LAGUILLER
Prénom	Jacques	Jean-Marie	Lionel	François	Arlette
Partie	RPR	FN	PS	UDF	LO
% des voix exprimés	19,88%	16,86%	16,18%	6,84%	5,72%

1. Déterminer le nombre de voix obtenues par chacun de ces cinq candidats.

Voici les résultats de sondages effectués avant ce premier tour sur un panel représentatif de 1.000 électeurs. Chaque sondage correspond ainsi à un échantillon de 1.000 personnes.

Nom	CHIRAC	LE PEN	JOSPIN
Prénom	Jacques	Jean-Marie	Lionel
10-11 avril - CSA	21%	12%	19%
13-15 avril - Sofres	20%	13%	18%
17-18 avril - CSA	19,5%	14%	18%

2. Déterminer, en utilisant la loi binomiale, l'intervalle de fluctuation au seuil de 95% pour chaque candidat.
3. Déterminer les échantillons que l'on peut valider au seuil de 95%.
4. Conclure.

Exercice 4:

Un groupe de citoyens demande à la municipalité d'une ville la modification d'un carrefour en affirmant que 40% des automobilistes tournent en utilisant une mauvaise file. Un officier de police constate que sur 500 voitures prises au hasard, 190 prennent une mauvaise file.

1. Déterminer, en utilisant la loi binomiale sous l'hypothèse $p = 0,4$, l'intervalle de fluctuation au seuil de 95%.
2. D'après l'échantillon, peut-on considérer, au seuil de 95%, comme exacte l'affirmation du groupe de citoyens ?

2 Loi normale

On rappelle ci-dessous l'énoncé du théorème de Moivre-Laplace (hors-programme) :

Théorème:

Soit, pour tout entier n , une variable aléatoire X_n qui suit une loi binomiale $\mathcal{B}(n, p)$ et $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$, la variable centrée réduite associée à X_n . Alors pour tous nombres a et b tels que $a \leq b$,

$$\lim_{n \rightarrow +\infty} P(a \leq Z_n \leq b) = P(a \leq Z \leq b)$$

où Z suit la loi normale centrée réduite $\mathcal{N}(0, 1)$.

Exercice 5:

Soit la variable aléatoire Z qui suit la loi normale centrée réduite $\mathcal{N}(0, 1)$ et $\alpha \in]0; 1[$. On peut démontrer qu'il existe un unique réel u_α tel que $P(-u_\alpha \leq Z \leq u_\alpha) = 1 - \alpha$. Déterminer $u_{0,05}$ et $u_{0,01}$.

Théorème:

Soit, pour tout entier n , une variable aléatoire X_n qui suit une loi binomiale $\mathcal{B}(n, p)$ alors $\lim_{n \rightarrow +\infty} P\left(\frac{X_n}{n} \in I_n\right) = 1 - \alpha$ où

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

Définition:

Sous les conditions énoncées dans le théorème ci-dessus, on dit que l'intervalle I_n est un intervalle de fluctuation asymptotique de la variable aléatoire fréquence $F_n = \frac{X_n}{n}$ au seuil de $1 - \alpha$. Ainsi, pour n grand, la variable aléatoire F_n prend ses valeurs dans l'intervalle I_n avec une probabilité proche de $1 - \alpha$. En pratique, on utilise cet intervalle de fluctuation de F_n dès que $n \geq 30$; $np \geq 5$ et $n(1-p) \geq 5$.

Exercice 6:

Dans l'exercice 3, déterminer les intervalles de fluctuation asymptotique au seuil de 0,95 et au seuil de 0,99 pour chaque candidat. Conclure.

Exercice 7:

Une compagnie aérienne possède des avions d'une capacité de 300 places. Cette compagnie a vendu n billets pour le vol 2013. La probabilité pour qu'un acheteur se présente à l'embarquement est p et les comportements des acheteurs sont indépendants les uns des autres. On note X_n la variable aléatoire désignant le nombre d'acheteurs d'un billet se présentant à l'embarquement. La compagnie cherche à optimiser le remplissage de l'avion en vendant éventuellement plus de places que la capacité totale de l'avion (surréservation ou surbooking) soit ici $n > 300$. Comme il y a évidemment un risque que le nombre de passagers munis d'un billet se présentant à l'embarquement excède 300, la compagnie veut maîtriser ce risque.

- Déterminer la loi de X_n .
- On suppose $p \in [0, 5; 0, 95]$. Écrire l'intervalle de fluctuation asymptotique I_n de $\frac{X_n}{n}$ au seuil de 0,95.
- Montrer que si $I_n \subset \left[0; \frac{300}{n}\right]$ alors la probabilité que le nombre de passagers se présentant à l'embarquement excède 300 est proche de 0,05.
- On cherche à déterminer la valeur de n maximale permettant de satisfaire la condition de l'inclusion $I_n \subset \left[0; \frac{300}{n}\right]$.
 - Montrer que $I_n \subset \left[0; \frac{300}{n}\right] \implies pn + 1,96\sqrt{n}\sqrt{p(1-p)} - 300 \leq 0$.
 - On pose $f(x) = px + 1,96\sqrt{x}\sqrt{p(1-p)} - 300$. Montrer qu'il existe un entier n_0 tel que si $n \leq n_0$ alors $f(n) \leq 0$ et si $n > n_0$ alors $f(n) > 0$.
 - Tracer la courbe représentative de f pour les valeurs $p = 0,85$; $p = 0,9$; $p = 0,95$.
 - Déterminer à la calculatrice les valeurs de n_0 pour $p = 0,85$; $p = 0,9$; $p = 0,95$.