Chapitre 14 : Intervalle de fluctuation et estimation.

I Attendus

- Savoir déterminer un intervalle de fluctuation (A partir donc de la proportion supposée et de l'effectif) *(1 page 423)*
- Prise de décision : Déterminer à partir de la proportion supposée et de la fréquence de l'échantillon savoir déterminer si la valeur supposée et *Acceptée* ou *Rejetée*. *(1 page 423)*
- Savoir déterminer un intervalle de confiance (A partir de la fréquence d'un échantillon et de son effectif). Ainsi faire une estimation de la proportion dans la population. *(Ex 3 page 425)*

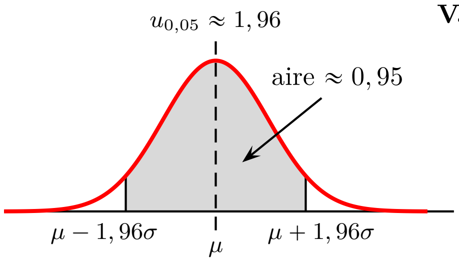
II Compléments sur les lois normale et binomiale

A Loi normale : intervalles centrés sur l'espérance

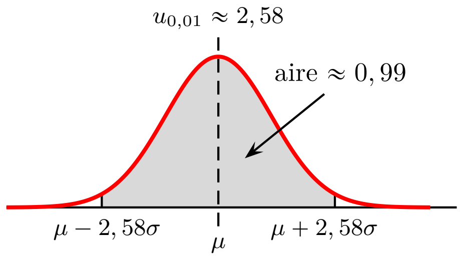
Théorème 1

Soit X une variable aléatoire qui suit la loi $\mathcal{N}(\mu;\sigma^2)$.
Pour tout nombre α de $]0;1[$, $P(X \in [\mu - \sigma u_\alpha \ ; \ \mu + \sigma u_\alpha]) = 1 - \alpha$, où u_α est l'unique nombre strictement positif tel que $P\left(-u_\alpha \leqslant \dfrac{X-\mu}{\sigma} \leqslant u_\alpha\right) = 1 - \alpha$.

Valeurs remarquables

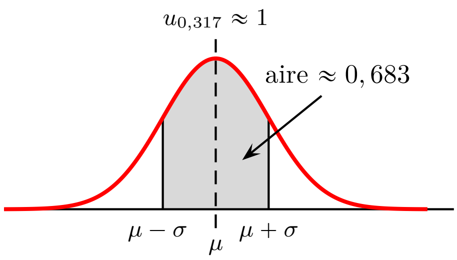


$P(\mu - 1,96\sigma \leqslant X \leqslant \mu + 1,96\sigma) \approx 0,95$

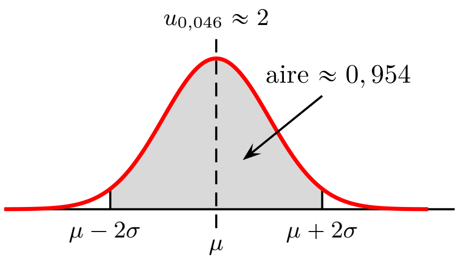


$P(\mu - 2,58\sigma \leqslant X \leqslant \mu + 2,58\sigma) \approx 0,99$

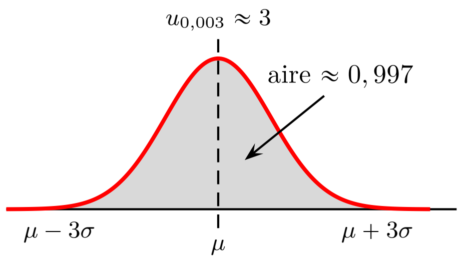
Valeurs approchées à connaître



$P(X \in [\mu - \sigma; \mu + \sigma]) \approx 0,683$



$P(X \in [\mu - 2\sigma; \mu + 2\sigma]) \approx 0,954$



$P(X \in [\mu - 3\sigma; \mu + 3\sigma]) \approx 0,997$

B Variable aléatoire fréquence F_n

Soit X_n la variable aléatoire qui, à tout échantillon de taille n étudié, associe le nombre d'individus qui possèdent le caractère étudié.
X_n suit la loi . . . .

Définition 1

La variable aléatoire F_n qui, à tout échantillon de taille n, associe la fréquence f_n du caractère étudié dans cet échantillon est appelée **variable aléatoire fréquence** et elle est définie par $F_n = \dfrac{X_n}{n}$.

Remarque : X_n prend les valeurs entières de 0 à n, donc F_n prend les valeurs ....
Par conséquent, F_n est une variable aléatoire .... mais pour $n \geqslant 2$, cette variable aléatoire ne suit pas une loi binomiale (valeurs non toutes entières).

Pour tout entier k compris entre 0 et n, $P(F_n = \dfrac{k}{n}) = P(\ldots\ldots\ldots\ldots) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$

L'espérance de la variable aléatoire fréquence est $E(F_n) = \ldots\ldots$ et son écart-type est $\sigma(F_n) = \ldots\ldots\ldots\ldots\ldots\ldots$

C Approximation de la loi de F_n par une loi normale

Proposition 2

Sous les conditions $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$, la loi de la variable aléatoire fréquence F_n peut être approchée par la loi normale de même espérance et de même variance $\mathcal{N}\left(p\ ;\ \dfrac{p(1-p)}{n}\right)$.

Exemple 1. On lance 100 fois une pièce équilibrée et on observe la fréquence d'apparition du Pile.

F_{100} indique la fréquence de succès lors des 100 épreuves et on a $p = \ldots\ldots$

$E(F_{100}) = \ldots\ldots$ et $\sigma(F_{100}) = \ldots\ldots\ldots\ldots\ldots\ldots$

$n = \ldots\ldots \geqslant 30$, $np = \ldots\ldots \geqslant 5$ et $n(1-p) = \ldots\ldots \geqslant 5$.

Les 3 conditions de l'approximation sont vérifiées donc on peut approcher la loi de F_{100} par la loi ....

III Intervalle de fluctuation et prise de décision

A Intervalle de fluctuation asymptotique au seuil de $1-\alpha$

Théorème 3

Pour tout nombre α de $]0;1[$, on pose $I_n = \left[p - u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$.

Alors, la probabilité $P(F_n \in I_n)$ tend vers $1-\alpha$ lorsque n tend vers $+\infty$: $\lim\limits_{n\to+\infty} P(F_n \in I_n) = 1-\alpha$.

Définition 2

Pour tout nombre α de $]0;1[$, $I_n = \left[p - u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ est appelé **intervalle de fluctuation asymptotique de F_n au seuil de** $1-\alpha$.

B Intervalle de fluctuation asymptotique au seuil de 95%

Comme $u_{\ldots\ldots} \approx \ldots\ldots$, un intervalle de fluctuation asymptotique de F_n au seuil de 95% est :

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

Remarque : L'intervalle de fluctuation asymptotique au seuil de 95% est inclus dans l'intervalle de fluctuation étudié en classe de Seconde, qui est défini par : $\left[p - \dfrac{1}{\sqrt{n}}\ ;\ p + \dfrac{1}{\sqrt{n}}\right]$.

C Prise de décision

Dans une population d'effectif N, **on suppose que la proportion d'un certain caractère est** p.
On souhaite valider ou non cette hypothèse.

Pour cela, on prélève dans la population, au hasard et avec remise, un échantillon de taille n, sur lequel on observe la fréquence f_{obs} de ce caractère.

f_{obs} est une réalisation de la variable aléatoire donnant la fréquence de succès $F_n = \dfrac{X_n}{n}$, où X_n, qui indique le nombre d'individus présentant le caractère étudié, suit la loi binomiale $\mathscr{B}(n,p)$.

Si les conditions d'approximation $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$ sont remplies, alors dans environ 95% des cas, la fréquence F_n est dans l'intervalle de fluctuation asymptotique : $I_n = \left[p - 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$.

Règle de prise de décision

- Si $f_{obs} \in I$, alors on **accepte** l'hypothèse que la proportion est p.
- Si $f_{obs} \notin I$, alors on **rejette** cette hypothèse, au seuil de 5%.

Cela signifie que le risque d'erreur est d'environ 5%.

Ex 11 à 18 page 427

IV Estimation et intervalle de confiance

La proportion p est ici inconnue.
On cherche à estimer p à partir d'un échantillon de taille n (tirage au hasard et avec remise).
On suppose que $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$.
Soit F_n la variable aléatoire qui, à chacun des échantillons de taille n, associe la fréquence du caractère dans cet échantillon.

Dans environ 95% des cas, F_n prend ses valeurs dans l'intervalle $\left[p - \dfrac{1}{\sqrt{n}}\ ;\ p + \dfrac{1}{\sqrt{n}}\right]$. On obtient alors le résultat suivant.

Proposition 4

La proportion inconnue p est telle que $P\left(F_n - \dfrac{1}{\sqrt{n}} \leqslant p \leqslant F_n + \dfrac{1}{\sqrt{n}}\right) \approx 0,95$.

Définition 3

On observe une fréquence f_{obs} sur un échantillon de taille n.
On appelle **intervalle de confiance de p au niveau de confiance de 95% l'intervalle :**

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

Règle

Pour estimer une proportion inconnue p à partir d'un échantillon, on utilise en général l'encadrement fourni par un intervalle de confiance au niveau de confiance de 95%.

Remarque : L'intervalle $\left[f_{obs} - 1,96\dfrac{\sqrt{f_{obs}(1-f_{obs})}}{\sqrt{n}}\ ;\ f_{obs} + 1,96\dfrac{\sqrt{f_{obs}(1-f_{obs})}}{\sqrt{n}}\right]$ est aussi un intervalle de confiance de p au niveau de confiance de 95%.

Ex 19 à 23 page 428