EXERCICE 1 :

Dans un pays, 10% des plages étaient atteintes par des algues toxiques.
On a modifié le processus de rejets chimiques : on admet que le nouveau processus de rejet, très différent du précédent, pourrait modifier cette proportion.
On prend un échantillon aléatoire de 150 prélèvements, on constate que 18 présentent des traces d'algues toxiques. Peut-on penser que le nouveau traitement a un impact sur le pourcentage de plages polluées ?

Pour répondre à la question, va-t-on utiliser un intervalle de fluctuation ou un intervalle de confiance ?

Pour répondre à la question, va-t-on utiliser un intervalle de fluctuation ou un intervalle de confiance ?

Réponse : On connaît la proportion de plages atteintes.

Pour répondre à la question, va-t-on utiliser un intervalle de fluctuation ou un intervalle de confiance ?

Réponse : On connaît la proportion de plages atteintes.

DONC

Fluctuation d'échantillonnage

Intervalles de fluctuation :
Dans une population, un caractère est présent dans une proportion p.
A chaque échantillon (avec remise) de taille n, on associe la fréquence de ce caractère dans l'échantillon. On définit ainsi une variable aléatoire F_n.
Définir un intervalle de fluctuation de F_n au seuil $1-\alpha$, c'est déterminer un intervalle I_n tel que $P(F_n \in I_n) \geqslant 1-\alpha$.

Remarque : on choisit souvent $\alpha = 5\%$.

Peut-on utiliser le cours de seconde pour répondre à la question ?
Pourquoi ?

Peut-on utiliser le cours de première pour répondre à la question ?
Inconvénient ?

Réponse :

X suit une loi binomiale de paramètres $n = 150$ et $p = 0,1$

$P(X \leqslant 7) \approx 0,014$

$P(X \leqslant 8) \approx 0,0307$

La borne inférieure de l'intervalle de fluctuation est donc $\dfrac{8}{150}$.

Réponse :

X suit une loi binomiale de paramètres $n = 150$ et $p = 0,1$

$P(X \leqslant 7) \approx 0,014$

$P(X \leqslant 8) \approx 0,0307$

La borne inférieure de l'intervalle de fluctuation est donc $\dfrac{8}{150}$.

Réponse :

X suit une loi binomiale de paramètres $n = 150$ et $p = 0,1$

$P(X \leqslant 7) \approx 0,014$

$P(X \leqslant 8) \approx 0,0307$

La borne inférieure de l'intervalle de fluctuation est donc $\dfrac{8}{150}$.

X suit une loi binomiale de paramètres $n = 150$ et $p = 0,1$

$P(X \leqslant 22) \approx 0,974$

$P(X \leqslant 23) \approx 0,985$

La borne supérieure de l'intervalle de fluctuation est donc $\dfrac{23}{150}$.

Réponse :

X suit une loi binomiale de paramètres $n = 150$ et $p = 0,1$

$P(X \leqslant 7) \approx 0,014$

$P(X \leqslant 8) \approx 0,0307$

La borne inférieure de l'intervalle de fluctuation est donc $\dfrac{8}{150}$.

X suit une loi binomiale de paramètres $n = 150$ et $p = 0,1$

$P(X \leqslant 22) \approx 0,974$

$P(X \leqslant 23) \approx 0,985$

La borne supérieure de l'intervalle de fluctuation est donc $\dfrac{23}{150}$.

L'intervalle de fluctuation est donc $\left[\dfrac{8}{150}; \dfrac{23}{150}\right]$

On considère l'échantillon de $n = 150$ plages et on détermine la fréquence de plages polluées : $f = \dfrac{18}{150} = 0,12$.

On considère l'échantillon de $n = 150$ plages et on détermine la fréquence de plages polluées : $f = \dfrac{18}{150} = 0,12$.

On prend une décision :

$f = \dfrac{18}{150}$ appartient à l'intervalle : $\left[\dfrac{8}{150}; \dfrac{23}{150}\right]$

On considère l'échantillon de $n = 150$ plages et on détermine la fréquence de plages polluées : $f = \dfrac{18}{150} = 0,12$.

On prend une décision :

$f = \dfrac{18}{150}$ appartient à l'intervalle : $\left[\dfrac{8}{150}; \dfrac{23}{150}\right]$

On en déduit que :

au seuil de 95%,on ne rejette pas l'hypothèse que p $= 10\%$.

Le hasard seul peut expliquer la différence entre les valeurs 10% de p et 12% de f.

Avantage :
L'intervalle obtenu est exact.

Avantage :
L'intervalle obtenu est exact.

Inconvénient :
La méthode est longue.

Définition

En classe de terminale, on utilise un intervalle de fluctuation asymptotique.
La **variable aléatoire** $F_n = \dfrac{X_n}{n}$ donne donc la fréquence du nombre de « succès ».

Un intervalle de fluctuation asymptotique de la variable aléatoire $F_n = \dfrac{X_n}{n}$ au seuil de 95 %,est un intervalle déterminé à partir de p et de n et qui contient F_n avec une probabilité d'autant plus proche de 95 % que n est grand.

Définition

En classe de terminale, on utilise un intervalle de fluctuation asymptotique.
La **variable aléatoire** $F_n = \frac{X_n}{n}$ donne donc la fréquence du nombre de « succès ».

Un intervalle de fluctuation asymptotique de la variable aléatoire $F_n = \frac{X_n}{n}$ au seuil de 95 %,est un intervalle déterminé à partir de p et de n et qui contient F_n avec une probabilité d'autant plus proche de 95 % que n est grand.
On peut considérer que les intervalles de fluctuation asymptotique sont des intervalles de fluctuation « approchés ».

Conditions d'utilisation

A vérifier impérativement !

- $n \geqslant 30$.

Conditions d'utilisation

A vérifier impérativement !

- $n \geqslant 30$.
- $np \geqslant 5$.

Conditions d'utilisation

A vérifier impérativement !

- $n \geqslant 30$.
- $np \geqslant 5$.
- $n(1-p) \geqslant 5$.

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

Retour à l'exercice

Un intervalle de fluctuation asymptotique à 95 % est :
$I = \left[0,1 - 1,96\frac{\sqrt{0{,}1\times0{,}9}}{\sqrt{150}} ; 0,1 + 1,96\frac{\sqrt{0{,}1\times0{,}9}}{\sqrt{150}}\right]$.

Soit :
$I = [0,051 ; 0,149]$.

Retour à l'exercice

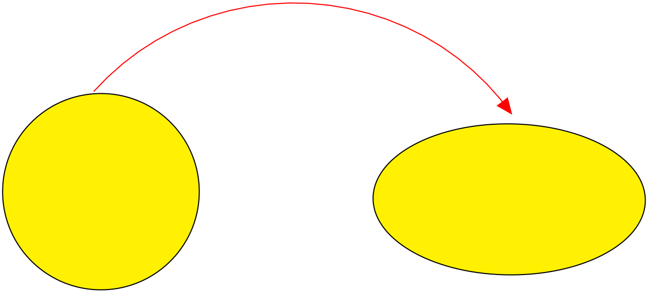
Un intervalle de fluctuation asymptotique à 95 % est :
$I = \left[0,1 - 1,96\frac{\sqrt{0{,}1\times0{,}9}}{\sqrt{150}} ; 0,1 + 1,96\frac{\sqrt{0{,}1\times0{,}9}}{\sqrt{150}}\right]$.

Soit :
$I = [0,051; 0,149]$.
La conclusion est la même que précédemment.

Intervalle de fluctuation

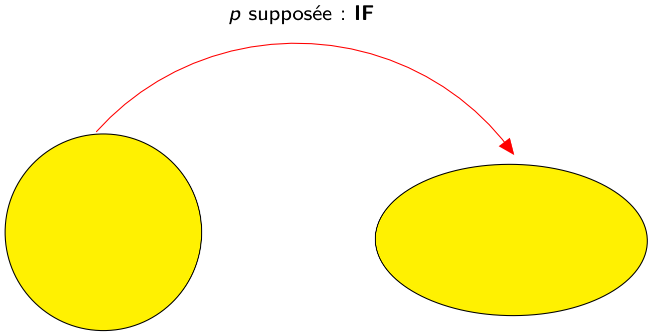
Intervalle de fluctuation = prise de décision sur la qualité d'un échantillon.

p connue : **IF**



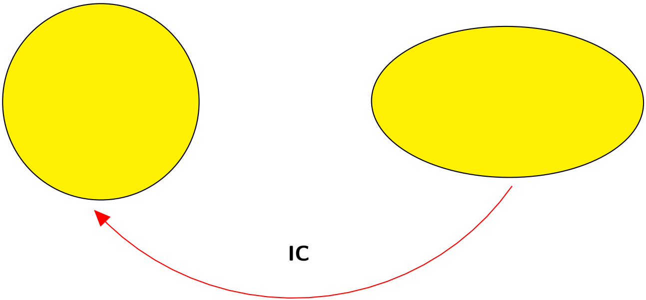
Intervalle de fluctuation

Intervalle de fluctuation = prise de décision (confirmation ou non de la supposition).



Intervalle de confiance

Intervalle de confiance = estimation de p.



- ***Soit :***
  on connaît la proportion p de présence du caractère **ou on fait une hypothèse sur cette valeur**.
  Exemple 1 : Le tirage successif avec remise de n boules d'une couleur donnée dont on connaît la proportion dans l'urne.
  Exemple 2 : Pour décider si une pièce est équilibrée ou non , on fait l'hypothèse que la fréquence d'apparition du « pile »est égale à 0,5 et teste cette hypothèse.
  Dans ce cas, on est dans le domaine de **l'échantillonnage** et on utilise un **intervalle de fluctuation**.

Cours

Echantillonage et estimation

- ***Soit :***
  on connaît la proportion p de présence du caractère **ou on fait une hypothèse sur cette valeur**.
  Exemple 1 : Le tirage successif avec remise de n boules d'une couleur donnée dont on connaît la proportion dans l'urne.
  Exemple 2 : Pour décider si une pièce est équilibrée ou non , on fait l'hypothèse que la fréquence d'apparition du « pile »est égale à 0,5 et teste cette hypothèse.
  Dans ce cas, on est dans le domaine de **l'échantillonnage** et on utilise un **intervalle de fluctuation**.
- ***Soit :***
  on ignore la proportion et on ne formule pas d'hypothèse sur cette valeur. Dans ce cas, on est dans le domaine de **l'estimation** et on utilise un **intervalle de confiance**.
  Exemple : Proportion d'objets défectueux dans une production.

Lancer de pièce :

Exemple

Un enfant a effectué 254 *lancers d'une pièce de monnaie et il a obtenu* 129 *pile. La pièce est-elle équilibrée ?*

Lancer de pièce :

Exemple

Un enfant a effectué 254 *lancers d'une pièce de monnaie et il a obtenu* 129 *pile. La pièce est-elle équilibrée ?*

On fait l'hypothèse que la pièce est équilibrée. La proportion est donc supposée égale à $\dfrac{1}{2}$.

Lancer de pièce :

Exemple

Un enfant a effectué 254 *lancers d'une pièce de monnaie et il a obtenu* 129 *pile. La pièce est-elle équilibrée ?*

On fait l'hypothèse que la pièce est équilibrée. La proportion est donc supposée égale à $\frac{1}{2}$.

DONC :

Fluctuation

Définition 1.

Un *échantillon* de taille n est la liste des résultats d'un schéma de Bernoulli de paramètres $(n; p)$. (répétition de n expériences à deux issues, de probabilité de succès p, identiques et indépendantes).
La *fréquence observée* des résultats d'un échantillon de taille n est

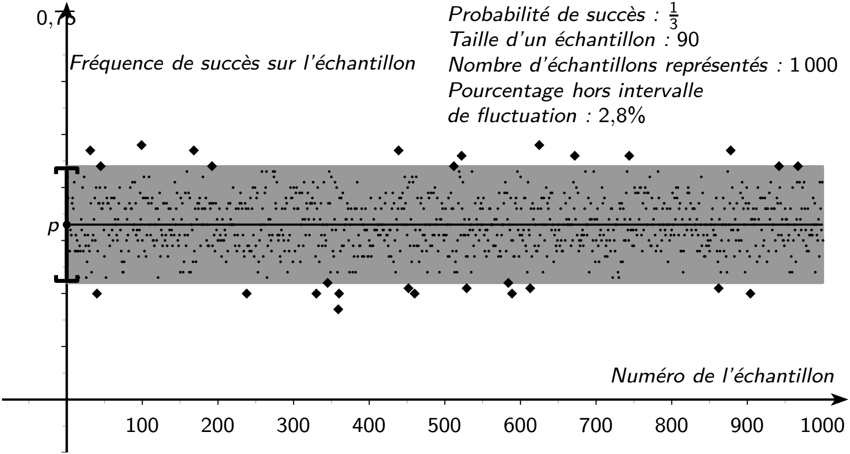
$$f = \frac{\text{nombre de succès}}{n}$$

Remarque

si $X_n \sim \mathcal{B}(n; p)$, *la variable aléatoire égale à la fréquence est* $F_n = \dfrac{X_n}{n}$.

Lancer d'un dé équilibré : obtention du 1 ou du 6.

Exemple



Théorème 1.

Pour $n \geqslant 1$ entier, soit $X_n \sim \mathcal{B}(n; p)$.
Soit $\alpha \in]0; 1[$ et u_α l'unique réel tel que $P(Y \in [-u_\alpha; u_\alpha]) = 1 - \alpha$ où $Y \sim \mathcal{N}(0; 1)$.

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

Définition 2.

On considère $X_n \sim \mathcal{B}(n; p)$ et α un réel de l'intervalle $]0; 1[$.

L'intervalle $I_n = \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}; p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ est un *intervalle de fluctuation asymptotique*, au seuil $1-\alpha$, de la fréquence de succès d'un schéma de Bernoulli de paramètres $(n; p)$.Cet intervalle est centré sur p.

Remarque

Pour $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$, on considérera que $P(F_n \in I_n) \approx 1 - \alpha$.

Remarque

Pour $n \geqslant 30$*,* $np \geqslant 5$ *et* $n(1-p) \geqslant 5$*, on considérera que* $P(F_n \in I_n) \approx 1 - \alpha$.

Remarque

Pour $\alpha = 0,05$*,* $u_\alpha \approx 1,96$*. On prend pour intervalle de fluctuation au seuil de 95 % de* $\dfrac{X_n}{n}$ *l'intervalle :* $J_n = \left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$.

Exemple : On tire 50 boules avec remise dans une urne qui contient des boules rouges dans une proportion de $0,4$. Déterminer un intervalle de fluctuation au seuil $0,9$ de la variable aléatoire fréquence de boules rouges.

Exemple : On tire 50 boules avec remise dans une urne qui contient des boules rouges dans une proportion de $0,4$. Déterminer un intervalle de fluctuation au seuil $0,9$ de la variable aléatoire fréquence de boules rouges.

Même question pour un tirage de 500 boules.

Théorème 2.

Pour n assez grand, $P\left(\frac{X_n}{n} \in [p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}]\right) \geqslant 0{,}95$.
L'intervalle $J_n = \left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}; p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ est inclus dans l'intervalle de fluctuation au seuil de 95% vu en classe de seconde.

Règle de décision : La proportion du caractère étudié est supposé égale à p. La prise de décision consiste à valider ou pas cette hypothèse.
Pour cela,

Règle de décision : La proportion du caractère étudié est supposé égale à p. La prise de décision consiste à valider ou pas cette hypothèse.
Pour cela,

- on calcule la fréquence f du caractère étudié.

Règle de décision : La proportion du caractère étudié est supposé égale à p. La prise de décision consiste à valider ou pas cette hypothèse.
Pour cela,

- on calcule la fréquence f du caractère étudié.
- on vérifie si les conditions sur les paramètres sont vérifiées : $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$(sinon, on procède comme en seconde ou en première)

Règle de décision : La proportion du caractère étudié est supposé égale à p. La prise de décision consiste à valider ou pas cette hypothèse.
Pour cela,

- on calcule la fréquence f du caractère étudié.
- on vérifie si les conditions sur les paramètres sont vérifiées : $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$(sinon, on procède comme en seconde ou en première)
- ensuite, si f appartient à l'intervalle de fluctuation asymptotique au seuil $0,95$, on accepte l'hypothèse sur la proportion p (on peut se tromper mais le risque n'est pas quantifié)

Règle de décision : La proportion du caractère étudié est supposé égale à p. La prise de décision consiste à valider ou pas cette hypothèse.
Pour cela,

- on calcule la fréquence f du caractère étudié.
- on vérifie si les conditions sur les paramètres sont vérifiées : $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$(sinon, on procède comme en seconde ou en première)
- ensuite, si f appartient à l'intervalle de fluctuation asymptotique au seuil $0,95$, on accepte l'hypothèse sur la proportion p (on peut se tromper mais le risque n'est pas quantifié)
- si f n'appartient pas à l'intervalle au seuil 0,95, on rejette l'hypothèse faite sur la proportion p car dans 5% des cas, la fréquence n'est pas dedans : on en conclut donc que c'est étrange...

Intervalle de confiance

Définition 3.

Quel que soit $\alpha \in]0;1[$, un intervalle de confiance pour une proportion p, au niveau de confiance $1-\alpha$, est un intervalle aléatoire (puisqu'il dépend de l'échantillon obtenu lors d'un sondage) et qui contient p dans $100 \times (1-\alpha)\%$ des sondages de ce type.

Théorème 3.

Soit $X_n \sim \mathcal{B}(n;p)$ et p la proportion inconnue d'apparition d'un caractère.Pour n assez grand, $P\left(\dfrac{X_n}{n} - \dfrac{1}{\sqrt{n}} \leqslant p \leqslant \dfrac{X_n}{n} + \dfrac{1}{\sqrt{n}}\right) \geqslant 0{,}95$.

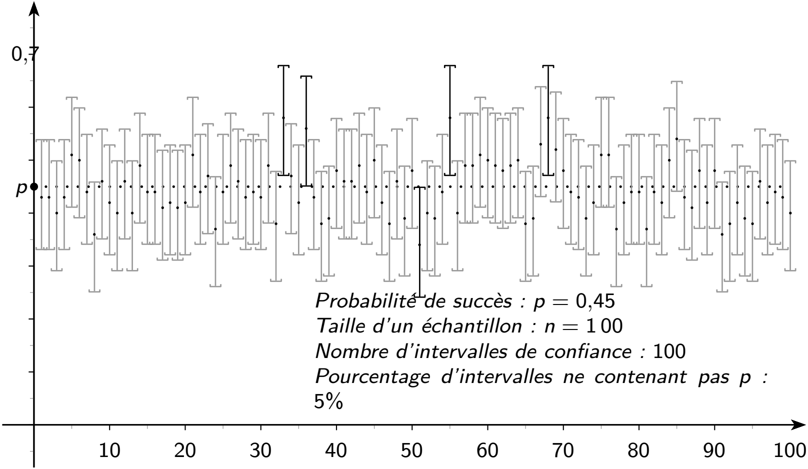
Définition 4.

Soit f la fréquence observée d'un caractère dans un échantillon de taille n extrait d'une population dans lequel la proportion de ce caractère est p.
Alors $I_{95\%} = \left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$ est un *intervalle de confiance* au seuil de 0,95. L'amplitude est de $\dfrac{2}{\sqrt{n}}$.

Remarque

Le nombre p appartiendra à 95% *des intervalles de confiance ainsi obtenus.*

Exemple



0,7
p
Probabilité de succès : $p = 0{,}45$
Taille d'un échantillon : $n = 1\,00$
Nombre d'intervalles de confiance : 100
Pourcentage d'intervalles ne contenant pas p : 5%
10
20
30
40
50
60
70
80
90
100

Exemple

Avant le second tour d'une élection présidentielle, on effectue deux sondages sur des échantillons représentatifs de la population. On ne retient que les suffrages exprimés, c'est-à-dire les personnes qui ont l'intention de voter pour l'un ou l'autre des candidats.

- *le premier sondage concerne $1\,003$ personnes qui votent à* $53{,}1\%$ *pour le candidat A.*
- *le seconde sondage concerne $9\,851$ personnes qui votent à* $51{,}2\%$ *pour le candidat A.*

Lequel de ces sondages permet d'affirmer, au seuil de confiance de 95%, que le candidat A va l'emporter ?