Cours 2: Intervalles de confiance d'une moyenne

Clément Rau
Laboratoire de Mathématiques de Toulouse
Université Paul Sabatier-IUT GEA Ponsan

Module: Prise en compte de l'aléa

1. Background théorique
   - Point de départ : la loi des grands nombres
   - Théorème central limite
2. Estimation d'une moyenne
   - Vocabulaire
   - Estimateurs
     - Comment estimer une moyenne
     - Comment estimer une variance (ou un écart-type) ?
   - Intervalle de confiance (asymptotique) d'une moyenne
     - Exemple d'obtention d'IC d'une espérance μ, lorsque σ connu.
     - Tableau récapitulatif
     - Exemples
   - Intervalle de fluctuation (asymptotique) d'une moyenne empirique (échantillonnage)

1. Background théorique
   - Point de départ : la loi des grands nombres
   - Théorème central limite
2. Estimation d'une moyenne
   - Vocabulaire
   - Estimateurs
     - Comment estimer une moyenne
     - Comment estimer une variance (ou un écart-type) ?
   - Intervalle de confiance (asymptotique) d'une moyenne
     - Exemple d'obtention d'IC d'une espérance μ, lorsque σ connu.
     - Tableau récapitulatif
     - Exemples
   - Intervalle de fluctuation (asymptotique) d'une moyenne empirique (échantillonnage)

Un point de départ

Supposons que l'on lance un « grand » nombre de fois une pièce (équilibrée) en l'air, il y aura en moyenne 50% de piles (et donc aussi 50% de face).

Un point de départ

Supposons que l'on lance un « grand » nombre de fois une pièce (équilibrée) en l'air, il y aura en moyenne 50% de piles (et donc aussi 50% de face).

Précisons les choses...

Un point de départ

On joue n fois au pile ou face, avec proba p de tomber sur pile.

Un point de départ

On joue n fois au pile ou face, avec proba p de tomber sur pile.
Pour $1 \leq i \leq n$ on pose

$$X_i = 1_{\{pile\}} = \begin{cases} 1 & \text{si on obtient pile,} \\ 0 & \text{sinon} \end{cases}$$

Un point de départ

On joue n fois au pile ou face, avec proba p de tomber sur pile.
Pour $1 \leq i \leq n$ on pose

$$X_i = 1_{\{pile\}} = \begin{cases} 1 & \text{si on obtient pile,} \\ 0 & \text{sinon} \end{cases}$$

Alors :

$$\frac{\sum_{i=1..n} X_i}{n} = \frac{nb\ de\ piles}{n}.$$

Un point de départ

On joue n fois au pile ou face, avec proba p de tomber sur pile.
Pour $1 \leq i \leq n$ on pose

$$X_i = 1_{\{pile\}} = \begin{cases} 1 & \text{si on obtient pile,} \\ 0 & \text{sinon} \end{cases}$$

Alors :

$$\frac{\sum_{i=1..n} X_i}{n} = \frac{nb\ de\ piles}{n}.$$

Et il semble assez naturel que lorsque n est grand le rapport $nb\ de\ piles/n$ tend vers la proba de tomber sur pile,

Un point de départ

On joue n fois au pile ou face, avec proba p de tomber sur pile.
Pour $1 \leq i \leq n$ on pose

$$X_i = 1_{\{pile\}} = \begin{cases} 1 & \text{si on obtient pile,} \\ 0 & \text{sinon} \end{cases}$$

Alors :

$$\frac{\sum_{i=1..n} X_i}{n} = \frac{nb\ de\ piles}{n}.$$

Et il semble assez naturel que lorsque n est grand le rapport $nb\ de\ piles/n$ tend vers la proba de tomber sur pile, c'est à dire précisément $p = \mathbb{E}(X_1)$.

Un point de départ

On joue n fois au pile ou face, avec proba p de tomber sur pile.
Pour $1 \leq i \leq n$ on pose

$$X_i = 1_{\{pile\}} = \begin{cases} 1 & \text{si on obtient pile,} \\ 0 & \text{sinon} \end{cases}$$

Alors :

$$\frac{\sum_{i=1..n} X_i}{n} = \frac{nb\ de\ piles}{n}.$$

Et il semble assez naturel que lorsque n est grand le rapport $nb\ de\ piles/n$ tend vers la proba de tomber sur pile, c'est à dire précisément $p = \mathbb{E}(X_1)$. Ainsi dans ce cas particulier, il semble que lorsque n grand,

$$\frac{\sum_{i=1..n} X_i}{n} \to \mathbb{E}(X_1).$$

Un point de départ

Même constatation, en lancant un dé et en regardant la proportion du nombre d'apparitions d'une face.

Loi faible des grands nombres

A l'aide d'outils simples (inégalité de Bienaymé Tcebychev), on pourrait montrer (...) que :

Loi faible des grands nombres

A l'aide d'outils simples (inégalité de Bienaymé Tcebychev), on pourrait montrer (...) que :

Théorème

Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de v.a. réelles deux à deux indépendantes et de même loi tel que $\mathbb{E}(X_1^2) < \infty$. Alors,

$$\forall\epsilon > 0 \quad \lim_n \mathbb{P}(\ |\frac{1}{n}\sum_{i=1..n} X_i - \mathbb{E}(X_1)| > \epsilon\ ) = 0$$

Ce type de convergence s'appelle la convergence en probabilité.

Loi faible des grands nombres

A l'aide d'outils simples (inégalité de Bienaymé Tcebychev), on pourrait montrer (...) que :

Théorème

Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de v.a. réelles deux à deux indépendantes et de même loi tel que $\mathbb{E}(X_1^2) < \infty$. Alors,

$$\forall \epsilon > 0 \quad \lim_n \mathbb{P}(\, |\frac{1}{n} \sum_{i=1..n} X_i - \mathbb{E}(X_1)| > \epsilon \,) = 0$$

Ce type de convergence s'appelle la convergence en probabilité.
Autrement dit, la moyenne arithmétique de $X_1, ..., X_n$ converge en probabilité vers l'espérance de X_1.

Loi forte des grands nombres

Théorème

Soit $(X_n)_{n\in\mathbb{N}^\star}$ *une suite de v.a. réelles deux à deux indépendantes et de même loi tel que* $\mathbb{E}(|X_1|) < \infty$. *Alors,*

$$\textit{pour presque tout } \omega, \quad \lim_n \frac{1}{n} \sum_{i=1..n} X_i = \mathbb{E}(X_1).$$

Loi forte des grands nombres

Théorème

Soit $(X_n)_{n\in\mathbb{N}^\star}$ *une suite de v.a. réelles deux à deux indépendantes et de même loi tel que* $\mathbb{E}(|X_1|) < \infty$. *Alors,*

$$\textit{pour presque tout } \omega, \quad \lim_n \frac{1}{n} \sum_{i=1..n} X_i = \mathbb{E}(X_1).$$

On parle de convergence presque sûre (p.s en abrégé). Cela signifie que pour presque chaque réalisation ω, la quantité

Moyenne arithmétique des X_i converge vers $\mathbb{E}(X_1)$.

Loi forte des grands nombres

Théorème

Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de v.a. réelles deux à deux indépendantes et de même loi tel que $\mathbb{E}(|X_1|) < \infty$. Alors,

$$\text{pour presque tout } \omega, \quad \lim_n \frac{1}{n}\sum_{i=1..n} X_i = \mathbb{E}(X_1).$$

On parle de convergence presque sûre (p.s en abrégé). Cela signifie que pour presque chaque réalisation ω, la quantité

$\boxed{\textit{Moyenne arithmétique} \text{ des } X_i \text{ converge vers } \mathbb{E}(X_1).}$

Remarques :

-Attention, la "vitesse" de convergence dépend du ω.

-On admet ce Théorème (LGN) fondamental dont les preuves sont beaucoup plus complexes que celles de sa version faible.

Quelques idées d'applications immédiates de LGN

- Estimer la proba de tomber sur face d'une pièce. (Vérifier si la pièce est équilibrée)

Quelques idées d'applications immédiates de LGN

- Estimer la proba de tomber sur face d'une pièce. (Vérifier si la pièce est équilibrée)
- Estimer la proportion d'électeurs votant pour un candidat.

Quelques idées d'applications immédiates de LGN

- Estimer la proba de tomber sur face d'une pièce. (Vérifier si la pièce est équilibrée)
- Estimer la proportion d'électeurs votant pour un candidat.

Rapidement, on tombe sur le probléme suivant :combien de lancers de pièces (ou combien de personnes doit on interroger), pour que le résultat que l'on avance soit correct avec grosse probabilité...

Quelques idées d'applications immédiates de LGN

- Estimer la proba de tomber sur face d'une pièce. (Vérifier si la pièce est équilibrée)
- Estimer la proportion d'électeurs votant pour un candidat.

Rapidement, on tombe sur le probléme suivant :combien de lancers de pièces (ou combien de personnes doit on interroger), pour que le résultat que l'on avance soit correct avec grosse probabilité...

La section suivante permet de répondre à ce genre de questions.

TCL

Posons

$$\bar{X}_n = \frac{\sum_i X_i}{n}.$$

TCL

Posons

$$\bar{X}_n = \frac{\sum_i X_i}{n}.$$

- Sous certaines conditions, on sait donc que $\bar{X}_n - \mathbb{E}(X_1)$ tend vers 0. Cette quantité représente les écarts, les fluctuations entre la moyenne que l'on observe $\bar{X}_n$ et la "vraie" moyenne $\mathbb{E}(X_1)$.

TCL

Posons

$$\bar{X}_n = \frac{\sum_i X_i}{n}.$$

- Sous certaines conditions, on sait donc que $\bar{X}_n - \mathbb{E}(X_1)$ tend vers 0. Cette quantité représente les écarts, les fluctuations entre la moyenne que l'on observe $\bar{X}_n$ et la "vraie" moyenne $\mathbb{E}(X_1)$.
- On aimerait aller à l'ordre supérieur et connaitre "la vitesse" de convergence vers 0.

TCL

Posons

$$\bar{X}_n = \frac{\sum_i X_i}{n}.$$

- Sous certaines conditions, on sait donc que $\bar{X}_n - \mathbb{E}(X_1)$ tend vers 0. Cette quantité représente les écarts, les fluctuations entre la moyenne que l'on observe $\bar{X}_n$ et la "vraie" moyenne $\mathbb{E}(X_1)$.
- On aimerait aller à l'ordre supérieur et connaitre "la vitesse" de convergence vers 0.
- Le Theoreme central limite (TCL) indique "grosso modo" que si l'on multiplie les écarts $\bar{X}_n - \mathbb{E}(X_1)$ par $\sqrt{n}$, on obtient une v.a dont la loi s'approche d'une Normale (lorsque n est grand).

TCL

Posons

$$\bar{X}_n = \frac{\sum_i X_i}{n}.$$

- Sous certaines conditions, on sait donc que $\bar{X}_n - \mathbb{E}(X_1)$ tend vers 0. Cette quantité représente les écarts, les fluctuations entre la moyenne que l'on observe $\bar{X}_n$ et la "vraie" moyenne $\mathbb{E}(X_1)$.
- On aimerait aller à l'ordre supérieur et connaitre "la vitesse" de convergence vers 0.
- Le Theoreme central limite (TCL) indique "grosso modo" que si l'on multiplie les écarts $\bar{X}_n - \mathbb{E}(X_1)$ par $\sqrt{n}$, on obtient une v.a dont la loi s'approche d'une Normale (lorsque n est grand).

Plus précisément,

Théorème (TCL)

Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles indépendantes et de même loi, de moyenne m et d'écart type σ. Notons

$$\bar{X}_n = \frac{X_1 + ... + X_n}{n}$$

et Z_n les v.a. associées centrées réduites :

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}.$$

Alors pour tout intervalle $[a; b]$, on a :

$$\lim_n \mathbb{P}(Z_n \in [a; b]) = \mathbb{P}(Y \in [a; b]) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt,$$

où Y suit une $\mathcal{N}(0, 1)$.

- On dit que la loi de la v.a. $Z_n = \frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$ converge en loi vers une Normale centreé réduite $\mathcal{N}(0, 1)$.

- On dit que la loi de la v.a. $Z_n = \frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$ converge en loi vers une Normale centreé réduite $\mathcal{N}(0,1)$.
- Vous pouvez retenir le Théorème sous une forme plus "digeste" ainsi :

$$\boxed{\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1),}$$

pour des X_i indépendants et identiquement distribués.

Remarques sur le TCL

- Caractère universel : Quelque soit la loi des X_i (moment d'ordre 1 fini), les sommes renormalisées convergent vers une meme loi limite, la loi Normale, ce qui explique le nom de cette loi.

Remarques sur le TCL

- Caractère universel : Quelque soit la loi des X_i (moment d'ordre 1 fini), les sommes renormalisées convergent vers une meme loi limite, la loi Normale, ce qui explique le nom de cette loi.
- Importance capitale de ce Théorème en statistiques.

Une première application immédiate

- Connaissant la densité de la loi Normale, le TCL nous dit que : si n est assez grand alors Z_n est très probablement compris entre -3 et 3 (la probabilité est 0.9973). Soit encore :

$$\frac{X_1 + ... + X_n}{n} - \mathbb{E}(X_1) \in [-\frac{3\sigma}{\sqrt{n}}; \frac{3\sigma}{\sqrt{n}}],$$

avec grosse probabilité.

Une seconde application

En pratique, lorsque l'on considère un grand nombre de v.a. indépendantes et de même loi $X_1, ..., X_n$, on approxime leur somme S_n ou leur moyenne $\bar{X}_n$ par des variables Normales suivantes :

$$\boxed{S_n \sim \mathcal{N}(nm, \sigma\sqrt{n}) \;\; et \;\; \bar{X}_n \sim \mathcal{N}(m, \sigma/\sqrt{n}),}$$

où $m = \mathbb{E}(X_1)$ et $\sigma^2 = var(X_1)$.

Une seconde application

En pratique, lorsque l'on considère un grand nombre de v.a. indépendantes et de même loi $X_1, ..., X_n$, on approxime leur somme S_n ou leur moyenne $\bar{X}_n$ par des variables Normales suivantes :

$$\boxed{S_n \sim \mathcal{N}(nm, \sigma\sqrt{n}) \;\; et \;\; \bar{X}_n \sim \mathcal{N}(m, \sigma/\sqrt{n}),}$$

où $m = \mathbb{E}(X_1)$ et $\sigma^2 = var(X_1)$.

- Si l'on prend $X_i \sim Bernoulli(p)$, on retrouve qu'une Binomiale s'approche par une Normale.

Une seconde application

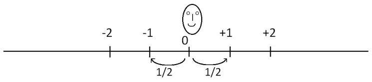
En pratique, lorsque l'on considère un grand nombre de v.a. indépendantes et de même loi $X_1, ..., X_n$, on approxime leur somme S_n ou leur moyenne $\bar{X}_n$ par des variables Normales suivantes :

$$\boxed{S_n \sim \mathcal{N}(nm, \sigma\sqrt{n}) \;\; et \;\; \bar{X}_n \sim \mathcal{N}(m, \sigma/\sqrt{n}),}$$

où $m = \mathbb{E}(X_1)$ et $\sigma^2 = var(X_1)$.

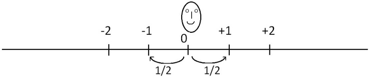
- Si l'on prend $X_i \sim Bernoulli(p)$, on retrouve qu'une Binomiale s'approche par une Normale.
- On a donc deux approximations possibles pour une $Binomiale(n, p)$:
  - celle par une loi de Poisson $\mathcal{P}(np)$ lorsque n est grand, p petit,
  - et celle par $\mathcal{N}(np, \sqrt{np(1-p)})$ lorsque n est grand (p quelconque).

D'autres applications originales...pour vous motiver ?



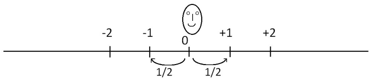
D'autres applications originales...pour vous motiver ?

- La trajectoire de l'ivrogne



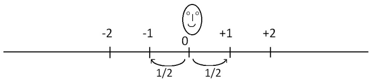
D'autres applications originales...pour vous motiver ?

- La trajectoire de l'ivrogne Soit un marcheur aléatoire (imaginez un bonhomme ivrogne) qui se déplace sur l'axe $\mathbb{Z}$ en sautant aléatoirement à chaque unité de temps (à chaque seconde par exemple) sur un de ces 2 voisins (droite ou gauche).



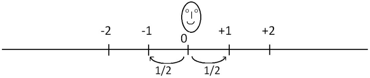
D'autres applications originales...pour vous motiver ?

- La trajectoire de l'ivrogne Soit un marcheur aléatoire (imaginez un bonhomme ivrogne) qui se déplace sur l'axe $\mathbb{Z}$ en sautant aléatoirement à chaque unité de temps (à chaque seconde par exemple) sur un de ces 2 voisins (droite ou gauche).



D'autres applications originales...pour vous motiver ?

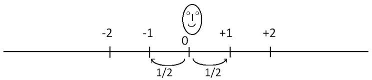
- La trajectoire de l'ivrogne Soit un marcheur aléatoire (imaginez un bonhomme ivrogne) qui se déplace sur l'axe $\mathbb{Z}$ en sautant aléatoirement à chaque unité de temps (à chaque seconde par exemple) sur un de ces 2 voisins (droite ou gauche).



Notons X_i sa position à l'instant i.

D'autres applications originales...pour vous motiver ?

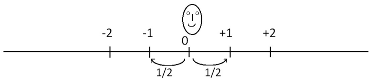
- La trajectoire de l'ivrogne Soit un marcheur aléatoire (imaginez un bonhomme ivrogne) qui se déplace sur l'axe $\mathbb{Z}$ en sautant aléatoirement à chaque unité de temps (à chaque seconde par exemple) sur un de ces 2 voisins (droite ou gauche).



Notons X_i sa position à l'instant i. On suppose qu'à $t = 0$, on a $X_0 = 0$.

D'autres applications originales...pour vous motiver ?

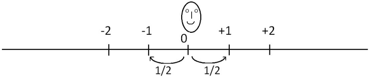
- La trajectoire de l'ivrogne Soit un marcheur aléatoire (imaginez un bonhomme ivrogne) qui se déplace sur l'axe $\mathbb{Z}$ en sautant aléatoirement à chaque unité de temps (à chaque seconde par exemple) sur un de ces 2 voisins (droite ou gauche).



Notons X_i sa position à l'instant i. On suppose qu'à $t = 0$, on a $X_0 = 0$. Le TCL nous permet de dire qu'avec grosse probabilité, le marcheur se trouvera dans la boule de centre 0 et de rayon $\sqrt{n}$, au bout d'un temps n.

D'autres applications originales...pour vous motiver ?

- La trajectoire de l'ivrogne Soit un marcheur aléatoire (imaginez un bonhomme ivrogne) qui se déplace sur l'axe $\mathbb{Z}$ en sautant aléatoirement à chaque unité de temps (à chaque seconde par exemple) sur un de ces 2 voisins (droite ou gauche).

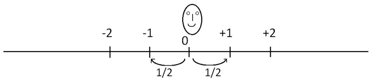


Notons X_i sa position à l'instant i. On suppose qu'à $t = 0$, on a $X_0 = 0$. Le TCL nous permet de dire qu'avec grosse probabilité, le marcheur se trouvera dans la boule de centre 0 et de rayon $\sqrt{n}$, au bout d'un temps n.

- Intervalle de confiance lors d'élection

D'autres applications originales...pour vous motiver ?

- La trajectoire de l'ivrogne Soit un marcheur aléatoire (imaginez un bonhomme ivrogne) qui se déplace sur l'axe $\mathbb{Z}$ en sautant aléatoirement à chaque unité de temps (à chaque seconde par exemple) sur un de ces 2 voisins (droite ou gauche).

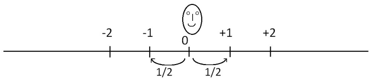


Notons X_i sa position à l'instant i. On suppose qu'à $t = 0$, on a $X_0 = 0$. Le TCL nous permet de dire qu'avec grosse probabilité, le marcheur se trouvera dans la boule de centre 0 et de rayon $\sqrt{n}$, au bout d'un temps n.

- Intervalle de confiance lors d'élection Deux candidats A et B sont en course pour une élection.

D'autres applications originales...pour vous motiver ?

- La trajectoire de l'ivrogne Soit un marcheur aléatoire (imaginez un bonhomme ivrogne) qui se déplace sur l'axe $\mathbb{Z}$ en sautant aléatoirement à chaque unité de temps (à chaque seconde par exemple) sur un de ces 2 voisins (droite ou gauche).



Notons X_i sa position à l'instant i. On suppose qu'à $t = 0$, on a $X_0 = 0$. Le TCL nous permet de dire qu'avec grosse probabilité, le marcheur se trouvera dans la boule de centre 0 et de rayon $\sqrt{n}$, au bout d'un temps n.

- Intervalle de confiance lors d'élection Deux candidats A et B sont en course pour une élection. Le TCL nous permet de donner un intervalle dans lequel la probabilité de gens votant pour A, doit se trouver à l'issue d'un sondage sur n personnes, avec une certaine proba. (cf exo en Td)

L'une des fonctions des statistiques est donc de proposer, à partir d'observations d'un phénomène aléatoire, une estimation d'un des paramètres du phénomène.

L'une des fonctions des statistiques est donc de proposer, à partir d'observations d'un phénomène aléatoire, une estimation d'un des paramètres du phénomène. But de la section suivante.

Introduction : les 3 grandes lignes

Les statistiques peuvent permettre :

Introduction : les 3 grandes lignes

Les statistiques peuvent permettre :

- d'estimer un paramètre inconnu,

Introduction : les 3 grandes lignes

Les statistiques peuvent permettre :

- d'estimer un paramètre inconnu,
- de donner une zone dans laquelle un paramètre, a de grande chance de se trouver

Introduction : les 3 grandes lignes

Les statistiques peuvent permettre :

- d'estimer un paramètre inconnu,
- de donner une zone dans laquelle un paramètre, a de grande chance de se trouver
- de prendre des décisions.

Introduction : les 3 grandes lignes

Les statistiques peuvent permettre :

- d'estimer un paramètre inconnu,
- de donner une zone dans laquelle un paramètre, a de grande chance de se trouver
- de prendre des décisions.

Chacune de ses questions correspond à une thématique en statistique.

Introduction : les 3 grandes lignes

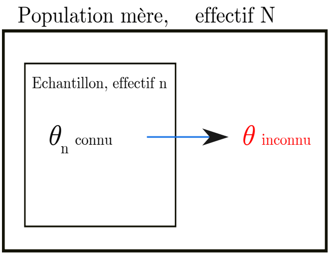
Les statistiques peuvent permettre :

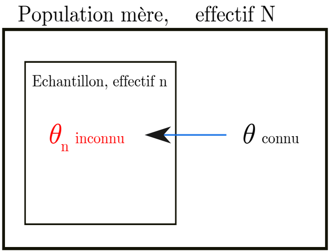
- d'estimer un paramètre inconnu,
- de donner une zone dans laquelle un paramètre, a de grande chance de se trouver
- de prendre des décisions.

Chacune de ses questions correspond à une thématique en statistique.

Dans ce module, on se cantonnera essentiellement sur les 2 premiers points.

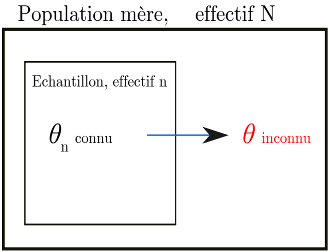
L'estimation au sens familier, se scinde en deux questions distinctes en Statistique :

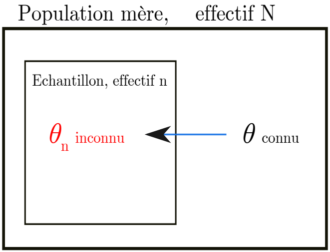




L'estimation au sens familier, se scinde en deux questions distinctes en Statistique :

- l'échantillonnage
- l'estimation





L'estimation au sens familier, se scinde en deux questions distinctes en Statistique :

- l'échantillonnage
- l'estimation

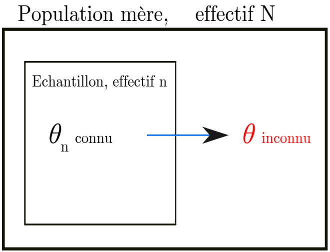


Figure – Principe de l'estimation.

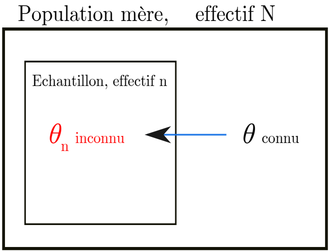
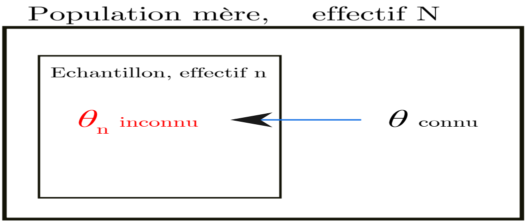


Figure – Principe de l'échantillonnage.

Echantillonage

L'échantillonnage permet de passer (de la loi connue d'un paramètre θ dans une population de taille N) à une estimée d'une quantité θ_n fabriquée à partir seulement d'une population de taille n plus petite (échantillon).



Echantillonage

L'échantillonnage permet de passer (de la loi connue d'un paramètre θ dans une population de taille N) à une estimée d'une quantité θ_n fabriquée à partir seulement d'une population de taille n plus petite (échantillon).

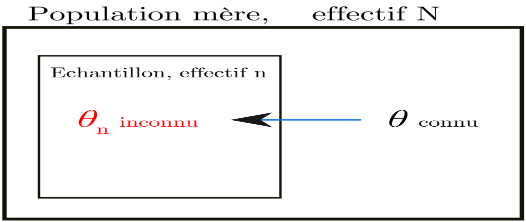


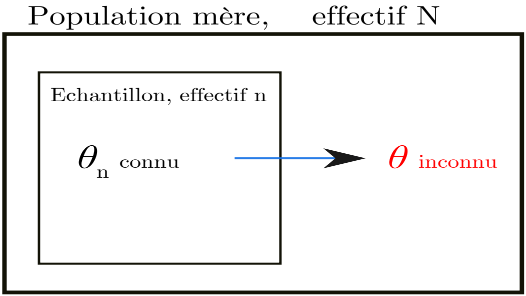
Figure – Principe de l'échantillonnage.

Exemple : échantillonnage

Dans une entreprise qui comptent 659 employés, on sait que $0,03\%$ *des employés sont mécontents. On pioche un échantillon de 15 employés. Quel est l'ordre de grandeur des employés mécontents dans cet échantillon ?*

Estimation statistique

L'estimation statistique permet d'induire, à partir des résulats observés sur un échantillon, des informations sur la population totale.



Estimation statistique

L'estimation statistique permet d'induire, à partir des résulats observés sur un échantillon, des informations sur la population totale.

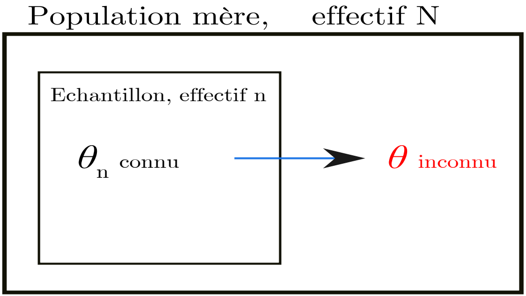


Figure – Principe de l'estimation.

Exemple : estimation

Dans un échantillon de 15 employés d'une entreprise, 7% s'estiment sous pression. Quel est l'ordre de grandeur des employés sous pression parmi tout le personnel de l'entreprise ?

Formulations possibles d'une "estimation"

En Statistique, on distingue essentiellement deux formes de présentation d'une estimation :

- Une estimation ponctuelle par une quantité que l'on appelle estimateur, qui doit être "proche" du paramètre cherché.

Formulations possibles d'une "estimation"

En Statistique, on distingue essentiellement deux formes de présentation d'une estimation :

- Une estimation ponctuelle par une quantité que l'on appelle estimateur, qui doit être "proche" du paramètre cherché.
- Une estimation par intervalle de confiance (IC). C'est une zone dans laquelle, on peut affirmer que le paramètre cherché se trouve avec une certaine probabilité.

Décor, notations valables pour toute la suite du cours

Sur un échantillon de taille n, on s'intéresse aux valeurs d'un paramètre X_i, (pour $1 \leq i \leq n$,), que prennent les n individus de l'échantillon.

Les X_i sont des v.a supposées :

- iid,
- d'espérance μ (connue ou pas)
- et d'écart-type σ (connus ou pas).

La moyenne empirique $\bar{X}_n$

On pose naturellement :

La moyenne empirique $\bar{X}_n$

On pose naturellement :

$$\bar{X}_n = \frac{\sum_{i=1\dots n} X_i}{n}.$$

Moyenne empirique des X_i

La moyenne empirique $\bar{X}_n$

On pose naturellement :

$$\boxed{\bar{X}_n = \frac{\sum_{i=1\ldots n} X_i}{n}.}$$

Moyenne empirique des X_i

Variance empirique

Au regard de la définition de la variance et de la loi des grands nombres, il est naturel d'introduire :

$$\boxed{S_n^2 = \frac{1}{n}\sum_{i=1\ldots n}(X_i - \bar{X}_n)^2 = \frac{1}{n}\Big[\sum_{i=1\ldots n} X_i^2\Big] - \bar{X}_n^{\,2}.}$$

Variance empirique

Au regard de la définition de la variance et de la loi des grands nombres, il est naturel d'introduire :

$$S_n^2 = \frac{1}{n} \sum_{i=1 \ldots n} (X_i - \bar{X}_n)^2 = \frac{1}{n} [\sum_{i=1 \ldots n} X_i^2] - \bar{X_n}^2.$$

Si l'on connaît la "vraie" valeur de l'espérance μ des X_i, on peut penser à introduire cet estimateur :

$$\hat{S_n}^2 = \frac{1}{n} \sum_{i=1 \ldots n} (X_i - \mu)^2 = \frac{1}{n} [\sum_{i=1 \ldots n} X_i^2] - \mu^2$$

Remarque sur S_n^2, biais

On peut montrer que :

$$\boxed{\mathbb{E}(\hat{S}_n{}^2) = \sigma^2 \quad \text{et} \quad \mathbb{E}(S_n^2) = \frac{n-1}{n}\sigma^2}$$ (petit calcul).

Remarque sur S_n^2, biais

On peut montrer que :

$$\boxed{\mathbb{E}(\hat{S}_n{}^2) = \sigma^2 \quad \text{et} \quad \mathbb{E}(S_n^2) = \frac{n-1}{n}\sigma^2}$$ (petit calcul).

On dit que S_n^2 a un biais, car $\mathbb{E}(S_n^2) \neq \sigma^2$. Parcontre $\hat{S}_n{}^2$ n' a pas de biais. Ainsi le "bon estimateur" sans biais de la variance (lorsque μ est inconnu) est :

$$\boxed{\frac{n}{n-1} S_n^2 = \frac{1}{n-1} \sum_{i=1\ldots n} (X_i - \bar{X}_n)^2.}$$

Remarque sur S_n^2, biais

On peut montrer que :

$$\boxed{\mathbb{E}(\hat{S_n}^2) = \sigma^2 \quad \text{et} \quad \mathbb{E}(S_n^2) = \frac{n-1}{n}\sigma^2} \text{ (petit calcul).}$$

On dit que S_n^2 a un biais, car $\mathbb{E}(S_n^2) \neq \sigma^2$. Parcontre $\hat{S_n}^2$ n' a pas de biais. Ainsi le "bon estimateur" sans biais de la variance (lorsque μ est inconnu) est :

$$\boxed{\frac{n}{n-1} S_n^2 = \frac{1}{n-1} \sum_{i=1\dots n} (X_i - \bar{X}_n)^2.}$$

Remarque : C'est d'ailleurs cette "variance" que renvoie certains logiciel (comme R) lorsqu'on tape `var`.

Trame de l'obtention d'estimation

- Pour estimer une quantité, à l'aide des théorémes limites (type Loi des grands nombres), nous venons de fabriquer une fonction des observations, (une statistique/une v.a) qui tend vers la quantité souhaitée . On préférera une statistique sans biais.

Trame de l'obtention d'estimation

- Pour estimer une quantité, à l'aide des théorémes limites (type Loi des grands nombres), nous venons de fabriquer une fonction des observations, (une statistique/une v.a) qui tend vers la quantité souhaitée . On préférera une statistique sans biais.
- Par des Théorème limite (type TCL), il est possible de connaître la loi limite de ces statistiques.

Trame de l'obtention d'estimation

- Pour estimer une quantité, à l'aide des théorémes limites (type Loi des grands nombres), nous venons de fabriquer une fonction des observations, (une statistique/une v.a) qui tend vers la quantité souhaitée . On préférera une statistique sans biais.
- Par des Théorème limite (type TCL), il est possible de connaître la loi limite de ces statistiques.
  $\Rightarrow$ Nous allons donc être capable de fournir une zone dans laquelle notre paramètre a de grande chance de se trouver.

Trame de l'obtention d'estimation

- Pour estimer une quantité, à l'aide des théorémes limites (type Loi des grands nombres), nous venons de fabriquer une fonction des observations, (une statistique/une v.a) qui tend vers la quantité souhaitée . On préférera une statistique sans biais.
- Par des Théorème limite (type TCL), il est possible de connaître la loi limite de ces statistiques.
  $\Rightarrow$ Nous allons donc être capable de fournir une zone dans laquelle notre paramètre a de grande chance de se trouver.

Notion d'intervalle de confiance (asymptotique).

Trame de l'obtention d'estimation

- Pour estimer une quantité, à l'aide des théorémes limites (type Loi des grands nombres), nous venons de fabriquer une fonction des observations, (une statistique/une v.a) qui tend vers la quantité souhaitée . On préférera une statistique sans biais.
- Par des Théorème limite (type TCL), il est possible de connaître la loi limite de ces statistiques.
  $\Rightarrow$ Nous allons donc être capable de fournir une zone dans laquelle notre paramètre a de grande chance de se trouver.

Notion d'intervalle de confiance (asymptotique).

Nous allons détailler la mécanique pour obtenir un IC dans un cas simple et nous admettrons les autres cas (preuves similaires...)

Trame de l'obtention d'estimation

- Pour estimer une quantité, à l'aide des théorémes limites (type Loi des grands nombres), nous venons de fabriquer une fonction des observations, (une statistique/une v.a) qui tend vers la quantité souhaitée . On préférera une statistique sans biais.
- Par des Théorème limite (type TCL), il est possible de connaître la loi limite de ces statistiques.
  $\Rightarrow$ Nous allons donc être capable de fournir une zone dans laquelle notre paramètre a de grande chance de se trouver.

Notion d'intervalle de confiance (asymptotique).

Nous allons détailler la mécanique pour obtenir un IC dans un cas simple et nous admettrons les autres cas (preuves similaires...)

Note : Nous developperons uniquement des IC bilatéraux dans ce cours. Voir les exos en TD pour des IC unilatéraux

Estimation d'une moyenne, dans le cas où la variance σ^2 est connue

- On estime la moyenne μ par $\bar{X}_n$.

Estimation d'une moyenne, dans le cas où la variance σ^2 est connue

- On estime la moyenne μ par $\bar{X}_n$.
- Par le TCL, on sait que $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0;1)$.

Estimation d'une moyenne, dans le cas où la variance σ^2 est connue

- On estime la moyenne μ par $\bar{X}_n$.
- Par le TCL, on sait que $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0;1)$.
- Etant donné une marge d'erreur α, (par ex 5%), on détermine alors un certain u à l'aide de la table de la $\mathcal{N}(0;1)$, tel que

$$\mathbb{P}(|Z| \leq u) \geq 1 - \alpha, \quad \text{où } Z \sim \mathcal{N}(0;1).$$

Estimation d'une moyenne, dans le cas où la variance σ^2 est connue

- On estime la moyenne μ par $\bar{X}_n$.
- Par le TCL, on sait que $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1)$.
- Etant donné une marge d'erreur α, (par ex 5%), on détermine alors un certain u à l'aide de la table de la $\mathcal{N}(0; 1)$, tel que

$$\mathbb{P}(|Z| \leq u) \geq 1 - \alpha, \quad \text{où } Z \sim \mathcal{N}(0; 1).$$

Remarque : Ce u dépend de α et sera noté $u_{1-\frac{\alpha}{2}}$, (cf explication ci après).

Estimation d'une moyenne, dans le cas où la variance σ^2 est connue

- On estime la moyenne μ par $\bar{X}_n$.
- Par le TCL, on sait que $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0;1)$.
- Etant donné une marge d'erreur α, (par ex 5%), on détermine alors un certain u à l'aide de la table de la $\mathcal{N}(0;1)$, tel que

$$\mathbb{P}(|Z| \leq u) \geq 1 - \alpha, \quad \text{où } Z \sim \mathcal{N}(0;1).$$

Remarque : Ce u dépend de α et sera noté $u_{1-\frac{\alpha}{2}}$, (cf explication ci après).

- Ainsi avec proba $\geq 1 - \alpha$, on a :$|\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)| \leq u_{1-\frac{\alpha}{2}}$.

Estimation d'une moyenne, dans le cas où la variance σ^2 est connue

- On estime la moyenne μ par $\bar{X}_n$.
- Par le TCL, on sait que $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0;1)$.
- Etant donné une marge d'erreur α, (par ex 5%), on détermine alors un certain u à l'aide de la table de la $\mathcal{N}(0;1)$, tel que

$$\mathbb{P}(|Z| \leq u) \geq 1 - \alpha, \quad \text{où } Z \sim \mathcal{N}(0;1).$$

<u>Remarque :</u> Ce u dépend de α et sera noté $u_{1-\frac{\alpha}{2}}$, (cf explication ci après).

- Ainsi avec proba $\geq 1 - \alpha$, on a :$|\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)| \leq u_{1-\frac{\alpha}{2}}$.

Estimation d'une moyenne, dans le cas où la variance σ^2 est connue

- On estime la moyenne μ par $\bar{X}_n$.
- Par le TCL, on sait que $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0;1)$.
- Etant donné une marge d'erreur α, (par ex 5%), on détermine alors un certain u à l'aide de la table de la $\mathcal{N}(0;1)$, tel que

$$\mathbb{P}(|Z| \leq u) \geq 1 - \alpha, \quad \text{où } \; Z \sim \mathcal{N}(0;1).$$

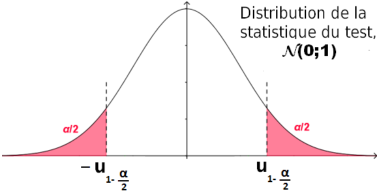
  Remarque : Ce u dépend de α et sera noté $u_{1-\frac{\alpha}{2}}$, (cf explication ci après).

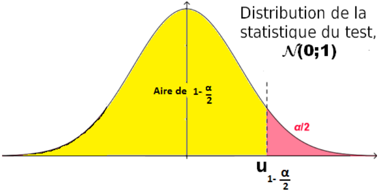
- Ainsi avec proba $\geq 1-\alpha$, on a :$|\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)| \leq u_{1-\frac{\alpha}{2}}$.

$$ie : \quad \boxed{\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}$$

Explication de la notation $u_{1-\frac{\alpha}{2}}$

Cette notation provient de "l'inverse de la fonction de répartition" (quantile).





L'aire avant le seuil $u_{1-\frac{\alpha}{2}}$ vaut $1-\frac{\alpha}{2}$

Remarque : Si vous trouvez la notation trop lourde, vous pouvez simplement retenir comment trouver ce seuil et le noter "juste" u.

IC de l'espérance μ, lorsque σ connu.

On retiendra donc qu'avec proba $\geq 1 - \alpha$, on a :

IC de l'espérance μ, lorsque σ connu.

On retiendra donc qu'avec proba $\geq 1 - \alpha$, on a :

ie : $$\boxed{\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}$$

IC de l'espérance μ, lorsque σ connu.

On retiendra donc qu'avec proba $\geq 1-\alpha$, on a :

$$ie : \quad \boxed{\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}$$

Avec les mêmes types d'idées, on peut trouver un IC pour une espérance lorsque σ est inconnu, ou un IC d'une proportion. On résume cela dans le tableau suivant à connaître.

Récapitulatif IC espérance/proportion

Paramètre à estimer	**IC au seuil α**		
Espérance μ	Si variance σ^2 connue $[\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$	Si variance σ^2 inconnue $[\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n-1}} ; \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n-1}}]$	
Condition *Lecture seuil*	$n \geq 30$ $\mathcal{N}(0,1)$	X_i Gaussien et n qqc $Student(n-1)$	$n \geq 30$ $\mathcal{N}(0,1)$
Proportion p	$[\bar{X}_n - u_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} ; \bar{X}_n + u_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}]$		
Condition *Lecture seuil*	$X_i \sim Bernoulli(p)$, $n \geq 30$ et $\min(n\bar{X}_n \,;\, n(1-\bar{X}_n)) \geq 5$ $\mathcal{N}(0,1)$		

Récapitulatif IC espérance/proportion

Paramètre à estimer	IC au seuil α		
Espérance μ	Si variance σ^2 connue $[\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} ; \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$	Si variance σ^2 inconnue $[\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n-1}} ; \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n-1}}]$	
Condition *Lecture seuil*	$n \geq 30$ $\mathcal{N}(0,1)$	X_i Gaussien et n qqc $Student(n-1)$	$n \geq 30$ $\mathcal{N}(0,1)$
Proportion p	$[\bar{X}_n - u_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} \; ; \; \bar{X}_n + u_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}]$		
Condition *Lecture seuil*	$X_i \sim Bernoulli(p)$, $n \geq 30$ et $\min(n\bar{X}_n \; ; \; n(1-\bar{X}_n)) \geq 5$ $\mathcal{N}(0,1)$		

Note : *La lecture du seuil correspond au choix de la table pour lire la valeur du u.*

Remarque

1. *Si* $n \geq 30$*, la loi de Student(n) est proche de la loi Normale.*

Remarque

1. *Si $n \geq 30$, la loi de Student(n) est proche de la loi Normale.*
2. *Dans le cas d'une proportion, vous pouvez aussi trouver dans la littérature l'IC suivant* $[\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{1}{2\sqrt{n}} \; ; \; \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{1}{2\sqrt{n}}]$ *qui est un peu moins précis, mais plus utile lorsqu'on cherche la taille d'un échantillon pour obtenir une amplitude donnée de l'IC.*

Exemple 1

Une machine produit en grande série des objets de masse théorique 180g. On admet que la variable aléatoire qui associe à un objet sa masse a pour ecart-type 0,92g. On préléve un échantillon de 100 objets et on mesure la masse de chacun, on obtient une moyenne de 179,93g.
Déterminer un intervalle de confiance au seuil de risque de 1%, de la masse μ d'un objet.

Exemple 1

On cherche à estimer une espérance dans le cas où la variance est connue.

Exemple 1

On cherche à estimer une espérance dans le cas où la variance est connue.

- Soit X_i, la v.a qui renvoit la masse de l'objet i de l'échantillon. On cherche un intervalle de confiance de $\mu = \mathbb{E}(X_i)$

Exemple 1

On cherche à estimer une espérance dans le cas où la variance est connue.

- Soit X_i, la v.a qui renvoit la masse de l'objet i de l'échantillon. On cherche un intervalle de confiance de $\mu = \mathbb{E}(X_i)$
- On sait qu'avec proba $1 - \alpha$,

$$\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Exemple 1

On cherche à estimer une espérance dans le cas où la variance est connue.

- Soit X_i, la v.a qui renvoit la masse de l'objet i de l'échantillon. On cherche un intervalle de confiance de $\mu = \mathbb{E}(X_i)$
- On sait qu'avec proba $1 - \alpha$,

$$\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

- $\alpha = 0,01$ donne un $u_{1-\frac{\alpha}{2}} = 2,58$. (table 2 de la $\mathcal{N}(0,1)$)

Exemple 1

On cherche à estimer une espérance dans le cas où la variance est connue.

- Soit X_i, la v.a qui renvoit la masse de l'objet i de l'échantillon. On cherche un intervalle de confiance de $\mu = \mathbb{E}(X_i)$
- On sait qu'avec proba $1-\alpha$,

$$\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

- $\alpha = 0,01$ donne un $u_{1-\frac{\alpha}{2}} = 2,58$. (table 2 de la $\mathcal{N}(0,1)$)
- D'où,

$$179,93 - 2,58 \times \frac{0,92}{\sqrt{100}} \leq \mu \leq 179,93 + 2,58 \times \frac{0,92}{\sqrt{100}}$$

ie : Avec proba $0,99$ on a, $\mu \in [179,69; 180,17]$.

Exemple 2

Le chiffre d'affaire mensuel d'une entreprise suit une loi Normale de moyenne μ et d'écart-type σ inconnus. Sur les 12 derniers mois, on a observé une moyenne des chiffres d'affaires égale à 10 000 euros avec un écart-type de 2000 euros.

Exemple 2

Le chiffre d'affaire mensuel d'une entreprise suit une loi Normale de moyenne μ et d'écart-type σ inconnus. Sur les 12 derniers mois, on a observé une moyenne des chiffres d'affaires égale à 10 000 euros avec un écart-type de 2000 euros.

Donner une estimation de μ par intervalle de confiance au niveau $0,98$.

Exemple 2

On cherche à estimer une espérance dans le cas où la variance est inconnue.

Exemple 2

On cherche à estimer une espérance dans le cas où la variance est inconnue.

- Soit X_i le chiffre d'affaire de l'entreprise le mois i,

Exemple 2

On cherche à estimer une espérance dans le cas où la variance est inconnue.

- Soit X_i le chiffre d'affaire de l'entreprise le mois i, suit une Normale.

Exemple 2

On cherche à estimer une espérance dans le cas où la variance est inconnue.

- Soit X_i le chiffre d'affaire de l'entreprise le mois i, suit une Normale. Nous n'avons donc pas de condition sur la taille n de l'échantillon.

Exemple 2

On cherche à estimer une espérance dans le cas où la variance est inconnue.

- Soit X_i le chiffre d'affaire de l'entreprise le mois i, suit une Normale. Nous n'avons donc pas de condition sur la taille n de l'échantillon.
- A l'aide de la table de la loi de Student, on trouve $t_{1-\frac{\alpha}{2}} = t_{0,99} \simeq 2,718$ tel que $\mathbb{P}(|T| \leq 2,718) \geq 0,98$

Exemple 2

On cherche à estimer une espérance dans le cas où la variance est inconnue.

- Soit X_i le chiffre d'affaire de l'entreprise le mois i, suit une Normale. Nous n'avons donc pas de condition sur la taille n de l'échantillon.
- A l'aide de la table de la loi de Student, on trouve $t_{1-\frac{\alpha}{2}} = t_{0,99} \simeq 2,718$ tel que $\mathbb{P}(|T| \leq 2,718) \geq 0,98$
- Donc avec proba $\geq 0,98$.

Exemple 2

On cherche à estimer une espérance dans le cas où la variance est inconnue.

- Soit X_i le chiffre d'affaire de l'entreprise le mois i, suit une Normale. Nous n'avons donc pas de condition sur la taille n de l'échantillon.
- A l'aide de la table de la loi de Student, on trouve $t_{1-\frac{\alpha}{2}} = t_{0,99} \simeq 2,718$ tel que $\mathbb{P}(|T| \leq 2,718) \geq 0,98$
- Donc avec proba $\geq 0,98$.

$$\mu \in [\bar{X}_{12} - 2,718 \times \frac{S_{12}}{\sqrt{11}}; \bar{X}_{12} + 2,718 \times \frac{S_{12}}{\sqrt{11}}].$$

Exemple 2

On cherche à estimer une espérance dans le cas où la variance est inconnue.

- Soit X_i le chiffre d'affaire de l'entreprise le mois i, suit une Normale. Nous n'avons donc pas de condition sur la taille n de l'échantillon.
- A l'aide de la table de la loi de Student, on trouve $t_{1-\frac{\alpha}{2}} = t_{0,99} \simeq 2,718$ tel que $\mathbb{P}(|T| \leq 2,718) \geq 0,98$
- Donc avec proba $\geq 0,98$.

$$\mu \in [\bar{X_{12}} - 2,718 \times \frac{S_{12}}{\sqrt{11}}; \bar{X_{12}} + 2,718 \times \frac{S_{12}}{\sqrt{11}}].$$

Avec $\bar{X_{12}} = 10000$ et $S_{12} = 2000$, on obtient

$$\mu \in [8360,9; 11639,02.], \quad \text{avec proba } 0,98.$$

Exemple 3

Lors d'une éléction, deux candidats A et B s'affrontent. Lors d'un sondage effectué sur $n = 500$ personnes, 387 personnes se déclarent voter pour A.

Exemple 3

Lors d'une éléction, deux candidats A et B s'affrontent. Lors d'un sondage effectué sur $n = 500$ personnes, 387 personnes se déclarent voter pour A.
Donner une estimation du pourcentage de votants pour A par intervalle de confiance au niveau $0,95$.

Exemple 3

On cherche à estimer la proportion p_A de votants pour A.

Exemple 3

On cherche à estimer la proportion p_A de votants pour A.

- Soit X_i la v.a égale à 1 si la personne i a voté pour A et 0 sinon.

Exemple 3

On cherche à estimer la proportion p_A de votants pour A.

- Soit X_i la v.a égale à 1 si la personne i a voté pour A et 0 sinon.
- Par hypothèse $\bar{X}_{500} = \frac{387}{500}$

Exemple 3

On cherche à estimer la proportion p_A de votants pour A.

- Soit X_i la v.a égale à 1 si la personne i a voté pour A et 0 sinon.
- Par hypothèse $\bar{X}_{500} = \frac{387}{500}$
- Les conditions $n = 500 \geq 30,\ n\bar{X}_{500} \geq 5\ et\ n(1 - \bar{X}_{500}) \geq 500$ sont vérifiées.

Exemple 3

On cherche à estimer la proportion p_A de votants pour A.

- Soit X_i la v.a égale à 1 si la personne i a voté pour A et 0 sinon.
- Par hypothèse $\bar{X}_{500} = \frac{387}{500}$
- Les conditions $n = 500 \geq 30,\ n\bar{X}_{500} \geq 5\ et\ n(1 - \bar{X}_{500}) \geq 500$ sont vérifiées.
- $\alpha = 0,05$ et on sait que pour une Normale, $u_{1-\frac{\alpha}{2}} = u_{0,975} \simeq 1,96$

Exemple 3

On cherche à estimer la proportion p_A de votants pour A.

- Soit X_i la v.a égale à 1 si la personne i a voté pour A et 0 sinon.
- Par hypothèse $\bar{X}_{500} = \frac{387}{500}$
- Les conditions $n = 500 \geq 30,\ n\bar{X}_{500} \geq 5\ et\ n(1 - \bar{X}_{500}) \geq 500$ sont vérifiées.
- $\alpha = 0,05$ et on sait que pour une Normale, $u_{1-\frac{\alpha}{2}} = u_{0,975} \simeq 1,96$
- Donc, avec proba $\geq 0,95$

Exemple 3

On cherche à estimer la proportion p_A de votants pour A.

- Soit X_i la v.a égale à 1 si la personne i a voté pour A et 0 sinon.
- Par hypothèse $\bar{X}_{500} = \frac{387}{500}$
- Les conditions $n = 500 \geq 30$, $n\bar{X}_{500} \geq 5$ *et* $n(1 - \bar{X}_{500}) \geq 500$ sont vérifiées.
- $\alpha = 0,05$ et on sait que pour une Normale, $u_{1-\frac{\alpha}{2}} = u_{0,975} \simeq 1,96$
- Donc, avec proba $\geq 0,95$

$$p_A \in [\bar{X}_{500} - 1,96 \times \frac{\sqrt{\bar{X}_{500}(1 - \bar{X}_{500})}}{\sqrt{500}} ; \bar{X}_{500} + 1,96 \times \frac{\sqrt{\bar{X}_{500}(1 - \bar{X}_{500})}}{\sqrt{500}}].$$

Exemple 3

On cherche à estimer la proportion p_A de votants pour A.

- Soit X_i la v.a égale à 1 si la personne i a voté pour A et 0 sinon.
- Par hypothèse $\bar{X}_{500} = \frac{387}{500}$
- Les conditions $n = 500 \geq 30,\ n\bar{X}_{500} \geq 5$ *et* $n(1 - \bar{X}_{500}) \geq 500$ sont vérifiées.
- $\alpha = 0,05$ et on sait que pour une Normale, $u_{1-\frac{\alpha}{2}} = u_{0,975} \simeq 1,96$
- Donc, avec proba $\geq 0,95$

$$p_A \in [\bar{X}_{500} - 1,96 \times \frac{\sqrt{\bar{X}_{500}(1-\bar{X}_{500})}}{\sqrt{500}}; \bar{X}_{500} + 1,96 \times \frac{\sqrt{\bar{X}_{500}(1-\bar{X}_{500})}}{\sqrt{500}}].$$

Avec $\bar{X}_{500} = \frac{387}{500}$, on obtient

$$p_A \in [0,737; 0,811], \quad \text{avec proba } 0,95.$$

Remarque IC lors d'élection

De l'expression littérale, on remarque que si l'on augmente la taille n de l'échantillon, l'intervalle (de confiance) se "resserre", ce qui permet de lever éventuellement une indetermination dans le cas où $1/2 \in [\bar{X}_n - \frac{1{,}96}{2\sqrt{n}}; \bar{X}_n + \frac{1{,}96}{2\sqrt{n}}]$.

Trame de l'obtention d'estimation

Rappel de la problématique : nous connaissons le (ou les) paramètre(s) de la population et nous voulons estimer la moyenne empirique d'un échantillon.

Trame de l'obtention d'estimation

Rappel de la problématique : nous connaissons le (ou les) paramètre(s) de la population et nous voulons estimer la moyenne empirique d'un échantillon.

- Il n'y a pas d'ingrédients mathématiques supplémentaires, il suffit de reprendre les inégalités du tableau précédent en disposant $\bar{X}_n$ au centre.
- A noter que pour une proportion empirique, nous obtenons une inégalité un peu plus "fine".

Récapitulatif intervalle de fluctuation moyenne/proportion empirique

Paramètre à estimer	**IF au seuil α**
moyenne $\bar{X}_n$	$[\mu - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \; ; \; \mu + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$
Condition *Lecture seuil*	$n \geq 30$ $\mathcal{N}(0,1)$
Proportion p_n	$[p - u_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \; ; \; p + u_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}]$
Condition *Lecture seuil*	$n \geq 30, \; et \; \min(np; n(1-p)) \geq 5$ $\mathcal{N}(0,1)$

Récapitulatif intervalle de fluctuation moyenne/proportion empirique

Paramètre à estimer	**IF au seuil α**
moyenne $\bar{X}_n$	$[\mu - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \; ; \; \mu + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$
Condition *Lecture seuil*	$n \geq 30$ $\mathcal{N}(0,1)$
Proportion p_n	$[p - u_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \; ; \; p + u_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}]$
Condition *Lecture seuil*	$n \geq 30$, *et* $\min(np; n(1-p)) \geq 5$ $\mathcal{N}(0,1)$