Probabilités et observations économiques

C. Tuleau-Malot

Université de Nice - Sophia Antipolis

Plan

Introduction

Ce que l'on a pû voir, avec les 10 jeux de données, c'est que même lorsque l'on simule des observations qui suivent la même loi, les estimations changent d'un jeu de données à un autre.

$\Rightarrow$ fluctuation d'échantillonnage

Introduction

Ce que l'on a pû voir, avec les 10 jeux de données, c'est que même lorsque l'on simule des observations qui suivent la même loi, les estimations changent d'un jeu de données à un autre.

$\Rightarrow$ fluctuation d'échantillonnage

Comment quantifier cela?

Introduction

Ce que l'on a pû voir, avec les 10 jeux de données, c'est que même lorsque l'on simule des observations qui suivent la même loi, les estimations changent d'un jeu de données à un autre.

$\Rightarrow$ fluctuation d'échantillonnage

Comment quantifier cela?

En construisant un intervalle de confiance!

Objectif

- par l'estimation ponctuelle, on obtient une valeur approchée d'une caractéristique de la population

Objectif

- par l'estimation ponctuelle, on obtient une valeur approchée d'une caractéristique de la population
- estimation obtenue à l'aide d'un échantillon de taille n

Objectif

- par l'estimation ponctuelle, on obtient une valeur approchée d'une caractéristique de la population
- estimation obtenue à l'aide d'un échantillon de taille n

$\Rightarrow$ quelle est l'influence de n?

Objectif

- par l'estimation ponctuelle, on obtient une valeur approchée d'une caractéristique de la population

- estimation obtenue à l'aide d'un échantillon de taille n

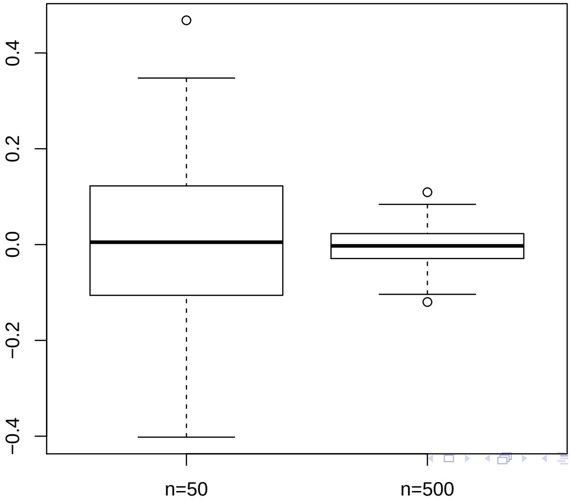
$\Rightarrow$ quelle est l'influence de n?

$\Rightarrow$ comment prendre en compte l'erreur du fait que l'on ne dispose que d'un échantillon et non de la population entière?

Visualisation

- On simule 100 jeux de données de taille respective $n = 50$ et $n = 500$
- les jeux de données ont été simulés à partir de loi loi normale centrée et réduite
- Pour chaque jeu de données, on calcule la moyenne empirique
- On représente la boîte à moustache associée aux jeux de données de taille 50, puis de taille 500
  Sur une boîte à moustache, on visualise les 1er, 3ème quartiles (limites inférieure et suprieure du rectangle), la médiane (trait en gras) et les limites.

Visualisation



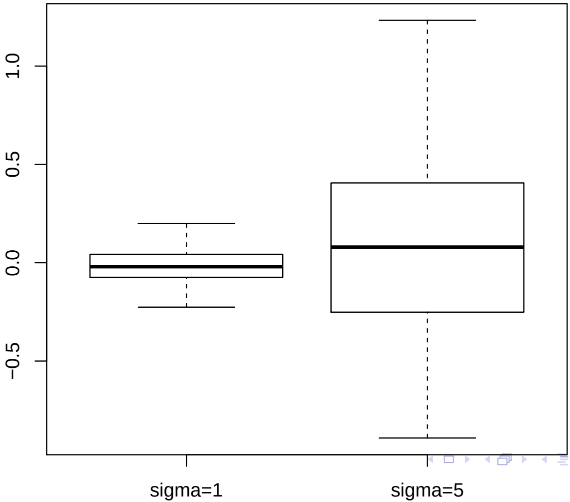
Visualisation

- on constate que quand n augmente, l'étalement des données diminue
- $\bar{X}_n$ est donc plus précis quand n est grand

Visualisation

- On simule 100 jeux de données de taille $n = 500$ dont on va faire bouger la variance.
- les jeux de données ont été simulés à partir de loi loi normale centrée de variance σ^2. On prend $\sigma = 1$ et $\sigma = 5$.
- Pour chaque jeu de données, on calcule la moyenne empirique
- On représente la boîte à moustache associée aux jeux de données associés à $\sigma = 1$ d'une part et $\sigma = 5$ d'autre part.

Visualisation



Visualisation

- on constate que quand σ augmente, l'étalement des données augmente
- $\bar{X}_n$ est donc moins précis quand σ est grand

Visualisation

L'intervalle de confiance va :

- quantifier ces dépendances
- donner une marge d'erreur quant à l'estimation ponctuelle

Visualisation

L'intervalle de confiance va :

- quantifier ces dépendances
- donner une marge d'erreur quant à l'estimation ponctuelle

Il va falloir poser un cadre : le cadre gaussien

Particularité du cadre gaussien

Proposition :

Si $X_1, \ldots, X_n$ forment un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, alors :

$$\bar{X}_n := \frac{X_1 + \ldots + X_n}{n} \sim \mathcal{N}(\mu, \sigma^2/n)$$

Particularité du cadre gaussien

Proposition :

Si $X_1, \ldots, X_n$ forment un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, alors :

$$\bar{X}_n := \frac{X_1 + \ldots + X_n}{n} \sim \mathcal{N}(\mu, \sigma^2/n)$$

Ceci n'est pas une conséquence du théorème central limit !

Pourquoi cela?

$$\begin{array}{rcll}
\mathbb{E}[\bar{X}_n] & = & \mathbb{E}[\frac{X_1+\ldots+X_n}{n}] & \\
 & = & \frac{\mathbb{E}[X_1+\ldots+X_n}{]} n & \text{par linéarité de l'espérance} \\
 & = & \frac{\mathbb{E}[X_1]+\ldots+\mathbb{E}[X_n]}{n} & \text{par linéarité de l'espérance} \\
 & = & \frac{\mu+\ldots+\mu}{n} & \text{car toutes de même loi} \\
 & = & \mu &
\end{array}$$

Pourquoi cela?

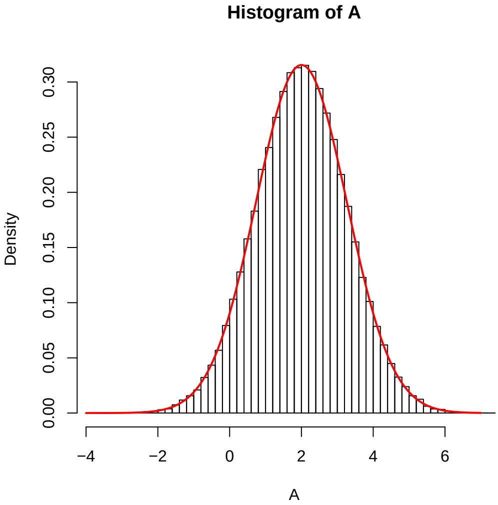
$$
\begin{array}{rcll}
\mathbb{E}[\bar{X}_n] & = & \mathbb{E}[\frac{X_1+\ldots+X_n}{n}] & \\
 & = & \frac{\mathbb{E}[X_1+\ldots+X_n}{]} n & \text{par linéarité de l'espérance} \\
 & = & \frac{\mathbb{E}[X_1]+\ldots+\mathbb{E}[X_n]}{n} & \text{par linéarité de l'espérance} \\
 & = & \frac{\mu+\ldots+\mu}{n} & \text{car toutes de même loi} \\
 & = & \mu & \\
\mathbb{V}[\bar{X}_n] & = & \mathbb{V}[\frac{X_1+\ldots+X_n}{n}] & \\
 & = & \frac{\mathbb{V}[X_1+\ldots+X_n}{]} n^2 & \text{par propriété de la variance} \\
 & = & \frac{\mathbb{E}[X_1]+\ldots+\mathbb{E}[X_n]}{n^2} & \text{par indépendance des variables} \\
 & = & \frac{\sigma^2+\ldots+\sigma^2}{n^2} & \text{car toutes de même loi} \\
 & = & \frac{\sigma^2}{n} &
\end{array}
$$

Le caractère gaussien car la somme de variables gaussiennes indépendantes est une variable gaussienne.

visualisation

- on a pris $n = 10$ et on a simulé 100000 jeux de donnés de taille $n = 10$ associés à une loi normale de paramètre $\mu = 2$ et $\sigma = 4$
- pour chacun des 100000 jeux de données, on évalue $\bar{X}_n$ et on trace l'histogramme que l'on compare à la distribution théorique signifiée par le théorème central limit.
- on constate que la densité théorique de la loi normale et l'histogramme des $\bar{X}_n$ sont très proches

visualisation



Principe d'un intervalle de confiance pour l'espérance

- On veut construire un intervalle dans lequel va se trouver très probablement le paramètre μ

Principe d'un intervalle de confiance pour l'espérance

- On veut construire un intervalle dans lequel va se trouver très probablement le paramètre μ
- on ne connaît pas μ mais une valeur approchée avec $\bar{X}_n$

Principe d'un intervalle de confiance pour l'espérance

- On veut construire un intervalle dans lequel va se trouver très probablement le paramètre μ
- on ne connaît pas μ mais une valeur approchée avec $\bar{X}_n$
- idée : on construit un intervalle autour de $\bar{X}_n$: $[\bar{X}_n - \epsilon; \bar{X}_n + \epsilon]$

Principe d'un intervalle de confiance pour l'espérance

- On veut construire un intervalle dans lequel va se trouver très probablement le paramètre μ
- on ne connaît pas μ mais une valeur approchée avec $\bar{X}_n$
- idée : on construit un intervalle autour de $\bar{X}_n$: $[\bar{X}_n - \epsilon; \bar{X}_n + \epsilon]$

il faut construire ϵ

$\bar{X}_n$ est une variable aléatoire $\Rightarrow$ l'intervalle est aléatoire

Principe d'un intervalle de confiance pour l'espérance

- On veut construire un intervalle dans lequel va se trouver très probablement le paramètre μ
- on ne connaît pas μ mais une valeur approchée avec $\bar{X}_n$
- idée : on construit un intervalle autour de $\bar{X}_n$: $[\bar{X}_n - \epsilon; \bar{X}_n + \epsilon]$

$\bar{X}_n$ est une variable aléatoire $\Rightarrow$ l'intervalle est aléatoire

l'intervalle va changer quand l'échantillon va changer!

niveau de confiance

- μ sera tantôt dans l'intervalle de confiance et tantôt en dehors de l'intervalle

niveau de confiance

- μ sera tantôt dans l'intervalle de confiance et tantôt en dehors de l'intervalle
- on aimerait que quand on crée beaucoup d'intervalles, très souvent μ soit dans l'intervalle

niveau de confiance

- μ sera tantôt dans l'intervalle de confiance et tantôt en dehors de l'intervalle
- on aimerait que quand on crée beaucoup d'intervalles, très souvent μ est dans l'intervalle

par exemple : dans 95% des cas

niveau de confiance

- μ sera tantôt dans l'intervalle de confiance et tantôt en dehors de l'intervalle
- on aimerait que quand on crée beaucoup d'intervalles, très souvent μ est dans l'intervalle

  par exemple : dans 95% des cas
- ce taux est appelé **niveau de confiance** (dans l'exemple 95%)

niveau de confiance

- On note α la probabilité que μ ne soit pas dans l'intervalle de confiance

niveau de confiance

- On note α la probabilité que μ ne soit pas dans l'intervalle de confiance
- le niveau est alors $1 - \alpha$

niveau de confiance

- On note α la probabilité que μ ne soit pas dans l'intervalle de confiance
- le niveau est alors $1 - \alpha$
- trouver ϵ de sorte que μ soit dans $[\bar{X}_n - \epsilon; \bar{X}_n + \epsilon]$ avec probabilité $1 - \alpha$

visualisation

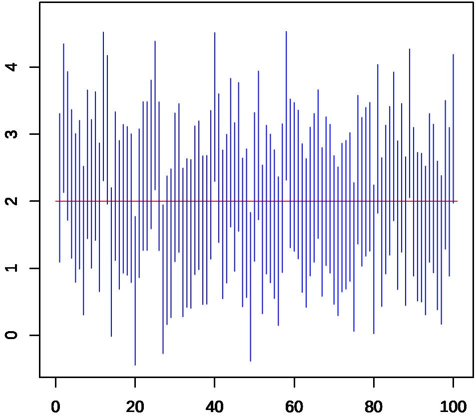
Pour visualiser le fait que le vrai paramètre peut être en dehors, nous avons simulés 100 jeux de données associés à des lois normales d'espérance 2.

Pour chacun des jeux de données, nous avons construit l'intervalle de confiance pour le paramètre d'espérance. Ils sont représentés par les traits verticaux bleus.

La valeur cible est matérialisée par le trait horizontal rouge.

On voit que très souvent le trait rouge coupe les traits bleus, mais il arrive que cela ne soit pas le cas. Le nombre de fois où tel n'est pas le cas est de l'ordre de 100.α.

visualisation



niveau de confiance

<u>influence de ϵ :</u>

- plus ϵ est grand $\Rightarrow$ plus μ a des chances d'être dans l'intervalle

  $\Rightarrow$ niveau de confiance élevé

niveau de confiance

<u>influence de ϵ :</u>

- plus ϵ est grand $\Rightarrow$ plus μ a des chances d'être dans l'intervalle

  $\Rightarrow$ niveau de confiance élevé

- plus ϵ est grand $\Rightarrow$ moins $\bar{X}_n$ est précis

  $\Rightarrow$ ϵ : marge d'erreur

construction d'un intervalle de confiance

On veut :

$$P(\mu \in [\bar{X}_n - \epsilon; \bar{X}_n + \epsilon]) = 1 - \alpha$$

construction d'un intervalle de confiance

On veut :

$$P(\mu \in [\bar{X}_n - \epsilon; \bar{X}_n + \epsilon]) = 1 - \alpha$$

Or :

$$\begin{aligned} P(\mu \in [\bar{X}_n - \epsilon; \bar{X}_n + \epsilon]) &= P(\bar{X}_n - \epsilon \leq \mu \leq \bar{X}_n + \epsilon) \\ &= P(-\epsilon \leq \mu - \bar{X}_n \leq \epsilon) \\ &= P(-\epsilon \leq \bar{X}_n - \mu \leq \epsilon) \end{aligned}$$

construction d'un intervalle de confiance

Or, dans le cadre gaussien, on a :

$$\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$$

Soit :

$$\sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

construction d'un intervalle de confiance

On veut :

$$P(\mu \in [\bar{X}_n - \epsilon ; \bar{X}_n + \epsilon]) = 1 - \alpha$$

Or :

$$P(\mu \in [\bar{X}_n - \epsilon ; \bar{X}_n + \epsilon]) = P(-\epsilon \leq \bar{X}_n - \mu \leq \epsilon)$$

mais encore :

$$P(\mu \in [\bar{X}_n - \epsilon ; \bar{X}_n + \epsilon]) = P(\sqrt{n}.\frac{-\epsilon}{\sigma} \leq \sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma} \leq \sqrt{n}.\frac{\epsilon}{\sigma})$$

(on a centré et réduit la variable $\bar{X}_n$)

construction d'un intervalle de confiance

Or :

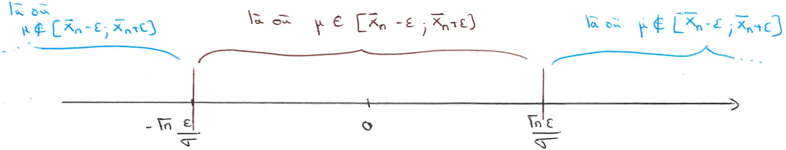
$$\alpha = P(\mu \notin [\bar{X}_n - \epsilon; \bar{X}_n + \epsilon]) = P(\sqrt{n}.\frac{-\epsilon}{\sigma} \geq \sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma} \text{ ou } \sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma} \geq \sqrt{n}.\frac{\epsilon}{\sigma})$$

construction d'un intervalle de confiance

Or :

$$P(\mu \notin [\bar{X}_n - \epsilon; \bar{X}_n + \epsilon]) = P(\sqrt{n}.\frac{-\epsilon}{\sigma} \geq \sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma} \text{ ou } \sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma} \geq \sqrt{n}.\frac{\epsilon}{\sigma})$$

En effet :



construction d'un intervalle de confiance

Or :

$$\begin{aligned}
P(\mu \notin [\bar{X}_n - \epsilon; \bar{X}_n + \epsilon]) &= P(\sqrt{n}.\frac{-\epsilon}{\sigma} \geq \sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma} \\
&\quad \text{ou } \sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma} \geq \sqrt{n}.\frac{\epsilon}{\sigma}) \\
&= P(\sqrt{n}.\frac{-\epsilon}{\sigma} \geq \sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma}) \\
&\quad + P(\sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma} \geq \sqrt{n}.\frac{\epsilon}{\sigma})
\end{aligned}$$

En effet, les événements que l'on considère sont disjoints. (Sinon cela serait faux car il ne faut pas oublier que $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ mais que dans le cas d'événements disjoints $P(A \cap B) = 0$)

construction d'un intervalle de confiance

Or :

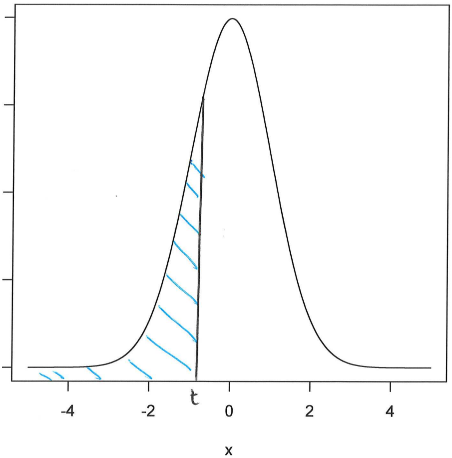
$$\begin{aligned} P(\mu \notin [\bar{X}_n - \epsilon; \bar{X}_n + \epsilon]) &= P(\sqrt{n}.\frac{-\epsilon}{\sigma} \geq \sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma} \\ &\quad \text{ou } \sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma} \geq \sqrt{n}.\frac{\epsilon}{\sigma}) \\ &= P(\sqrt{n}.\frac{-\epsilon}{\sigma} \geq \sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma}) \\ &\quad + P(\sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma} \geq \sqrt{n}.\frac{\epsilon}{\sigma}) \end{aligned}$$

Et par symétrie de la distribution d'une $\mathcal{N}(0, 1)$, on a :

$$P(\mu \notin [\bar{X}_n - \epsilon; \bar{X}_n + \epsilon]) = 2.P(\sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma} \geq \sqrt{n}.\frac{\epsilon}{\sigma})$$

Rappel de la symétrie d'une loi normale

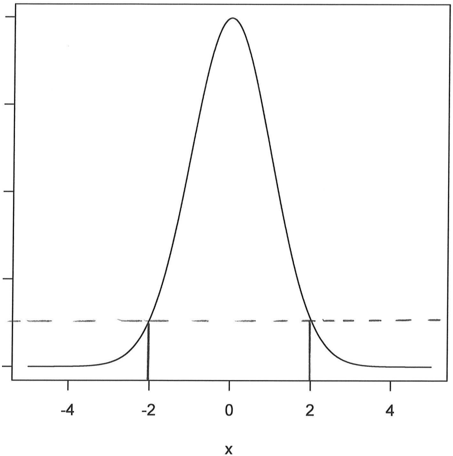
fonction de densité de ~~loi~~ la loi $\mathcal{N}(0;1)$



aire hachurée $= P(X \leq t)$ avec $X \sim \mathcal{N}(0;1)$

Rappel de la symétrie d'une loi normale

fonction de densité de la loi $\mathcal{N}(0;1)$

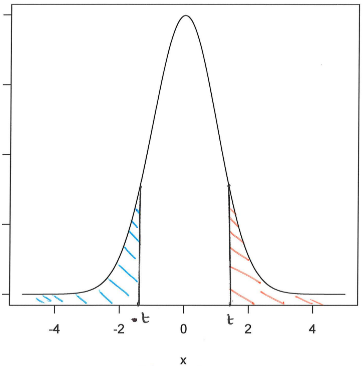


parité de la fonction :

pour tout $x \in \mathbb{R}$, $f(x) = f(-x)$

Rappel de la symétrie d'une loi normale

fonction de densité de la loi $\mathcal{N}(0;1)$



aire bleue = $P(X \leq -t)$

aire orange = $P(X \geq t)$

on a $P(X \leq -t) = P(X \geq t)$

construction d'un intervalle de confiance

Or on veut que :

$$P(\mu \notin [\bar{X}_n - \epsilon ; \bar{X}_n + \epsilon]) = \alpha$$

Soit :

$$P(\sqrt{n}.\frac{\bar{X}_n - \mu}{\sigma} \geq \sqrt{n}.\frac{\epsilon}{\sigma}) = \alpha/2$$

construction d'un intervalle de confiance

Dans la table de la loi normale centrée réduite, on cherche $z_{\alpha/2}$ tel que

$$P(\mathcal{N}(0,1) \geq z_{1-\alpha/2}) = \alpha/2$$

$z_{1-\alpha/2}$ est de fractile d'ordre $1 - \alpha/2$ associé à une loi normale centrée réduite

construction d'un intervalle de confiance

On pose alors :

$$\sqrt{n}.\epsilon/\sigma = z_{1-\alpha/2}$$

Soit :

$$\epsilon = \frac{z_{1-\alpha/2}.\sigma}{\sqrt{n}}$$

construction d'un intervalle de confiance

On en déduit l'intervalle de confiance :

$$\left[\bar{X}_n - \frac{z_{1-\alpha/2}.\sigma}{\sqrt{n}}, \bar{X}_n + \frac{z_{1-\alpha/2}.\sigma}{\sqrt{n}}\right]$$

construction d'un intervalle de confiance

On en déduit l'intervalle de confiance :

$$\left[\bar{X}_n - \frac{z_{1-\alpha/2}.\sigma}{\sqrt{n}}, \bar{X}_n + \frac{z_{1-\alpha/2}.\sigma}{\sqrt{n}}\right]$$

- plus σ est grand $\Rightarrow$ plus la marge d'erreur est grande

construction d'un intervalle de confiance

On en déduit l'intervalle de confiance :

$$\left[\bar{X}_n - \frac{z_{1-\alpha/2}.\sigma}{\sqrt{n}}, \bar{X}_n + \frac{z_{1-\alpha/2}.\sigma}{\sqrt{n}}\right]$$

- plus σ est grand $\Rightarrow$ plus la marge d'erreur est grande
- plus n est grand $\Rightarrow$ plus la marge d'erreur est petite

Exemple

On souhaite déterminer la dépense moyenne des clients d'un centre commercial.

On suppose que la distribution des dépenses des clients suit une loi normale d'écart-type $\sigma = 30$. On considère un échantillon de taille

$n = 16$ pour lequel $\bar{x}_n = 140$.

Déterminer l'intervalle de confiance pour les dépenses moyennes au niveau 90%.

Exemple : solution

On cherche un intervalle de confiance pour le paramètre d'espérance de la loi.

Exemple : solution

On cherche un intervalle de confiance pour le paramètre d'espérance de la loi. D'après ce qui précède, on sait que

l'intervalle de confiance a pour expression:

$$\left[140 - \frac{30.z_{1-\alpha/2}}{4}, 140 + \frac{30.z_{1-\alpha/2}}{4}\right]$$

On a utilisé la formule précédent dans laquelle on a remplacé les paramètre par leur valeur quand on les connaissaient.

Il suffit à présent d'aller chercher la valeur de $z_{1-\alpha/2}$ dans la table de la loi normale centrée réduite.

table loi normale centrée réduite

- il n'existe qu'une table celle de la loi normale centrée réduite
- cela résulte du lien entre une loi normale quelconque et la loi normale centrée réduite
  Si $X \sim \mathcal{N}(\mu, \sigma^2)$ alors $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$
- Dans la table, ce sont les probabilités $P(X \leq z)$ qui sont retranscrites, avec $X \sim \mathcal{N}(0,1)$

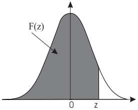
table loi normale centrée réduite

- Si on cherche $P(X \leq 1.54)$, on se regarde la valeur dans la table à l'intersection de la ligne associée à 1.5 et de la colonne associée à 0.04.
  Sur la table, on lit 0.9382.

- Si on cherche t tel que $P(X \leq t) = 0.975$, on cherche dans le corps de la table la valeur 0.975 et on en déduit la valeur de t, en l'occurrence 1.96 sur notre exemple.

table loi normale centrée réduite

Fonction de répartition de la loi normale centrée réduite
(probabilité F(z) de trouver une valeur inférieure à z)



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981

Exemple : solution

En quoi cela va nous aider?

On cherche $z_{\alpha/2}$ tel que

$$P(\mathcal{N}(0,1) \geq z_{1-\alpha/2}) = \alpha/2$$

Autrement dit tel que :

$$P(\mathcal{N}(0,1) \leq z_{1-\alpha/2}) = F(z_{1-\alpha/2}) = 1 - \alpha/2$$

Dans la table, on lit pour $\alpha = 0.1$, 1.645
On prend en effet le milieu entre 1.64 et 1.65, puisque 0.95 est le milieu de 0.9495 et de 0.9505.

Exemple : solution

On en déduit que :

$$\left[140 - \frac{30.1,645}{4}, 140 + \frac{30.1,645}{4}\right]$$

Soit

$$[127.66; 152.34]$$

construction d'un intervalle de confiance : cadre non gaussien

- tout ce qui précède a été fait dans le cadre gaussien

construction d'un intervalle de confiance : cadre non gaussien

- tout ce qui précède a été fait dans le cadre gaussien
- et si pas dans le cadre gaussien

construction d'un intervalle de confiance : cadre non gaussien

- tout ce qui précède a été fait dans le cadre gaussien
- et si pas dans le cadre gaussien

  $\Rightarrow$ utilisation du Théorème Central Limit

Théorème Central Limit

Soit $X_1, \dots, X_n$ un n-échantillon dont on note μ l'espérance et σ^2 la variance. Alors, pour n grand, on a :

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \rightsquigarrow \mathcal{N}(0, 1)$$

construction d'un intervalle de confiance : cadre non gaussien

On va utiliser le fait que $\bar{X}_n$ suit approximativement une loi normale d'espérance μ et de variance σ^2/n.

construction d'un intervalle de confiance : cadre non gaussien

On va utiliser le fait que $\bar{X}_n$ suit approximativement une loi normale d'espérance μ et de variance σ^2/n.

En menant les calculs comme précédemment, on retrouve l'intervalle précédent :

$$\left[\bar{X}_n - \frac{z_{1-\alpha/2}.\sigma}{\sqrt{n}}, \bar{X}_n + \frac{z_{1-\alpha/2}.\sigma}{\sqrt{n}}\right]$$

construction d'un intervalle de confiance : cadre non gaussien

On va utiliser le fait que $\bar{X}_n$ suit approximativement une loi normale d'espérance μ et de variance σ^2/n.

En menant les calculs comme précédemment, on retrouve l'intervalle précédent :

$$\left[\bar{X}_n - \frac{z_{1-\alpha/2}.\sigma}{\sqrt{n}}, \bar{X}_n + \frac{z_{1-\alpha/2}.\sigma}{\sqrt{n}}\right]$$

La seule différence avec ce qui précède c'est que pour déterminer $z_{1-\alpha/2}$, on remplace la vraie loi pour l'approximation gaussienne.
$\Rightarrow$ on obtient un intervalle de confiance asymptotique et non plus exact comme précédemment puisque :

$$\lim_{n\to+\infty} P(\mu \in \text{intevalle de confiance}) = 1 - \alpha$$

Exemple

On souhaite déterminer la dépense moyenne des clients d'un centre commercial.

On suppose que la distribution des dépenses des clients a un écart-type $\sigma = 30$. On considère un échantillon de taille $n = 100$

pour lequel $\bar{x}_n = 140$.

Déterminer l'intervalle de confiance pour les dépenses moyennes au niveau 90%.

Exemple : solution

On cherche un intervalle de confiance pour le paramètre d'espérance de la loi.

Exemple : solution

On cherche un intervalle de confiance pour le paramètre d'espérance de la loi. D'après ce qui précède, on sait que

l'intervalle de confiance a pour expression:

$$\left[140 - \frac{30.z_{1-\alpha/2}}{10}, 140 + \frac{30.z_{1-\alpha/2}}{10}\right]$$

Exemple : solution

On en déduit que :

$$\left[140 - \frac{30.1,645}{4}, 140 + \frac{30.1,645}{10}\right]$$

Soit

$$[135.06; 144.94]$$

variance inconnue

- tout ce qui précède a été fait à variance connue

variance inconnue

- tout ce qui précède a été fait à variance connue
- et si la variance n'est pas connue?

variance inconnue

- tout ce qui précède a été fait à variance connue
- et si la variance n'est pas connue?

  $\Rightarrow$ on va l'estimer

construction d'un intervalle de confiance : cadre gaussien et variance inconnue

Quelques définitions :

- Soit $Y_1, \ldots, Y_n$ un n-échantillon de loi $\mathcal{N}(0,1)$.
  On considère $C = Y_1^2 + \ldots + Y_n^2$.
  Ainsi définie, on dit que C suit une loi du Chi-Deux (χ^2) à n degré de liberté.

construction d'un intervalle de confiance : cadre gaussien et variance inconnue

Quelques définitions :

Soit $Y_1, \ldots, Y_n$ un n-échantillon de loi $\mathcal{N}(0,1)$.
On considère $C = Y_1^2 + \ldots + Y_n^2$.
Ainsi définie, on dit que C suit une loi du Chi-Deux (χ^2) à n degré de liberté.

- C n peut prendre que des valeurs positives
- le degré de liberté est le nombre de variables $Y_1, \ldots, Y_n$ que l'on peut choisir librement pour entièrement définir C.

construction d'un intervalle de confiance : cadre gaussien et variance inconnue

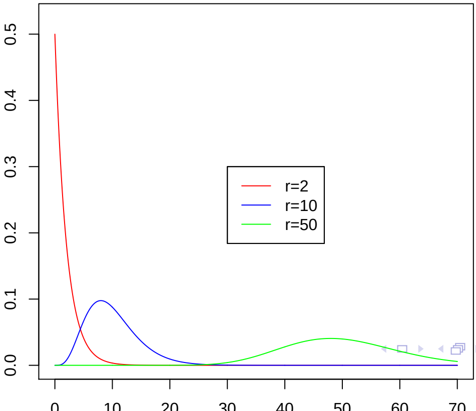
Quelques définitions :

Soit $Y_1, \ldots, Y_n$ un n-échantillon de loi $\mathcal{N}(0,1)$.
On considère $C = Y_1^2 + \ldots + Y_n^2$.
Ainsi définie, on dit que C suit une loi du Chi-Deux (χ^2) à n degré de liberté.

- C n peut prendre que des valeurs positives
- le degré de liberté est le nombre de variables $Y_1, \ldots, Y_n$ que l'on peut choisir librement pour entièrement définir C.

construction d'un intervalle de confiance : cadre gaussien et variance inconnue

Quelques définitions : Visualisation de la fonction de densité d'une loi du χ^2 selon la valeur du degré de liberté



construction d'un intervalle de confiance : cadre gaussien et variance inconnue

Quelques définitions :

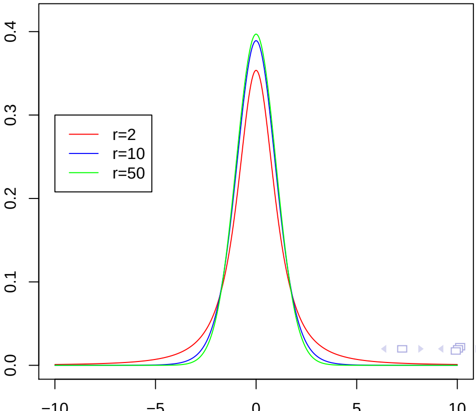
Soit Y une loi $\mathcal{N}(0,1)$ et V une loi du χ^2 à r degré de liberté. On suppose de plus que Y et V sont indépendantes.

On considère $T = \frac{Y}{\sqrt{V/r}}$. Alors ainsi définie, T suit une loi de Student à r degré de liberté.

- la fonction de densité d'une loi de Student est symétrique par rapport à l'axe des ordonnées
- on a ainsi les mêmes propriétés de symétries que pour la loi normale

construction d'un intervalle de confiance : cadre gaussien et variance inconnue

Quelques définitions : Visualisation de la fonction de densité d'une loi de Student en fonction de son degré de liberté



construction d'un intervalle de confiance : cadre gaussien et variance inconnue

Quelques résultats :

Soit $X_1, \ldots, X_n$ un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$ avec μ et σ^2 inconnus.

- Soit $U = \frac{n-1}{\sigma^2}\bar{S}_n^2$ avec
  $\bar{S}_n^2 = \frac{1}{n-1}\left((X_1 - \bar{X}_n)^2 + \ldots + (X_n - \bar{X}_n)^2\right)$
  On peut montrer et on admet que U suit une loi du Chi-Deux (χ^2) à $(n-1)$ degré de liberté.

construction d'un intervalle de confiance : cadre gaussien et variance inconnue

Quelques résultats :

Soit $X_1, \ldots, X_n$ un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$ avec μ et σ^2 inconnus.

- Soit $U = \frac{n-1}{\sigma^2}\bar{S}_n^2$ avec $\bar{S}_n^2 = \frac{1}{n-1}\left((X_1 - \bar{X}_n)^2 + \ldots + (X_n - \bar{X}_n)^2\right)$
  On peut montrer et on admet que U suit une loi du Chi-Deux (χ^2) à $(n-1)$ degré de liberté.

- On admet aussi que U (ou $\bar{S}_n^2$) et $Z = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}$ sont des variables indépendantes

construction d'un intervalle de confiance : cadre gaussien et variance inconnue

Quelques résultats :

- Si on considère

$$T = \frac{Z}{\sqrt{U/(n-1)}}$$

alors T suit une loi de Student à $(n-1)$ degré de liberté, par définition de la loi de Student.

construction d'un intervalle de confiance : cadre gaussien et variance inconnue

<u>Quelques résultats :</u>

- Si on considère

$$T = \frac{Z}{\sqrt{U/(n-1)}}$$

alors T suit une loi de Student à $(n-1)$ degré de liberté, par définition de la loi de Student.

On montre, après avoir remplacé Z par sa définition et procédé à des simplifications que :

$$T = \sqrt{n}\frac{\bar{X}_n - \mu}{\bar{S}_n}$$

construction d'un intervalle de confiance : cadre gaussien et variance inconnue

Quelques résultats :
Si l'on compare les expressions de Z et T on voit que T est Z dans lequel on a remplacé σ par $\bar{S}_n$.

Quand on remplace un vrai paramètre par un estimateur, on peut changer la distribution. Par exemple, ici on passe d'une loi normale à une loi de Student.

construction d'un intervalle de confiance : cadre gaussien et variance inconnue

En menant les calculs comme précédemment, on cherche ϵ tel que :

$$
\begin{aligned}
P(\mu \notin [\bar{X}_n - \epsilon; \bar{X}_n + \epsilon]) &= P(\sqrt{n}.\frac{-\epsilon}{\bar{S}_n} \geq \sqrt{n}.\frac{\bar{X}_n - \mu}{\bar{S}_n} \\
&\quad \text{ou } \sqrt{n}.\frac{\bar{X}_n - \mu}{\bar{S}_n} \geq \sqrt{n}.\frac{\epsilon}{\bar{S}_n}) \\
&\quad \text{on réduit par } \bar{S}_n/\sqrt{n} \\
&\quad \text{au lieu de réduire par } \sigma/\sqrt{n} \\
&= P(\sqrt{n}.\frac{-\epsilon}{\bar{S}_n} \geq \sqrt{n}.\frac{\bar{X}_n - \mu}{\bar{S}_n}) \\
&\quad + P(\sqrt{n}.\frac{\bar{X}_n - \mu}{\bar{S}_n} \geq \sqrt{n}.\frac{\epsilon}{\bar{S}_n})
\end{aligned}
$$

construction d'un intervalle de confiance : cadre gaussien et variance inconnue

En menant les calculs comme précédemment, on cherche ϵ tel que :

$$
\begin{aligned}
P(\mu \notin [\bar{X}_n - \epsilon ; \bar{X}_n + \epsilon]) &= P(\sqrt{n}.\frac{-\epsilon}{\bar{S}_n} \geq \sqrt{n}.\frac{\bar{X}_n - \mu}{\bar{S}_n} \\
& \quad \text{ou } \sqrt{n}.\frac{\bar{X}_n - \mu}{\bar{S}_n} \geq \sqrt{n}.\frac{\epsilon}{\bar{S}_n}) \\
&= P(\sqrt{n}.\frac{-\epsilon}{\bar{S}_n} \geq \sqrt{n}.\frac{\bar{X}_n - \mu}{\bar{S}_n}) \\
& \quad + P(\sqrt{n}.\frac{\bar{X}_n - \mu}{\bar{S}_n} \geq \sqrt{n}.\frac{\epsilon}{\bar{S}_n}) \\
&= P(T \leq \sqrt{n}.\frac{-\epsilon}{\bar{S}_n}) + P(T \geq \sqrt{n}.\frac{\epsilon}{\bar{S}_n})
\end{aligned}
$$

construction d'un intervalle de confiance : cadre gaussien et variance inconnue

En menant les calculs comme précédemment, on cherche ϵ tel que :

$$\begin{aligned} P(\mu \notin [\bar{X}_n - \epsilon; \bar{X}_n + \epsilon]) &= P(\sqrt{n}.\frac{-\epsilon}{\bar{S}_n} \geq \sqrt{n}.\frac{\bar{X}_n - \mu}{\bar{S}_n} \\ &\quad \text{ou } \sqrt{n}.\frac{\bar{X}_n - \mu}{\bar{S}_n} \geq \sqrt{n}.\frac{\epsilon}{\bar{S}_n}) \\ &= P(\sqrt{n}.\frac{-\epsilon}{\bar{S}_n} \geq \sqrt{n}.\frac{\bar{X}_n - \mu}{\bar{S}_n}) \\ &\quad + P(\sqrt{n}.\frac{\bar{X}_n - \mu}{\bar{S}_n} \geq \sqrt{n}.\frac{\epsilon}{\bar{S}_n}) \\ &= 2.P(T \geq \sqrt{n}.\frac{\epsilon}{\bar{S}_n}) \end{aligned}$$

dernière égalité : par symétrie de la loi de Student (même phénomène que pour la loi normale)

construction d'un intervalle de confiance : cadre gaussien et variance inconnue

On lit ainsi dans la table associée $t_{1-\alpha/2,n-1}$ de sorte que :

$$P(T \geq t_{1-\alpha/2,n-1}) = \alpha/2$$

construction d'un intervalle de confiance : cadre gaussien et variance inconnue

On lit ainsi dans la table associée $t_{\alpha/2,n-1}$ de sorte que :

$$P(T \geq t_{1-\alpha/2,n-1}) = \alpha/2$$

Au final, on obtient comme intervalle :

$$\left[\bar{X}_n - \frac{t_{1-\alpha/2,n-1}.\bar{S}_n}{\sqrt{n}}, \bar{X}_n + \frac{t_{1-\alpha/2,n-1}.\bar{S}_n}{\sqrt{n}}\right]$$

Exemple

On souhaite déterminer la dépense moyenne des clients d'un centre commercial.

On suppose que la distribution des dépenses des clients suit une loi normale. On considère un échantillon de taille $n = 16$ pour lequel

$\bar{x}_n = 140$ et $\bar{s}_n = 30$.

Déterminer l'intervalle de confiance pour les dépenses moyennes au niveau 90%.

Exemple : solution

On cherche un intervalle de confiance pour le paramètre d'espérance de la loi.

Exemple : solution

On cherche un intervalle de confiance pour le paramètre d'espérance de la loi. D'après ce qui précède, on sait que

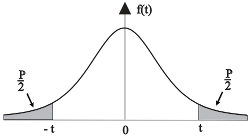
l'intervalle de confiance a pour expression:

$$\left[140 - \frac{30.t_{1-\alpha/2,15}}{4}, 140 + \frac{30.t_{1-\alpha/2;15}}{4}\right]$$

Exemple : solution

Table de la loi de Student

Valeurs de **T** ayant la probabilité **P** d'être dépassées en valeur absolue



ν	P =0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,128	0,260	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921

Exemple : solution

D'où :

$$\left[140 - \frac{30.1,753}{4}, 140 + \frac{30.1,753}{4}\right]$$

Soit :

$$[126,85; 153,15]$$