EXAM de STAT L2 S2 ECO FI DU CNI E1

Questions de Cours

① Aléatoire : échantillon, moment empirique centré d'ordre 2

Non aléatoire : écart-type théorique

② La méthode des moments est une méthode permettant de trouver une forme fonctionnelle pour un estimateur $f(X_1, \dots, X_n)$. Elle consiste simplement à utiliser comme estimateur du moment théorique d'ordre k le moment empirique d'ordre k. C'est ainsi que la moyenne empirique (moment empirique d'ordre 1) est un estimateur de l'espérance (moment théorique d'ordre 1)

Exercice 2

Pour poser

Population mère : L'ensemble des clients de l'enseigne

Variable aléatoire parente X_i : consentement à payer du client i.

$X_i \sim \mathcal{D}(\mu, \sigma)$ inconnues

Echantillon simple au hasard constitué de tirages indépendants de X_i pour $i = 1, \dots n$, $n = 138$

On observe la moyenne et l'écart-type d'échantillon :

$$\overline{X}_{138} = 26 \quad \text{et} \quad S_{c\,138} = 5$$

Formules des stats d'éch.

où $$\overline{X}_{138} = \frac{1}{138} \sum_{i=1}^{138} X_i$$

$$S_{c\,138} = \frac{1}{137} \sum_{i=1}^{138} \left(X_i - \overline{X}_{138}\right)^2$$

On cherche à construire $IC_{90\%}(\mu) = [a ; b]$

tel que $P[a \leq \mu \leq b] = 99$

• estimateur pour μ : moyenne empirique $\overline{X}_{138}$.

• loi de proba de $\overline{X}_{138}$.

on ne connaît pas la distribution de la v.a parent

on utilise le <u>théorème central limite</u> :

soient $X_1, \dots, X_n$ une suite iid de v.a $X_i \sim D(\mu, \sigma)$.

$$\sqrt{n}\ \frac{\overline{X}_n - \mu}{\sigma} \underset{\infty}{\sim} N(0,1)$$

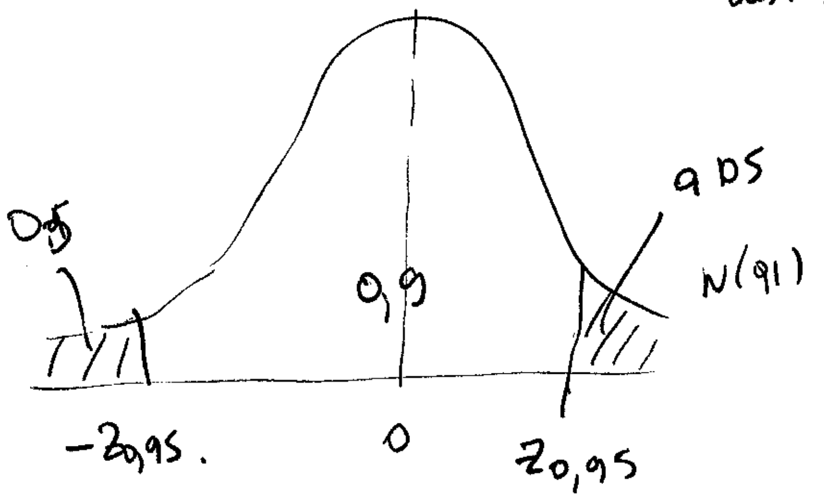
Ainsi, $\overline{X}_{138} \underset{\infty}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{138}}\right)$

• σ est inconnu il faut l'estimer avec $S_{c\,138}$

Dans ce cas, n étant grand, on a $\sqrt{n}\ \frac{\overline{X}_{138} - \mu}{S_{c138}} \underset{\infty}{\sim} N(0,1)$.

$$Z = \sqrt{138}\ \frac{\overline{X}_{138} - \mu}{S_{c138}} \underset{\infty}{\sim} N(0,1).$$

distribution de Z



$z_{0,95} = \Phi^{-1}(0,95)$
$= 1,64$ ou $1,65$.

ainsi $P(-1,645 \leq Z \leq 1,645) = 0,9$

Ex 1 suite

Donc $P\left(-1{,}645 \leq \sqrt{138}\,\dfrac{\overline{X}_{138}-\mu}{Sc_{138}} \leq 1{,}645\right) = 0{,}9$

$\times \dfrac{Sc_{138}}{\sqrt{138}}$ $\quad P\left(-1{,}645 \times \dfrac{Sc_{138}}{\sqrt{138}} \leq \overline{X}_{138}-\mu \leq 1{,}645 \times \dfrac{Sc_{138}}{\sqrt{138}}\right) = 0{,}9$

$-\overline{X}_{138}$ $\quad P\left(-\overline{X}_{138} - 1{,}645 \times \dfrac{Sc_{138}}{\sqrt{138}} \leq -\mu \leq -\overline{X}_{138} + 1{,}645 \times \dfrac{Sc_{138}}{\sqrt{138}}\right) = 0{,}9$

$\times(-1)$ $\quad P\left(\underbrace{\overline{X}_{138} - 1{,}645 \times \dfrac{Sc_{138}}{\sqrt{138}}}_{a = 26 - \underbrace{1{,}645 \times \frac{5}{\sqrt{138}}}_{0{,}7}} \leq \mu \leq \underbrace{\overline{X}_{138} + 1{,}645\, \dfrac{Sc_{138}}{\sqrt{138}}}_{b = 26 + 1{,}645 \times \frac{5}{\sqrt{138}}}\right) = 0{,}9$

ainsi $\boxed{IC_{90\%}(\mu) = [25{,}3\ ;\ 26{,}7]}$

- « Il y a 90 % de chances pour que μ soit compris entre 25,3 € et 26,7 € ».

L'introduction du nouveau produit est donc rentable puisque $25{,}3 > 20$ et $26{,}7 > 20$ avec 10 % de chances de se tromper.

Exercice 2

① Modélisation :

Nous avons 2 populations mères

A : personnes inscrits depuis moins de 3 mois ciblées

B : " " " non ciblées

On dispose d'un échantillon simple au hasard pour chaque popᵒ de tailles

$m_A = 160$ et $m_B = 4276 = 4436 - 160$.

On définit les variables aléatoires X_i^A et X_i^B qui associent à chaque chômeur i une valeur 0 s'il n'a pas retrouvé d'emploi ou 1 s'il a retrouvé un emploi. Ces variables aléatoires suivent une loi de Bernoulli de paramètres respectivement p_A et p_B. $X_i^A \sim b(p_A)$ $X_i^B \sim b(p_B)$

On suppose l'indépendance entre les populations A et B (tirage aléatoire). On réalise le test suivant :

$$\begin{cases} H_0 : p_A - p_B = 0 & \text{(les probas de retour à l'emploi sont identiques pour les popᵒ A et B.)} \\ H_1 : p_A - p_B > 0 & \text{(les probas sont} \neq \text{: la popᵒ A connaît de meilleures proba de retour à l'emploi).} \end{cases}$$

au seuil $\alpha = 10\%$.

Estimation

On doit estimer les paramètres inconnus $p_A - p_B$ à l'aide de l'estimateur écart de fréquences empiriques $F_n^A - F_n^B$

où $F_n^A = \frac{1}{m_A} \sum_{i=1}^{m_A} X_i^A$ et $F_n^B = \frac{1}{m_B} \sum_{i=1}^{m_B} X_i^B$

$\text{plim } F_n^A = p_A$ et $\text{plim } F_n^B = p_B$ donc $\underset{n \to +\infty}{\text{plim}} F_n^A - F_n^B = p_A - p_B$

Distribution de l'estimateur

Par le Théorème Central limite les X_i^A sont iid $E(X_i^A) = p_A$
$V(X_i^A) = p_A(1-p_A)$.

$$\sqrt{n_A}\,\frac{\overline{X_{n_A}} - p_A}{\sqrt{p_A(1-p_A)}} \underset{\infty}{\sim} N(0,1)$$

d'où $F_{n_A} = \overline{X}_{n_A} \underset{\infty}{\sim} N\left(p_A, \sqrt{\frac{p_A(1-p_A)}{n_A}}\right)$

de la même façon pour la pop B : $F_{n_B} = \overline{X}_{n_B} \underset{\infty}{\sim} N\left(p_B, \sqrt{\frac{p_B(1-p_B)}{n_B}}\right)$

Donc $F_{n_A} - F_{n_B} \underset{\infty}{\sim} N(\cdot,\cdot)$

$E(f_{n_A} - f_{n_B}) = p_A - p_B$

$V(f_{n_A} - f_{n_B}) = V(f_{n_A}) + V(f_{n_B})$ car les échantillons A et B sont indépendants

$$= \frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}$$

Ainsi : $F_{n_A} - F_{n_B} \underset{\infty}{\sim} N\left(p_A - p_B, \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}\right)$

on remplace l'écart type de F_n^A par $\sqrt{\frac{F_n^A(1-F_n^A)}{n_A}}$ idem pour F_n^B.

on centre et on réduit :

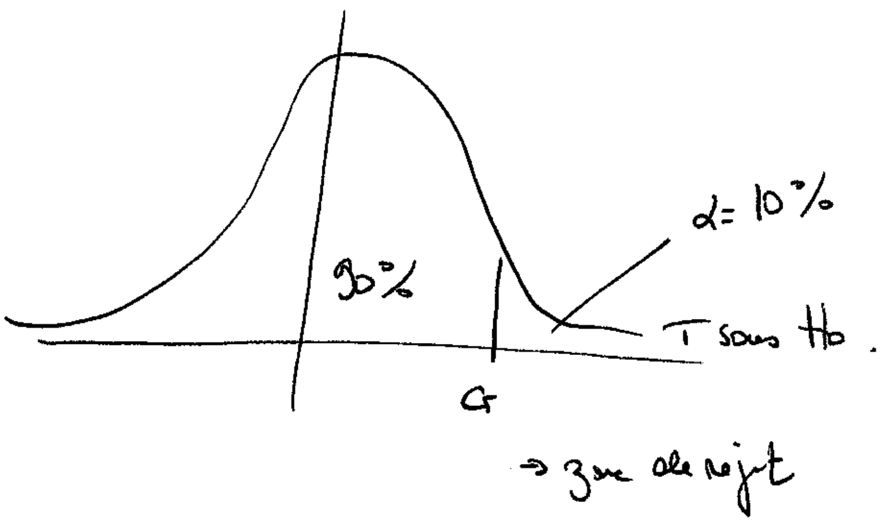
$$\frac{F_{n_A} - F_{n_B} - (p_A - p_B)}{\sqrt{\frac{F_{n_A}(1-F_{n_A})}{n_A} + \frac{F_{n_B}(1-F_{n_B})}{n_B}}} \underset{\infty}{\sim} N(0,1)$$

Test

statistique de test (sous H_0 : $p_A - p_B = 0$)

$$T = \frac{f_{m_A} - f_{m_B}}{\sqrt{\frac{f_{m_A}(1-f_{m_A})}{m_A} + \frac{f_{m_B}(1-f_{m_B})}{m_B}}} \underset{H_0}{\overset{\infty}{\sim}} N(0,1)$$

$c = \Phi^{-1}(1-\alpha) = \Phi^{-1}(0,9)$
$= 1,285$.



D'où la région critique

$$\boxed{W_r =]1,285 ; +\infty[}$$

réalisation de $t = \frac{0,2 - 0,16}{\underbrace{\sqrt{\frac{0,2 \times 0,8}{160} + \frac{0,16 \times 0,84}{4276}}}_{\simeq 0,032}} = 1,25$

$t \notin W_r$. Donc on ne rejette pas H_0

Le dispositif n'est pas efficace avec 10% de chances de se tromper.

② Si les conseillers sélectionnaient les personnes qu'ils préfèrent et si ces préférences sont reliées aux chances d'emploi (chômeurs par exemple) alors l'évaluation du dispositif serait biaisée (surévaluation de l'impact dans cet exemple).