Questions de Cours

① Aléatoire : fréquence empirique, erreur de première espèce
Non aléatoire : espérance μ

② La méthode du maximum de vraisemblance est une méthode de construction d'estimateur : fournit une forme fonctionnelle $f(X_1, \dots, X_n)$ où $X_1, \dots, X_n$ est l'échantillon.
Le principe est de chercher le(s) paramètre(s) qui maximisent la fonction de vraisemblance représentant la probabilité d'obtenir l'échantillon réalisé

Exercice 1

① La variable aléatoire parente $X \sim \mathcal{D}(\mu, \sigma = 5)$
$E(X) = \mu$ est inconnue on l'estime en utilisant l'estimateur moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ où $(X_1, \dots, X_n)$ constitue l'échantillon simple au hasard.
Dans notre cas, $n = 250$ Donc estimateur : $\underline{\bar{X}_{250}}$

② Un estimateur $f(X_1, \dots, X_n)$ est sans biais si $E(f(X_1, \dots, X_n)) = E(X_i)$. Dans notre cas :

$$E(\bar{X}_{250}) = E\left(\frac{1}{250}\sum_{i=1}^{250} X_i\right) = \frac{1}{250}\sum_{i=1}^{250} E(X_i) \quad \text{or } E(X_i) = \mu \ \forall i$$

Donc $E(\bar{X}_{250}) = \frac{1}{250}\sum \mu = \frac{250}{250}\mu = \mu$ $\quad \bar{X}_{250}$ $\underline{\text{est sans biais}}$

(3) $V(\overline{X}_{250}) = V\left(\frac{1}{250}\sum_{i=1}^{250} X_i\right) = \frac{1}{250^2} V\left(\sum_{i=1}^{250} X_i\right)$

Comme on a un échantillon simple au hasard, les X_i sont indépendants donc $V(\sum X_i) = \sum V(X_i)$ ainsi

$V(\overline{X}_{250}) = \frac{1}{250^2}\sum_{i=1}^{250} V(X_i)$ or $V(X_i) = \sigma^2 = 25 \quad \forall i$

$\underline{V(\overline{X}_{250})} = \frac{1}{250^2}\sum_{i=1}^{250} 25 = \frac{250 \times 25}{250^2} = \frac{25}{250} = \boxed{\frac{1}{10}}$

(4) Soit une suite iid $X_1, \dots, X_n$ où $X_i \sim \mathcal{D}(\mu, \sigma)$

le théorème central limite indique que :

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

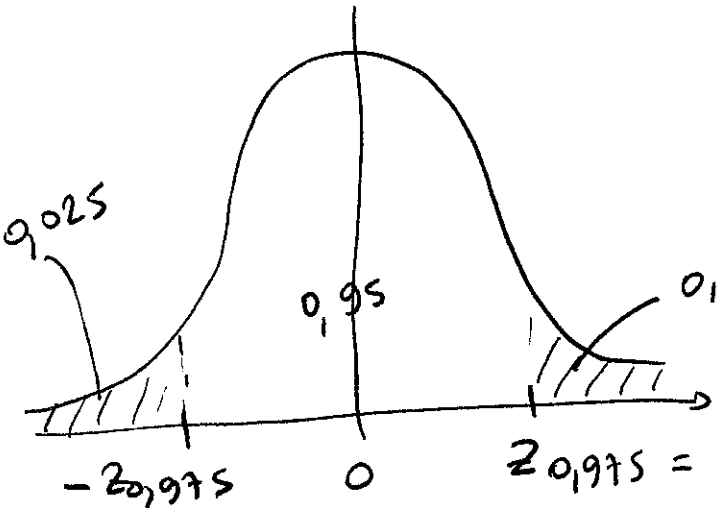
Ainsi, $\overline{X}_n \underset{\infty}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ Dans notre cas :

$\boxed{\overline{X}_{250} \underset{\infty}{\sim} \mathcal{N}\left(\mu, \frac{1}{\sqrt{10}}\right)}$ (c'est une distribution asymptotique)

(5) On cherche $IC_{95\%}(\mu) = [a, b]$ tels que

$P(a \leq \mu \leq b) = 0,95$

je centre et réduis $\overline{X}_{250}$: $Z = \frac{\overline{X}_{250} - \mu}{1/\sqrt{10}} \underset{\infty}{\sim} N(0,1)$



loi normale $N(0,1)$

distribution de Z

$P(Z < z_{0,975}) = 0,975$

$z_{0,975} = \Phi^{-1}(0,975)$

Ex 1 5 suite

on sait (ou table) que $z_{0,975} = 1,96$

Ainsi $P(-1,96 \leq Z \leq 1,96) = 0,95$

$$P\left(-1,96 \leq \frac{\overline{X}_{250} - \mu}{1/\sqrt{10}} \leq 1,96\right) = 0,95$$

$\times 1/\sqrt{10}$ $$P\left(-1,96 \times \frac{1}{\sqrt{10}} \leq \overline{X}_{250} - \mu \leq 1,96 \times \frac{1}{\sqrt{10}}\right) = 0,95$$

$-\overline{X}_{250} = -38,7$ $$P\left(-38,7 - 1,96 \times \frac{1}{\sqrt{10}} \leq -\mu \leq -38,7 + 1,96 \times \frac{1}{\sqrt{10}}\right) = 0,95$$

$\times(-1)$ $$P\Big(\underbrace{38,7 - 1,96 \times \frac{1}{\sqrt{10}}}_{a = 38,08} \leq \mu \leq \underbrace{38,7 + 1,96 \times \frac{1}{\sqrt{10}}}_{b = 39,32}\Big) = 0,95$$

$$\boxed{IC_{95\%}(\mu) = [38,08 ; 39,32]}$$

"Il y a 95% de chances pour que μ soit compris entre 38,08 et 38,856."

Exercice 2

(1) Il s'agit de vérifier que le nombre d'individus ciblés (dans l'échantillon) est compatible avec le processus générateur des cibles.

Formalisons :

Soit la variable aléatoire parente $X_i \left\langle \begin{array}{l} 1 \text{ l'individu } i \text{ est ciblé} \\ 0 \text{ l'individu } i \text{ n'est pas ciblé} \end{array} \right.$

$X_i \sim \mathcal{B}(p)$ où $p = 2\% = 0{,}02$
énoncé : processus informatique.

$n = 4436$ inscrits depuis moins de 3 mois ont été soumis à ce processus de ciblage.

~~1~~60 personnes ont été ciblées ce qui correspond à une fréquence empirique $F_{4436} = \frac{\cancel{1}60}{4436} = \cancel{10{,}37}\%$ environ.
$3{,}6\%$

→ Construisons l'intervalle de pari pour la fréquence empirique.

$$F_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

→ Distribution de F_n :

Par le théorème central limite X_i $i = 1, \dots, n$ est une suite i.i.d de v.a d'espérance μ d'écart type σ

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

ici $X_i \sim \mathcal{B}(p)$ $E(X_i) = \mu = p$
$V(X_i) = p(1-p) = \sigma^2 \Rightarrow \sigma = \sqrt{p(1-p)}$

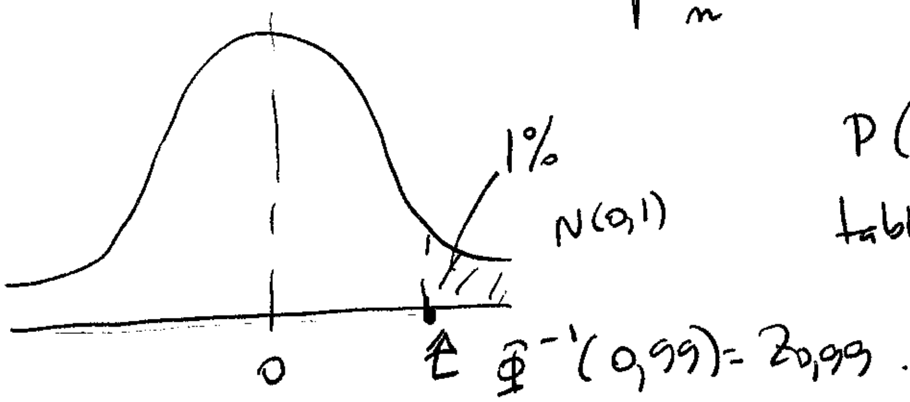
D'où $\sqrt{n}\,\dfrac{\bar{X}_n - p}{\sqrt{p(1-p)}} \underset{\infty}{\sim} N(0,1)$

Ainsi $\boxed{\bar{X}_n \underset{\infty}{\sim} N\left(p, \sqrt{\dfrac{p(1-p)}{n}}\right)}$

Les conseillers ne peuvent pas souhaiter des individus, on cherche l'intervalle de pari unilatéral de niveau 99 %

$$P(\bar{X}_n < b) = 0{,}99$$

on centre et on réduit : $Z = \dfrac{\bar{X}_n - p}{\sqrt{\dfrac{p(1-p)}{n}}} \underset{\infty}{\sim} N(0,1)$



$P(Z < z_{0,99}) = 0{,}99$

table : $z_{0,99} \simeq 2{,}33$ (ou 2,32).

Donc $P(Z < 2{,}33) = 0{,}99$

$$P\left(\dfrac{\bar{X}_n - p}{\sqrt{\dfrac{p(1-p)}{n}}} \leq 2{,}33\right) = 0{,}99$$

$\times\sqrt{\dfrac{p(1-p)}{n}}$

$$P\left(\bar{X}_n - p \leq 2{,}33\sqrt{\dfrac{p(1-p)}{n}}\right) = 0{,}99$$

$+p$

$$P\left(\bar{X}_n \leq \underbrace{p + 2{,}33\sqrt{\dfrac{p(1-p)}{n}}}_{b}\right) = 0{,}99$$

le processus aléatoire parent est connu (système automatique) E6

$p = 2\% = 0{,}02$

$n = 4436$

Ainsi $$P\Big(\overline{X}_n \leq \underbrace{0{,}02 + 2{,}33\sqrt{\frac{0{,}02 \times 0{,}98}{4436}}}_{b = 0{,}0249}\Big) = 0{,}99$$

Il y a 99% de chances pour que la moyenne empirique soit inférieure à 2,49 % si le processus aléatoire parent est bien $X_i \sim \mathcal{B}(0{,}02)$.

Or on obtient $\overline{X}_n = \frac{160}{4436} = 3{,}6\%$ en dehors de l'intervalle de pari. Il y avait moins de 1% de chances pour que cela arrive. On en conclut donc qu'il y a fraude avec 1% de chances de se tromper.

② Si les conseillers sélectionnent les personnes qu'ils préfèrent et si ces préférences sont reliées à leurs chances de retrouver un emploi (le charisme par exemple impliquerait un lien positif). Alors l'évaluation de l'effet du dispositif serait biaisée (surévaluation dans l'exemple donné).