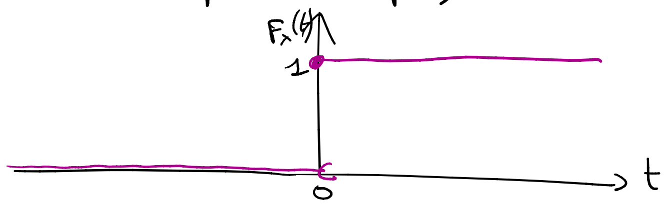
TD 9 - mercredi

mercredi 7 avril 2021 06:41

Exercice 1

1) $\forall t \in \mathbb{R} \quad F_X(t) = \mathbb{P}(X \leq t) \in [0,1]$. F_X est croissante, $\lim_{-\infty} F_X = 0$, $\lim_{+\infty} F_X = 1$.

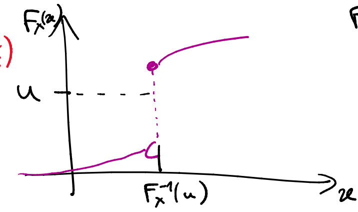
Soit X la variable aléatoire qui vaut 0 p.s., alors $F_X(t) = \mathbb{1}_{t \geq 0}$.

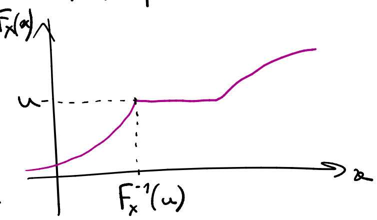


F_X n'est ni injective ni surjective.

Si $X \sim \mathcal{N}(0,1)$, alors F_X est bijective $\mathbb{R} \to]0,1[$ (F_X' est la densité de $\mathcal{N}(0,1)$, qui est >0)

2) $F_X^{-1}(u) = \inf\{x \in \mathbb{R} : F_X(x) \geq u\}$ (*)





Remarque Si $x \in \mathbb{R}$, $u \in (0,1)$ vérifient $F_X(x) \geq u$, alors $F_X^{-1}(u) \leq x$.

3 a) C'est faux, sinon tous les F_X seraient bijectifs, et on a un contre-exemple.

$F_X(x) \geq \underbrace{F_X(x)}_{u}$ donc $\boxed{F_X^{-1}(F_X(x)) \leq x}$

b) C'est faux (même contre-exemple : on a $F_X^{-1}(u) = \begin{cases} -\infty & \text{si } u < 0 \\ 0 & \text{si } u \geq 0 \end{cases}$, pas bijectif)

Par définition, $t = F^{-1}(q)$ est tel que $F(t) \geq u$

donc $\boxed{F_X(F_X^{-1}(q)) \geq q.}$

4) Voir cours (équivalence $q \leq F(x) \Leftrightarrow F^{-1}(q) \leq x$ donc $\mathbb{P}(F_X^{-1}(U) \leq x) = \mathbb{P}(U \leq F_X(x))$
$= F_X(x)$
$= \mathbb{P}(X \leq x)$)

5) Si la loi de X n'a pas d'atome, F_X est continue, donc tout $u \in (0,1)$ a un antécédent.

Soit $y \in (0,1)$. Il existe z t.q. $F_X(z) = y$. Prenons le plus grand tel z. Alors

$$\mathbb{P}(F_X(X) \leq y) = \mathbb{P}(F_X(X) \leq F_X(z)) = \mathbb{P}(X \leq z)$$

$X \leq z \Rightarrow F_X(X) \leq F_X(z)$ par croissance de F_X donc $\mathbb{P}(F_X(X) \leq F_X(z)) \geq \mathbb{P}(X \leq z)$

Pour le sens $F_X(X) \leq F_X(z) \Rightarrow X \leq z$, par contraposée : montrons que $X > z \Rightarrow F_X(X) > F_X(z)$. Cela vient de la définition de z.

$\hookrightarrow$ donc $\mathbb{P}(F_X(X) \leq F_X(z)) \leq \mathbb{P}(X \leq z)$

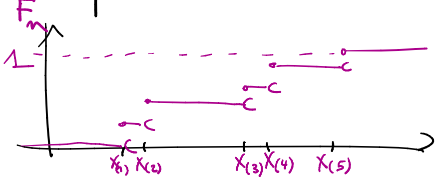
D'où l'égalité.

Donc $\mathbb{P}(F_X(X) \leq y) = \mathbb{P}(X \leq z) = F_X(z) \underset{\text{déf. de } z}{=} y$ et ce pour tout $y \in (0,1)$, donc $F_X(X) \sim \mathcal{U}([0,1])$

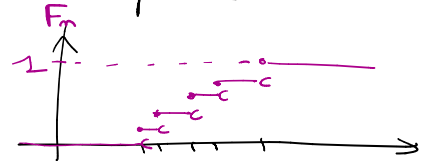
Exercice 2

Exercice 2

Expérience 1



Expérience 2



F_n = fonction de répartition empirique

F_n est une variable aléatoire à valeur dans l'ensemble des fonctions de répartition.

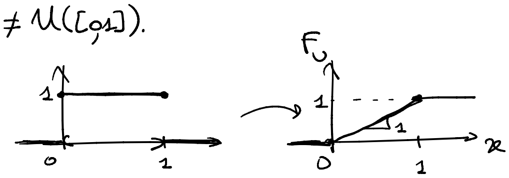
3) $$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \le x} = \frac{1}{n}\max\{i : X_{(i)} \le x\} = \begin{cases} 0 & \text{si } x < X_{(1)} \\ \frac{i}{n} & \text{si } x \in [X_{(i)}, X_{(i+1)}) \\ 1 & \text{si } x \geq X_{(n)} \end{cases}$$

Notation $X_{(1)} \le \dots \le X_{(n)}$ sont les statistiques d'ordre de l'échantillon $X_1, \dots, X_n$, c'est-à-dire l'échantillon trié par ordre croissant.

1) Formalisation du problème : les observations sont i.i.d. de loi P.

H_0 : $P = \mathcal{U}([0,1])$ $\qquad$ H_1 : $P \neq \mathcal{U}([0,1])$.

2) La densité de $\mathcal{U}([0,1])$ est $x \mapsto \mathbb{1}_{[0,1]}(x)$



donc $F_U(x) = \begin{cases} 0 & \text{si } x \le 0 \\ x & \text{si } x \in (0,1) \\ 1 & \text{si } x \geq 1 \end{cases}$.

4) Sous H_0, $\|F_n - F_U\|_\infty \xrightarrow[n\to+\infty]{} 0$ p.s.

5) Sous H_0, $\sqrt{n}\|F_n - F_U\|_\infty$ converge vers la loi de Kolmogorov K (= la loi de $\sup_{t\in[0,1]} |B(t)|$) $\quad$ B : pont brownien

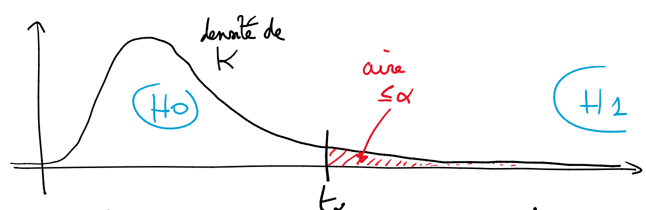
6) Si les observations sont iid de loi de fonction de répartition F^*, alors par le théorème de Glivenko-Cantelli, $\|F_n - F^*\|_\infty \xrightarrow[n\to+\infty]{} 0$

Comme par l'inégalité triangulaire, $\big| \|F_n - F_U\|_\infty - \|F^* - F_U\|_\infty \big| \le \|F_n - F^*\|_\infty$,

$$\|F_n - F_U\|_\infty \xrightarrow[n\to+\infty]{} \|F^* - F_U\|_\infty \underbrace{> 0}_{\text{car } F^* \neq F_U}$$

D'où $$\sqrt{n}\|F_n - F_U\|_\infty \xrightarrow[n\to+\infty]{} +\infty$$

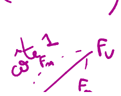
7)

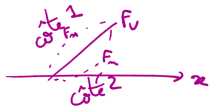


Région de rejet $R = [t_\alpha, +\infty[$. $\quad$ t_α est tel que le niveau du test est $\le \alpha$, c-à-d. $\mathbb{P}_{X\sim K}(X \ge t_\alpha) \le \alpha$: t_α est le quantile de niveau $1-\alpha$ de la loi de Kolmogorov (tableau A.14)

Ce tableau fournit les quantiles de la loi de $\|F_n - F_U\|$ ...

Two sided : $\max\left(\sup(F_n - F_U), -\inf(F_n - F_U)\right)$





Ce tableau fournit les quantiles de la loi de $\|F_n - F_U\|_\infty$, pas $\sqrt{n}\|F_n - F_U\|_\infty$!

8) Pour un test de niveau 5%, la région de rejet est $[0{,}294 ; +\infty[$.

9) $\|F_n - F_U\|_\infty = \max\limits_{1 \leq i \leq n} \left\{ \max\left(X_{(i)} - \frac{i}{n}, \; X_{(i)} - \frac{i-1}{n} \right) \right\}$.

Spoiler : le max est atteint pour $i=2$.

donc $\|F_n - F_U\|_\infty = 0{,}402$

10) $\|F_n - F_U\|_\infty \geq 0{,}294$ donc est dans la région de rejet.

donc on rejette H0 : l'échantillon ne suit pas la loi uniforme.

11)

$[0, \frac{1}{3}]$	$[\frac{1}{3}, \frac{2}{3}]$	$[\frac{2}{3}, 1]$
1	10	9

Test du χ^2 d'ajustement : $H0 : p = p_0$ vs $H1 : p \neq p_0$ où $p_0 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

$$T^{obs} = \frac{(1 - 20 \cdot \frac{1}{3})^2}{20 \cdot \frac{1}{3}} + \frac{(10 - 20 \cdot \frac{1}{3})^2}{20 \cdot \frac{1}{3}} + \frac{(9 - 20 \cdot \frac{1}{3})^2}{20 \cdot \frac{1}{3}} = 7{,}3 \quad , \quad T \sim \chi^2(2) \text{ sous H0}.$$

Exercice 3

$$h_{n,m}(X,Y) = \|F_n - G_m\|_\infty = \sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i \leq x} - \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}_{Y_j \leq x} \right|$$

Formalisation : les X_i, $1 \leq i \leq n$, sont iid de loi de fonction de répartition F.
les Y_j, $1 \leq j \leq m$, ——— G.

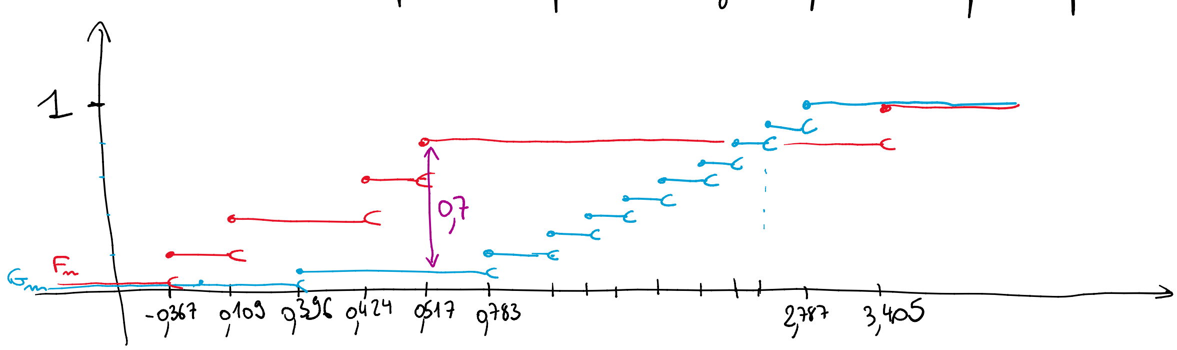
$H0 : F = G$ $\qquad$ $H1 : F \neq G$.

$$h_{n,m}(X,Y) = \sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i \leq x} - \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}_{Y_j \leq x} \right|$$

$$\overset{\text{sous H0}}{\sim} \sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{F^{-1}(U_i) \leq x} - \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}_{F^{-1}(V_j) \leq x} \right| \quad \text{avec } U_i, V_j \text{ iid} \sim \mathcal{U}([0,1]).$$

$$\overset{\text{sous H0}}{\sim} \sup_{q \in \mathrm{Im}(F)} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{U_i \leq q} - \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}_{V_j \leq q} \right|.$$

$=$ la statistique dont les quantiles sont fournis par la table pourvu que $\mathrm{Im}(F) = (0,1)$.



Donc $h_{n,m}^{obs} = 0{,}7$. $\qquad$ Région de rejet : $R = [t_\alpha, +\infty[$.

Quantile d'ordre 95% de $h_{5,10}$: table A.16 : $t_\alpha = 0{,}7$.

$h^{obs}_{n,m} \in R$ donc <u>on rejette H0</u> : les deux échantillons n'ont pas été générés par la même loi.

<u>Exercice 4</u> $\quad X_i \sim \mathcal{N}(\mu, \sigma^2)$, $Z_i = \frac{X_i - \overline{X}_n}{\hat{\sigma}_n}$, la loi de Z ne dépend pas de μ et σ.

2) $$D = \sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \leq x} - F(\mu+\sigma x) \right| = \sup_{x\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{Z_i \leq \frac{x-\overline{X}_n}{\hat{\sigma}_n}} - F^Z_n\left(\frac{x-\overline{X}_n}{\hat{\sigma}_n}\right) \right|$$

$$= \sup_{z\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{Z_i \leq z} - F^Z_n(z) \right|$$ ne dépend plus de μ et σ.

(Cf. cours pour tester H0 : $X \sim \mathcal{E}(\lambda)$ pour un $\lambda > 0$ vs. H1 : X ne suit pas une loi exponentielle).