$X \to f_\theta(x) = \frac{2}{\theta^2} x \mathbb{1}_{[0,\theta]}(x)$, $\theta > 0$.

Exo 4

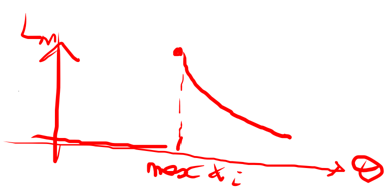
5) EMV $\quad L_n(\theta) = \prod_{i=1}^{n} \frac{2}{\theta^2} X_i \underbrace{\mathbb{1}_{[0,\theta]}(X_i)}$

[l'indicatrice dépend de θ
↳ on ne dérive pas.

. $\theta \mapsto L_n(\theta)$ prend 2 valeurs 0 ou $\underbrace{\frac{2^n}{\theta^{2n}} \prod_{i=1}^{n} X_i}$

quand $\mathbb{1}_{[0,\theta]}(X_i) = 1 \ \forall i$.

càd $\forall i, \theta \geq X_i \Leftrightarrow \theta \geq X_{(n)} = \max_i X_i$



. De plus $\theta \mapsto L_n(\theta)$ est décroissante sur $[\max_i X_i ; +\infty[$

Le maximum est donc atteint en $\max_i X_i$.

L'EMV est donc $\hat{\theta}^{MV} = \max_i X_i$.

6) Fdr de X

$$F_X(t) = \int_0^t f_X(x)\,dx = \int_0^t \frac{2}{\theta^2} x \,\mathbb{1}_{[0,\theta]}(x)\,dx$$

$= \mathbb{P}(X \leq t)$

$$= \begin{cases} 0 & \text{si } t \leq 0 \\ \frac{1}{\theta^2}\left[x^2\right]_0^t & \text{si } 0 \leq t \leq \theta \\ 1 & \text{si } t > \theta \end{cases}$$

$$= \begin{cases} 0 & \text{si } t < 0 \\ t^2/\theta^2 & \text{si } 0 \leq t \leq \theta \\ 1 & \text{si } t > \theta \end{cases}$$

7) Fdr de $X_{(m)} = \max X_i$

$\forall t \in [0, \theta]$,

$$F_{X_{(m)}}(t) = \mathbb{P}(\underbrace{\forall i,\ X_i \leq t}_{\max_i X_i \leq t}) \underset{\text{Var. indép.}}{=} \prod_{i=1}^{m} \mathbb{P}(X_i \leq t) \underset{X_i \text{ ident. distribuées}}{=} \mathbb{P}(X_1 \leq t)^m$$

$\mathbb{P}(\forall i,\ X_i \leq t) = \mathbb{P}\left(\bigcap_{i=1}^{m} X_i \leq t\right)$

$$= \frac{t^{2m}}{\theta^{2m}}$$

8) Densité de $X_{(n)}$:

$$\forall x \in \mathbb{R}, \quad f_{X_{(n)}}(x) = \frac{d \, \mathbb{P}(X_{(n)} \leq x)}{dx}$$

$$= \frac{2n}{\theta} \left(\frac{x}{\theta}\right)^{2n-1} \mathbb{1}_{[0,\theta]}(x)$$

9) Biais de l'estimateur. $\mathbb{E}[X_{(n)} - \theta]$? $= \mathbb{E}[X_{(n)}] - \theta$.

$$\mathbb{E}[X_{(n)}] = \int_0^\theta \frac{2n}{\theta} x \left(\frac{x}{\theta}\right)^{2n-1} dx = \int_{\mathbb{R}} x f_{X_{(n)}}(x) \, dx$$

$$= \int_0^\theta 2n \left(\frac{x}{\theta}\right)^{2n} dx = \left[\frac{2n}{2n+1} \frac{x^{2n+1}}{\theta^{2n}}\right]_0^\theta$$

$$= \boxed{\frac{2n}{2n+1} \theta}$$

↳ $X_{(n)}$ est biaisé.

10) Risque quadratique : $\text{Biais}^2(X_{(n)}) + \text{Var}(X_{(n)})$

$$= \mathbb{E}\left[(X_{(n)} - \theta)^2\right] = \underbrace{\mathbb{E}\left[X_{(n)}^2\right]}_{\text{on doit calculer le moment d'ordre 2}} - 2\theta\,\mathbb{E}\left[X_{(n)}\right] + \theta^2 .$$

$$\mathbb{E}\left[X_{(n)}^2\right] = \int_0^\theta \frac{2n}{\theta} x^2 \left(\frac{x}{\theta}\right)^{2n-1} dx = \int_{\mathbb{R}} x^2 f_{X_{(n)}}(x)\, dx$$

$$= \theta \int_0^\theta 2n \left(\frac{x}{\theta}\right)^{2n+1} dx$$

$$= \theta \left[\frac{2n}{2n+2} \frac{x^{2n+2}}{\theta^{2n+1}}\right]_0^\theta = \theta \frac{2n}{2n+2} \theta . = \frac{2n}{2n+2}\theta^2 .$$

$$\mathbb{E}\left[(X_{(n)}-\theta)^2\right] = \frac{2n}{2n+2}\theta^2 - 2\,\frac{2n}{2n+1}\theta^2 + \theta^2$$

$$= \frac{n}{n+1}\theta^2 - \frac{4n}{2n+1}\theta^2 + \theta^2$$

$$= \frac{n(2n+1) - 4n(n+1) + (n+1)(2n+1)}{(n+1)(2n+1)}\theta^2$$

$$= \frac{\cancel{2n^2} - \cancel{4n^2} + \cancel{2n^2} + \cancel{n} - \cancel{4n} + \cancel{n} + \cancel{2n} + 1}{(n+1)(2n+1)}\theta^2$$

$$= \frac{\theta^2}{(n+1)(2n+1)}$$

11) Convergence en loi de $n(\theta - X_{(n)})$.

$$\left\{\begin{array}{l} Z_n \xrightarrow[n\to+\infty]{\mathcal{L}} Z \\ F_{Z_n}(z) \xrightarrow[n\to+\infty]{} F_Z(z) \quad \forall z. \end{array}\right.$$

Astuces

$\boxed{\begin{array}{l}\text{DL} \\ \bullet \ \log(1+u) \underset{u\sim 0}{\sim} u\end{array}}$

$\boxed{\bullet \ u^v = \exp v \log u}$

$$\mathbb{P}(n(\theta - X_{(n)}) \leq t) = \mathbb{P}\left(-X_{(n)} \leq \frac{t}{n} - \theta\right)$$

$$= \mathbb{P}\left(X_{(n)} \geq \theta - \frac{t}{n}\right).$$

$$= 1 - \mathbb{P}\left(X_{(n)} \leq \theta - \frac{t}{n}\right)$$

$$= 1 - \underbrace{\frac{\left(\theta - \frac{t}{n}\right)^{2n}}{\theta^{2n}}}_{\text{Forme exp !}} = 1 - \left(1 - \frac{t}{n\theta}\right)^{2n}$$

$$= 1 - \exp\left(2n \log\left(1 - \frac{t}{n\theta}\right)\right)$$

$$\underset{n\to+\infty}{\sim} 1 - \exp\left(\frac{-2n\,t}{n\theta}\right) = 1 - \exp\left(-\frac{2t}{\theta}\right)$$

$$= 1 - \exp\left(-\frac{2t}{\theta}\right).$$

↳ $n(\theta - X_{(n)})$ converge en loi vers une loi exp de param $\frac{2}{\theta}$.

(→ Q12 ~~exercice~~ exercice 3 / Q4 QCM)

similaire)

12) Intervalle de confiance au niveau $1-\alpha$ pour θ.

- IC non asymptotique → on ne peut pas utiliser la 11).
  (on a une loi asymptotique)

- On utilise : $X_{(n)} \leqslant \theta$ p.s.

- On cherche c_α tq $\mathbb{P}(\theta \in [X_{(n)} ; X_{(n)} + c_\alpha]) = 1-\alpha$

c_α, $\mathbb{P}(\theta \leqslant X_{(n)} + c_\alpha) = 1-\alpha$

$\Leftrightarrow \mathbb{P}(\theta - c_\alpha \leqslant X_{(n)}) = 1-\alpha$

$\Leftrightarrow 1 - \mathbb{P}(X_{(n)} \leqslant \theta - c_\alpha) = 1-\alpha$

$\Leftrightarrow \mathbb{P}(X_{(n)} \leqslant \theta - c_\alpha) = \alpha$

$$F_{X_{(m)}}\left(\theta - c_\alpha\right) = \alpha .$$

$$\iff \quad \frac{(\theta - c_\alpha)^{2m}}{\theta^{2m}} = \alpha .$$

$$\iff \quad \left(1 - \frac{c_\alpha}{\theta}\right)^{2m} = \alpha .$$

$$\iff \quad \exp 2m \log\left(1 - \frac{c_\alpha}{\theta}\right) = \alpha$$

$$\iff \quad \log\left(1 - \frac{c_\alpha}{\theta}\right) = \frac{\log \alpha}{2m} .$$

$$\iff \quad 1 - \frac{c_\alpha}{\theta} = \exp\left(\frac{\log \alpha}{2m}\right)$$

$$\iff \quad c_\alpha = \left(1 - \exp\left(\frac{\log \alpha}{2m}\right)\right)\theta .$$

ne pas poser : quantile d'une loi qu'on ne connait pas.

$$\underset{\text{(intermédiaire)}}{IC} \overset{\text{de } \theta}{=} \left[X_{(n)} \,;\, \cancel{X}_{(n)} + \left(1 - \exp\left(\frac{\ln \alpha}{2n}\right)\right)\theta \right]$$

θ inconnu

$$\mathbb{P}\Big(\underbrace{X_{(n)} \leq \theta}_{=1 \text{ p.s.}} \leq X_{(n)} + \left(1 - \exp\left(\frac{\ln \alpha}{2n}\right)\right)\theta\Big) = 1-\alpha.$$

$$= \mathbb{P}\left(\theta \leq X_{(n)} + \left(1 - \exp\left(\frac{\ln \alpha}{2n}\right)\right)\theta\right)$$

$$= \mathbb{P}\left(\theta \exp\left(\frac{\ln \alpha}{2n}\right) \leq X_{(n)}\right) = \mathbb{P}\left(\theta \leq X_{(n)} \exp\left(-\frac{\ln \alpha}{2n}\right)\right)$$

$$IC = \left[X_{(n)} \,;\, X_{(n)} \exp\left(-\frac{\ln \alpha}{2n}\right) \right].$$

13) Comparaison

Q4) → IC asymptotique ($n \to +\infty$)

Q12) → IC non asymptotique ($\forall n$)
n fixé

↳ On choisit l'IC de Q12) car il sera plus fiable (pour tout nombre d'obs fixé)

Exercice 1 TD5

"Grand truth" 5.2.2.

On a une valeur que l'on considère vraie

- quantité d'androgènes / ℓ de sang

3,22 , ...

- H_0 : « les athlètes de RDA ne sont dopées ».

1) Test à utiliser ?

Test d'inégalité des moyennes (slide 13)

H_0 : « $\underbrace{\mu_A}_{\text{moyenne de quantité d'androgène chez les athlètes de RDA}} \leq \underbrace{3.1}_{\text{quantité d'androgène chez une femme lambda}}$ » vs H_1 : « $\mu_A > 3.1$ ».

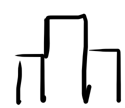
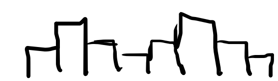
2) (Slide 13 « Contexte »)

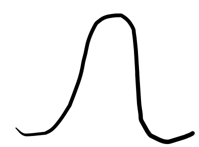
Hypothèses pour le test :

- $X_1, \dots, X_n$ (les données) doivent être i.i.d
- $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ les données sont Gaussiennes.

3) Commenter un histogramme → hypothèses à vérifier

↳ Valider la distribution Gaussienne

- symétrique
- 1 mode →  (si on avait  2 modes)

Gaussienne  1 mode + symétrique

↳ L'hypothèse Gaussienne semble cohérente, l'histogramme est symétrique et n'a qu'un mode.

(↳ De plus, le nombre de données est faible, donc les hypothèses sont difficiles à vérifier)

(↳ indépendance ⇝ protocole permet de valider l'indépendance des données.)

4) Rédaction importante (+++)

. On teste au risque 5% :

H_0 : " $\mu_A \leq 3,1$ " contre H_1 : " $\mu_A > 3,1$ "

. Sous H_0, il existe $\mu_A' \leq 3,1$ tel que :

$$\underbrace{Z(\mu_A') = \underbrace{\sqrt{n}\,\frac{\overline{X}_n - \mu_A'}{S_n}}_{\text{statistique de test}} \sim T_{n-1}}_{\text{démontré dans Chap Cochran (Chap 4)}}$$

, avec T_{n-1} la loi de Student de degré $n-1$

$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, X_i v.a. (associées aux données)

↳ $x_i = 3,22$, ...

$S_n = \sqrt{\text{variance empirique}}$

$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$

. On obtient la région de rejet :

$$ZR = \left\{ z(\mu_A') > t_{n-1, 1-\alpha} \right\},$$

où $t_{n-1, 1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi de Student de degrés $n-1$.

. Ici la sortie du test R nous donne une p-valeur de 0,0146, qui est inférieure à 0,05 donc on rejette H_0 (au niveau 5%)

Réduction :

1) Rappeler le test
2) Donner la stat. de test (cours par ♡)
3) Donner la région de rejet (cours par ♡)
4) Conclusion

- $\boxed{Z_{obs}} > \underbrace{t_{n-1, 1-\alpha}}_{\text{quantile de la Student de degré } n-1}$ (cours théo)

$\Leftrightarrow$

- $\underbrace{F(Z_{obs})}_{} > \underbrace{F(t_{n-1, 1-\alpha})}_{= 1-\alpha}$, ($\underline{R}$)

avec F fdR de la Student de degré n-1

$\Leftrightarrow 1 - \mathbb{P}(T \geq Z_{obs}) > 1 - \alpha$

$\Leftrightarrow 1 - \text{pvalue} > 1 - \alpha$

$\Leftrightarrow \boxed{\text{pvalue} \leq \alpha}$

(pas à savoir, pas à écrire)

6) On a comparé la moyenne d'androgène de femme sportives avec celle de femme "λ" (potentiellement pas sportive)

On ne répond donc pas à la question du dopage, on répond juste à la question est-ce que une femme sportive a plus d'androgène qu'une femme non sportive.