

3.3 Tests statistiques

Démarche scientifique

Toute **démarche scientifique** s'effectue selon le même schéma. Afin d'analyser la plausibilité d'une théorie, on itère les étapes suivantes :

- Enoncé d'une hypothèse (théorie) pouvant être contredite par des données.
- Récolte de données
- Comparaison des données avec les prédictions/implications de l'hypothèse.
- Non-rejet, rejet ou modification éventuelle de l'hypothèse.

Dans un cadre statistique, en supposant que l'on dispose d'un modèle pour le phénomène étudié, on itère les étapes suivantes :

- Enoncé d'une hypothèse (typiquement sur les paramètres du **modèle statistique**). Cette hypothèse peut être contredite par des données (via une statistique, appelée **statistique de test**).
- Récolte de données
- **Rejet (ou non) de l'hypothèse** à partir de la comparaison entre les données et les implications de l'hypothèse. En cas d'écart, à partir de quel seuil juge-t-on cet écart **significatif**, i.e., suffisamment important pour justifier le rejet de l'hypothèse ?

Exemple

Exemple Question : L'alcool ralentit-il les réflexes ?

Afin d'étudier l'effet de l'alcool sur les réflexes, on fait passer à 14 sujets un test de dextérité avant et après qu'ils aient consommé 100 ml de vin. Leurs temps de réaction (en ms) avant et après sont donnés dans le tableau suivant :

Sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Avant	57	54	62	64	71	65	70	75	68	70	77	74	80	83
Après	55	60	68	69	70	73	74	74	75	76	76	78	81	90

Cadre statistique : [1] Hypothèse nulle et alternative

Etant donné un modèle statistique (de densité $f(x; \theta)$), nous voulons choisir entre deux théories concurrentes à propos du paramètre θ . Ces dernières forment une paire d'hypothèses :

H_0 : l'hypothèse nulle vs H_1 : l'hypothèse alternative.

Exemple. Dans une population décrite par la loi $\mathcal{N}(\mu, \sigma^2)$, nous pouvons former des hypothèses sur μ comme suit :

$$\underbrace{\left\{ \begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{array} \right\}}_{\text{paire bilatérale}} \quad \text{ou} \quad \underbrace{\left\{ \begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{array} \right\} \quad \text{ou} \quad \left\{ \begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{array} \right\}}_{\text{paires unilatérales}}.$$

Cadre statistique : [2] Statistique de test

Comment choisir entre les deux hypothèses ?

- Nous tirons un échantillon $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} f(x; \theta)$ tiré de la population.
Comment l'utiliser pour prendre notre décision ?
- Nous choisissons une statistique $T = T_n = g(X_1, \dots, X_n)$ qui a tendance à prendre des valeurs “typiques” sous l'hypothèse nulle H_0 (i.e., si H_0 est vraie) et “extrêmes” (dans la direction de l'hypothèse alternative H_1) sous H_1
- Ainsi, si on observe une valeur plutôt “extrême” (“extrême” dans la direction de l'hypothèse alternative H_1) de T , nous avons de l'évidence contre H_0 .

Notre règle de décision est donc :

- Rejeter H_0 si la valeur observée de T est **assez extrême** (au-delà d'une **valeur critique** à déterminer).
- Ne pas rejeter H_0 si la valeur observée de T n'est **pas assez extrême**.

Exemple : paire bilatérale

Soient $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, où σ^2 est inconnue, et considérons la paire d'hypothèses :

$$\left\{ \begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{array} \right\}.$$

On parle de paire bilatérale car $\mu \neq \mu_0$ est équivalent à $\mu < \mu_0$ ou $\mu > \mu_0$.

Considérons la statistique de test $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$.

- Si H_0 est vraie, alors $T \sim t_{n-1}$ (donc si H_0 est vraie, T prend typiquement des valeurs proches de 0).
- Compte tenu de H_1 , nous considérons donc les valeurs de T comme "extrêmes" si elles sont "éloignées" de 0. Notons qu'ici, la notion d'"extrême" dans la direction de l'hypothèse alternative H_1 signifie une valeur "extrême" de la valeur absolue de T .
- Nous allons rejeter H_0 si $|T|$ est **suffisamment élevée**, i.e., $|T| > v^*$, où $v^* > 0$ est une valeur critique à déterminer.

Exemple : paire unilatérale

Soient $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, où σ^2 est inconnu, et considérons la paire d'hypothèses :

$$\left\{ \begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{array} \right\}.$$

Considérons la statistique de test $T_n = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$.

- Si H_0 est vraie, alors $T \sim t_{n-1}$.
- Compte tenu de H_1 , nous considérons donc les valeurs de T comme “extrêmes” si elles sont fortement négatives. Donc ici, la notion d’“extrême” dans la direction de l’hypothèse alternative H_1 signifie une valeur “extrême” de $|\min(T, 0)|$ et non de $|T|$.
- Nous allons donc rejeter H_0 si T est **suffisamment négative**, i.e., $T < v_*$, où $v_* < 0$ est la valeur critique à déterminer.

Cadre statistique : [3] Significativité statistique

Choix de la valeur critique (par exemple v^* et v_*) : Comment définir **suffisamment élevée** ou **suffisamment négative**. En d'autres termes, quelle ampleur est considérée comme **significative** ?

Pour répondre à cette question, il faut considérer les deux types d'erreurs que l'on peut commettre lorsque l'on se décide en faveur de l'une des hypothèses :

Décision \ Vérité	H_0 vraie	H_0 fausse
Non-rejet de H_0	😊	Erreur de type II
Rejet de H_0	Erreur de type I	😊

Erreur de type I (**faux positif**) considérée plus grave que l'erreur de type II (**faux négatif**) — filtre de spam

Cadre statistique : [3] Significativité statistique

- On ne peut pas contrôler les deux erreurs à la fois :

Pr(erreur type I) petite



rejet uniquement pour des valeurs très extrêmes



difficile de rejeter



Pr(erreur type II) grande

- L'asymétrie entre la gravité des erreurs nous aide à choisir H_0 et H_1
- On va contrôler l'erreur de type I, qui est plus grave
- Parfois on choisit H_0 par convenance mathématique (de sorte que mathématiquement il serait plus facile de contrôler l'erreur de type I)

Cadre statistique : [3] Significativité statistique

- Nous choisissons la valeur maximale que l'on tolère pour la probabilité d'erreur de type I (éventuellement en tenant compte de l'avis d'un spécialiste). Cette quantité est notée α et appelée **niveau/seuil de significativité du test** ; $\alpha \in (0, 1)$. On choisit généralement une valeur faible pour α . Typiquement, $\alpha = 0.1, 0.05, 0.01, 0.001$; le plus souvent, $\alpha = 0.05$.

- La valeur critique est déterminée de manière à ce que

$$\Pr_{H_0}[\text{Rejet de } H_0] = \alpha.$$

- Ainsi, la **valeur critique** est telle que

$$\Pr_{H_0}[|T| > \text{valeur critique}] = \alpha \quad (\text{cas bilatéral}),$$

$$\Pr_{H_0}[T < \text{valeur critique}] = \alpha \quad \text{ou}$$

$$\Pr_{H_0}[T > \text{valeur critique}] = \alpha \quad (\text{cas unilatéral}).$$

- Les probabilités sont sous l'hypothèse que H_0 **est vraie** !

Cadre statistique : [3] Significativité statistique

Exemple, paire bilatérale : Soient $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, où σ^2 est inconnu, et considérons la paire $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$.

Nous allons rejeter H_0 si $|T| = \left| \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \right|$ est assez large, c'est à dire $|T| > v^*$.

Soit α le niveau de significativité. La valeur critique v^* satisfait

$$\Pr_{H_0}[|T| > v^*] = \alpha,$$

i.e.,

$$\Pr_{H_0}[T < -v^* \text{ ou } T > v^*] = \alpha.$$

Quand H_0 est vraie $T \sim t_{n-1}$. On doit donc choisir

$$v^* = t_{n-1, 1-\alpha/2},$$

où $t_{n-1, 1-\alpha/2}$ est le $(1 - \alpha/2)$ quantile de la loi de Student t_{n-1} .

Cadre statistique : [4] La p -valeur

Au lieu d'utiliser des valeurs critiques pour choisir entre H_0 et H_1 , nous pouvons utiliser une autre approche, basée sur la notion de p -valeur.

- La p -valeur (notée p_{obs}) est la probabilité d'obtenir une valeur de la statistique de test au moins aussi élevée (élevée dans la direction de H_1) que celle que nous avons observée si H_0 était vraie.
- Supposons que la réalisation de la statistique de test sur nos données est $T = t_{\text{obs}}$. Alors :
 - Cas bilatéral : $p_{\text{obs}} = \Pr_{H_0}[|T| > |t_{\text{obs}}|]$,
 - Cas unilatéral à gauche : $p_{\text{obs}} = \Pr_{H_0}[T < t_{\text{obs}}]$,
 - Cas unilatéral à droite : $p_{\text{obs}} = \Pr_{H_0}[T > t_{\text{obs}}]$.
- Petite valeurs de p_{obs} s'opposent à H_0 car elles démontrent que la réalité observée serait très improbable si l'hypothèse nulle H_0 était vraie.
- Cas bilatéral (203) : $p_{\text{obs}} = 2(1 - F_{t_{n-1}}(|t_{\text{obs}}|))$, où $F_{t_{n-1}}$ est la fonction de répartition de la loi de Student t_{n-1} .

Cadre statistique : [4] La p -valeur

On peut utiliser la p -valeur pour faire un test d'hypothèse :

$$\boxed{\text{rejetter } H_0 \iff p_{obs} < \alpha}$$

Exemple bilatérale (203) $p_{obs} = 2(1 - F_{t_{n-1}}(t_{obs}))$ donc

$$p_{obs} < \alpha \iff F_{t_{n-1}}(|t_{obs}|) > 1 - \alpha/2 \iff |t_{obs}| > t_{n-1, 1-\alpha/2}$$

De manière générale, l'approche de la p -valeur est équivalente à l'approche des valeurs critiques. Cependant, la p -valeur p_{obs} fournit une information plus facilement interprétable que la valeur t_{obs} . Il s'agit d'une mesure de l'évidence contre H_0 contenue dans les données.

Attention : la p -valeur **n'est pas** la probabilité que H_0 soit vraie.

Résumé : les éléments d'un test

- A Une **hypothèse nulle** H_0 à tester contre une hypothèse alternative H_1 .
- B Une **statistique de test** T , choisie de telle sorte que des valeurs "extrêmes" de T (en direction de H_1) suggèrent que H_0 est fautive. La valeur observée de T est t_{obs} .
- C Un **niveau de significativité** α , qui la probabilité d'erreur de type I (rejet de H_0 quand H_0 est vraie) maximale que nous allons tolérer.
- D1 Des **valeurs critiques**, telles que quand T tombe au-delà de ces valeurs, nous rejetons H_0 en faveur de H_1 . Les valeurs critiques sont choisies pour respecter le niveau de significativité α .

Au lieu de D1, nous pouvons utiliser l'approche équivalente D2 :
- D2 Une **valeur** p_{obs} donnant la probabilité d'observer une valeur de T aussi élevée que t_{obs} sous H_0 . On rejette alors H_0 en faveur de H_1 quand $p_{\text{obs}} < \alpha$.

Exemple

Exemple On a contrôlé 10 compteurs d'électricité nouvellement fabriqués et obtenu les valeurs suivantes (en MW) :

983 1002 998 996 1002 983 994 991 1005 986.

On suppose qu'il s'agit de réalisation d'un échantillon iid d'une loi normale. On aimerait savoir s'il y a un écart entre la moyenne attendue de 1000 MW et la moyenne réelle des compteurs qui sortent de la fabrication. Nous avons obtenu $\bar{x} = 994 < 1000$. S'agit-il d'un hasard ou une faute de production ?

On va prendre $\alpha = 5\%$.

Solution Exemple 208

Supposons que nos observations x_1, \dots, x_n soient des réalisations de variables aléatoires $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, avec σ^2 inconnu. On veut tester : $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$, où $\mu_0 = 1000$. On prend comme statistique de test

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \text{ sous } H_0 : \mu = \mu_0.$$

Dans notre cas $n = 10$, $\mu_0 = 1000$, $\bar{x} = 994$, et

$$s^2 = \frac{1}{9} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{9} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) = 64.88,$$

donc $t_{\text{obs}} = -2.35$.

On rejette H_0 si et seulement si $|t_{\text{obs}}| > t_{n-1, 1-\alpha/2}$ (cas bilatéral). , $t_{n-1, 1-\alpha/2} = 2.262$ (voir les tables), et comme $t_{\text{obs}} = -2.35 < -2.262$, on rejette l'hypothèse H_0 .

Intervalles de confiance et tests

- Soient $Y_1, \dots, Y_n \sim \mathcal{N}(\mu, \sigma^2)$ avec μ, σ inconnus
- Considérons $H_0 : \mu = 0$ et $H_1 : \mu \neq 0$
- On rejette H_0 au niveau α si et seulement si

$$|T_n| = \left| \sqrt{n} \frac{\bar{Y}_n}{S_n} \right| > t_{n-1, 1-\alpha/2}$$

- Intervalle de confiance (IC) de niveau $1 - \alpha$

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

- Cet intervalle contient zéro si et seulement si

$$|\bar{Y}_n| \leq t_{n-1, 1-\alpha/2} S_n / \sqrt{n} \text{ et donc}$$

on rejette $H_0 \iff 0$ n'est pas dans l'intervalle de confiance

- De manière générale, rejet de $H_0^{\theta_0} : \theta = \theta_0$ est équivalent à {l'IC ne contient pas θ_0 } si on se base sur la même statistique de test
- α petit \iff difficile à rejeter \iff IC de niveau $1 - \alpha$ large

- Rejet $\iff p_{obs} \leq \alpha \iff \theta_0 \notin \text{IC de niveau } 1 - \alpha$
- **IC** : α fixé, pour quels $\theta_0 \in \mathbb{R}$ $H_0^{\theta_0}$ n'est pas rejetée ?
- **p-valeur** : θ_0 fixé, pour quels $\alpha \in (0, 1)$ $H_0^{\theta_0}$ est rejetée ?

3.4 Tests khi-deux

Le test khi-deux

- On se pose la question de l'adéquation d'une distribution théorique à des données
- Supposons que dans une expérience on observe n résultats différents avec des
 - **fréquences observées** dans k classes disjointes o_1, \dots, o_k , alors que les
 - **fréquences théoriques** correspondantes sont e_1, \dots, e_k ,
 - où $\sum_{i=1}^k o_i = \sum_{i=1}^k e_i = n$
- On a H_0 : "les observations proviennent de la loi théorique spécifiée"
- Une mesure de l'écart entre les o_j et les e_j est donnée par la **statistique khi-deux**

$$T_n = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$$

- Si n est grand et les e_i ne sont pas trop petites (règle de pouce : $e_i \geq 5$ pour la plupart), alors $T_n \stackrel{\text{app}}{\sim} \chi_\nu^2$ sous H_0 , où
 - χ_ν^2 est la loi **khi-carré**, la loi de $\sum_{i=1}^\nu Z_i^2$ où $Z_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$
 - $\nu = k - 1$ si les e_i peuvent être calculés sans devoir estimer des paramètres inconnus
 - $\nu = k - 1 - c$ si les e_i sont calculés après avoir estimé c paramètres

Exemple

$$\text{rejet de } H_0 \text{ au niveau } \alpha \iff t_{\text{obs}} = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} > \chi_{\nu, 1-\alpha}^2$$

Exemple $n = 60$ jets d'un dé ont donné la répartition suivante :

Valeur x_i	1	2	3	4	5	6	
Valeur o_i	8	10	9	16	13	4	60

Ici $k = 6$ et H_0 : "équilibre du dé" est équivalente au modèle

$$\Pr(X = x) = 1/6, \quad x \in \{1, \dots, 6\}.$$

Exemple

Exemple L'intelligence (QI) X de $n = 1000$ personnes est testée :

Score X	[0, 70)	[70, 85)	[85, 100)	[100, 115)	[115, 130)	[130, ∞)
Nombre o_i	34	114	360	344	120	28

Est-il plausible que $X \sim \mathcal{N}(100, 15^2)$?

On a

$$\begin{aligned}e_i &= n \Pr_{\mathcal{N}(100, 15^2)}(a_i \leq X \leq b_i) = n \Pr \left(\frac{a_i - 100}{15} \leq \frac{X_i - 100}{15} \leq \frac{b_i - 100}{15} \right) \\ &= n \Phi \left(\frac{b_i - 100}{15} \right) - n \Phi \left(\frac{a_i - 100}{15} \right)\end{aligned}$$

Tableaux de contingence

Un **tableau de contingence** est une classification de n objets ou individus selon plusieurs critères

- Une question fondamentale concerne **l'indépendance** des critères
- Supposons qu'on observe deux caractères A (h classes) et B (k classes) sur chacun de n individus, donnant le **tableau de contingence** suivant :

	B						
A	1	2	...	j	...	k	Σ
1	n_{11}	n_{12}	...	n_{1j}	...	n_{1k}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2j}	...	n_{2k}	$n_{2.}$
...
i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ik}	$n_{i.}$
...
h	n_{h1}	n_{h2}	...	n_{hj}	...	n_{hk}	$n_{h.}$
Σ	$n_{.1}$	$n_{.2}$...	$n_{.j}$...	$n_{.k}$	$n_{..} = n$

Indépendance

- Soit n_{ij} le nombre de personnes tombant dans la classe i du caractère A et dans la classe j du caractère B , et soit

$$n_{i.} = \sum_{j=1}^k n_{ij}, \quad n_{.j} = \sum_{i=1}^h n_{ij}, \quad i \in \{1, \dots, h\}, j \in \{1, \dots, k\}$$

- On désire tester H_0 : **A et B sont indépendants**. Dans ce cas

$$\Pr(A = i, B = j) = \Pr(A = i) \times \Pr(B = j), \quad i \in \{1, \dots, h\}, j \in \{1, \dots, k\},$$

et les probabilités empiriques sont

$$\Pr(\widehat{A = i}) = \frac{n_{i.}}{n}, \quad \Pr(\widehat{B = j}) = \frac{n_{.j}}{n}$$

Ainsi, sous H_0 , l' (i, j) ième élément du tableau de contingence est

$$E_{ij} = n \times \Pr(A = i, B = j) = n \times \Pr(A = i) \times \Pr(B = j)$$

qu'on va estimer par

$$e_{ij} = n \frac{n_{i.}}{n} \frac{n_{.j}}{n} = \frac{n_{i.} n_{.j}}{n}$$

Calcul sous H_0

- Sous H_0 et pour n grand, la statistique T_n suit une distribution χ^2_ν avec $\nu = (h-1)(k-1)$, car on a dû estimer $c = (h-1) + (k-1)$ probabilités, et
$$hk - 1 - c = kh - 1 - (h-1) - (k-1) = (h-1)(k-1)$$
- Pour tester à un niveau de significativité α fixé, on rejette H_0 si $t_{\text{obs}} > \chi^2_{(h-1)(k-1), 1-\alpha}$, sinon on ne rejette pas H_0

Exemple On a relevé parmi 95 personnes la couleur de leurs yeux (caractère A) et celle de leurs cheveux (caractère B) et on a obtenu les résultats suivants :

A	B		
	Cheveux clairs	Cheveux sombres	
Yeux bleus	$n_{11} = 32$	$n_{12} = 12$	$n_{1.} = 44$
Yeux bruns	$n_{21} = 14$	$n_{22} = 22$	$n_{2.} = 36$
Autres	$n_{31} = 6$	$n_{32} = 9$	$n_{3.} = 15$
Σ	$n_{.1} = 52$	$n_{.2} = 43$	$n = 95$

On désire tester si la couleur des cheveux est indépendante de celle des yeux

Donc on a H_0 : indépendance entre couleur des cheveux et couleur des yeux

Régularité (non-examinable)

Les conditions de régularité sont compliquées. Elles sont fausses le plus souvent quand

- un des paramètres est discret
- le support de $f(y; \theta)$ dépend de θ
- le vrai θ se trouve sur une borne des valeurs possibles

Elles sont satisfaites dans la grande majorité des cas rencontrés en pratique

Exemple Soient $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} U(0, \theta)$, trouver la vraisemblance $L(\theta)$ et le $\hat{\theta}_{\text{ML}}$. Montrer que la loi limite de $n(\theta - \hat{\theta}_{\text{ML}})/\theta$ quand $n \rightarrow \infty$ est $\exp(1)$.
Discuter.

Preuve (non-examinable)

- Les fonctions de densité et de répartition de y_j sont

$$f(y; \theta) = \theta^{-1} I(0 \leq y \leq \theta), \quad F(y) = y/\theta, \quad 0 \leq y \leq \theta.$$

- L'indépendance donne

$$L(\theta) = \prod \theta^{-1} I(Y_j < \theta) = \theta^{-n} I(\max Y_j \leq \theta), \quad \theta > 0,$$

qui est maximisée au point $\hat{\theta}_{\text{ML}} = M_n = \max Y_j$

- On a

$$\Pr(M_n \leq x) = \prod_{j=1}^n \Pr(Y_j \leq x) = (x/\theta)^n, \quad 0 \leq x \leq \theta$$

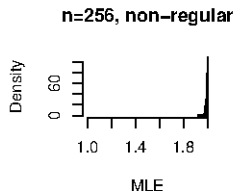
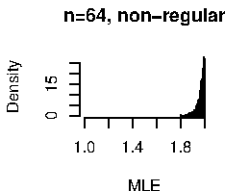
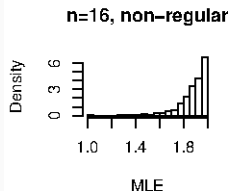
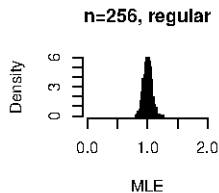
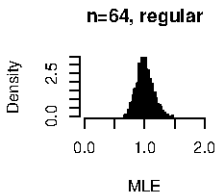
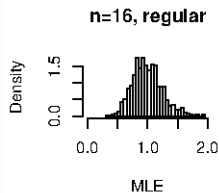
- Donc pour $x \geq 1$ et $n \geq x$,

$$\begin{aligned} \Pr \left\{ n(\theta - \hat{\theta}_{\text{ML}})/\theta \leq x \right\} &= \Pr(\hat{\theta}_{\text{ML}} \geq \theta - x\theta/n) = 1 - \{(\theta - x\theta/n)/\theta\}^n \\ &\rightarrow 1 - \exp(-x), \end{aligned}$$

comme requis

Exemple (non examinable)

Comparaison des lois de $\hat{\theta}$ dans un cas régulier (en haut, avec écart-type $\propto n^{-1/2}$ et loi limite normale) et dans un cas non-régulier (en bas, avec écart-type $\propto n^{-1}$ et loi limite non-normale)



4. Régression

4.1 Introduction