Retour sur le maximum de vraisemblance

- Soient $x_1, \ldots, x_n$ des données supposées être une réalisation d'un échantillon aléatoire $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x;\theta)$ (densité / masse). La **vraisemblance** pour θ est la fonction

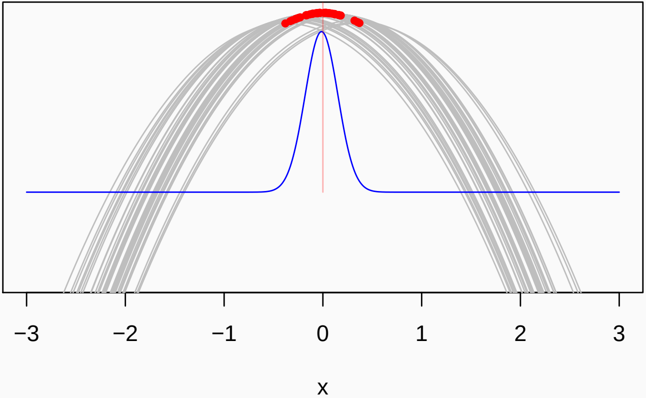
$$L(\theta) = f(x_1;\theta) \times f(x_2;\theta) \times \cdots \times f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

- $\widehat{\theta}_{\mathrm{ML}}$ satisfait

$$L(\widehat{\theta}_{\mathrm{ML}}) \geq L(\theta) \quad \text{pour tout } \theta$$

- **Interprétation** : dans le cas discret, on maximise $\mathrm{Pr}_\theta(X_1 = x_1, \ldots, X_n = x_n)$
- Il est plus facile de maximiser $\ell(\theta) := \log L(\theta)$, souvent en résolvant $d\ell(\theta)/d\theta = 0$, et vérifiant qu'il s'agit bien d'un maximum

La vraisemblance est une fonction aléatoire



Fonctions de vraisemblances correspondantes à 50 échantillons de taille $n = 40$. La vraie valeur est $\theta = 0$ et les 50 maxima sont en rouge.

Distribution asymptotique/approximative de $\widehat{\theta}_{\mathrm{ML}}$

- Dans des modèles "réguliers" (normal, exponentiel, Poisson, ...) on a $\ell'(\widehat{\theta}_{\mathrm{ML}}) = 0$
- Comme $\ell(\theta)$ est une somme de variables aléatoires iid, le théorème central limite s'applique
- Grâce à la méthode delta $\widehat{\theta}_{\mathrm{ML}} \overset{\mathrm{app}}{\sim} \mathcal{N}(\theta, 1/I_n(\theta))$ avec **l'information de Fisher**

  $$I_n(\theta) = -\mathbb{E}_\theta(\ell''(\theta)) = \mathbb{E}_\theta(J_n(\theta)), \qquad J_n(\theta) = -\ell''(\theta)$$

  $1/I_n(\theta)$ est la **variance asymptotique** de $\widehat{\theta}_{\mathrm{ML}}$, $I_n(\theta)$ est la courbure
- **L'information observée** est $J_n(\widehat{\theta}_{\mathrm{ML}})$
- **Attention !** Certains modèles, tel que $U[0, \theta]$, ne sont pas réguliers

Exemple $x_1, \ldots, x_n$ réalisations d'une loi $\exp(\lambda)$ avec densité $\lambda e^{-\lambda x} I(x \geq 0)$, $\lambda > 0$

Exemple

Exemple $x_1, \ldots, x_n$ réalisations d'une loi $\exp(\lambda)$ avec densité $\lambda e^{-\lambda x} I(x \geq 0)$, $\lambda > 0$

3.2 Estimation par intervalle

Les intervalles de confiance

Un élément clé de la statistique est de donner une idée de l'incertitude d'un constat

Soit θ un paramètre inconnu, et soit $\tilde{\theta}_n = 1$ une estimation de θ basée sur $y_1, \ldots, y_n$:

- si $n = 10^5$ on est beaucoup plus sûr que $\theta \approx \tilde{\theta}_n$ que si $n = 10$
- pour exprimer ceci on aimerait donner un intervalle qui serait plus large quand $n = 10$ que quand $n = 10^5$, pour expliciter l'incertitude liée à $\tilde{\theta}_n$

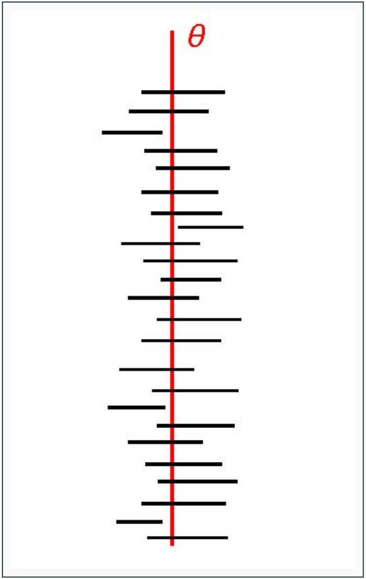
Définition: Soient $Y \equiv Y_1, \ldots, Y_n$ des données issues d'une loi F de paramètre $\theta \in \mathbb{R}$. Un **intervalle de confiance** (IC, 'confidence interval' en anglais) (L_n, U_n) pour θ est une statistique sous forme d'intervalle qui contient θ avec une probabilité spécifiée $1 - \alpha$

$$\mathrm{Pr}_\theta(L_n \leq \theta \leq U_n) = 1 - \alpha \qquad \forall\theta$$

- $1 - \alpha \in (0, 1)$ est le **niveau**, souvent $\alpha \in \{0.05, 0.01, 0.1\}$
- Les bornes $L_n = L_n(Y), U_n = U_n(Y)$ sont des statistiques et non pas des inconnus
- Un IC **bilatéral**, de la forme (L_n, U_n), est le plus souvent utilisé
- Un IC **unilatéral à droite** est $(-\infty, U_n]$ tel que $\mathrm{Pr}_\theta(U_n \geq \theta) = 1 - \alpha$
- Un IC **unilatéral à gauche** est $[L_n, \infty)$ tel que $\mathrm{Pr}_\theta(L_n \leq \theta) = 1 - \alpha$

Interprétation d'un intervalle de confiance

- (L_n, U_n) est un intervalle aléatoire qui contient θ avec probabilité $(1-\alpha)$
- Si on répète l'expérience avec d'autres données, on aura un autre intervalle de confiance
- Nous ne savons pas si notre IC contient θ, mais cette procédure nous fournit une garantie statistique que cet événement a une probabilité $(1-\alpha)$



Construire un IC dans le cas normal

- Soient $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 1)$, et

$$\overline{Y}_n = \frac{1}{n} \sum_{j=1}^{n} Y_j$$

la moyenne de l'échantillon. On a $\overline{Y}_n \sim \mathcal{N}(\mu, 1/n)$ et donc

$$Z_n = n^{1/2}(\overline{Y}_n - \mu) \sim \mathcal{N}(0, 1)$$

a une distribution qui ne dépend pas de μ. On obtient

$$\Pr_\mu \left(\overline{Y}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}} \leq \mu \leq \overline{Y}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}} \right) = 1 - \alpha$$

avec $z_\beta = \Phi^{-1}(\beta)$ les quantiles de la loi $\mathcal{N}(0, 1)$

- Si $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, et **on connaît** σ^2, alors

$$\frac{Z_n}{\sigma} = \frac{n^{1/2}(\overline{Y}_n - \mu)}{\sigma} \sim \mathcal{N}(0, 1)$$

et l'intervalle de confiance pour μ est

$$\left[\overline{Y}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}} \sigma, \overline{Y}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}} \sigma \right]$$

- La largeur $\downarrow$ avec n, $\uparrow$ avec σ et avec le niveau $(1-\alpha)$, car $\Pr_\theta(L_n \leq \theta \leq U_n) \uparrow$

Exemple

Exemple On suppose que la résistance Y d'un certain type d'équipements électriques est distribuée selon une loi normale avec $\sigma = 0.12$ ohm.

Un échantillon de taille $n = 64$ a donné comme moyenne la valeur $\bar{y}_n = 5.34$ ohm.

Trouver un intervalle de confiance pour μ au niveau 95%.

Intervalle de Student

- Soient $Y_1, \dots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ avec σ^2 **inconnue**
- $\left[\overline{Y}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}}\sigma, \overline{Y}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}}\sigma\right]$ n'est plus un intervalle de confiance
- On suppose $n > 1$ et estime σ^2 par

$$S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \overline{Y}_n)^2$$

- On remplace σ par S_n : soit

$$T_n = \frac{Z_n}{S_n} = n^{1/2}\frac{\overline{Y}_n - \mu}{S_n} = \frac{Z_n/\sigma}{\sqrt{S_n^2/\sigma^2}}$$

- Nominateur et dénominateur indép, leurs lois ne dépendent pas de (μ, σ^2)
- T_n suit une **loi de Student avec $n-1$ degrés de liberté**, $T_n \sim t_{n-1}$, qui ne dépend de $\theta = (\mu, \sigma^2)$. C'est une loi symétrique, on dénote les quantiles par $t_{n-1,\beta}$
- L'intervalle de confiance qui en résulte est

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

- La largeur $\downarrow$ avec n, $\uparrow$ avec S_n et avec le niveau $(1-\alpha)$

Exemple

Exemple Pour déterminer le point de fusion μ d'un certain alliage, on a procédé à $n = 9$ observations qui ont donné une moyenne $\overline{y}_n = 1040^\circ$ avec $s_n = 16^\circ$. On suppose que les données suivent une loi normale.

Trouver un intervalle de confiance pour μ à niveau 95%.

Intervalles de confiance pour μ dans le cas normal : résumé

Soient $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ avec μ inconnu

- Pour σ **connue** $n^{1/2}\frac{\overline{Y}_n - \mu}{\sigma} \sim \mathcal{N}(0,1)$ et on a les IC de niveau $1-\alpha$

$$\left[\overline{Y}_n - \frac{z_{1-\alpha/2}}{\sqrt{n}}\sigma, \overline{Y}_n + \frac{z_{1-\alpha/2}}{\sqrt{n}}\sigma\right]$$

$$\left[\overline{Y}_n - \frac{z_{1-\alpha}}{\sqrt{n}}\sigma, \infty\right] \qquad \left[-\infty, \overline{Y}_n + \frac{z_{1-\alpha}}{\sqrt{n}}\sigma\right]$$

  avec $z_\beta = \Phi^{-1}(\beta)$ les quantiles de la loi $\mathcal{N}(0,1)$

- Pour σ **inconnue** (et $n > 1$) $n^{1/2}\frac{\overline{Y}_n - \mu}{S_n} \sim t_{n-1}$ et on a les IC de niveau $1-\alpha$

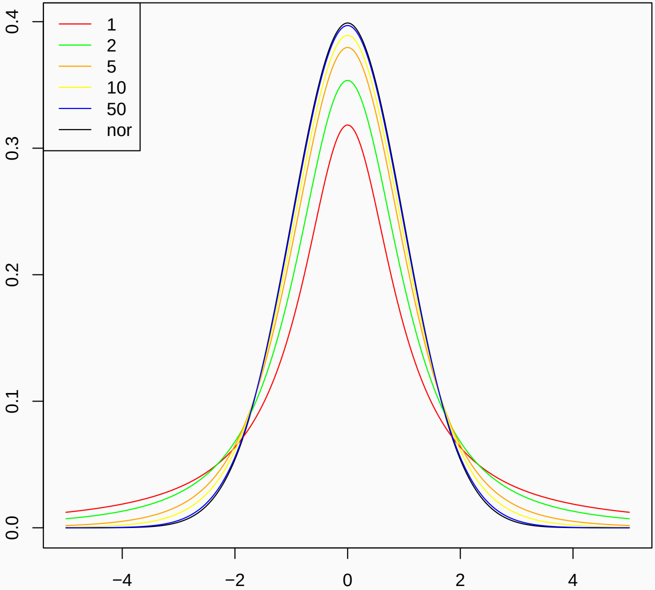
$$\left[\overline{Y}_n - \frac{t_{n-1,1-\alpha/2}}{\sqrt{n}}S_n, \overline{Y}_n + \frac{t_{n-1,1-\alpha/2}}{\sqrt{n}}S_n\right]$$

$$\left[\overline{Y}_n - \frac{t_{n-1,1-\alpha}}{\sqrt{n}}S_n, \infty\right] \qquad \text{ou} \qquad \left[-\infty, \overline{Y}_n + \frac{t_{n-1,1-\alpha}}{\sqrt{n}}S_n\right]$$

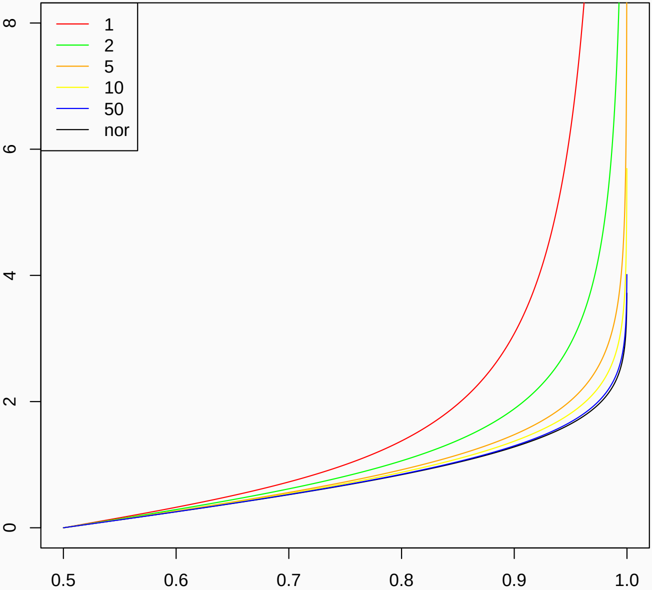
  avec t_{n-1} la **loi de Student avec $n-1$ degrés de liberté**, $t_{n-1,\beta}$ sont les quantiles de cette distribution, S_n définie à la diapositive 182.

t_{n-1} symétrique comme $\mathcal{N}(0,1)$, mais quantiles plus grands : $|t_{n-1,\beta}| > |z_\beta|$, $\beta \neq \frac{1}{2}$

Densités de t_k et $\mathcal{N}(0,1)$



Quantiles de t_k et $\mathcal{N}(0,1)$



Pivots

- Une fonction $T(Y_1, \ldots, Y_n, \theta)$ dont la loi est connue et ne dépend pas de θ s'appelle un **pivot**
- **Attention !** Ce n'est pas une statistique car c'est une fonction de θ
- Si $a \leq b$,

prob $= \alpha_1$ F_T prob $= \alpha_2$

$$\alpha_1 = \Pr(T < a), \quad \alpha_2 = \Pr(T > b), \quad \alpha_1 + \alpha_2 = \alpha \in [0, 1]$$

a b

(souvent $\alpha_1 = \alpha_2 = \alpha/2$) alors

$$\Pr_\theta(a \leq T \leq b) = \Pr_\theta(T \leq b) - \Pr_\theta(T < a) = (1 - \alpha_2) - \alpha_1 = 1 - \alpha$$

- Si on peut isoler θ, on peut trouver des variables aléatoires L et U telles que

$$\Pr_\theta(L \leq \theta \leq U) = 1 - \alpha$$

Exemples T_n (diapositive 182), Z_n (diapositive 179).

$\sqrt{n}(\bar{Y}_n - \mu)/\sigma \sim N(0,1)$

Exemple Soient $Y_1, \dots, Y_n \overset{ind}{\sim} U(0,\theta)$, $M_n = \max(Y_i)$. Alors $T(Y_1, \dots, Y_n, \theta) = M_n/\theta$ est un pivot.

$$\Pr_\theta(T \le t) = \Pr_\theta(M_n \le t\theta)$$

$$= \Pr_\theta(Y_1 \le t\theta, \dots, Y_n \le t\theta)$$

$$\overset{ind}{=} \prod_{i=1}^{n} \Pr_\theta(Y_i \le t\theta) = \left(\Pr_\theta(Y_1 \le t\theta)\right)^n$$

$$= \begin{cases} 1 & \text{si } t > 1 \\ 0 & \text{si } t < 0 \\ t^n = \left(\frac{t\theta}{\theta}\right)^n & \text{si } t \in [0,1] \end{cases}$$

ne dépend pas de θ $\Rightarrow$ T pivot

Intervalles de confiance approximatifs

- En pratique la plupart des intervalles de confiance se basent sur des approximations fournies par le théorème central limite, étant de la forme
$$(L_n, U_n) = (\widehat{\theta}_n - \sqrt{V_n} z_{1-\alpha/2}, \widehat{\theta}_n + \sqrt{V_n} z_{1-\alpha/2}),$$
où V_n est une estimateur de $\mathrm{var}_\theta(\widehat{\theta}_n)$ dont la racine s'appelle **erreur type (standard error)** de $\widehat{\theta}_n$. Sa réalisation $v_n^{1/2}$ est aussi appelée erreur type
- L'intervalle de confiance est approximatif dans le sens que
$$\mathrm{Pr}_\theta(L_n \leq \theta \leq U_n) \to 1 - \alpha, \qquad n \to \infty$$
- Dans des modèles réguliers, la variance asymptotique de $\widehat{\theta}_{\mathrm{ML}}$ (voir diapositive 174) est $1/I_n(\theta)$, estimée par $1/J_n(\widehat{\theta}_{\mathrm{ML}})$ et donc
$$(L_n, U_n) = \left(\widehat{\theta}_{\mathrm{ML}} - z_{1-\alpha/2}/\sqrt{J_n(\widehat{\theta}_{\mathrm{ML}})}, \widehat{\theta}_{\mathrm{ML}} + z_{1-\alpha/2}/\sqrt{J_n(\widehat{\theta}_{\mathrm{ML}})}\right)$$

Exemple Soient $Y_1, \dots, Y_n \overset{\text{iid}}{\sim} \mathrm{Poiss}(\lambda)$, $\lambda > 0$ inconnu. Trouver une erreur type pour $\widehat{\lambda}_{\mathrm{ML}}$, et ainsi donner un intervalle de confiance approximatif de niveau 90% pour λ. $\alpha = 0{,}1 \to 1 - \frac{\alpha}{2} = 0{,}95$

$$L(\lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{y_i}}{y_i!} = e^{-\lambda n} \lambda^{n\overline{Y}_n} \prod_{i=1}^{n} \frac{1}{y_i!}$$

$$\ell(\lambda) = -n\lambda + n\overline{Y}_n \log\lambda + \sum_{i=1}^{n} \log\frac{1}{y_i!}$$

$$\ell'(\lambda) = -n + \frac{n\overline{Y}_n}{\lambda} = 0 \Leftrightarrow \lambda = \overline{Y}_n$$

$$\ell''(\lambda) = -\frac{n\overline{Y}_n}{\lambda^2} < 0 \Rightarrow \hat{\lambda}_{ML} = \overline{Y}_n$$

$$J_n(\lambda) = -\ell''(\lambda) = \frac{n\overline{Y}_n}{\lambda^2}$$

$$J_n(\hat{\lambda}_{ML}) = \frac{n\overline{Y}_n}{\hat{\lambda}_{ML}^2} = \frac{n}{\overline{Y}_n}$$

IC approximatif $\left[\overline{Y}_n - z_{0.95}\sqrt{\frac{\overline{Y}_n}{n}}, \overline{Y}_n + z_{0.95}\sqrt{\frac{\overline{Y}_n}{n}}\right]$

Exemple Soient $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} U(0,\theta)$ avec θ inconnu et $\overline{Y}_n = n^{-1}\sum_{j=1}^n Y_j$. Utiliser le théorème central limite pour trouver un intervalle de confiance approximatif de niveau 95% pour θ.

$$\mathbb{E}_\theta Y_i = \frac{\theta}{2} \qquad \mathrm{Var}_\theta Y_i = \frac{\theta^2}{12}$$

$$T_n = \sqrt{n}\,\frac{\overline{Y}_n - \frac{\theta}{2}}{\sqrt{\frac{\theta^2}{12}}} \overset{app}{\sim} \mathcal{N}(0,1)$$

pivot approximatif

$$T_n = \sqrt{12n}\left(\frac{\overline{Y}_n}{\theta} - \frac{1}{2}\right)$$

$$1-\alpha \approx P_\theta\left(-z_{1-\frac{\alpha}{2}} \le T_n \le z_{1-\frac{\alpha}{2}}\right) = P_\theta\left(\frac{1}{2} - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{12n}} \le \frac{\overline{Y}_n}{\theta} \le \frac{1}{2} + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{12n}}\right)$$

$$\Rightarrow IC \left[\frac{\overline{Y}_n}{\frac{1}{2} + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{12n}}}, \frac{\overline{Y}_n}{\frac{1}{2} - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{12n}}}\right]$$

approximatif

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 ?

hypothèse : oui

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
X	Avant	57	54	62	64	71	65	70	75	68	70	77	74	80	83
Y	Après	55	60	68	69	70	73	74	74	75	76	76	78	81	90
		-2	6	6	5	-1			-1						

$$D = Y - X \qquad \bar{d}_n = 3,5 > 0$$

$$W = \begin{cases} 1 & \text{si } D > 0 \\ 0 & \text{sinon} \end{cases} \qquad \overline{w}_n = \frac{10}{14}$$

si l'alcool n'a pas d'effet

$\bar{d}_n \approx 0$

$\overline{w}_n \approx \frac{1}{2}$

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 :

$$H_0 : \text{l'hypothèse } \underline{\text{nulle}} \qquad \text{vs} \qquad H_1 : \text{l'hypothèse } \underline{\text{alternative}}.$$

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

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

Comment choisir entre les deux hypothèses ?

- Nous tirons un échantillon $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x;\theta)$ tiré de la population. Comment l'utiliser pour prendre notre décision ?
- Nous choisissons une statistique $T = T_n = g(X_1, ..., X_n)$ 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.

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \rightarrow \text{extrême} = \text{grandes valeurs}$$

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).

négatives ou positives

- Ne pas rejeter H_0 si la valeur observée de T n'est pas assez extrême.

On n'accepte pas H_0

Exemple : paire bilatérale

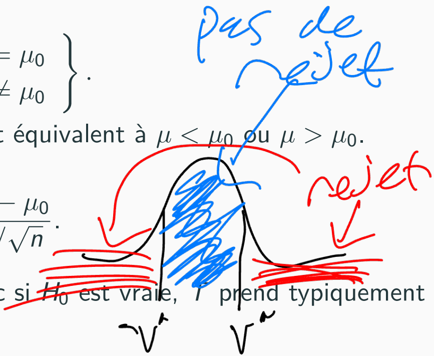
Soient $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, où σ^2 est inconnue, et considérons la paire d'hypothèses :

μ_0 connue

$$\begin{Bmatrix} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{Bmatrix}.$$

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 $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$.



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

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

pas de rejet

rejet

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

$V_n \neq -V^*$

- 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 :

$$\text{Pr(erreur type I) petite}$$
$$\Updownarrow$$
$$\text{rejet uniquement pour des valeurs très extrêmes}$$
$$\Updownarrow$$
$$\text{difficile de rejetter}$$
$$\Updownarrow$$
$$\text{Pr(erreur type II) grande}$$

- 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 convenience mathématique (de sorte que mathématiquement il serait plus facile de contrôler l'erreur de type I)

H_0 ! $\neq H_0$ compliqué

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, $\alpha = 0.05$.

  le même α de l'IC $(1-\alpha)$

- La valeur critique est déterminée de manière à ce que

$$\mathrm{Pr}_{H_0}[\text{Rejet de } H_0] = \alpha.$$

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

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

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

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

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

Cadre statistique : [3] Significativité statistique

Exemple, paire bilatérale : Soient $X_1, ..., X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, 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

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

$= Pr(T > v^*) + Pr(T < -v^*)$

i.e.,

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

symétrie

$= 2Pr(T < -v^*)$

$= \alpha$

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 $T = t_{\mathrm{obs}}$. Alors :

  bilatéral

  - Cas bilatéral : $p_{\mathrm{obs}} = \mathrm{Pr}_{H_0}[|T| > |t_{\mathrm{obs}}|]$,
  - Cas unilatéral à gauche : $p_{\mathrm{obs}} = \mathrm{Pr}_{H_0}[T < t_{\mathrm{obs}}]$,
  - Cas unilatéral à droite : $p_{\mathrm{obs}} = \mathrm{Pr}_{H_0}[T > t_{\mathrm{obs}}]$.
- Petite valeurs de p_{obs} s'opposent à H_0 car elles démontrent que la realité observée serait très improbable si l'hypothèse nulle H_0 était vraie.
- Cas bilatéral (203) : $p_{obs} = 2(1 - F_{t_{n-1}}(|t_{obs}|))$, où $F_{t_{n-1}}$ est la fonction de répartition de la loi de Student t_{n-1}.

t_{obs} est aléatoire $\Longrightarrow$ P_{obs} aléatoire

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

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

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 fausse. 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 $x_1, \ldots, x_n$ soient des réalisations de variables aléatoires $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, 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 $Y_1, \ldots, Y_n \sim \mathcal{N}(\mu, \sigma^2)$ 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

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

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

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

- Cet intervalle contient zéro si et seulement si $|\overline{Y}_n| \leq t_{n-1,1-\alpha/2} S_n / \sqrt{n}$ et donc

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

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

Intervalles de confiance (IC) et tests

- Rejet $\iff p_{obs} \leq \alpha \iff \theta_0 \notin$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 $o_1, \ldots, o_k$, alors que les
  - **fréquences théoriques** correspondantes sont $e_1, \ldots, e_k$,
  - 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 $T_n \overset{\text{app}}{\sim} \chi^2_\nu$ sous H_0, où
  - χ^2_ν est la loi **khi-carré**, la loi de $\sum_{i=1}^{\nu} Z_i^2$ où $Z_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$
  - $\nu = k - 1$ 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

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

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

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

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

$$e_i = n\mathrm{Pr}_{\mathcal{N}(100,15^2)}(a_i \leq X \leq b_i) = n\mathrm{Pr}\left(\frac{a_i - 100}{15} \leq \frac{X_i - 100}{15} \leq \frac{b_i - 100}{15}\right)$$

$$= n\Phi\left(\frac{b_i - 100}{15}\right) - n\Phi\left(\frac{a_i - 100}{15}\right)$$

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	$\cdots$	j	$\ldots$	k	Σ
1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1k}	$n_{1\cdot}$
2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2k}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{ik}	$n_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
h	n_{h1}	n_{h2}	$\cdots$	n_{hj}	$\cdots$	n_{hk}	$n_{h\cdot}$
Σ	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot k}$	$n_{\cdot\cdot} = n$

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

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

- 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

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

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

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

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 :

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

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 $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} U(0,\theta)$, trouver la vraisemblance $L(\theta)$ et le $\widehat{\theta}_{\mathrm{ML}}$. Montrer que la loi limite de $n(\theta - \widehat{\theta}_{\mathrm{ML}})/\theta$ quand $n \to \infty$ est $\exp(1)$. Discuter.

Preuve (non-examinable)

- Les fonctions de densité et de répartition de y_j sont

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

- L'indépendance donne

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

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

- On a

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

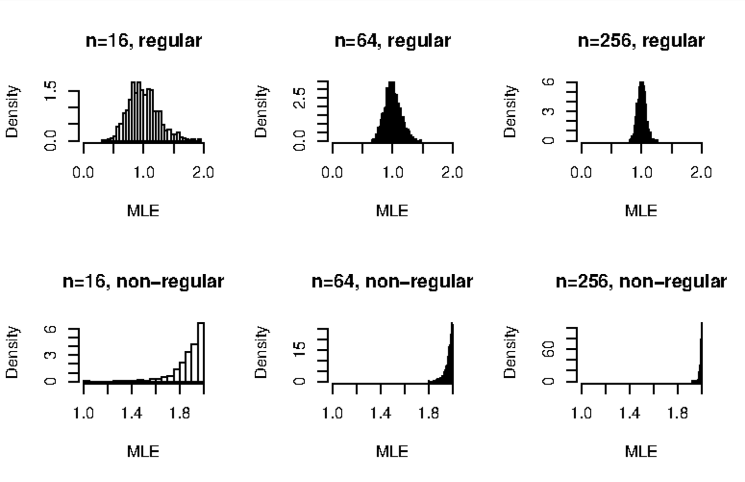
- Donc pour $x \ge 1$ et $n \ge x$,

$$\mathrm{Pr}\left\{ n(\theta - \widehat{\theta}_{\mathrm{ML}})/\theta \le x \right\} = \mathrm{Pr}(\widehat{\theta}_{\mathrm{ML}} \ge \theta - x\theta/n) = 1 - \{(\theta - x\theta/n)/\theta\}^n$$
$$\to 1 - \exp(-x),$$

comme requis

Exemple (non examinable)

Comparaison des lois de $\widehat{\theta}$ 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