Cours 8 – Estimateur par intervalle et Intervalle de confiance.

Eya ZOUGAR *

Institut National des Sciences appliqués-INSA

Génie mathématiques GM3
Wednesday 22nd March, 2023

INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ROUEN NORMANDIE

*Basé sur le cours de Bruno PORTIER

1. Introduction.
1.1. Le cadre.

On dispose de n données réelles $x_1, x_2, \dots, x_n$ qui sont les mesures d'une variable quantitative.

Hypothèse: Ces données sont en fait les réalisations de n variables aléatoires $X_1, X_2, \dots, X_n$ que l'on suppose indépendantes et de même loi F.

On s'intéresse à une caractéristique de cette loi F (espérance, variance, etc ...) ou bien à un paramètre de cette loi, en supposant que cette loi soit paramétrée.

Notons θ cette caractéristique ou ce paramètre, supposé inconnu.

Soit T_n un estimateur de ce paramètre inconnu, construit à partir des variables aléatoires $X_1, X_2, \dots, X_n$.

But: On souhaite pouvoir quantifier la confiance que l'on peut accorder à l'estimation du paramètre θ.

Comment faire cela ? Construire un intervalle de confiance.

1.2. Intervalle de confiance.

Définition

Un intervalle de confiance pour le paramètre θ au niveau de confiance $(1-\alpha)$ où $\alpha \in]0,1[$, est un intervalle, q'on le note $\mathrm{IC}_{1-\alpha}(\theta)$, de la forme:

$$\mathrm{IC}_{1-\alpha}(\theta) = [b_1(x_1,\cdots,x_n)\,;\, b_2(x_1,\cdots,x_n)]$$

avec

$$\mathbb{P}\Big[\theta \in [b_1(X_1,\cdots,X_n)\,;\, b_2(X_1,\cdots,X_n)]\Big] \;=\; 1-\alpha$$

- L'intervalle aléatoire $[b_1(X_1,\cdots,X_n)\,;\, b_2(X_1,\cdots,X_n)]$ est un estimateur par intervalle du paramètre θ au niveau de confiance $(1-\alpha)$.
- L'intervalle de confiance est donc la réalisation sur les données de l'estimateur par intervalle.
- Mais attention, cela ne signifie pas que θ appartienne à $\mathrm{IC}_{1-\alpha}(\theta)$. En fait, on ne saura pas. On affirmera que $\theta \in \mathrm{IC}_{1-\alpha}(\theta)$ en prenant le risque α de se tromper.

1.3. Estimateur par intervalle.

- La construction d'un intervalle de confiance au niveau de confiance $(1-\alpha)$, repose d'abord sur la construction d'un estimateur par intervalle du paramètre θ.
- Il s'agit donc pour un niveau de confiance donné de trouver un intervalle aléatoire qui contient le paramètre θ avec une probabilité de $(1-\alpha)$.
- Généralement, cet intervalle est construit à partir de la loi de l'estimateur ponctuel T_n du paramètre θ. Il faut bien entendu que cette loi soit tabulée pour pouvoir en extraire un quantile.

1.4. Remarque sur la répartition du risque α.

Pour le niveau de confiance donné $(1-\alpha)$, on a :

$$
\begin{aligned}
& \mathbb{P}\Big[\theta \in [b_1(X_1,\cdots,X_n)\,;\, b_2(X_1,\cdots,X_n)]\Big] = 1-\alpha \\
\iff \quad & \mathbb{P}\Big[\theta \notin [b_1(X_1,\cdots,X_n)\,;\, b_2(X_1,\cdots,X_n)]\Big] = \alpha \\
\iff \quad & \underbrace{\mathbb{P}\left[\theta < b_1(X_1,\ldots,X_n)\right]}_{\alpha_1} + \underbrace{\mathbb{P}\left[\theta > b_2(X_1,\ldots,X_n)\right]}_{\alpha_2} = \alpha
\end{aligned}
$$

Toute la question est alors de définir la répartition du risque α entre les 2 probabilités α_1 et α_2.

En général, on répartit le risque α de manière égale sur les 2 probabilités ($\alpha/2$ pour chacune), mais on peut décider de faire autrement par exemple, pour diminuer la largeur de l'intervalle, lorsque la loi n'est pas symétrique. On parlera :

- d'intervalle bilatéral symétrique si $\alpha_1 = \alpha_2 = \alpha/2$;
- d'intervalle bilatéral si $\alpha_1 \neq 0$ et $\alpha_2 \neq 0$;
- d'intervalle unilatéral à gauche si $\alpha_2 = 0$;
- d'intervalle unilatéral à droite si $\alpha_1 = 0$.

2. Notion de Quantile.
2.1. Définition.

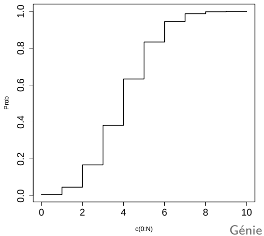
Soit X une variable aléatoire à valeurs dans un ensemble E, de fonction de répartition F.

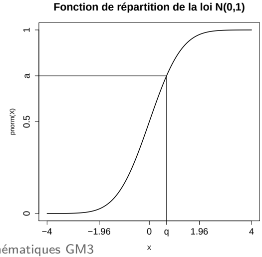
Quantile

On appelle quantile d'ordre a avec $a \in]0,1[$, le réel q défini par :

$$q = \inf\{x \in E \text{ tel que } F(x) \geq a\} \tag{1}$$

On trouvera dans le graphique ci-dessous la Fonction de répartition d'une variable aléatoire de loi binomiale $B(n = 10, p = 0.4)$ (à gauche) et celle d'une gaussienne centrée réduite (à droite).





2.2. Quantile dans le cas de lois continues.

Lorsque la variable aléatoire X est à valeurs réelles, de loi absolument continue par rapport à la mesure de Lebesgue, sa fonction de répartition est strictement croissante sur E.

La fonction F est alors bijective. Elle admet une fonction réciproque, notée F^{-1}.

Quantile

Le quantile q est alors solution de :

$$F(q) = \mathbb{P}\left[X \leq q]\right) = a \Longleftrightarrow q = F^{-1}(a) \tag{2}$$

Lorsque la fonction F est connue et que sa réciproque est explicite, on peut calculer très facilement le quantile q_a.

C'est le cas de la loi exponentielle, Weibull, Pareto, etc ...

Lorsque la fonction réciproque de F n'est pas explicite, on utilise des tables statistiques ou des logiciels pour déterminer le quantile recherché. C'est le cas en particulier pour la loi Normale, de Student, de Fisher, du Khi-deux, les lois usuellement utilisées en statistique.

2.3. Utilisation pour la construction d'intervalles de Pari.

2.3.1. Définition.

Intervalle de Pari

On appelle Intervalle de Fluctuation, ou parfois Intervalle de Pari, au niveau de confiance $1-\alpha$, d'une variable aléatoire X l'intervalle $[a, b]$ tel que

$$\mathbb{P}\left[X \in [a, b]\right] = 1 - \alpha$$

Comment trouve-t-on *a* et *b* ?

On a :

$$\mathbb{P}\Big[X \in [a, b]\Big] = 1 - \alpha \iff \mathbb{P}\Big[X \notin [a, b]\Big] = \alpha \iff \underbrace{\mathbb{P}\left[X < a\right]}_{=\alpha_1} + \underbrace{\mathbb{P}\left[X > b\right]}_{=\alpha_2} = \alpha$$

Lorsque $\alpha_1 \neq 0$ et $\alpha_2 \neq 0$, les réels a et b sont respectivement les quantiles d'ordre α_1 et $1-\alpha_2$ de la variable X puisque :

$$\mathbb{P}\left[X < a\right] = \alpha_1 \ \text{ et } \ \mathbb{P}\left[X \leq b\right] = 1 - \alpha_2$$

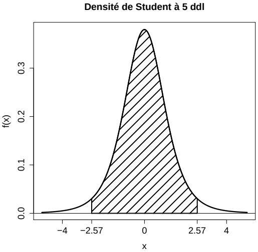
2.3.2. Remarques.

- L'intervalle de pari bilatéral symétrique contient les valeurs les plus vraisemblables de X.
- Si on répartit le risque α de manière égale a et b sont respectivement les quantiles d'ordre $\alpha/2$ et $1-\alpha/2$ de X.
- Lorsque la loi de X est symétrique et centrée et que l'on cherche un intervalle bilatéral symétrique, alors $a=-b$ (avec $b>0$). Il suffit alors de trouver le réel b qui n'est autre que le quantile d'ordre $1-\alpha/2$ de X.
- Nous verrons dans la partie du cours portant sur les tests statistiques d'hypothèse que la zone de rejet d'un test est le complémentaire de l'intervalle de pari de la loi de la statistique de test.

2.3.3. Exemple.

Considérons une variable aléatoire X de loi de Student à 5 degrés de liberté.
On souhaite trouver le réel t tel que:

$$\mathbb{P}\left[X \in [-t\,,\; t]\right] = \mathbb{P}\left[|X| \leq t\right] = 0.95$$



Le réel t est donc le quantile d'ordre 0,975 d'une loi de Student à 5 ddl.
On trouve dans la table de la loi de Student la valeur $t = 2,57$.

L'intervalle $[-2,57\,;\,2,57]$ est l'intervalle de Pari au niveau 95% de la Student à 5 ddl.
Il contient les valeurs les plus probables d'une Student à 5 ddl.

3. IC dans le cas de l'échantillon gaussien.
3.1. Le cadre.

On considère n données réelles $x_1, x_2, \ldots, x_n$.

On se place dans le cadre de l'échantillon gaussien, c'est à dire que l'on suppose que les données $x_1, x_2, \ldots, x_n$ sont les réalisations de n variables aléatoires $X_1, X_2, \ldots, X_n$ indépendantes et de même loi $\mathcal{N}(\mu, \sigma^2)$, avec μ et σ^2 inconnus.

On sait estimer les paramètres μ et σ^2 par $\overline{X}_n$ et S^2 respectivement.

cependant, quelle confiance accorder à ces estimations ?

Pour répondre à cette question, il suffit de construire des intervalles de confiance pour ces deux paramètres.
Nous ne considérerons que le cas des IC bilatéral symétrique.

3.2. Intervalle de confiance pour la moyenne.
3.2.1. Le résultat clé

Commençons par préciser le résultat théorique qui permettra de construire un estimateur par intervalle du paramètre μ.

On sait d'après le cours précédent, que : $Z = \dfrac{\sqrt{n}(\overline{X}_n - \mu)}{S} \sim T_{n-1}$.

Comme on l'a vu, on peut, pour tout $\alpha \in]0,1[$ et lorsque $Z \sim T_{n-1}$, trouver le réel positif t tel que

$$\mathbb{P}\Big[|Z| \leq t\Big] \; = \; 1 - \alpha \Longleftrightarrow \mathbb{P}\Big[Z \leq t\Big] \; = \; 1 - \alpha/2$$

Ce réel t est en fait le quantile d'ordre $(1 - \alpha/2)$ de la loi de Student à $(n-1)$ ddl.

C'est à partir de ce résultat que nous allons pouvoir construire un estimateur par intervalles, pour en déduire un intervalle de confiance pour le paramètre μ.

3.2.2. Construction de l'estimateur par intervalle.

On construit l'estimateur par intervalle du paramètre μ au niveau de confiance $(1-\alpha)$ de la manière suivante.

Soit t le quantile d'ordre $(1-\alpha/2)$ d'une loi de Student à $(n-1)$ ddl. Puisque $Z \sim T_{n-1}$, on a

$$
\begin{aligned}
\mathbb{P}\Big[|Z| \leq t\Big] = 1-\alpha \quad &\Longleftrightarrow \quad \mathbb{P}\left[\left|\frac{\sqrt{n}(\overline{X}_n - \mu)}{S}\right| \leq t\right] = 1-\alpha \\
&\Longleftrightarrow \quad \mathbb{P}\left[-t \leq \frac{\sqrt{n}(\overline{X}_n - \mu)}{S} \leq t\right] = 1-\alpha \\
&\Longleftrightarrow \quad \mathbb{P}\left[-t\,\frac{S}{\sqrt{n}} \leq \overline{X}_n - \mu \leq t\,\frac{S}{\sqrt{n}}\right] = 1-\alpha \\
&\Longleftrightarrow \quad \mathbb{P}\left[\overline{X}_n - t\,\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X}_n + t\,\frac{S}{\sqrt{n}}\right] = 1-\alpha
\end{aligned}
$$

Un estimateur par intervalle du paramètre μ au niveau de confiance $(1-\alpha)$ est donc donné par

$$
\left[\overline{X}_n - t\,\frac{S}{\sqrt{n}}\,;\,\overline{X}_n + t\,\frac{S}{\sqrt{n}}\right]
$$

3.2.3. L'intervalle de confiance.

On obtient l'intervalle de confiance en calculant l'estimateur par intervalle sur les données.

L'intervalle de confiance de la moyenne μ au niveau de confiance $(1-\alpha)$ est donc défini par

$$\mathrm{IC}_{1-\alpha}(\theta) = \left[\overline{x}_n - t\,\frac{s}{\sqrt{n}}\,;\,\overline{x}_n + t\,\frac{s}{\sqrt{n}}\right]$$

où $\overline{x}_n$ et s sont les réalisations sur les données de $\overline{X}_n$ et S, c'est à dire

$$\overline{x}_n = \frac{1}{n}\sum_{j=1}^{n} x_j \quad \text{et} \quad s = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(x_j - \overline{x}_n\right)^2}$$

cet intervalle est un intervalle de confiance bilatéral symétrique.

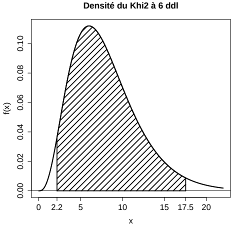
3.3. Intervalle de confiance pour la variance.
3.3.1. Le résultat clé

Pour construire un intervalle de confiance pour la variance σ^2, on utilise le fait que $Z = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

Si $Z \sim \chi^2_{n-1}$, alors pour tour réel $\alpha \in]0,1[$, on sait trouver les quantiles k_{α_1} et $k_{1-\alpha_2}$ tels que :

$$\mathbb{P}\left[k_{\alpha_1} \leq Z \leq k_{1-\alpha_2}\right] = 1-\alpha$$

avec $\alpha_1 + \alpha_2 = \alpha$. Généralement, on prend $\alpha_1 = \alpha/2$ et $\alpha_2 = \alpha/2$. Autrement dit, on encadre les valeurs de Z par les quantiles d'ordre $\alpha/2$ et $1-\alpha/2$ d'une loi du Khi-deux à $(n-1)$ ddl.



3.3.2. Construction de l'estimateur par intervalle.

On peut ainsi construire un estimateur par intervalle de la variance σ^2 au niveau de confiance $(1-\alpha)$.

On peut en effet montrer que

$$
\begin{aligned}
& \mathbb{P}\left[k_{\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq k_{1-\alpha/2}\right] = 1-\alpha \\
\iff \quad & \mathbb{P}\left[\frac{(n-1)S^2}{k_{1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{k_{\alpha/2}}\right] = 1-\alpha
\end{aligned}
$$

Un estimateur par intervalle de la variance σ^2 au niveau de confiance $(1-\alpha)$ est donc :

$$
\left[\frac{(n-1)S^2}{k_{1-\alpha/2}}\,;\,\frac{(n-1)S^2}{k_{\alpha/2}}\right]
$$

3.3.3. L'intervalle de confiance pour la variance σ^2.

On obtient l'intervalle de confiance en calculant l'estimateur par intervalle sur les données.

Un intervalle de confiance au niveau de confiance $(1-\alpha)$ pour la variance σ^2 est alors donné par :

$$\mathrm{IC}_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)s^2}{k_{1-\alpha/2}} \,;\, \frac{(n-1)s^2}{k_{\alpha/2}} \right]$$

où s^2 est la réalisation de S^2 sur les données.

Il s'agit ici d'un intervalle bilatéral non symétrique.

4. I.C. construit à partir d'un TLC.
4.1. Le cadre et le problème.

Considérons un paramètre inconnu θ et son estimateur T_n.

Dans de nombreuses situations, on ne dispose pas de la loi de l'estimateur T_n.

Cependant, on dispose parfois d'un thèorème de limite centrale (TLC), généralement un résultat de la forme

$$V_n(T_n - \theta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

où V_n est une variable aléatoire positive qui tend vers l'infini.

comment construire un intervalle de confiance bilatéral symétrique pour le paramètre θ à partir de ce TLC ?

4.2. Principe de Construction de l'intervalle de confiance.

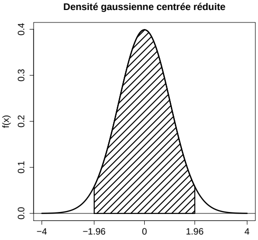
Pour construire un intervalle de confiance pour le paramètre θ, l'idée va être d'approximer la loi de $Q_n = V_n(T_n - \theta)$ par la loi $\mathcal{N}(0,1)$.

On considère donc que pour n fini, mais assez grand, $Q_n \sim \mathcal{N}(0,1)$.

Comme la loi normale $\mathcal{N}(0,1)$ est symétrique et tabulée, on peut pour un niveau de confiance $(1-\alpha)$ donné, trouver le réel $t > 0$ tel que

$$\mathbb{P}\left[-t \leq U \leq t\right] = \mathbb{P}\left[|U| \leq t\right] = 1 - \alpha$$

où $U \sim \mathcal{N}(0,1)$.
Ce réel est le quantile d'ordre $(1-\alpha/2)$ de la loi normale que l'on notera $u_{1-\alpha/2}$.



4.3. Construction de l'intervalle de confiance.
4.3.1. L'estimateur par intervalle.

En considérant que $Q_n \sim \mathcal{N}(0,1)$, nous allons pouvoir construire un intervalle de confiance pour le paramètre θ mais qui sera au niveau de confiance approché $(1-\alpha)$:

Pour le niveau de confiance $(1-\alpha)$ donné, on a :

$$
\begin{aligned}
& \mathbb{P}\left[|Q_n| \leq u_{1-\alpha/2}\right] = 1-\alpha \\
\Longleftrightarrow \quad & \mathbb{P}\left[T_n - \frac{u_{1-\alpha/2}}{V_n} \leq \theta \leq T_n + \frac{u_{1-\alpha/2}}{V_n}\right] = 1-\alpha
\end{aligned}
$$

L'estimateur par intervalle du paramètre θ est donc défini par :

$$
\left[T_n \ \pm \ \frac{u_{1-\alpha/2}}{V_n}\right]
$$

4.3.2. L'intervalle de confiance.

A partir de l'estimateur par intervalle ainsi obtenu, on déduit l'intervalle de confiance du paramètre θ au niveau de confiance approché $(1-\alpha)$:

$$\mathrm{IC}_{1-\alpha}(\theta) = \left[t_n \pm \frac{u_{1-\alpha/2}}{v_n} \right]$$

où t_n et v_n désignent les réalisations sur les données de T_n et V_n.

Toute la question sera alors de connaitre la confiance que l'on pourra accorder à l'approximation de Q_n par une loi normale centrée réduite.

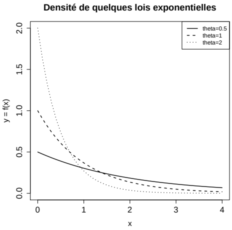
4. Exemple 1: Loi exponentielle. Densité.

On considère n données réelles positives $x_1, x_2, \ldots, x_n$. On suppose que ces données sont les réalisations de n variables aléatoires $X_1, X_2, \ldots, X_n$ i.i.d de loi exponentielle de paramètre $\theta > 0$.

La densité de la loi exponentielle est définie pour tout $x \geq 0$ par:

$$f(x) = \theta \exp(-\theta x)$$

Sa fonction de répartition F est donnée par: $F(x) = 1 - \exp(-\theta x)$.
On trouvera dans le graphique ci-dessous quelques courbes de densité avec différents paramétres (θ).



Vraisemblance et log-vraisemblance.

On souhaite estimer le paramètre θ par la méthode du maximum de vraisemblance.

1. On commence donc par expliciter la vraisemblance de l'échantillon.

   Puisque les variables aléatoires $X_1, X_2, \ldots, X_n$ sont indépendantes et de loi continue, la vraisemblance s'écrit ici :

$$L(x_1, x_2, \ldots, x_n, \theta) = \prod_{j=1}^{n} (\theta \exp(-\theta x_j)) = \theta^n \exp\left(-\theta \sum_{j=1}^{n} x_j\right).$$

2. La log-vraisemblance est alors égale à: $\forall \mathbf{x} = (x_1, x_2, \ldots, x_n)$.

$$\mathcal{L}(\mathbf{x}, \theta) = \log\left(L(\mathbf{x}, \theta)\right) = n \log(\theta) - \theta\, s_n.$$

   Avec $s_n = \sum_{j=1}^{n} x_j$.

Calcul de l'estimateur du maximum de vraisemblance.

L'estimation du maximum de vraisemblance du paramètre θ est donné par:

$$t_n = \arg\max_{z\in\mathbb{R}_+} L(\mathbf{x}, z) = \arg\max_{z\in\mathbb{R}_+} \mathcal{L}(\mathbf{x}, z)$$

• On pose $g(z) = \mathcal{L}(\mathbf{x}, z) = n\log(z) - z\, s_n$.
On montre facilement que :

$$g'(z) = \frac{n}{z} - s_n \quad \text{et} \quad g''(z) = -\frac{n}{z^2}$$

Ainsi, puisque $g'(z) = 0 \iff z = n/s_n$ et $g''(z) < 0$, la valeur t_n qui maximise la vraisemblance est égale à $t_n = 1/\overline{x}_n$.

L'estimateur T_n du maximum de vraisemblance est, alors, défini par :

$$T_n = \frac{1}{\overline{X}_n} \quad \text{avec} \quad \overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

Propriétés.

On démontre facilement que l'estimateur T_n du maximum de vraisemblance est fortement consistant :

$$T_n \xrightarrow[n\to\infty]{p.s.} \theta$$

En effet, gràce à la loi forte des grands nombres :

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow[n\to\infty]{p.s.} \mathbb{E}\left[X_1\right] = \frac{1}{\theta}$$

De la même manière, puisque gràce au théorème de limite centrale, on a :

$$\sqrt{n}\left(\overline{X}_n - \theta^{-1}\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \mathbb{V}\mathrm{ar}(X_1) = \frac{1}{\theta^2}\right)$$

on déduit par la delta-méthode que :

$$\sqrt{n}\left(T_n - \theta\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \theta^2\right)$$

Estimateur par intervalle

On pose $Q_n = \dfrac{\sqrt{n}(T_n - \theta)}{T_n}$.

Pour n assez grand, on décide d'approximer la loi de Q_n par une loi $\mathcal{N}(0,1)$.

Soit $u_{1-\alpha/2}$ le quantile d'ordre $(1-\alpha/2)$ de la loi normale centrée réduite. On a alors

$$\begin{aligned} & \mathbb{P}\left[|Q_n| \leq u_{1-\alpha/2}\right] = 1-\alpha \\ \iff \quad & \mathbb{P}\left[\left|\frac{\sqrt{n}(T_n-\theta)}{T_n}\right| \leq u_{1-\alpha/2}\right] = 1-\alpha \end{aligned}$$

L'estimateur par intervalle.

En travaillant par équivalence, on aboutit à :

$$\begin{aligned} & \mathbb{P}\left[\left|\frac{\sqrt{n}(T_n-\theta)}{T_n}\right| \leq u_{1-\alpha/2}\right] = 1-\alpha \\ \iff \quad & \mathbb{P}\left[|T_n-\theta| \leq u_{1-\alpha/2}\,\frac{T_n}{\sqrt{n}}\right] = 1-\alpha \\ \iff \quad & \mathbb{P}\left[T_n - u_{1-\alpha/2}\,\frac{T_n}{\sqrt{n}} \leq \theta \leq T_n + u_{1-\alpha/2}\,\frac{T_n}{\sqrt{n}}\right] = 1-\alpha \end{aligned}$$

Intervalle de confiance

L'estimateur par intervalle du paramètre θ au niveau de confiance $(1-\alpha)$ est :

$$\left[T_n - u_{1-\alpha/2} \frac{T_n}{\sqrt{n}} \; ; \; T_n + u_{1-\alpha/2} \frac{T_n}{\sqrt{n}} \right]$$

L'intervalle de confiance.

En prenant la réalisation de l'estimateur par intervalle sur les données, on obtient l'intervalle de confiance du paramètre θ :

$$\mathrm{IC}_{1-\alpha}(\theta) = \left[t_n - u_{1-\alpha/2} \frac{t_n}{\sqrt{n}} \; ; \; t_n + u_{1-\alpha/2} \frac{t_n}{\sqrt{n}} \right]$$

où t_n est la réalisation de T_n sur les données, c'est à dire

$$t_n = \frac{1}{\frac{1}{n} \sum_{j=1}^{n} x_j}$$

4. Exemple 2: loi de Weibull.

Reprenons le cadre de l'échantillon de Weibull.

On souhaite construire un intervalle de confiance pour le paramètre θ que l'on a estimé avec l'estimateur du maximum de vraisemblance T_n.

Pour cet estimateur, on ne dispose que du TLC suivant :

$$Q_n = \sqrt{g''(T_n)}\,(T_n - \theta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\, 1)$$

avec $g(z) = -LL(\mathbf{x}, z)$. Ainsi, pour n assez grand, on peut approximer la loi de Q_n par une loi $\mathcal{N}(0, 1)$ et en déduire un intervalle de confiance pour le paramètre θ au niveau de confiance $(1 - \alpha)$:

$$\mathrm{IC}_{1-\alpha}(\theta) = \left[t_n \;\pm\; u_{1-\alpha/2}\,\frac{1}{\sqrt{g''(t_n)}} \right]$$

Toute la question sera alors de savoir à partir de quelle taille d'échantillon, on peut considérer que l'approximation par la loi $\mathcal{N}(0, 1)$ est bonne.

4.1. Illustration par simulations avec le logiciel R.

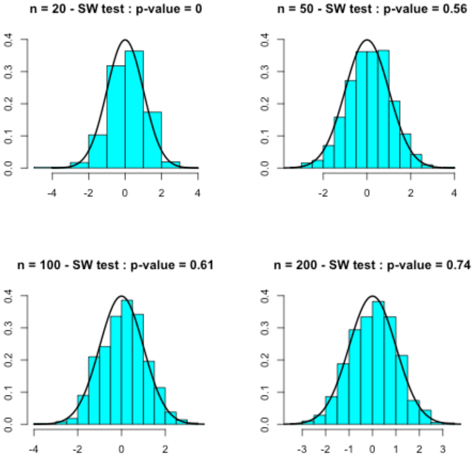
4.1.1. Pourquoi faire des simulations ?

Toute la question, c'est de savoir à partir de quelle taille d'échantillon l'approximation de la loi de Q_n par la loi normale centrée réduite est de bonne qualité.
On peut étudier cela par simulation, en simulant, pour différentes tailles d'échantillon, un grand nombre de réalisations de la variable Q_n afin de comparer la distribution de ces réalisations à celle de la loi normale centrée réduite.

On peut aussi envisager de mettre en oeuvre un test d'adéquation à la loi normale $\mathcal{N}(0,1)$ (test de Kolmogorov-Smirnov).

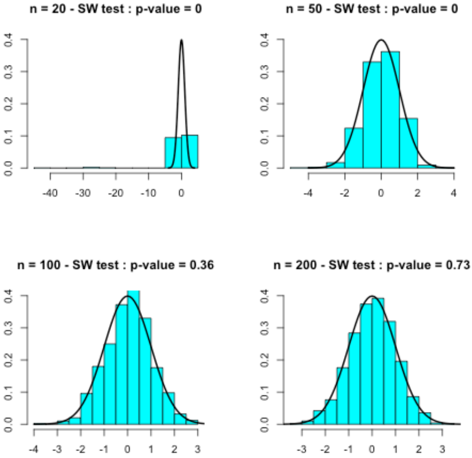
4.1.2. Illustration 1 : $\theta = 1,5$.

Le graphique ci-dessous permet d'illustrer la qualité de l'approximation fournie par le TLC dans le cas d'une Weibull standard de paramètre $\theta = 1,5$ et pour de petites tailles d'échantillon ($n = 20, 50, 100, 200$).



4.1.3. Illustration 2 : $\theta = 0,2$.

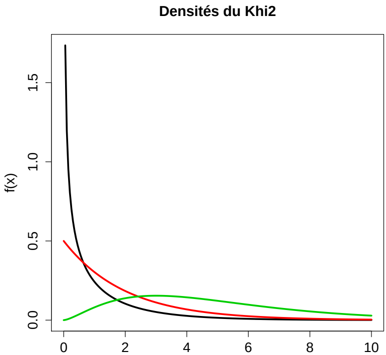
Le graphique ci-dessous permet d'illustrer la qualité de l'approximation fournie par le TLC dans le cas d'une Weibull standard de paramètre $\theta = 0,2$ et pour de petites tailles d'échantillon ($n = 20, 50, 100, 200$).

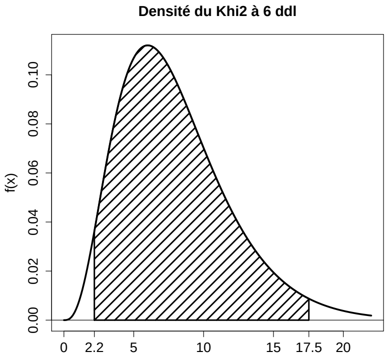


5. Rappels sur quelques lois utiles.
5.1. Loi du Khi-deux.

On appelle loi du chi-deux à p degrés de liberté (ddl) la loi de la variable

$Z = \sum_{j=1}^{p} X_j^2$ où les variables $X_1, X_2, \ldots, X_p,\ p \geq 1$ sont des variables aléatoires indépendantes et de même loi $\mathcal{N}(0,1)$.
Oa a de plus, $\mathbb{E}[Z] = p$ et $\mathbb{V}ar[Z] = 2p$.

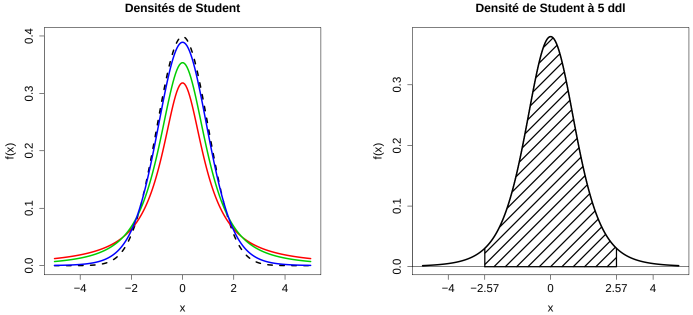




5.2. Loi de Student.

On appelle loi de Student à n degrés de liberté, le rapport d'une gaussienne centrée réduite et de la racine carrée d'un khi-deux à n degrés de liberté, la gaussienne et le khi-deux étant indépendants.

Plus précisément, si $U \sim \mathcal{N}(0,1)$ et $V \sim \chi^2_n$ avec U et V indépendantes, alors $Z_n = \dfrac{U}{\sqrt{V/n}}$ suit une loi de Student à n ddl et on note $Z \sim T_n$ et on a $\mathbb{E}(Z_n) = 0$ et $Z_n \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$.



5.3. Loi de Fisher.

On appelle loi de Fisher à p et q degrés de liberté le rapport de 2 khi-deux indépendants à p et q degrés de libertés respectivement.

Plus précisément, si $U \sim \chi^2_p$ et $V \sim \chi^2_q$ avec U et V indépendantes, alors $Z = \dfrac{U/p}{V/q}$ suit une loi de Fisher à p et q ddl et on note $Z \sim F(p,q)$.