Chapitre 6 : Estimation
MA 361 : Probabilités continues

Pierre-Alain TOUPANCE
pierre-alain.toupance@esisar.grenoble-inp.fr

Grenoble INP - ESISAR
3ième année

14 octobre 2020

Le but de l'estimation consiste à partir d'un échantillon de prévoir des informations sur la population totale.
On effectue deux types d'estimations :

- estimation ponctuelle
- estimation par intervalle de confiance

Soit Ω dont on considère un caractère :
On prend un échantillon de n individus et l'on obtient les données $x_1, x_2, ..., x_n$
On pose :

$$\widehat{m} = \frac{x_1 + x_2 + ... + x_n}{n}$$

$$\widehat{s}^2 = \frac{(x_1 - \widehat{m})^2 + ...(x_n - \widehat{m})^2}{n}$$

Soit $X_1, X_2, ..., X_n$ les variables aléatoires correspondant à ces données de moyenne m et de variance s^2, on pose :

$$\overline{X}_n = \frac{X_1 + ... + X_n}{n}$$

$$S_n^2 = \frac{(X_1 - \overline{X_n})^2 + ... + (X_n - \overline{X_n})^2}{n}$$

Estimateur

Soit (X_i) une suite de VA dépendant d'un paramètre θ. Soit $Y_n = g_n(X_1, ..., X_n)$ où g_n est une fonction de n variables.

1. On dit que Y_n est un estimateur de θ lorsque :

$$\lim_{n \to +\infty} \mathbb{E}(Y_n) = \theta$$

   $g_n(x_1, x_2, ... x_n)$ est appelé estimation de θ.

2. On dit que l'estimateur Y_n est sans biais lorsque $\forall n \in \mathbb{N}^*, \ \mathbb{E}(Y_n) = \theta$
3. On dit que l'estimateur est convergent lorsque $\lim \mathbb{V}(Y_n) = 0$

Estimation ponctuelle

On a :

- $E[\overline{X}_n] = m$, $\widehat{m}$ est donc une estimation ponctuelle sans biais de $\mathbb{E}[X]$
- $\mathbb{E}(S_n^2) = \dfrac{n-1}{n}s^2$, ainsi $\dfrac{n}{n-1}\widehat{s}^2$ est une estimation ponctuelle sans biais de $V(X)$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\sum_{i=1}^{n}(X_i - m) = \sum_{i=1}^{n} X_i - nm = n(\overline{X}_n - m)$$

Ainsi

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

$$\sum_{i=1}^{n}(X_i - m) = \sum_{i=1}^{n} X_i - nm = n(\overline{X}_n - m)$$

Ainsi

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

$$\sum_{i=1}^{n}(X_i - m) = \sum_{i=1}^{n} X_i - nm = n(\overline{X}_n - m)$$

Ainsi

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

$$\sum_{i=1}^{n}(X_i - m) = \sum_{i=1}^{n} X_i - nm = n(\overline{X}_n - m)$$

Ainsi

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

$$\sum_{i=1}^{n}(X_i - m) = \sum_{i=1}^{n} X_i - nm = n(\overline{X}_n - m)$$

Ainsi

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

On a :

$$\mathbb{E}[\frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2] =$$

et

$$\mathbb{E}[(\overline{X}_n - m)^2] =$$

On a :

$$\mathbb{E}[\frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2] = \frac{1}{n}\sum_{i=1}^{n}V(X_i) = s^2$$

et

$$\mathbb{E}[(\overline{X}_n - m)^2] =$$

On a :

$$\mathbb{E}[\frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2] = \frac{1}{n}\sum_{i=1}^{n} V(X_i) = s^2$$

et

$$\mathbb{E}[(\overline{X}_n - m)^2] = V(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n}s^2$$

Par conséquent :

$$\mathbb{E}[S_n^2] = s^2 - \frac{s^2}{n}$$

D'où

$$\mathbb{E}[S_n^2] = \frac{n-1}{n} s^2$$

On choisit comme estimation de θ la valeur qui maximise la probabilité de provoquer l'apparition de l'échantillon effectivement observé.

Maximum de vraisemblance

Soit $x_1,\ x_2, ..., x_n$ un échantillon, on note :

$$p = P(x_1, \theta)...P(x_n, \theta) = L(x_1, ..., x_n, \theta)$$

Dans le cas continue,

$$p = f(x_1, \theta)...f(x_n, \theta)dx_1...dx_n \text{ où } f \text{ est la densité des VAR}$$

On cherche alors à résoudre :

$$L(x_1, ..., x_n, \widehat{\theta}) = \max_{\theta} L(x_1, ..., x_n, \theta)$$

En général, on cherche à maximiser $\ln(L)$

Propriété

L'estimateur du maximum de vraisemblance du paramètre θ est la solution de :

$$\frac{\partial Ln(L(X,\theta))}{\partial \theta} = 0$$

$$\frac{\partial^2 Ln(L(X,\theta))}{\partial \theta^2} < 0$$

Exemple

Soit une population, avec $X \rightsquigarrow \mathcal{N}(m,\sigma)$.
On cherche à estimer m et σ

$$L = \Pi_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i-m)^2/(2\sigma^2)}$$

Propriété

L'estimateur du maximum de vraisemblance du paramètre θ est la solution de :

$$\frac{\partial Ln(L(X,\theta))}{\partial\theta} = 0$$

$$\frac{\partial^2 Ln(L(X,\theta))}{\partial\theta^2} < 0$$

Exemple

Soit une population, avec $X \rightsquigarrow \mathcal{N}(m,\sigma)$.
On cherche à estimer m et σ

$$L = \Pi_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x_i-m)^2/(2\sigma^2)}$$

$$L = \frac{1}{} e^{-\sum(x_i-m)^2/(2\sigma^2)}$$

$$\ln L = -n\ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2$$

$$\frac{\partial \ln(L)}{\partial m} = 0 \Leftrightarrow \frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(x_i - m) = 0$$

$$\Leftrightarrow m = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\partial \ln(L)}{\partial \sigma} = 0 \Leftrightarrow -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - m)^2 = 0$$

$$\Leftrightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$$

$$\ln L = -n \ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - m)^2$$

$$\frac{\partial \ln(L)}{\partial m} = 0 \Leftrightarrow \frac{1}{2\sigma^2} \sum_{i=1}^{n} 2(x_i - m) = 0$$

$$\Leftrightarrow m = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\frac{\partial \ln(L)}{\partial \sigma} = 0 \Leftrightarrow -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - m)^2 = 0$$

$$\Leftrightarrow \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - m)^2$$

$$\ln L = -n\ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2$$

$$\frac{\partial \ln(L)}{\partial m} = 0 \Leftrightarrow \frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(x_i - m) = 0$$

$$\Leftrightarrow m = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\partial \ln(L)}{\partial \sigma} = 0 \Leftrightarrow -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - m)^2 = 0$$

$$\Leftrightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$$

$$\ln L = -n \ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2$$

$$\frac{\partial \ln(L)}{\partial m} = 0 \Leftrightarrow \frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(x_i - m) = 0$$

$$\Leftrightarrow m = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\partial \ln(L)}{\partial \sigma} = 0 \Leftrightarrow -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - m)^2 = 0$$

$$\Leftrightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$$

Loi du Khi deux

Soit $X_1, X_2, ..., X_n$ n variables aléatoires normales centrées réduites indépendantes.

On pose $U = \sum_{i=1}^{n} X_i^2$

On dit que U suit une loi du Khi deux à n degrés de liberté (ddl), on note $U \rightsquigarrow \chi_n^2$.

On a $E[U] = n$ et $V(U) = 2n$

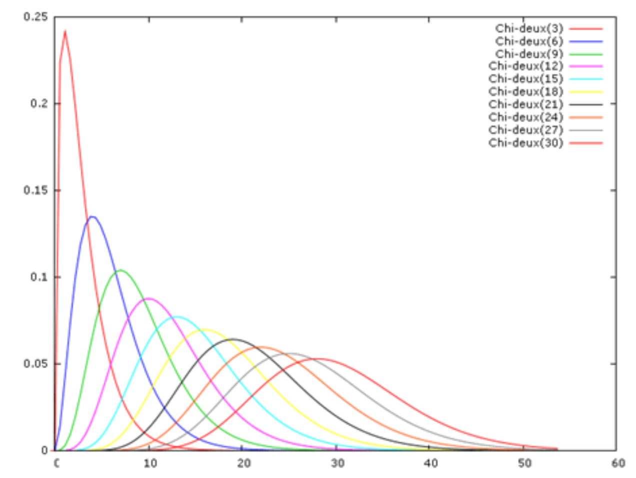

FIGURE – Densités de plusieurs lois du Khi deux

Loi du Student

Soit $X \rightsquigarrow \mathcal{N}(0,1)$ et $Y_n \rightsquigarrow \chi^2_n$

Soit $T_n = \dfrac{X}{\sqrt{Y_n/n}}$

On dit que T_n suit une loi du Student à n degrés de liberté.

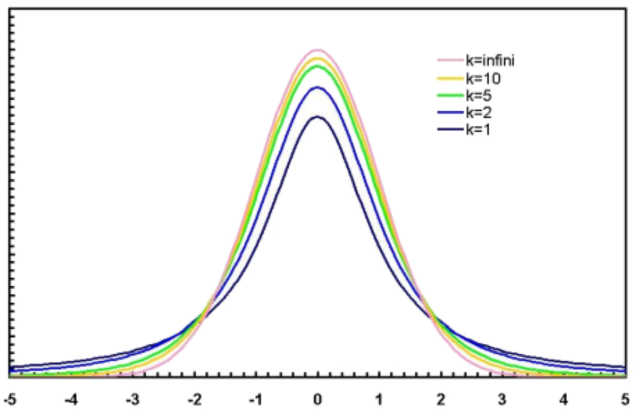

Figure – Densités de plusieurs lois du Student

intervalle de confiance

Il est préférable de compléter l'estimation ponctuelle par un intervalle, c'est à dire, nous cherchons a et b tel que :

$$P(a \leq \theta \leq b) = 1 - \alpha$$

où θ est la valeur que l'on souhaite estimée (moyenne ou variance par exemple)

- $1 - \alpha$ est appelé **niveau de confiance** de l'intervalle.
- On dit aussi que $[a; b]$ est un intervalle de confiance de θ **au risque** α
- On cherchera a et b tel que $P(\theta \leq a) = P(\theta \geq b) = \dfrac{\alpha}{2}$

intervalle de confiance d'une proportion

Soit K_n la variable aléatoire qui compte le nombre d'individu ayant la propriété P dans un échantillon de taille n non exhaustif.
On a $K_n \rightsquigarrow \mathcal{B}(n, f)$

Si les conditions d'approximation sont vérifiées, $F_n = \frac{K_n}{n}$ peut être approchée par une loi normale $U \rightsquigarrow \mathcal{N}\left(f, \sqrt{\frac{f(1-f)}{n}}\right)$

On pose $U^* = \dfrac{F-f}{\sqrt{\frac{f(1-f)}{n}}} \rightsquigarrow \mathcal{N}(0,1)$.

A l'aide de la table de la loi normale centrée réduite, on détermine $t_{1-\alpha/2}$ tel que :

$$\mathbb{P}\left(-t_{1-\alpha/2} \leqslant \frac{F-f}{\sqrt{\frac{f(1-f)}{n}}} \leqslant t_{1-\alpha/2}\right) = 1-\alpha$$

On a ainsi :

$$\mathbb{P}\left(F - t_{1-\alpha/2}\sqrt{\frac{f(1-f)}{n}} \leqslant f \leqslant F + t_{1-\alpha/2}\sqrt{\frac{f(1-f)}{n}}\right) = 1-\alpha$$

Par conséquent, on prendra : $\begin{cases} a = \widehat{f} - t_{1-\alpha/2}\sqrt{\dfrac{\widehat{f}(1-\widehat{f})}{n}} \\ b = \widehat{f} + t_{1-\alpha/2}\sqrt{\dfrac{\widehat{f}(1-\widehat{f})}{n}} \end{cases}$

Intervalle de confiance d'une proportion

On obtient ainsi l'intervalle de confiance au risque α

$$\boxed{I_\alpha = \left[\widehat{f} - t_{1-\alpha/2}\sqrt{\frac{\widehat{f}(1-\widehat{f})}{n}}\,;\,\widehat{f} + t_{1-\alpha/2}\sqrt{\frac{\widehat{f}(1-\widehat{f})}{n}}\right]}$$

où $\widehat{f}$ est la proportion de l'échantillon

Exemple :

1000 électeurs choisis au hasard ont été interrogés avant une élection. 520 se sont déclarés favorables au candidat Robert. Déterminer un intervalle de confiance à 95% de la proportion p d'électeurs favorables à Robert dans la population.

Exemple :
Soit F la variable aléatoire égale à la proportion de personnes qui votent pour A. On a

$$F \rightsquigarrow \mathcal{N}\left(f, \sqrt{\frac{0,52 \times 0,48}{1000}}\right)$$

On pose

$$F^* = \frac{F - f}{\sqrt{\frac{0,52 \times 0,48}{1000}}}$$

On cherche $t_{1-\alpha/2}$ tel que $P(-t_{1-\alpha/2} \leq F^* \leq t_{1-\alpha/2}) = 0,95$

Exemple :

Soit F la variable aléatoire égale à la proportion de personnes qui votent pour A. On a

$$F \rightsquigarrow \mathcal{N}\left(f, \sqrt{\frac{0,52 \times 0,48}{1000}}\right)$$

On pose

$$F^* = \frac{F - f}{\sqrt{\frac{0,52\times 0,48}{1000}}}$$

On cherche $t_{1-\alpha/2}$ tel que $P(-t_{1-\alpha/2} \leq F^* \leq t_{1-\alpha/2}) = 0,95$

Exemple :

Soit F la variable aléatoire égale à la proportion de personnes qui votent pour A. On a

$$F \rightsquigarrow \mathcal{N}\left(f, \sqrt{\frac{0,52 \times 0,48}{1000}}\right)$$

On pose

$$F^* = \frac{F - f}{\sqrt{\frac{0,52 \times 0,48}{1000}}}$$

On cherche $t_{1-\alpha/2}$ tel que $P(-t_{1-\alpha/2} \leq F^* \leq t_{1-\alpha/2}) = 0,95$

En utilisant la table de la loi normale centrée réduite, on obtient : $t_{1-\alpha/2} = 1,96$

On a que

$$P(F - 1,96\sqrt{\frac{0,52 \times 0,48}{1000}} \leq f \leq F + 1,96\sqrt{\frac{0,52 \times 0,48}{1000}}) = 0,95$$

On prendra $\begin{cases} a = 0,52 - 1,96\sqrt{\frac{0,52\times0,48}{1000}} \\ b = 0,52 + -1,96\sqrt{\frac{0,52\times0,48}{1000}} \end{cases}$

Ainsi l'intervalle de confiance au seuil de 5 % est :

$$I = [0,489; 0,551]$$

En utilisant la table de la loi normale centrée réduite, on obtient : $t_{1-\alpha/2} = 1,96$
On a que

$$P(F-1,96\sqrt{\frac{0,52\times 0,48}{1000}} \leq f \leq F+1,96\sqrt{\frac{0,52\times 0,48}{1000}}) = 0,95$$

On prendra $\begin{cases} a = 0,52 - 1,96\sqrt{\frac{0,52\times0,48}{1000}} \\ b = 0,52 + -1,96\sqrt{\frac{0,52\times0,48}{1000}} \end{cases}$
Ainsi l'intervalle de confiance au seuil de 5 % est :

$$I = [0,489; 0,551]$$

En utilisant la table de la loi normale centrée réduite, on obtient : $t_{1-\alpha/2} = 1,96$
On a que

$$P(F-1,96\sqrt{\frac{0,52\times 0,48}{1000}} \leq f \leq F+1,96\sqrt{\frac{0,52\times 0,48}{1000}}) = 0,95$$

On prendra $\begin{cases} a = 0,52 - 1,96\sqrt{\frac{0,52\times0,48}{1000}} \\ b = 0,52 + -1,96\sqrt{\frac{0,52\times0,48}{1000}} \end{cases}$

Ainsi l'intervalle de confiance au seuil de 5 % est :

$$I = [0,489; 0,551]$$

En utilisant la table de la loi normale centrée réduite, on obtient : $t_{1-\alpha/2} = 1,96$
On a que

$$P(F-1,96\sqrt{\frac{0,52\times 0,48}{1000}} \leq f \leq F+1,96\sqrt{\frac{0,52\times 0,48}{1000}}) = 0,95$$

On prendra $\begin{cases} a = 0,52 - 1,96\sqrt{\frac{0,52\times 0,48}{1000}} \\ b = 0,52 + -1,96\sqrt{\frac{0,52\times 0,48}{1000}} \end{cases}$
Ainsi l'intervalle de confiance au seuil de 5 % est :

$$I = [0,489; 0,551]$$

intervalle de confiance de la moyenne

1er cas : σ est connu

- Si les variables aléatoires $X_1, X_2, ..., X_n$ suivent des lois normales de paramètre (m, σ) alors $\overline{X}_n \rightsquigarrow \mathcal{N}(m, \frac{\sigma}{\sqrt{n}})$.
- si $n > 30$ alors $\overline{X}_n$ peut être approchée par $Y \rightsquigarrow \mathcal{N}(m, \frac{\sigma}{\sqrt{n}})$

intervalle de confiance de la moyenne

1er cas : σ est connu On pose

$$U = \frac{\overline{X}_n - m}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0,1)$$

on détermine $t_{1-\alpha/2}$ tel que :

$$\mathbb{P}\Big(-t_{1-\alpha/2} \leqslant \frac{\overline{X}_n - m}{\sigma/\sqrt{n}} \leqslant t_{1-\alpha/2}\Big) = 1-\alpha$$

C'est équivalent à

$$\mathbb{P}\Big(\overline{X}_n - t_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leqslant m \leqslant \overline{X}_n + t_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\Big) = 1-\alpha$$

1er cas : σ est connu

L'intervalle de confiance de m au risque α est :

$$\boxed{I_\alpha = [\widehat{m} - \frac{\sigma}{\sqrt{n}} t_{1-\alpha/2} ; \widehat{m} + \frac{\sigma}{\sqrt{n}} t_{1-\alpha/2}]}$$

2ième cas : σ est inconnu

Dans ce cas, on pose

$$T = \frac{\overline{X}_n - m}{\sqrt{S_n^2/(n-1)}} = \frac{(\overline{X}_n - m)/(\sigma/\sqrt{n})}{\sqrt{(nS_n^2/(\sigma^2))/(n-1)}}$$

T suit une loi du Student à n-1 degré de liberté, on utilise la table numérique du Student pour obtenir $t_{1-\alpha/2}$ tel que $P(-t_{1-\alpha/2} \leq T \leq t_{1-\alpha/2}) = 1 - \alpha$

Remarque : D'après le théorème de Cochran

$\frac{nS_n^2}{\sigma^2} = \sum_{i=1}^n (\frac{X_i - \overline{X}_n}{\sigma})^2$ suit une loi du Khi deux à $n-1$ ddl.

2ième cas : σ est inconnu

On a

$$\mathbb{P}\Big(-t_{1-\alpha/2} \leqslant T \leqslant_{1-\alpha/2} \Big) = 1-\alpha$$

Par conséquent :

$$\mathbb{P}\Big(\overline{X}_n - t_{1-\alpha/2}\sqrt{S_n^2/(n-1)} \leqslant m \leqslant \overline{X}_n + t_{1-\alpha/2}\sqrt{S_n^2/(n-1)}\Big) = 1-\alpha$$

L'intervalle de confiance de m au risque α est :

$$\boxed{I_\alpha = \left[\widehat{m} - \frac{\widehat{s}}{\sqrt{n-1}}t_{1-\alpha/2}; \widehat{m} + \frac{\widehat{s}}{\sqrt{n-1}}t_{1-\alpha/2}\right]}$$

2ième cas : σ est inconnu : Exemple

On réceptionne des pièces et on sait que la longueur des pièces à une distribution normale.
On prélève un échantillon de 20 pièces qui a une moyenne de 10cm et un écart type de 2.
Déterminons un intervalle de confiance de la moyenne de la longueur des pièces au risque de 5 %

2ième cas : σ est inconnu :

Soit $X_1, X_2, ...X_{20}$ les variables aléatoires égales à la longueur des 20 pièces prélevées.

On pose

$$\overline{X}_{20} = \frac{\sum X_i}{20}$$

$$T = \frac{\overline{X}_{20} - m}{\sqrt{S_{20}^2/19}}$$

T suit une loi du Student à 19 ddl.

On cherche $t_{1-\alpha/2}$ tel que

$$P(t_{1-\alpha/2} \leq \overline{X}_{20} \leq t_{1-\alpha/2}) = 0,95$$

dans la table du Student à 19 ddl. on obtient $t_{1-\alpha/2} = 2,093$
On a :

$$\mathbb{P}\left(\overline{X}_{20} - 2,093\sqrt{S_{20}^2/19} \leqslant m \leqslant \overline{X}_{20} + 2,093\sqrt{S_n^2/(n-1)}\right) = 0,95$$

Ainsi l'intervalle de confiance est :

$$I = [10 - 2,093 \times \sqrt{\frac{4}{19}}; 10 + 2,093 \times \sqrt{\frac{4}{19}}]$$

$$I = [9,0397; 10,9603]$$

intervalle de confiance de la variance

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

On pose :

$$Y = (n-1)\frac{S_n^2}{\sigma^2}$$

Y suit une loi du Khi deux à (n-1) degrés de liberté.

On détermine $t_{\alpha/2}$ et $t_{1-\alpha/2}$ dans la table du Khi deux tel que :

$$P\left(t_{\alpha/2} \leq (n-1)\frac{S_n^2}{\sigma^2} \leq t_{1-\alpha/2}\right) = 1-\alpha$$

On a donc

$$P\left((n-1)\frac{S_n^2}{t_{1-\alpha/2}} \leq \sigma^2 \leq (n-1)\frac{S_n^2}{t_{\alpha/2}}\right) = 1-\alpha$$

On obtient ainsi l'intervalle de confiance de la variance :

$$\boxed{I_\alpha = \left[\frac{(n-1)\widehat{s}^2}{t_{1-\alpha/2}} ; \frac{(n-1)\widehat{s}^2}{t_{\alpha/2}}\right]}$$

Exemple :

On a un échantillon de 16 chiffres d'affaires tel que $s'^2_{16} = 72,53$. Déterminer l'intervalle de confiance de la variance de niveau 0,95.

Sur la table de la loi du Khi deux à 15 ddl, on a : $t_{0,025} = 6,27$ et $t_{0,975} = 27,49$

Ainsi l'intervalle de confiance au risque de 5% de σ^2 est

$$I = [15 \times \frac{72,53}{27,49}; 15 \times \frac{72,53}{6,27}] = [39,57; 173,52].$$

Pour σ on obtient donc l'intervalle de confiance : $[6,29; 13,17]$.

Exemple :

On a un échantillon de 16 chiffres d'affaires tel que $s'^2_{16} = 72,53$. Déterminer l'intervalle de confiance de la variance de niveau 0,95.

Sur la table de la loi du Khi deux à 15 ddl, on a : $t_{0,025} = 6,27$ et $t_{0,975} = 27,49$

Ainsi l'intervalle de confiance au risque de 5% de σ^2 est

$$I = [15 \times \frac{72,53}{27,49}; 15 \times \frac{72,53}{6,27}] = [39,57; 173,52].$$

Pour σ on obtient donc l'intervalle de confiance : $[6,29; 13,17]$.

Exemple :

On a un échantillon de 16 chiffres d'affaires tel que $s'^2_{16} = 72,53$. Déterminer l'intervalle de confiance de la variance de niveau 0,95.

Sur la table de la loi du Khi deux à 15 ddl, on a : $t_{0,025} = 6,27$ et $t_{0,975} = 27,49$

Ainsi l'intervalle de confiance au risque de 5% de σ^2 est

$$I = [15 \times \frac{72,53}{27,49}; 15 \times \frac{72,53}{6,27}] = [39,57; 173,52].$$

Pour σ on obtient donc l'intervalle de confiance : $[6,29; 13,17]$.