Chapitre 3 : Intervalles de confiance

N. Biheng

1 Cadre général

> *J'ai encore une fois oublié que le hasard est le résultat d'une immense équation dont nous ne connaissons pas toutes les racines.*
> (Honoré de Balzac in *Z. Marcas*)

> *On définit la « qualité » d'une estimation en lui associant un intervalle ayant une forte probabilité de contenir la vraie valeur du paramètre[...].*
> (André Vessereau in [7])

Lors du chapitre précédent, nous avons étudié l'estimation ponctuelle d'un paramètre θ (ou plus généralement de $g(\theta)$). En revanche, nous ne disposons pas d'éléments pour quantifier la qualité d'une estimation.
Rappelons deux définitions du chapitre précédent.

Définition 1
Un *modèle statistique* est la donnée de l'ensemble des valeurs que peut prendre la variable aléatoire X noté $\mathcal{H}$, d'un ensemble $\Theta \subset \mathbb{R}^d$ appelé l'ensemble des paramètres et d'une famille $\mathcal{P} := \{\mathbb{P}_\theta | \theta \in \Theta\}$ de lois indexée par Θ.

Définition 2
Un échantillon de taille n est la donnée de n variables aléatoires indépendantes et identiquement distribuées (i.i.d.) qui suivent la même loi que la variable aléatoire X.

Ce chapitre reposera sur la notion de *fonction pivotale* (*asymptotiquement pivotale* pour les intervalles de confiance asymptotiques).
Une *fonction pivotale* est une fonction Ψ définie sur $\mathcal{H}^n \times \Theta$ telle que $\Psi(X_1, \ldots, X_n, \theta)$ suit une loi « indépendante » du paramètre θ. Nous pourrons ainsi encadrer $\Psi(X_1, \ldots, X_n, \theta)$ à l'aide de fractiles [1] de cette loi et en déduire un encadrement de θ qui sera fonction de l'échantillon.

1. Soit X une variable aléatoire. Un *fractile* d'ordre α de la loi de X est un réel u_α tel que $\mathbb{P}(u_\alpha \leq X) = \alpha$.

Exemple 1
Pour un échantillon $(X_1, \ldots, X_n)$ de loi normale $\mathcal{N}(m, \sigma^2)$ avec σ^2 connu, $\Psi(X_1, \ldots, X_n, m) := \dfrac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$ suit une loi normale centrée réduite donc la fonction Ψ est une fonction pivotale dans ce cas.

Nous exploiterons différentes fonctions pivotales dans ce chapitre sans les expliciter.

2 Intervalle de confiance pour la moyenne m

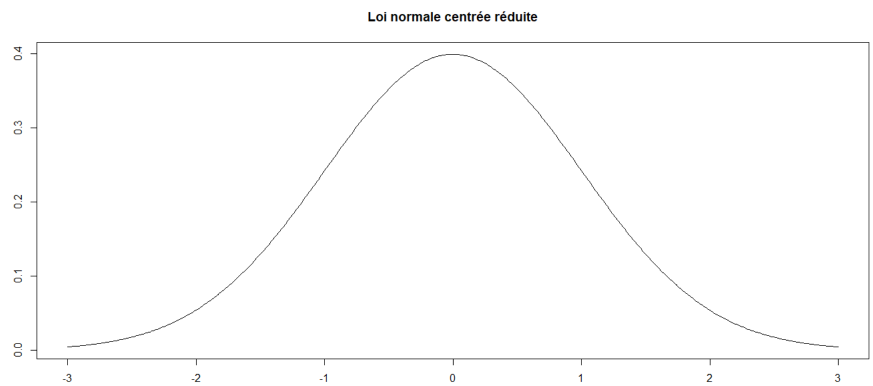
2.1 Cas 1 : Variance connue

Rappelons que nous nous placerons dans le cas gaussien, i.e. nous chercherons à estimer la moyenne m d'une variable aléatoire X dont nous connaissons la variance σ^2(et donc l'écart-type σ). La proposition suivante, vue dans le chapitre 1, nous sera d'une grande utilité :

Proposition 1
Soit X une variable aléatoire suivant une loi normale centrée réduite ($\mathcal{N}(0, 1)$).

1. $\mathbb{P}(X \leq 0) = \mathbb{P}(X \geq 0) = 0,5$
2. Plus généralement, pour tout réel a, $\mathbb{P}(X \leq -a) = \mathbb{P}(X \geq a)$
3. $\mathbb{P}(-1,96 \leq X \leq 1,96) \approx 0,95$ et $\mathbb{P}(-2,58 \leq X \leq 2,58) \approx 0,99$.

Rappelons également que la loi normale est symétrique par rapport à sa moyenne m et donc, que la loi normale centrée réduite l'est par rapport à 0 comme l'indique l'illustre la proposition précédente.



Soit $(X_1, \ldots, X_n)$ un échantillon de taille $n \geq 1$ de la loi $\mathcal{N}(m, \sigma^2)$ i.e. la donnée de n variables aléatoires indépendantes suivant cette loi.

Les variables aléatoires X_i sont indépendantes donc $\displaystyle\sum_{i=1}^{n} X_i$ suit une loi $\mathcal{N}(nm, n\sigma^2)$ et $\bar{X}_n = \displaystyle\frac{1}{n}\sum_{i=1}^{n} X_i$ suit une loi $\mathcal{N}(m, \frac{\sigma^2}{n})$.

Ceci repose sur le résultat suivant bien connu[2] :

Proposition 2

1. Soient X_1 et X_2 deux variables aléatoires indépendantes suivant respectivement des lois $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$ alors la variable aléatoire $X_1 + X_2$ suit une loi $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

2. Soit X une variable aléatoire suivant une loi $\mathcal{N}(m, \sigma^2)$ alors la variable aléatoire $Y = X - m$ suit une loi $\mathcal{N}(0, \sigma^2)$.

3. Soit X une variable aléatoire suivant une loi $\mathcal{N}(m, \sigma^2)$ et a un réel différent de 0, la variable aléatoire aX suit une loi $\mathcal{N}(am, a^2\sigma^2)$.

D'après le deuxième point, $\bar{X}_n - m$ suit une loi $\mathcal{N}(0, \frac{\sigma^2}{n})$.

Donc, en utilisant le troisième point, $\displaystyle\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$ suit une loi normale centrée réduite.

2. Nous avons démontré le premier point lors du chapitre 1 à l'aide de la fonction caractéristique. Les autres points peuvent être obtenus de manière similaire.

Ex supra,

$$\mathbb{P}\left(-1,96 \leq \frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \leq 1,96\right) \approx 0,95$$

Notons que ceci ne dépend pas de la valeur de n.
Par conséquent, le réel m que nous cherchons à estimer vérifie :

$$m \in \left[\bar{X}_n - 1,96\frac{\sigma}{\sqrt{n}}; \bar{X}_n + 1,96\frac{\sigma}{\sqrt{n}}\right]$$

au niveau de confiance 0,95, au sens où il appartient à cet intervalle avec une probabilité égale à 0,95. En d'autres termes, il appartient à cet intervalle pour 95% des échantillons.
En considérant les réalisations d'un échantillon $(X_1, \ldots, X_n)$, c'est à dire des observations $(x_1, \ldots, x_n)$ dont nous noterons la moyenne $\bar{x}_n$, nous pourrons définir l'intervalle de confiance au niveau de confiance $0,95$ pour la moyenne m.

Définition 3
L'*intervalle de confiance* au niveau $0,95$ pour la moyenne m est :

$$\left[\bar{x}_n - 1,96\frac{\sigma}{\sqrt{n}}; \bar{x}_n + 1,96\frac{\sigma}{\sqrt{n}}\right].$$

De même, pour le niveau de confiance 0,99, nous utiliserons le dernier point de la proposition 1.

Définition 4
L'*intervalle de confiance* au niveau $0,99$ pour la moyenne m est :

$$\left[\bar{x}_n - 2,58\frac{\sigma}{\sqrt{n}}; \bar{x}_n + 2,58\frac{\sigma}{\sqrt{n}}\right].$$

Plus généralement en notant $z_{1-\frac{\alpha}{2}}$ le fractile d'ordre $1 - \dfrac{\alpha}{2}$ de la loi $\mathcal{N}(0,1)$ [3], nous pouvons définir :

Définition 5
L'intervalle de confiance au niveau de confiance $1 - \alpha$ pour la moyenne m est :

$$\left[\bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \bar{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right].$$

3. Nous laissons $\frac{\alpha}{2}$ à gauche de l'intervalle et $\frac{\alpha}{2}$ à droite de l'intervalle. La forme de l'intervalle est simple car la loi normale centrée réduite est symétrique par rapport à 0.

Exemple 2
Proposer un intervalle de confiance au niveau 0,90 pour la moyenne m pour une variable aléatoire gaussienne de variance 2 dont nous connaissons les observations suivantes : 3,1 ; 2,4 ; 5 ; 7 et 2,8.

Calculons tout d'abord $\bar{x}$. On a :

$$\bar{x} = \frac{3,1 + 2,4 + 5 + 7 + 2,8}{5} = 4,06$$

Nous aurons recours au fractile $z_{0,95} \approx 1,64$.
L'intervalle de confiance au niveau 0,90 pour la moyenne est :

$$\left[4,06 - 1,64\frac{\sqrt{2}}{\sqrt{5}}; 4,06 + 1,64\frac{\sqrt{2}}{\sqrt{5}}\right]$$

i.e. approximativement l'intervalle $[3,02; 5,10]$.

2.2 Cas 2 : Variance inconnue

Nous aurons besoin d'introduire la loi du Khi-deux et la loi de Student.

Définition 6
Soient $(X_1, \ldots, X_n)$ la donnée de n variables aléatoires indépendantes normales centrées réduites.
La variable aléatoire $U_n := \sum_{i=1}^{n} X_i^2$ suit une loi du Khi-deux à n degrés de liberté notée $\chi^2(n)$.

Enonçons ses propriétés principales dans la proposition suivante :

Proposition 3
1. Sa densité est : $f_{U_n}(x) = \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} e^{-\frac{x}{2}} x^{\frac{n}{2}-1}$ pour $x \geq 0$,
2. $E(U_n) = n$,
3. $V(U_n) = 2n$,
4. Sa fonction caractéristique est : $\phi_{U_n}(t) = \dfrac{1}{(1-2it)^{\frac{n}{2}}}$.

Le théorème suivant sera également utile :

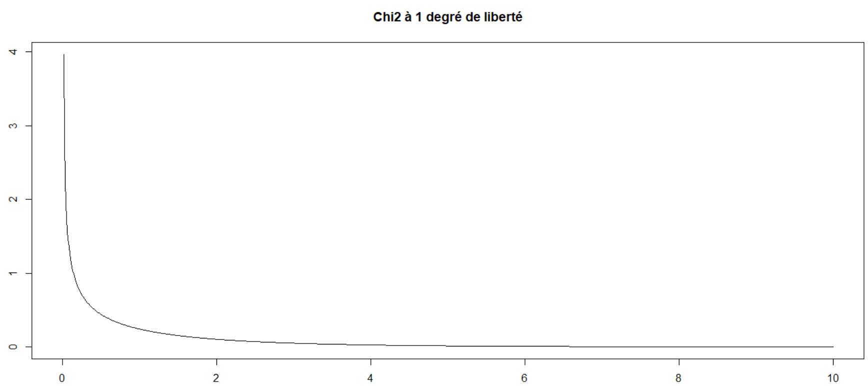
Théorème 1
Soient X et Y deux variables aléatoires indépendantes suivant respectivement des lois $\chi^2(m)$ et $\chi^2(n)$ alors la variable aléatoire $X+Y$ suit une loi $\chi^2(m+n)$.

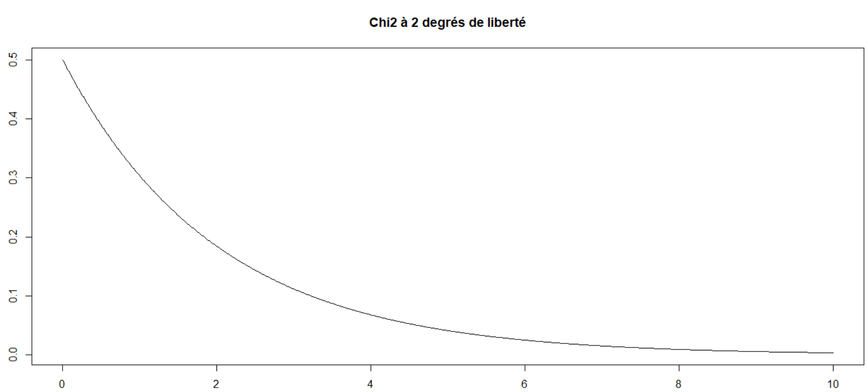
Preuve
Les variables aléatoires X et Y sont indépendantes donc :

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) = \frac{1}{(1-2it)^{\frac{m}{2}}} \times \frac{1}{(1-2it)^{\frac{n}{2}}} = \frac{1}{(1-2it)^{\frac{m+n}{2}}}$$

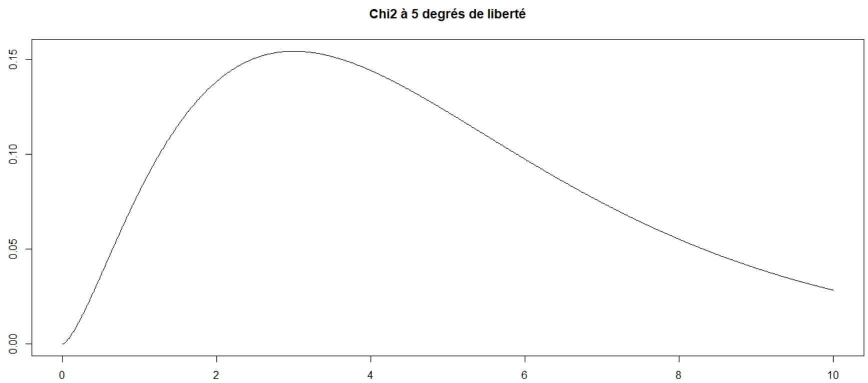
d'où le résultat.

Voici les représentations graphiques pour 1 et 2 degrés de liberté.



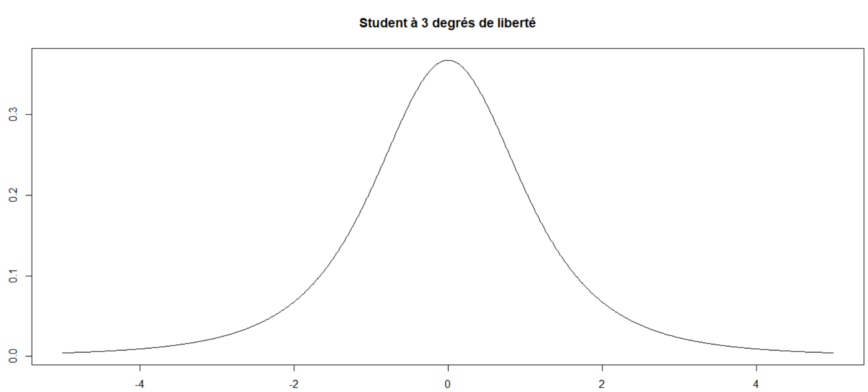


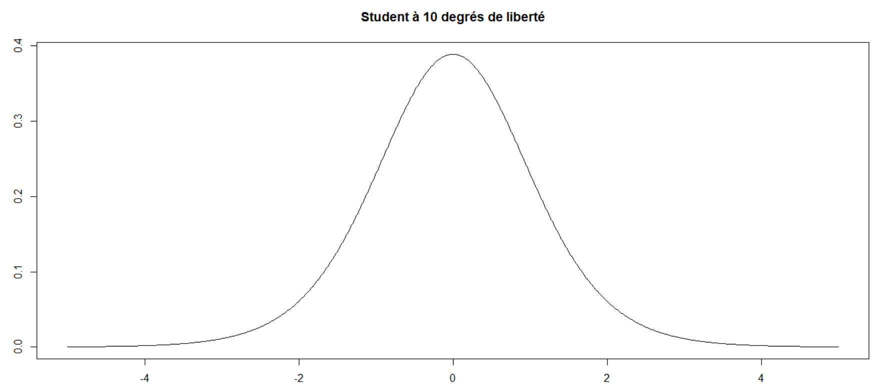
Pour $n > 2$, la forme est la suivante :



Définition 7
Soient X et Y deux variables aléatoires indépendantes suivant respectivement des lois $\mathcal{N}(0,1)$ et $\chi^2(n)$.
La variable $T_n = \dfrac{X}{\sqrt{\frac{Y}{n}}}$ suit une loi de Student à n degrés de liberté. Elle est notée $\mathcal{T}_n$.

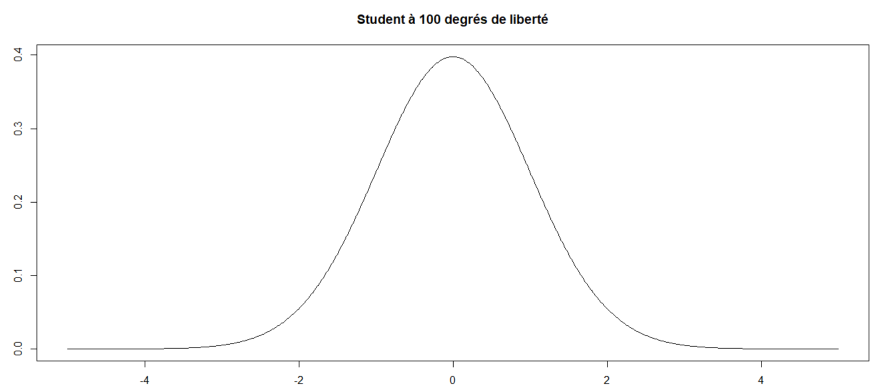
Les propriétés que nous utiliserons sont résumées dans la proposition suivante :
Proposition 4
1. $E(T_n) = 0$ (**symétrie** comme la loi normale centrée réduite),
2. $V(T_n) = \dfrac{n}{n-2}$ pour $n > 2$.





Théorème 2
T_n converge en loi vers $\mathcal{N}(0,1)$ lorsque n tend vers $+\infty$.

Pour $n = 100$, la représentation graphique de la densité de la loi de Student à n degrés de libertés est « très proche » de la courbe en cloche.



Rappelons maintenant un résultat du chapitre 1.

Théorème 3
Pour tout échantillon $(X_1, \ldots, X_n)$,
$S_n^2 := \dfrac{1}{n-1} \displaystyle\sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ est un estimateur sans biais de σ^2.

Nous admettrons le résultat suivant :

Théorème 4
La variable aléatoire $T_n := \dfrac{\sqrt{n}(\bar{X}_n - m)}{\sqrt{S_n^2}}$ suit une loi de Student à $n-1$ degrés de liberté.
Remarque 1
Pourquoi cette perte d'un degré de liberté ? Il est aisément compréhensible que les n variables aléatoires $X_i - \bar{X}_n$ ne sont pas indépendantes car $\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X}_n) = 0$ d'où la perte d'un degré de liberté.
En suivant la même méthode que dans le cas précédent,

$$\mathbb{P}\left(-t_{1-\frac{\alpha}{2}} \leq T_n \leq t_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

où $t_{1-\frac{\alpha}{2}}$ désigne le fractile d'ordre $1 - \dfrac{\alpha}{2}$ de la loi de Student à $n-1$ degrés de liberté. Nous exploitons ici la symétrie de la loi de Student.
Définition 8
L'intervalle de confiance au niveau $1-\alpha$ pour la moyenne m est :

$$\left[\bar{x}_n - t_{1-\frac{\alpha}{2}}\frac{\sqrt{s_n^2}}{\sqrt{n}}; \bar{x}_n + t_{1-\frac{\alpha}{2}}\frac{\sqrt{s_n^2}}{\sqrt{n}}\right].$$

Exemple 3
Considérons l'échantillon de 6 observations extraites d'une loi normale de moyenne et de variance inconnues : 7, 6, 4, 8, 5 et 9.
Déterminer un intervalle de confiance niveau 0,95 pour la moyenne m.

Calculons tout d'abord $\bar{x}$. On a :

$$\bar{x} = \frac{7+6+4+8+5+9}{6} = 6.5$$

Il faut maintenant calculer s^2.

$$s^2 = \frac{1}{6-1}\left((7-6.5)^2 + (6-6.5)^2 + (4-6.5)^2 + (8-6.5)^2 + (5-6.5)^2 + (9-6.5)^2\right) = 3.5$$

Nous aurons recours au fractile $t_{0,975} \approx 2,57$ où $t_{0,975}$ désigne le fractile de la loi de Student à $6-1=5$ degrés de liberté d'ordre 0.975.
L'intervalle de confiance pour la moyenne au niveau 0,95 est :

$$\left[6.5 - 2.57\frac{\sqrt{3.5}}{\sqrt{6}}; 6.5 + 2.57\frac{\sqrt{3.5}}{\sqrt{6}}\right]$$

i.e. approximativement l'intervalle $[4.54; 8.46]$.

3 Intervalle de confiance pour la variance σ^2

3.1 Cas 1 : Moyenne connue

Dans cette partie, nous considérons toujours $(X_1, \ldots, X_n)$ un échantillon de taille $n \geq 1$ de la loi $\mathcal{N}(m, \sigma^2)$. Maintenant, nous cherchons à estimer la variance σ^2.
Nous savons que les variables aléatoires $X_i - m$ sont des variables aléatoires normales indépendantes et centrées donc les variables aléatoires $\dfrac{X_i - m}{\sigma^2}$ sont des variables aléatoires indépendantes normales centrées réduites.
Par conséquent, en considérant $S_n^{2\star} := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - m)^2$, $nS_n^{2\star}$ suit une loi Khi-deux à n degrés de liberté.
Comme précédemment,

$$\mathbb{P}\left(\chi^2_{\frac{\alpha}{2}} \leq \frac{nS_n^{2\star}}{\sigma^2} \leq \chi^2_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

où $\chi^2_{\frac{\alpha}{2}}$ et $\chi^2_{1-\frac{\alpha}{2}}$ désignent respectivement le fractile d'ordre $\frac{\alpha}{2}$ et le fractile d'ordre $1 - \frac{\alpha}{2}$ de la loi $\chi^2(n)$.
Notons qu'ici, nous avons besoin des deux fractiles $\chi^2_{\frac{\alpha}{2}}$ et $\chi^2_{1-\frac{\alpha}{2}}$ car la loi $\chi^2(n)$ n'est pas symétrique.

Définition 9
L'*intervalle de confiance* au niveau $1 - \alpha$ pour la variance σ^2 est :

$$\left[n\frac{s_n^{2\star}}{\chi^2_{1-\frac{\alpha}{2}}}; n\frac{s_n^{2\star}}{\chi^2_{\frac{\alpha}{2}}}\right]$$

Exemple 4
Considérons l'échantillon de 8 observations extraites d'une loi normale de moyenne 2 et de variance inconnue : 3, 1, 5, 8, 5, 10, 12 et 9.
Déterminer un intervalle de confiance niveau 0,90 pour la variance σ^2.

Calculons $s^{2\star}$.
$s^{2\star} = \dfrac{1}{8}\left((3-2)^2 + (1-2)^2 + (5-2)^2 + (8-2)^2 + (5-2)^2 + (10-2)^2 + (12-2)^2 + (9-2)^2\right)$
donc $s^{2\star} \approx 33,63$.

Nous aurons recours aux fractiles de la loi $\chi^2(8)$:
$\chi^2_{0,05} \approx 2,733$ et $\chi^2_{0,95} \approx 15,51$.
Nous en déduisons l'intervalle de confiance niveau 0,90 pour la variance σ^2 :
$\left[8 \times \dfrac{33,63}{15,51}; 8 \times \dfrac{33,63}{2,733}\right]$ i.e. $[17,35; 98,44]$.

3.2 Cas 2 : Moyenne inconnue

Ne connaissant pas m, nous allons le remplacer par $\bar{X}_n$ et remplacer $S_n^{2\star}$ par S_n^2. Rappelons que $S_n^2 := \dfrac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Nous admettrons le théorème suivant :
Théorème 5
Pour tout échantillon $(X_1, \ldots, X_n)$,
$\dfrac{(n-1)S_n^2}{\sigma^2}$ suit une loi $\chi^2(n-1)$.
L'intervalle de confiance sera de forme similaire à celui du cas précédent.
Définition 10
L'*intervalle de confiance* au niveau $1 - \alpha$ pour la variance σ^2 est :

$$\left[(n-1)\frac{s_n^2}{\chi^2_{1-\frac{\alpha}{2}}}; (n-1)\frac{s_n^2}{\chi^2_{\frac{\alpha}{2}}}\right]$$

où $\chi^2_{\frac{\alpha}{2}}$ et $\chi^2_{1-\frac{\alpha}{2}}$ désignent respectivement le fractile d'ordre $\frac{\alpha}{2}$ et le fractile d'ordre $1 - \frac{\alpha}{2}$ de la loi $\chi^2(n-1)$.
Exemple 5
Considérons l'échantillon de 8 observations extraites d'une loi normale de moyenne et de variance inconnues : 3, 1, 5, 8, 5, 10, 12 et 9.
Déterminer un intervalle de confiance au niveau 0,95 pour la variance σ^2.

Calculons tout d'abord $\bar{x}$.
On a :

$$\bar{x} = \frac{3 + 1 + 5 + 8 + 5 + 10 + 12 + 9}{8} = 6,625$$

Il faut maintenant calculer s^2.

$$s^2 = \frac{1}{7}((3 - 6,625)^2 + (1 - 6,625)^2 + (5 - 6,625)^2 + (8 - 6,625)^2 + (5 - 6,625)^2 + (10 - 6,625)^2 + (12 - 6,625)^2 + (9 - 6,625)^2)$$

donc $s^2 \approx 13,98$.

Nous aurons recours aux fractiles de la loi $\chi^2(7)$: $\chi^2_{0,025} \approx 1,690$ et $\chi^2_{0,975} \approx 16,01$.

Nous en déduisons l'intervalle de confiance pour la variance σ^2 au niveau 0,95 : $\left[7 \times \frac{13,98}{16,01}; 7 \times \frac{13,98}{1,690}\right]$ i.e. approximativement $[6,11; 57,91]$.

4 Intervalles de confiance asymptotiques

4.1 Intervalle de confiance pour la moyenne m

Dans ce paragraphe, la variance est supposée connue. Ainsi, nous nous placerons dans le cas général où nous chercherons à estimer la moyenne m d'une variable aléatoire X de carré intégrable dont nous connaissons la variance σ^2(et donc l'écart-type σ). Rappelons le théorème central limite.

Théorème 6 (Théorème Central Limite)
Soit (X_i) une suite de v.a. i.i.d. telle que $E(X_1^2) < +\infty$.
Notons $m := E(X_i)$ et $\sigma^2 = V(X_i)$.
$\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$ converge en loi vers une loi normale centrée réduite.

Par conséquent, pour n assez grand [4],

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \leq z_{1-\frac{\alpha}{2}}\right) \approx 1 - \alpha$$

avec les notations précédentes. Ainsi, nous pourrons définir la notion d'intervalle de confiance asymptotique.

Définition 11 (Variance connue)
L'*intervalle de confiance asymptotique* au niveau de confiance $1 - \alpha$ pour la moyenne m est :

$$\left[\bar{x}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \bar{x}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right].$$

Notons qu'ici, nous utilisons la variance σ^2 que nous connaissons.
Dans le cas où elle est inconnue, nous aurons recours au théorème suivant que nous admettrons :

4. En pratique, l'approximation est souvent jugée acceptable pour $n \geq 30$.

Théorème 7
Soit (X_i) une suite de v.a. i.i.d. telle que $E(X_1^2) < +\infty$.
$\dfrac{\sqrt{n}(\bar{X}_n - m)}{\sqrt{S_n^2}}$ converge en loi vers une loi normale centrée réduite.

Par conséquent, nous obtenons l'intervalle de confiance asymptotique suivant :

Définition 12 (Variance inconnue)
L'*intervalle de confiance asymptotique* au niveau de confiance $1 - \alpha$ pour la moyenne m est :

$$\left[\bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sqrt{s_n^2}}{\sqrt{n}}; \bar{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sqrt{s_n^2}}{\sqrt{n}}\right].$$

Nous pourrons utiliser des intervalles plus simples lorsque la variance est fonction de la moyenne. Nous aborderons deux cas :

- estimation du paramètre p d'une variable aléatoire de Bernoulli,
- estimation du paramètre λ d'une loi de Poisson.

4.2 Intervalle de confiance pour la proportion

Dans cette partie, nous proposerons un intervalle de confiance pour la proportion de personnes porteuses d'un caractère dans une population donnée. Chaque personne est porteuse ou non donc une sous-population individus pourra être modélisée [5] par un échantillon de taille n de la loi de Bernoulli où $X_i = 1$ si la personne est porteuse du caractère et 0 sinon.
Nous notons $\hat{p}$ la proportion de personnes porteuses du caractère dans l'échantillon, elle correspond à la moyenne empirique des variables aléatoires de Bernoulli associées.
Rappelons le théorème suivant vu dans le chapitre 1 :

Théorème 8 (Théorème de Moivre-Laplace)
Soit X_n une variable aléatoire suivant une loi $\mathcal{B}(n, p)$.
Posons $q := 1 - p$.

5. Cette modélisation n'est valide que lorsque la population est suffisamment importante.cf. [5]

Pour tout réel x,

$$\lim_{n \longrightarrow +\infty} \mathbb{P}\left(\frac{X_n - np}{\sqrt{npq}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt = F(x)$$

où F désigne la fonction de répartition de la loi normale centrée réduite.

Ce théorème implique que la loi binomiale peut être approchée par la loi normale pour n suffisamment grand. Rappelons le jeu de conditions d'approximation donné dans le chapitre 1 :

1. $n \geq 30$,
2. $n\hat{p} \geq 5$,
3. $n(1-\hat{p}) \geq 5$.

Définition 13
L'*intervalle de confiance asymptotique* au niveau $1-\alpha$ pour la proportion p est :

$$\left[\hat{p} - z_{1-\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} ; \hat{p} + z_{1-\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right]$$

où $z_{1-\frac{\alpha}{2}}$ désigne le fractile d'ordre $1-\dfrac{\alpha}{2}$ de la loi normale centrée réduite.

Exemple 6
François prélève 300 serpents dans une forêt et constate que 70 d'entre eux sont venimeux.
Déterminer un intervalle de confiance asymptotique pour la proportion de serpents venimeux dans cette forêt au niveau de confiance $0,95$.

Calculons tout d'abord $\hat{p}$.
Par définition, $\hat{p} = \dfrac{70}{300} \approx 0,23$.
Vérifions les conditions d'approximation par la loi normale :

1. $n = 300 \geq 30$,
2. $n\hat{p} = 300 \times 0,23 = 69 \geq 5$,
3. $n(1-\hat{p}) = 231 \geq 5$.

Nous aurons recours au fractile $z_{0,975} \approx 1,96$.
L'intervalle de confiance asymptotique pour la proportion de serpents venimeux dans cette forêt au niveau 0,95 est :

$$\left[0,23 - 1,96\frac{\sqrt{0,23 \times 0,77}}{\sqrt{300}}; 6 + 1,96\frac{\sqrt{0,23 \times 0,77}}{\sqrt{300}}\right]$$

i.e. approximativement l'intervalle $[0,18; 0,28]$.

4.3 Intervalle de confiance pour le paramètre λ d'une loi de Poisson

Rappelons que, pour, une variable aléatoire X qui suit une loi $\mathcal{P}(\lambda)$, $E(X) = V(X) = \lambda$.
Par conséquent, le théorème central limite appliqué dans ce cas donne le théorème suivant :

Théorème 9
Soit (X_i) une suite de variables aléatoires i.i.d. de loi $\mathcal{P}(\lambda)$,
$\sqrt{n}\frac{\bar{X}_n - \lambda}{\sqrt{\lambda}}$ converge en loi vers une loi normale centrée réduite.

Par conséquent, pour n assez grand

$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n}\frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \leq z_{1-\frac{\alpha}{2}}\right) \approx 1 - \alpha$$

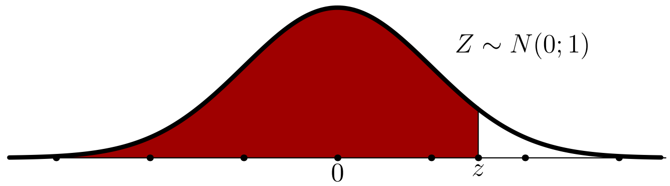
où $z_{1-\frac{\alpha}{2}}$ désigne le fractile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale centrée réduite.

Définition 14
L'*intervalle de confiance asymptotique* au niveau $1 - \alpha$ pour le paramètre λ est :

$$\left[\bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sqrt{\bar{x}_n}}{\sqrt{n}}; \bar{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sqrt{\bar{x}_n}}{\sqrt{n}}\right]$$

Tableau N [1]

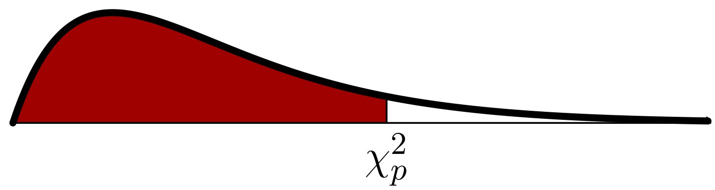
Aire sous la courbe normale à gauche de z, c'est à dire $P[Z \leq z]$, ou $Z \sim N(0;1)$.



	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.00	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.10	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.20	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.30	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.40	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.50	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.60	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.70	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.80	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.90	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09

Tableau C [1/2]

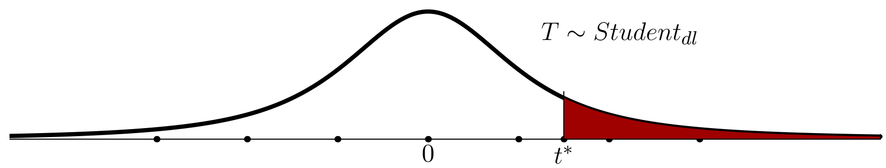
Percentiles de la distribution du χ^2. Valeurs de χ^2_P correspondant à P



dl	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.1}$	$\chi^2_{0.9}$	$\chi^2_{0.95}$	$\chi^2_{0.975}$	$\chi^2_{0.99}$	$\chi^2_{0.995}$
1	.0000	.0002	.0010	.0039	.0158	2.706	3.841	5.024	6.635	7.879
2	.0100	.0201	.0506	.1026	.2107	4.605	5.991	7.378	9.210	10.60
3	.0717	.1148	.2158	.3518	.5844	6.251	7.815	9.348	11.34	12.84
4	.2070	.2971	.4844	.7107	1.064	7.779	9.488	11.14	13.28	14.86
5	.4117	.5543	.8312	1.145	1.610	9.236	11.07	12.83	15.09	16.75
6	.6757	.8721	1.237	1.635	2.204	10.64	12.59	14.45	16.81	18.55
7	.9893	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48	20.28
8	1.344	1.646	2.180	2.733	3.490	13.36	15.51	17.53	20.09	21.95
9	1.735	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67	23.59
10	2.156	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21	25.19
11	2.603	3.053	3.816	4.575	5.578	17.28	19.68	21.92	24.72	26.76
12	3.074	3.571	4.404	5.226	6.304	18.55	21.03	23.34	26.22	28.30
13	3.565	4.107	5.009	5.892	7.042	19.81	22.36	24.74	27.69	29.82
14	4.075	4.660	5.629	6.571	7.790	21.06	23.68	26.12	29.14	31.32
15	4.601	5.229	6.262	7.261	8.547	22.31	25.00	27.49	30.58	32.80
16	5.142	5.812	6.908	7.962	9.312	23.54	26.30	28.85	32.00	34.27
17	5.697	6.408	7.564	8.672	10.09	24.77	27.59	30.19	33.41	35.72
18	6.265	7.015	8.231	9.390	10.86	25.99	28.87	31.53	34.81	37.16
dl	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.1}$	$\chi^2_{0.9}$	$\chi^2_{0.95}$	$\chi^2_{0.975}$	$\chi^2_{0.99}$	$\chi^2_{0.995}$

Tableau T1 [1/2]

Tableau de t^* tel qu'une variable de Student à dl degrés de liberté ait probabilité p d'être supérieure à t^*



	$P[T \geq t^*] = p$											
dl	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.8165	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.7649	.9785	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.7407	.9410	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.7267	.9195	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.7176	.9057	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.7111	.8960	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.7064	.8889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.7027	.8834	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.6998	.8791	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.6974	.8755	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.6955	.8726	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.6938	.8702	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.6924	.8681	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.6912	.8662	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.6901	.8647	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.6892	.8633	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.6884	.8620	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005

Références

[1] Dacunha-Castelle, D. : *Chemins de l'aléatoire.* Flammarion (2000)

[2] Foata, D., Franchi, J. Fuchs, A. : *Calcul de probabilités.* Dunod (2012)

[3] Monfort, A. : *Cours de statistique mathématique.* Economica (1997)

[4] Revuz, D. : *Probabilités*. Hermann (1997)

[5] Tassi, P. : *Méthodes statistiques*. Economica (2004)

[6] Toulouse, P.S. : *Thèmes de probabilités et de statistiques*. Dunod (1999)

[7] Vessereau, A. : *La Statistique*. Que sais-je ? Presses universitaires de France (1947)

[8] Williams, D. : *Probability with Martingales*. Cambridge Mathematical Textbooks (1991)