

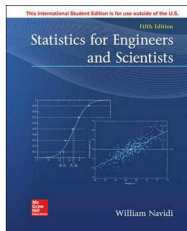
Introduction aux statistiques pour l'Ingénieur

Estimation par intervalle

Stéphane Canu

asi.insa-rouen.fr/enseignants/~scanu

scanu@insa-rouen.fr



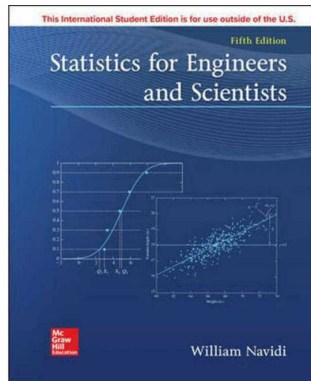
ITI 3, INSA Rouen Normandie, avril 2024

Lecture road map

1 Rappels

2 Estimation par intervalle

- Estimation par intervalle
- La loi du χ^2
 - Définition
 - Propriétés et approximation
- La loi de Student
 - Définition
 - Propriétés et approximation



<https://moodle.insa-rouen.fr/course/view.php?id=93>

Estimateur et vraisemblance

Le modèle : X de loi parente de distribution $f_\theta(x)$

Échantillon i.i.d.

$$(X_1, \dots, X_n)$$

Vraisemblance :

$$L(\theta, X_1, \dots, X_n) = \prod_{i=1}^n f_\theta(X_i)$$

Log vraisemblance :

$$\ell(\theta, X_1, \dots, X_n) = \log L(\theta, X_1, \dots, X_n) = \sum_{i=1}^n \log f_\theta(X_i)$$

Estimateur du max de vraisemblance

$$\hat{\theta}_{MV} = \arg \max_{\theta} \ell(\theta, X_1, \dots, X_n) \Leftrightarrow \left. \frac{\partial \ell(\theta, X_1, \dots, X_n)}{\partial \theta} \right|_{\theta = \hat{\theta}_{MV}} = 0$$

Information de Fisher

X de loi parente de distribution $f_{\theta}(x)$, échantillon i.i.d. et vraisemblance

$$(X_1, \dots, X_n) \quad L(\theta, X_1, \dots, X_n) = \prod_{i=1}^n f_{\theta}(X_i)$$

Log vraisemblance :

$$\ell(\theta, X_1, \dots, X_n) = \log L(\theta, X_1, \dots, X_n) = \sum_{i=1}^n \log f_{\theta}(X_i)$$

Estimateur du max de vraisemblance

$$\hat{\theta}_{MV} = \arg \max_{\theta} \ell(\theta, X_1, \dots, X_n) \quad \Leftrightarrow \quad \left. \frac{\partial \ell(\theta, X_1, \dots, X_n)}{\partial \theta} \right|_{\theta = \hat{\theta}_{MV}} = 0$$

Information de Fisher

$$I_n(\theta) = \text{Var} \left(\frac{\partial \ell(\theta, X_1, \dots, X_n)}{\partial \theta} \right) = -\mathbb{E} \left(\frac{\partial^2 \ell(\theta, X_1, \dots, X_n)}{\partial \theta^2} \right)$$

Fiche recette

❶ Vraisemblance : $L(\theta, X_1, \dots, X_n) = \prod_{i=1}^n f_{\theta}(X_i)$

❷ Log vraisemblance : $\ell(\theta, X_1, \dots, X_n) = \log L = \sum_{i=1}^n \log f_{\theta}(X_i)$

❸ Calcul de la dérivée de la log vraisemblance $\frac{\partial \ell(\theta, X_1, \dots, X_n)}{\partial \theta}$

S'il existe un estimateur $\hat{\theta}(X_1, \dots, X_n)$ de θ tel que

$$\frac{\partial \ell(X_1, \dots, X_n)}{\partial \theta} = A(n, \theta)(\hat{\theta}(X_1, \dots, X_n) - \theta)$$

Alors $\hat{\theta}(X_1, \dots, X_n)$ est un estimateur sans biais de risque minimal (UMUE) et son information de Fisher est :

$$I_n(\theta) = A(n, \theta)$$

Exemples

- Loi de Bernouilli
- Loi exponentielle
- Loi de Poisson
- Loi normale

Exemple (1) : la loi de Bernouilli

$$f(x) = p^x(1-p)^{1-x}$$

$$L(\lambda, X_1, \dots, X_n) = \prod_{i=1}^n p^{X_i}(1-p)^{1-X_i}$$

$$\ell(\lambda, X_1, \dots, X_n) = \log p \sum_{i=1}^n X_i + \log(1-p)(n - \sum_{i=1}^n X_i)$$

$$\begin{aligned} \frac{\partial \ell(\lambda, X_1, \dots, X_n)}{\partial \lambda} &= \frac{\sum_{i=1}^n X_i}{p} - \frac{n - \sum_{i=1}^n X_i}{1-p} \\ &= \frac{n}{p(1-p)} \left(\frac{1}{n} \sum_{i=1}^n X_i - p \right) \end{aligned}$$

$$\hat{\theta}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i, \quad A(n, p) = \frac{n}{p(1-p)}, \quad \text{Var}(\hat{\theta}) = \frac{p(1-p)}{n}$$

Exemple (2) : la loi exponentielle

$$f(x) = \lambda \exp^{-\lambda x}$$

$$L(\lambda, X_1, \dots, X_n) = \lambda^n \prod_{i=1}^n \exp^{-\lambda X_i}$$

$$\ell(\lambda, X_1, \dots, X_n) = n \log \lambda - \lambda \sum_{i=1}^n X_i$$

$$\begin{aligned} \frac{\partial \ell(\lambda, X_1, \dots, X_n)}{\partial \lambda} &= \frac{n}{\lambda} - \sum_{i=1}^n X_i \\ &= -n \left(\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{\lambda} \right) \end{aligned}$$

$$u(\lambda) = \frac{1}{\lambda}, \quad \hat{u}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\text{Var}(\hat{u}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = -\frac{u'(\lambda)}{A(n, \lambda)} = \frac{1}{n\lambda^2},$$

Exemple (3) : la loi de poisson

$$\begin{aligned}\mathbb{P}(X = x) &= \frac{\lambda^x}{x!} \exp^{-\lambda} \\ L(\lambda, X_1, \dots, X_n) &= \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} \exp^{-n\lambda} \\ \ell(\lambda, X_1, \dots, X_n) &= \log \lambda \sum_{i=1}^n X_i - \sum_{i=1}^n X_i! - n\lambda \\ \frac{\partial \ell(\lambda, X_1, \dots, X_n)}{\partial \lambda} &= \frac{1}{\lambda} \sum_{i=1}^n X_i - n \\ &= \frac{n}{\lambda} \left(\frac{1}{n} \sum_{i=1}^n X_i - \lambda \right)\end{aligned}$$

$$\hat{\lambda}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i, \quad A(n, p) = \frac{n}{\lambda}, \quad \text{Var}(\hat{\lambda}) = \frac{\lambda}{n}$$

Exemple (4) : la loi normale

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$L(\mu, X_1, \dots, X_n) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

$$\ell(\mu, X_1, \dots, X_n) = n \log \frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

$$\begin{aligned} \frac{\partial \ell(\mu, X_1, \dots, X_n)}{\partial \mu} &= \frac{1}{\sigma^2} \left(\sum_{i=1}^n x_i - n\mu \right) \\ &= \frac{n}{\sigma^2} \left(\frac{1}{n} \sum_{i=1}^n X_i - \mu \right) \end{aligned}$$

$$\hat{\mu}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i, \quad A(n, p) = \frac{n}{\sigma^2}, \quad \text{Var}(\hat{\mu}) = \frac{\sigma^2}{n}$$

Exemple (5) : la loi normale

$$\begin{aligned}f(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \\L(\sigma^2, X_1, \dots, X_n) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} \\\ell(\sigma^2, X_1, \dots, X_n) &= n \log \frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \\\frac{\partial \ell(\sigma^2, X_1, \dots, X_n)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \\&= \frac{n}{2\sigma^4} \left(\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - \sigma^2 \right)\end{aligned}$$

$$\hat{\sigma}^2(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2, \quad A(n, p) = \frac{n}{2\sigma^4}, \quad \text{Var}(\hat{\sigma}) = \frac{2\sigma^4}{n}$$

Propriété des estimateurs max de vraisemblance

Si

- la probabilité ou de la densité dépend d'un paramètre θ
- le support de la probabilité ou de la densité ne dépend pas de θ
- si I existe, est inversible et quelques autres conditions plus techniques

Asymptotiquement sans biais

$$\hat{\theta}_{MV} \xrightarrow[n \rightarrow \infty]{} \theta^*$$

Asymptotiquement efficace

$$\text{Var}(\hat{\theta}_{MV}) \xrightarrow[n \rightarrow \infty]{} I_n^{-1}(\theta^*)$$

Asymptotiquement normal

$$\sqrt{n}(\hat{\theta}_{MV} - \theta^*) \xrightarrow[n \rightarrow \infty]{} \mathcal{N}(0, I_1^{-1}(\theta^*))$$

On a pour l'information de Fisher : $I_n(\theta^*) = nI_1(\theta^*)$

Propriété des estimateurs max de vraisemblance

Asymptotiquement normal

$$\sqrt{n}(\hat{\theta}_{MV} - \theta^*) \xrightarrow[n \rightarrow \infty]{} \mathcal{N}(0, I_1^{-1}(\theta^*))$$

Approximation normale, si

- n est grand,

la loi de $\hat{\theta}_{MV}$ peut être approchée par

$$\mathcal{N}(\theta^*, I_n^{-1}(\theta^*))$$

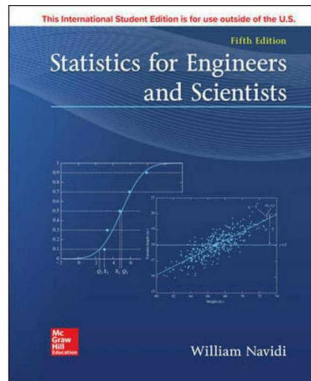
I_n étant l'information de Fisher

Lecture road map

1 Rappels

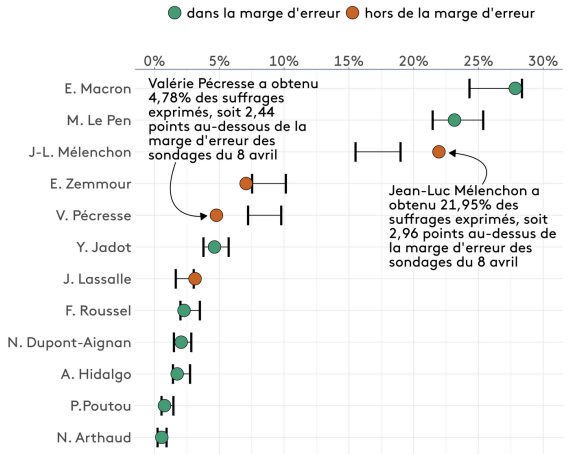
2 Estimation par intervalle

- Estimation par intervalle
- La loi du χ^2
 - Définition
 - Propriétés et approximation
- La loi de Student
 - Définition
 - Propriétés et approximation



<https://moodle.insa-rouen.fr/course/view.php?id=93>

Estimation par intervalle



Liste des sondages utilisés : Opinion Way (Les Echos), Ifop (Paris Match), Elabe (Bfm), Ipsos (Le Monde), Harris (Challenges), Ipsos (Le Parisien, Franceinfo)

Sources : NspPolls, ministère de l'Intérieur - Crédits : franceinfo

Principe de l'estimation par intervalle

Au lieu de chercher **un seul** estimateur (comme $\hat{\theta}_{MV}(x_1, \dots, x_n)$),
On en cherche **deux estimateurs**

$$\hat{B}_i(X_1, \dots, X_n) \quad \text{et} \quad \hat{B}_s(X_1, \dots, X_n)$$

de sorte que **avec grande probabilité**

$$\hat{B}_i(X_1, \dots, X_n) \leq \theta^* \leq \hat{B}_s(X_1, \dots, X_n)$$

Objectif : un intervalle de confiance

- est plus « réaliste » qu'une seule valeur (*cf les sondages*)
- *aide à prendre une décision*
- *permet la détection d'observations aberrantes*

C'est ce qu'on appelle

Intervalle de confiance

Qu'est-ce qu'un « bon » intervalle de confiance ?

Intervalle de confiance
Intervalle de prédiction

$$\hat{\theta} \pm f_{\alpha/2} \times \text{Var}(\hat{\theta})$$

estimation valeur critique variabilité

Estimation de l'espérance σ^2 connue (loi normale)

$$\hat{\theta}_{MV}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

On a donc

$$\mathbb{P}\left(-1,96 \leq \frac{\hat{\theta}_{MV} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1,96\right) = 0,95$$

soit

$$\mathbb{P}\left(\underbrace{\hat{\theta}_{MV} - 1,96 \frac{\sigma}{\sqrt{n}}}_{\hat{B}_i(X_1, \dots, X_n)} \leq \mu \leq \underbrace{\hat{\theta}_{MV} + 1,96 \frac{\sigma}{\sqrt{n}}}_{\hat{B}_s(X_1, \dots, X_n)}\right) = 0,95$$

Que peut-on en déduire ?

Si l'on répète de nombreuses fois l'enquête, on obtiendra des intervalles différents qui contiendront, dans 95% des cas, la valeur du paramètre μ .

Estimateurs max de vraisemblance (n grand)

Soit $\hat{\theta}_{MV}$ l'estimateur max de vraisemblance d'un paramètre θ .

Approximation normale, si n est grand,
la loi de $\hat{\theta}_{MV}$ peut être approchée par

$$\mathcal{N}(\theta^*, I_n^{-1}(\theta^*))$$

I_n étant l'information de Fisher

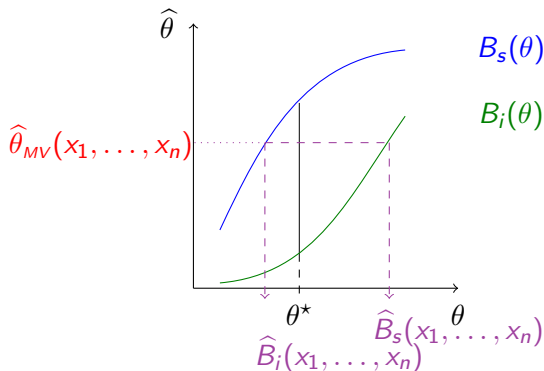
$$\mathbb{P}\left(\underbrace{\hat{\theta}_{MV} - 1,96 \sqrt{I_n^{-1}(\theta^*)}}_{\hat{B}_i(X_1, \dots, X_n)} \leq \theta^* \leq \underbrace{\hat{\theta}_{MV} + 1,96 \sqrt{I_n^{-1}(\theta^*)}}_{\hat{B}_s(X_1, \dots, X_n)}\right) \approx 0,95$$

$$\mathbb{P}\left(\underbrace{\hat{\theta}_{MV} + u_{\alpha/2} \sqrt{I_n^{-1}(\theta^*)}}_{\hat{B}_i(X_1, \dots, X_n)} \leq \theta^* \leq \underbrace{\hat{\theta}_{MV} + u_{1-\alpha/2} \sqrt{I_n^{-1}(\theta^*)}}_{\hat{B}_s(X_1, \dots, X_n)}\right) \approx 1-\alpha$$

Interprétation graphique

On veut
$$\mathbb{P}\left(\underbrace{\hat{\theta}_{MV} - 1,96 \frac{\sigma}{\sqrt{n}}}_{\hat{B}_i(X_1, \dots, X_n)} \leq \mu \leq \underbrace{\hat{\theta}_{MV} + 1,96 \frac{\sigma}{\sqrt{n}}}_{\hat{B}_s(X_1, \dots, X_n)}\right) = 0,95$$

On a
$$\mathbb{P}\left(\underbrace{\mu - 1,96 \frac{\sigma}{\sqrt{n}}}_{B_i(\theta)} \leq \hat{\theta}_{MV} \leq \underbrace{\mu + 1,96 \frac{\sigma}{\sqrt{n}}}_{B_s(\theta)}\right) = 0,95$$



Principe général : fonction pivotale

$$\phi(\hat{\theta}, \theta) \sim \mathcal{L}$$

Exemple :

- pour la loi parente Normale (variance connue, $\sigma = 1$)

$$X \sim \mathcal{N}(\mu, \sigma)$$

- la fonction pivotale de la moyenne dépend de la variance

$$\phi(\hat{\theta} = \bar{X}, \mu) = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$$

La réalité est (malheureusement) souvent plus complexe...

Définition de l'intervalle de confiance

Définition (Intervalle de confiance)

Soit $X_1, \dots, X_i, \dots, X_n$ un échantillon i.i.d dont la loi parente dépend d'un paramètre θ . Soit $\alpha \in [0, 1]$ un niveau de défiance fixé.

On appelle intervalle de confiance du paramètre θ tout intervalle $[\hat{B}_i, \hat{B}_s]$ où $\hat{B}_i(X_1, \dots, X_n)$ et $\hat{B}_s(X_1, \dots, X_n)$ sont deux statistiques telles que

$$\mathbb{P} \left(\hat{B}_i(X_1, \dots, X_n) \leq \theta \leq \hat{B}_s(X_1, \dots, X_n) \right) = 1 - \alpha$$

- On ne peut pas dire qu'une réalisation de l'intervalle de confiance à une probabilité de $1 - \alpha$ de contenir θ .
- Il n'y a pas d'unicité de l'intervalle. Il est raisonnable de rechercher l'intervalle de plus petite taille.
- Parfois on peut s'intéresser à l'une des deux bornes, on parle alors d'un intervalle unilatéral.

Plan

1 Rappels

2 Estimation par intervalle

- Estimation par intervalle
- La loi du χ^2
 - Définition
 - Propriétés et approximation
- La loi de Student
 - Définition
 - Propriétés et approximation

La loi du χ^2

Soit $Y \sim \mathcal{N}(0, 1)$ une variable aléatoire normale centrée réduite. Soit Y_1, Y_2, \dots, Y_n un échantillon de n réalisations i.i.d. de cette variable aléatoire.

Definition (La loi du χ^2)

On appelle loi du χ^2 à n degrés de libertés la loi de la variable aléatoire Z_n

$$Z_n = \sum_{i=1}^n Y_i^2$$

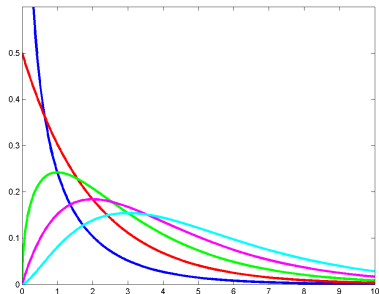


Figure: Exemples de loi du chi 2 pour 1 (bleu), 2 (rouge), 3 (vert), 4 (violet) et 5 (bleu ciel) degrés de liberté

Propriétés et approximation

$$\mathbb{E}(Z_n) = n \quad \text{Var}(Z_n) = 2n \quad \text{mode}(Z_n) = n - 2$$

En effet :

$$\mathbb{E}(Z_n) = \mathbb{E}\left(\sum_{i=1}^n Y_i^2\right) = n\mathbb{E}(Y^2) = nV(Y) = n$$

Lorsque le nombre de degrés de libertés est important on peut utiliser une approximation asymptotique de la loi du chi 2. Les plus utilisées sont les approximations de Paul Levy et de Fisher :

$$\text{Levy : } \frac{Z_n - n}{\sqrt{2n}} \rightarrow \mathcal{N}(0, 1) \quad \text{Fisher : } \sqrt{2Z_n} - \sqrt{2n - 1} \rightarrow \mathcal{N}(0, 1)$$

Il existe aussi une loi du chi 2 dite décentrée. C'est la somme de carrés de carrés d'une variable gaussienne non centrée. Nous ne l'utiliserons pas dans ce cours.

si X_1, X_2, \dots, X_n est un échantillon de n réalisations i.i.d. d'une variable aléatoire normale $\mathcal{N}(\mu, \sigma^2)$ d'espérance μ et la variance σ^2 .

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$$

puisque $Y_i = \frac{X_i - \mu}{\sigma}$ suit une loi normale centrée réduite.

Il est moins évident de montrer que : $\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$

Lorsque l'on remplace le paramètre μ par son estimation $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ on perd un degré de liberté. En effet on a la décomposition suivante :

$$\underbrace{\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2}}_{\chi_n^2} = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} + \underbrace{n \frac{(\bar{X} - \mu)^2}{\sigma^2}}_{\chi_1^2}$$

Qui permet de conclure en invoquant le théorème de Cochran sur l'additivité des degrés de liberté.

L'estimation de la variance d'une loi normale

$X \sim \mathcal{N}(\mu, \sigma^2)$ peut σ^2 inconnus

$\hat{\sigma}_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ avec $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

QUELLE est la Loi de Cet estimateur ?

$\frac{(n-1)}{\sigma^2} \hat{\sigma}_{n-1}^2 \sim \chi_{n-1}^2$

$P\left(\chi_{n-1}^2 \leq \frac{n-1}{\sigma^2} \hat{\sigma}_{n-1}^2 \leq \chi_{n-1}^2\right) = 1-\alpha$

$P\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{n-1}^2} \leq \sigma^2 \leq \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{n-1}^2}\right) = 1-\alpha$

Plan

1 Rappels

2 Estimation par intervalle

- Estimation par intervalle
- La loi du χ^2
 - Définition
 - Propriétés et approximation
- La loi de Student
 - Définition
 - Propriétés et approximation

La loi de Student : définition

- Soit $N \sim \mathcal{N}(0, 1)$ une variable aléatoire normale centrée réduite.
- Soit X_n la variable aléatoire distribuée suivant une loi du χ^2 à n ddl
 - ▶ C'est le cas par exemple, si N_1, N_2, \dots, N_n un échantillon de n réalisation i.i.d. une variable aléatoire normale centrée réduite quand $X_n = \sum_{i=1}^n N_i^2$
- supposons que N et X_n sont indépendantes (*i.e.* $\text{cov}(Y, X_n) = 0$)

Definition (La loi de student)

On appelle loi de student à n degrés de libertés la loi de la variable aléatoire \mathcal{T}_n

$$\mathcal{T}_n = \frac{N}{\sqrt{\frac{X_n}{n}}}$$

$$N \sim \mathcal{N}(0, 1)$$

$$X_n \sim \chi_n^2$$

La loi de Student : $\mathcal{T}_n = \frac{N}{\sqrt{\frac{X_n}{n}}}$

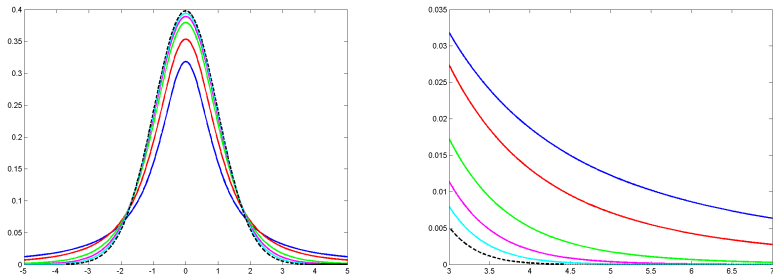


Figure: Exemples de loi de student pour 1 (bleu), 2 (rouge), 5 (vert), 10 (violet) et 20 (bleu ciel) degrés de liberté. La courbe en pointillés noir est la courbe de Gauss donnée comme référence. La figure de droite montre un zoom sur la « queue » de la distribution.

Loi de Student et loi normale

$$\mathcal{T}_n \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, 1)$$

Propriétés et approximation

- Publiée pour la première fois en 1908 par William Sealy Gosset qui travaillait chez Guinness (la brasserie de Dublin). Pour des raisons commerciales, il a du utiliser le pseudonyme de Student, qui restera attaché à cette loi.
- tend vers une loi normale $n > 30$
- attention la différence est plus importante dans les « queue » de la distribution :

▶ $N \sim \mathcal{N}(0, 1) : \mathbb{P}(N > 2) = 0,023$	$p1 = 1 - \text{cdf}('norm', 2, 0, 1)$
▶ $T \sim \mathcal{T}_1 : \mathbb{P}(T > 2) = 0,148$	$p2 = 1 - \text{cdf}('t', 2, 1)$
▶ $T \sim \mathcal{T}_2 : \mathbb{P}(T > 2) = 0,092$	$p2 = 1 - \text{cdf}('t', 2, 2)$
▶ $T \sim \mathcal{T}_{10} : \mathbb{P}(T > 2) = 0,038$	$p2 = 1 - \text{cdf}('t', 2, 10)$

$$U \sim \mathcal{N}(0, \sigma^2) \quad N = \frac{U}{\sigma} \sim \mathcal{N}(0, 1)$$

$$T = \frac{N}{\hat{\sigma}} = \frac{N}{\sqrt{\frac{N_1^2 + N_2^2}{2}}} \sim \mathcal{T}_2$$

Intervalle sur la moyenne d'une loi normale

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$$

$$\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$$

Rappel

$$\mathcal{T}_n = \frac{N}{\sqrt{\frac{X_n}{n}}}$$

$$N \sim \mathcal{N}(0, 1)$$

$$X_n \sim \chi_n^2$$

La fonction pivotale

$$\frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{(n-1)\sigma^2}}} = \frac{\bar{X} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim \mathcal{T}_{n-1}$$

avec

$$\hat{\sigma} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Estimation par intervalle d'une proportion n grand

D'après l'approximation normale de la binomiale $\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$,
et donc

$$\mathbb{P}\left(u_{\alpha/2} \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq -u_{\alpha/2}\right) \approx 1 - \alpha$$

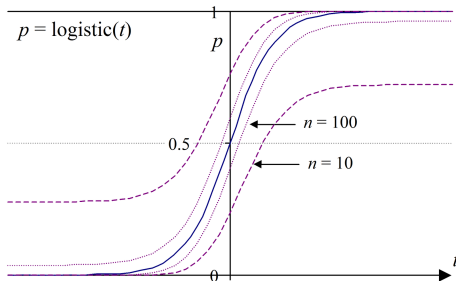
ce qui suggère d'utiliser comme fonction pivotale

$$u_{\alpha/2} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

qui donne comme bornes (Wilson score interval):

$$\frac{\hat{p} + \frac{u_{\alpha/2}^2}{2n}}{1 + \frac{u_{\alpha/2}^2}{n}} \pm \frac{u_{\alpha/2}}{1 + \frac{u_{\alpha/2}^2}{n}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n} + \frac{u_{\alpha/2}^2}{4n^2}}$$

Intervalle sur l'estimation d'une proportion



https://en.wikipedia.org/wiki/Binomial_proportion_confidence_interval

Intervalle sur l'estimation d'une proportion



statsmodels 0.14.0 (+380)

Search

statsmodels 0.14.0 (+380)

Installing statsmodels

Getting started

User Guide

Background

Regression and Linear Models

Time Series Analysis

Other Models

Statistics and Tools

Statistics **stats**

Contingency tables

Multiple Imputation with
Chained Equations

Empirical Likelihood
eml

Distributions

Graphics

Input-Output **io**

Tools

Working with Large Data Sets

Optimization

Data Sets

Sandbox

Examples

API Reference

About statsmodels

Developer Page

Release Notes

statsmodels.stats.proportion.proportion_confint

`statsmodels.stats.proportion.proportion_confint(count, nobs, alpha=0.05, method='normal')`[\[source\]](#)

Confidence interval for a binomial proportion

Parameters

count : {int, array_like}

number of successes, can be pandas Series or DataFrame. Arrays must contain integer values.

nobs : {int, array_like}

total number of trials. Arrays must contain integer values.

alpha : float

Significance level, default 0.05. Must be in (0, 1)

method : {"normal", "agresti_coull", "beta", "wilson", "binom_test"}

default: "normal" method to use for confidence interval. Supported methods:

- *normal* : asymptotic normal approximation
- *agresti_coull* : Agresti-Coull interval
- *beta* : Clopper-Pearson interval based on Beta distribution
- *wilson* : Wilson Score interval
- *jeffreys* : Jeffreys Bayesian Interval
- *binom_test* : Numerical inversion of binom_test

Returns

ci_low, ci_upper : {float, ndarray, Series DataFrame}