



UNIVERSITÉ DE PARAKOU (UP)
ECOLE NATIONALE DE STATISTIQUE, DE PLANIFICATION
ET DE DÉMOGRAPHIE (ENSPD)



Master 1 Statistique, Planification et Démographie (Semestre 8)

ENSPD 2023-2024

Cours de Statistique Nonparamétrique

Enseignant :

Raymond M M AFFOSSOGBE
Docteur en Statistique-Probabilité
Email : rafmoqui@yahoofr
+229 96877413

Table des matières

1	Estimation nonparamétrique de la densité	4
1.1	Estimateur empirique de la fonction de répartition	4
1.1.1	Quantile	5
1.2	Estimateur de Rosenblatt	5
1.3	Estimateur à noyau de Parzen-Rosenblatt	7
1.3.1	Propriétés ponctuelles	7
1.3.2	Propriétés globales	9
1.3.3	Choix optimal de la fenêtre par validation croisée	10
1.4	Estimation par projection	11
1.4.1	Cas de la base trigonométrique	12
1.4.2	Estimateur par ondelettes	13
2	Estimation de la fonction de régression	14
2.1	Estimateur à noyau	15
2.2	Propriétés de l'estimateur de Nadaraya-Waston	16
2.3	Estimation par polynômes locaux	17
2.3.1	Calcul de l'estimateur	18
2.4	Choix du paramètre de lissage	20
3	Tests nonparamétrique	22
3.1	Différents test nonparamétrique	23
3.2	Test sur une population	24
3.2.1	Test de Kolmogorov	24
3.2.2	Cramér-Von Mises et d'Anderson-Darling	25
3.2.3	Le test de Spearman	26
3.2.4	Test de Kendall	27
3.2.5	Test de médiane ou de symétrie	28

a- Test du signe :	28
3.3 Test sur deux populations	30
3.3.1 Test de Kolmogorov Smirnov (KS)	30
3.3.2 Cramér-Von Mises et d'Anderson-Darling	30
3.3.3 Test de Corrélation de rang de Spearman	31
3.3.4 Test de Wilcoxon (Wilcoxon signed rank test)	32
3.3.5 Test de Mann-Whitney	34
3.3.6 Test de Kruskal et Wallis	36
3.4 Test du χ^2	36
3.4.1 Construction du test	36
3.4.2 Quatre applications du test du Chi deux	38
3.5 Exercices	40
4 Bibliographie	42
5 Annexe	43

Introduction

La statistique non paramétrique est une branche de la statistique qui se concentre sur l'analyse des données sans faire d'hypothèses spécifiques sur la forme ou les paramètres de la distribution sous-jacente des données. Contrairement à la statistique paramétrique, qui repose sur des modèles de distribution spécifiques (par exemple, la distribution normale), la statistique non paramétrique offre une approche plus flexible et adaptée à des situations où les hypothèses paramétriques ne sont pas satisfaites.

Les méthodes non paramétriques sont utilisées lorsque les données sont difficiles à caractériser par des paramètres numériques ou lorsque la forme de la distribution est inconnue. Elles sont également utiles lorsque les échantillons sont petits ou lorsque les données sont sujettes à des valeurs aberrantes (valeurs extrêmes) qui pourraient biaiser les résultats des méthodes paramétriques traditionnelles.

Dans la statistique non paramétrique, on fait souvent appel à des statistiques d'ordre, telles que les rangs, les médianes, les statistiques U de Mann-Whitney, etc., pour effectuer des tests d'hypothèses. Ces tests permettent de comparer des échantillons ou de déterminer si deux groupes de données proviennent de distributions similaires.

De plus, la statistique non paramétrique comprend des techniques d'estimation de la densité, qui permettent de représenter graphiquement la distribution des données sans supposer une forme spécifique.

Les modèles de régression non paramétrique sont également utilisés pour modéliser les relations entre variables sans imposer de contraintes sur la forme fonctionnelle de la

relation.

Les méthodes non paramétriques peuvent être appliquées dans divers domaines, tels que les sciences sociales, les sciences de la santé, l'économie, l'ingénierie et bien d'autres. Elles offrent une approche plus souple et robuste pour analyser les données dans des situations réelles complexes.

Il est essentiel de comprendre que la statistique non paramétrique ne signifie pas l'absence de tout paramètre dans les analyses, mais plutôt qu'elle ne dépend pas de l'identification précise de paramètres spécifiques pour caractériser la distribution des données. Cela rend les méthodes non paramétriques plus adaptées à un large éventail de scénarios et les rend très utiles pour les problèmes complexes où les modèles paramétriques ne sont pas appropriés.

La problématique de la statistique non paramétrique réside dans le défi d'analyser et de tirer des conclusions à partir de données sans faire d'hypothèses spécifiques sur la distribution sous-jacente. Bien que cette approche soit plus souple et adaptée à un large éventail de situations, elle présente également des défis et des limitations qui nécessitent une attention particulière.

1. Efficacité statistique : Les méthodes non paramétriques peuvent être moins efficaces que les méthodes paramétriques lorsque des hypothèses spécifiques sur la distribution des données sont satisfaites. En l'absence de connaissances préalables sur la distribution réelle, les tests non paramétriques peuvent avoir moins de puissance pour détecter des différences significatives entre les groupes.

2. Complexité computationnelle : Certaines méthodes non paramétriques peuvent être plus complexes et exiger plus de ressources informatiques que les méthodes paramétriques. Cela peut être un inconvénient pour des ensembles de données massifs ou dans des situations où les analyses doivent être réalisées en temps réel.

3. Choix de la méthode : La statistique non paramétrique offre une grande variété de méthodes, et choisir la méthode appropriée peut être délicat sans connaissances préalables sur les données. Certaines méthodes peuvent mieux fonctionner dans certains contextes que d'autres, et il est essentiel de sélectionner celle qui répond le mieux à la question de recherche.

4. Biais de classement : Les statistiques non paramétriques, comme les tests de rangs, se basent sur les ordres des observations plutôt que sur leurs valeurs réelles. Cela peut entraîner une perte d'information et une diminution de la sensibilité des tests, en parti-

culier avec des échantillons de petite taille.

5. Estimation de densité : L'estimation non paramétrique de la densité peut être difficile, en particulier dans des régions où les données sont rares. L'obtention d'une représentation précise de la distribution des données peut devenir un défi.

6. Overfitting : Dans les modèles de régression non paramétrique, il existe un risque d'overfitting, c'est-à-dire que le modèle s'adapte trop étroitement aux données d'apprentissage et ne généralise pas bien sur de nouvelles données.

Malgré ces défis, la statistique non paramétrique offre de nombreux avantages et est souvent le meilleur choix dans des situations réelles où les hypothèses paramétriques ne sont pas réalistes. Elle permet une analyse plus souple et plus robuste des données, en faisant moins de suppositions sur leur nature sous-jacente. Les statisticiens et chercheurs doivent être conscients de ces problématiques et choisir judicieusement les méthodes non paramétriques en fonction du contexte de leur analyse.

ESTIMATION NONPARAMÉTRIQUE DE LA DENSITÉ

1.1 Estimateur empirique de la fonction de répartition

Rappelons que la fonction de répartition empirique \mathbb{F}_n de l'échantillon est définie par :

$$\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}} = \text{pourcentage d'observations inférieures à } x$$

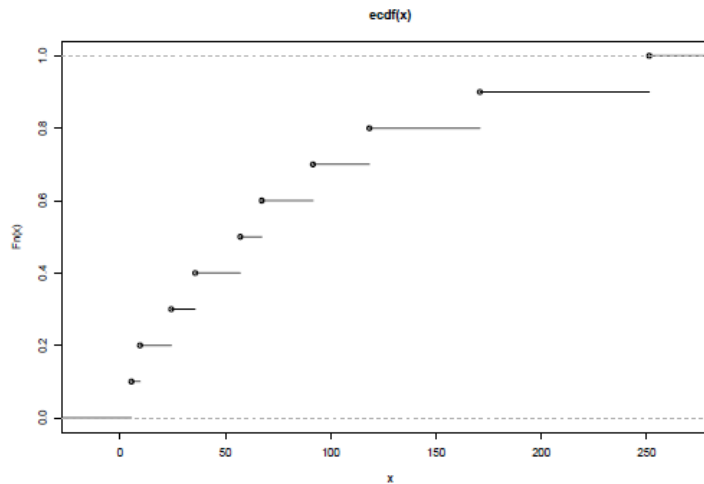


FIGURE 1.1 – Représentation de la fonction de répartition empirique

Propriété 1.1 $\forall x \in \mathbb{R}, n\mathbb{F}_n(x)$ est de loi binomiale $\mathcal{B}(n, F(x))$

Propriété 1.2 — $\forall x \in \mathbb{R}, \mathbb{F}_n(x)$ est un estimateur sans biais et convergent en moyenne quadratique de $F(x)$.

$$— \forall x \in \mathbb{R}, \quad \mathbb{F}_n(x) \xrightarrow{PS} F(x).$$

Preuve 1 Il suffit d'appliquer la loi des grands nombres aux variables aléatoires de loi de Bernoulli $\mathbf{1}_{\{X_i \leq x\}}$:

$$\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}} \xrightarrow{PS} E(\mathbf{1}_{\{X_i \leq x\}}) = F(x).$$

Théorème 1.1 (Glivenko-Catelli)

$$D_n = \sup \{ |\mathbb{F}_n(x) - F(x)| ; x \in \mathbb{R} \} \xrightarrow{PS} 0.$$

$$\sqrt{n} \frac{\mathbb{F}_n(x) - F(x)}{\sqrt{\mathbb{F}_n(x)(1 - \mathbb{F}_n(x))}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Propriété 1.3 (Intervalle de confiance) $\forall x \in \mathbb{R}$, un intervalle de confiance asymptotique de seuil α pour $F(x)$ est :

$$\left[\mathbb{F}_n(x) - \frac{u_\alpha}{\sqrt{n}} \sqrt{\mathbb{F}_n(x)(1 - \mathbb{F}_n(x))} \quad , \quad \mathbb{F}_n(x) + \frac{u_\alpha}{\sqrt{n}} \sqrt{\mathbb{F}_n(x)(1 - \mathbb{F}_n(x))} \right]$$

1.1.1 Quantile

Pour $p \in [0, 1]$, le quantile d'ordre p d'une v.a.r. X est donné par

$$F^{-1}(p) = \inf \{ x : F(x) \geq p \} \quad 0 < p < 1$$

et peut être estimé par

$$F_n^{-1}(p) = \inf \{ x : F_n(x) \geq p \},$$

le p ème quantile de la fonction de répartition empirique.

1.2 Estimateur de Rosenblatt

Soit F la fonction de répartition (f.d.r.) de la loi commune des v.a.r continue X_i ; on sait que f est la dérivée de F ; on peut alors exprimer f à partir de F par la formule :

$$f(x) = F'(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h};$$

ceci montre que, pour h assez petit, on a l'approximation suivante :

$$f(x) \approx \frac{F(x+h) - F(x-h)}{2h} \quad (1.1)$$

La f.d.r. F étant elle aussi inconnue, une idée pour estimer f consiste à remplacer dans 1.1 la fonction F par la f.d.r. empirique \hat{F}_n calculée sur la base de l'échantillon X_1, \dots, X_n par

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{]-\infty, x]}(X_i)$$

ceci conduit à l'estimateur simple \hat{f}_n de f , appelé [estimateur de Rosenblatt](#), donné par

$$\begin{aligned} \hat{f}_n(x) &= \frac{1}{2nh} \sum_{i=1}^n (\mathbf{I}_{]-\infty, x+h]}(X_i) - \mathbf{I}_{]-\infty, x-h]}(X_i)) \\ \hat{f}_n(x) &= \frac{1}{2nh} \sum_{i=1}^n \mathbf{I}_{]x-h, x+h]}(X_i) \end{aligned}$$

Théorème 1.2 (Glivenko-Cantelli)

$$\lim_{n \rightarrow +\infty} \left(\sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F(x) \right| \right) = 0 \quad P_s$$

Puisque

$$X_i \in]x-h, x+h] \Leftrightarrow x-h < X_i \leq x+h \Leftrightarrow -1 < \frac{X_i - x}{h} \leq 1 \Leftrightarrow \frac{X_i - x}{h} \in \left] -1, 1 \right],$$

on a $\mathbb{I}_{]x-h, x+h]}(X_i) = \mathbf{I}_{]-1, 1]} \left(\frac{X_i - x}{h} \right)$, et

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n K_0 \left(\frac{X_i - x}{h} \right)$$

où $K_0(t) = \frac{1}{2} \mathbb{I}_{]-1, 1]}(t)$.

Remarque 1.1 *L'estimateur de Rosenblatt est une fonction discontinue. Ceci est un handicap puisqu'il est censé estimer une densité qui, dans la plupart des cas, est une fonction continue. Comment se fait en pratique le choix de la fenêtre h qui garantit une bonne estimation de la densité.*

1.3 Estimateur à noyau de Parzen-Rosenblatt

Définition 1.1 On appelle noyau toute fonction intégrable $K : \mathbb{R} \rightarrow \mathbb{R}$ telle que

$$\int_{\mathbb{R}} K(t) dt = 1$$

Remarque 1.2 Il n'est pas rare de trouver une autre définition du noyau dans la littérature. Cette situation est due aux objectif théorique d'utilisation d'estimateur à noyau de paramètre inconnu donné.

Exemple 1.1 Nous donnons ci-dessous des exemples de noyaux.

1. Noyau rectangulaire. $K(t) = \frac{1}{2} \mathbb{I}_{[-1,1]}(t)$.
2. Noyau triangulaire. $K(t) = (1 - |t|) \mathbb{I}_{[-1,1]}(t)$.
3. Noyau parabolique ou d'Epanechnikov. $K(t) = \frac{3}{4} (1 - t^2) \mathbb{I}_{[-1,1]}(t)$.
4. Noyau biweight. $K(t) = \frac{15}{16} (1 - t^2)^2 \mathbb{I}_{[-1,1]}(t)$.
5. Noyau gaussien. $K(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$.
6. Noyau de Silverman. $K(t) = \frac{1}{2} \exp\left(-\frac{|t|}{\sqrt{2}}\right) \sin\left(-\frac{|t|}{\sqrt{2}} + \frac{\pi}{4}\right)$.

La méthode du noyau consiste à généraliser cette approche de Rosenblatt d'estimation de la densité à d'autres fonctions K ce qui donne l'estimateur suivant :

Définition 1.2 (Estimateur par noyau) Un estimateur à noyau de la densité f est une fonction \hat{f} définie par :

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)$$

1.3.1 Propriétés ponctuelles

Nous allons, à présent, étudier les qualités de l'estimateur à noyau en un point $x_0 \in \mathbb{R}$ fixé. Notons que dans ce cas, $\hat{f}_n(x_0)$ est un estimateur de $f(x_0)$ dont la qualité peut, par conséquent, être mesurée par le risque quadratique :

$$\text{MSE}(x_0) = \mathbb{E} \left(\left(\hat{f}_n(x_0) - f(x_0) \right)^2 \right) = \sigma^2(x_0) + b^2(x_0)$$

où $\sigma^2(x_0) = \text{Var}(\hat{f}_n(x_0))$ est la variance de l'estimateur, et $b(x_0) = \mathbb{E}(\hat{f}_n(x_0)) - f(x_0)$ est le biais. Nous allons étudier séparément la variance et le biais, puis en déduire l'étude du risque quadratique.

Propriété 1.4 On suppose f bornée, et que le noyau K vérifie :

$$\int_{\mathbb{R}} K^2(t) dt < +\infty$$

Alors, pour tous $h > 0$ et $n \geq 1$, on a

$$\sigma^2(x_0) \leq \frac{C_1}{nh},$$

où $C_1 = \|f\|_{\infty} \int_{\mathbb{R}} K^2(t) dt$.

Définition 1.3 — Soient I un intervalle de \mathbb{R} , β et L des réels strictement positifs. On pose $\ell = \lfloor \beta \rfloor$. La classe de Hölder $\Sigma(\beta, L)$ sur I est l'ensemble des fonctions $\varphi : I \rightarrow \mathbb{R}$ dérivables jusqu'à l'ordre ℓ et telles la dérivée ℓ -ième $\varphi^{(\ell)}$ vérifie :

$$\forall (x, y) \in I^2, |\varphi^{(\ell)}(x) - \varphi^{(\ell)}(y)| \leq L|x - y|^{\beta - \ell}.$$

— Soient $\ell \in \mathbb{N}^*$ et un noyau K . On dit que K est d'ordre ℓ si, pour tout $j \in \{0, \dots, \ell\}$, la fonction $t \mapsto t^j K(t)$ est intégrable, et si :

$$\int_{\mathbb{R}} t^j K(t) dt = 0, j = 1, \dots, \ell$$

On a alors :

Propriété 1.5 On suppose que $f \in \Sigma(\beta, L)$ où $\beta > 0$ et $L > 0$, et que le noyau K est d'ordre $\ell = \lfloor \beta \rfloor$ et est tel que :

$$\int_{\mathbb{R}} |t|^{\beta} |K(t)| dt < +\infty$$

Alors, pour tous $h > 0$ et $n \geq 1$, on a

$$|b(x_0)| \leq C_2 h^{\beta},$$

où $C_2 = \frac{L}{\ell!} \int_{\mathbb{R}} |t|^{\beta} |K(t)| dt$.

Propriété 1.6 On suppose que f est bornée et appartient à $\Sigma(\beta, L)$ où $\beta > 0$ et $L > 0$, et que le noyau K est d'ordre $\ell = \lfloor \beta \rfloor$ et est tel que :

$$\int_{\mathbb{R}} K^2(t) dt < +\infty, \int_{\mathbb{R}} |t|^{\beta} |K(t)| dt < +\infty$$

Alors, pour tous $h > 0$ et $n \geq 1$, on a

$$\text{MSE}(x_0) \leq \frac{C_1}{nh} + C_2^2 h^{2\beta}.$$

Cette proposition montre que si la fenêtre $h = h_n$ est telle que $nh_n \rightarrow +\infty$ et $h_n \rightarrow 0$ lorsque $n \rightarrow +\infty$, alors $\text{MSE}(x_0)$ tend vers 0 lorsque $n \rightarrow +\infty$. Cependant, on peut obtenir la fenêtre optimale, c'est à dire celle qui conduit à une minimisation du risque quadratique, et en déduire la vitesse de convergence correspondante de ce risque.

Propriété 1.7 *Sous les conditions de la Proposition 1.6 :*

1. La fenêtre optimale est donnée par $h_n^* = \left(\frac{C_1}{2\beta C_2^2}\right)^{\frac{1}{2\beta+1}} n^{-\frac{1}{2\beta+1}}$.
2. La vitesse de convergence du risque quadratique est donnée par $\text{MSE}(x_0) = O\left(n^{-\frac{2\beta}{2\beta+1}}\right)$.

1.3.2 Propriétés globales

Pour étudier la qualité de l'estimateur à noyau non plus ponctuellement, mais globalement, on va considérer le risque quadratique intégré défini par

$$\text{MISE} = \int_{\mathbb{R}} \text{MSE}(x)dx = \int_{\mathbb{R}} \sigma^2(x)dx + \int_{\mathbb{R}} b^2(x)dx.$$

Nous allons étudier séparément l'intégrale de la variance et celle du carré du biais, puis en déduire l'étude du MISE.

Propriété 1.8 *On suppose que f est bornée et appartient à $\mathcal{H}(\beta, L)$ où $\beta > 0$ et $L > 0$, et que le noyau K est d'ordre $\ell = \lfloor \beta \rfloor$ et est tel que :*

$$\int_{\mathbb{R}} K^2(t)dt < +\infty, \int_{\mathbb{R}} |t|^\beta |K(t)|dt < +\infty$$

Alors, pour tous $h > 0$ et $n \geq 1$, on a

$$\text{MISE} \leq \frac{C_3}{nh} + C_2^2 h^{2\beta}.$$

$$C_3 = \int_{\mathbb{R}} K^2(t)dt$$

Cette proposition montre que si la fenêtre $h = h_n$ est telle que $nh_n \rightarrow +\infty$ et $h_n \rightarrow 0$ lorsque $n \rightarrow +\infty$, alors MISE tend vers 0 lorsque $n \rightarrow +\infty$. Tout comme cela avait été fait pour le risque quadratique, la fenêtre optimale et la vitesse de convergence correspondante du MISE peuvent être déterminées. Par un raisonnement identique à celui adopté pour le risque quadratique, on obtient :

Propriété 1.9 *Sous les conditions de la Proposition 1.8 :*

1. La fenêtre optimale est donnée par $h_n^* = \left(\frac{C_3}{2\beta C_2^2}\right)^{\frac{1}{2\beta+1}} n^{-\frac{1}{2\beta+1}}$.

2. La vitesse de convergence du risque quadratique intégré est donnée par $\text{MISE} = O\left(n^{-\frac{2\beta}{2\beta+1}}\right)$.

1.3.3 Choix optimal de la fenêtre par validation croisée

Les valeurs optimales de la fenêtre obtenues ci-dessus ne sont pas utiles en pratique car elles ne peuvent être calculées vu qu'elles dépendent de constantes qui, elles mêmes, dépendent de la densité inconnue, éventuellement via β . Il y a donc nécessité de rechercher une autre méthode permettant de déterminer une fenêtre optimale. Une première idée est de déterminer h qui minimise le MISE, mais cette approche est, elle aussi, impraticable car le MISE dépend de la densité. Il convient alors de déterminer un estimateur sans biais du MISE, puis d'opérer comme choix de fenêtre celle qui minimise cet estimateur. Remarquons tout d'abord que

$$\begin{aligned}\text{MISE}(h) &= \int_{\mathbb{R}} \mathbb{E} \left(\left(\hat{f}_n(x) - f(x) \right)^2 \right) dx \\ &= \mathbb{E} \left(\int_{\mathbb{R}} \left(\hat{f}_n(x) - f(x) \right)^2 dx \right) \\ &= \mathbb{E} \left(\int_{\mathbb{R}} \hat{f}_n^2(x) dx - 2 \int_{\mathbb{R}} \hat{f}_n(x) f(x) dx \right) + \int_{\mathbb{R}} f^2(x) dx.\end{aligned}$$

Puisque $\int_{\mathbb{R}} f^2(x) dx$ ne dépend pas de h , minimiser $\text{MISE}(h)$ équivaut à minimiser $J(h)$ où

$$J(h) = \mathbb{E} \left(\int_{\mathbb{R}} \hat{f}_n^2(x) dx - 2 \int_{\mathbb{R}} \hat{f}_n(x) f(x) dx \right) = \text{MISE}(h) - \int_{\mathbb{R}} f^2(x) dx.$$

Nous allons, en fait, déterminer un estimateur sans biais de $J(h)$ par la méthode de validation croisée, et c'est cet estimateur qui devra être minimisé. Pour $i \in \{1, \dots, n\}$, on considère

$$\hat{f}_{n,-i}(x) = \frac{1}{(n-1)h} \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{X_j - x}{h}\right)$$

et

$$\widehat{CV}(h) = \int_{\mathbb{R}} \hat{f}_n^2(x) dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{n,-i}(X_i)$$

On a alors :

Propriété 1.10 *On suppose que :*

$$\int_{\mathbb{R}} f^2(x)dx < +\infty \quad \text{et} \quad \int_{\mathbb{R}^2} \left| K\left(\frac{t-x}{h}\right) \right| f(t)f(x)dtdx < +\infty$$

Alors, pour tout $h > 0$, on a

$$\mathbb{E}(\widehat{CV}(h)) = J(h) = \text{MISE}(h) - \int_{\mathbb{R}} f^2(x)dx.$$

Le choix optimal de la fenêtre s'opère donc par une minimisation de $\widehat{CV}(h)$. Plus précisément, on détermine \widehat{h}_{CV} qui vérifie :

$$\widehat{h}_{CV} = \arg \min_{h>0} \widehat{CV}(h).$$

1.4 Estimation par projection

On suppose ici que la densité f appartient à $L^2(I)$ où $I \subset \mathbb{R}$. Cet espace est un espace de Hilbert séparable pour le produit scalaire défini par

$$\langle f, g \rangle = \int_I f(x)g(x)dx$$

Il existe donc une base orthonormée $\{\phi_j\}_{j \in \mathbb{N}^*}$ de $L^2(I)$, et f s'écrit sous la forme

$$f = \sum_{j=1}^{+\infty} \theta_j \phi_j$$

où

$$\theta_j = \langle f, \phi_j \rangle = \int_I \phi_j(x)f(x)dx = \mathbb{E}(\phi_j(X)),$$

X étant une v.a. de densité f .

Définition 1.4 *On appelle estimateur par projection l'estimateur \widehat{f}_{nD} défini par*

$$\widehat{f}_{nD}(x) = \sum_{j=1}^D \widehat{\theta}_j \phi_j(x), \quad \text{où} \quad \widehat{\theta}_j = \frac{1}{n} \sum_{i=1}^n \phi_j(X_i).$$

L'entier D est appelé dimension de l'estimateur.

Remarque 1.3 Remarquons que $\widehat{\theta}_j$ est un estimateur sans biais de θ_j car $\mathbb{E}(\widehat{\theta}_j) = \theta_j$, et que sa variance est :

$$\text{Var}(\widehat{\theta}_j) = \frac{1}{n} \text{Var}(\phi_j(X_1)) = \frac{1}{n} (\mathbb{E}(\phi_j^2(X_1)) - \mathbb{E}(\phi_j(X_1))^2) = \frac{1}{n} (\mathbb{E}(\phi_j^2(X_1)) - \theta_j^2).$$

Puisque $\int_I f^2(x)dx$ ne dépend pas de D , la formule (1.25) montre que minimiser $\text{MISE}(D)$ équivaut à minimiser $J(D)$ où

$$J(D) = \text{MiSE}(D) - \int_I f^2(x)dx = \frac{1}{n} \sum_{j=1}^D \mathbb{E}(\phi_j^2(X)) - \frac{n+1}{n} \sum_{j=1}^D \theta_j^2.$$

Nous allons, en fait, déterminer un estimateur sans biais de $J(D)$, et c'est cet estimateur qui devra être minimisé. On a :

Propriété 1.11 *L'estimateur :*

$$\widehat{J}(D) = \frac{1}{n-1} \sum_{j=1}^D \left[\frac{2}{n} \sum_{i=1}^n \phi_j^2(X_i) - (n+1)\widehat{\theta}_j^2 \right]$$

est un estimateur sans biais de $J(D)$.

Le choix optimal de la dimension s'opère donc par une minimisation de $\widehat{J}(D)$. Plus précisément, on détermine \widehat{D}^* qui vérifie :

$$\widehat{D}^* = \arg \min_{D \geq 1} \widehat{J}(D)$$

1.4.1 Cas de la base trigonométrique

On suppose ici que $I = [0, 1]$; on considère alors la base trigonométrique $\{\phi_j\}_{j \in \mathbb{N}^*}$ de $L^2([0, 1])$ définie par

$$\phi_1(x) = 1, \quad \phi_{2k}(x) = \sqrt{2} \cos(2\pi kx), \quad \phi_{2k+1}(x) = \sqrt{2} \sin(2\pi kx), \quad k \in \mathbb{N}^*.$$

Dans ce cas, on a pour tout $j \in \mathbb{N}^*$ et tout $x \in [0, 1]$, $\phi_j^2(x) \leq 2$. Ainsi, (1.28) implique

$$\text{MISE}(D) \leq \frac{2D}{n} + \sum_{j=D+1}^{+\infty} \theta_j^2.$$

Cette inégalité montre que si $D = D_n$ est une suite croissante telle que $\lim_{n \rightarrow +\infty} D_n = +\infty$ et $\lim_{n \rightarrow +\infty} \frac{D_n}{n} = 0$, alors $\text{MISE}(D_n)$ tend vers 0 lorsque $n \rightarrow +\infty$ puisque $\lim_{n \rightarrow +\infty} \sum_{j=D_n+1}^{+\infty} \theta_j^2 = 0$ vu que la série $\sum_{j \geq 1} \theta_j^2$ converge.

1.4.2 Estimateur par ondelettes

On suppose ici que $f \in \mathcal{L}^2(\mathbb{R})$. On considère le produit scalaire de $\mathcal{L}^2(\mathbb{R})$ défini par

$$\langle f, g \rangle = \int_{\mathbb{R}} f(x)g(x)dx$$

Soit $\varphi \in \mathcal{L}^2(\mathbb{R})$ une ondelette père et ψ l'ondelette mère associée; ce sont des fonctions de \mathbb{R} vers \mathbb{R} vérifiant :

- le système $\{\varphi_k, k \in \mathbb{Z}\}$, avec $\varphi_k(x) = \varphi(x - k)$, est un système orthonormé; on note V_0 le sous-espace de $\mathcal{L}^2(\mathbb{R})$ engendré par ce système.
- Pour tout $j \in \mathbb{Z}$, le système $\{\psi_{jk}, k \in \mathbb{Z}\}$, avec $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$, est un système orthonormé, et le sous-espace de $\mathcal{L}^2(\mathbb{R})$ qu'il engendre est noté W_j .
- $\mathcal{L}^2(\mathbb{R}) = V_0 \oplus \bigoplus_{j=0}^{+\infty} W_j$.

Alors $\{\varphi_k, \psi_{jk}/(j, k) \in \mathbb{Z}^2\}$ est une base orthonormée de $\mathcal{L}^2(\mathbb{R})$, et f peut s'écrire sous la forme :

$$f(x) = \sum_{k \in \mathbb{Z}} \alpha_k \varphi_k(x) + \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \gamma_{jk} \psi_{jk}(x),$$

où

$$\alpha_k = \int_{\mathbb{R}} f(x) \varphi_k(x) dx \quad \text{et} \quad \gamma_{jk} = \int_{\mathbb{R}} f(x) \psi_{jk}(x) dx$$

L'estimateur par ondelettes de f est alors défini par :

$$\hat{f}_n(x) = \sum_{k \in \mathbb{Z}} \hat{\alpha}_k^{(n)} \varphi_k(x) + \sum_{j=0}^{j_n} \sum_{k \in \mathbb{Z}} \hat{\gamma}_{jk}^{(n)} \psi_{jk}(x),$$

où

$$\hat{\alpha}_k^{(n)} = \frac{1}{n} \sum_{i=1}^n \varphi_k(X_i), \quad \hat{\gamma}_{jk}^{(n)} = \frac{1}{n} \sum_{i=1}^n \psi_{jk}(X_i).$$

et $(j_n)_{n \in \mathbb{N}}$ est une suite croissante d'entiers telle que $j_n \nearrow +\infty$ lorsque $n \rightarrow +\infty$.

Exemple 1.2 *Ondelettes de Haar. Elles sont définies par les ondelettes père et mère données par*

$$\varphi = \mathbb{I}_{[0,1[} \quad \text{et} \quad \psi = \mathbb{I}_{[1/2,1[} - \mathbb{I}_{[0,1/2]}.$$

ESTIMATION DE LA FONCTION DE RÉGRESSION

Deux grandes approches de modèles de régression sont présentes dans la littérature. La première approche est le modèle de régression paramétrique, qui peut se définir par :

$$Y = r_\theta(X) + \varepsilon,$$

où la fonction de lien r_θ appartient à une famille bien connue de fonctions paramétrées par θ , vecteur de paramètres réels. ε le terme d'erreur est aléatoire et porte le reste des informations sur Y que X n'arrive pas à expliquer. Plusieurs méthodes existantes permettent d'estimer le paramètre θ du modèle comme la méthode du maximum de vraisemblance, la méthode des moindres carrés, etc. et sont très efficaces lorsque la famille des fonctions g_θ est bien spécifiée. Cependant, dans beaucoup d'applications, la mise en évidence d'une fonction de lien adéquate est très difficile voir impossible. Ce défaut de spécification du modèle est la difficulté majeure de ce modèle. Ainsi, alternativement, il a été proposé le modèle de régression non-paramétrique qui se définit par la relation :

$$Y = r(X) + \varepsilon. \tag{2.1}$$

où Y et X sont des v.a.r, ε est une v.a. réelle centrée et de variance ν^2 , indépendante de X , et r est une fonction inconnue qui établit un lien entre X et Y . Remarquons que si 2.1 est vérifié, on a

$$\mathbb{E}(Y \mid X) = \mathbb{E}(r(X) \mid X) + \mathbb{E}(\varepsilon \mid X) = \mathbb{E}(r(X) \mid X) = r(X);$$

donc

$$r(x) = \mathbb{E}(Y \mid X = x) \quad (2.2)$$

$$r(x) = \int_{\mathbb{R}} y f_{Y|X=x}(y) dy = \int_{\mathbb{R}} y \frac{f_{(X,Y)}(x, y)}{f_X(x)} dy = \frac{\int_{\mathbb{R}} y f_{(X,Y)}(x, y) dy}{f_X(x)} \quad (2.3)$$

$$r(x) = \frac{m(x)}{f_X(x)} \quad (2.4)$$

On veut proposer un estimateur de r sur la base d'un échantillon i.i.d. $(X_1, Y_1), \dots, (X_n, Y_n)$ de (X, Y) .

2.1 Estimateur à noyau

Un estimateur par noyau de f est

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)$$

un estimateur par noyau de $f_{(X,Y)}(x, y)$ est

$$\hat{f}_{n,XY}(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) K\left(\frac{Y_i - y}{h}\right),$$

où K est un noyau tel que $\int_{\mathbb{R}} tK(t)dt = 0$. Ainsi un estimateur par noyau de r est

$$\hat{r}_n(x) = \frac{\frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \int_{\mathbb{R}} y K\left(\frac{Y_i - y}{h}\right) dy}{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}$$

en effectuant le changement de variable $t = \frac{Y_i - y}{h}$, il vient

$$\int_{\mathbb{R}} y K\left(\frac{Y_i - y}{h}\right) dy = h \int_{\mathbb{R}} (Y_i - ht) K(t) dt = hY_i \int_{\mathbb{R}} K(t) dt - h \int_{\mathbb{R}} tK(t) dt = hY_i.$$

D'où

$$\hat{r}_n(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}.$$

Définition 2.1 Soit K un noyau ; on appelle estimateur à noyau, ou encore estimateur de Nadaraya-Watson de r , l'estimateur \hat{r}_n défini par

$$\hat{r}_n(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}.$$

Le réel h est appelé fenêtre de l'estimateur.

2.2 Propriétés de l'estimateur de Nadaraya-Waston

Pour étudier les propriétés de cet estimateur, certaines conditions de régularités sont nécessaires. Nous définissons certaines comme suit :

Définition 2.2 — Dans la suite, nous désignons par $\lfloor \beta \rfloor$ le plus grand entier strictement inférieure à β .

— Supposons T un intervalle de \mathbb{R} β et L deux réels positifs. On appelle **classe de Hölder** $\Sigma(\beta, L)$ sur T l'ensemble des fonctions $\ell = \lfloor \beta \rfloor$ fois différentiables $f : T \rightarrow \mathbb{R}$ dont les dérivées $f^{(\ell)}$ vérifient

$$|f^{(\ell)}(x) - f^{(\ell)}(x')| \leq L |x - x'|^{\beta - \ell}, \quad \forall x, x' \in T.$$

On va considérer le risque quadratique intégré défini par

$$\text{MISE} = \int_{\mathbb{R}} \sigma^2(x) dx + \int_{\mathbb{R}} b^2(x) dx,$$

où $\sigma^2(x) = \text{Var}(\hat{r}_n(x))$ et $b(x) = \mathbb{E}(\hat{r}_n(x)) - r(x)$.

Propriété 2.1 On suppose que r est bornée et appartient à $\Sigma(\beta, L)$ où $\beta > 0$ et $L > 0$, et que le noyau K est d'ordre $\ell = \lfloor \beta \rfloor$ et est tel que :

$$\int_{\mathbb{R}} K^2(t) dt < +\infty, \quad \int_{\mathbb{R}} |t|^\beta |K(t)| dt < +\infty$$

Alors, pour tous $h > 0$ et $n \geq 1$, on a

$$\text{MISE} \leq \frac{C_4}{nh} + C_2^2 h^{2\beta}.$$

Cette proposition montre que si la fenêtre $h = h_n$ est telle que $nh_n \rightarrow +\infty$ et $h_n \rightarrow 0$ lorsque $n \rightarrow +\infty$, alors MISE tend vers 0 lorsque $n \rightarrow +\infty$. La fenêtre optimale et la vitesse de convergence correspondante du MISE peuvent être déterminées. On obtient :

Propriété 2.2 Sous les conditions de la Proposition 2.1 :

- La fenêtre optimale est donnée par $h_n^* = \left(\frac{C_4}{2\beta C_2^2} \right)^{\frac{1}{2\beta+1}} n^{-\frac{1}{2\beta+1}}$.
- La vitesse de convergence du risque quadratique intégré est donnée par $\text{MISE} = O\left(n^{-\frac{2\beta}{2\beta+1}}\right)$.

2.3 Estimation par polynômes locaux

L'estimateur de Nadaraya-Watson est un cas particulier d'estimation par polynômes locaux. En effet, cette méthode d'estimation est définie par une généralisation de la propriété suivante :

Propriété 2.3 *L'estimateur \hat{r}_n de Nadaraya-Watson associé à un noyau strictement positif K vérifie :*

$$\hat{r}_n(x) = \arg \min_{\theta \in \mathbb{R}} \left(\sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) (Y_i - \theta)^2 \right). \quad (2.5)$$

Preuve 2 *Soit le polynôme de second degré*

$$p(\theta) = \sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) (Y_i - \theta)^2.$$

On a

$$p'(\theta) = -2 \sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) (Y_i - \theta) = 2 \left\{ \theta \sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) - \sum_{i=1}^n Y_i K \left(\frac{X_i - x}{h} \right) \right\}$$

et

$$p''(\theta) = 2 \sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) > 0.$$

$$\begin{aligned} p'(\hat{\theta}(x)) = 0 &\iff \theta \sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) - \sum_{i=1}^n Y_i K \left(\frac{X_i - x}{h} \right) = 0 \\ &\iff \hat{\theta}(x) = \frac{\sum_{i=1}^n Y_i K \left(\frac{X_i - x}{h} \right)}{\sum_{i=1}^n K \left(\frac{X_i - x}{h} \right)} \\ &\iff \hat{\theta}(x) = \hat{r}_n(x) \end{aligned}$$

L'estimateur par polynôme local s'obtient en généralisant le problème d'optimisation (2.5) par utilisation d'un polynôme de degré supérieur à 2 en θ . Cela donne :

Si $r \in \Sigma(\beta, L)$, $\beta > 1$, $\ell = \lfloor \beta \rfloor$, alors pour z suffisamment proche de x nous pouvons écrire :

$$r(z) \approx r(x) + r'(x)(z - x) + \cdots + \frac{r^{(\ell)}(x)}{\ell!} (z - x)^\ell = \theta^T(x) U \left(\frac{z - x}{h} \right)$$

si

$$\begin{aligned} U(u) &= (1, u, u^2/2!, \dots, u^\ell/\ell!)^T, \\ \theta(x) &= (r(x), hr'(x), h^2r''(x), \dots, h^\ell r^{(\ell)}(x))^T. \\ \theta(x) &= (\theta_0(x), \dots, \theta_\ell(x))^T \end{aligned}$$

$$\begin{aligned} r(X_i) &\approx \theta^T(x) U\left(\frac{X_i - x}{h}\right) \\ r(X_i) &\approx \sum_{k=0}^{\ell} \frac{\theta_k}{k!} \left(\frac{X_i - x}{h}\right)^k \end{aligned}$$

Définition 2.3 Soient K un noyau positif, $\ell \in \mathbb{N}$ et $h > 0$. On définit, pour $x \in \mathbb{R}$, :

$$\hat{\theta}(x) = \operatorname{argmin}_{\theta=(\theta_0, \dots, \theta_\ell) \in \mathbb{R}^{\ell+1}} \left\{ \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \left(Y_i - \sum_{k=0}^{\ell} \frac{\theta_k}{k!} \left(\frac{X_i - x}{h}\right)^k\right)^2 \right\}. \quad (2.6)$$

Posant $\hat{\theta}(x) = (\hat{\theta}_0(x), \dots, \hat{\theta}_\ell(x))$, l'estimateur par polynôme local d'ordre ℓ de r est donné par

$$\hat{r}_n^{(\ell)}(x) = \hat{\theta}_0(x) \quad (2.7)$$

Remarque 2.1 L'estimateur de Nadaraya-Watson est retrouvé comme cas particulier en prenant $\ell = 0$ voir la preuve dans la proposition 2.3.

2.3.1 Calcul de l'estimateur

La fonction à minimiser définie en (2.6) peut s'écrire comme un critère des moindres carrés ; cela facilite le calcul de l'estimateur donné en (2.7). Plus précisément, posant

$$w_i(x) = K\left(\frac{X_i - x}{h}\right)$$

et considérant les matrices

$$\mathbb{X}_x = \begin{pmatrix} 1 & X_1 - x & \dots & \frac{(X_1 - x)^\ell}{\ell!} \\ 1 & X_2 - x & \dots & \frac{(X_2 - x)^\ell}{\ell!} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_n - x & \dots & \frac{(X_n - x)^\ell}{\ell!} \end{pmatrix} \text{ et } W_x = \begin{pmatrix} w_1(x) & 0 & 0 & \dots & 0 \\ 0 & w_2(x) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & w_n(x) \end{pmatrix}$$

Propriété 2.4 Si la matrice $\mathbb{X}_x^T W_x \mathbb{X}_x$ est inversible, alors l'estimateur par polynôme local d'ordre ℓ de r vérifie :

$$\hat{r}_n^{(\ell)}(x) = \sum_{i=1}^n l_i(x) Y_i$$

où $l_i(x)$ est la i -ème composante du vecteur $l(x)$ donné par

$$l(x) = W_x \mathbb{X}_x (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} e_1,$$

avec

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{\ell+1}$$

Preuve 3 Le polynôme

$$p(\theta) = \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \left(Y_i - \sum_{k=0}^{\ell} \frac{\theta_k}{k!} \left(\frac{X_i - x}{h}\right)^k \right)^2 = \sum_{i=1}^n w_i(x) \left(Y_i - \sum_{k=0}^{\ell} \frac{\theta_k}{k!} \left(\frac{X_i - x}{h}\right)^k \right)^2$$

peut être explicité comme un critère des moindres carrés. En effet, on a tout d'abord

$$Y_i - \sum_{k=0}^{\ell} \frac{\theta_k}{k!} \left(\frac{X_i - x}{h}\right)^k = (\mathbb{Y} - \mathbb{X}_x a)_i$$

où

$$\mathbb{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad \text{et} \quad a = \begin{pmatrix} \theta_0 \\ \frac{\theta_1}{h} \\ \vdots \\ \frac{\theta_\ell}{h^\ell} \end{pmatrix} = V_h \theta$$

avec

$$V_h = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{h} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{h^\ell} \end{pmatrix} \quad \text{et} \quad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_\ell \end{pmatrix}$$

Donc

$$\begin{aligned}
p(\theta) &= \sum_{i=1}^n w_i(x) (\mathbb{Y} - \mathbb{X}_x V_h \theta)_i^2 = (\mathbb{Y} - \mathbb{X}_x V_h \theta)^T W_x (\mathbb{Y} - \mathbb{X}_x V_h \theta) \\
&= \mathbb{Y}^T W_x \mathbb{Y} - 2\mathbb{Y}^T W_x \mathbb{X}_x V_h \theta + \theta^T V_h \mathbb{X}_x^T W_x \mathbb{X}_x V_h \theta,
\end{aligned}$$

et sa différentielle en un vecteur $u \in \mathbb{R}^{\ell+1}$ est

$$Dp(\theta)(u) = -2\mathbb{Y}^T W_x \mathbb{X}_x V_h u + 2\theta^T V_h \mathbb{X}_x^T W_x \mathbb{X}_x V_h u.$$

L'estimateur $\hat{\theta}(x)$ défini en (2.6) vérifie donc $Dp(\hat{\theta}(x))(u) = 0$ pour tout vecteur $u \in \mathbb{R}^{\ell+1}$, ce qui donne

$$(\hat{\theta}(x))^T V_h \mathbb{X}_x^T W_x \mathbb{X}_x V_h u = \mathbb{Y}^T W_x \mathbb{X}_x V_h u, \forall u \in \mathbb{R}^{\ell+1}$$

et, par conséquent,

$$(\hat{\theta}(x))^T V_h \mathbb{X}_x^T W_x \mathbb{X}_x = \mathbb{Y}^T W_x \mathbb{X}_x$$

c'est à dire

$$V_h \hat{\theta}(x) = (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} \mathbb{X}_x^T W_x \mathbb{Y}$$

De (2.7), il vient alors

$$\hat{r}_n^{(\ell)}(x) = \hat{\theta}_0(x) = e_1^T V_h \hat{\theta}(x) = e_1^T (\mathbb{X}_x^T W_x \mathbb{X}_x)^{-1} \mathbb{X}_x^T W_x \mathbb{Y} = l(x)^T \mathbb{Y} = \sum_{i=1}^n l_i(x) Y_i$$

2.4 Choix du paramètre de lissage

On doit choisir la fenêtre h pour l'estimateur à noyau et les estimateurs par polynômes locaux. Une technique commune à ces différentes méthodes pour déterminer une valeur optimale du paramètre est la validation croisée que nous allons maintenant décrire. Notons λ le paramètre à choisir et $\hat{r}_{n,\lambda}$ l'estimateur de r associé à ce paramètre. Pour avoir une meilleure estimation du risque, on doit construire son estimateur avec des données qui n'ont pas été utilisées pour construire $\hat{r}_{n,\lambda}$. Pour cela, on utilise la méthode de validation croisée qui consiste à partitionner l'échantillon en V blocs notés B_1, \dots, B_V . Soit $v : \{1, \dots, n\} \rightarrow \{1, \dots, V\}$ telle que $v(i)$ désigne le bloc auquel appartient la i -ème observation (X_i, Y_i) ; pour tout $k \in \{1, \dots, V\}$, on note $\hat{r}_{n,\lambda}^{(-k)}$ l'estimateur obtenu en supprimant de l'échantillon les données appartenant au bloc B_k . On a alors :

Définition 2.4 *Le score de validation croisée V-fold est la quantité*

$$\widehat{R}(\lambda) = \frac{1}{n} \sum_{i=1}^n \left(Y_i - \widehat{r}_{n,\lambda}^{(-v(i))}(X_i) \right)^2$$

Un cas particulier est donnée par la validation croisée leave-one-out, obtenue quand on considère n blocs réduits à une observation, et définie par :

Définition 2.5 *Le score de validation croisée leave-one-out est la quantité*

$$\widehat{R}(\lambda) = \frac{1}{n} \sum_{i=1}^n \left(Y_i - \widehat{r}_{n,\lambda}^{(-i)}(X_i) \right)^2$$

Le choix de λ s'opère alors par une minimisation de $\widehat{R}(\lambda)$, c'est à dire qu'on détermine $\widehat{\lambda}$ qui vérifie

$$\widehat{\lambda} = \arg \min_{\lambda} (\widehat{R}(\lambda)).$$

Le calcul de $\widehat{R}(\lambda)$ peut être long car on doit recalculer, pour chaque i de 1 à n l'estimateur de la fonction de régression. Toutefois, il peut être amélioré dans certains cas. En effet, si l'estimateur est de la forme

$$\widehat{r}_{n,\lambda}(x) = \sum_{i=1}^n l_i(x) Y_i$$

avec $\sum_{i=1}^n l_i(x) = 1$, on a :

Propriété 2.5 *Sous les hypothèses ci-dessus, le score de validation croisée leave-one-out est égal à :*

$$\widehat{R}(\lambda) = \frac{1}{n} \sum_{i=1}^n \left(\frac{Y_i - \widehat{r}_{n,\lambda}(X_i)}{1 - l_i(X_i)} \right)^2.$$

Dans ce cas, le calcul est moins long car on ne calcule l'estimateur de r qu'une fois et on l'applique ensuite aux différentes valeurs de l'échantillon.

Exemple 2.1 *L'estimateur de Nadaraya-Watson a la forme (2.21) avec*

$$l_i(x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{X_j - x}{h}\right)}.$$

TESTS NONPARAMÉTRIQUE

Un test non paramétrique est une méthode statistique utilisée pour comparer des groupes ou vérifier des relations entre variables sans faire d'hypothèses spécifiques sur la distribution sous-jacente des données. Contrairement aux tests paramétriques qui supposent une distribution particulière (par exemple, la distribution normale), les tests non paramétriques offrent une approche plus souple et robuste qui convient à un large éventail de situations.

Les tests non paramétriques sont souvent utilisés dans les cas suivants :

1. Lorsque les données ne suivent pas une distribution normale ou toute autre distribution connue.
2. Lorsque les échantillons sont petits et que les tests paramétriques ne sont pas appropriés.
3. Lorsque les données sont ordinales ou de nature non continue.
4. Lorsque les hypothèses paramétriques ne peuvent pas être vérifiées ou ne sont pas justifiées.

Dans les tests non paramétriques, les données sont souvent converties en rangs, et les comparaisons sont effectuées sur les rangs plutôt que sur les valeurs réelles. Cela permet de se débarrasser des contraintes de la distribution et de rendre les tests plus robustes aux valeurs aberrantes.

Les tests non paramétriques sont fréquemment utilisés pour comparer des groupes ou des échantillons indépendants (comme le test de Mann-Whitney), pour comparer des échantillons appariés (comme le test de Wilcoxon), pour comparer plus de deux groupes (comme le test de Kruskal-Wallis), ou pour vérifier des relations entre variables (comme le test de corrélation de rang de Spearman).

Il est important de noter que la statistique non paramétrique ne signifie pas l'absence de tout paramètre dans les tests, mais plutôt qu'elle ne dépend pas de l'identification précise de paramètres spécifiques pour caractériser la distribution des données. Ces tests

sont très utiles lorsque les hypothèses paramétriques ne sont pas satisfaites et offrent une approche plus flexible pour analyser les données dans des situations réelles complexes.

3.1 Différents test nonparamétrique

Il existe plusieurs tests non paramétriques couramment utilisés pour comparer des échantillons ou des groupes, ou pour vérifier des relations entre variables sans faire d'hypothèses spécifiques sur la distribution des données. Voici quelques-uns des tests non paramétriques les plus fréquemment utilisés :

1. Test de Wilcoxon (test de la somme des rangs de Wilcoxon) : Utilisé pour comparer deux échantillons appariés ou pour tester si la médiane d'une distribution est égale à une valeur spécifiée.
2. Test de Mann-Whitney (test U de Mann-Whitney) : Permet de comparer deux échantillons indépendants et de déterminer s'ils proviennent de la même distribution.
3. Test de Kruskal-Wallis : Utilisé pour comparer plus de deux échantillons indépendants. Il permet de vérifier s'il existe des différences significatives dans les distributions des groupes.
4. Test de Friedman : Une extension du test de Wilcoxon pour comparer plus de deux échantillons appariés.
5. Test de rangs signés de Wilcoxon : Une version non paramétrique du test t de Student, utilisée pour comparer deux échantillons appariés.
6. Test de la médiane (test de Hodges-Lehmann) : Un test pour comparer deux groupes en se basant sur la différence médiane plutôt que sur la moyenne.
7. Test de rangs signés de Wald-Wolfowitz (test runs) : Permet de tester l'indépendance séquentielle dans une série de données ordonnées.
8. Test de Kruskal-Wallis généralisé : Une extension du test de Kruskal-Wallis qui tient compte des liens de dépendance dans les données.
9. Test de Kolmogorov-Smirnov : Utilisé pour comparer deux distributions empiriques ou pour tester si une distribution suit une distribution spécifique.
10. Test de permutation : Une méthode générale qui consiste à échanger les valeurs entre les groupes pour obtenir une distribution nulle et calculer des p-values.

Ces tests non paramétriques couvrent un large éventail de scénarios et sont adaptés à différents types de données. Ils sont largement utilisés dans la recherche scientifique, les sciences sociales, la médecine, l'économie et bien d'autres domaines où les hypothèses

paramétriques ne sont pas satisfaites ou appropriées. Le choix du test dépendra du type de données, du nombre de groupes à comparer et de la nature de la question de recherche.

3.2 Test sur une population

Soit X_1, \dots, X_n un échantillon i.i.d. de v.a. X de fdr F , et soit F_0 une fdr fixée.

3.2.1 Test de Kolmogorov

On suppose que X est continue et on veut tester si

- **Hypothèse du test** : H_0 : " $F = F_0$ " contre H_1 : " $F \neq F_0$ "

- **Statistique de KS** :

$$D_n = \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_0(x) \right|$$

où $\forall x \in \mathbb{R}$, $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq x)$ est la fonction de répartition empirique de X .

Théorème 3.1 Soit $(X_{(1)}, \dots, X_{(n)})$ un n -échantillon issu de X . On note F la fonction de répartition de X et \hat{F}_n la fonction de répartition empirique. Si F est continue alors la loi de

$$D_n = \sup_{t \in \mathbb{R}} \left| \hat{F}_n(t) - F_0(t) \right|$$

ne dépend pas de F .

Preuve 4 On a, en posant $x = F(t)$,

$$\begin{aligned} D(F_n, F) &= \sup_{t \in \mathbb{R}} |F_n(t) - F_0(t)| = \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq t\}} - F_0(t) \right| \\ &= \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n 1_{\{F_0(X_i) \leq F_0(t)\}} - F_0(t) \right| = \sup_{x \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^n 1_{\{U_i \leq x\}} - x \right| \end{aligned}$$

d'où le résultat.

Théorème 3.2 On suppose que l'on a les mêmes hypothèses que ci-dessus. Alors la variable aléatoire $\sqrt{n}D_n$ converge en loi, vers une loi limite qui ne dépend pas de F et dont la fonction de répartition est égale à :

$$\forall t \geq 0, \quad F_{KS}(t) = 1 + 2 \sum_{k=1}^{+\infty} (-1)^k \exp(-2k^2 t^2)$$

On a donc, pour $\lambda > 0$,

$$\mathbb{P}(\sqrt{n}D_n \leq \lambda) \rightarrow 1 + 2 \sum_{k=1}^{+\infty} (-1)^k \exp(-2k^2\lambda^2)$$

Remarque 3.1 remarquer que si on réordonne de manière croissante l'échantillon, $(X_{(1)}, \dots, X_{(n)})$ alors $F_n(X_{(j)}) = j/n$ et

$$D_n = \max_{1 \leq j \leq n} \left(\max \left(\left| \frac{j}{n} - F_0(X_{(j)}) \right|, \left| F_0(X_{(j)}) - \frac{j-1}{n} \right| \right) \right) = \max_{1 \leq j \leq n} \left\{ F_0(X_{(j)}) - \frac{j-1}{n} \right\}$$

Règle de décision du test : On cherche (dans une table) le quantile $k_{1-\alpha}$ de la loi de Kolmogorov. Si $D_n > k_{1-\alpha}$, On rejette H_0 au risque α de ce tromper. Ce test est asymptotiquement de niveau α et sa puissance tend vers 1 quand n tend vers $+\infty$.

3.2.2 Cramér-Von Mises et d'Anderson-Darling

- **Hypothèse du test :** $H_0 : "F = F_0"$ contre $H_1 : "F \neq F_0"$

- **Statistique de KS :**

$$Q = n \int_{-\infty}^{+\infty} (\hat{F}(x) - F_0(x))^2 \Psi(x) dF_0(x),$$

où $\Psi(x)$ est une fonction de pondération qui va caractériser l'un ou l'autre test. Une conséquence majeure de ces variations est que la distributions des nouvelles statistiques, et donc leurs valeurs critiques vont dépendre de la fonction $F_0(x)$ retenue par l'hypothèse nulle.

1. le test de Cramer-von Mises La fonction de pondération est $\Psi(x) = 1$

$$W^2 = Q = n \int_{-\infty}^{+\infty} (\hat{F}(x) - F_0(x))^2 dF_0(x),$$

. En pratique, on remplace les intégrales par des sommes et la statistique est la suivante :

$$W^2 = \sum_{i=1}^n \left(\frac{2i-1 - F_0(X_{(i)})}{2n} \right)^2 + \frac{1}{12n}$$

où $X_{(i)}$ est la i ème plus petite valeur de l'échantillon.

Règle de décision : Si W^2 est supérieure à sa valeur critique, on rejette H_0 .

2. le test de Anderson-Darling La fonction de pondération est $\Psi(x) = [F(x)(1 - F(x))]^{-1}$ et

$$Q = A^2 = n \int_{-\infty}^{+\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} dF(x)$$

en pratique ces intégrales sont remplacées par une somme et la statistique donne

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i - 1) \log F(x_{(i)}) + (2n - 2i + 1) \log (1 - F(x_{(i)}))]$$

$$A^2 = -n - \sum_{i=1}^n \frac{2i - 1}{n} [\log(F(x_{(i)})) + \log(1 - F(x_{n+1-i}))]$$

Règle de décision : Si A^2 est supérieure à sa valeur critique, on rejette H_0 .

3.2.3 Le test de Spearman

Hypothèse du test : $H_0 : X_1, \dots, X_n$ sont indépendantes et de même loi i.i.d.
Contre $H_1 = \bar{H}_0 : X_1, \dots, X_n$ ne sont pas i.i.d." qui est trop vaste. Les alternatives les plus fréquemment retenues sont celles qui portent sur l'existence d'une tendance :

- H_1 : "Les X_i sont stochastiquement croissantes"
- H_2 : "Les X_i sont stochastiquement décroissantes"

D'autres alternatives sont possibles aussi comme par exemple :

- "Les X_i sont stochastiquement périodiques"
- "Les X_i sont de même moyenne mais de variances croissantes" Sous H_0 , les X_i sont i.i.d. donc leur ordre n'a aucune importance. Ce n'est évidemment pas le cas sous H_1 et H_2 . Il semble donc logique d'utiliser les statistiques d'ordre et de rang pour construire les tests.

Définition 3.1 On dira que la suite de variables aléatoires $\{X_i\}_{i \geq 1}$ est stochastiquement croissante (resp. décroissante) si et seulement si les fonctions de répartition des X_i diminuent (resp. augmentent) au sens où :

$$\forall x \in \mathbb{R}, \quad i < j \Rightarrow F_{X_i}(x) \geq F_{X_j}(x) \text{ (resp. } \leq \text{)}$$

Statistique du test : $R_{RI,n} = \frac{12}{n(n^2-1)} \sum_{i=1}^n R_i i - 3 \frac{n+1}{n-1}$ où R_i est le rang de X_i .

Règle de décision :

- Pour H_1 (test de croissance) : on rejette H_0 si $R_{RI,n} > s_{n,\alpha}$,

- Pour H_2 (test de décroissance) : on rejette H_0 si $R_{RI,n} < s_{n,\alpha}$, où $s_{n,\alpha}$ est le quantile d'ordre $1 - \alpha$ de la loi de Spearman de paramètre n .

Quand la taille de l'échantillon est suffisamment grande, on utilise les résultats suivants.

Propriété 3.1 — Sous H_0 , pour $n > 10$,

$$\sqrt{n-2} \frac{R_{RI,n}}{\sqrt{1-R_{RI,n}^2}}$$

est approximativement de loi de Student $St(n-2)$.

- Sous H_0 ,

$$\sqrt{n-1} R_{RI,n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

3.2.4 Test de Kendall

Hypothèse du test : $H_0 : X_1; \dots; X_n$ sont indépendantes et de même loi i.i.d

- H_1 : "Les X_i sont stochastiquement croissantes"
- H_2 : "Les X_i sont stochastiquement décroissantes".

Statistique du test :

$$\tau_n = \frac{4}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbf{1}_{\{X_i < X_j\}} - 1.$$

Règle de décision : Pour $n \leq 10$

- Pour H_1 (test de croissance), on rejette H_0 si $\tau_n > k_{n,\alpha}$
- Pour H_2 (test de décroissance), on rejette H_0 si $\tau_n < k_{n,\alpha}$ où $k_{n,\alpha}$ est le quantile d'ordre $1 - \alpha$ de la loi de Kendall de paramètre n .

Propriété 3.2 Sous H_0 , $\sqrt{\frac{9n(n-1)}{2(2n+5)}} \tau_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$.

En pratique, pour $n \leq 10$, on utilise une table de quantiles de la loi de Kendall, et pour $n > 10$, on utilise l'approximation normale.

3.2.5 Test de médiane ou de symétrie

Hypothèse du test : $H_0 : q_{1/2} = m$ où $\mathbb{P}(X \leq q_{1/2}) = 1/2$, i.e. "la médiane de la distribution est m " contre

$$H_1 : "q_{1/2} \neq m", \text{ ou } H_2 : "q_{1/2} \geq m", \text{ ou } H_3 : "q_{1/2} \leq m".$$

Exemple 3.1 On veut tester si le fait d'emménager en couple a une influence sur le poids. Pour cela, on mesure la différence de poids avant et un an après l'emménagement pour 100 individus.

a- Test du signe :

Statistique du test : $S_n^+ = \sum_{i=1}^n \mathbf{1}_{\{X_i > m\}}$.

Règle de décision :

Propriété 3.3 — $S_n^+ \sim \mathcal{B}(n, 1 - F(m))$.

— Sous H_0 , $S_n^+ \sim \mathcal{B}(n, \frac{1}{2})$.

— Pour H_1 , on rejette H_0 si $S_n^+ > k_{1-\alpha}$ où $k_{1-\alpha}$ est le quantile d'ordre $1 - \alpha$ de la loi Binomiale $\mathcal{B}(n, 1 - F(m))$.

— Cependant la fonction de répartition de cette loi n'étant pas inversible, ce quantile n'est pas défini pour tout α . Ainsi pour les grandes valeurs de n on utilise une approximation de la loi Binomiale $\mathcal{B}(n, \frac{1}{2})$ par celle de la loi normale $\mathcal{N}(\frac{n}{2}, \frac{n}{4})$.

On a alors

$$\frac{S_n^+ - \frac{n}{2}}{\sqrt{\frac{n}{4}}} = \frac{2S_n^+ - n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

— Pour H_2 on rejette H_0 si $\frac{2S_n^+ - n}{\sqrt{n}} > u_{1-\alpha}$.

— Pour H_3 on rejette H_0 si $\frac{2S_n^+ - n}{\sqrt{n}} < -u_{1-\alpha}$.

— Pour H_1 on rejette H_0 si $\left| \frac{2S_n^+ - n}{\sqrt{n}} \right| > u_\alpha$.

où $u_{1-\alpha}$ est le quantile d'ordre $1 - \alpha$. En pratique, on admet que l'approximation normale est valide dès que $n > 10$.

b- Le test des rangs signés de Wilcoxon :

Hypothèse du test : $H_0 : q_{1/2} = 0$

Statistique du test :

$$W_n^+ = \sum_{i=1}^n R_i^+ \mathbf{1}_{\{X_i > 0\}}$$

où

$$\begin{aligned} \forall i \in \{1, \dots, n\}, \quad R_i^+ &= 1 + \sum_{j=1}^n \mathbf{1}_{\{|X_j| < |X_i|\}} \\ &= \text{rang de } |X_i| \text{ dans la suite } |X_1|^*, \dots, |X_n|^* \end{aligned}$$

Remarque 3.2 *Cas des ex-æquos*

- Normalement le test s'applique à des variables diffuses, donc pas d'ex-æquos en théorie.
- En pratique, on affecte des rangs moyens en cas d'égalité et il existe des corrections des lois limites dans ce cas.

Exemple 3.2 $n = 5$.

x_i	2.3	-3.5	1.7	0.5	-1.4
x_i^*	-3.5	-1.4	0.5	1.7	2.3
r_i	5	1	4	3	2
$ x_i $	2.3	3.5	1.7	0.5	1.4
$ x_i ^*$	0.5	1.4	1.7	2.3	3.5
r_i^+	4	5	3	1	2

Règle de décision :

Propriété 3.4 — W_n^+ est à valeurs dans $\left\{0, \dots, \frac{n(n+1)}{2}\right\}$.

- $W_n^+ = \sum_{1 \leq i \leq j \leq n} \mathbf{1}_{\{X_i + X_j > 0\}}$.
- $W_n^+ = \sum_{1 \leq i < j \leq n} \mathbf{1}_{\{X_i + X_j > 0\}} + S_n^+$.
- Sous H_0 , $E(W_n^+) = \frac{n(n+1)}{4}$ et $\text{Var}(W_n^+) = \frac{n(n+1)(2n+1)}{24}$.
- Sous H_0 , $\frac{W_n^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$.

En pratique, pour $n \leq 20$, on utilise une table de Wilcoxon. Pour $n > 20$, on utilise l'approximation gaussienne.

Remarque 3.3 *On montre que le test des rangs signés est plus puissant que le test du signe. De plus, il est utilisable sans problèmes même pour les très petits échantillons. Donc il est conseillé d'utiliser le test des rangs signés plutôt que le test du signe.*

3.3 Test sur deux populations

Nous allons construire un test d'homogénéité. On observe deux échantillons de taille respective n et m , (X_1, \dots, X_n) et (Y_1, \dots, Y_m) . On note F la fonction de répartition de chacune des variables X_i et G la fonction de répartition de chacune des variables Y_i .

Hypothèse du test : $H_0 : F = G$ contre $H_1 : F \neq G$.

Mais on peut aussi s'intéresser aux hypothèses :

- $H_2 : "F > G"$, qui signifie que les X_i sont stochastiquement inférieurs aux Y_j .
- $H_3 : "F < G"$, qui signifie que les X_i sont stochastiquement supérieurs aux Y_j .

C'est ce genre d'hypothèses que l'on utilisera si on cherche à déterminer si un traitement est plus efficace qu'un autre.

3.3.1 Test de Kolmogorov Smirnov (KS)

Statistique du test :

$$D_{m,n} = \sup_{t \in \mathbb{R}} |F_n(t) - G_m(t)|.$$

où,

$$\forall t \in \mathbb{R}, \quad F_n(t) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq t\}} \quad \text{et} \quad G_m(t) = \frac{1}{m} \sum_{i=1}^m 1_{\{Y_i \leq t\}}$$

Théorème 3.3 Avec les hypothèses données ci-dessus on a, sous " $H_0 : F = G$ " :

$$\mathbb{P} \left(\sqrt{\frac{nm}{n+m}} D_{m,n} \leq \lambda \right) \rightarrow 1 + 2 \sum_{k=1}^{+\infty} (-1)^k \exp(-2k^2 \lambda^2)$$

Règle de décision du test : On cherche (dans une table) le quantile $k_{1-\alpha}$ de la loi de Kolmogorov. Si $D_{m,n} > k_{1-\alpha}$, On rejette H_0 au risque α de ce tromper. Ce test est asymptotiquement de niveau α et sa puissance tend vers 1 quand n tend vers $+\infty$.

3.3.2 Cramér-Von Mises et d'Anderson-Darling

Statistique du test : La statistique de Cramér-von Mises est définie par

$$C_{n,m} = \frac{mn}{m+n} \int_{\mathbb{R}} \{F_n(x) - G_m(x)\}^2 dH_{n,m}(x),$$

où

$$H_{n,m}(x) = \frac{1}{n+m} \left\{ \sum_{i=1}^n \mathbb{I}(X_i \leq x) + \sum_{i=1}^m \mathbb{I}(Y_i \leq x) \right\}$$

est la fonction de répartition empirique combinée. À noter que

$$H_{n,m}(x) = \frac{n}{n+m} F_n(x) + \frac{m}{n+m} G_m(x).$$

Pour faciliter le calcul de $C_{n,m}$, soient Z_1, \dots, Z_{n+m} , l'échantillon obtenu en combinant X_1, \dots, X_n et Y_1, \dots, Y_m , et en plaçant les observations en ordre croissant. $H_{n,m}$ est la fonction de répartition de Z_1, \dots, Z_{n+m} , on a

$$C_{n,m} = \frac{mn}{(n+m)^2} \sum_{i=1}^{n+m} \{F_n(Z_i) - G_m(Z_i)\}^2.$$

Le lemme suivant donne une formule de calcul utile pour $C_{n,m}$.

Lemme 3.1 Si R_{X_i} est le rang de X_i parmi $X_1, \dots, X_n, Y_1, \dots, Y_m$ et R_{Y_i} est le rang de Y_i parmi $X_1, \dots, X_n, Y_1, \dots, Y_m$, alors

$$C_{n,m} = \frac{A_{n,m}}{nm(n+m)^2} - \frac{4nm-1}{6(m+n)},$$

où

$$A_{n,m} = n \sum_{i=1}^n (R_{X_i} - i)^2 + m \sum_{i=1}^m (R_{Y_i} - i)^2.$$

Décision du test : L'hypothèse nulle H_0 est rejeté lorsque $C_{n,m} > q_{1-\alpha}$, où les valeurs de $q_{1-\alpha}$ se retrouvent dans l'article de Buning (2002).

3.3.3 Test de Corrélation de rang de Spearman

Pour valider l'existence d'un lien entre deux variables, on réalise ordinairement une régression linéaire simple, voire une régression non linéaire. La qualité du lien supposé est mesurée par le coefficient de corrélation (dit « de Pearson »). Cependant, il existe des situations pour lesquelles une mesure de la corrélation sur les valeurs est inadaptée. Si les variables sont ordinales, discrètes, ou si des valeurs extrêmes risquent de biaiser les résultats, ou encore que les valeurs en elles-mêmes n'ont que peu d'importance, ou enfin qu'elles ne suivent pas une loi normale, il nous reste un joker : les corrélations des rangs.

On n'utilise alors pas les VALEURS des observations dans les calculs mais leur RANG.

Le rang de chaque élément dans la série croissante de X et de Y sera calculé. On calcule ensuite le coefficient de corrélation r_s entre la série des rangs de X et la série des rangs de Y . On peut retrouver cette valeur ainsi :

1. calculer la différence de classement d_i pour chaque couple de valeur (r_{x_i}, r_{y_i}) .
2. la valeur r_s sera donnée par

$$r_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n d_i^2$$

La variable R_s sous l'hypothèse d'indépendance des deux variables a pour espérance $E(R_s) = 0$ et pour variance $V(R_s) = \frac{1}{n-1}$.

Si $n > 30$ alors $Z = \frac{R_s - E(R_s)}{\sqrt{V(R_s)}} = R_s \sqrt{n-1}$ suit la loi normale centrée réduite. Si $n \leq 30$, les valeurs théoriques sont données dans la table du coefficient r de Spearman.

Exemple 3.3 Placez les enfants dans une classe, par ordre ascendant en fonction de leur taille, en prenant note du rang de chaque enfant (premier, deuxième, troisième, etc.), du plus court au plus grand. Vous les placez ensuite en fonction de leur poids, puis vous prenez note de leur rang. Est-ce que chaque enfant occupe le même rang, dans chacune des mesures ? Peut-être que oui, dans l'ensemble, bien qu'un enfant court puisse également être au-dessus de son poids ou qu'un enfant grand, être, lui aussi, en-dessous de son poids, ce qui les classerait dans un rang différent pour chaque variable. La corrélation des rangs démontre le degré de correspondance entre le classement hiérarchique d'un échantillonnage d'observations sur deux variables. Les formules de Kendall ou Spearman sont les variétés communes de ce type de corrélations, car elles donnent toutes les deux une valeur de $-1,0$ (classement inverse parfait) à $0,0$ (aucun accord) à $+1,0$ (classement identique des deux variables).

On ordonne la taille et le poids sur 10 enfants. On obtient les résultats suivants :

Enfant °	1	2	3	4	5	6	7	8	9	10
Taille	1	5	3	8	10	4	2	7	6	9
Poids	5	3	9	10	2	1	6	8	7	4

On trouve $r_s = -0,07$. On accepte H_0 : indépendance des deux caractères.

3.3.4 Test de Wilcoxon (Wilcoxon signed rank test)

Le test de Wilcoxon est un test non paramétrique qui permet de tester si deux populations appariées ont même moyenne μ_1 et μ_2 en se basant sur deux échantillons.

1. Hypothèses du test

- $H_0 : \mu_1 = \mu_2$
- $H_1 : \text{à définir}$

2. Statistique du test

Soient $d_i = x_i - y_i$ (pour $i = 1$ à n) les différences non nulles entre chaque paire d'observations. Nous cherchons les rangs r_i des $|d_i|$. La statistique de rangs signés de Wilcoxon s'écrit :

$$W = \min \left(\sum_{d_i > 0} r_i, \sum_{d_i < 0} r_i \right)$$

3. Règle de décision

On détermine $W_{\alpha/2}$ dans une table de Wilcoxon. Pour un test bilatéral, on rejette H_0 , si On ne peut rejeter H_0 si $W < W_{\alpha/2}$. On ne peut pas rejeter par contre H_0 si $W \in]W_{\alpha/2}, W_{1-\alpha/2}[$ avec $W_{1-\alpha/2} = n(n+1)/2 - W_{\alpha/2}$. Par ailleurs lorsque la taille de l'échantillon est grandes ($n > 25$) on utilise alors la statistique suivante

$$Z = \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \sim \mathcal{N}(0, 1)$$

qui est une approximation normale pour décider.

Exemple 3.4 *Un échantillon aléatoire de dix étudiants est consulté pour noter, dans un test à l'aveugle, la qualité de deux types de bière, l'une locale, l'autre importée. Les notes sont sur une échelle de 1 (pas bon) à 10 (excellent). Utiliser le test de Wilcoxon pour tester l'hypothèse nulle "la distribution des différences entre paires est centrée sur zéro" contre l'hypothèse alternative "La population des étudiants buveurs de bières préfère la catégorie d'importation."*

Etudiant	Locale	Import	Etudiant	Locale	Import
A	2	6	F	4	8
B	3	5	G	3	9
C	7	6	H	4	6
D	8	8	I	5	4
E	7	5	J	6	9

Différences : -4; -2; 1; 0; 2; -4; -6; -2; 1; -3

Tri	0	1	1	-2	2	-2	-3	-4	-4	-6
rang	-	1	2	3	4	5	6	7	8	9
rang moyen		1,5	1,5	4	4	4	6	7,5	7,5	9
$r_i > 0$		1,5	1,5		4					
$r_i < 0$				4		4	6	7,5	7,5	9

Ainsi $W = \min(7, 38) = 7$

On a $W_{0.05} = 8$ (test unilatéral).

3.3.5 Test de Mann-Whitney

Le test de Mann-Whitney est un test non paramétrique qui permet de tester si deux échantillons issus de populations indépendantes ont même moyenne. Les valeurs doivent être numériques (i.e. pouvoir être ordonnées). Il ne nécessite pas de connaître les valeurs des échantillons mais leurs rangs. On suppose donc toujours que l'on dispose de deux échantillons x_1, \dots, x_n et y_1, \dots, y_m issus de variables numériques ou ordinales indépendantes, de lois respectives \mathcal{L}_X et \mathcal{L}_Y .

1. Hypothèse à tester

- $H_0 : \mathcal{L}_X = \mathcal{L}_Y$ ou encore par rapport aux fonctions de distribution $H_0 : F_X = F_Y$.
- $H_1 : \mathcal{L}_X \neq \mathcal{L}_Y$ ou $F_X \neq F_Y$

2. Statistique du test :

$$U_{n,m} = \sum_{i=1}^n \sum_{j=1}^m 1_{\{x < y\}}(X_i, Y_j)$$

où $1_{\{x < y\}}(X_i, Y_j)$ vaut 1 si $X_i < Y_j$, 0.5 si $X_i = Y_j$ et 0 sinon.

On notera U_1 cette valeur obtenue à partir du premier échantillon et U_2 la valeur trouvée en échangeant les rôles des échantillons. Seule la plus petite des deux valeurs trouvées sera comparée aux tables.

On notera U_1 cette valeur obtenue à partir du premier échantillon et U_2 la valeur trouvée en échangeant les rôles des échantillons. Seule la plus petite des deux valeurs trouvées sera comparée aux tables.

On peut également calculer cette statistique en considérant la somme R_1 de tous les rangs après ordonnancement des observations de la première population. On a alors

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2}$$

On aura de même,

$$U_2 = R_2 - \frac{n_2(n_2 + 1)}{2}$$

où R_2 est la somme des rangs du deuxième échantillon.

En sachant que $R_1 + R_2 = N(N+1)/2$ avec $N = n_1 + n_2$, on trouve que

$$U_1 + U_2 = n_1 n_2$$

Ce qui permet de vérifier le calcul de U_1 et U_2 .

3. **Règle de décision** Dans le cas d'un test bilatéral, on détermine $u = \min(u_1, u_2)$. On rejette H_0 si $u \in [0, m_\alpha]$ avec m_α donné par la table de Mann et Whitney.

Remarque 3.4 Si on suppose que $U_1 = U_2$ On utilise plutôt la statistique

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

dont la loi approximative est la loi normale centrée réduite pour décider.

Exemple 3.5 La taille des feuilles de ronces ont été mesurées pour voir si il y a une différence entre la taille des feuilles qui poussent en plein soleil et celles qui poussent à l'ombre. Les résultats sont les suivants (Largeur des feuilles en cm)

Soleil	6.0	4.8	5.1	5.5	4.1	5.3	4.5	5.1
Ombre	6.5	5.5	6.3	7.2	6.8	5.5	5.9	5.5

Valeurs ordonnées

E1	4.1	4.5	4.8	5.1	5.1	5.3	5.5	6.0								
E2							5.5	5.5	5.5	5.9	6.3	6.5	6.8	7		
rang	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
rang moyen	1	2	3	4.5	4.5	6	8.5	8.5	8.5	8.5	11	12	13	14	15	16

$$U_1 = 8 + 8 + 8 + 8 + 8 + 8 + 6.5 + 4 = 58.5$$

$$U_2 = 1.5 + 1.5 + 1.5 + 1 + 0 + 0 + 0 + 0 = 5.5$$

$$R_1 = 1 + 2 + 3 + 4.5 + 4.5 + 6 + 8.5 + 12 = 41.5$$

$$R_2 = 8.5 + 8.5 + 8.5 + 11 + 13 + 14 + 15 + 16 = 94.5$$

Dans tous les cas, on obtient la valeur $U = \min(U_1, U_2) = 5.5$.

Ensuite on utilise la table de Mann-Whitney au risque de 5% ($n_1 = 8, n_2 = 8$), pour obtenir une valeur $m_\alpha = 13$.

On rejettera l'hypothèse nulle si U est inférieure à la valeur m_α . Dans le cas de l'exemple, comme $U < m_\alpha$, on rejette H_0 . La différence entre la taille des feuilles à l'ombre et au soleil est donc significative au risque $\alpha = 5\%$.

3.3.6 Test de Kruskal et Wallis

On souhaite comparer plus de k échantillons, avec $k \geq 2$. Pour cela on désigne par $X_1^i, \dots, X_{n_i}^i$ le $i^{\text{ème}}$ échantillon $i \in [1, k]$ de fonction de répartition F_i et par $n = \sum_{i=1}^k n_i$.

Hypothèse du test : On veut tester $F_1 = \dots, F_k$ contre $H_1 = \bar{H}_0$.

Statistique du test : On ordonne l'ensemble des n observations et on note :

- R_j^i = rang de X_j^i dans l'échantillon global.
- $R^i = \sum_{j=1}^{n_i} R_j^i$ = somme des rangs des observations du $i^{\text{ème}}$ échantillon dans l'échantillon global.

La statistique du test est

$$K_n = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R^i{}^2}{n_i} - 3(n+1)$$

Propriété 3.5 Sous H_0 , $K_n \xrightarrow{\mathcal{L}} \chi_{k-1}^2$.

En pratique, l'approximation par la loi du χ^2 est valide dès qu'il y a au moins 5 observations par échantillon.

On rejette H_0 si $K_n > z_{k-1, \alpha}$, où $z_{k-1, \alpha}$ est le quantile d'ordre $1 - \alpha$ de la loi χ_{k-1}^2 .

3.4 Test du χ^2

3.4.1 Construction du test

On considère des variables aléatoires $(X_k)_{1 \leq k \leq n}$ indépendantes identiquement distribuées, à valeurs dans $\{1, \dots, m\}$ et on note $\mathbb{P}(X_k = i) = \pi_i$. On définit

$$N_i(n) = \sum_{k=1}^n 1_{\{X_k=i\}}$$

c'est-à-dire le nombre fois que la valeur i apparaît dans l'échantillon (X_1, X_2, \dots, X_n) . On s'intéresse à la répartition de $N(n) = (N_1(n), N_2(n), \dots, N_m(n))$. On a pour tout (i_1, i_2, \dots, i_m) tels que $0 \leq i_1, i_2, \dots, i_m \leq n$ et $i_1 + \dots + i_m = n$

$$\mathbb{P}(N_1(n) = i_1, \dots, N_m(n) = i_m) = C_n^{i_1} C_{n-i_1}^{i_2} C_{n-i_1-i_2}^{i_3} \dots 1 \cdot \pi_1^{i_1} \dots \pi_m^{i_m} = \frac{n!}{i_1! \dots i_m!} \pi_1^{i_1} \dots \pi_m^{i_m}$$

on obtient une répartition multinomiale.

Nous allons appliquer la convergence en loi au vecteur $N(n)$. Donnons d'abord une version vectorielle du théorème central limite.

Théorème 3.4 Soit $(X_l)_{l \geq 1}$ une suite de vecteurs de aléatoires de \mathbb{R}^k , indépendants identiquement distribués tels que $\mathbb{E}(X_1) = \mu \in \mathbb{R}^k$ et $\Gamma = \mathbb{E}[(X_1 - \mu)^t (X_1 - \mu)^t]$ matrice de covariance $k \times k$. Alors on a

$$\frac{\sum_{j=1}^n X_j - n\mu}{\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{L} \mathcal{N}(0, \Gamma)$$

Remarquons que $X \in \mathbb{R}^k$ suit une loi $\mathcal{N}_k(0, \Gamma)$ si sa fonction caractéristique est

$$\mathbb{E}(e^{i\langle t, X \rangle}) = \exp(-1/2 \langle t, \Gamma t \rangle)$$

et sa densité, lorsque Γ^{-1} est alors

$$f(x) = \frac{1}{(2\pi)^{k/2} \det(\Gamma)} \exp\left\{-\frac{1}{2} \langle x, \Gamma^{-1} x \rangle\right\}$$

En appliquant ce résultat on obtient :

Théorème 3.5 Avec les notations du début du paragraphe, et en posant $\pi^t = (\pi_1, \dots, \pi_m)$ on obtient :

$$\frac{N(n) - n\pi^t}{\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{L} \mathcal{N}(0, \Delta_\pi - \pi^t \pi) \quad \text{où} \quad \Delta_\pi = \begin{pmatrix} \pi_1 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & \pi_m \end{pmatrix}$$

On obtient alors le résultat qui suit, qui nous donne une convergence vers une loi du χ^2 .

Théorème 3.6 Avec les notations précédentes on a

$$D_n(\pi) = \sum_{i=1}^m \frac{(N_i(n) - n\pi_i)^2}{n\pi_i} \xrightarrow[n \rightarrow +\infty]{L} \chi^2(m-1)$$

3.4.2 Quatre applications du test du Chi deux

Première application : test d'ajustement (loi discrète)

Soit (X_1, \dots, X_n) un n -échantillon de la variable X à valeurs dans $\{1, \dots, m\}$ et de loi inconnue π , où $\mathbb{P}(X = i) = \pi_i$. On teste $H_0 : \pi = p$ contre $H_1 : \pi \neq p$ (p loi de probabilité discrète sur $\{1, \dots, m\}$).

Supposons que H_1 soit vraie, alors :

$$\frac{N(n)}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} p' \neq p$$

d'après la loi forte des grands nombres. Par conséquent on obtient

$$D_n(p) = \sum_{i=1}^m \frac{n}{p_i} \left(\frac{N_i(n)}{n} - p_i \right)^2 \xrightarrow[n \rightarrow +\infty]{p.s.} +\infty$$

car sous $p' \neq p$ et donc il existe i tel que

$$\left(\frac{N_i(n)}{n} - p_i \right)^2 \xrightarrow[n \rightarrow +\infty]{} (p'_i - p_i)^2 > 0$$

Donc si H_0 est fausse on a $D_n(p) \rightarrow +\infty$, on peut donc utiliser une région de rejet de type $D_n(p) > C$ (unilatérale). Si on choisit un risque de première espèce inférieur à α , c'est-à-dire que l'on veut à la limite $\mathbb{P}(\chi^2(m-1) > \chi^2_{1-\alpha}(m-1)) \leq \alpha$ car d'après la convergence en loi :

$$\mathbb{P}(D_n(p) > \chi^2_{1-\alpha}(m-1)) \rightarrow \mathbb{P}(\chi^2(m-1) > \chi^2_{1-\alpha}(m-1)) \leq \alpha$$

Le test est donc asymptotiquement de niveau α , de plus sa puissance est 1 car sous H_1 , $D_n(p) \rightarrow +\infty$ et donc

$$\mathbb{P}(D_n(p) > \chi^2_{1-\alpha}(m-1)) \rightarrow 1$$

Exemple 3.6 On lance 200 fois une pièce et on obtient 110 piles. Le nombre de piles obtenu est une variable aléatoire $\mathcal{B}(200, p)$. On veut tester : $H_0 : p = 1/2$ contre $H_1 : p \neq 1/2$.

	Effectifs observés $N_i(n)$	Effectifs théoriques np_i	Ecart $N_i(n) - np_i$	$(N_i(n) - np_i)^2 / np_i$
<i>pile</i>	110	100	10	1
<i>face</i>	90	100	-10	1
<i>total</i>	200	200	0	2

On a $D_n(p) = 2$ et pour $\alpha = 0.05$ on trouve $\chi_{1-0.05}^2(2-1) = \chi_{1-0.05}^2(1) = 3,84$. On trouve donc $D_n(p) < \chi_{1-\alpha}^2(m-1)$ et on accepte l'hypothèse H_0 .

Deuxième application : test d'ajustement (loi continue)

Soit (X_1, \dots, X_n) un n -échantillon de la variable X qui suit une loi absolument continue F inconnue. On remplace la loi des X_i par une loi discrète. Pour ce faire on procède de la manière suivante : on considère des intervalles I_j tels que

$$\mathbb{R} = \cup_{j=1}^m I_j \quad \text{avec} \quad I_j \cap I_l = \emptyset \text{ si } j \neq l$$

Si on veut alors tester $H_0 : F = F_0$ contre $H_0 : F \neq F_0$ on se ramène au cas précédent en posant

$$\pi_j = \mathbb{P}_F(X \in I_j) \quad p_j = \mathbb{P}_{F_0}(X \in I_j)$$

Troisième application : test d'ajustement (famille de lois)

Soit (X_1, \dots, X_n) un n -échantillon de la variable X à valeurs dans $\{1, \dots, m\}$, qui suit une loi $(P_\theta)_{\theta \in \Theta}$, avec θ inconnue. On a

$$D_n(p(\theta)) = \sum_{i=1}^m \frac{(N_i(n) - np_i(\theta))^2}{np_i(\theta)}$$

On remplace θ par un estimateur, par exemple l'estimateur du maximum de vraisemblance $\hat{\theta}$, puis on procède comme dans la première application en remarquant que si $\Theta \in \mathbb{R}^k$

$$D_n(p(\hat{\theta})) \rightarrow \chi^2(m - k - 1)$$

Exemple 3.7 On a 200 séries de 400 pièces chacune. Pour $k = 1, \dots, 200$, on note X_k le nombre de pièces défectueuses dans la k -ième série. On veut tester l'hypothèse $H_0 : \pi_i = p_i = \frac{e^{-\lambda} \lambda^i}{i!}, i \geq 0$. Le tableau ci-dessous nous permet de déterminer

$$\hat{\lambda}_{200} = \frac{1}{200} \sum_{i=1}^{200} i N_i(200) = 4$$

i	≤ 1	2	3	4	5	6	7	≥ 8
$N_i(200)$	17	31	37	41	30	23	18	8
$np_i(\hat{\lambda}_{200})$	18.4	29.3	39.1	39.1	31.3	20.8	11.9	10.2

On obtient

$$D_{200}(\hat{\lambda}_{200}) = \frac{(17 - 18.4)^2}{18.4} + \frac{(31 - 29.3)^2}{29.3} + \dots + \frac{(8 - 10.2)^2}{10.2} = 1.283$$

On a $m = 8$ et $k = 1$ d'où il faut regarder $\chi^2(6)$. Pour $\alpha = 0.05$ le quantile est $\chi_{0.95}^2(6) = 12.69$. On accepte donc H_0 .

Quatrième application : test d'indépendance

On observe un échantillon $((X_1, Y_1), \dots, (X_n, Y_n))$ à valeur dans $\{(x(i), y(j)), 1 \leq i \leq r, 1 \leq j \leq s\}$. On veut tester $H_0 : X$ et Y sont indépendants. Soit $N_{i,j}$ le nombre de couples observés $(x(i), y(j))$ parmi n observations. On pose $N_{i.} = \sum_{j=1}^s N_{i,j}$ et $N_{.j} = \sum_{i=1}^r N_{i,j}$. On peut alors montrer que sous ces hypothèses

$$\xi_n = n \sum_{i=1}^r \sum_{j=1}^s \frac{(N_{i,j} - N_{i.}N_{.j}/n)^2}{N_{i.}N_{.j}} \xrightarrow[n \rightarrow +\infty]{L} \chi^2(r-1)(s-1)$$

On procède ensuite comme dans les paragraphes précédents.

3.5 Exercices

Exercice 3.1 Deux groupes de 10 étudiants ayant suivi une formation différente ont subi le même examen. Le classement de l'examen est le suivant :

groupe A	1		3	4	5		7	8	8ex			12			15		17			
groupe B		2				6				10	11		13	14		15ex		18	19	20

On veut déterminer si les différences de formation influencent significativement les résultats.

1. Montrer qu'il faut utiliser un test non paramétrique.
2. Appliquer le test de Mann et Whitney et conclure.

Exercice 3.2 On a dosé la teneur en calcium de trois types d'eau issus d'origines géographiques différentes. Chacun d'eux a fait l'objet de quatre prélèvements, dont les résultats sont exprimés ici en mg/l.

Eau 1 : 18; 20; 22; 25,

Eau 2 : 15; 16; 17; 21,

Eau 3 : 15; 20; 21; 25.

L'origine géographique de ces eaux a-t-elle une influence significative sur leur teneur en calcium ?

Exercice 3.3 On a étudié l'activité d'une enzyme, l'actylcholinestérase, chez des animaux soumis à l'action d'un insecticide organophosphoré. Elle est exprimé ici en micromoles de substrat hydrolysé par minute et par mg de protéines. Les résultats obtenus sur des échantillons indépendants en fonction du temps d'exposition sont fournis par le tableau suivant.

<i>aucune exposition</i>	<i>1 jour</i>	<i>2 jours</i>	<i>3jours</i>
15,0	15,0	2,0	0,5
8,5	9,0	2,2	3,0
10,0	8,0	4,0	2,3
10,0	2,0	2,4	0,6
7,6	5,0	1,1	0,9
5,0	3,0	0,7	0,5

L'insecticide entraîne-t-il une diminution significative de l'activité de l'enzyme ? (On comparera globalement les quatre échantillons.)

BIBLIOGRAPHIE

BüNING, H. (2002). Robustness and power of modified Lepage, KolmogorovSmirnov and Cramér-von Mises two-sample tests. J. Appl. Stat. 29, 907924.

Fan, J. et Gijbels, I. (1996). Local polynomial modelling and its applications. Chapman & Hall, New York.

Green, P.J. et Silverman, B.W. (2000). Nonparametric regression and generalized linear models. Chapman & Hall, New York.

Härdle, W. (1990) : Applied Nonparametric Regression. Cambridge University Press, Cambridge.

Hart, J.D. (1997). Nonparametric smoothing and lack-of-fit tests. Springer, New York.

Loader, C. (1999). Local regression and likelihood. Springer, New York.

Silverman, B.W. (1986) : Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

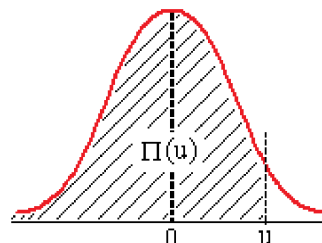
Simonoff, J.S. (1996). Smoothing methods in Statistics. Springer.

ANNEXE

Fonction de répartition Π de la loi normale centrée
réduite : $U \rightarrow N(0, 1)$.

Probabilité de trouver une valeur inférieure à u .

$\Pi(u) = P(U \leq u)$; $\Pi(-u) = P(U \leq -u) = 1 - \Pi(u)$



u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992

Exemple : $\Pi(1.26) = P(U \leq 1.26) = 0.89617 = 89.62\%$

Pour $P < 0.5$ (colonne de gauche et ligne supérieure). les fractiles sont négatifs.

Pour $P > 0.5$ (colonne de droite et ligne inférieure). les fractiles sont positifs.

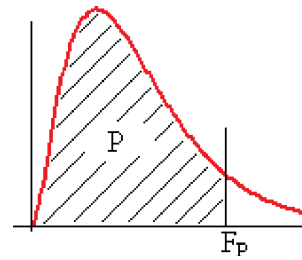
P	0	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009	0.01	
0	infini	3.0902	2.8782	2.7478	2.6521	2.5758	2.5121	2.4573	2.4089	2.3656	2.3263	0.99
0.01	2.3263	2.2904	2.2571	2.2262	2.1973	2.1701	2.1444	2.1201	2.0969	2.0748	2.0537	0.98
0.02	2.0537	2.0335	2.0141	1.9954	1.9774	1.9600	1.9431	1.9268	1.9110	1.8957	1.8808	0.97
0.03	1.8808	1.8663	1.8522	1.8384	1.8250	1.8119	1.7991	1.7866	1.7744	1.7624	1.7507	0.96
0.04	1.7507	1.7392	1.7279	1.7169	1.7060	1.6954	1.6849	1.6747	1.6646	1.6546	1.6449	0.95
0.05	1.6449	1.6352	1.6258	1.6164	1.6072	1.5982	1.5893	1.5805	1.5718	1.5632	1.5548	0.94
0.06	1.5548	1.5464	1.5382	1.5301	1.5220	1.5141	1.5063	1.4985	1.4909	1.4833	1.4758	0.93
0.07	1.4758	1.4684	1.4611	1.4538	1.4466	1.4395	1.4325	1.4255	1.4187	1.4118	1.4051	0.92
0.08	1.4051	1.3984	1.3917	1.3852	1.3787	1.3722	1.3658	1.3595	1.3532	1.3469	1.3408	0.91
0.09	1.3408	1.3346	1.3285	1.3225	1.3165	1.3106	1.3047	1.2988	1.2930	1.2873	1.2816	0.90
0.10	1.2816	1.2759	1.2702	1.2646	1.2591	1.2536	1.2481	1.2426	1.2372	1.2319	1.2265	0.89
0.11	1.2265	1.2212	1.2160	1.2107	1.2055	1.2004	1.1952	1.1901	1.1850	1.1800	1.1750	0.88
0.12	1.1750	1.1700	1.1650	1.1601	1.1552	1.1503	1.1455	1.1407	1.1359	1.1311	1.1264	0.87
0.13	1.1264	1.1217	1.1170	1.1123	1.1077	1.1031	1.0985	1.0939	1.0893	1.0848	1.0803	0.86
0.14	1.0803	1.0758	1.0714	1.0669	1.0625	1.0581	1.0537	1.0494	1.0451	1.0407	1.0364	0.85
0.15	1.0364	1.0322	1.0279	1.0237	1.0194	1.0152	1.0110	1.0069	1.0027	0.9986	0.9945	0.84
0.16	0.9945	0.9904	0.9863	0.9822	0.9782	0.9741	0.9701	0.9661	0.9621	0.9581	0.9542	0.83
0.17	0.9542	0.9502	0.9463	0.9424	0.9385	0.9346	0.9307	0.9269	0.9230	0.9192	0.9154	0.82
0.18	0.9154	0.9116	0.9078	0.9040	0.9002	0.8965	0.8927	0.8890	0.8853	0.8816	0.8779	0.81
0.19	0.8779	0.8742	0.8706	0.8669	0.8632	0.8596	0.8560	0.8524	0.8488	0.8452	0.8416	0.80
0.20	0.8416	0.8381	0.8345	0.8310	0.8274	0.8239	0.8204	0.8169	0.8134	0.8099	0.8064	0.79
0.21	0.8064	0.8030	0.7995	0.7961	0.7926	0.7892	0.7858	0.7824	0.7790	0.7756	0.7722	0.78
0.22	0.7722	0.7688	0.7655	0.7621	0.7588	0.7554	0.7521	0.7488	0.7454	0.7421	0.7388	0.77
0.23	0.7388	0.7356	0.7323	0.7290	0.7257	0.7225	0.7192	0.7160	0.7128	0.7095	0.7063	0.76
0.24	0.7063	0.7031	0.6999	0.6967	0.6935	0.6903	0.6871	0.6840	0.6808	0.6776	0.6745	0.75
0.25	0.6745	0.6713	0.6682	0.6651	0.6620	0.6588	0.6557	0.6526	0.6495	0.6464	0.6433	0.74
0.26	0.6433	0.6403	0.6372	0.6341	0.6311	0.6280	0.6250	0.6219	0.6189	0.6158	0.6128	0.73
0.27	0.6128	0.6098	0.6068	0.6038	0.6008	0.5978	0.5948	0.5918	0.5888	0.5858	0.5828	0.72
0.28	0.5828	0.5799	0.5769	0.5740	0.5710	0.5681	0.5651	0.5622	0.5592	0.5563	0.5534	0.71
0.29	0.5534	0.5505	0.5476	0.5446	0.5417	0.5388	0.5359	0.5330	0.5302	0.5273	0.5244	0.70
0.30	0.5244	0.5215	0.5187	0.5158	0.5129	0.5101	0.5072	0.5044	0.5015	0.4987	0.4958	0.69
0.31	0.4958	0.4930	0.4902	0.4874	0.4845	0.4817	0.4789	0.4761	0.4733	0.4705	0.4677	0.68
0.32	0.4677	0.4649	0.4621	0.4593	0.4565	0.4538	0.4510	0.4482	0.4454	0.4427	0.4399	0.67
0.33	0.4399	0.4372	0.4344	0.4316	0.4289	0.4261	0.4234	0.4207	0.4179	0.4152	0.4125	0.66
0.34	0.4125	0.4097	0.4070	0.4043	0.4016	0.3989	0.3961	0.3934	0.3907	0.3880	0.3853	0.65
0.35	0.3853	0.3826	0.3799	0.3772	0.3745	0.3719	0.3692	0.3665	0.3638	0.3611	0.3585	0.64
0.36	0.3585	0.3558	0.3531	0.3505	0.3478	0.3451	0.3425	0.3398	0.3372	0.3345	0.3319	0.63
0.37	0.3319	0.3292	0.3266	0.3239	0.3213	0.3186	0.3160	0.3134	0.3107	0.3081	0.3055	0.62
0.38	0.3055	0.3029	0.3002	0.2976	0.2950	0.2924	0.2898	0.2871	0.2845	0.2819	0.2793	0.61
0.39	0.2793	0.2767	0.2741	0.2715	0.2689	0.2663	0.2637	0.2611	0.2585	0.2559	0.2533	0.60
0.40	0.2533	0.2508	0.2482	0.2456	0.2430	0.2404	0.2378	0.2353	0.2327	0.2301	0.2275	0.59
0.41	0.2275	0.2250	0.2224	0.2198	0.2173	0.2147	0.2121	0.2096	0.2070	0.2045	0.2019	0.58
0.42	0.2019	0.1993	0.1968	0.1942	0.1917	0.1891	0.1866	0.1840	0.1815	0.1789	0.1764	0.57
0.43	0.1764	0.1738	0.1713	0.1687	0.1662	0.1637	0.1611	0.1586	0.1560	0.1535	0.1510	0.56
0.44	0.1510	0.1484	0.1459	0.1434	0.1408	0.1383	0.1358	0.1332	0.1307	0.1282	0.1257	0.55
0.45	0.1257	0.1231	0.1206	0.1181	0.1156	0.1130	0.1105	0.1080	0.1055	0.1030	0.1004	0.54
0.46	0.1004	0.0979	0.0954	0.0929	0.0904	0.0878	0.0853	0.0828	0.0803	0.0778	0.0753	0.53
0.47	0.0753	0.0728	0.0702	0.0677	0.0652	0.0627	0.0602	0.0577	0.0552	0.0527	0.0502	0.52
0.48	0.0502	0.0476	0.0451	0.0426	0.0401	0.0376	0.0351	0.0326	0.0301	0.0276	0.0251	0.51
0.49	0.0251	0.0226	0.0201	0.0175	0.0150	0.0125	0.0100	0.0075	0.0050	0.0025	0.0000	0.50
	0.01	0.009	0.008	0.007	0.006	0.005	0.004	0.003	0.002	0.001	0	P

Exemples : $\Pi(u) = P(U \leq u) = P = 0.6340 \Rightarrow u = 0.3425$; $\Pi(u) = P(U \leq u) = P = 0.4020 \Rightarrow u = -0.2482$

1 : Fractiles de la loi du χ^2_v

Cette table donne les fractiles F_P de la loi de khi-deux

à v degrés de liberté : $P = P(\chi^2_v \leq F_P)$



$v \quad P$	0.010	0.020	0.025	0.050	0.100	0.150	0.200	0.800	0.900	0.950	0.975	0.980	0.990
1	0.000	0.001	0.001	0.004	0.016	0.036	0.064	1.642	2.706	3.841	5.024	5.412	6.64
2	0.020	0.040	0.051	0.103	0.211	0.325	0.446	3.219	4.605	5.991	7.378	7.824	9.21
3	0.115	0.185	0.216	0.352	0.584	0.798	1.005	4.642	6.251	7.815	9.348	9.837	11.35
4	0.297	0.429	0.484	0.711	1.064	1.366	1.649	5.989	7.779	9.488	11.143	11.668	13.28
5	0.554	0.752	0.831	1.145	1.610	1.994	2.343	7.289	9.236	11.070	12.833	13.388	15.09
6	0.872	1.134	1.237	1.635	2.204	2.661	3.070	8.558	10.645	12.592	14.449	15.033	16.81
7	1.239	1.564	1.690	2.167	2.833	3.358	3.822	9.803	12.017	14.067	16.013	16.622	18.48
8	1.646	2.032	2.180	2.733	3.490	4.078	4.594	11.030	13.362	15.507	17.535	18.168	20.09
9	2.088	2.532	2.700	3.325	4.168	4.817	5.380	12.242	14.684	16.919	19.023	19.679	21.67
10	2.558	3.059	3.247	3.940	4.865	5.570	6.179	13.442	15.987	18.307	20.483	21.161	23.21
11	3.053	3.609	3.816	4.575	5.578	6.336	6.989	14.631	17.275	19.675	21.920	22.618	24.73
12	3.571	4.178	4.404	5.226	6.304	7.114	7.807	15.812	18.549	21.026	23.337	24.054	26.22
13	4.107	4.765	5.009	5.892	7.042	7.901	8.634	16.985	19.812	22.362	24.736	25.472	27.69
14	4.660	5.368	5.629	6.571	7.790	8.696	9.467	18.151	21.064	23.685	26.119	26.873	29.14
15	5.229	5.985	6.262	7.261	8.547	9.499	10.307	19.311	22.307	24.996	27.488	28.259	30.58
16	5.812	6.614	6.908	7.962	9.312	10.309	11.152	20.465	23.542	26.296	28.845	29.633	32.00
17	6.408	7.255	7.564	8.672	10.085	11.125	12.002	21.615	24.769	27.587	30.191	30.995	33.41
18	7.015	7.906	8.231	9.390	10.865	11.946	12.857	22.760	25.989	28.869	31.526	32.346	34.81
19	7.633	8.567	8.907	10.117	11.651	12.773	13.716	23.900	27.204	30.144	32.852	33.687	36.19
20	8.260	9.237	9.591	10.851	12.443	13.604	14.578	25.038	28.412	31.410	34.170	35.020	37.57
21	8.897	9.915	10.283	11.591	13.240	14.439	15.445	26.171	29.615	32.671	35.479	36.343	38.93
22	9.542	10.600	10.982	12.338	14.041	15.279	16.314	27.301	30.813	33.924	36.781	37.659	40.29
23	10.196	11.293	11.689	13.091	14.848	16.122	17.187	28.429	32.007	35.172	38.076	38.968	41.64
24	10.856	11.992	12.401	13.848	15.659	16.969	18.062	29.553	33.196	36.415	39.364	40.270	42.98
25	11.524	12.697	13.120	14.611	16.473	17.818	18.940	30.675	34.382	37.652	40.646	41.566	44.31
26	12.198	13.409	13.844	15.379	17.292	18.671	19.820	31.795	35.563	38.885	41.923	42.856	45.64
27	12.879	14.125	14.573	16.151	18.114	19.527	20.703	32.912	36.741	40.113	43.195	44.140	46.96
28	13.565	14.847	15.308	16.928	18.939	20.386	21.588	34.027	37.916	41.337	44.461	45.419	48.28
29	14.256	15.574	16.047	17.708	19.768	21.247	22.475	35.139	39.087	42.557	45.722	46.693	49.59
30	14.953	16.306	16.791	18.493	20.599	22.110	23.364	36.250	40.256	43.773	46.979	47.962	50.89
40	22.164	23.838	24.433	26.509	29.051	30.856	32.345	47.269	51.805	55.758	59.342	60.436	63.69
50	29.707	31.664	32.357	34.764	37.689	39.754	41.449	58.164	63.167	67.505	71.420	72.613	76.15
60	37.485	39.699	40.482	43.188	46.459	48.759	50.641	68.972	74.397	79.082	83.298	84.580	88.38
70	45.442	47.893	48.758	51.739	55.329	57.844	59.898	79.715	85.527	90.531	95.023	96.388	100.42
80	53.540	56.213	57.153	60.391	64.278	66.994	69.207	90.405	96.578	101.88	106.63	108.07	112.33

Exemple : $v = 10$ d.d.l. $P = P(\chi^2_{10} \leq F_P) = 0.95 \Rightarrow F_P = 18.307$

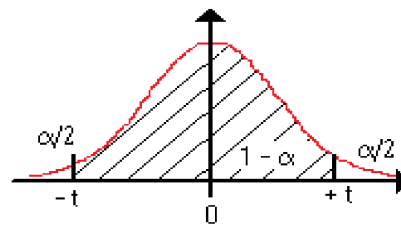
Approximation : Pour $v > 100$ d.l.l. $\chi^2(v) \cong N(v; \sqrt{2v})$ ou $\sqrt{2}\chi^2 - \sqrt{2v-1} \cong N(0,1)$

Table de la Loi de Student

Cette table donne les fractiles de la loi de Student
à ν degrés de liberté : valeur t ayant la probabilité
 α d'être dépassée en valeur absolue :

$$P(|T_\nu| \leq t) = P(-t \leq T_\nu \leq t) = 1 - \alpha.$$

$$P(|T_\nu| > t) = 1 - P(|T_\nu| \leq t) = \alpha$$



ν	α	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.005	0.001
1		0.1584	0.3249	0.5095	0.7265	1	1.3764	1.9626	3.0777	6.3137	12.706	31.821	63.656	127.32	636.58
2		0.1421	0.2887	0.4447	0.6172	0.8165	1.0607	1.3862	1.8856	2.92	4.3027	6.9645	9.925	14.089	31.6
3		0.1366	0.2767	0.4242	0.5844	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8408	7.4532	12.924
4		0.1338	0.2707	0.4142	0.5686	0.7407	0.941	1.1896	1.5332	2.1318	2.7765	3.7469	4.6041	5.5975	8.6101
5		0.1322	0.2672	0.4082	0.5594	0.7267	0.9195	1.1558	1.4759	2.015	2.5706	3.3649	4.0321	4.7733	6.8685
6		0.1311	0.2648	0.4043	0.5534	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074	4.3168	5.9587
7		0.1303	0.2632	0.4015	0.5491	0.7111	0.896	1.1192	1.4149	1.8946	2.3646	2.9979	3.4995	4.0294	5.4081
8		0.1297	0.2619	0.3995	0.5459	0.7064	0.8889	1.1081	1.3968	1.8595	2.306	2.8965	3.3554	3.8325	5.0414
9		0.1293	0.261	0.3979	0.5435	0.7027	0.8834	1.0997	1.383	1.8331	2.2622	2.8214	3.2498	3.6896	4.7809
10		0.1289	0.2602	0.3966	0.5415	0.6998	0.8791	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693	3.5814	4.5868
11		0.1286	0.2596	0.3956	0.5399	0.6974	0.8755	1.0877	1.3634	1.7959	2.201	2.7181	3.1058	3.4966	4.4369
12		0.1283	0.259	0.3947	0.5386	0.6955	0.8726	1.0832	1.3562	1.7823	2.1788	2.681	3.0545	3.4284	4.3178
13		0.1281	0.2586	0.394	0.5375	0.6938	0.8702	1.0795	1.3502	1.7709	2.1604	2.6503	3.0123	3.3725	4.2209
14		0.128	0.2582	0.3933	0.5366	0.6924	0.8681	1.0763	1.345	1.7613	2.1448	2.6245	2.9768	3.3257	4.1403
15		0.1278	0.2579	0.3928	0.5357	0.6912	0.8662	1.0735	1.3406	1.7531	2.1315	2.6025	2.9467	3.286	4.0728
16		0.1277	0.2576	0.3923	0.535	0.6901	0.8647	1.0711	1.3368	1.7459	2.1199	2.5835	2.9208	3.252	4.0149
17		0.1276	0.2573	0.3919	0.5344	0.6892	0.8633	1.069	1.3334	1.7396	2.1098	2.5669	2.8982	3.2224	3.9651
18		0.1274	0.2571	0.3915	0.5338	0.6884	0.862	1.0672	1.3304	1.7341	2.1009	2.5524	2.8784	3.1966	3.9217
19		0.1274	0.2569	0.3912	0.5333	0.6876	0.861	1.0655	1.3277	1.7291	2.093	2.5395	2.8609	3.1737	3.8833
20		0.1273	0.2567	0.3909	0.5329	0.687	0.86	1.064	1.3253	1.7247	2.086	2.528	2.8453	3.1534	3.8496
21		0.1272	0.2566	0.3906	0.5325	0.6864	0.8591	1.0627	1.3232	1.7207	2.0796	2.5176	2.8314	3.1352	3.8193
22		0.1271	0.2564	0.3904	0.5321	0.6858	0.8583	1.0614	1.3212	1.7171	2.0739	2.5083	2.8188	3.1188	3.7922
23		0.1271	0.2563	0.3902	0.5317	0.6853	0.8575	1.0603	1.3195	1.7139	2.0687	2.4999	2.8073	3.104	3.7676
24		0.127	0.2562	0.39	0.5314	0.6848	0.8569	1.0593	1.3178	1.7109	2.0639	2.4922	2.797	3.0905	3.7454
25		0.1269	0.2561	0.3898	0.5312	0.6844	0.8562	1.0584	1.3163	1.7081	2.0595	2.4851	2.7874	3.0782	3.7251
26		0.1269	0.256	0.3896	0.5309	0.684	0.8557	1.0575	1.315	1.7056	2.0555	2.4786	2.7787	3.0669	3.7067
27		0.1268	0.2559	0.3894	0.5306	0.6837	0.8551	1.0567	1.3137	1.7033	2.0518	2.4727	2.7707	3.0565	3.6895
28		0.1268	0.2558	0.3893	0.5304	0.6834	0.8546	1.056	1.3125	1.7011	2.0484	2.4671	2.7633	3.047	3.6739
29		0.1268	0.2557	0.3892	0.5302	0.683	0.8542	1.0553	1.3114	1.6991	2.0452	2.462	2.7564	3.038	3.6595
30		0.1267	0.2556	0.389	0.53	0.6828	0.8538	1.0547	1.3104	1.6973	2.0423	2.4573	2.75	3.0298	3.646
50		0.1263	0.2547	0.3875	0.5278	0.6794	0.8489	1.0473	1.2987	1.6759	2.0086	2.4033	2.6778	2.937	3.496
60		0.1262	0.2545	0.3872	0.5272	0.6786	0.8477	1.0455	1.2958	1.6706	2.0003	2.3901	2.6603	2.9146	3.4602
70		0.1261	0.2543	0.3869	0.5268	0.678	0.8468	1.0442	1.2938	1.6669	1.9944	2.3808	2.6479	2.8987	3.435
80		0.1261	0.2542	0.3867	0.5265	0.6776	0.8461	1.0432	1.2922	1.6641	1.9901	2.3739	2.6387	2.887	3.4164
infini (loi normale)		0.1257	0.2533	0.3853	0.5244	0.6744	0.8416	1.0364	1.2816	1.6449	1.96	2.3264	2.5759	2.8072	3.2908

Exemples : $\nu = 10$ d.d.l. $P(|T_{10}| \leq t) = 0.95 \Rightarrow t = \pm 2.2281$

$P(T_{10} \leq t) = 0.95 \Rightarrow t = + 1.8125$

Table de Fisher pour $\alpha = 0.05$ (5%)

$v_1 \rightarrow$ $v_2 \downarrow$	1	2	3	4	5	6	8	12	24	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,01	1,83	1,61	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

Table de Shapiro-Francia

Tables des valeurs critiques du test de Normalité de Shapiro-Francia, pour trois seuils α différents, en fonction de la taille n de l'échantillon.

$n \backslash \alpha$	0,10	0,05	0,01
5	0,9045	0,8866	0,8336
6	0,9119	0,8942	0,8485
7	0,9193	0,9018	0,8488
8	0,9248	0,9023	0,8676
9	0,9302	0,9113	0,8715
10	0,9351	0,9186	0,8724
11	0,9371	0,9224	0,8889
12	0,9409	0,9261	0,8917
13	0,9505	0,9288	0,8919
14	0,9514	0,9377	0,9063
15	0,9523	0,9398	0,9075
16	0,9532	0,9400	0,9141
17	0,9558	0,9431	0,9143
18	0,9561	0,9439	0,9161
19	0,9590	0,9493	0,9235
20	0,9595	0,9500	0,9283
21	0,9617	0,9542	0,9291
22	0,9626	0,9544	0,9311
23	0,9655	0,9554	0,9329
24	0,9662	0,9557	0,9388
25	0,9669	0,9572	0,9432
26	0,9675	0,9576	0,9437
27	0,9688	0,9599	0,9459
28	0,9701	0,9637	0,9466
29	0,9712	0,9641	0,9490
30	0,9723	0,9651	0,9506
31	0,9734	0,9654	0,9524
32	0,9745	0,9669	0,9538
33	0,9756	0,9678	0,9546
34	0,9758	0,9686	0,9548
35	0,9760	0,9691	0,9551
36	0,9762	0,9695	0,9551
37	0,9763	0,9698	0,9556
38	0,9765	0,9709	0,9560
39	0,9770	0,9712	0,9567
40	0,9775	0,9724	0,9576
41	0,9779	0,9726	0,9604
42	0,9780	0,9730	0,9610
43	0,9782	0,9732	0,9611
44	0,9785	0,9734	0,9631
45	0,9789	0,9743	0,9636
46	0,9793	0,9758	0,9643
47	0,9801	0,9764	0,9646
48	0,9806	0,9767	0,9653
49	0,9807	0,9768	0,9655
50	0,9810	0,9769	0,9667
51	0,9815	0,9772	0,9669
52	0,9818	0,9773	0,9673

$n \backslash \alpha$	0,10	0,05	0,01
53	0,9822	0,9777	0,9685
54	0,9824	0,9780	0,9694
55	0,9825	0,9788	0,9695
56	0,9826	0,9792	0,9697
57	0,9828	0,9796	0,9704
58	0,9830	0,9797	0,9705
59	0,9833	0,9800	0,9707
60	0,9836	0,9801	0,9710
61	0,9838	0,9810	0,9718
62	0,9839	0,9811	0,9722
63	0,9842	0,9812	0,9722
64	0,9843	0,9813	0,9728
65	0,9844	0,9813	0,9732
66	0,9849	0,9817	0,9734
67	0,9852	0,9820	0,9736
68	0,9852	0,9821	0,9739
69	0,9854	0,9821	0,9744
70	0,9855	0,9822	0,9755
71	0,9856	0,9827	0,9757
72	0,9857	0,9828	0,9759
73	0,9859	0,9830	0,9762
74	0,9862	0,9832	0,9764
75	0,9864	0,9834	0,9765
76	0,9866	0,9834	0,9771
77	0,9867	0,9837	0,9774
78	0,9868	0,9839	0,9776
79	0,9869	0,9841	0,9779
80	0,9871	0,9841	0,9780
81	0,9872	0,9845	0,9782
82	0,9873	0,9847	0,9785
83	0,9874	0,9848	0,9786
84	0,9876	0,9848	0,9787
85	0,9877	0,9849	0,9788
86	0,9877	0,9851	0,9790
87	0,9878	0,9853	0,9791
88	0,9880	0,9854	0,9792
89	0,9880	0,9855	0,9793
90	0,9882	0,9857	0,9794
91	0,9883	0,9859	0,9796
92	0,9887	0,9859	0,9800
93	0,9888	0,9861	0,9802
94	0,9888	0,9862	0,9805
95	0,9889	0,9862	0,9807
96	0,9889	0,9863	0,9811
97	0,9890	0,9867	0,9812
98	0,9893	0,9868	0,9814
99	0,9894	0,9872	0,9818
100	0,9896	0,9873	0,9820

Valeurs critiques de la statistique de Kruskal-Wallis pour trois échantillons

Effectifs			Risque de première espèce α				
n_1	n_2	n_3	.10	.05	.01	.005	.001
2	2	2	4.25				
3	2	1	4.29				
3	2	2	4.71	4.71			
3	3	1	4.57	5.14			
3	3	2	4.56	5.36			
3	3	3	4.62	5.60	7.20	7.20	
4	2	1	4.50				
4	2	2	4.46	5.33			
4	3	1	4.06	5.22			
4	3	2	4.51	5.44	6.44	7.00	
4	3	3	4.71	5.73	6.75	7.32	8.02
4	4	1	4.17	4.97	6.67		
4	4	2	4.55	5.60	7.14	7.59	8.32
4	4	3	4.65	5.69	7.66	8.00	8.65
4	4	4	4.20	5.00			
5	2	1	4.20	5.00			
5	2	2	4.36	5.16	6.53		
5	3	1	4.02	4.96			
5	3	2	4.65	5.25	6.82	7.18	
5	3	3	4.53	5.65	7.08	7.51	8.24
5	4	1	3.99	4.99	6.95	7.36	
5	4	2	4.54	5.27	7.12	7.57	8.11
5	4	3	4.55	5.63	7.44	7.91	8.50
5	4	4	4.62	5.62	7.76	8.14	9.00
5	5	1	4.11	5.13	7.31	7.75	
5	5	2	4.62	5.34	7.27	8.13	8.68
5	5	3	4.54	5.71	7.54	8.24	9.06
5	5	4	4.53	5.64	7.77	8.37	9.32
5	5	5	4.56	5.78	7.98	8.72	9.68

Si tous les effectifs des k groupes sont supérieurs ou égaux à 5, nous utiliserons l'approximation $KW \sim \chi_{k-1}^2$.

Remarque : Si votre logiciel le permet, préférez toujours le calcul de p -valeurs « exactes » même par la méthode de Monte Carlo.

Valeur critique pour le test de Mann-Whitney

Test bilatéral pour $\alpha = 5\%$.

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	*	*	*	*	*	*	0	0	0	0	1	1	1	1	1	2	2	2	2
3		*	*	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4			0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14
5				2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6					5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7						8	10	12	14	16	18	20	22	24	26	28	30	32	34
8							13	15	17	19	22	24	26	29	31	34	36	38	41
9								17	20	23	26	28	31	34	37	39	42	45	48
10									23	26	29	33	36	39	42	45	48	52	55
11										30	33	37	40	44	47	51	55	58	62
12											37	41	45	49	53	57	61	65	69
13												45	50	54	59	63	67	72	76
14													55	59	64	69	74	78	83
15														64	70	75	80	85	90
16															75	81	86	92	98
17																87	93	99	105
18																	99	106	112
19																		113	119
20																			127

Une * indique que le test ne peut être significatif au niveau $\alpha = 5\%$.

Test de Wilcoxon pour $\alpha = 5\%$

La table donne le plus grand nombre entier w_α tel que

$$\mathbb{P}_{(\mathcal{H}_0)} [W_n^+ \leq w_\alpha] \leq \alpha/2.$$

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
w_α	*	*	*	*	*	1	2	4	6	8	11	14	17	21	25

Une * indique que le test ne peut être significatif au niveau $\alpha = 5\%$.

Table de Shapiro-Francia

Tables des valeurs critiques du test de Normalité de Shapiro-Francia, pour trois seuils α différents, en fonction de la taille n de l'échantillon.

$n \backslash \alpha$	0,10	0,05	0,01
5	0,9045	0,8866	0,8336
6	0,9119	0,8942	0,8485
7	0,9193	0,9018	0,8488
8	0,9248	0,9023	0,8676
9	0,9302	0,9113	0,8715
10	0,9351	0,9186	0,8724
11	0,9371	0,9224	0,8889
12	0,9409	0,9261	0,8917
13	0,9505	0,9288	0,8919
14	0,9514	0,9377	0,9063
15	0,9523	0,9398	0,9075
16	0,9532	0,9400	0,9141
17	0,9558	0,9431	0,9143
18	0,9561	0,9439	0,9161
19	0,9590	0,9493	0,9235
20	0,9595	0,9500	0,9283
21	0,9617	0,9542	0,9291
22	0,9626	0,9544	0,9311
23	0,9655	0,9554	0,9329
24	0,9662	0,9557	0,9388
25	0,9669	0,9572	0,9432
26	0,9675	0,9576	0,9437
27	0,9688	0,9599	0,9459
28	0,9701	0,9637	0,9466
29	0,9712	0,9641	0,9490
30	0,9723	0,9651	0,9506
31	0,9734	0,9654	0,9524
32	0,9745	0,9669	0,9538
33	0,9756	0,9678	0,9546
34	0,9758	0,9686	0,9548
35	0,9760	0,9691	0,9551
36	0,9762	0,9695	0,9551
37	0,9763	0,9698	0,9556
38	0,9765	0,9709	0,9560
39	0,9770	0,9712	0,9567
40	0,9775	0,9724	0,9576
41	0,9779	0,9726	0,9604
42	0,9780	0,9730	0,9610
43	0,9782	0,9732	0,9611
44	0,9785	0,9734	0,9631
45	0,9789	0,9743	0,9636
46	0,9793	0,9758	0,9643
47	0,9801	0,9764	0,9646
48	0,9806	0,9767	0,9653
49	0,9807	0,9768	0,9655
50	0,9810	0,9769	0,9667
51	0,9815	0,9772	0,9669
52	0,9818	0,9773	0,9673

$n \backslash \alpha$	0,10	0,05	0,01
53	0,9822	0,9777	0,9685
54	0,9824	0,9780	0,9694
55	0,9825	0,9788	0,9695
56	0,9826	0,9792	0,9697
57	0,9828	0,9796	0,9704
58	0,9830	0,9797	0,9705
59	0,9833	0,9800	0,9707
60	0,9836	0,9801	0,9710
61	0,9838	0,9810	0,9718
62	0,9839	0,9811	0,9722
63	0,9842	0,9812	0,9722
64	0,9843	0,9813	0,9728
65	0,9844	0,9813	0,9732
66	0,9849	0,9817	0,9734
67	0,9852	0,9820	0,9736
68	0,9852	0,9821	0,9739
69	0,9854	0,9821	0,9744
70	0,9855	0,9822	0,9755
71	0,9856	0,9827	0,9757
72	0,9857	0,9828	0,9759
73	0,9859	0,9830	0,9762
74	0,9862	0,9832	0,9764
75	0,9864	0,9834	0,9765
76	0,9866	0,9834	0,9771
77	0,9867	0,9837	0,9774
78	0,9868	0,9839	0,9776
79	0,9869	0,9841	0,9779
80	0,9871	0,9841	0,9780
81	0,9872	0,9845	0,9782
82	0,9873	0,9847	0,9785
83	0,9874	0,9848	0,9786
84	0,9876	0,9848	0,9787
85	0,9877	0,9849	0,9788
86	0,9877	0,9851	0,9790
87	0,9878	0,9853	0,9791
88	0,9880	0,9854	0,9792
89	0,9880	0,9855	0,9793
90	0,9882	0,9857	0,9794
91	0,9883	0,9859	0,9796
92	0,9887	0,9859	0,9800
93	0,9888	0,9861	0,9802
94	0,9888	0,9862	0,9805
95	0,9889	0,9862	0,9807
96	0,9889	0,9863	0,9811
97	0,9890	0,9867	0,9812
98	0,9893	0,9868	0,9814
99	0,9894	0,9872	0,9818
100	0,9896	0,9873	0,9820

Table de Kolmogorov-Smirnov

Seuils critiques $D_\alpha(n)$					
n	$\alpha = 0.20$	$\alpha = 0.15$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
1	0.900	0.925	0.950	0.975	0.995
2	0.684	0.726	0.776	0.842	0.929
3	0.565	0.597	0.642	0.708	0.828
4	0.494	0.525	0.564	0.624	0.733
5	0.446	0.474	0.510	0.565	0.669
6	0.410	0.436	0.470	0.521	0.618
7	0.381	0.405	0.438	0.486	0.577
8	0.358	0.381	0.411	0.457	0.543
9	0.339	0.360	0.388	0.432	0.514
10	0.322	0.342	0.368	0.410	0.490
11	0.307	0.326	0.352	0.391	0.468
12	0.295	0.313	0.338	0.375	0.450
13	0.284	0.302	0.325	0.361	0.433
14	0.274	0.292	0.314	0.349	0.418
15	0.266	0.283	0.304	0.338	0.404
16	0.258	0.274	0.295	0.328	0.392
17	0.250	0.266	0.286	0.318	0.381
18	0.244	0.259	0.278	0.309	0.371
19	0.237	0.252	0.272	0.301	0.363
20	0.231	0.246	0.264	0.294	0.356
25	0.210	0.220	0.240	0.270	0.320
30	0.190	0.200	0.220	0.240	0.290
35	0.180	0.190	0.210	0.230	0.270
> 35	$1.07/\sqrt{n}$	$1.14/\sqrt{n}$	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.63/\sqrt{n}$

Table de Wilcoxon

n	P = 0,05	P = 0,02	P = 0,01
5	18	16	15
6	27	24	23
7	37	34	32
8	49	46	43
9	63	59	56
10	79	74	71
11	97	91	87
12	116	110	105
13	137	130	125
14	160	152	147
15	185	176	170
16	212	202	196
17	241	230	223
18	271	259	252
19	303	291	282
20	338	324	315

Valeurs critiques de la statistique de Kruskal-Wallis pour 3 échantillons

Effectifs			Risque de première espèce α				
n_1	n_2	n_3	.10	.05	.01	.005	.001
2	2	2	4.25				
3	2	1	4.29				
3	2	2	4.71	4.71			
3	3	1	4.57	5.14			
3	3	2	4.56	5.36			
3	3	3	4.62	5.60	7.20	7.20	
4	2	1	4.50				
4	2	2	4.46	5.33			
4	3	1	4.06	5.22			
4	3	2	4.51	5.44	6.44	7.00	
4	3	3	4.71	5.73	6.75	7.32	8.02
4	4	1	4.17	4.97	6.67		
4	4	2	4.55	5.60	7.14	7.59	8.32
4	4	3	4.65	5.69	7.66	8.00	8.65
4	4	4	4.20	5.00			
5	2	1	4.20	5.00			
5	2	2	4.36	5.16	6.53		
5	3	1	4.02	4.96			
5	3	2	4.65	5.25	6.82	7.18	
5	3	3	4.53	5.65	7.08	7.51	8.24
5	4	1	3.99	4.99	6.95	7.36	
5	4	2	4.54	5.27	7.12	7.57	8.11
5	4	3	4.55	5.63	7.44	7.91	8.50
5	4	4	4.62	5.62	7.76	8.14	9.00
5	5	1	4.11	5.13	7.31	7.75	
5	5	2	4.62	5.34	7.27	8.13	8.68
5	5	3	4.54	5.71	7.54	8.24	9.06
5	5	4	4.53	5.64	7.77	8.37	9.32
5	5	5	4.56	5.78	7.98	8.72	9.68

A Table de Mann-Whitney

Référence : Table A5.07 : Critical Values for the Wilcoxon/Mann-Whitney Test (U)

n_1	n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2		—	—	—	—	—	—	—	0	0	0	0	1	1	1	1	1	2	2	2	2
3		—	—	—	—	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4		—	—	—	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5		—	—	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6		—	—	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7		—	—	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8		—	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9		—	0	2	4	7	10	12	15	17	21	23	26	28	31	34	37	39	42	45	48
10		—	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11		—	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12		—	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13		—	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14		—	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15		—	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16		—	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17		—	2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18		—	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19		—	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20		—	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

Bilatéral $\alpha = .05$ (Unilatéral $\alpha = .025$)

n_1	n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0	0
3		—	—	—	—	—	—	—	—	0	0	0	1	1	1	2	2	2	2	3	3
4		—	—	—	—	—	0	0	1	1	2	2	3	3	4	5	5	6	6	7	8
5		—	—	—	—	0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
6		—	—	—	0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
7		—	—	—	0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
8		—	—	—	1	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
9		—	—	0	1	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
10		—	—	0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	42
11		—	—	0	2	5	7	10	13	16	18	21	24	27	30	33	36	39	42	45	46
12		—	—	1	3	6	9	12	15	18	21	24	27	31	34	37	41	44	47	51	54
13		—	—	1	3	7	10	13	17	20	24	27	31	34	38	42	45	49	53	56	60
14		—	—	1	4	7	11	15	18	22	26	30	34	38	42	46	50	54	58	63	67
15		—	—	2	5	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	73
16		—	—	2	5	9	13	18	22	27	31	36	41	45	50	55	60	65	70	74	79
17		—	—	2	6	10	15	19	24	29	34	39	44	49	54	60	65	70	75	81	86
18		—	—	2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	92
19		—	0	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	99
20		—	0	3	8	13	18	24	30	36	42	46	54	60	67	73	79	86	92	99	105

Bilatéral $\alpha = .01$ (Unilatéral $\alpha = .005$)

B Table de Wilcoxon

Critical Values of the Wilcoxon Signed Ranks

n	Test bilatéral		Test unilatéral	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
5	-	-	0	-
6	0	-	2	-
7	2	-	3	0
8	3	0	5	1
9	5	1	8	3
10	8	3	10	5
11	10	5	13	7
12	13	7	17	9
13	17	9	21	12
14	21	12	25	15
15	25	15	30	19
16	29	19	35	23
17	34	23	41	27
18	40	27	47	32
19	46	32	53	37
20	52	37	60	43
21	58	42	67	49
22	65	48	75	55
23	73	54	83	62
24	81	61	91	69
25	89	68	100	76
26	98	75	110	84
27	107	83	119	92
28	116	91	130	101
29	126	100	140	110
30	137	109	151	120

Ref : http://facultyweb.berry.edu/vbissonnette/tables/wilcox_t.pdf

Calcul des valeurs sur : <http://comp9.psych.cornell.edu/Darlington/wilcoxon/wilcox0.htm>

C Table du coefficient de rang de Spearman

Valeurs critiques pour un test unilatéral utilisant ρ .

n	5%	1%	n	5%	1%
4	1.000	*	18	.401	.550
5	.900	1.000	19	.391	.535
6	.829	.943	20	.380	.522
7	.714	.893	21	.370	.509
8	.643	.833	22	.361	.497
9	.600	.783	23	.353	.486
10	.564	.745	24	.344	.476
11	.536	.709	25	.337	.466
12	.503	.678	26	.331	.457
13	.484	.648	27	.324	.449
14	.464	.626	28	.318	.441
15	.446	.604	29	.312	.433
16	.429	.582	30	.306	.425
17	.414	.566	40	.264	.368

Les données de la table sont les plus petites valeurs de ρ (jusqu'à 3 décimales) qui correspondent à une probabilité $\leq 5\%$ (ou 1%) sur un seul côté. La valeur observée est significative si elle est supérieure ou égale à la valeur de la table. Le niveau de signification exact ne dépasse jamais la valeur nominale (5% ou 1%). La table peut également être utilisée pour les valeurs critiques à 10% et 2% d'un test bilatéral. L'étoile indique que la signification associée au risque proposé ne peut être calculée dans ce cas.

Valeurs critiques pour un test bilatéral utilisant ρ .

n	5%	1%	n	5%	1%
4	*	*	18	.472	.600
5	1.000	*	19	.460	.584
6	.886	1.000	20	.447	.570
7	.786	.929	21	.436	.556
8	.738	.881	22	.425	.544
9	.700	.883	23	.416	.532
10	.648	.794	24	.407	.521
11	.618	.755	25	.398	.511
12	.587	.727	26	.390	.501
13	.560	.703	27	.383	.492
14	.538	.679	28	.375	.483
15	.521	.654	29	.368	.475
16	.503	.635	30	.362	.467
17	.488	.618	40	.313	.405

Les données de la table sont les plus petites valeurs de ρ (jusqu'à 3 décimales) qui correspondent à une probabilité $\leq 5\%$ (ou 1%) sur les deux côtés. La valeur observée est significative si elle est supérieure ou égale à la valeur de la table. Le niveau de signification exact ne dépasse jamais la valeur nominale (5% ou 1%). La table peut également être utilisée pour les valeurs critiques à 2.5% et 0.5% d'un test unilatéral. L'étoile indique que la signification associée au risque proposé ne peut être calculée dans ce cas.

Ref : <http://www.answers.com/topic/critical-values-for-spearman-s-rank-correlation-coefficient>

