Statistique inférentielle II: Tests statistiques

Nathalie Cheze[1]

December 2, 2005

1 Introduction

Dans de nombreux domaines, on est amené à prendre des décisions au vu de données collectées. La statistique inférentielle (estimation) fournit des éléments permettant de spécifier du mieux possible, à partir de ces observations, le modèle probabiliste qui a engendré les données: détermination du modèle, estimation des paramètres inconnus et validation du modèle. Nous allons présenter sur un exemple concret les différentes questions que l'on peut se poser, qui amènent à des problèmes de tests de nature différente.

Supposons qu'une usine fabrique des pièces sur une machine. Le diamètre de chaque pièce dépend du réglage de la machine, le réglage est d'autant meilleur que le diamètre de la pièce est proche d'une norme d_0; mais comme le réglage ne peut être parfait, on n'a jamais exactement d_0. Le diamètre de chaque piéce peut être modélisé par une variable aléatoire X. De plus, chaque pièce fabriquée a une probabilité θ inconnue, mais la même pour toutes les pièces, d'être défectueuse, c'est à dire de diamètre supérieur à la norme d_0. Ce nombre θ dépend du réglage de la machine, le réglage est d'autant meilleur que θ est proche de 0; mais comme le réglage ne peut être parfait, on n'a jamais $\theta = 0$. On fabrique un certain nombre n de pièces qui servent à tester le réglage. L'observation consiste à mesurer le diamètre X des pièces de l'échantillon de taille n et compter le nombre Y de pièces défectueuses parmi ces n pièces. On peut alors se poser plusieurs types de problèmes.

1) On veut vérifier si la machine est "bien réglée", c'est-à-dire si θ est suffisamment petit. Cela revient à s'assurer que la vraie valeur de θ ne dépasse un seuil critique θ_0 fixé à l'avance (sinon, il faut refaire le réglage de la machine): cela s'appelle **tester** le fait que $\theta \leq \theta_0$. Ce type de tests **paramétriques** sera étudié au troisième chapitre.

2) On veut préciser la loi de probabilité inconnue de X. Cela revient à résoudre un problème d'ajustement de la loi empirique observée sur l'échantillon à une loi théorique: il s'agit d'un test **d'ajustement** ou **d'adéquation** étudié dans le quatrième chapitre.

3) Il peut arriver que la production se fasse sur deux chaînes différentes. Le diamètre des pièces issues de la seconde chaîne est modélisé par une variable aléatoire Z.

• On peut se demander si Z suit la même loi de probabilité que X. Cela revient à comparer deux échantillons dont les lois de probabilité des populations-mères sont inconnues, ceci rentre dans le cadre des tests **non paramétriques** traités dans le quatrième chapitre.

[1]Université Paris 10, MODALX, 200 av. de la République, 92000 Nanterre, Nathalie.Cheze@u-paris10.fr

• Supposons que les lois de probabilité de X et Z soient connues, de moyenne m_X et m_Z inconnues. On se demande si les deux chaînes sont réglées de la même façon, c'est à dire si $m_X = m_Y$. C'est un exemple de tests de comparaisons de paramètres, étudiés dans le paragraphe 3.3.

Le chapitre 2 présente le formalisme commun de ces différents tests avant d'étudier, dans les chapitres suivants, les spécificités de chacun d'eux.

2 Théorie des tests

2.1 Modèle d'échantillonnage

se reporter au cours sur l'estimation.

2.2 Tests paramétriques

On suppose que la probabilité P_θ est connue, au paramètre θ inconnu près. Nous supposons que Θ est divisé en une partie Θ_0 et son complémentaire Θ_1. L'objectif d'un test paramétrique est, au vu de l'observation de l'échantillon $(x_1, ..., x_n)$, de "décider" si la vraie valeur de θ se trouve dans Θ_0 ou dans Θ_1.

On note $\mathbf{H_0}$, l'hypothèse **nulle** selon laquelle $\theta \in \Theta_0$ et H_1 l'hypothèse **alternative** selon laquelle $\theta \in \Theta_1$.

2.3 Tests non paramétriques

Dans la pratique, on ne peut pas toujours se placer dans le cadre précédent. La loi de probabilité de X notée P_X dans ce paragraphe, est supposée inconnue. L'objectif des tests non paramétriques est de décider, au vu des observations de l'échantillon $(x_1, ..., x_n)$, si X suit une loi de probabilité P_0 donné.

On note $\mathbf{H_0}$, l'hypothèse **nulle** selon laquelle $P_X = P_0$ et H_1 l'hypothèse **alternative** selon laquelle $P_X \neq P_0$.

Nous allons maintenant donner le vocabulaire indispensable à la théorie des tests.

2.4 Erreurs et risques

Dans un problème de décision, deux types d'erreurs sont possibles:
• l'**erreur de première espèce** est l'erreur commise lorsqu'on décide de rejeter H_0 alors que celle-ci est vraie.
• l'**erreur de deuxième espèce** est l'erreur commise lorsqu'on décide de ne pas rejeter H_0 alors que celle-ci est fausse.

Les conséquences de ces erreurs sont d'importances diverses. Revenons à l'exemple introductif concernant les pièces défectueuses, pour bien comprendre cette remarque.

Soit θ_0 la valeur limite de la proportion des pièces défectueuses acceptable. On veut décider, au vu du nombre Y de pièces défectueuses observées dans un échantillon de taille n, si l'on accepte ou non l'hypothèse $H_0 : \ \theta > \theta_0$.
Rejeter à tort l'hypothèse $\theta > \theta_0$ revient à ne pas régler de nouveau la machine alors qu'elle devrait l'être. La conséquence est le mécontentement des clients, voire des problèmes de sécurité si des pièces défectueuses sont utilisées. Ne pas rejeter à tort que $\theta > \theta_0$ revient à effectuer un réglage supplémentaire -inutile- de la machine. La conséquence de cette décision est moins grave que la précédente.

A toute décision correspond une probabilité de prendre la bonne décision et une probabilité de se tromper:

• le **risque de première espèce** est la probabilité de l'erreur de première espèce, noté α, c'est-à-dire la probabilité de rejeter à tort H_0.
• le **risque de deuxième espèce**, noté β, est la probabilité de l'erreur de deuxième espèce, c'est-à-dire la probabilité de ne pas rejeter à tort H_0.
• la **puissance** du test, notée $1 - \beta$, est la probabilité de rejeter avec raison H_0.

Dans l'exemple introductif, le risque de première espèce correspond à la probabilité de ne pas régler de nouveau la machine alors qu'elle devrait l'être et le risque de deuxième espèce correspond à la probabilité de régler de nouveau la machine alors qu'elle était déjà bien réglée.

On peut résumer ces différents risques sous forme de tableau:

Vérité Décision	H_0 est vraie	H_1 est vraie
ne pas rejeter H_0	bonne décision $1 - \alpha$	mauvaise décision β
rejeter H_0	mauvaise décision α	bonne décision $1 - \beta$

L'idéal est d'avoir les risques de première et de deuxième espèce les plus proches possibles de zéro. Malheureusement, on peut montrer que ces deux risques varient en sens inverse: quand l'un diminue, l'autre augmente. Dans la pratique, on construit les tests d'hypothèses de telle sorte que l'erreur de première espèce soit celle qui a les conséquences plus graves. Pour cela, on choisira pour hypothèse nulle H_0, l'hypothèse pour laquelle le rejet à tort est le plus préjudiciable.

Dans notre exemple introductif, l'erreur la plus grave est de ne pas régler de nouveau la machine alors qu'elle devrait l'être, c'est-à-dire rejeter l'hypothèse $\theta > \theta_0$ à tort. Cela revient donc à poser $H_0 : \theta > \theta_0$ et $H_1 : \theta \leq \theta_0$.

Cette façon de procéder revient à fixer le niveau du test α à une valeur d'autant plus petite que l'erreur de première espèce peut avoir des conséquences graves. Les valeurs usuelles de α sont $10\%, 5\%, 1\%$. Cette optique rend le problème dissymétrique. On priviligie l'hypothèse H_0. Par exemple, on ne prendra pas nécessairement la même décision, au vu des obervations, suivant si l'on teste $H_0 : \theta > \theta_0$ contre $H_1 : \theta \leq \theta_0$ ou $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$. On verra un exemple dans le paragraphe 3.1.

Remarque: il vaut mieux dire "ne pas rejeter H_0" que "accepter H_0". En effet, si on rejette H_0 c'est que les observations sont telles qu'il est improbable que H_0 soit vraie. Par contre, si on ne rejette pas H_0, c'est que les observations ne permettent pas de dire que H_0 est fausse, mais cela ne veut pas dire pour autant que H_0 est vraie.

2.5 Construction d'un test d'hypothèses

On appelle **règle de décision**, une règle qui permet de décider H_0 ou H_1 au vu des observations, sous la contrainte que le risque de première espèce du test α soit fixé.

Pour établir cette régle, l'idée est de construire une **région critique ou de rejet** du test, notée $W \subset \Omega$. . Elle sera déterminée grâce à la modélisation probabiliste, développée dans la suite.

Récapitulons la démarche à suivre pour effectuer un test d'hypothèses:

a- choisir les hypothèses H_0 et H_1 en priviligiant H_0.
b- fixer le risque de première espèce du test α selon la gravité des conséquences de l'erreur de première espèce.
c- construire la région critique.
d- regarder si les observations se trouvent ou non dans W.
e- conclure au rejet ou non-rejet de H_0.

Pour un α fixé, on peut construire différentes région critiques, donc différents tests. Le meilleur des tests à α fixé, est celui qui minimise le risque de seconde espèce β (de façon équivalente, celui qui maximise la puissance $1 - \beta$). Il s'agit d'un problème mathématique assez complexe. Pour le résoudre, on introduit le critère de qualité suivant:

Définition 3.
Soit un test T de région de rejet W, de niveau α.
a) Un test T' de région de rejet W' est dit **plus puissant** que T si il est de niveau $\alpha' \leq \alpha$ et si $P_\theta(W) < P_\theta(W')$, $\forall \theta \in \theta_1$.

b) Le test T est **uniformément plus puissant** (en abrégé: UPP) s'il n'existe pas de test T' plus puissant que T.

Dans certains cas, on peut déterminer un test le plus puissant, ce qui n'est pas vrai en général. On verra quelques exemples dans la suite.

3 Tests paramétriques

Considérons la modélisation statistique $(\Omega, \mathcal{A}, (P_\theta)_{\theta \in \Theta})$. On suppose que la loi P_θ de X est connue, de paramètre $\theta \in \Theta \in I\!R$ inconnu.

Nous supposons que Θ est divisé en une partie Θ_0 et son complémentaire Θ_1. Les tests paramétriques permettent de tester l'hypothèse $H_0 : \theta \in \Theta_0$ contre l'hypothèse $H_1 : \theta \in \Theta_1$. Sous ces hypothèses, le risque de première espèce (resp. de deuxième espèce) peut s'écrire $\alpha(\theta) = P_\theta(W), \theta \in \Theta_0$ (resp. $\beta(\theta) = 1 - P_\theta(W), \theta \in \Theta_1$). On appelle **niveau** du test la constante $\sup_{\theta \in \Theta_0} P_\theta(W)$.

Les ensembles Θ_0 et Θ_1 peuvent prendre plusieurs formes, ce qui nous amène à étudier plusieurs situations:

- Cas (1): test d'une hypothèse simple contre une hypothèse simple: $\Theta = \{\theta_0, \theta_1\}$.

$$\begin{array}{c} H_0 : \theta = \theta_0 \\ \text{contre} \\ H_1 : \theta = \theta_1 \end{array}$$

$(\theta_1 \neq \theta_0)$
On peut remarquer que, dans ce cas, le niveau du test est égal à α et le risque de deuxième espèce vaut $\beta = 1 - P_{\theta_1}(W)$.

- Cas (2): test d'une hypothèse simple contre une hypothèse multiple: $\Theta \subset I\!R$.

$$\begin{array}{c} H_0 : \theta = \theta_0 \\ \text{contre} \\ H_1 : \theta \neq \theta_0 \end{array}$$

- Cas (3): test d'une hypothèse multiple contre une hypothèse multiple: $\Theta \subset I\!R$.

$$\begin{array}{c} H_0 : \theta_1 \leq \theta \leq \theta_2 \\ \text{contre} \\ H_1 : \theta < \theta_1 \text{ ou } \theta > \theta_2 \end{array}$$

- Cas (4): test d'une hypothèse multiple contre une hypothèse multiple: $\Theta \subset I\!R$.

$$\begin{array}{c} H_0 : \theta \leq \theta_0 \\ \text{contre} \\ H_1 : \theta > \theta_0 \end{array}$$

- Cas (5): test d'une hypothèse multiple contre une hypothèse multiple: $\Theta \subset I\!R$.

$$H_0 : \theta \geq \theta_0$$
$$\text{contre}$$
$$H_1 : \theta < \theta_0$$

• Cas (6): test d'une hypothèse multiple contre une hypothèse multiple: $\Theta \subset I\!R$.

$$H_0 : \theta \leq \theta_1 \text{ ou } \theta \geq \theta_2$$
$$\text{contre}$$
$$H_1 : \theta_1 < \theta < \theta_2$$

Remarquons que le cas (2) est un cas particulier du cas (3) en prenant $\theta_1 = \theta_2 = \theta_0$. Le cas (1) correspond au cas (rare) où Θ peut être réduit à deux valeurs θ_0 et θ_1. Il est cependant d'un grand intérêt pédagogique et permet de présenter la méthode, qui sera ensuite généralisée dans les autres cas. Dans ce cours, on étudiera seulement les cas 1) et 2).

3.1 Test d'une hypothèse simple contre une hypothèse simple

.

Soit deux nombres réelles $\theta_0 \neq \theta_1$. On cherche à tester, au niveau α

$$H_0 : \theta = \theta_0$$
$$\text{contre}$$
$$H_1 : \theta = \theta_1$$

au vu d'observations $x = (x_1, ..., x_n)$ d'un n-échantillon de X.
On note $L(x, \theta)$ la fonction de vraisemblance de x en θ dont on rappelle la définition ci-dessous

Définition .
On se place dans le cadre d'hypothèses suivant :
on appelle vraisemblance du paramètre θ, l'application $L : \Theta \to I\!R^+$ définie par:

- cas (a): si E est fini ou dénombrable, $L(x_1, x_2, ..., x_n, \theta) = \prod_{i=1}^n P_\theta(X = x_i)$

- cas (b): si $E = I\!R$ ou $I\!R^d$ et P_θ admet une densité f_θ, $L(x_1, x_2, ..., x_n, \theta) = \prod_{i=1}^n f_\theta(x_i, \theta)$.

Le théorème suivant de Neyman-Pearson permet de construire un test UPP dans ce cadre.

Théorème de Neyman-Pearson.
Pour tout $\alpha \in]0, 1[$ donné, il existe un test UPP de niveau α, défini par la région de rejet$W$:

$$W = (x \in \Omega : \frac{L(x, \theta_0)}{L(x, \theta_1)} \leq k)$$

Remarques.

• La signification intuitive de ce théorème est claire: il s'agit de refuser H_0 lorsque la vraisemblance de l'échantillon en θ_0 est plus "petite" que celle en θ_1.

• Le risque de première espèce est égal, dans les tests d'hypothèses simples, au niveau du test $\alpha = P_{\theta_0}(W) = P_{\theta_0}(\frac{L(X,\theta_0)}{L(X,\theta_1)} \leq k)$. Ce qui nous permet d'obtenir la constante k, comme quantile de la v.a. $\frac{L(X,\theta_0)}{L(X,\theta_1)}$. Si cette v.a. a une densité, on pourra trouver k pour toute valeur de α.

• Le risque de deuxième espèce vaut, dans ce cas, $\beta = P_{\theta_1}(\bar{W})$, où $\bar{W}$ représente le complémentaire de W, c'est-à-dire la région de non-rejet de H_0 .

Explicitons de manière détaillée sur un exemple simple, cette méthode qui permet d'obtenir un test UPP et donnons ensuite des exemples, sans calcul, de tests pour des lois usuelles.

Exemples.

• **Test sur l'espérance d'une loi normale $N(m, \sigma^2)$, σ connu**

Considérons un $n-$échantillon $(X_1, ..., X_n)$ de loi normale $N(m, \sigma^2)$, σ connu. On veut tester, au niveau α, les hypothèses suivantes:

$$\begin{array}{c} H_0 : m = m_0 \\ \text{contre} \\ H_1 : m = m_1 \end{array}$$

$(m_0 \neq m_1)$

Pour déterminer la **région critique** W, on utilise le théorème de Neyman-Pearson et pour cela, on doit écrire le rapport des vraisemblances.

$$\frac{L(x,\theta_0)}{L(x,\theta_1)} = exp(-\frac{1}{2\sigma^2}(\sum_{i=1}^{n}((x_i - m_0)^2 - (x_i - m_1)^2))$$

Le logarithme étant monotone, étudier $\frac{L(x,\theta_0)}{L(x,\theta_1)} \leq k$ revient à étudier l'inégalité suivante:

$$\sum_{i=1}^{n}((x_i - m_1)^2 - (x_i - m_0)^2) \leq 2\sigma^2 \ Logk$$

En simplifiant les termes de gauche dans l'inégalité, on obtient finalement:

$$(m_0 - m_1)\sum_{i=1}^{n} x_i \leq k'$$

avec $k' = \sigma^2 \; Logk \; + \; \frac{n}{2}(m_0^2 - m_1^2)$.
Deux cas se présentent:

i) $m_0 > m_1$: on déduit de l'inégalité précédente la forme de la région de rejet:

$$W \; = \; (x \in \Omega : \; \bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \leq \; K)$$

(où $K \; = \; \frac{k'}{n(m_0 - m_1)}$, donné pour indication, car seule la forme de la région de rejet nous intéresse).

ii) $m_0 < m_1$: dans ce cas, on obtient

$$W \; = \; (x \in \Omega : \; \bar{x}_n \geq \; K)$$

c'est-à-dire qu'on rejette l'hypothèse H_0 au risque α si l'échantillon observé $\bar{x}_n \geq K$.

Il reste donc à déterminer la valeur de la constante K. Expliquons, toujours sur cet exemple, la méthode.

Détermination de K
Limitons nous au cas i): $m_0 > m_1$. Puisque le risque de première espèce est égal à α, on a

$$\alpha \; = \; P_{m_0}(\bar{X}_n \leq \; K)$$

La variable aléatoire $\sqrt{n}\frac{\bar{X}_n - m_0}{\sigma}$ suit une loi $N(0,1)$, sous P_{m_0}. On peut donc écrire

$$\alpha \; = \; P_{m_0}(\sqrt{n}\frac{\bar{X}_n - m_0}{\sigma} \leq \; \sqrt{n}\frac{K - m_0}{\sigma})$$

D'où, $K \; = \; m_0 - u_{1-\alpha}\frac{\sigma}{\sqrt{n}}$ avec $u_{1-\alpha}$ quantile d'ordre $1-\alpha$ de la loi $N(0,1)$, lu dans une table statistiqu.

La région de rejet du test est donc:

$$W \; = \; (x \in \Omega : \; \bar{x}_n \leq \; m_0 \; - \; u_{1-\alpha}\frac{\sigma}{\sqrt{n}})$$

Il ne nous reste plus qu'à étudier la puissance du test $1 - \beta = P_{m_1}(W)$

$$1 - \beta \; = \; P_{m_1}(\bar{X}_n \leq \; K)$$

ce qui peut s'écrire, si on centre et on réduit la variable $\bar{X}_n$ sous l'hypothèse H_1:

$$1 - \beta \; = \; P_{m_1}(\sqrt{n}\frac{\bar{X}_n - m_1}{\sigma} \leq \; \sqrt{n}\frac{K - m_1}{\sigma})$$

Le calcul de la puissance se réduit donc au calcul de la fonction de répartition de la loi $N(0,1)$ en $\sqrt{n}\frac{K-m_1}{\sigma}$, qui peut, par exemple, se lire facilement sur une table statistique .

Donnons maintenant un exemple numérique pour illustrer l'importance du choix des hypothèses nulle et alternative. Le fait d'inverser H_0 et H_1 peut amener à changer de décison.

Exemple numérique
La législation en vigueur impose aux aéroports certaines normes concernant les bruits émis par les avions au décollage et à l 'atterrissage. Ainsi, pour les zones habitées proches d'un aéroport, la limite tolérée se situe à 80 décibels. A partir de cette limite, l'aéroport doit indemniser les riverains. Les habitants d'un village voisin assurent que le bruit au dessus d'une certaine partie de leur village atteint la valeur limite de 80 décibels. L'aéroport affirme qu'il n'est que de 78 décibels. Afin de trancher entre les deux parties, on mesure l'intensité du bruit provoqué par le passage de 100 avions. Le bruit moyen émis par ces 100 avions est de 79.1. On admet que l'intensité du bruit causé par un avion suit une loi normale de moyenne m décibels et de variance 49 décibels^2. On admet aussi que les intensités du bruit causé par différents avions sont indépendants entre eux.
1- On choisit de tester

$$\begin{aligned} H_0 : m &= 80 \\ &\text{contre} \\ H_1 : m &= 78 \end{aligned}$$

au niveau 5%.

Le risque de première espèce représente la probabilité de décider que l'aéroport n'indemnise pas les riverains alors qu'ils devraient l'être.
Le risque de deuxième espèce représente la probabilité de décider que l'aéroport indemnise les riverains alors qu'ils ne devraient pas l'être.
On voit bien que le test, formulé de cette façon privilégie les riverains.

D'après ce qui précéde, La région de rejet du test est:

$$W = (x \in \Omega : \bar{x}_n \leq 80 - 1,645\sqrt{\frac{49}{100}} = 78,83)$$

Comme $\bar{x}_n = 79,1 > 78,83$, on ne rejette pas H_0 et on indemnise les riverains.

La puissance du test vaut

$$1-\beta = P_{m_1}(10\frac{\bar{X}_n - 78}{\sigma} \leq 10\frac{78,83-78}{7}) = 0,8869$$

2- On choisit maintenant de tester

$$\begin{array}{rcl} H_0 : m & = & 78 \\ & \text{contre} & \\ H_1 : m & = & 80 \end{array}$$

au niveau 5%.

Le risque de première espèce représente la probabilité de décider que l'aéroport indemnise les riverains alors qu'ils ne devraient pas l'être.
Le risque de deuxième espèce représente la probabilité de décider que l'aéroport n'indemnise pas les riverains alors qu'ils devraient l'être.
On voit bien que le test, formulé de cette façon privilégie l'aéroport.

D'après ce qui précède, la région de rejet du test est:

$$W = (x \in \Omega : \bar{x}_n \geq 78 + 1,645\frac{7}{10} = 79,17)$$

Comme $\bar{x}_n = 79,1 < 79,17$, on ne rejette pas H_0 et on n'indemnise pas les riverains.

La puissance du test vaut

$$1 - \beta = P_{m_1}(10\frac{\bar{X}_n - 80}{\sigma} \leq 10\frac{79,17 - 80}{7}) = 0,8869$$

3- Conclusion: au niveau 5%, les deux tests précédents de même puissance nous donne des conclusions contradictoires. Si on regarde les deux régions de rejet, on voit qu'on prend les mêmes décisons si $\bar{x}_n \leq 78,81$ ou $\bar{x}_n \geq 79,17$. Par contre, on prend des décisions contradictoires si $78,83 < \bar{x}_n < 79,17$. Ceci est du à la méthode de Neyman-Pearson, qui permet de construire des tests UPP mais qui privilégie H_0 en fixant le niveau α.

Dans ce cas, un moyen de départager les deux parties est de construire un troisième test pour lequel les risques de première et de deuxième espèce sont égaux. Comme cela, on ne priviligie pas d'hypothèses. Pour cela, il suffit de prendre comme région de rejet, pour le cas 1), $W = (x \in \Omega, \bar{x}_n \leq \frac{78+80}{2})$, respectivement pour le cas 2), $W = (x \in \Omega, \bar{x}_n \geq \frac{78+80}{2})$. Dans les deux cas, on prend la décision $m = 80$ et on indemnise les riverains. On peut calculer le niveau du test $\alpha = 0,0764 = \beta$. Ce test n'est pas comparable aux deux premiers car pour pouvoir comparer des tests entre eux, il faut qu'ils soient de même niveau.

Nous donnons ci-après des exemples usuels, mais sans détailler les calculs, la méthode étant identique à celle qui vient d'être exposée.

• **Test sur l'espérance d'une loi normale $N(m, \sigma^2)$, σ inconnu**
Considérons un n-échantillon $(X_1, ..., X_n)$ de loi normale $N(m, \sigma^2)$, σ inconnu. On veut tester les hypothèses suivantes, pour un niveau α donné:

$$H_0 : m = m_0$$
$$\text{contre}$$
$$H_1 : m = m_1$$

On se limite au cas $m_0 > m_1$, le cas $m_0 < m_1$ s'en déduisant en changeant le sens des inégalités.

On peut reprendre les étapes de construction du test de l'exemple précédent, en utilisant la v.a $\sqrt{n}\frac{\bar{X}_n - m}{\hat{S}_n}$ plutôt que la v.a. $\sqrt{n}\frac{\bar{X}_n - m}{\sigma}$, car σ est inconnu. La v.a $\sqrt{n}\frac{\bar{X}_n - m}{\hat{S}_n}$ suit une loi de Student $St(n-1)$.

On obtient donc la région critique:

$$W = (x \in \Omega : \ \bar{x}_n \leq \ m_0 - t_{1-\alpha}\frac{\hat{s}_n}{\sqrt{n}})$$

où $t_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi $St(n-1)$, lu dans une table statistique.

• **Test sur la variance d'une loi normale $N(m, \sigma^2)$, m connu**
Considérons un $n-$échantillon $(X_1, ..., X_n)$ de loi normale $N(m, \sigma^2)$, m connu. On veut tester les hypothèses suivantes, pour un niveau α donné:

$$H_0 : \sigma^2 = \sigma_0^2$$
$$\text{contre}$$
$$H_1 : \sigma^2 = \sigma_1^2$$

On se limite au cas $\sigma_0^2 > \sigma_1^2$, le cas $\sigma_0^2 < \sigma_1^2$ s'en déduisant en changeant le sens des inégalités.

Le calcul du rapport des vraisemblances donne une région de rejet de la forme :

$$W = (x \in \Omega : \ \sum_{i=1}^{n}(x_i - m)^2 \leq \ K)$$

La v. a. $\frac{n}{\sigma^2}\hat{\sigma}_n^2$ suit une loi du $\chi^2(n)$, avec $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$.

On peut donc écrire

$$\alpha = P_{\sigma_0^2}(\frac{n\hat{\sigma}_n^2}{\sigma_0^2} \leq \ \frac{K}{\sigma_0^2})$$

D'où, $K = c\,\sigma_0^2$ où c est le quantile d'ordre α de la loi $\chi^2(n)$, qui peut être lu, par exemple, dans une table statistique associée.
On obtient finalement comme région de rejet

$$W = (x \in \Omega : \ \frac{n\hat{\sigma}_n^2}{\sigma_0^2} \leq c)$$

Dans le cas où m est inconnue, on utilise la v.a. $\frac{n-1}{\sigma^2}\hat{S}_n^2$ qui suit une loi du $\chi^2(n-1)$ au lieu de la v.a. $\frac{n}{\sigma^2}\hat{\sigma}_n^2$. On obtient la même région de rejet, mais dans ce cas, c représente le quantile d'ordre α de la loi $\chi^2(n-1)$.

• **Test sur la proportion d'une loi de Bernoulli** $Be(p)$**.**
Considérons un $n-$échantillon $(X_1, ..., X_n)$ de loi de Bernoulli $Be(p)$. On veut tester les hypothèses suivantes, au niveau α fixé, pour $p_0 \neq p_1 \in]0,1[$:

$$\begin{array}{rcl} H_0: p & = & p_0 \\ & \text{contre} & \\ H_1: p & = & p_1 \end{array}$$

On se limite au cas $p_0 > p_1$, le cas $p_0 < p_1$ s'en déduisant en changeant le sens des inégalités.

La forme de la région de rejet s'obtient facilement par le calcul du rapport des vraisemblances:

$$W = (x \in \Omega : \sum_{i=1}^{n} x_i \leq K)$$

Puisque le risque de première espèce est égal à α, on a

$$\alpha = P_{p_0}(\sum_{i=1}^{n} X_i \leq K)$$

La fonction de répartition de $\sum_{i=1}^{n} X_i$ n'étant pas bijective , on ne trouvera pas toujours une région de rejet de niveau exactement égale à α, car il n'existe pas forcément une constante K telle que $\alpha = P_{p_0}(\sum_{i=1}^{n} X_i \leq K)$. Deux possibilités s'ouvrent à nous: soit on cherche un test de niveau non pas égal à α mais de niveau $\leq \alpha$, soit on cherche un test mixte qu'on définira ci-après.

Cas(i): test de niveau inférieur ou égal à α

Ecrivons que le niveau de test est inférieur à α

$$P_{p_0}(\sum_{i=1}^{n} X_i \leq K) \leq \alpha$$

La variable aléatoire $\sum_{i=1}^{n} X_i$ suit une loi binomiale $B(n, p)$. Ainsi, sa fonction de répartition est constante par morceaux. La constante K va donc être l'unique nombre tel que

$$P(Y \leq K) = \alpha_1$$
$$P(Y \leq K+1) = \alpha_2, \text{ avec } \alpha_1 \leq \alpha < \alpha_2$$

où la probabilité $P(Y \leq y)$ peut être, par exemple, lue dans une table statistique de la loi $B(n, p_0)$.

Cas(ii): test de niveau de niveau égal à α: test mixte

Dans le cas précédent, on a obtenu $P_{p_0}(\sum_{i=1}^{n} X_i \leq K) \leq \alpha$. Pour obtenir un test de niveau égal à α, on peut choisir une règle de décision aléatoire de la façon suivante:
soit une constante $\gamma \in]0,1[$ telle que

$$\alpha = P_{p_0}(\sum_{i=1}^{n} X_i \leq K) + \gamma \, P_{p_0}(\sum_{i=1}^{n} X_i = K+1)$$

On procède de la façon suivante
- on rejette H_0 si $\sum_{i=1}^{n} x_i \leq K$
- on accepte H_0 si $\sum_{i=1}^{n} x_i > K+1$.

Dans le cas où l'observation tombe sur le point frontière $(\sum_{i=1}^{n} x_i = K+1)$, on réalise un tirage au sort. Avec une probabilité γ, on rejette H_0 et avec une probabilité $1-\gamma$, on ne rejette pas H_0.

Si la taille de l'échantillon est suffisamment grande (en pratique, $n \geq 30$), on peut construire une région critique asymptotique. Par le théorème de la limite centrale, la v.a. $\frac{\bar{X}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ suit approximativement une loi $N(0,1)$, sous H_0. On obtient donc la région de rejet:

$$W = (x \in \Omega : \bar{x}_n \leq p_0 - u_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{n}} \,)$$

où $u_{1-\alpha}$ est le quantile d'ordre α de la loi $N(0,1)$, lu dans une table statistique.

3.2 Hypothèse simple contre hypothèse multiple

• Considérons tout d'abord le test *unilatéral.* On veut tester, au niveau α :

$$H_0 : \theta = \theta_0$$
contre
$$H_1 : \theta > \theta_0 \text{ (ou bien } \theta < \theta_0)$$

Pour chaque $\theta_1 > \theta_0$, on doit donc tester une hypothèse simple contre une hypothèse simple:

$$\begin{aligned} &H_0: \ \theta \ = \ \theta_0 \\ &\text{contre} \\ &H_1: \ \theta \ = \ \theta_1, \ \text{avec } \theta_1 > \theta_0, \ (\text{ou bien } \theta \ = \ \theta_1, \ \text{avec } \theta_1 < \theta_0) \end{aligned}$$

En reprenant les étapes de la méthode précédente pour le test sur l'espérance d'une loi normale $N(m, \sigma^2)$, σ connu, on voit que la forme de la région critique dépend seulement du sens de l'inégalité $m_1 > m_0$ ou $m_1 < m_0$. La région de rejet ne dépend donc pas de m_1, on obtient donc la même région de rejet que dans le cas (1) et le test obtenu est donc UPP dans un sens que l'on va préciser. Pour cela, il faut introduire une nouvelle notion.

Définition 4 On appelle **courbe d'efficacité** d'un test sur le paramètre θ, la courbe représentative de la fonction $\beta : \Theta \to [0, 1]$, définie par:

$$\beta(\theta) = 1 - P_\theta(W)$$

$\beta(\theta)$ est la probabilité de ne pas rejeter H_0, quel que soit θ.

Remarque: Par définition, $\beta(\theta_0) = 1 - \alpha$ et pour $\theta \in \Theta_1$, $\beta(\theta)$ représente le risque de deuxième espèce.

On dit que le test précédent est UPP c'est-à-dire que sa courbe d'efficacité est, en tout point θ, en dessous de la courbe d'efficacité de n'importe quel autre test.

• Considérons maintenant le cas (2) qui est un test bilatéral:

$$\begin{aligned} &H_0: \ \theta \ = \ \theta_0 \\ &\text{contre} \\ &H_1: \ \theta \ \neq \ \theta_0 \end{aligned}$$

au niveau α.

Cela revient à tester une hypothèse simple contre une hypothèse multiple:

$$\begin{aligned} &H_0: \ \theta \ = \ \theta_0 \\ &\text{contre} \\ &H_1: \ \theta \ = \ \theta_1, \ \text{avec } \theta_1 \neq \theta_0 \end{aligned}$$

Reprenons l'exemple du test sur l'espérance d'une loi normale $N(m, \sigma^2)$, σ connu pour donner une idée de l'approche générale de ce test.
Considérons un $n-$échantillon $(X_1, ..., X_n)$ de loi normale $N(m, \sigma^2)$, σ connu. On veut tester les hypothèses suivantes au niveau α:

$$
\begin{aligned}
H_0 : \; m \; &= \; m_0 \\
&\text{contre} \\
H_1 : \; m \; &\neq \; m_0
\end{aligned}
$$

On décompose ce test en deux tests de niveau $\alpha/2$:

$$
\begin{aligned}
H_0 : \; m \; &= \; m_0 \\
&\text{contre} \\
H_1 : \; m \; &= \; m_1, \; m_1 < m_0
\end{aligned}
$$

et

$$
\begin{aligned}
H_0 : \; m \; &= \; m_0 \\
&\text{contre} \\
H_1 : \; m \; &= \; m_1, \; m_1 > m_0
\end{aligned}
$$

La méthode de Neyman-Pearson nous donne comme région de rejet de niveau $\alpha/2$, respectivement pour chaque test, $W_1 = (x \in \Omega : \bar{x}_n \leq m_0 - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$ et $W_2 = (x \in \Omega : \bar{x}_n \geq m_0 + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})$ où $u_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi $N(0,1)$, lu dans une table statistique.

On peut montrer que ce test n'est pas UPP. Il n'existe pas de test UPP pour ce type d'hypothèses car le choix de couper le test de niveau α, en deux tests de niveau $\alpha/2$ est arbitaire. Il n'y a pas de choix optimal pour le découpage du risque α en $\alpha_1 + \alpha_2$ qui permet d'obtenir un test uniformément le plus puissant parmi tous les tests.

On va construire sur un exemple numérique, deux tests de même niveau et de même puissance de régions de rejet différentes.

Exemple numérique
Une machine produit des pièces circulaires. Du fait des imperfections du processus de fabrication, le diamètre d'une pièce donnée correspond à la réalisation d'une variable aléatoire distribuée selon une loi normale de moyenne m et d'écart-type $0,02mm$. Pour qu'une pièce soit utilisable, son diamètre doit être égal à $23,65mm$. On admet que le diamètre des pièces produites sont indépendantes les unes des autres. Afin de vérifier si la machine est bien réglée, c'est-à-dire si $m = 23,65mm$, on mesure le diamètre de 10 pièces et on obtient une moyenne de $23.661mm$. On veut effectuer le test $H_0 : m = 23,65$ contre $H_1 : m \neq 23,65$ pour une erreur de première espèce $\alpha = 0,1$. On calculera la puissance de ce test pour $m = 23,63$ et $m = 23,67$.

1- Par la méthode de Neyman-Pearson, on obtient la région de rejet, au niveau 10%, $W = (x \in \Omega : \bar{x}_n \leq 23,65 - 1,645\frac{0,02}{\sqrt{10}}$ ou $\bar{x}_n \geq 23,65 + 1,645\frac{0,02}{\sqrt{10}})$, c'est-à-dire $W = (x \in \Omega : \bar{x}_n \leq 23,64$ ou $\bar{x}_n \geq 23,66)$. Comme $\bar{x}_n = 23,661 \in W$, on rejette H_0. On peut calculer la puissance de ce test en $m = 23,63$ et $m = 23,67$ qui vaut

$\eta(23,63) = \eta(23,67) = 0,9429$.

2-Pour les mêmes hypothèses, on considère la règion de rejet suivante:$W = (x \in \Omega : \bar{x}_n \leq 23,6409$ ou $\bar{x}_n \geq 23,6624)$. On peut calculer α qui vaut 10%. Comme $\bar{x}_n = 23,661 \in \bar{W}$, on ne rejette pas H_0. On peut comparer les deux tests car ils sont de même niveau. Pour cela, calculons la puissance de ce test en $m = 23,63$ et $m = 23,67$. Elles valent respectivement $\eta'(23,63) = 0,9573$ et $\eta(23,67) = 0,8849$.

Finalement, au niveau 10%, on a

$$\eta'(23,67) < \eta(23,63) = \eta(23,67) < \eta'(23,63)$$

Aucun des deux tests n'est plus puissant que l'autre.

3.3 Tests de comparaison de paramètres

Comme on l'a vu dans l'exemple introductif concernant la fabrication de pièces en acier, il est fréquent d'avoir à comparer entre elles des caractéristiques de deux variables aléatoires X et Y. Pour cela, on considère un échantillon de X de taille n et un échantillon de Y de taille m. Pour donner la démarche générale des tests de comparaison, nous allons, dans un premier temps, nous limiter à l'étude d'échantillons gaussiens (de loi normale) indépendants entre eux. On désire comparer l'espérance et la variance de X et Y. Dans un deuxième temps, nous généraliserons ces résultats au cas de deux échantillons indépendants entre eux et de grande taille, de v.a. de loi de probabilité quelconque. Si les tailles des deux échantillons ne sont pas suffisament grandes, ces techniques ne sont plus adaptées. Une alternative, dans ce cas, sera étudiée dans le chapitre concernant les tests non paramétriques. Pour finir, nous étudierons le cas apparié, c'est-à-dire quand les deux échantillons de taille n, $(X_1, ..., X_n)$ et $(Y_1, ..., Y_n)$, ne sont pas indépendants entre eux. On se limitera au cas gaussien.

3.3.1 Comparaison de deux échantillons gaussiens indépendants

Considérons une v.a. X de loi $N(m_X, \sigma_X^2)$ et une v.a. Y de loi $N(m_Y, \sigma_Y^2)$. Notons $\bar{X}_n$ la moyenne empirique et $\hat{S}_X^2$ la variance empirique sans biais de l'échantillon $(X_1, ..., X_n)$de X et $\bar{Y}_m$ la moyenne empirique et $\hat{S}_Y^2$ la variance empirique sans biais de l'échantillon $(Y_1, ..., Y_m)$ de Y. On suppose que les deux échantillons sont indépendants entre eux.

3.3.2 Test de l'égalité des variances.

On se place dans le cas où m_X et m_Y sont inconnues. On veut tester les hypothèses suivantes, au niveau α:

$$\begin{aligned} H_0 : \sigma_X^2 &= \sigma_Y^2 \\ \text{contre} & \\ H_1 : \sigma_X^2 &\neq \sigma_Y^2 \end{aligned}$$

La v.a. $\frac{\hat{S}_X^2}{\hat{S}_Y^2}\frac{\sigma_Y^2}{\sigma_X^2}$ suit une loi de Fisher $F(n-1,m-1)$.

Le problème de test au niveau α peut se réécrire, en considérant la v.a. $F=\frac{\hat{S}_X^2}{\hat{S}_Y^2}$ et $\theta=\frac{\sigma_X^2}{\sigma_Y^2}$ de la manière suivante:

$$\begin{aligned} H_0:\ \theta &= 1 \\ \text{contre} & \\ H_1:\ \theta &\neq 1 \end{aligned}$$

On est ainsi ramené au cas (2) du paragraphe précédent qui donne comme région de rejet

$$W=(f\leq F_{\alpha_1} \text{ ou } f\geq F_{1-\alpha_2})$$

où F_α représente le quantile d'ordre α et $\alpha=\alpha_1+\alpha_2$ de la loi de Fisher $F(n-1,m-1)$.

• **Test de l'égalité des espérances, variances connues.**

On suppose maintenant les variances connues. On veut tester les hypothèses suivantes, au niveau α:

$$\begin{aligned} H_0:\ m_X &= m_Y \\ \text{contre} & \\ H_1:\ m_X &\neq m_Y \end{aligned}$$

La v.a. $\frac{\bar{X}_n-\bar{Y}_m-(m_X-m_Y)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}$ suit une loi normale $N(0,1)$.

Le problème de test au niveau α peut se réécrire, en considérant la v.a. $\bar{Z}_n=\bar{X}_n-\bar{Y}_m$ et $\theta=m_X-m_Y$ de la manière suivante:

$$\begin{aligned} H_0:\ \theta &= 0 \\ \text{contre} & \\ H_1:\ \theta &\neq 0 \end{aligned}$$

On est ramené au cas (2) du paragraphe précédent qui donne comme région de rejet

$$W=\left(\bar{z}_n\leq -u_{1-\alpha/2}\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}} \text{ ou } \bar{z}_n\geq u_{1-\alpha/2}\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}\right)$$

où $u_{1-\alpha/2}$ représente le quantile d'ordre $1-\alpha/2$ de la loi $N(0,1)$.

3.3.3 Test de l'égalité des espérances, variances inconnues.

Pour pouvoir construire ce test, il faut supposer que les variances sont égales pour obtenir la loi de la statistique de test. Cette condition revient, à suppposer que l'on a, au préalable, effectué le test d'égalité des variances et que l'on a retenu l'hypothèse $H_0 : \ \sigma_X^2 = \sigma_Y^2$.

Notons $\sigma^2 = \sigma_X^2 = \sigma_Y^2$ la valeur commune et inconnue des variances.

On peut montrer qu'un estimateur sans biais de σ^2 meilleur que $\hat{S}_X^2$ et $\hat{S}_Y^2$ est donné par:

$$S^2 \ = \ \frac{(n-1)\hat{S}_X^2 + (m-1)\hat{S}_Y^2}{n+m-2}$$

La v.a. $(n+m-2)\frac{S^2}{\sigma^2}$ suit une loi du $\chi^2(n+m-2)$ et la v.a. $\frac{\bar{X}_n - \bar{Y}_m - (m_X - m_Y)}{S\sqrt{\frac{1}{n}+\frac{1}{m}}}$ suit une loi de Student $St(n+m-2)$.

Le problème de test au niveau α peut se réécrire, en considérant la v.a. $\bar{Z}_n = \bar{X}_n - \bar{Y}_m$ et $\theta = m_X - m_Y$ de la manière suivante:

$$\begin{array}{c} H_0 : \ \theta \ = \ 0 \\ \text{contre} \\ H_1 : \ \theta \ \neq \ 0 \end{array}$$

On est ramené au cas (2) du paragraphe précédent qui donne comme région de rejet

$$W = \left(\bar{z}_n \leq -t_{1-\alpha/2} \ s \ \sqrt{\frac{1}{n} + \frac{1}{m}} \text{ ou } \bar{z}_n \geq t_{1-\alpha/2} \ s \ \sqrt{\frac{1}{n} + \frac{1}{m}} \right)$$

où $t_{1-\alpha/2}$ représente le quantile d'ordre $1-\alpha/2$ de la loi $St(n+m-2)$.

3.3.4 Comparaison de deux proportions pour des échantillons indépendants de grande taille

On considère deux variables X et Y. On tire un échantillon de taille n de X et un échantillon de taille m de Y. On suppose que les deux échantillons sont indépendants entre eux et que $n, m \geq 30$. On note X_i (resp. Y_i) la v.a. qui vaut 1 si le $i^{\text{ème}}$ individu interrogé dans le premier échantillon (resp. dans le deuxième échantillon) présente une certaine caractéristique, que l'on souhaite étudiée, avec la probabilité p_X, (resp. avec la probabilité p_Y) et 0 sinon, avec la probabilité $1 - p_X$ (resp. avec la probabilité $1 - p_Y$). On peut voir que la v.a. X_i (resp. Y_i) suit une loi de Bernoulli $Be(p_X)$ (resp. $Be(p_Y)$). La v.a. $\sum_{i=1}^{n} X_i$ (resp. $\sum_{i=1}^{m} Y_i$), représentant le nombre d'individus dans le premier échantillon (resp. dans le deuxième échantillon) ayant la caractéristique, suit une loi binomiale $B(n, p_X)$ (resp. $B(m, p_Y)$). On souhaite tester si les proportions p_X et p_Y sont égales.

Le problème de test au niveau α peut donc s'écrire:

$$\begin{aligned} H_0 : \ p_X &= p_Y \\ \text{contre} & \\ H_1 : \ p_X &\neq p_Y \end{aligned}$$

On peut réécrire ce test, en considérant la v.a. $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{m}\sum_{i=1}^{m} Y_i$ et $\theta = p_X - p_Y$ de la manière suivante:

$$\begin{aligned} H_0 : \ \theta &= 0 \\ \text{contre} & \\ H_1 : \ \theta &\neq 0 \end{aligned}$$

On est ramené au cas (2) du paragraphe 3.2.1. De plus, comme $n, m \geq 30$, par le théorème de la limite centrale, la v.a. $\dfrac{\bar{Z}_n - \ (m_X - m_Y)}{\sqrt{\frac{p_X(1-p_X)}{n} + \frac{p_Y(1-p_Y)}{m}}}$ suit approximativement une loi $N(0,1)$.
On en déduit la région de rejet

$$W = \left(\frac{\bar{z}_n}{\sqrt{\frac{p_X(1-p_X)}{n} + \frac{p_Y(1-p_Y)}{m}}} \leq -t_{1-\alpha/2} \text{ ou } \frac{\bar{z}_n}{\sqrt{\frac{p_X(1-p_X)}{n} + \frac{p_Y(1-p_Y)}{m}}} \geq t_{1-\alpha/2} \right)$$

où $t_{1-\alpha/2}$ représente le quantile d'ordre $1 - \alpha/2$ de la loi $N(0,1)$.

Sous $H_0 : p_X = p_Y = p$, on estime la valeur commune p par $\hat{P} = \dfrac{\sum_{i=1}^{n} X_i + \sum_{i=1}^{m} Y_i}{n+m}$ et on remplace dans la région de rejet W, p_X et p_Y par $\hat{p}$. On obtient finalement comme région de rejet:

$$W = (\bar{z}_n \leq -t_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n} + \frac{1}{m})} \text{ ou } \bar{z}_n \geq t_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n} + \frac{1}{m})})$$

3.3.5 Cas apparié

On dit que deux échantillons de taille n sont appariés si ils sont constitués de deux observations de la même variable sur les mêmes individus. Par exemple, pour vérifier l'efficacité d'un médicament, on fait des mesures sur un échantillon d'individus avant et après traitement. On dispose donc de deux échantillons de taille n notés $(X_1, X_2, ..., X_n)$ et $(Y_1, Y_2, ..., Y_n)$ La v.a. X, représentant la première mesure, est supposée de loi $N(m_X, \sigma_X^2)$ et la v.a. Y, représentant la deuxième mesure, est supposée de loi $N(m_Y, \sigma_Y^2)$. Les (X_i) sont donc indépendantes entre eux ainsi que les (Y_i). Mais X_i et Y_i ne sont pas indépendantes. On suppose, que le vecteur $(X_i, Y_i)_{i=1,...,n}$ est gaussien.
Pour tout $i = 1, ..., n$, on pose $Z_i = X_i - Y_i$. Les Z_i sont indépendantes entre elles du fait que (X_i, Y_i) est un vecteur gaussien et de même loi normale $N(m_X - m_Y, \sigma_X^2 + \sigma_Y^2)$. L' hypothèse d'indépendance des Z_i est indispensable pour construire le test suivant. On veut tester si il y a eu un changement en moyenne entre X et Y, ce qui revient à tester $m_Z = m_X - m_Y = 0$ à partir de l'échantillon différence $(Z_1, Z_2, ..., Z_n)$.

Le problème de test au niveau α peut donc s'écrire:

$$\begin{aligned} H_0 : m_Z &= 0 \\ \text{contre} & \\ H_1 : m_Z &\neq 0 \end{aligned}$$

On est ramené au cas (2) du paragraphe 3.2 qui donne comme région de rejet

$$W = \left(\bar{z}_n \leq -u_{1-\alpha/2} \frac{\hat{S}_Z}{\sqrt{n}} \text{ ou } \bar{z}_n \geq u_{1-\alpha/2} \frac{\hat{S}_Z}{\sqrt{n}} \right)$$

où $u_{1-\alpha/2}$ représente le quantile d'ordre $1-\alpha/2$ de la loi $N(0,1)$, $\bar{z}_n$ la réalisation de la moyenne empirique et $\hat{S}_Z^2$ la réalisation de la variance empirique sans biais de l'échantillon $(Z_1, Z_2, ..., Z_n)$.

3.4 Remarques

3.4.1 Lien entre tests d'hypothèses bilatérales et intervalles de confiance

Reprenons l'exemple du paragraphe 3.2.1 pour expliquer ce lien. Considérons un $n-$échantillon $(X_1, ..., X_n)$ de loi normale $N(m, \sigma^2)$, σ connu. On veut tester les hypothèses suivantes au niveau α:

$$\begin{aligned} H_0 : m &= m_0 \\ \text{contre} & \\ H_1 : m &\neq m_0 \end{aligned}$$

La région de rejet est donnée par: $W = (x \in \Omega : \bar{x}_n \leq m_0 - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$ ou $\bar{x}_n \geq m_0 + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})$ où $u_{1-\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de la loi $N(0,1)$.

On rejette donc au risque α, l'hypothèse $H_0 : m = m_0$ si $(\bar{x}_n \leq m_0 - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$ ou $\bar{x}_n \geq m_0 + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})$, ce qui est équivalent à la condition $(m_0 \leq \bar{x}_n - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$ ou $m_0 \geq \bar{x}_n + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})$.
L'intervalle $[\bar{x}_n - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x}_n + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}]$ est la fourchette d'estimation au niveau de confiance $1-\alpha$ de l'espérance m d'une loi normale $N(m, \sigma^2)$, σ connu.
Il existe donc un lien entre test d'hypothèses bilatérales et intervalle de confiance. Ce lien est facile à comprendre. La signification de l'intervalle de confiance est qu'on a $1-\alpha\%$ de chances de dire vrai lorsqu'on affirme que m appartient à cet intervalle. Si m_0 n'appartient pas à cet intervalle, il est peu probable que $m = m_0$. On a même $1-\alpha\%$ de chances de dire vrai lorqu'on affirme que $m \neq m_0$. Ce qui permet de conclure le test.

3.4.2 Degré de signification

On a vu précédemment le rôle fondamental du niveau de test α. Pour expliquer la notion de degré de signification , plaçons-nous dans le cas du test sur l'espérance d'une loi normale normale $N(m, \sigma^2)$, σ connu. On veut tester, au niveau α, les hypothèses suivantes:

$$H_0 : m = m_0$$
$$\text{contre}$$
$$H_1 : m = m_1$$

$(m_0 > m_1)$

On a vu, au paragraphe 3.1 la forme de la région de rejet:

$$W = (x \in \Omega : \bar{x}_n \leq K)$$

La constante K est calculée grâce au niveau $\alpha = P_{H_0}(\bar{X}_n \leq K)$. Si l'observation $\bar{x}_n$ est inférieure ou égale à K, on rejette H_0, sinon on ne rejette pas H_0. On voit que plus α est petit, plus la région de rejet diminue et on a donc moins tendance à rejeter H_0. Ainsi, il existe α_1 pour lequel on ne rejette pas H_0 et un $\alpha_2 > \alpha_1$ pour lequel on rejette H_0. Cette remarque met en évidence l'existence d'un niveau de test critique, appelé degré de signification, compris entre α_1 et α_2, à partir duquel on rejette toujours H_0. Comme c'est l'observation $\bar{x}_n$ qui permet de conclure au rejet ou non-rejet de H_0, il suffit de prendre comme seuil K, cette observation.

Définition: Le degré de signification est défini par $P_{H_0}(\bar{X}_n \leq \bar{x}_n)$.

On rejette H_0 au niveau α dés que α est supérieur ou égal au degré de signification.
Ce degré de signification a une grande importance car elle permet de savoir si il est raisonnable de rejeter ou non H_0, plutôt que de fixer α à l'avance et, de plus, c'est cette valeur qui est donnée par la majorité des logiciels de statistiques (sous le nom de p-value). Si le degré de signification est petit, il faut mieux rejeter H_0, sinon on ne rejettera pas H_0.

Exemple numérique: Reprenons l'exemple du paragraphe 3.1, cas 1, la région de rejet est de la forme $W = (x \in \Omega, \bar{x}_n \leq K)$. Donc Le degré de signification est donné par $P_{80}(\bar{X}_n \leq 79,1) = P_{80}(10\frac{\bar{X}_n - 80}{7} \leq 10\frac{79,1-80}{7}) = 0,344$. On rejette donc H_0 dès que $\alpha \geq 0,344$. Il est donc raisonnable de ne pas rejeter H_0.

4 Quelques tests non paramétriques

Un test non paramétrique est un test qui porte sur des échantillons de variables aléatoires dont la loi de probabilité est inconnue. Il en existe beaucoup. Nous nous limiterons ici aux tests les plus usuels.

4.1 Tests d'ajustement

4.1.1 Tests du χ^2

Après étude d'un échantillon prélevé dans une population, on est amené à faire le choix d'un modèle théorique. La nature de la distribution est, selon les cas, supposée normale ou binomiale,... Nous avons fait des hypothèses de ce type dans le paragraphe portant sur les tests paramétriques. Il est nécessaire de tester la légitimité de ce choix. Le principe du test du χ^2 consiste à construire un échantillon "idéal" à partir du modèle théorique de même taille que l'échantillon prélevé et comparer les deux en s'assurant que la valeur de l'écart obtenu reste dans les limites compatibles avec les fluctuations d'échantillonnage.

• **Test d'ajustement à une loi donnée complétement spécifiée**
On est ici dans un cadre "non paramétrique" où Θ est l'ensemble de toutes les probabilités sur un espace donné Ω fini ou dénombrable, ou $\Omega = \mathbb{R}^d$. Plus précisément, on considère un n-échantillon $(X_1, ..., X_n)$ de loi P sur Ω. Cette loi est décrite par m modalités.

On veut tester au niveau α:

$$H_0 : \ P = \ P_0, \text{ loi donnée}$$
$$\text{contre}$$
$$H_1 : \ P \neq P_0$$

On suppose, qu'on possède le nombre de réalisations n_i de la v.a. N_i, $i = 1, ..., m$ des m modalités de X, au cours de $n = \sum_{i=1}^{m} n_i$ expériences de même nature et indépendantes. C'est-à-dire qu'on posséde les effectifs observés d'un échantillon de n individus dont l'espace d'observations est divisé en m catégories.

Notons $p_1, ..., p_m$ les probabilités de chaque éventualité de 1 à m, calculées à partir de la loi donnée P_0, parfaitement spécifé et $\sum_{i=1}^{m} p_i = 1$.
Le problème est de décider si les observations $n_1, n_2, .., n_m$ sont compatibles avec l'hypothèse H_0 que les probabilités sont issues de $p_1, p_2, ..., p_m$, c'est-à-dire de vérifier si les n_i sont 'significativement' proches de np_i, au sens de la distance du khi-deux définie ci-après.

On appelle *distance du khi-deux* entre la loi théorique et la loi empirique observée, la statistique

$$T_n = \sum_{i=1}^{m} \frac{(N_i - np_i)^2}{np_i}$$

Théorème de Pearson: Sous l'hypothèse nulle, la v.a. T_n suit approximativement une loi $\chi^2(m-1)$ dès que $n \geq 30$ et $np_i \geq 5, \ i = 1, ..., m$.

Remarque.
La condition restrictive $np_i \geq 5$ a une signification intuitive: si np_i est faible, la contribution de la modalité i à la valeur de T_n peut être grande et fausser le résultat. Si dans une classe, cette condition n'est pas vérifiée, il faut la regrouper avec la classe suivante ou précédente, suivant le contexte, pour pouvoir appliquer le théorème.

Illustrons sur un exemple simple ce test.

Exemple.
On lance un dé 450 fois et on enregistre le nombre de fois où chacune des faces apparaît :

face	1	2	3	4	5	6
n_i	62	50	76	68	111	83

Peut-on, au niveau $\alpha = 5\%$, rejeter l'hypothèse que le dé est bien équilibré ?

Cela revient à tester

$$H_0 : \ P = P_0 = \ \{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \frac{1}{6}, \frac{1}{6}\}$$
$$\text{contre}$$
$$H_1 : \ P \neq P_0$$

On construit le tableau des effectifs théoriques en calculant np_i avec $p_i = 1/6, \ i = 1, ..., 6$:

face	1	2	3	4	5	6
np_i	75	75	75	75	75	75

Les conditions d'application du test sont vérifiées: $n \geq 30$ et $np_i \geq 5$. Sous H_0, la v.a. T_n suit une loi de $\chi^2(5)$.

La région de rejet est donnée par

$$W = (x \in \Omega, t_n \geq K) \tag{1}$$

où $K = 11,07$ est lue dans une table de loi de $\chi^2(5)$.

On calcule la réalisation de T_n sur l'échantillon et on obtient la valeur $t_n = \frac{(62-75)^2}{75} + \frac{(50-75)^2}{75} + \frac{(76-75)^2}{75} + \frac{(68-75)^2}{75} + \frac{(111-75)^2}{75} + \frac{(83-75)^2}{75} = 18,134$.
Comme $t_n \geq 11,07$, on est dans la région de rejet. On rejette donc au risque 0.05, l'hypothèse que le dé est équilibré.

• Test d'ajustement à une loi dont certains paramètres ont été estimés

Dans le test précédent, la loi de référence est parfaitement connue. En pratique, tous les paramètres de la loi théorique ne sont pas connus, il faut en estimer un certain nombre. Notons $\hat{p}_i$ l'estimation de p_i théorique. En reprenant les notations précédentes, on appelle *distance du khi-deux* entre la loi théorique, dont les paramètres inconnus ont étés estimés, et la loi empirique observée, la statistique

$$T_n = \sum_{i=1}^{m} \frac{(N_i - n\hat{p}_i)^2}{n\hat{p}_i}$$

La loi de T_n est donnée par le théorème suivant.

Théorème. Supposons que le nombre de paramètres à estimer soit égal à k. Sous l'hypothèse nulle, la v.a. T_n suit approximativement une loi $\chi^2(m-k-1)$ dès que $n \geq 30$ et $np_i \geq 5,\ i = 1, ..., m$.

Donnons un exemple simple de ce test.

Exemple.

On étudie l'arrivée de clients à un guichet, pendant la durée de temps unitaire d'une minute. On fait 1000 mesures. Peut-on supposer que le nombre d'arrivées par minute suit une loi de Poisson?

Soit X la v.a. comptant le nombre d'arrivées par minute. On veut tester, au niveau 5%:

$$\begin{aligned} H_0 : X \text{ suit une loi de Poisson de paramètre } \lambda \\ \text{contre} \\ H_1 : X \text{ ne suit pas une loi de Poisson de paramètre } \lambda \end{aligned}$$

où le paramètre $\lambda > 0$ est inconnu.

La répartition observée des arrivées est la suivante:

X	0	1	2	3	4	5
effectif n_i	676	262	51	8	3	0

On peut montrer qu'un bon estimateur de λ est la moyenne empirique $\bar{X}_n$. Sur l'échantillon observé, l'estimation de λ vaudra donc $\hat{\lambda} = \frac{1}{1000}(0\times676+1\times262+2\times51+3\times8+4\times3+5\times0) = 0.4$.

On construit le tableau des effectifs théoriques en calculant la probabilité de la loi $P(0,4)$ par $P(X=x) = e^{-0,4}\frac{0,4^x}{x!}$:

X	0	1	2	3	4	5
p_i	0.6703	0.2681	0.0536	0.0072	0.0007	0.0001
effectif np_i	670.3	268.1	53.6	7.2	0.7	0.1

Pour que les conditions d'application du test soient vérifiées, c'est-à-dire: $n \geq 30$ et $np_i \geq 5$, il faut regrouper les trois dernières classes. On ajoute donc les effectifs de ces trois classes.

On considère donc le tableau suivant:

X	0	1	2	3,4,5
effectif n_i observé	676	262	51	11
effectif $n\hat{p}_i$ théorique estimé	670.3	268.1	53.6	8

La v.a. X a finalement 4 modalités et on a estimé un paramètre, on sait donc,en appliquant le théorème précédent, que sous H_0, la v.a. T_n suit une loi de $\chi^2(4-1-1)$.

La région de rejet est donnée par $W = (x \in \Omega, t_n \geq K)$ où $K = 5.99$ est lue dans une table de loi de $\chi^2(2)$.

On calcule la valeur de T_n sur l'échantillon et on obtient $t_n = 1.44$. Comme $t_n < 5.99$, on n'est pas dans la région de rejet et donc, on accepte au niveau 0.05 l'hypothèse de l'ajustement par une loi $P(0.4)$.

- **Autre utilisation: test d'indépendance**

Le test du χ^2 est très souvent employé dans le problème suivant: on considère une population dont les éléments possèdent deux caractères X et Y. On se demande si ces deux caractères peuvent être considérés comme indépendants.

Evidemment, ce test ne rentre pas dans le cadre des tests d'ajustement mais l'importance de son utilisation dans la pratique justifie pleinement sa place dans cet article. De plus, ce test permet de donner une justification théorique du critère d'indépendance.

On considère un échantillon de taille n sur lequel on étudie deux caractères statistiques X et Y. On suppose que la variable X (resp. Y) peut prendre k modalités (resp. l modalités), notées $E_1, ..., E_k$ (resp. $F_1, ..., F_l$). On note n_{ij} le nombre d'observations possédant le caractère X avec la modalité E_i et le caractère Y avec la modalité avec F_j. On peut représenter ces données dans un tableau de contingence:

X/Y	F_1	...	F_j	...	F_l
E_1	n_{11}	...	n_{1j}	...	n_{1l}
...	...	...	...	...	...
E_i	n_{i1}	...	n_{ij}	...	n_{il}
...	...	...	...	...	...
E_k	n_{k1}	...	n_{kj}	...	n_{kl}

Notons $p_{i.} = P[X \in E_i]$, $p_{.j} = P[Y \in F_j]$ et $p_{ij} = P[X \in E_i, Y \in F_j]$.

L'indépendance de X et Y s'exprime par la propriété $p_{ij} = p_{i.}\, p_{.j}, \forall i = 1, ..., k,\ j = 1, ..., l$.

Les probabilités $p_{i.}$ et $p_{.j}$ étant inconnues, on peut les estimer sur l'échantillon par

$$\hat{p}_{i.} = \frac{n_{i.}}{n} \text{ et } \hat{p}_{.j} = \frac{n_{.j}}{n} \text{ respectivement}$$

avec $n_{i.} = \sum_{j=1}^{l} n_{ij}$, $n_{.j} = \sum_{i=1}^{k} n_{ij}$ et $n = \sum_{j=1}^{l}\sum_{i=1}^{k} n_{ij}$ où $n_{i.}$ (resp. $n_{.j}$) représente le nombre d'individus ayant la modalité E_i (resp. la modalité F_j).
On note N_{ij} la v.a. qui, à un échantillon de taille n, associe le nombre n_{ij}, $N_{i.}$ la v.a. qui, à un échantillon de taille n, associe le nombre $n_{i.}$ et $N_{.j}$ la v.a. qui, à un échantillon de taille n, associe le nombre $n_{.j}$.

On veut tester au niveau α:

$$\begin{aligned} H_0 : X \text{ et } Y \text{ sont indépendantes} &\Longleftrightarrow P_{X,Y} = P_X \otimes P_Y \\ &\quad \text{contre} \\ H_1 : X \text{ et } Y \text{ ne sont pas indépendantes} &\Longleftrightarrow P_{X,Y} \neq P_X \otimes P_Y \end{aligned}$$

Le problème est de décider si les effectifs observés n_{ij} sont compatibles avec l'hypothèse H_0 sous laquelle les variables X et Y sont indépendantes, c'est-à-dire que chaque effectif théorique np_{ij} est égal à $np_{i.}\, p_{.j}$. En remplaçant les $p_{i.}$ et $p_{.j}$ par leurs estimateurs, cette condition d'indépendance est donnée par $np_{ij} = \frac{n_{i.}\, n_{.j}}{n}$. Il faut donc vérifier si les n_{ij} sont "significativement" proches de $\frac{n_{i.}\, n_{.j}}{n}$, au sens de la distance au khi-deux définie ci-dessous.

On appelle *distance du khi-deux* entre le tableau observé et le tableau théorique sous H_0 , la statistique

$$T_n = \sum_{i=1}^{k}\sum_{j=1}^{l} \frac{(N_{ij} - \frac{n_{i.}n_{.j}}{n})^2}{\frac{n_{i.}n_{.j}}{n}}$$

Théorème: Sous l'hypothèse nulle, la v.a. T_n suit approximativement une loi $\chi^2((k-1)(l-1))$ dès que $n \geq 30$ et que les effectifs théoriques sont supérieurs ou égaux à 5.

Intuitivement, si X et Y sont indépendants, T_n doit être "petit". Cela conduit à choisir une région de rejet, au niveau α de la forme

$$W = (x \in \Omega, t_n \geq K)$$

où K représente le quantile d'ordre $1-\alpha$ de la loi du $\chi^2((k-1)(l-1))$.

Donnons un exemple simple de ce test.
Exemple.
Un enquête a été réalisée auprès d'un échantillon de 250 personnes au sujet de l'abaissement à 16 ans du droit de vote. Les réponses ont été classées suivant le niveau d'instruction des individus:

X/Y	pour	contre	$n_{i.}$
BEP	10	15	25
BAC	21	84	105
Universitaire	20	100	120
$n_{.j}$	51	199	250

Peut-on au niveau $\alpha = 1\%$, accepter l'hypothèse d'une liaison entre l'opinion d'une personne et son niveau d'instruction?

On construit le tableau des effectifs théoriques:

X/Y	pour	contre
BEP	5.1	19.9
BAC	21.42	83.53
Universitaire	24.48	95.52

Par exemple, le nombre 5.1 a été obtenu en calculant $\frac{n_{1.}n_{.1}}{n} = \frac{25\times 51}{250}$.

Les conditions d'application du théorème sont vérifiées. Les v.a. X et Y ont respectivement 2 et 3 modalités. En appliquant le théorème précédent, sous H_0, la v.a. T_n suit une loi $\chi^2((2-1)(3-1))$.

La région de rejet est donnée par

$$W = (x \in \Omega, t_n \geq K)$$

où $K = 9.21$ est lue dans une table de loi $\chi^2(2)$.

On calcule la valeur de T_n sur l'échantillon et on obtient $t_n = 7$. Comme $t_n < 9.21$ n'est pas dans la région de rejet, on ne rejette pas au niveau 0.01 l'hypothèse d'une liaison entre X et Y.

4.1.2 Tests basés sur la fonction de répartition empirique

Il existe évidemment beaucoup d'autres tests d'ajustement que nous n'exposerons pas ici, du fait de la spécificité de chacun. L'important est de savoir qu'il en existe d'autres et de pouvoir y accéder en cas de besoin.

Certains sont basés sur la comparaison de la fonction de répartition théorique et la fonction de répartition observée. Nous allons étudier, dans ce paragraphe, les deux les plus utilisés dans la pratique: le test de Kolmogorov-Smirnov et le test Cramer-von Mises.

Il existe un grand nombre de tests spécifiques à l 'ajustement à la loi normale, du fait du rôle important de cette loi dans les procédures classiques de la statistique, comme on a pu le voir dans les chapitres précédents . C'est pour cette raison qu'un grand nombre de tests de normalité se sont développés à coté des tests généraux précédemment étudiés. On peut en citer quelques uns: droite de Henry, test de Shapiro-Wilk,....

• **Fonction de répartition et hypothèses du test**

On considère un n-échantillon $(X_1, ..., X_n)$ de loi P sur Ω. On définit la fonction de répartition de X par $F(x) = P(X \leq x)$. On désire ajuster la loi P inconnue à une loi spécifiée P_0, de fonction de répartition F_0. Ce qui revient à tester au niveau α

$$H_0 : \; F = \; F_0,$$

contre

$$H_1 : F \neq F_0$$

Pour construire ce test, on considère la fonction de répartition empirique définie, pour $x \in \mathbb{R}$, par $f_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{]-\infty,x]}(x_i)$. On peut facilement vérifier que $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{]-\infty,x]}(X_i)$ est un estimateur sans biais de $F(x)$ et qu'il converge presque sûrement vers $F(x)$. On va donc construire les statistiques de test grâce à cet estimateur. On rejettera H_0 si les fonctions F_n et F_0 sont "significativement" éloignées, dans un sens précisé ci-après. Il y a plusieurs façons de mesurer lécart entre F et F_0, qui donnent lieu à différentes statistiques de tests:

• statistique de Kolmogorov-Smirnov: $K_n = \sqrt{n} \; sup_{x \in \mathbf{R}} |F_n(x) - F_0(x)|$

• statistique de Cramer-von Mises: $CVM_n = n \int_{-\infty}^{+\infty} (F_n(x) - F_0(x))^2 \; dF_0(x)$

On peut montrer que sous H_0, ces statistiques de test convergent en loi vers des lois de probabilité indépendantes de F_0, d'expressions complexes mais tabulées.

Intuitivement, si $F = F_0$, K_n et CVM_n doivent être "petites". Cela conduit à choisir une région de rejet, au niveau α de la forme

$$W = (x \in \Omega, k_n \geq K) \text{ ou } W = (x \in \Omega, cvm_n \geq L)$$

où K et L représentent les quantiles d'ordre $1-\alpha$ lus dans les tables associées aux lois limites.

4.2 Tests non paramétriques de comparaison de lois

On va considérer ici des tests de comparaison de lois, dans le cas où les lois à comparer sont inconnues. Nous allons exposer trois tests usuels de ce type dans ce paragraphe. On pourra trouver une littérature abondante concernant ce sujet dans le livre de Lecoutre-Tassi[5]. Ces tests sont assez faciles à mettre en oeuvre, d'où leur popularité mais ils existent peu de résultats concernant leur puissance.

On suppose qu'on a deux échantillons indépendants entre eux $(X_1, ..., X_m)$ et $(Y_1, ..., Y_n)$, de fonctions de répartition respectives F_X et F_Y. On désire savoir si les v.a. X et Y ont même loi, sans faire d'hypothèses sur leur loi . On veut donc tester au niveau α:

$$H_0 : \; F_X \; = F_Y$$
$$\text{contre}$$
$$H_1 : \; F_X \neq F_Y$$

4.2.1 Test de Kolmogorov-Smirnov

Cest une généralisation du test de Kolmogorov-Smirnov pour un échantillon, étudié dans le paragraphe précédent. Pour construire ce test, on considère les fonctions de répartition empiriques de X et Y, notées respectivement $F_m^X(x)$ et $F_n^Y(x)$. On suppose que les deux échantillons sont de grande taille . Si les deux échantillons proviennent de la même loi, ils ont la même fonction de répartition et leurs fonctions de répartition empiriques doivent être proches. On va donc construire la statistique de test grâce à ces estimateurs. On rejettera H_0 si les fonctions F_m^X et F_n^Y sont "significativement" éloignées, dans le sens suivant:

- statistique de Kolmogorov-Smirnov: $K_{m,n} = \sqrt{\frac{mn}{m+n}} \; sup_{x\in \boldsymbol{R}}|F_m^X(x) - F_n^Y(x)|$

On peut montrer que sous H_0, cette statistique de test converge en loi vers la même loi de probabilité que la loi limite obtenue pour K_n au paragraphe précédent.

Intuitivement, si $F_X = F_Y$, K_n doit être "petite". Cela conduit à choisir une région de rejet, au niveau α de la forme

$$W = (x \in \Omega, k_n \geq K)$$

où K représente le quantile d'ordre $1 - \alpha$ lu dans la table associée à la loi limite.

4.2.2 Test de Mann-Whitney

Le principe de ce test est le suivant:
si X et Y snt indépendantes et de même loi alors $P(X \leq Y) = P(Y \leq X) = \frac{1}{2}$. Comme les échantillons sont indépendants, on a alors, $\forall i, j$, $P(X_i \leq Y_j) = \frac{1}{2}$.
Notons Z le nombre de couples (i, j) tels que $X_i \leq Y_j$ parmi les nm couples possibles. Donc, la réalisation de Z doit être proche de $\frac{nm}{2}$,sous l'hypothèse $H_0 : F_X = F_Y$.

On peut montrer que, sous H_0, la v.a. Z suit une certaine loi pour laquelle il existe une table statistique. Pour des grands échantillons, on utilise une approximation de la loi de Z par la loi normale: sous H_0, la v.a. $\dfrac{Z - \frac{nm}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}}$ suit approximativement une loi $N(0, 1)$.

Intuitivement, si $F_X = F_Y$, la différence $|Z - \frac{nm}{2}|$ doit être "petite". Cela conduit à choisir une région de rejet, au niveau α de la forme

$$W = \left(x \in \Omega, \; \left| \frac{z - \frac{nm}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}} \right| \geq u_{1-\alpha/2} \right)$$

où $u_{1-\alpha/2}$ représente le quantile d'ordre $1-\alpha/2$ lu dans la table $N(0,1)$ dans le cas des grands échantillons.

4.2.3 Test de Wilcoxon

Une autre façon de procéder est de ranger les $n+m$ éléments $(x_1, ..., x_m, y_1, ..., y_n)$ par ordre croissant et on note R_i le rang de x_i, $i = 1, ..., m$ parmi les $n+m$ valeurs.

On considère la v.a. $R = \sum_{i=1}^{m} R_i$.

On peut montrer que la statistique de test de Mann-Whitney Z est égale à $\frac{m(m+2n+1)}{2} - R$ et ainsi les tests de Wilcoxon et Mann-Whitney sont équivalents. Seule la procédure change.